Spatial and Spectral Joint Super-Resolution Using Convolutional Neural Network

Shaohui Mei, *Member, IEEE*, Ruituo Jiang, Xu Li, *Member, IEEE*, and Qian Du, *Fellow, IEEE*

***Abstract*— Many applications have benefited from the images with both high spatial and spectral resolution, such as mineralogy and surveillance. However, it is difficult to acquire such images due to the limitation of sensor technologies. Recently, super-resolution (SR) techniques have been proposed to improve the spatial or spectral resolution of images, e.g., improving the spatial resolution of hyperspectral images (HSIs) or improving spectral resolution of color images (reconstructing HSIs from RGB inputs). However, none of the researches attempted to improve both spatial and spectral resolution together. In this article, these two types of resolution are jointly improved using convolutional neural network (CNN). Specifically, two kinds of CNN-based SR are conducted, including a simultaneous spatial–spectral joint SR (SimSSJSR) that conducts SR in spectral and spatial domain simultaneously and a separated spatial–spectral joint SR (SepSSJSR) that considers spectral and spatial SR sequentially. In the proposed SimSSJSR, a full 3-D CNN is constructed to learn an end-to-end mapping between a low spatial-resolution mulitspectral image (LR-MSI) and the corresponding high spatial-resolution HSI (HR-HSI). In the proposed SepSSJSR, a spatial SR network and a spectral SR network are designed separately, and thus two different frameworks are proposed for SepSSJSR, namely SepSSJSR1 and SepSSJSR2, according to the order that spatial SR and spectral SR are applied. Furthermore, the least absolute deviation, instead of mean square error (MSE) in traditional SR networks, is chosen as the loss function for the proposed networks. Experimental results over simulated images from different sensors demonstrated that the proposed SepSSJSR1 is most effective to improve spatial and spectral resolution of MSIs sequentially by conducting spatial SR prior to spectral SR. In addition, validation on real Landsat images also indicates that the proposed SSJSR techniques can make full use of available MSIs for high-resolution-based analysis or applications.**

***Index Terms*— Convolutional neural network (CNN), hyperspectral image (HSI), multispectral image (MSI), spatial–spectral, super-resolution (SR).**

Manuscript received July 17, 2019; revised November 7, 2019; accepted December 30, 2019. Date of publication January 29, 2020; date of current version June 24, 2020. This work was supported in part by the National Natural Science Foundation of China under Grant 61671383, in part by the Natural Science Foundation of Shaanxi Province under Grant 2018JM6005, in part by the Fundamental Research Funds for the Central Universities under Grant 3102018AX001, and in part by the Seed Foundation of Innovation and Creation for Graduate Students in NPU under Grant ZZ2019164. *(Corresponding author: Shaohui Mei.)*

Shaohui Mei, Ruituo Jiang, and Xu Li are with the School of Electronics and Information, Northwestern Polytechnical University, Xi'an 710129, China (e-mail: meish@nwpu.edu.cn).

Qian Du is with the Department of Electrical and Computer Engineering, Mississippi State University, Starkville, MS 39762 USA.

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2020.2964288

I. Introduction

REMOTE sensing using satellite and airborne sensors is a powerful and operational tool for Earth observation. Such technology has been widely used in vegetation investigation, environment protection, precision agriculture, etc. With the development of sensor technologies, the spectral resolution of sensors increases continuously, from hundreds of nanometers in panchromatic sensing to several nanometers in hyperspectral sensing. However, due to the limitation in signal-to-noise ratio (SNR) and time constraint, there is a tradeoff between spatial and spectral resolution in remote sensing. As a consequence, panchromatic sensing produces extremely high spatial resolution yet low spectral resolution. On the contrary, hyperspectral sensing results in very high spectral resolution yet low spatial resolution. In practical applications, e.g., mineralogy, manufacturing, and surveillance, images with both high spatial resolution and high spectral resolution are beneficial. Therefore, it is very necessary to improve both the spatial and spectral resolution of remote sensing images using signal processing techniques.

Of all the satellite-based remote sensing systems, many multispectral data have been acquired, such as IKONOS, QuickBird, Geo Eye-1, EOS-Terra, World View, MODIS, SPOT, SkySat, Sentinel-2A, Landsat series, and GF series. The spatial resolution of these multispectral images (MSIs) is lower than the panchromatic and RGB images. Moreover, their spectral resolution is much lower than that of hyperspectral images (HSIs). Neither the spectral resolution nor the spatial resolution of these MSIs can meet the requirement of fine remote sensing in many applications. Therefore, in order to make full use of available multispectral data for high-resolution-based analysis or applications, spatial and spectral joint super-resolution (SR) techniques are required to enhance both spatial and spectral resolution of these multispectral remote sensing images.

Single-image spatial SR, which aims to reconstruct a high spatial resolution image only from a low spatial resolution image, can break the limitation of the inherent spatial resolution in imaging systems without any other prior or auxiliary information. The basic method of single-image spatial SR is through a nonlinear interpolator, such as bilinear and bi-cubic interpolation [1], which directly exploits the information of neighboring pixels. However, these methods often lead to edge blur and ringing effect. In the past decades, SR of color images has gained great attention and many algorithms have been developed, such as the iterative back projection (IBP) for adaptive image enlargement [2], [3] and sparse

0196-2892 © 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: Zhejiang University. Downloaded on August 20,2020 at 13:09:20 UTC from IEEE Xplore. Restrictions apply.

representation-based methods [4], [5]. Recently, deep learning-based methods have been applied to the SR of color images and demonstrated great superiority. SR convolutional neural network (SRCNN) [6] was a pioneering work for deep learning in SR reconstruction, which first used bicubic interpolation to enlarge the low-resolution image to target size and then fit the nonlinear mapping through a three-layer convolutional network. CNN-based SR framework has been proven to be based on the sparse coding pipeline, according to which sparse coding-based network (SCN) [7] replaced the mapping layer by a set of sparse coding subnetworks. Fast SRCNN [8] improved SRCNN by replacing bicubic interpolation with a deconvolution layer to enlarge the image to the target size. A robust SR method [9] was also proposed by combining the domain expertise of the conventional sparse coding and the merits of deep learning to achieve better results. Efficient subpixel CNN [10] extracted features directly on low-resolution image by convolutional layers to generate r^2 feature maps with the same size of the original low resolution image. Then a subpixel convolutional layer was applied to these feature maps to rearrange the r^2 channels of each pixel of the feature image into a $r \times r$ area, corresponding to an $r \times r$ subblock in the high-resolution image, such that the size of $H \times W \times r^2$ feature image was rearranged into a high-resolution image with size of $rH \times rW \times 1$. The dense convolutional network (DenseNet) [11], which concatenated the features of all layers by feeding the features of each layer to all subsequent layers in a dense block, has also been used for SR problem [12]. All of these CNNs for the SR of color images can be directly applied to HSIs in a band-by-band or 3-band-group manner. However, spectral distortion is often induced in such extensions since spectral correlation in contiguous bands is ignored. Therefore, a 3-D full CNN (3D-FCNN) [13] was constructed for spatial SR of HSIs by using 3-D convolution to explore the information in both spatial context of neighboring pixels and spectral correlation of neighboring bands.

