HyperPNN: Hyperspectral Pansharpening via Spectrally Predictive Convolutional Neural Networks

Lin He, *Member, IEEE*, Jiawei Zhu, Jun Li, *Senior Member, IEEE*, Antonio Plaza, *Fellow, IEEE*, Jocelyn Chanussot, *Fellow, IEEE*, and Bo Li

***Abstract*—Hyperspectral (HS) pansharpening intends to synthesize a HS image with a registered panchromatic image, to generate an enhanced image with simultaneous high spectral resolution and high spatial resolution. However, the spectral range gap between the two kinds of images and the need to resolve details for many continuous narrow bands make the technique prone to spectral distortion and spatial blurring. To mitigate the problems, we propose a new HS pansharpening framework via spectrally predictive convolutional neural networks (HyperPNN). In our proposed HyperPNN, spectrally predictive structure is introduced to strengthen the spectral prediction capability of a pansharpening network. Following the concept of the proposed HyperPNN, two specific pansharpening convolutional neural network (CNN) models, i.e., HyperPNN1 and HyperPNN2, are designed. Experimental results from three datasets suggest the excellent performance of our CNN-based HS pansharpening methods.**

***Index Terms*—Convolutional neural networks (CNN), hyperspectral (HS) image, pansharpening, spectral prediction.**

I. Introduction

With the development of imaging spectroscopy technique, hyperspectral (HS) image processing has become a flourishing topic in the fields of remote sensing and information acquisition with a wide range of applications [1]–[3]. However, due to inherent limitations of imaging systems, HS images, characterized by many continuous bands with high spectral resolution, usually suffer from low spatial resolution. This weakness leads to relatively poor performance in some practical applications [3]–[5]. Therefore, full-resolution HS images with both high spatial and spectral resolution are desired. One way to obtain such kind of ideal images is to fuse high spectral resolution HS images with high spatial resolution panchromatic (PAN) images, which is usually called HS pansharpening [3].

Following traditional pansharpening approaches which are mainly used for combining multispectral (MS) images with PAN images, some kinds of strategies therein could be possibly considered for the HS pansharpening task [6]–[8]. The first one is component substitution where the spatial component of an HS image is replaced with that of the connected PAN image, usually utilizing a mathematical image detail formulation involving, such as principal component analysis [9], [10], intensity-hue-saturation method [11], [12] and guided filtering [13]. The second strategy is multiresolution analysis, which extracts and injects spatial details in a multiresolution manner, some typical examples of which are smoothing filter-based intensity modulation (SFIM)[14], the MTF-generalized Laplacian pyramid (MTF-GLP) method [15] and regression-based high-pass modulation[16]. Another one is matrix factorization, which formulates the observed data with signal subspace representation and then leverages optimization tool to factorize related matrices [17]. The fourth one is Bayesian inference, which depends on Bayesian statistics to infer probabilistic parameters used for pansharpening models [18]–[20].

Over recent years, convolutional neural networks (CNNs) start prevailing in improving image spatial resolution. Dong *et al.* proposed a super-resolution CNN (SRCNN) which is a three-layer CNN to learn the mapping from the input low-resolution image to the output high-resolution image [21]. Kim *et al.* also designed a CNN structure for super-resolution, where the residual component is learned [22]. Such two CNNs are in fact representatives of two kinds of basic CNN structures, respectively. One structure is flat network where the input signals mainly pass through an only path of convolutional layers; the other structure could be called component network which comprises of parallel pathways, where individual signal components pass the pathways and then add up at the end of the parallel structure. A component network is usually equipped with skip connections. After the works of Dong *et al.* and Kim *et al.*, CNNs lent themselves to multispectral pansharpening task. Masi *et al.* presented a flat pansharpening CNN (PNN) [23], where the

Manuscript received December 9, 2018; revised April 12, 2019; accepted May 5, 2019. Date of publication June 4, 2019; date of current version September 16, 2019. This work was supported in part by the National Natural Science Foundation of China under Grants 61571195, 61771496, 61633010, and 61836003, in part by the Guangdong Provincial Natural Science Foundation under Grants 2016A030313254, 2016A030313516, and 2017A030313382, and in part by the National Key Research and Development Program of China under Grant 2017YFB0502900. *(Corresponding authors: Lin He; Jun Li.)*

L. He and J. Zhu are with the School of Automation Science and Engineering, South China University of Technology, Guangzhou 510640, China (e-mail: helin@scut.edu.cn; zjw7342@163.com).

J. Li is with the Guangdong Provincial Key Laboratory of Urbanization and Geo-simulation, School of Geography and Planning, Sun Yat-sen University, Guangzhou 510275, China (e-mail: lijun48@mail.sysu.edu.cn).

A. Plaza is with the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, Escuela Politécnica, University of Extremadura, 10003 Cáceres, Spain (e-mail: aplaza@unex.es).

J. Chanussot is with the University Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, 38000 Grenoble, France (e-mail: jocelyn.chanussot@gipsa-lab.grenoble-inp.fr).

B. Li is with the Beijing Key Laboratory of Digital Media, School of Computer Science and Engineering, and the State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, Beijing 100191, China (e-mail: boli@buaa.edu.cn).

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

Digital Object Identifier 10.1109/JSTARS.2019.2917584

pre-interpolated low-resolution MS image is stacked with the associated PAN image to go through a CNN process. Wei *et al.* adopted a CNN method with residual component learning for pansharpening [24], and also a component pansharpening network involving multiscale kernels trick [25]. Later, Giuseppe *et al.* bought target-adaptive modality into CNN design to further improve the network performance[26].

HS pansharpening is faced with challenges significantly different from other image resolution enhancement tasks, including traditional multispectral pansharpening. The challenges mainly lie on two aspects. First, the spectral range of an HS image is usually much wider than that of a registered PAN image, which makes the HS bands out of the PAN spectral range difficult to sharpen. Second, the HS image contains many continuous bands with high spectral resolution, which obviously means that resolving details from the PAN image for all the bands at one time is a hard task. Therefore, HS pansharpening requires a strong spectral-prediction ability in the adopted method. Due to the fact that existing CNN-based multispectral pansharpening methods do not specially consider the spectral prediction problem, there are obvious limitations when we straightforwardly apply these methods to HS pansharping. Motivated by this point, in this paper we present a new framework of HS pansharpening via spectrally predictive CNN (HyperPNN), which exploits the spectral convolution structure to strengthen spectral prediction. Following that, two specific realizations of HyperPNN, i.e., HyperPNN1, which is flat type, and HyperPNN2, which is component type, are developed in this paper. Each of the two proposed HyperPNN methods is mainly composed of a spectral prediction sub-network and a spectral–spatial inference sub-network, being able to achieve remarkable HS pansharpening performance.

The rest of this paper is organized as follows. Brief background knowledge about CNN is provided in Section II. Detailed descriptions of the proposed HyperPNN methods are presented in Section III. The results of our experiments are shown and analyzed in Section IV. Finally, the conclusion is drawn in Section V.

II. Convolutional Neural Networks

Normally, CNNs are designed for processing data residing on regular lattices. The data are convoluted on successive multiple layers with slide windows or kernels. During the process, the kernels are updated to minimize the discrepancy between the expected and the current outputs where forward-propagate and back-propagate processes are involved.

