

Spectral-Fidelity Convolutional Neural Networks for Hyperspectral Pansharpening

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

Abstract—Hyperspectral (HS) pansharpening aims at fusing a low-resolution HS (LRHS) image with a panchromatic image to obtain a full-resolution HS image. Most of the existing HS pansharpening approaches are usually based on traditional multispectral pansharpening techniques, which are not especially tailored for two inherent challenges of the HS pansharpening, i.e., much wider spectral range gap between the two kinds of images and having to recover details in many continuous spectral bands simultaneously. In this article, we develop new spectral-fidelity convolutional neural networks (called HSpEtnets) for HS pansharpening to keep the fidelity of a pansharpened image to its true spectra as much as possible. Our methods particularly focus on the decomposability of HS details, accordingly synthesizing these details progressively, and meanwhile introduce a spectral-fidelity loss. We give theoretical justifications and provide detailed experimental results, showing the superiorities of the proposed HSpEtnets with regard to other state-of-the-art pansharpening approaches.

Index Terms—Convolutional neural networks (CNNs), hierarchical detail reconstruction, hyperspectral image, pansharpening, spectral-fidelity loss.

I. INTRODUCTION

IMAGING spectroscopy captures the electromagnetic spectrum in many narrow contiguous bands, yielding hyperspectral (HS) images, which comprise both spatial and abundant spectral information from the observed scenes. Such a characteristic makes HS image data very valuable in many application

domains, receiving great attention from the remote sensing and image processing communities [1]–[5]. However, due to the physical limitations of sensors, improving the spectral resolution during image collection tends to negatively affect the spatial resolution. Therefore, HS images usually have lower spatial resolution when compared to panchromatic (PAN) images, which may not be sufficient in some practical applications where both high spatial and spectral resolutions are desired simultaneously [6]. One way to overcome this obstacle is to perform HS pansharpening, which is dedicated to generate high-spatial-resolution HS (HRHS) images by fusing low-resolution HS (LRHS) images with their PAN (high-resolution) counterparts [4], while traditional pansharpening, in contrast, merges multispectral (MS) images with PANs.

Existing pansharpening methods can be roughly classified into two categories: component substitution (CS) and multiresolution analysis (MRA). The CS category, which substitutes the spatial component of an HS image for that of the PAN image, involves methods such as principal component analysis [7]–[9], intensity-hue-saturation [10]–[12], Gram–Schmidt spectral sharpening [13], [14], or guided filtering [15]. The MRA group, which first extracts and then injects spatial details from the PAN images to the HS images in a multiresolution manner, includes approaches based on wavelet transform [16]–[18], smoothing filter-based intensity modulation (SFIM) [19], the modulation transfer function-generalized Laplacian pyramid [20], or regression-based high-pass modulation [21]. Apart from these classic MS pansharpening approaches, there are specific strategies designed for improving the spatial resolution of HS images. One of them is matrix factorization, which first models the observed data with a signal subspace representation and then leverages an optimization tool to factorize the related matrices [22]. The other is Bayesian inference, which utilizes Bayesian statistics to estimate probabilistic parameters used for pansharpening models [23]–[25].

Over recent years, convolutional neural networks (CNNs) have experienced significant success in tasks related to the enhancement of spatial resolution [26]–[36]. Dong *et al.* originally proposed a three-layer super-resolution CNN to learn the mapping between LR and HR images [26]. Kim *et al.* designed a very deep CNN structure with residual learning for the same problem [27]. Zhang *et al.* addressed the same topic by combining dense and residual networks to construct a deeper architecture [28]. CNNs have achieved good performance not only in single image super-resolution, but also in MS pansharpening.

Manuscript received May 26, 2020; revised August 10, 2020; accepted September 4, 2020. Date of publication September 18, 2020; date of current version October 8, 2020. This work was supported in part by the National Natural Science Foundation of China under Grant 62071184, Grant 61771496, and Grant 61571195 and in part by the Guangdong Provincial Natural Science Foundation under Grant 2016A030313254, Grant 2016A030313516, and Grant 2017A030313382. (Corresponding author: Jun Li.)

Lin He and Jiawei 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).

Jun 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).

Deyu Meng is with the School of Mathematics and Statistics and the Ministry of Education Key Laboratory of Intelligent Networks and Network Security, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: dymeng@mail.xjtu.edu.cn).