The high spectral-resolution images, i.e., HSIs, are beneficial to numerous applications [14]. However, commercially available hyperspectral imaging cameras are usually expensive. Therefore, spectral SR has also gained lots of attention in the past decade. For example, Nguyen *et al.* [15] applied a radial basis function (RBF) network to model the mapping between white-balancing RGB values and illumination-free reflectance spectra. Robles-Kelly [16] learned a set of prototypes from a large number of spectral reflectance distributions and their corresponding RGBs using a constrained sparse coding approach, which is used for spectral reconstruction and illuminance recovery. Arad and Ben-Shahar [17] learned a couple of dictionaries between hyperspectral signatures and their RGB projections from a large database of HSIs of natural scenes using K-SVD algorithm [18]. This upsampling prior information is then used to reconstruct the spectrum with orthogonal matching pursuit (OMP) [19]. Different spectral sensitivities of different cameras were also used to reconstruct the hyperspectral signal at different scene points [20]. Akhtar and Mian [21] proposed to recover spectral details from RGB images by modeling natural spectra under Gaussian Processes and using clusters in the data processing. In these algorithms, the camera spectral response function is assumed to be known. Recently, CNN is adopted to directly learn the mapping between RGB images and their corresponding HSIs, according to which the spectral SR is conducted without the spectral response function [22]. For example, Wu *et al.* [23] improved the sparse coding shallow method in [17] by introducing an "A+"-based shallow method [24], which can provide comparable performance to that in [22]. Xiong *et al.* [25] proposed a unified deep learning framework for HSI recovery from spectrally undersampled projections, namely HSCNN. Such HSCNN was then extended into HSCNN+ [26] by removing the handcrafted upsampling and utilizing residual blocks to provide a more accurate solution. Alvarez-Gila *et al.* [27] adopted an adversarial framework-based generative model to learn the end-to-end mapping between pairs of input RGB images and their hyperspectral counterparts. Yan *et al.* [28] designed a multiscale symmetrically cascaded downsampling-upsampling CNN architecture to jointly encode the local and nonlocal image information for spectral reconstruction.

Though the SR problems in the spatial or spectral domain have been widely explored in the past decades, none of them considered SR in these two domains together. Therefore, in this article, the SR in both spectral and spatial domains is jointly considered using CNN-based framework. Specifically, a high spatial-resolution HSI (HR-HSI) is reconstructed from a low spatial-resolution MSI (LR-MSI) based on deep learning techniques. Two different kinds of SR methods are proposed and compared in this article, including a simultaneous spatial–spectral joint SR (SimSSJSR) that conducts SR in spectral and spatial domain simultaneously and a separated spatial–spectral joint SR (SepSSJSR) that considers spectral and spatial SR sequentially. In the proposed SimSSJSR, a full 3-D CNN is constructed to learn an end-to-end mapping between the LR-MSI and the HR-HSI. In the proposed SepSSJSR, the spatial SR network and the spectral SR network are designed separately, and thus two different frameworks are proposed for SepSSJSR, namely SepSSJSR1 and SepSSJSR2, according to the order that spectral SR and spatial SR are applied. In the spatial SR, subpixel convolutional layer [10] is first used for spatial upsampling instead of traditional bicubic interpolation, and then CNN is further used to fine-tune spectral distortion, while a full 3-D CNN is directly designed for spectral SR. Finally, extensive experiments over benchmark data sets from ROSIS and AVIRIS sensors are conducted to compare the three proposed CNN-based strategies for SSJSR.

In summary, the main contributions of this article can be summarized as follows.

1) For the first time, both spatial resolution and spectral resolution are jointly enhanced in one framework. Note that, by using the proposed framework, both spatial and spectral resolution of these multispectral remote sensing images can be improved together, which can make full use of available multispectral data for high-resolution-based analysis or applications.
2) Two different kinds of CNN-based frameworks (three specific CNN-based networks) are proposed and constructed for SSJSR, i.e., SimSSJSR and SepSSJSR.

Authorized licensed use limited to: Zhejiang University. Downloaded on August 20,2020 at 13:09:20 UTC from IEEE Xplore. Restrictions apply.

3) Extensive experiments over benchmark data sets from ROSIS and AVIRIS sensors are conducted to compare the three proposed CNN-based strategies for SSJSR and different parameters involved in these CNN-based SR are discussed. Moreover, experiments over both classification application and real Landsat images are carried out to illustrate the superiority of the proposed SSJSR.

The remainder of this article is organized as follows. Section II formulated the spatial and spectral SR problem and proposed two kinds of CNN-based frameworks for spatial and spectral joint SR. In Section III, extensive experiments are conducted to compare the proposed three CNN-based SR algorithms. Finally, conclusions and future research directions are present in Section IV.

II. Proposed Methods

In this section, two different frameworks are proposed for SSJSR, i.e., SimSSJSR that conducts SR in spectral and spatial domain simultaneously and SepSSJSR that considers spectral and spatial SR sequentially.

A. SimSSJSR

Generally, an LR-MSI can be viewed as a downsampled version of HR-HSI in both spectral and spatial dimensions. Let $\mathbf{X} \in \mathbb{R}^{d\times h\times w}$ represent an LR-MSI in which d is the number of spectral bands, h is the image height, and w is the image width. Its corresponding HR-HSI is denoted as $\mathbf{Y} \in \mathbb{R}^{D\times H\times W}$, where D is the number of spectral bands, H is the image height, and W is the image width. Without loss of generality, $H = c\cdot h$, $W = c\cdot w$, in which c is the spatial scale factor, and $d < D$ [1]. Therefore, SSJSR aims to find a mapping function $\mathcal{F}$ to restore a HR-HSI $\widehat{\mathbf{Y}} \in \mathbb{R}^{D\times H\times W}$

$$\widehat{\mathbf{Y}} = \mathcal{F}(\mathbf{X}) \tag{1}$$

such that $\widehat{\mathbf{Y}}$ approaches $\mathbf{Y}$ as accurate as possible.

Deep neural network (DNN) has been widely used to estimate unknown from the input, e.g., estimating a high-resolution image in SR problem or predicting class label in classification problem. Therefore, it can be directly used to find an end-to-end mapping between LR-MSI and HR-HSI. Generally, the problem of SimSSJSR using DNN can be formulated as

$$\begin{aligned} &\widehat{\mathbf{Y}} = \mathcal{F}(\mathbf{X}, \Theta) \\ &\text{s.t. } \widehat{\mathbf{Y}} \approx \mathbf{Y} \end{aligned} \tag{2}$$

where $\Theta = (\mathbf{W}, \mathbf{B})$ consists of the weights $\mathbf{W}$ and the biases $\mathbf{B}$ of the DNN.

In this article, a full 3-D CNN is first constructed to directly learn an end-to-end mapping between LR-MSI and HR-HSI, by which SR in both spectral and spatial dimensions is conducted simultaneously. The structure of the proposed SimSSJSR is shown in Fig. 1. It contains several convolutional layers for spatial–spectral feature extraction from LR-MSI and one subpixel convolutional layer for the reconstruction of HR-HSI. In order to fully explore spatial–spectral features in LR-MSI, 3-D convolution is adopted by stacking dozens of 3-D convolution kernels in one layer. When these 3-D convolution layers are connected sequentially, the 3-D convolution should be conducted with one extra fixed dimension to handle these inputs from multiple feature cubes simultaneously [29]. Therefore, 3-D convolution kernel is defined as $\mathbf{W} \in \mathscr{R}^4$ with size of $D_c \times D_x \times D_y \times D_z$, where the extra fourth dimension D_c represents the number of kernels in a convolutional layer. Suppose the input to the ith convolution layer (denoted as "Conv$_i$") is defined as $\mathbf{I}_i \in \mathscr{R}^4$ with size of $D_i \times D_x^{(\mathbf{I})_i} \times D_y^{(\mathbf{I})_i} \times D_z^{(\mathbf{I})_i}$. Without loss of generality, if the LR-MSI image cube is fed as input, $D_0 = d$. As a result, the 3-D convolution in the ith convolution layer is represented as

Fig. 1. Structure of the proposed SimSSJSR.