Consider the HS pansharpening case. The PAN image with the size $L \times W$ is denoted as $\mathbf{P} \in \mathbb{R}^{L \times W}$, while the low-resolution HS image with $l \times w$ pixels and B spectral bands is indicated as $\widetilde{\mathbf{H}} \in \mathbb{R}^{l \times w \times B}$. By fusing $\mathbf{P}$ and $\widetilde{\mathbf{H}}$, we expect a high-resolution HS image $\widehat{\mathbf{H}} \in \mathbb{R}^{L \times W \times B}$. Thus, CNN-based HS pansharpening can be formulated as minimizing the following loss function:

$$\ell(\boldsymbol{\theta}) = \|M(\mathbf{P}, \widetilde{\mathbf{H}}; \boldsymbol{\theta}) - \widehat{\mathbf{H}}\|_F^2 \tag{1}$$

where $M(\cdot)$ represents the mapping from the input to the output of a CNN, $\boldsymbol{\theta}$ denotes parameters to be optimized, and $\|\cdot\|_F$ refers to the Frobenius norm. The CNN can learn the involved kernels from data, offering the possibility for HS pansharpening.

III. Proposed Methods

As aforementioned, HS pansharpening is faced with two special challenges. One is the spectral range gap between the HS image and the PAN image; the other is resolving details for many continuous and high spectral resolution bands simultaneously. Thus, approaches with high-spectra-predictive capability are desirable.

Inspired by such an analysis, we here propose the concept of HS pansharpening via spectrally predictive CNNs, namely HyperPNN, to accommodate HS pansharpening challenges, where spectrally predictive layers are added to strengthen the spectral prediction ability of a network. In a practical HyperPNN network, it is expected to consist of two sub-networks, i.e., a spectral prediction sub-network and a spatial–spectral inference sub-network. The spectral prediction sub-network comprises spectral convolutional layers with convolution operations only along spectral dimension, being mainly for the purpose of spectral prediction, while the spatial–spectral inference sub-network contains normal spatial–spectral convolutional layers, mainly aiming at exploiting spectral and spatial contextual information, simultaneously.

In the following, we present two new specific structures based on the HyperPNN concept, i.e., HyperPNN1 and HyperPNN2.

A. HyperPNN1

Our HyperPNN1, shown in Fig. 1, is a flat CNN, with the purpose of building a spectrally predictive CNN suitable for HS pansharpening. As shown in Fig. 1, HyperPNN1 contains seven layers, including two parts—the spectral prediction sub-network with four layers and the spatial–spectral inference sub-network with three layers, respectively. Notice that, in Fig. 1, the spatial sizes of adopted convolutional kernels are specified for each layer.

In the beginning of the network, the pre-interpolated HS image is connected to two serialized convolutional layers, where the image is the input of the first convolutional layer whose output is then fed to the second convolutional layer. With the help of such two convolutional layers, the pre-interpolated HS image is transformed to a feature cube with lower dimension, where the spectral prediction is performed. Then, that feature cube is concatenated with the PAN image along the spectral dimension on the third layer, forming the input of the next convolutional layer which is serialized with another two convolutional layers to establish a spectral–spatial inference sub-network. In the end, the output of the spectral–spatial inference module is connected to another two serialized spectrally predictive convolutional layers, strengthening the spectral prediction to yield the pansharpened HS image.

The spectral prediction sub-network of our HyperPNN1 is constructed by the first two layers and the last two ones, where convolutions with kernels in the spatial size of 1×1 are applied. That is, we perform operation per pixel along the only spectral dimension, aiming at yielding spectral representative features and

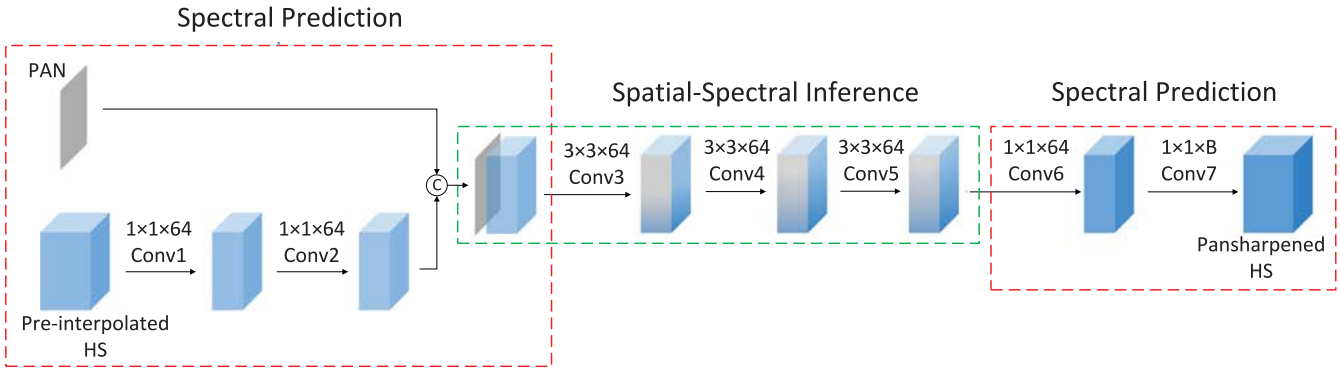

Fig. 1. Graphical illustration of the proposed HyperPNN1. ⓒ denotes concatenation operation along spectral dimension. Notice that the sizes of adopted convolutional kernels are specified in each layer.

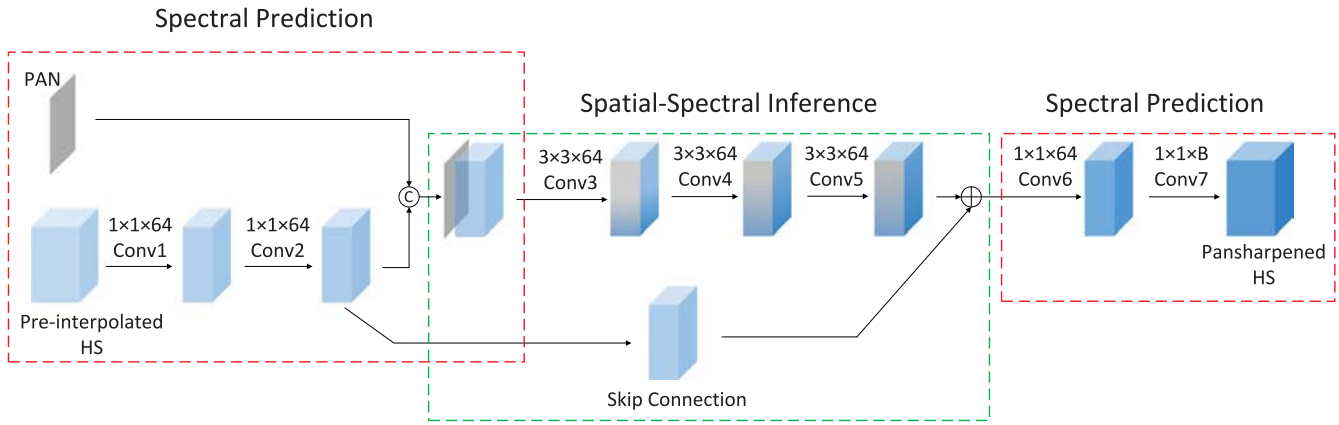

Fig. 2. Graphical illustration of the proposed HyperPNN2.

hence strengthening the spectral prediction and learning. Meanwhile, the lower dimension feature cube means the decrease of the number of parameters fed into the connected spatial–spectral inference network, thus reducing the risk of training instability and overfitting. In the final part of the whole network, there are two additional spectral layers to further dispose information along spectral dimension. Through such a design of the spectral prediction sub-network, sufficient spectral information useful for pansharpening will be collected during the procedure of successive forward- and back-propagation, thus achieving strong spectral predictive ability for the images to be pansharpened.