Jocelyn Chanussot is with GIPSA Lab, Grenoble Institute of Technology, Centre National de la Recherche Scientifique, Université Grenoble Alpes, 38000 Grenoble, France (e-mail: jocelyn.chanussot@grenoble-inp.fr).

Antonio Plaza is with the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, Escuela Politécnica, University of Extremadura, E-10071 Cáceres, Spain (e-mail: aplaza@unex.es).

Digital Object Identifier 10.1109/JSTARS.2020.3025040

For instance, Masi *et al.* developed a lightweight architecture for pansharpening, which uses preinterpolated low-resolution MS images (stacked with coregistered PAN images) as input [29]. Wei *et al.* adopted a CNN method with residual learning for pansharpening [31], as well as a more complex one with a multiscale kernel trick [32]. Scarpa *et al.* brought the target-adaptive modality into CNN design to further improve the network performance [33]. More recently, some researches have made attempt to apply CNN in HS pansharpening. He *et al.* adopted CNNs in HS pansharpening in an end-to-end manner, outperforming other traditional methods thanks to the use of spectral predictive structures [34]. Xie *et al.* treated HS pansharpening as a constrained minimization problem with additional priors learned by the CNN [35]. He *et al.* proposed a spectral constrained adversarial autoencoder to extract deep features of HS images, designed an adaptive fusion approach ruled by feature selection, and estimated HRHS images through solving an optimization equation [36].

Differently from super-resolution and MS pansharpening, HS pansharpening incurs inherent challenges, which is chiefly consisted of 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 reconstructing the whole details from the PAN image for all the bands at one time is a hard task. To deal with the challenging situation that the HS pansharpening has to be faced with, in this article, we present novel spectral-fidelity CNNs (called HSpEtnets), which intend to make the pansharpened HS image consistent with the true spectra as much as possible. The methods are developed with considerations in two aspects. On the one hand, our networks particularly focus on the decomposability of HS details and accordingly synthesize these details progressively to relief the difficulty to reconstruct a full-resolution HS image in one step; on the other hand, a spectral-fidelity loss is employed in the networks to further strengthen the spectral synthesis.

More specifically, the innovative contributions of this article can be summarized as follows.

- 1) We introduce two novel HSpEtnet methods, i.e., HSpEtnet1 and HSpEtnet2, for HS pansharpening. They hierarchically learn the HS details from the original data, while guided by a spectral-fidelity loss. Each HSpEtnet is composed of three successive subnetworks: feature extraction, feature fusion, and detail reconstruction, where some strategies such as two-stream feature extractors are additionally adopted to improve spectral prediction.
- 2) We propose a CNN-based unsupervised learning framework that can guide the design of HS pansharpening CNNs. The framework comprises structures that guarantee the fulfillment of an HS pansharpening task in an unsupervised manner. Moreover, it is able to enhance the spectral inferring to ensure a reliable reconstruction of the original image spectra.
- 3) We prove theoretically that the global structure of HSpEtnets is able to decrease the initial losses, while the

local structure of the HSpEtnets is capable of reducing gradient vanishing.

The remainder of this article is organized as follows. Section II presents detailed descriptions of the proposed methods, including a pansharpening framework and two HS pansharpening CNNs (called HSpEtnet1 and HSpEtnet2). Section III provides a detailed experimental validation of our methods. Section IV concludes this article with some remarks and hints at plausible future research lines.

II. PROPOSED METHODS

Our proposed HSpEtnets are dedicated to dealing with the main challenges of HS pansharpening, brought in by the special characteristics of HS images. For convenience, the following notations will be used throughout the following sections:

- $\mathbf{P}_o$: observed PAN image;
- $\mathbf{H}_o$: observed HS image;
- $\mathbf{P}$: low-resolution PAN (LRP) image;
- $\mathbf{H}'$: LRHS image;
- $\mathbf{H}$: interpolated LRHS image;
- $\mathbf{H}$: HRHS image; and
- $\mathbf{H}$: reconstructed HRHS image.