$$\begin{aligned} O_i^{j,x,y,z} &= \mathrm{Conv}(\mathbf{I}_i; \mathbf{W}_i) \\ &= \sum_{k=0}^{D_i-1}\sum_{p=0}^{D_x-1}\sum_{q=0}^{D_y-1}\sum_{r=0}^{D_z-1} W_i^{k,p,q,r} I_i^{k,x\cdot s_x+p,y\cdot s_y+q,z\cdot s_z+r} \\ &\quad + b_i^j \end{aligned} \tag{3}$$

where subscripts "j" indexes the convolutional kernels in ith layer, (s_x, s_y, s_z) represents the size of straddle in three dimensions, and b_i^j represents the bias for jth kernel in ith layer. As a result, the output of the ith convolution layer $\mathbf{O}_i$ is with the size $D_c \times D_x^{(\mathbf{I})_i} \times D_y^{(\mathbf{I})_i} \times D_z^{(\mathbf{I})_i}$. Except for the last convolutional layer, a parametric rectified linear unit (ReLU) layer is added after each convolutional layer, which uses the "ReLU" activation function to adaptively learn the parameters of the rectifiers [30]. Note that for the last convolutional layer, a subpixel convolutional layer f^{sp}, instead of ReLU layer, is adopted to convert the LR feature maps to a HR image $\widehat{Y}$

$$\widehat{Y} = f^{sp}(\mathbf{O}_i) = \mathrm{PS}(\mathrm{Conv}(\mathbf{O}_n; \mathbf{W}_{\mathrm{sp}})) \tag{4}$$

where $PS(\cdot)$ is a periodic shuffling operator that rearranges the elements of a $C\cdot r^2 \times D \times H \times W$ tensor to a tensor of shape $C \times D \times rH \times rW$.

The size of different convolutional layers and "ReLU" layers are listed in Table I. In the proposed SimSSJSR, a subimage cube of $1 \times d \times 64 \times 64$ is fed to the network to keep the network in a reasonable scale, where 64×64 represents spatial pixels and d is the spectral dimension depending on the sensors acquiring LR-MSI images. In the first Layer "Conv_1," D different convolution kernels with a size of $3 \times 3 \times 3$ are applied to the input to generate D feature maps of the size $d \times 64 \times 64$. After "Relu_1," a transpose layer is added to adjust

[1] In spectral dimension, we do not constrain SR is conducted with an upscale factor since spectral SR is generally restricted by the band configuration of sensors.

Authorized licensed use limited to: Zhejiang University. Downloaded on August 20,2020 at 13:09:20 UTC from IEEE Xplore. Restrictions apply.

Fig. 2. Structure of the proposed SepSSJSR1.

TABLE I
DETAILED CONFIGURATIONS OF THE PROPOSED SIMSSSR

Layer	Input Size	Kernel Size	Output size
Conv_1	$1 \times d \times 64 \times 64$	$D \times 3 \times 3 \times 3$	$D \times d \times 64 \times 64$
Relu_1	$D \times d \times 64 \times 64$	-	$D \times d \times 64 \times 64$
Transpose	$D \times d \times 64 \times 64$	-	$d \times D \times 64 \times 64$
Conv_2	$d \times D \times 64 \times 64$	$64 \times 3 \times 3 \times 3$	$64 \times D \times 64 \times 64$
...	...	...	...
Relu_n-1	$d \times D \times 64 \times 64$	-	$d \times D \times 64 \times 64$
Conv_n	$d \times D \times 64 \times 64$	$c^2 \times 3 \times 1 \times 1$	$c^2 \times D \times 64 \times 64$
Sub-pixel	$c^2 \times D \times 64 \times 64$	-	$1 \times D \times 64c \times 64c$

the dimensions of current output. Layers from "Conv_2" to "Conv_n-1" then sequentially apply 64 different convolutional kernels of size $3 \times 3 \times 3$ to generate 64 deeper feature maps of the size $D \times 64 \times 64$. In layer "Conv_n," c^2 different convolutional kernels with a size of $3 \times 1 \times 1$ are used to generate c^2 feature maps of the size $D \times 64 \times 64$. Finally, a subpixel convolutional layer is adopted for c-scale spatial SR, producing an output image $\widehat{\mathbf{Y}}$ with a size of $D \times 64c \times 64c$.

In order to train the proposed SimSSJSR, a loss function that measures the difference between $\widehat{\mathbf{Y}}$ and $\mathbf{Y}$ must be defined. Traditional mean-squared-error (MSE) measured by ℓ_2-norm is much larger in the case of outliers compared to the least absolute deviations measured by ℓ_1-norm. As a consequence, the MSE-based loss function may try to adjust the model according to these outlier values. On the contrary, ℓ_1-norm-based loss function is more robust to outliers, which is especially beneficial for training a network. In this article, the ℓ_1-norm-based loss function is adopted

$$\ell_1(\hat{\mathbf{Y}}, \mathbf{Y}) = \frac{1}{N}\sum_{n=1}^{N}\left(\frac{1}{\mathrm{HWD}}\sum_{i=1}^{H}\sum_{j=1}^{W}\sum_{k=1}^{D}|\hat{Y}_{i,j,k} - Y_{i,j,k}|\right) \quad (5)$$

where N is the batch size, indicating that the parameters in the proposed SSJSR are updated in batch-mode. The training procedure of the proposed SimSSJSR is summarized in Algorithm 1. Once the network is well-trained, it can be directly used for SR of LR-MSI in both spectral and spatial dimensions.

Algorithm 1 Training Process of SimSSJSR

Input: $N \times d \times 64 \times 64$ sub_image cube
Output: $N \times D \times 64c \times 64c$ sub_image cube
Initialize: Initializing the network with a semi orthogonal matrix, as described in [31].

1: **while** the SimSSJSR network hasn't converged yet **do**
2: computing loss according to eq. (5).
3: training the SimSSJSR_Model1(Training_data LR-MSI $\mathbf{X}$, Ground_truth HR-HSI $\mathbf{Y}$);
4: update weights $\mathbf{W}$ and biases $\mathbf{B}$ by minimizing the ℓ_1-norm-based loss function defined by (5).
5: **end while**
6: Saving parameters $\Theta = (\mathbf{W}, \mathbf{B})$ for the well-trained SimSSJSR.

B. SepSSJSR

The downsampling of HR-HSI to LR-MSI can be divided into two steps. First, a high-spatial-resolution MSI (HR-MSI) is spectrally downsampled from HR-HSI in spectral dimension, which is formulated as

$$\mathbf{P} = \mathbf{SY} \quad (6)$$

where $\mathbf{P} \in \mathbb{R}^{d\times H\times W}$ represent the HR-MSI and $\mathbf{S} \in \mathbb{R}^{D\times d}$ represents the spectral response function of the multispectral imaging sensor. Then the LR-MSI can be viewed as a spatially downsampled version of HR-MSI, which is formulated as

$$\mathbf{X} = \mathbf{down_c}(\mathbf{G} * \mathbf{P}) \quad (7)$$

where $\mathbf{G} \in \mathbb{R}^{c\times c}$ is a Gaussian filter with a kernel size c and the $\mathbf{down_c}()$ represents the downsampling operator with the factor c. Correspondingly, the SSJSR of LR-MSI can also be conducted in two sequential steps, in which spatial SR is first applied to LR-MSI to estimate HR-MSI and spectral SR is then conducted on HR-MSI to estimated HR-HSI. Such a

TABLE II
DETAILS OF THE SPATIAL SR NETWORK IN THE PROPOSED SEPSSJSR

Layer	Input Size	Kernel Size	Output size
Conv_1	$1 \times d \times 64 \times 64$	$64 \times 3 \times 3 \times 3$	$64 \times d \times 64 \times 64$
Relu_1	$64 \times d \times 64 \times 64$	-	$64 \times d \times 64 \times 64$
Conv_2	$64 \times d \times 64 \times 64$	$32 \times 3 \times 3 \times 3$	$32 \times d \times 64 \times 64$
...	...	...	...
Relu_n-1	$32 \times d \times 64 \times 64$	-	$32 \times d \times 64 \times 64$
Conv_n	$32 \times d \times 64 \times 64$	$c^2 \times 3 \times 3 \times 3$	$c^2 \times d \times 64 \times 64$
Sub-pixel	$c^2 \times d \times 64 \times 64$	-	$1 \times d \times 64c \times 64c$

Authorized licensed use limited to: Zhejiang University. Downloaded on August 20,2020 at 13:09:20 UTC from IEEE Xplore. Restrictions apply.