On the other hand, the spatial–spectral inference sub-network of HyperPNN1 is constituted by three spatial–spectral convolutional layers. The input of its first layer is the feature cube made by concatenating the PAN image and the spectrally convolved HS image. In the sub-network, kernels with certain spatial covering are used, aiming at introducing spatial contextual information into the spectral–spatial features. Specifically, we exemplify here 3×3 spatial covering.

Let $\boldsymbol{\theta} = (\mathbf{W}, \mathbf{b})$, where $\mathbf{W}$ and $\mathbf{b}$ represent the kernels and biases, respectively. When learning with HyperPNN1, both forward- and back-propagation processes are included. After each forward-propagation, the network output will be compared with the expected one, seeing the loss function $\ell(\boldsymbol{\theta})$ in (1). If the loss $\ell(\boldsymbol{\theta})$ exceeds the allow range, network parameters $\boldsymbol{\theta}$ will, in a subsequent backward-propagation process, be updated by a stochastic gradient descend (SGD) process with a learning rate of η. For the $(t+1)$th iteration, the update can be briefly formulated in the following:

$$\boldsymbol{\theta}^{t+1} = \boldsymbol{\theta}^{t} - \eta \cdot \frac{\partial \ell^{t}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{t}} \tag{2}$$

where

$$\frac{\partial \ell^{t}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{t}} = \left\{ \frac{\partial \ell^{t}(\boldsymbol{\theta})}{\partial \mathbf{W}^{t}}, \frac{\partial \ell^{t}(\boldsymbol{\theta})}{\partial \mathbf{b}^{t}} \right\}. \tag{3}$$

B. HyperPNN2

Now we present another HS pansharpening CNN, i.e., HyperPNN2. As demonstrated in, similar to HyperPNN1, the overall structure of our HyperPNN2 also consists of a spectrally predictive sub-network and a spectral–spatial inference sub-network, totally being with seven layers. Same as HyperPNN1, HyperPNN2 has four layers in the spectral prediction sub-networks and three layers in the spectral–spatial inference sub-network.

However, different from HyperPNN1, HyperPNN2 has a skip connection over the convolutional layers of the spatial–spectral inference sub-network. Therein, the skip connection plays mainly two roles. On the one hand, it makes the low level features extracted from low resolution HS image to be the baseline of learning in the convolutional pathway of the spatial–spectral inference sub-network, hence reducing its optimization

searching scope. Considering the output of Conv6 $\mathbf{O}_6$, it can be represented as

$$\mathbf{O}_6 = \varphi(\mathbf{W}_6 * (\mathbf{O}_5 + \mathbf{O}_2) + \mathbf{b}_6) \tag{4}$$

where $\mathbf{O}_2$, $\mathbf{O}_5$, and $\mathbf{O}_6$ represent the output of Conv2, Conv5, and Conv6, respectively, $\mathbf{W}_6$ and $\mathbf{b}_6$ are weights and biases of Conv6, while $\varphi(\cdot)$ is the activation function and $*$ denotes convolution operation. Equation (4) indicates that $\mathbf{O}_6$ is based on $\mathbf{O}_2$ so that the low level features $\mathbf{O}_2$ directly works on $\mathbf{O}_6$. This frees the convolutional pathway in the spatial–spectral inference network from learning repeatedly the feature component shared with the low level layer. On the other hand, gradient vanishing [27]–[29] tends to be mitigated under HyperPNN2, thanks to the skip connection. Consider kernel Conv2, its associated gradient is of the following form:

$$\begin{aligned}
&\frac{\partial \ell(\boldsymbol{\theta})}{\partial \mathbf{W}_2} \\
&= \frac{\partial \ell(\boldsymbol{\theta})}{\partial \mathbf{O}_2}\frac{\partial \mathbf{O}_2}{\partial \mathbf{W}_2} \\
&= \frac{\partial \ell(\boldsymbol{\theta})}{\partial \mathbf{O}_7}\frac{\partial \mathbf{O}_7}{\partial \mathbf{O}_6}\left(\frac{\partial \mathbf{O}_6}{\partial \mathbf{O}_5}\frac{\partial \mathbf{O}_5}{\partial \mathbf{O}_4}\frac{\partial \mathbf{O}_4}{\partial \mathbf{O}_3}\frac{\partial \mathbf{O}_3}{\partial \mathbf{O}_2} + \frac{\partial \mathbf{O}_6}{\partial \mathbf{O}_2}\right)\frac{\partial \mathbf{O}_2}{\partial \mathbf{W}_2} \\
&= \frac{\partial \ell(\boldsymbol{\theta})}{\partial \mathbf{O}_7}\mathbf{W}_7\varphi'(\mathbf{O}_6)\left(\mathbf{W}_6\prod_{i=1}^{3}\mathbf{W}_{6-i}\varphi'(\mathbf{O}_{5-i}) + \mathbf{W}_6\right)\frac{\partial \mathbf{O}_2}{\partial \mathbf{W}_2} \\
&= \frac{\partial \ell(\boldsymbol{\theta})}{\partial \mathbf{O}_7}\mathbf{W}_7\varphi'(\mathbf{O}_6)\mathbf{W}_6\left(\prod_{i=1}^{3}\mathbf{W}_{6-i}\varphi'(\mathbf{O}_{5-i}) + 1\right)\frac{\partial \mathbf{O}_2}{\partial \mathbf{W}_2}
\end{aligned} \tag{5}$$

where $\varphi'(\cdot)$ represents the derivative of the activation function. Equation (5) is used to update Conv2 during the back-propagation process with SGD. It should be noted that the term $(\prod_{i=1}^{3}\mathbf{W}_{6-i}\varphi'(\mathbf{O}_{5-i}) + 1)$ in (5) is attributed to the existence of the skip connection. Without the skip connection, (5) will be degenerated to

$$\begin{aligned}
&\frac{\partial \ell(\boldsymbol{\theta})}{\partial \mathbf{W}_2} \\
&= \frac{\partial \ell(\boldsymbol{\theta})}{\partial \mathbf{O}_2}\frac{\partial \mathbf{O}_2}{\partial \mathbf{W}_2} \\
&= \frac{\partial \ell(\boldsymbol{\theta})}{\partial \mathbf{O}_7}\frac{\partial \mathbf{O}_7}{\partial \mathbf{O}_6}\frac{\partial \mathbf{O}_6}{\partial \mathbf{O}_5}\frac{\partial \mathbf{O}_5}{\partial \mathbf{O}_4}\frac{\partial \mathbf{O}_4}{\partial \mathbf{O}_3}\frac{\partial \mathbf{O}_3}{\partial \mathbf{O}_2}\frac{\partial \mathbf{O}_2}{\partial \mathbf{W}_2} \\
&= \frac{\partial \ell(\boldsymbol{\theta})}{\partial \mathbf{O}_7}\mathbf{W}_7\varphi'(\mathbf{O}_6)\mathbf{W}_6\left(\prod_{i=1}^{3}\mathbf{W}_{6-i}\varphi'(\mathbf{O}_{5-i})\right)\frac{\partial \mathbf{O}_2}{\partial \mathbf{W}_2}
\end{aligned} \tag{6}$$

which is exactly the case in the HyperPNN1, and the multiple multiplication $(\prod_{i=1}^{3}\mathbf{W}_{6-i}\varphi'(\mathbf{O}_{5-i}))$ here implies that the gradient stack is prone to vanishing [27]–[29]. Thus, in comparison with the gradient of the HyperCNN2 in (6), the gradient of HyperCNN1 modeled in (5) tends to be lower. With the higher gradient, HyperPNN2, basically a component type CNN, is expected to achieve faster convergence than HyperPNN1.