A. HS Pansharpening Framework

For HS pansharpening, a CNN receives the observed HS image and the PAN image as input and provides an HRHS image as output. The major mathematical formulation for the CNN can be generally given by

$$\mathbf{O}_l = f(\mathbf{W}_l * \mathbf{O}_{l-1} + \mathbf{b}_l) \quad (1)$$

where $*$ is the convolutional operation, $\mathbf{O}_l$ denotes the feature map (i.e., the convolutional output) of the l th layer, $\mathbf{W}_l$ and $\mathbf{b}_l$ represent the parameters of that layer, respectively, and $f(\cdot)$ stands for the activation function. During the CNN optimization, kernel parameters are updated under a given criterion, which is often the minimization of the mean square error (MSE)

$$\ell(\boldsymbol{\theta}) = E \|M(\cdot; \boldsymbol{\theta}) - \mathbf{H}\|_F^2 \quad (2)$$

where $E(\cdot)$ is the expectation, $\boldsymbol{\theta}$ represent the CNN parameters to be solved, $\|\cdot\|_F$ is the Frobenius norm, $\mathbf{H}$ denotes the target HRHS, and $M(\cdot)$ represents the mapping from the LRHS image to the HRHS image. Once the mapping $M(\cdot; \boldsymbol{\theta})$ is learned and the new observed PAN and HS images $\mathbf{P}'_o$ and $\mathbf{H}'_o$ are given, the pansharpened HS can be generated.

Compared to conventional MS pansharpening and single image super-resolution tasks, HS pansharpening confronts two inherent challenges. One is the wide gap between the spectral range of an HS image and that of the corresponding PAN image. The other is how to reconstruct the complicated details present in the many continuous (and narrow) bands of the HS image simultaneously. These challenges demand the HS pansharpening methodology to exhibit a strong capacity of spectral prediction and reconstruction; otherwise, some problems such as spectral distortion may be incurred.

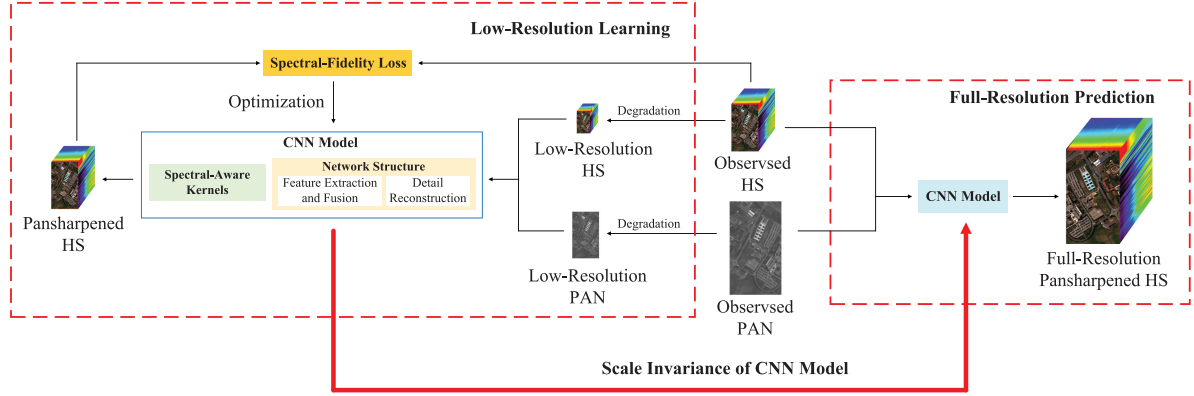


Fig. 1. HS pansharpening framework based on CNNs.

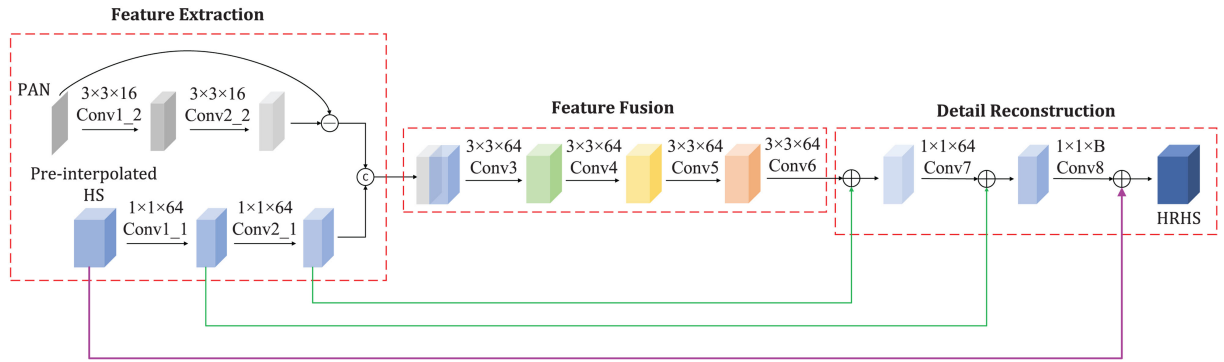


Fig. 2. Graphical illustration of the proposed HSPeNet1 method. Purple line represents global skip connections, while green lines represent local skip connections.