Fig. 3. Structure of the proposed SepSSJSR2.

TABLE III
DETAILS OF THE SPECTRAL SR NETWORK IN THE PROPOSED SEPSSJSR

Layer	Input Size	Kernel Size	Output size
Conv_1	$1 \times d \times 64c \times 64c$	$D \times 3 \times 3 \times 3$	$D \times d \times 64c \times 64c$
Relu_1	$D \times d \times 64c \times 64c$	-	$D \times d \times 64c \times 64c$
Transpose	$D \times d \times 64c \times 64c$	-	$d \times D \times 64c \times 64c$
Conv_2	$d \times D \times 64c \times 64c$	$64 \times 3 \times 3 \times 3$	$64 \times D \times 64c \times 64c$
...	...	...	...
Relu_n-1	$64 \times D \times 64c \times 64c$	-	$64 \times D \times 64c \times 64c$
Conv_n	$64 \times D \times 64c \times 64c$	$1 \times 3 \times 1 \times 1$	$1 \times D \times 64c \times 64c$

separative SR problem can be formulated as

$$\mathbf{Y} = \mathcal{F}_1(\mathbf{P}) \quad s.t. \quad \mathbf{P} = \mathcal{F}_2(\mathbf{X}) \tag{8}$$

in which $\mathcal{F}_1$ and $\mathcal{F}_2$, respectively, represent the spectral SR function of HR-MSI and spatial SR function of LR-MSI. Similar to previous SimSSJSR, these two SR problems can also be solved by DNNs. Therefore, two different CNNs are constructed for the spectral SR problem of HR-MSI and spatial SR problem of LR-MSI, and the SSJSR is proposed by sequentially connecting these two CNNs as illustrated in Fig. 2. In the proposed SepSSJSR shown in Fig. 2, the spatial SR network is adopted to learn the spatial SR function of LR-MSI

$$\widehat{\mathbf{P}} = \mathcal{F}_2(\mathbf{X}, \Theta_2) \tag{9}$$

where $\widehat{\mathbf{P}}$ represents reconstructed HR-MSI from the input of LR-MSI, and $\Theta_2 = (\mathbf{W}_2, \mathbf{B}_2)$ consists of the weights $\mathbf{W}_2$ and the biases $\mathbf{B}_2$ of the spatial SR network. In this article, the spatial SR network consists of several convolutional layers and one subpixel convolutional layers, where their parameters are listed in Table II. Layers "Conv_2" …"Conv_n-1" own the same structure, each of which has 32 different convolutional kernels with size $3 \times 3 \times 3$. Except for the last convolutional layer, every convolutional layer is followed by a ReLU layer as the activation function. In addition, padding is used to prevent shrink in the size of the image in all these convolutional layers. Similar to the SimSSJSR, the input data of the spatial SR network, i.e., the input to the SepSSJSR, is also set as $1 \times d \times 64 \times 64$ subimage cube. Padding is used to prevent shrink in the size of the image in all convolutional layers. In the Layer "Conv_1," 64 different convolutional kernels with size $3 \times 3 \times 3$ are applied on the LR-MSI to generate 64 feature maps of the size $d \times 64 \times 64$. Then layers "Conv_2" …"Conv_n-1" sequentially apply 32 different convolutional kernels of size $3 \times 3 \times 3$ on the output of the previous layer, resulting in 32 feature maps of size $d \times 64 \times 64$. Layer "Conv_n" generates c^2 feature maps of the size $d \times 64 \times 64$ by applying c^2 different convolutional kernels with size of $3 \times 3 \times 3$. Finally, a subpixel convolutional layer is applied for c-scale spatial SR to generate a HR-MSI subimage cube $\widehat{\mathbf{P}}$ with size $d \times 64c \times 64c$.

(a) (b)

Fig. 4. Pseudo color images of data sets acquired by ROSIS sensor. (a) Pavia Centre data set (Band 15, 30, and 60 are displayed as blue, green, and red, respectively). (b) Pavia University data set (Band 10, 35, and 65 are displayed as blue, green, and red, respectively).

TABLE IV
BAND SIMULATION RELATIONSHIP BETWEEN MSI OF IKONOS SENSOR AND HSI OF ROSIS SENSOR

MSI simulation bands	B	G	R	NIR
Corresponding HSI bands	3-22	19-41	50-66	80-102

The spectral SR network in the proposed SepSSJSR shown in Fig. 2 is proposed to learn spectral SR function which is formulated as

$$\widehat{\mathbf{Y}} = \mathcal{F}_1(\widehat{\mathbf{P}}, \Theta_1) \tag{10}$$

Authorized licensed use limited to: Zhejiang University. Downloaded on August 20,2020 at 13:09:20 UTC from IEEE Xplore. Restrictions apply.

Fig. 5. Pseudo color image of Cuprite data set (Band 5, 15, and 25 are displayed as blue, green, and red, respectively).

Fig. 6. Simulated LR-MSI of the Pavia data set.

where $\Theta_1 = (\mathbf{W}_1, \mathbf{B}_1)$ consists of the weights $\mathbf{W}_1$ and the biases $\mathbf{B}_1$ of the spectral SR network. The spectral SR network contains n convolutional layers, in which padding is also adopted to prevent shrink in the size of the image. The output of the spatial SR network is directly fed to this network. The first $n-1$ convolutional layers are used to learn joint features for all the band images in HR-HSI with a "Relu" layer. Specifically, Layer "Conv_1" has D different convolutional kernels with size $3 \times 3 \times 3$ and layers "Conv_2" … "Conv_n-1" have the same structure, each of them has 64 different convolutional kernels with size $3 \times 3 \times 3$. Then the final convolutional layer with a kernel size of $3 \times 1 \times 1$ is used to estimate all the band images of HR-HSI $\widehat{\mathbf{Y}}$ from these joint features. The parameters involved in this spectral SR network are listed in Table III.

In the proposed SepSSJSR, spatial SR network and spectral SR network are trained sequentially, which means that the spatial SR is trained first and then for the spectral SR network. Noted that ℓ_1-based loss function is also adopted to train these two networks. When these two SR networks are well-trained, the proposed SepSSJSR can be used to reconstruct HR-HSI directly from LR-MSI. The training procedure of the proposed SepSSJSR is summarized in Algorithm 2.

The LR-MSI can also be obtained from HR-HSI by spatially downsampling first and then spectral degradation. Therefore, the HR-MSI can also be reconstructed from LR-MSI by conducting spectral SR first and then spatial SR. Another version of SepSSJSR is proposed as illustrated in Fig. 3. Without loss of generality, we denote the SpeSSJSR in Fig. 2 as SpeSSJSR1 and that in Fig. 3 as SpeSSJSR2. The network parameters involved in the spectral SR and spatial SR networks of SpeSSJSR2 are similar to those in SpeSSJSR1.

Algorithm 2 Training Process of SepSSJSR1

Input: $1\times d\times 64\times 64$ sub image cube
Output: $1\times D\times 64c\times 64c$ sub image cube
Initialize: Initializing the network with a semi orthogonal matrix, as described in [31].

```
Train the spatial SR network first
while the spatial network hasn't converged yet do
    computing loss2 according to eq. (5).
    train the cnn_Model2(Training_data LR-MSI X, Training_label HR-MSI P)
    update W_2 and B_2 by minimizing the loss2 between the reconstructed HR-MSI P̂ and the corresponding label P
end while
save parameters Θ_2 of the cnn_Model2
Train the spectral SR network
while the spectral network hasn't converged yet do
    computing loss1 according to eq. (5).
    train the cnn_Model1(Training_data is the output of cnn_Model2 HR-MSI P̂, Training_label HR-HSI Y)
    update W_1 and B_1 by minimizing the loss1 between the reconstructed HR-MSI Ŷ and the corresponding label Y
end while
save parameters Θ_1 of the cnn_Model1
```

Fig. 7. Simulated LR-MSI of the PaviaU data set.