Finally, for illustrative purposes, Fig. 3 demonstrates the training losses of our proposed HyperPNN1 and HyperPNN2, respectively, for an experiment over the Salinas dataset. Detailed description of this dataset will be introduced in the following experimental section. From Fig. 3, we can observe that HyperPNN2 has a smaller loss and converges fast, in comparison with the HyperPNN1, which is consistent with our expectation based on aforementioned analysis.

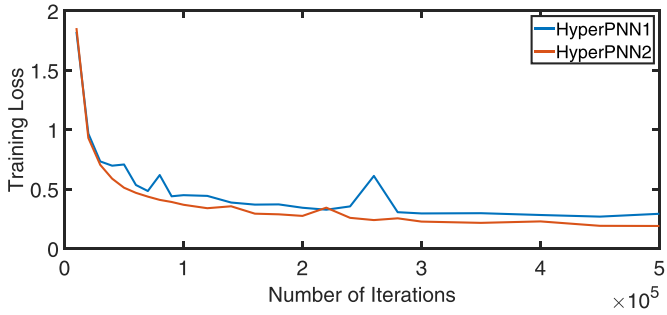

Fig. 3. Training losses of our HyperPNN1 and HyperPNN2 with the Salinas dataset.

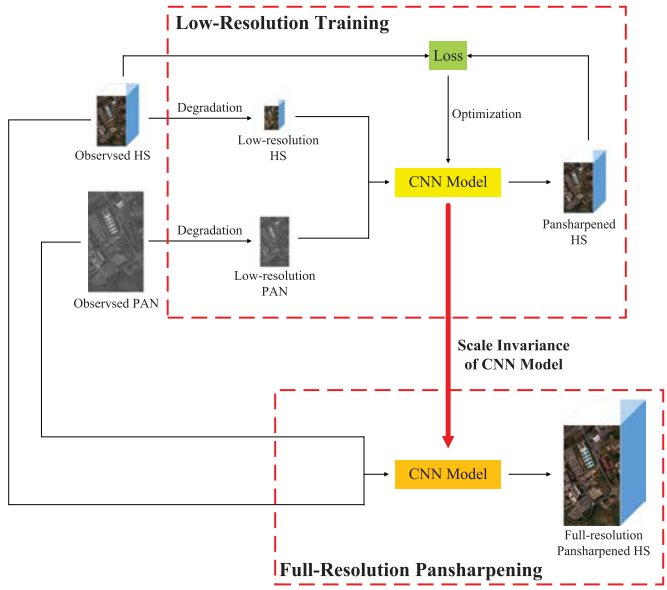

Fig. 4. Graphical illustration of HyperPNN implementation in a real scenario.

C. Real Scenario Implementation

In a real scenario, the implementation of our HyperPNN methods can be attributed to two major phases. The first stage is the low-resolution training phase, where the observed HS and PAN images are degraded to be low-resolution ones. Then, a pansharpening CNN model is trained, which involves the low-resolution HS and PAN images as its input and the original observed HS image as the reference. The second stage is the full-resolution pansharpening phase, where the pansharpening CNN model trained in the first phase is used to fuse the original observed HS and PAN images to yield the final full-resolution pansharpened HS image. Fig. 4 graphically illustrates the implementation procedure for our HyperPNN methods in the real scenario. The key to the feasibility of the implementation is that there exists a scale invariance for the pansharpening CNN model, which means that the pansharpening CNN learned in

the low-resolution level can be transferred to the full-resolution pansharpening.

IV. Experimental Results and Analysis

This section evaluates the performance of our proposed HS pansharpening methods, where three real-world remotely sensed image datasets are concerned, i.e., Washington DC Mall, Moffett Field, and Salina Scene. The Washington DC Mall dataset is acquired by the HS digital imagery collection experiment sensor which collects spectral signal with wavelength from 0.4 to 2.4 μm in 210 bands, while the Moffett Field and Salina Scene dataset are both obtained by the airborne visible/infrared imaging spectrometer sensor whose images cover spectra ranging from 0.4 to 2.5 μm with 224 bands. In the experiments, we remove noisy bands, with 191 bands for Washington DC Mall, 176 bands for Moffett Field, and 204 bands for Salina Scene remaining, respectively. Based on the Wald's protocol [30], the original HS image serves as the reference while the low-resolution one is degraded from the former by Gaussian blurring and downsampling. Moreover, the PAN image is simulated by averaging the visible spectral bands of the reference.

A. Experimental Setup

As aforementioned, our proposed model totally contains seven layers. Each of them includes 64 filters except for the top layer with B filters. The filter size setting is shown in Figs. 1 and 2 for HyperPNN1 and HyperPNN2, respectively. The adopted activation function is the rectified linear unit [31]. Specifically, for both methods, the kernel sizes for the spectral prediction sub-network and the spectral–spatial inference sub-network are the same. Furthermore, training samples are divided into 11 by 11 patches with the stride of 5 and 64 patches are regarded as a mini-batch for SGD. The training process is stopped at 5×10^5 iterations. Due to the fact that we empirically observe that with a number of 5×10^5 iterations, both HyperPNN1 and HyperPNN2 tend to converge, as shown in Fig. 3. The learning rate η in (2) is initially set to 0.0001 and it adaptively updates based on Adam[32]. We employed Caffe [33] to train our CNNs on a PC with a GPU (Nvidia GTX 1050Ti 4GB with CUDA 8.1 and CUDNN V5) and test them on MATLAB R2015a via CPU mode (computer with Intel I7 and 16 GB RAM) under the deep learning framework of Matconvnet [34]. The training process for each model cost roughly 6 h.

For the purposes of comparison, five state-of-the-art pansharpening methods are considered in our experiments, including SFIM[14], CNMF[17], Bayesian sparse [19], PNN[23], and DRPNN [24]. To objectively evaluate the performance of different HS pansharpening methods, four widely used quantitative pansharpening metrics are utilized, i.e., cross correlation (CC), spectral angle mapper (SAM), root-mean-squared error (RMSE), and erreur relative global adimensionnelle de synthese (ERGAS) [3]. Additionally, running time is also included in quantitative evalupation. CC, as a spatial index, characterizes the geometric distortion by the average of CCs for all bands. SAM is a spectral index which measures the spectral shape preservation by calculating the angle deviations for all pixels. RMSE and ERGAS both belong to global indexes. The former one measure the L_2 norm between the pansharpened image and the reference image, while the latter one is computed with the summation of RMSE values. The value of CC lies on the interval [0, 1]. The closer to 1 the index is, the higher the spatial quality of the pansharpened image is in. Lower SAM means higher spectral pansharpening quality. As for RMSE and ERGAS, they are expected to be lower, implying better pansharpening results. Besides quantitative evaluation, visual inspection is also conducted.