Fig. 1 graphically illustrates an HS pansharpening framework, which performs the HS pansharpening task in an unsupervised manner while accommodating the strengthening of spectral reconstruction. The following considerations need to be made about this general framework.

- 1) As shown in Fig. 1, an HS pansharpening process can be divided into two phases, i.e., low-resolution learning and full-resolution prediction. During the learning phase, the observed HS and PAN images are first degraded to low-resolution ones (i.e., LRHS and low-resolution PAN images). Then, the observed HS image is used as the ground-truth, and the loss between this ground truth and the CNN output drives the optimization of CNN parameters, yielding a CNN model. During the prediction phase, the new observed HS and PAN images go through the solved CNN to yield a final reconstructed HS image, taking advantage of the scale invariance of the CNN model from the low-resolution level to the full-resolution level.
- 2) As mentioned above, the HS pansharpening process has to face two significant dilemmas: the significant difference of spectral coverage between the HS and the PAN images, and the need to recover details in many continuous narrow bands simultaneously. To address these issues, HS pansharpening needs to further strengthen its ability to predict spectral characteristics while reconstructing spatial details. As indicated in Fig. 1, while suitable feature extraction/fusion needs to be used in order to improve the preservation of HS details, a spectral-fidelity loss and

spectral-aware convolution kernels are prioritized over spectral learning and inference.

Following the above HS pansharpening framework, we will develop two novel HS pansharpening CNNs in the subsections below, namely, HSPeNet1 and HSPeNet2.

B. HSPeNet1

Our HSPeNet1 is composed of three successive subnetworks: feature extraction subnetwork, feature fusion subnetwork, and detail reconstruction subnetwork (see Fig. 2).

1) *Feature Extraction Subnetwork*: Our HSPeNet1 has a two-stream feature extraction subnetwork. As mentioned before, there are significant differences between the HS and the PAN images in terms of spatial and spectral characteristics. HS pixels contain abundant spectral information but they are with lower spatial resolution, whereas PAN pixels are normally characterized by high spatial resolution but limited spectral information. Intuitively, we should cope with the two different image sources by adopting different strategies when extracting features from them. Here, we adopt a two-stream strategy for feature extraction, where one stream deals with the PAN image and the other works with the HS image, as shown in Fig. 2.

- 1) In the stream for the PAN image, a skip connection with subtraction operation, along with 3×3 convolutional kernels, is applied. Detail injection is one of fundamental principles for pansharpening, where spatial details are often obtained by the subtraction of low-frequency components

from the original image. Our new subtraction-involved skip connection can fulfill such a low-pass filtering in end-to-end learning fashion, thus producing reliable spatial details.

- 2) In the stream for the HS image, only 1×1 convolutional kernels are utilized on the raw images to generate pixel-wise features across different bands. Without the consideration of adjacent pixels, kernels are inclined to extract the inherent spectral characteristics of HS images.

Both the PAN stream and the HS stream can be regarded as spectral prediction (by band transform) operations; the former aims at spectral band generation and the latter performs spectral band redundancy reduction. After feature extractions in the PAN stream and HS stream, their features are concatenated along the spectral dimension and then are input to the succeeding feature fusion subnetwork.

2) *Feature Fusion Subnetwork*: The feature fusion subnetwork of HSpNet1 comprises four convolution layers (as can be seen in Fig. 2). Two streams of features from the feature extraction subnetwork are concatenated along spectral dimension and then as the input of this subnetwork. The convolution operation following the concatenation plays the role of weighted sum of the PAN and the HS sources. This process can be formulated with the convolutional output of the layer Conv3

$$\begin{aligned} \mathbf{O}_3 &= \mathbf{W}_3 * f(\mathbf{O}_2) + \mathbf{b}_3 \\ &= \mathbf{W}_3 * f([\mathbf{O}_{21}, \mathbf{O}_{22}]) + \mathbf{b}_3 \\ &= \mathbf{W}_{31} * f(\mathbf{O}_{21}) + \mathbf{W}_{32} * f(\mathbf{O}_{22}) + \mathbf{b}_3 \end{aligned} \quad (3)$$