III. Experiments and Results

A. Data Sets

In this experiment, data sets collected from two well-known hyperspectral sensors, namely ROSIS and AVIRIS, are selected. Two scenarios acquired by the ROSIS sensor are chosen from the flight activities at the Pavia Centre and the Pavia University. Both of these two data sets cover a spectral range of approximately 430–860 nm. The Pavia Centre has 102 spectral bands, while the Pavia University has 103 spectral bands. Though Pavia Centre scene contains 1096×1096 pixels, only 1096×715 effective pixels are selected by removing the area with no information. The pseudo color images for these two data sets are shown in Fig. 4. For the AVIRIS sensor, the Cuprite data set, which contains 512×614 pixels covering a spectral range of 400–2400 nm, is adopted. Of the 224 spectral bands, the 22 corrupted spectral bands were removed and the remaining 202 spectral bands were used. The pseudo color image for Cuprite data set is shown in Fig. 5.

For quantitative assessment, these three data sets are used as groundtruth of HR-HSI to simulate LR-MSI by performing

Authorized licensed use limited to: Zhejiang University. Downloaded on August 20,2020 at 13:09:20 UTC from IEEE Xplore. Restrictions apply.

TABLE V
BAND SIMULATION RELATIONSHIP BETWEEN MSI OF LANDSAT SENSOR AND HSI OF AVIRIS SENSOR

MSI simulation bands	B	G	R	NIR	SWIR1	SWIR2
Corresponding HSI bands	5-11	12-19	23-31	39-53	117-137	160-186

Fig. 8. Simulated LR-MSI of the Cuprite data set.

Fig. 9. Experimental results of the proposed spatial SR network with different number of layers over Pavia Centre data set.

Fig. 10. Experimental results of the proposed spectral SR network with different number of layers over Pavia Centre data set.

Fig. 11. Experimental results of the proposed SimSSJSR with different number of layers over Pavia Centre data set.

spectral downsampling and spatial downsampling. Due to the different spectral ranges covered by the data sets collected by these two sensors, the spectral downsampling is slightly different. The ROSIS sensor covers a spectral range from 430 to 860 nm, which is almost the same as the multispectral sensor of IKONOS that covers a spectral range from 450 to 860 nm. Therefore, for the Pavia Center and Pavia University data sets acquired by the ROSIS sensor, the multispectral sensor configuration of IKONOS is used to simulate their corresponding HR-MSIs. The simulated HR-MSIs of the Pavia Centre data set and Pavia University data set, which have four spectral bands, can be acquired by averaging the corresponding band images in HR-HSIs according to Table IV. Specifically, since the wavelength of band B of IKONOS sensor corresponds to the wavelength of band images from 3 to 22 from the ROSIS sensor, band B of IKONOS sensor is simulated by averaging band images from 3 to 22 from ROSIS sensor. Similarly, band G, R, and NIR can also be simulated according to the band matching relationship listed in Table IV.

For the HSI acquired by AVIRIS sensor, i.e., Cuprite data set in this experiment, the MSI sensor of Landsat TM is used for simulation. Bands B, G, R, NIR, SWIR1, and SWIR2 of Landsat TM are adopted and their matching relationship with AVIRIS sensor is illustrated in Table V. The same average strategy as that of IKONOS and ROSIS sensors is used to simulate these six band images of Landsat TM from HSIs of AVIRIS sensor. After the spectral downsampling, all these three simulated HR-MSIs are downsampled by a factor of 2 using Gaussian low-pass spatial filtering to generate the LR-MSIs. The simulated LR-MSI for Pavia Centre data set, Pavia University data set, and Cuprite data set are shown in Figs. 6–8, respectively.

For these three data sets, a 150 × 150 subregion, as highlighted in blue square in Figs. 4 and 5, is selected to validate the performance of our proposed models, whereas the remaining pixels are used for training. In order to form inputs

Authorized licensed use limited to: Zhejiang University. Downloaded on August 20,2020 at 13:09:20 UTC from IEEE Xplore. Restrictions apply.

Fig. 12. Sample band images of HR-HSI reconstructed for Pavia Centre data set by the proposed SepSSJSR1, SepSSJSR2, and SimSSJSR. The groundtruth maps are shown in the first row. The reconstructed band images for these three algorithms are shown in second–fourth rows, and their corresponding reconstruction error images are shown in fifth–seventh rows. The legend color bar is shown in the bottom row.

Fig. 13. Sample band images of HR-HSI reconstructed for Pavia University data set by the proposed SepSSJSR1, SepSSJSR2, and SimSSJSR, respectively. The groundtruth maps are shown in the first row. The reconstructed band images for these three algorithms are shown in second–fourth rows, and their corresponding reconstruction error images are shown in fifth–seventh rows. The legend color bar is shown in the bottom row.

for the proposed SSJSR networks, subimages with a size of $64 \times 64 \times d$ are cropped by using a 64×64 spatial window sliding on the simulated LR-MSIs. Their corresponding $128 \times 128 \times D$ HR-HSIs are also cropped as groundtruth. In particular, $d = 4$ and $D = 102, 103$ for Pavia Centre and Pavia University data sets, respectively, whereas $d = 6$ and $D = 202$ for Cuprite data set. All the experiments are implemented on the Pytorch framework with NVIDIA 1080 GPU, using Adam optimizer [32], where backpropagation (BP) [33] strategy is adopted to train the network with a learning rate of 0.0002.

B. Quantitative Metrics

In order to evaluate the performance of SSJSR quantitatively, three metrics are used to evaluate the quality of the reconstructed HR-HSIs in terms of both the spatial

Authorized licensed use limited to: Zhejiang University. Downloaded on August 20,2020 at 13:09:20 UTC from IEEE Xplore. Restrictions apply.

Fig. 14. Sample band images of HR-HSI reconstructed for Cuprite data set by the proposed SepSSJSR1, SepSSJSR2, and SimSSJSR, respectively. The groundtruth maps are shown in the first row. The reconstructed band images for these three algorithms are shown in the second–fourth rows, and their corresponding reconstruction error images are shown in fifth–seventh rows. The legend color bar is shown in the bottom row.

reconstruction quality of each band image and the spectral reconstruction quality of each pixel, including mean peak SNR (MPSNR), structural similarity index (MSSIM), and spectral angle mapper (SAM). The MPSNR, which measures the similarities between the reconstructed images $F(\mathbf{X})$ and groundtruth $\mathbf{Y}$ in a band by band manner, is defined as

$$\mathrm{MPSNR} = \frac{1}{D}\sum_{i=0}^{D} 10 \times \log_{10}\left(\frac{\mathrm{MAX}_j^2}{\mathrm{MSE}_j}\right) \tag{11}$$

Authorized licensed use limited to: Zhejiang University. Downloaded on August 20,2020 at 13:09:20 UTC from IEEE Xplore. Restrictions apply.

Fig. 15. Example spectra of several typical materials in Pavia Centre and Pavia University data sets. (a) Trees located at (965,38) in Pavia Centre. (b) Tiles located at (1048,15) in Pavia Centre. (c) Meadows located at (475,1) in Pavia University. (d) Gravel located at (461,5) in Pavia University.

where MAX_j represents the maximum pixel value in the jth band image of $\mathbf{Y}$ and MSE_j is the mean square error (MSE) of the jth band image. Generally, the higher the MPSNR, the better the reconstructed quality.