B. Experiment 1—Washington DC Mall Dataset

This dataset covers an urban area in Washington DC Mall. Fig. 5(a) exhibits the ground-truth HS image. The size of the degraded HS image is 256×60 and that of the PAN image is 1280×300. We choose a sub-scene with 256×128 pixels for test.

First, quantitative comparison is demonstrated in Table I. With the spectral and spatial pansharpening metrics being involved, i.e., CC, SAM, RMSE, and ERGAS, our HyperPNN2 yields the best performances under all cases while HyperPNN1 plays the role of the closest competitor except under CC. More specifically, CNMF has a little worse CC than other methods, whereas PNN and CNMF yield significantly inferior performances to others, consistently under SAM, RMSE, and ERGAS. Meanwhile, HyperPNN2 and HyperPNN1 are computationally efficient, ranking the second and the third places in terms of running time, respectively. The fact that HyperPNN2 runs faster than HyperPNN1 confirms the advantages of the component type CNN discussed in Section III-B. In a nutshsell, our proposed pansharpening methods achieve excellent performance here when evaluating quantitatively.

In addition to numeric metrics above for qualitative assessment, careful visual inspection is also needed, especially to identify artifacts and distortions that the quantitative analysis may fail to reveal. Fig. 5(b)–(h) show the pansharpened HS images by various methods, where two local square areas are enlarged in order to show the visual differences more clearly. By inspection on these figures, visual qualities can be identified. Our HyperPNN methods are the most similar to the ground truth, both in spatial detail and spectral fidelity. SFIM does well in spectral preservation but poor in spatial restoration, a little inconsistently with quantitative quality assessment. CNMF presents a significant spectral distortion, such as the road shown in the second enlarged box. Bayesian sparse incurs some artifacts, see the stripe ripple on the greens above the building. PNN provides image of low quality, with intense blurring, and obvious spectral distortion. DRPNN is also characterized by a little spectral distortion, such as the blue spots near the building in the first enlarged subimage and the spots on the roof of the square building in the second subimage.

C. Experiment 2—Moffett Field Dataset

This dataset covers a mixed urban/rural area in Moffett Field, California. Fig. 6(a) exhibits the ground-truth HS image. The size of the degraded HS image is 79×37 with 100 m resolution

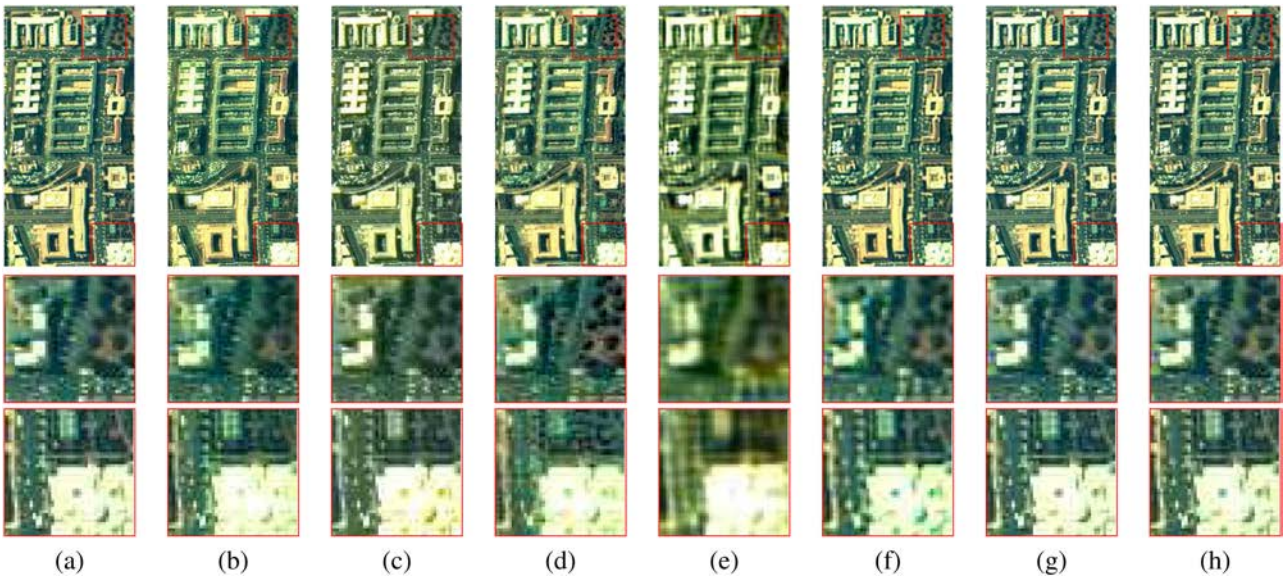

Fig. 5. HS pansharpening results for Washington DC Mall dataset: (a) Ground-truth; (b) SFIM; (c) CNMF; (d) Bayesian sparse; (e) PNN; (f) DRPNN; (g) HyperPNN1; (h) HyperPNN2.

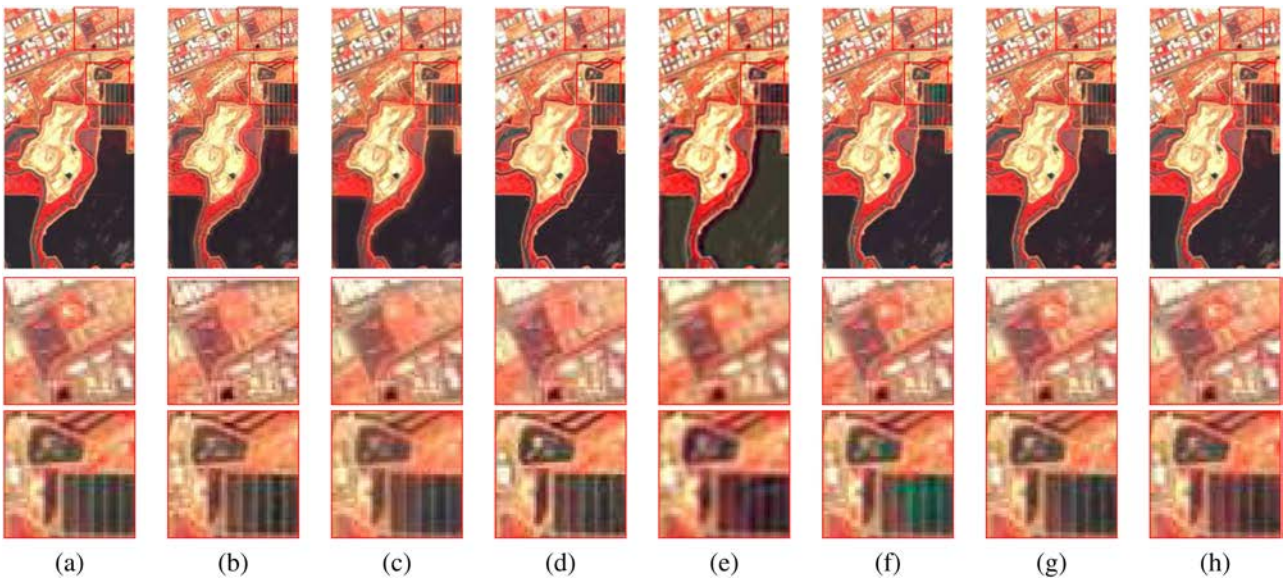

Fig. 6. HS pansharpening results for Moffett Field dataset: (a) Ground-truth; (b) SFIM; (c) CNMF; (d) Bayesian sparse; (e) PNN; (f) DRPNN; (g) HyperPNN1; (h) HyperPNN2.