where $\mathbf{O}_{21}$ and $\mathbf{O}_{22}$ denotes, respectively, the convolutional outputs of the layers Conv2_1 and Conv2_2, $[\mathbf{O}_{21}, \mathbf{O}_{22}]$ represents concatenating two feature cubes along the spectral dimension, and $\mathbf{W}_{31}$ and $\mathbf{W}_{32}$ are the weights connected to the two cubes. Different from directly pixelwise addition, concatenation fuses the input features with different weights, which is learned by the CNN automatically. After that in the layer Conv3, the fusion is further performed layerwise in the subsequent three layers, which can be mathematically represented as

$$\mathbf{O}_6 = \mathbf{W}_6 * f(\mathbf{W}_5 * f(\mathbf{W}_4 * f(\mathbf{O}_3) + \mathbf{b}_4) + \mathbf{b}_5) + \mathbf{b}_6. \quad (4)$$

3) *Detail Reconstruction Subnetwork*: The third subnetwork of HSpNet1 is the detail reconstruction subnetwork, as shown in Fig. 2. Three embedded skip connections are layerwise summed up therein, which is to recover the HS details progressively to relief the burden of reconstructing details at one time. In the final convolutional layer of the subnetwork, Conv8, no activation is involved, and HS details are output as

$$\mathbf{O}_8 = \mathbf{W}_8 * f(\mathbf{W}_7 * f(\mathbf{O}_6 + \mathbf{O}_{21}) + \mathbf{b}_7 + \mathbf{O}_{11}) + \mathbf{b}_8. \quad (5)$$

Then, the HRHS image is yielded from the summation of Conv8 output $\mathbf{O}_8$ and the interpolated LRHS image delivered by the outermost skip. The associated MSE loss is

$$\begin{aligned} \ell_1(\boldsymbol{\theta}) &= E\|M(\tilde{\mathbf{P}}, \tilde{\mathbf{H}}; \boldsymbol{\theta}) + \tilde{\mathbf{H}} - \mathbf{H}_o\|_F^2 \\ &= \frac{1}{N_p} \sum_{i=1}^{N_p} \|M(\tilde{\mathbf{P}}^{(i)}, \tilde{\mathbf{H}}^{(i)}; \boldsymbol{\theta}) + \tilde{\mathbf{H}}^{(i)} - \mathbf{H}_o^{(i)}\|_F^2 \end{aligned} \quad (6)$$

where i is the index of patches and N_p is the number of patches. Actually, if defining the HS details to recover as $\mathbf{D} = \mathbf{H}_o - \tilde{\mathbf{H}}$, (6) is then to be $\ell_1(\boldsymbol{\theta}) = E\|M(\tilde{\mathbf{P}}, \tilde{\mathbf{H}}; \boldsymbol{\theta}) - \mathbf{D}\|_F^2$. Therefore, by minimizing the loss (6) with a pansharpening training, HS details are learned by the network, which determines the CNN parameter set $\boldsymbol{\theta}$ and thus the estimation of HS details. Then, $\boldsymbol{\theta}$ can be transferred to the high-resolution scale under the postulate of the CNN model invariance, as guided in the aforementioned HS pansharpening framework shown in Fig. 2. It means that, if the observed data $\mathbf{P}_o$ and $\mathbf{H}_o$ are given and the mapping $M(\cdot)$ is obtained, the HRHS image can be reconstructed by

$$\hat{\mathbf{H}} = \mathbf{H}_o + M(\mathbf{P}_o, \mathbf{H}_o; \boldsymbol{\theta}). \quad (7)$$

For convenience, we use $\hat{\mathbf{D}}$ to represent the CNN learned details instead of $M(\tilde{\mathbf{P}}, \tilde{\mathbf{H}}; \boldsymbol{\theta})$ hereinafter, and then, the loss $\ell_1(\boldsymbol{\theta})$ can be rewritten as

$$\ell_1(\boldsymbol{\theta}) = E\|\hat{\mathbf{D}} + \tilde{\mathbf{H}} - \mathbf{H}_o\|_F^2. \quad (8)$$

Regarding the other two inner skip connections, they collaboratively learn certain kinds of details while helping to reduce the degradation of the network during the back-forward propagations [37], [38]. At this point, it is important to emphasize that 1×1 convolutional kernels are used in the progressive detail reconstruction to strengthen the spectral prediction capability.