The MSSIM measures the structural similarity of the reconstructed images $F(\mathbf{X})$ and groundtruth $\mathbf{Y}$. It is defined as

$$\mathrm{MSSIM} = \frac{1}{D}\sum_{i=0}^{D}\frac{(2\mu_{F(\mathbf{X})_i}\mu_{\mathbf{Y}_i}+c_1)(2\sigma_{F(\mathbf{X})_i\mathbf{Y}_i}+c_2)}{\left(\mu^2_{F(\mathbf{X})_i}+\mu^2_{\mathbf{Y}_i}+c_1\right)\left(\sigma^2_{F(\mathbf{X})_i}+\sigma^2_{\mathbf{Y}_i}+c_2\right)} \tag{12}$$

where $F(\mathbf{X})_i$ and $\mathbf{Y}_i$ represent the ith band image in $F(\mathbf{X})$ and $\mathbf{Y}_i$, respectively, $\sigma^2_{F(\mathbf{X})_i}$ and $\sigma^2_{\mathbf{Y}_i}$ are the variance of $F(\mathbf{X})_i$ and $\mathbf{Y}_i$, respectively, $\mu_{F(\mathbf{X})_i}$ and $\mu_{\mathbf{Y}_i}$ are the mean of $F(\mathbf{X})_i$ and $\mathbf{Y}_i$, respectively, $\sigma_{F(\mathbf{X})_i\mathbf{Y}_i}$ is the covariance of $F(\mathbf{X})_i$ and $\mathbf{Y}_i$, and c_1 and c_2 are constants used to maintain stability. The higher the MSSIM, the better preservation of the spatial structures.

The SAM is defined as the spectral angle between the reconstructed spectra $\widehat{\mathbf{x}}$ and its corresponding groundtruth spectra $\mathbf{x}$, which is defined as

$$\mathrm{SAM}(\mathbf{x},\widehat{\mathbf{x}}) = \arccos\left(\frac{\langle \mathbf{x},\widehat{\mathbf{x}}\rangle}{\|\mathbf{x}\|_2\|\widehat{\mathbf{x}}\|_2}\right) \tag{13}$$

where $\langle\cdot,\cdot\rangle$ denotes dot product of two vectors and $\|\mathbf{x}\|_2$ represents the l_2 norm of a vector. The smaller SAM, the less spectral distortion.

C. Parameter Analysis

1) Number of Convolutional Layers: First, the number of convolutional layer in the spatial SR network is analyzed by varying it from 3 to 5. The experimental results of the spatial SR network with a different number of convolutional layers over Pavia Centre data set are shown in Fig. 9. It is observed that the proposed spatial SR network with three convolutional layers slightly outperforms that with four or five convolutional layers. Therefore, three convolutional layers are adopted in the spatial SR network. In addition, the number of the convolutional layer in the proposed spectral SR is also discussed by varying it from 3 to 6, and the experimental results of the spectral SR network with different number of convolutional layer over Pavia Centre data set are shown in Fig. 10. It is observed that the proposed spectral SR network with five convolutional layers produces higher MPSNR than that with other numbers of convolutional layers. Although it has slightly smaller MSSIM and larger SAM than that with six convolutional layers, five convolutional layers are adopted since less parameters are to be learned and thus computational time is saved. Finally, the number of the convolutional layer in the proposed SimSSJSR is also discussed. Fig. 11 shows experimental results for the proposed SimSSJSR with different number of convolutional layers. Obviously, five convolutional layers are the best choice to achieve the balance between image quality and computational overhead for the proposed SimSSJSR.

2) Loss Function Selection: In this experiment, two different loss functions are compared, i.e., the ℓ_1 loss function measuring absolute reconstruction error and the ℓ_2 loss function measuring MSE. These two loss functions are used to train the proposed SimSSJSR, spatial SR, and spectral SR, respectively. The comparative results for these three networks with different loss functions over the Pavia Center data set are listed in Table VI. It can be seen that, for all the three CNN-based SR approaches, the average quantitative results of the ℓ_1 loss function obviously outperform that of the MSE loss function. Therefore, ℓ_1-based loss function is chosen for all the proposed SR approaches.

D. Comparative Results

In this experiment, the performance of the proposed three CNN-based SSJSR networks, including SimSSJSR,

Authorized licensed use limited to: Zhejiang University. Downloaded on August 20,2020 at 13:09:20 UTC from IEEE Xplore. Restrictions apply.

TABLE VI

AVERAGE QUANTITATIVE RESULTS OF THE PROPOSED SPATIAL SR, SPECTRAL SR, AND SIMSSJSR WITH THE DIFFERENT LOSS FUNCTIONS ON THE PAVIA CENTRE DATA SET

Algorithms	Loss Function (Best Values)	MPSNR ($+\infty$)	MSSIM (1)	SAM (0)
Spatial SR	ℓ_2 loss function	35.2448	0.9642	2.5603
	ℓ_1 loss function	**35.9290**	**0.9692**	**2.2918**
Spectral SR	ℓ_2 loss function	47.0060	**0.9966**	1.3364
	ℓ_1 loss function	**47.4015**	**0.9966**	**1.2508**
SimSSJSR	ℓ_2 loss function	34.7612	0.9545	4.4139
	ℓ_1 loss function	**34.9010**	**0.9565**	**4.3222**

TABLE VII

COMPARATIVE RESULTS OF DIFFERENT SSJSRS OVER PAVIA CENTRE DATA SET

Dataset	Algorithm (Best Values)	MPSNR ($+\infty$)	MSSIM (1)	SAM (0)
Pavia Centre	SimSSJSR	34.9010	0.9565	4.3222
	SepSSJSR1	**35.3745**	**0.9596**	**3.9738**
	SepSSJSR2	27.9022	0.7610	11.4821
Pavia University	SimSSJSR	32.2800	0.9586	4.0257
	SepSSJSR1	**32.5945**	**0.9612**	**3.6921**
	SepSSJSR2	24.6507	0.7011	13.8680
Cuprite	SimSSJSR	33.9145	0.9532	1.8986
	SepSSJSR1	**34.6180**	**0.9566**	**1.8429**
	SepSSJSR2	24.6117	0.7153	5.3332

SepSSJSR1, and SepSSJSR2, are verified on the three data sets. The comparative results of these three different proposed CNN-based SSJSR networks over different data sets are listed in Table VII, in which the best results over each data set are marked in bold. Obviously, the proposed SepSSJSR1 outperforms the other two proposed SepSSJSR2 and SimSSJSR over all these three data sets acquired by ROSIS or AVIRIS sensors with the highest MPSNR and MSSIM values and lowest SAM, demonstrating that it is better to conduct spatial SR prior to spectral SR in SSJSR. Compared with SimSSJSR, the proposed SepSSJSR1 divides the complex SSJSR problem into two separate problems so that its performance is superior. According to the network structure listed in Tables II and III, spatial SR often involves more parameters than spectral SR. Therefore, the performance of spectral SR would be more accurate than that of spatial SR if the same amount of training samples are used to train the network, which can also be demonstrated by the results in Table VI. Though SepSSJSR2 divides the SSJSR problem into two separate problems as that in SepSSJSR1, its performance is very poor since the complex spatial SR problem is conducted over the estimated LR-HSIs. Figs. 12–14 present several sample reconstructed band images and their corresponding reconstruction error maps for these three data sets, which also demonstrate that the proposed SepSSJSR1 can reconstruct band images with lower reconstruction error. Finally, Fig. 15 shows four example spectra reconstructed by different CNNs, where the proposed SepSSJSR1 can recovery spectra with higher precision. All these visual results also confirm that conducting spatial SR prior to spectral SR is better for SSJSR of LR-MSIs.