TABLE I
QUALITY METRICS OF DIFFERENT HS PANSHARPENING METHODS ON WASHINGTON DC MALL DATASET

	CC	SAM	RMSE	ERGAS	Time(s)
Ground-truth	1	0	0	1	
SFIM	0.892	0.686	696.455	0.380	**4.366**
CNMF	0.793	1.082	976.476	0.570	23.913
Bayesian Sparse	0.866	0.559	423.272	0.236	20.908
PNN	0.816	1.034	864.141	0.480	12.908
DRPNN	0.811	0.517	465.855	0.259	9.104
HyperPNN1	0.842	0.400	344.122	0.191	5.669
HyperPNN2	**0.898**	**0.307**	**284.479**	**0.158**	4.569

and that of the PAN image is 395×185 with 20 m resolution. We choose a sub-scene with 256×128 pixels for test.

Table II lists the results of quantitative assessment on the Moffett Field. The superiority of our HyperPNN methods to other compared methods is significant under all the four pansharpening metrics. More specifically, HyperPNN1 takes the first place under CC, SAM, RMSE, and ERGAS while HyperPNN2 is the closest follower, where both of them gain overwhelming advantages especially under the latter three metrics. Moreover, both HyperPNN1 and HyperPNN2 can achieve high computation efficacy in terms of running time.

For visual inspection, Fig. 6(b)–(h) show the HS pansharpening images yielded by different methods. Similar to what was demonstrated on the previous dataset, our HyperPNN methods exhibit an excellent quality on Moffett Field, with comparable

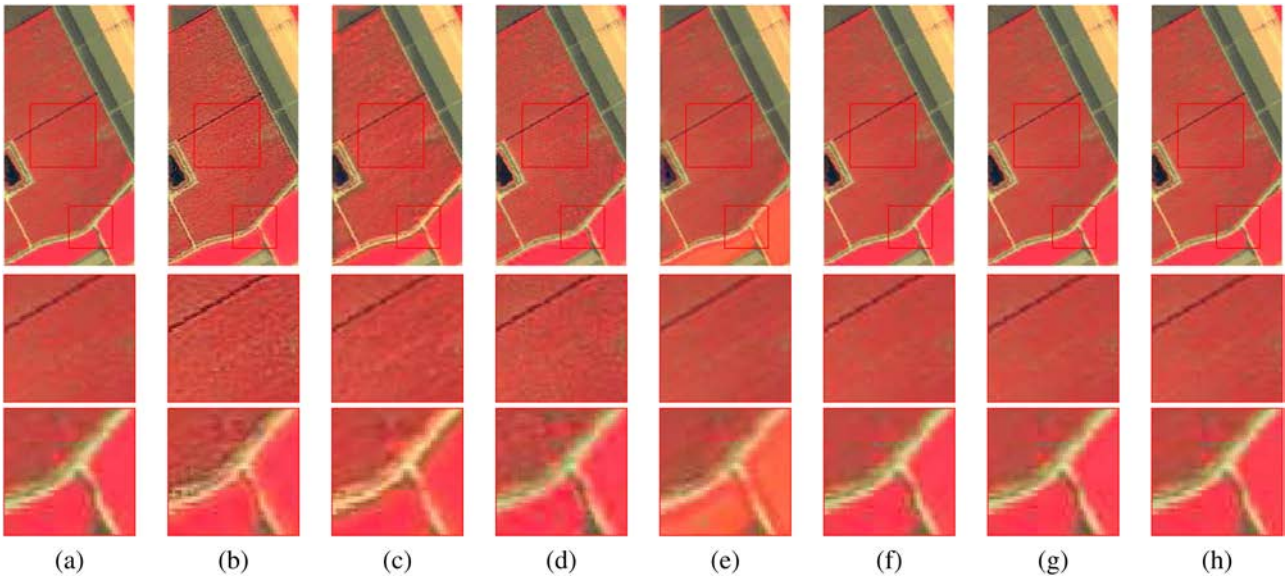

Fig. 7. HS pansharpening results for Salinas Scene dataset: (a) Ground-truth; (b) SFIM; (c) CNMF; (d) Bayesian sparse; (e) PNN; (f) DRPNN; (g) HyperPNN1; (h) HyperPNN2.

TABLE II
QUALITY METRICS OF DIFFERENT HS PANSHARPENING METHODS ON MOFFETT FIELD DATASET

	CC	SAM	RMSE	ERGAS	Time(s)
Ground-truth	1	0	0	1	
SFIM	0.971	6.336	266.495	3.775	**1.248**
CNMF	0.956	8.180	336.157	4.723	2.887
Bayesian Sparse	0.986	13.429	303.247	4.148	24.193
PNN	0.968	13.350	278.557	3.811	2.176
DRPNN	0.990	6.472	156.354	2.161	1.986
HyperPNN1	**0.993**	4.522	128.967	1.789	1.410
HyperPNN2	**0.993**	**4.391**	**125.472**	**1.747**	1.370

TABLE III
QUALITY METRICS OF DIFFERENT HS PANSHARPENING METHODS ON SALINAS SCENE DATASET

	CC	SAM	RMSE	ERGAS	Time(s)
Ground-truth	1	0	0	1	
SFIM	0.945	2.532	149.693	3.886	1.696
CNMF	0.915	2.789	217.299	2.794	5.562
Bayesian Sparse	0.957	4.487	157.358	6.294	29.908
PNN	0.857	2.599	235.884	18.084	3.564
DRPNN	0.980	1.049	55.405	2.753	2.658
HyperPNN1	0.983	0.995	60.851	1.321	1.607
HyperPNN2	**0.985**	**0.922**	**52.006**	**1.286**	**1.512**

spatial resolution and unnoticeable spectral distortions, especially as being visible in the enlarged boxes. SFIM, CNMF, and Bayesian sparse are characterized by some spectral distortions, while in the result from SFIM, some areas are over sharpened, with too much high-frequency details, for example in the bottom of the first enlarged boxes. PNN is unsatisfactory for the significant diffused blurring and a little lack of spectral fidelity. DRPNN suffers from severe spectral distortion, such as the water area shown in the lower part of the second enlarged subimage.

D. Experiment 3—Salinas Scene Dataset

This dataset covers a rural area in Salinas Valley, California. Fig. 7(a) exhibits the ground-truth HS image. The size of the degraded HS image is 102×43 and that of the PAN image is 510×215. We choose a sub-scene with 256×128 pixels for test.

Table III reports the results of quantitative assessment on the Salinas Scene. Similar to the quantitative results on the previous dataset, HyperPNN2 still achieves the highest scores in terms of all assessment metrics. HyperPNN1 also obtains very competitive results, ranking in the second place. It is noteworthy that our HyperPNN1 and HyperPNN2 run faster than all the other methods on this dataset, achieving very remarkable computation efficacy.

Visual inspection helps explaining these behaviors. Fig. 7(b)–(h) show the HS pansharpening results. Similar to previous visual results, our HyperPNN methods still attain the best performance in the Salinas Scene. SFIM, CNMF, and Bayesian sparse inject so abundant details that they incur over pansharpening, especially SFIM and Bayesian sparse, as seen in their first subimages. PNN looks subtle diffused spatial blurring and yield obvious spectral distortion, especially as shown in the second enlarged subimage, explaining clearly its poor performance of spectral preservation.

The above quantitative evaluations and visual inspections on three datasets verify the excellent performances of our proposed CNNs for HS pansharpening.

V. CONCLUSION

In this study, we have developed spectrally predictive CNN methods for HS pansharpening. HS pansharpening is faced with two major challenges: 1) the spectral range gap between the HS image and the PAN image; and 2) resolving details for many continuous narrow bands simultaneously. To deal with them, we developed the concept of HyperPNN and then accordingly

designed two specific networks, i.e., HyperPNN1 and HyperPNN2. Due to the inclusion of spectral predictive sub-network dedicated to learn along the spectral dimension, our networks are well suitable for the HS pansharpening task. Experimental results verify the remarkable performances of our HyperPNN methods.

REFERENCES

[1] G. Camps-Valls, D. Tuia, L. Bruzzone, and J. A. Benediktsson, "Advances in hyperspectral image classification: Earth monitoring with statistical learning methods," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 45–54, Jan. 2014.

[2] W.-K. Ma, J. M. Bioucas-Dias, J. Chanussot, and P. Gader, "Signal and image processing in hyperspectral remote sensing [from the guest editors]," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 22–23, Jan. 2014.

[3] L. Loncan *et al.*, "Hyperspectral pansharpening: A review," *IEEE Geosci. Remote Sens. Mag.*, vol. 3, no. 3, pp. 27–46, Sep. 2015.

[4] A. Mohammadzadeh, A. Tavakoli, and M. J. V. Zoej, "Road extraction based on fuzzy logic and mathematical morphology from pan-sharpened ikonos images," *Photogrammetric Rec.*, vol. 21, no. 113, pp. 44–60, 2006.

[5] G. A. Licciardi, A. Villa, M. M. Khan, and J. Chanussot, "Image fusion and spectral unmixing of hyperspectral images for spatial improvement of classification maps," in *Proc. Geosci. Remote Sens. Symp.*, 2012, pp. 7290–7293.

[6] H. Ghassemian, "A review of remote sensing image fusion methods," *Inf. Fusion*, vol. 32, pp. 75–89, 2016.

[7] G. Vivone *et al.*, "A critical comparison among pansharpening algorithms," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2565–2586, 2015.

[8] 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.

[9] V. K. Shettigara, "A generalized component substitution technique for spatial enhancement of multispectral images using a higher resolution data set," *Photogrammetric Eng. Remote Sens.*, vol. 58, no. 5, pp. 561–567, 1992.

[10] W. Liao *et al.*, "Processing of multiresolution thermal hyperspectral and digital color data: Outcome of the 2014 IEEE GRSS data fusion contest," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2984–2996, Jun. 2015.

[11] T. M. Tu, S. C. Su, H. C. Shyu, and P. S. Huang, "A new look at IHS-like image fusion methods," *Inf. Fusion*, vol. 2, no. 3, pp. 177–186, 2001.

[12] T. M. Tu, P. S. Huang, C. L. Hung, and C. P. Chang, "A fast intensity-hue-saturation fusion technique with spectral adjustment for IKONOS imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 1, no. 4, pp. 309–312, Oct. 2004.

[13] J. Qu, Y. Li, and W. Dong, "Hyperspectral pansharpening with guided filter," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 11, pp. 2152–2156, Nov. 2017.

[14] J. G. Liu, "Smoothing filter-based intensity modulation: A spectral preserve image fusion technique for improving spatial details," *Int. J. Remote Sens.*, vol. 21, no. 18, pp. 3461–3472, 2000.

[15] L. Alparone and B. Aiazzi, "MTF-tailored multiscale fusion of high-resolution MS and Pan imagery," *Photogrammetric Eng. Remote Sens.*, vol. 72, no. 5, pp. 591–596, 2006.

[16] G. Vivone, R. Restaino, and J. Chanussot, "A regression-based high-pass modulation pansharpening approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 56 no. 2, pp. 984–996, Feb. 2018.

[17] 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.

[18] Q. Wei, N. Dobigeon, J. Y. Tourneret, J. Bioucas-Dias, and S. Godsil, "R-FUSE: Robust fast fusion of multiband images based on solving a Sylvester equation," *IEEE Signal Process. Lett.*, vol. 23, no. 11, pp. 1632–1636, Nov. 2016.

[19] Q. Wei, N. Dobigeon, and J. Y. Tourneret, "Bayesian fusion of multi-band images," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 6, pp. 1117–1127, Sep. 2015.

[20] M. Simões, 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.

[21] 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.

[22] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. Comput. Vis. Pattern Recognit.*, 2016, pp. 1646–1654.

[23] G. Masi, D. Cozzolino, L. Verdoliva, and G. Scarpa, "Pansharpening by convolutional neural networks," *Remote Sens.*, vol. 8, no. 7, 2016, Art. no. 594.

[24] Y. Wei, Q. Yuan, H. Shen, and L. Zhang, "Boosting the accuracy of multispectral image pansharpening by learning a deep residual network," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1795–1799, Oct. 2017.

[25] Q. Yuan, Y. Wei, X. Meng, H. Shen, and L. Zhang, "A multiscale and multidepth convolutional neural network for remote sensing imagery pansharpening," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 3, pp. 978–989, Mar. 2018.

[26] G. Scarpa, S. Vitale, and D. Cozzolino, "Target-adaptive cnn-based pansharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 9, pp. 5443–5457, Sep. 2018.

[27] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 630–645.

[28] R. Pascanu, T. Mikolov, and Y. Bengio, "On the difficulty of training recurrent neural networks," in *Proc. Int. Conf. Mach. Learn.*, 2013, pp. 1310–1318.

[29] G. B. Goh, N. O. Hodas, and A. Vishnu, "Deep learning for computational chemistry," *J. Comput. Chem.*, vol. 38, no. 16, pp. 1291–1307, 2017.

[30] L. Wald, T. Ranchin, and M. Mangolini, "Fusion of satellite images of different spatial resolutions: Assessing the quality of resulting images," *Photogrammetric Eng. Remote Sens.*, vol. 63, no. 6, pp. 691–699, 1997.

[31] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. Int. Conf. Int. Conf. Mach. Learn.*, 2010, pp. 807–814.

[32] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.

[33] Y. Jia *et al.*, "Caffe: Convolutional architecture for fast feature embedding," in *Proc. 22nd ACM Int. Conf. Multimedia.*, 2014, pp. 675–678.

[34] A. Vedaldi and K. Lenc, "MatConvNet: Convolutional neural networks for MATLAB," in *Proc. 23rd ACM Int. Conf. Multimedia.*, 2015, pp. 689–692.