TABLE VIII

COMPUTATIONAL TIME OF DIFFERENT PROPOSED METHODS

Pavia Center			
Methods	SimSSJSR	SepSSJSR1	SepSSJSR2
Training(h)	10.82	29.25	14.56
Validating(ms)	235	273	251

(a) (b) (c) (d)

Fig. 16. Classification maps obtained by SVM over different images. (a) Groundtruth. (b) HR-MSIs reconstructed by spatial network in SepSSJSR1. (c) HR-HSIs reconstructed by SepSSJSR1. (d) Groundtruth Pavia University data set.

E. Computational Analysis

Table VIII lists the computational time of the proposed three CNN-based spatial and spectral jointly SR methods for Pavia Centre data set. All these networks are implemented on a computer with one NVIDIA 1080 GPU card and 16-GB memory. It is observed that SimSSJSR takes the least time for training because it conducts SR in spectral and spatial domain simultaneously. SepSSJSR1 takes more time for training because it conducts spatial SR first and then constructs spectral SR in a high-resolution feature space, resulting in more parameters to be trained. However, as shown in Table VIII, once the networks are well trained, they take almost the same time to reconstruct HR-HSIs.

F. Discussion I: Application to Improve the Performance of Classification for MSIs

The SSJSR technique is proposed to jointly enhance both spatial and spectral resolution of MSIs, aiming to extend the use of MSIs to a high-resolution level. For example, the performance of the classification of MSIs can be improved by the proposed SSJSR algorithms in two aspects: realizing subpixel mapping of MSIs according to high spatial resolution and improving the accuracy of subpixel mapping according to

Authorized licensed use limited to: Zhejiang University. Downloaded on August 20,2020 at 13:09:20 UTC from IEEE Xplore. Restrictions apply.

Fig. 17. Experimental results of the proposed SepSSJSR1 on real Landsat 8 MSIs. (a) Sample images of LR-MSI acquired from Landsat 8. (b) Reconstructed sample band images with paddy cover of HR-HSI on AVIRIS band configuration. (c) Reconstructed sample band images with soil cover of HR-HSI on AVIRIS band configuration. (d) Reconstructed spectrum of four sample points.

the reconstructed spectral curve. In order to validate such superiority, experiments of classification over Pavia University data set reconstructed by the best version of the proposed SSJSR (i.e., SepSSJSR1) in different stages are conducted, including HR-MSI reconstructed by the spatial network in SepSSJSR1, HR-HSI reconstructed by the SepSSJSR1, and groundtruth version of HSI. Note that the classification over original MSI is not conducted due to two reasons: the groundtruth of Pavia University data set is available at high-spatial level, and the superiority of the high spatial-resolution images over low spatial-resolution in terms of classification is obvious.

The support vector machine (SVM) with RBF kernel is adopted as classifier and about 10% of samples are randomly selected for training. Table IX lists the classification accuracy of SVM on different data sets in terms of per-class accuracy, average accuracy (AA), overall accuracy (OA), and Kappa coefficient (Kappa). It is observed that the classification of reconstructed HR-HSI significantly outperforms that on HR-MSI, indicating the proposed SSJSR can clearly improve the performance of classification by reconstructing spectral signature. Moreover, the classification performance over the reconstructed HR-HSI is just slightly inferior to its groundtruth version. The classification maps over these data, together with the groundtruth classification map, are shown in Fig. 16. These results also confirmed that the performance of HR-MSI can be improved by further reconstructing the spectral reflectance curves, especially in the highlighted part in Fig. 16. In summary, though the proposed SSJSR maybe cannot reproduce the exact spectral reflectance curve of pixels, e.g., the proposed SSJSR slightly smooths groundtruth spectra near band 80 in Fig. 15(a), the reconstructed spectra can certainly improve the performance of LR-MSIs by SR to HR-HSIs.

G. Discussion II: Application to Real MSIs Acquired by Landsat Satellite

In order to validate the effectiveness of the proposed SSJSR on MSIs in real applications, we also applied the best well-trained SSJSR model (i.e., SepSSJSR1) for AVIRIS band configuration in the previous experiment to real Landsat 8 data set shown in Fig. 17. Specifically, two areas with different land-cover and land-use are chosen for evaluation. As shown in Fig. 17, the left subimage is covered with a lot of paddy fields while the right subimage mainly contains soil. It is observed that the spectra of pixels in the Landsat 8 image are enhanced to own a spectral curve to character their spectral property. For example, the pixels with same land-cover own similar spectral curves, such as P1 and P2, P3 and P4, and the

Authorized licensed use limited to: Zhejiang University. Downloaded on August 20,2020 at 13:09:20 UTC from IEEE Xplore. Restrictions apply.

TABLE IX
CLASSIFICATION ACCURACY OF SVM OVER DIFFERENT RECONSTRUCTED IMAGES OF PAVIA UNIVERSITY DATA SET

Class	Class Name	Training	Testing	HR-MSIs	HR-HSIs	Orginal data
1	Asphalt	663	5968	89.70	91.92	**93.06**
2	Meadows	1865	16784	97.25	**97.49**	97.23
3	Gravel	210	1889	38.12	77.18	**78.83**
4	Trees	306	2758	89.23	**96.81**	94.66
5	Painted metal sheets	135	1210	99.34	**99.50**	99.29
6	Bare Soil	503	4526	48.48	76.51	**89.65**
7	Bitumen	133	1197	33.58	**85.71**	84.66
8	Self-Blocking Bricks	368	3314	81.83	**87.60**	84.89
9	Shadows	95	852	**100**	**100**	99.88
	AA			75.34	90.30	**91.35**
	OA			83.91	92.02	**93.28**
	Kappa			0.7805	0.8934	**0.9108**

pixels with different land-use own slightly different spectral curves, like P2 and P4. Such an enhancement in the spectral domain is especially useful to improve the performance of different applications of MSIs, such as classification, target detection, and recognition. Meanwhile, the spatial resolution of images is improved to better model spatial structure of objects. Consequently, our proposed SSJSR can make full use of available multispectral data for high-resolution-based analysis or applications.

IV. CONCLUSION AND FUTURE WORK

In this article, both spatial and spectral resolutions of HSI are improved using CNNs. Specifically, three CNNs are constructed, including SimSSJSR using a full 3-D CNN to learn an end-to-end mapping from LR-MSI to HR-HSI, and two separate networks executing the spatial SR and spectral SR sequentially. Extensive experimental results demonstrate that SepSSJSR that conduct spatial SR prior to spectral SR can reconstruct HR-HSI with higher accuracy. Our proposed work provides a feasible and efficient way to acquire HSI with both high spatial and spectral resolution. In addition, it is better to conduct spatial SR before spectral SR in SR implementations.

Our main concern is how to conduct both spatial SR and spectral SR together in just one framework for SSJSR. Two simple but effective CNN-based networks are constructed for spatial SR and spectral SR in SepSSJSR, respectively. Many state-of-the-art spatial SR network (e.g., SRCNN [6], [8] and robust SR in [9]) or spectral network (e.g., HSCNN [25] and HSCNN+ [26]) can be adopted to replace the corresponding network in the proposed SepSSJSR to further improve the reconstruction performance. Moreover, the prior formation, such as sparse prior [7], [9], can also be used to further alleviate the ill-posed mapping problem of SR.

REFERENCES

[1] X. Li and M. T. Orchard, "New edge-directed interpolation," *IEEE Trans. Image Process.*, vol. 10, no. 10, pp. 1521–1527, Oct. 2001.

[2] W. Dong, L. Zhang, G. Shi, and X. Wu, "Nonlocal back-projection for adaptive image enlargement," in *Proc. 16th IEEE Int. Conf. Image Process. (ICIP)*, Nov. 2009, pp. 349–352.

[3] H.-C. Chen and W.-J. Wang, "Two-stage interpolation algorithm based on fuzzy logics and edges features for image zooming," *EURASIP J. Adv. Signal Process.*, vol. 2009, no. 1, pp. 1–11, 2009.