Lin He (S'05–M'12) received the B.S. degree from the Xi'an Institute of Technology, Xi'an, China, in 1995, the M.S. degree from Chongqing University, Chongqing, China, in 2003, both in instrumentation engineering, and the Ph.D. degree in pattern recognition and intelligent systems from Northwestern Polytechnical University, Xi'an, China, in 2007.

Since 2007, he has been with the School of Automation Science and Engineering, South China University of Technology, Guangzhou, China, where he is currently an Associate Professor. His current research interests include statistical pattern recognition, hyperspectral image processing, and high-dimensional signal processing.

Jiawei Zhu received the B.E. degree in automation in 2018 from the South China University of Technology, Guangzhou, China, where he is currently working toward the M.S. degree. His research interests include degraded information reconstruction for remote sensed images, data fusion, and computer vision.

Jun Li received the Geographical Information Systems degree from Hunan Normal University in 2004, the M.Sc. degree in remote sensing and photogrammetry from Peking University in 2007, and the Ph.D. degree in electrical and computer engineering from Instituto Superior Tecnico, Technical University of Lisbon, Lisbon, Portugal, in 2011.

From 2011 to 2012, she was a Postdoctoral Researcher with the Department of Technology of Computers and Communications, University of Extremadura, Badajoz, Spain. From 2014–2018, she was a Professor with the School of Geography and Planning, Sun Yat-sen University, Guangzhou, China. She is currently with the College of Electrical and Information Engineering, Hunan University, Changsha, China. Since then, she has obtained several prestigious funding grants at national and international level. She has authored and coauthored a total of 69 journal citation report (JCR) papers, 48 international conference papers, and a book chapter. She has received a significant number of citations to her published works, with several papers distinguished as "Highly Cited Papers" in Thomson Reuters' Web of Science - Essential Science Indicators (WoS-ESI). Her main research interests include remotely sensed hyperspectral image analysis, signal processing, supervised/semisupervised learning, and active learning.

Dr. Li has served as a Guest Editor of a Special Issue in the prestigious PROCEEDINGS OF THE IEEE, *ISPRS Journal of Photogrammetry*, and *Remote Sensing*. She has been serving as Associate Editor of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING (JSTARS) since 2014. Her students have also obtained important distinctions and awards at international conferences and symposia.

Antonio Plaza received the M.Sc. and Ph.D. degrees in computer engineering from the University of Extremadura, Badajoz, Spain, in 1999 and 2002, respectively.

He is the Head of the Hyperspectral Computing Laboratory at the Department of Technology of Computers and Communications, University of Extremadura. He has authored and coauthored more than 600 publications, including 200 JCR journal papers (145 in IEEE journals), 23 book chapters, and 285 peer-reviewed conference proceeding papers. He has guest-edited ten special issues on hyperspectral remote sensing for different journals. He has reviewed more than 500 manuscripts for over 50 different journals. His main research interests include hyperspectral data processing and parallel computing of remote sensing data.

Prof. Plaza was a recipient of the recognition of the Best Reviewers of the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS (in 2009) and the Best Reviewers of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING (in 2010), for which he served as Associate Editor during 2007–2012. He was also a recipient of the Best Column Award of the IEEE SIGNAL PROCESSING MAGAZINE in 2015, the 2013 Best Paper Award of the JSTARS journal, and the most highly cited paper (2005–2010) in the *Journal of Parallel and Distributed Computing*. He received Best Paper Awards at the IEEE International Conference on Space Technology and the IEEE Symposium on Signal Processing and Information Technology. He is also an Associate Editor for IEEE ACCESS, and was a member of the Editorial Board of the IEEE GEOSCIENCE AND REMOTE SENSING NEWSLETTER (2011–2012) and the IEEE GEOSCIENCE AND REMOTE SENSING MAGAZINE (2013). He was also a member of the steering committee of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING (JSTARS). He served as the Director of Education Activities for the IEEE Geoscience and Remote Sensing Society (GRSS) in 2011–2012, and as the President of the Spanish Chapter of IEEE GRSS in 2012–2016. He is currently serving as the Editor-in-Chief of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING journal. Additional information: http://www.umbc.edu/rssipl/people/aplaza.

Jocelyn Chanussot (M'04–SM'04–F'12) received the M.Sc. degree in electrical engineering from the Grenoble Institute of Technology (Grenoble INP), Grenoble, France, in 1995, and the Ph.D. degree from the Université de Savoie, Annecy, France, in 1998.

In 1999, he was with the Geography Imagery Perception Laboratory for the Delegation Generale de l'Armement (DGA - French National Defense Department). Since 1999, he has been with Grenoble INP, where he is currently a Professor of Signal and Image Processing. He is conducting his research at the Grenoble Images Speech Signals and Automatics Laboratory (GIPSA-Lab). His research interests include image analysis, multicomponent image processing, nonlinear filtering, data fusion, and machine learning in remote sensing. He has been a Visiting Scholar at Stanford University, Stanford, CA, USA, KTH Royal Institute of Technology, and National University of Singapore, Singapore. Since 2013, he has been an Adjunct Professor with the University of Iceland, Reykjavik, Iceland. During 2015–2017, he was a Visiting Professor with the University of California, Los Angeles, CA, USA.

Dr. Chanussot is the founding President of IEEE Geoscience and Remote Sensing French chapter (2007–2010), which received the 2010 IEEE Geoscience and Remote Sensing Society (GRSS) Chapter Excellence Award. He was the corecipient of the NORSIG 2006 Best Student Paper Award, the IEEE GRSS 2011 and 2015 Symposium Best Paper Award, the IEEE GRSS 2012 Transactions Prize Paper Award, and the IEEE GRSS 2013 Highest Impact Paper Award. He was the 2018 Highly Cited Researcher (Clarivate Analytics). He was the General Chair of the first IEEE GRSS Workshop on Hyperspectral Image and Signal Processing, Evolution in Remote sensing (WHISPERS). He was the Chair (2009–2011) and Co-Chair of the GRS Data Fusion Technical Committee (2005–2008) and the Program Chair of the IEEE International Workshop on Machine Learning for Signal Processing (2009). He is an Associate Editor for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING and the IEEE TRANSACTIONS ON IMAGE PROCESSING. He was the Editor-in-Chief of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING (2011–2015). In 2013, he was a Guest Editor for the PROCEEDINGS OF THE IEEE and in 2014, a Guest Editor for the IEEE SIGNAL PROCESSING MAGAZINE. He is a member of the IEEE Geoscience and Remote Sensing Society AdCom, Machine Learning for Signal Processing Technical Committee of the IEEE Signal Processing Society (2006–2008), and the Institut Universitaire de France (2012–2017).

Bo Li received the B.S. degree in computer science from Chongqing University, Chongqing, China, in 1986, the M.S. degree in computer science from Xian Jiaotong University, China, in 1989, and the Ph.D. degree in computer science from Beihang University, China, in 1993.

Since then, he joined the School of Computer Science and Engineering, Beihang University, Beijing, China. He is currently the Director of Beijing Key Laboratory of Digital Media, and the Director of the Professional Committee of Multimedia Technology of the China Computer Federation. He has authored and coauthored more than 100 academic papers in diverse research fields, including intelligent perception, data mining, remote sensing image fusion, and intelligent hardware.