[4] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2861–2873, Nov. 2010.

[5] X. Gao, K. Zhang, D. Tao, and X. Li, "Image super-resolution with sparse neighbor embedding," *IEEE Trans. Image Process.*, vol. 21, no. 7, pp. 3194–3205, Jul. 2012.

[6] 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.*, 2014, pp. 184–199.

[7] Z. Wang, D. Liu, J. Yang, W. Han, and T. Huang, "Deeply improved sparse coding for image super-resolution," 2015, *arXiv:1507.08905*. [Online]. Available: https://arxiv.org/abs/1507.08905

[8] C. Dong, C. C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 391–407.

[9] D. Liu, Z. Wang, B. Wen, J. Yang, W. Han, and T. S. Huang, "Robust single image super-resolution via deep networks with sparse prior," *IEEE Trans. Image Process.*, vol. 25, no. 7, pp. 3194–3207, Jul. 2016.

[10] W. Shi *et al.*, "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)*, Jun. 2016, pp. 1874–1883.

[11] G. Huang, Z. Liu, L. V. D. Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4700–4708.

[12] T. Tong, G. Li, X. Liu, and Q. Gao, "Image super-resolution using dense skip connections," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 4799–4807.

[13] 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, Nov. 2017.

[14] Y. Jia *et al.*, "From RGB to spectrum for natural scenes via manifold-based mapping," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 4705–4713.

[15] R. M. H. Nguyen, D. K. Prasad, and M. S. Brown, "Training-based spectral reconstruction from a single RGB image," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 186–201.

[16] A. Robles-Kelly, "Single image spectral reconstruction for multimedia applications," in *Proc. 23rd ACM Int. Conf. Multimedia (MM)*, 2015, pp. 251–260.

[17] B. Arad and O. Ben-Shahar, "Sparse recovery of hyperspectral signal from natural RGB images," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 19–34.

[18] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

Authorized licensed use limited to: Zhejiang University. Downloaded on August 20,2020 at 13:09:20 UTC from IEEE Xplore. Restrictions apply.

[19] Y. Pati, R. Rezaiifar, and P. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. 27th Asilomar Conf. Signals, Syst. Comput.*, Dec. 2002, pp. 40–44.
[20] S. W. Oh, M. S. Brown, M. Pollefeys, and S. J. Kim, "Do it yourself hyperspectral imaging with everyday digital cameras," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2461–2469.
[21] N. Akhtar and A. Mian, "Hyperspectral recovery from RGB images using Gaussian processes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 1, pp. 100–113, Jan. 2020.
[22] S. Galliani, C. Lanaras, D. Marmanis, E. Baltsavias, and K. Schindler, "Learned spectral super-resolution," 2017, *arXiv:1703.09470*. [Online]. Available: https://arxiv.org/abs/1703.09470
[23] J. Wu, J. Aeschbacher, and R. Timofte, "In defense of shallow learned spectral reconstruction from RGB images," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCVW)*, Oct. 2017, pp. 471–479.
[24] R. Timofte, V. De Smet, and L. Van Gool, "A+: Adjusted anchored neighborhood regression for fast super-resolution," in *Proc. Asian Conf. Comput. Vis.*, 2014, pp. 111–126.
[25] Z. Xiong, Z. Shi, H. Li, L. Wang, D. Liu, and F. Wu, "HSCNN: CNN-based hyperspectral image recovery from spectrally undersampled projections," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCVW)*, Oct. 2017, pp. 518–525.
[26] Z. Shi, C. Chen, Z. Xiong, D. Liu, and F. Wu, "HSCNN+: Advanced CNN-based hyperspectral recovery from RGB images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2018, pp. 939–947.
[27] A. Alvarez-Gila, J. V. D. Weijer, and E. Garrote, "Adversarial networks for spatial context-aware spectral image reconstruction from RGB," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCVW)*, Oct. 2017, pp. 480–490.
[28] Y. Yan, L. Zhang, J. Li, W. Wei, and Y. Zhang, "Accurate spectral super-resolution from single RGB image using multi-scale CNN," in *Proc. Chin. Conf. Pattern Recognit. Comput. Vis. (PRCV)*, 2018, pp. 206–217.
[29] S. Mei, J. Ji, Y. Geng, Z. Zhang, X. Li, and Q. Du, "Unsupervised spatial–spectral feature learning by 3D convolutional autoencoder for hyperspectral classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6808–6820, Sep. 2019.
[30] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1026–1034.
[31] A. M. Saxe, J. L. Mcclelland, and S. Ganguli, "Exact solutions to the nonlinear dynamics of learning in deep linear neural networks," in *Proc. Int. Conf. Learn. Represent.*, 2014.
[32] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Represent.*, 2015.
[33] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, Oct. 1986.

Shaohui Mei (Member, IEEE) received the B.S. degree in electronics and information engineering and the Ph.D. degree in signal and information processing from Northwestern Polytechnical University, Xi'an, China, in 2005 and 2011, respectively.

He was a Visiting Student with The University of Sydney, Sydney, NSW, Australia, from October 2007 to October 2008. He is currently an Associate Professor with the School of Electronics and Information, Northwestern Polytechnical University. His research interests include hyperspectral remote sensing image processing and applications, intelligent signal and information acquisition and processing, neural networks, pattern recognition, and machine learning.

Ruituo Jiang received the B.E. degree in electronics and information engineering from Northwestern Polytechnical University, Xi'an, China, in 2017, where he is currently pursuing the master's degree in signal and information processing, with a focus on hyperspectral image super-resolution.

Xu Li (Member, IEEE) received the B.Sc. degree in electronics and information technology, the M.Sc. degree in signal and information processing, and the Ph.D. degree in circuits and systems from Northwestern Polytechnical University, Xi'an, China, in 2001, 2004, and 2010, respectively.

In 2004, he joined the School of Electronics and Information, Northwestern Polytechnical University. From April 2007 to April 2008, he was a Visiting Ph.D. Student with Télécom ParisTech, Paris, France. From November 2012 to November 2014, he was an Enterprise Post-Doctoral Fellow with the Jiangsu Research and Development Center for Internet of Things, Wuxi, China. From December 2015 to March 2016, he was a Marie Currie Research Fellow with the School of Computer Science, University of Lincoln, Lincoln, U.K. He is currently an Associate Professor with the School of Electronics and Information, Northwestern Polytechnical University. His research interests include image fusion, pansharpening, and image enhancement.

Qian Du (Fellow, IEEE) received the Ph.D. degree in electrical engineering from the University of Maryland at Baltimore County, Baltimore, MD, USA, in 2000.

She is currently the Bobby Shackouls Professor with the Department of Electrical and Computer Engineering, Mississippi State University, Starkville, MS, USA. Her research interests include hyperspectral remote sensing image analysis and applications, pattern classification, data compression, and neural networks.

Dr. Du is a fellow of the SPIE-International Society for Optics and Photonics. She was a recipient of the 2010 Best Reviewer Award from the IEEE Geoscience and Remote Sensing Society. She was the Co-Chair of the Data Fusion Technical Committee of the IEEE Geoscience and Remote Sensing Society from 2009 to 2013 and the Chair of the Remote Sensing and Mapping Technical Committee of the International Association for Pattern Recognition from 2010 to 2014. She is the General Chair of the 4th IEEE GRSS Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, Shanghai, China, in 2012. She served as an Associate Editor for the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, the *Journal of Applied Remote Sensing*, and the IEEE SIGNAL PROCESSING LETTERS. Since 2016, she has been the Editor-in-Chief of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING.

Authorized licensed use limited to: Zhejiang University. Downloaded on August 20,2020 at 13:09:20 UTC from IEEE Xplore. Restrictions apply.