4) *Spectral-Fidelity Loss*: MSE losses prevail in the design of CNN-based pansharpening methods. The MSE loss in (6) can (to some degree) measure the deviation of the pansharpened HS from the reference MS. However, as mentioned previously, HS pansharpening needs to strengthen the spectral prediction capability, and such an MSE alone may not be sufficient as a criterion to guide the CNN learning to relieve spectral distortion. In the field of HS image processing, the cosine similarity metric is often used as the spectral angle metric (SAM) to measure the similarity between HS pixels, which has been successful applied in many tasks such as unmixing (as exemplified in the following [39]–[41]). If the similarity is adopted between the ground-truth HS image and the pansharpened HS image, it can directly reflect spectral distortion existing in the pansharpened HS image. Thus, we introduce the cosine similarity metric to work collaboratively with MSE to form a spectral-fidelity loss. The cosine similarity metric is of the form

$$\begin{aligned} \ell_2(\boldsymbol{\theta}) &= E \sum_{i=1}^{HW} (1 - \cos \alpha_i) \\ &= E \sum_{i=1}^{HW} \left\{ 1 - \frac{(\hat{\mathbf{D}}_i + \tilde{\mathbf{H}}_i)^T \mathbf{H}_{oi}}{\|\hat{\mathbf{D}}_i + \tilde{\mathbf{H}}_i\| \|\mathbf{H}_{oi}\|} \right\} \\ &= \frac{1}{N_p} \sum_{j=1}^{N_p} \sum_{i=1}^{H_p W_p} \left(1 - \frac{(\hat{\mathbf{D}}_i^{(j)} + \tilde{\mathbf{H}}_i^{(j)})^T \mathbf{H}_{oi}^{(j)}}{\|\hat{\mathbf{D}}_i^{(j)} + \tilde{\mathbf{H}}_i^{(j)}\| \|\mathbf{H}_{oi}^{(j)}\|} \right) \end{aligned} \quad (9)$$

where H and W stand for the height and the width of the HS image, respectively; H_p and W_p denote the height and the width of a training patch, respectively; i and j are the indexes of pixels and patches, respectively; $\hat{\mathbf{D}}_i$, $\tilde{\mathbf{H}}_i$, and $\mathbf{H}_{oi}$ denote the i th pixel of $\hat{\mathbf{D}}$, $\tilde{\mathbf{H}}$, and $\mathbf{H}_o$, respectively; and $\alpha_i \in (-\pi, \pi]$ represents the

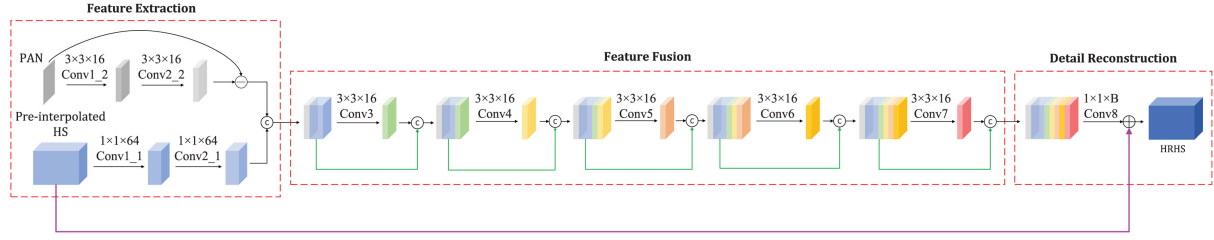


Fig. 3. Graphical illustration of the proposed HSpNet2 method. Purple line represents global skip connections, while green lines represent local skip connections.

angle between $(\hat{\mathbf{D}}_i + \tilde{\mathbf{H}}_i)$ and $\mathbf{H}_{oi}$. Besides, $\hat{\mathbf{D}}_i^{(j)}$, $\tilde{\mathbf{H}}_i^{(j)}$, and $\mathbf{H}_{oi}^{(j)}$ stand for the i th pixel of the j th patch of $\hat{\mathbf{D}}$, $\tilde{\mathbf{H}}$, and $\mathbf{H}_o$. The loss $\ell_2(\theta)$ can measure the cosine similarity between two pixels from different HS images. The more similar they are, the lower the absolute value of α is.

Leveraging the loss $\ell_1(\theta)$ in (8) and the loss $\ell_2(\theta)$ in (9), the final loss of our HSpNet1 is constructed as

$$\begin{aligned} \ell_3(\theta) &= \lambda \ell_1(\theta) + (1 - \lambda) \ell_2(\theta) \\ &= \lambda E \|\hat{\mathbf{D}} + \tilde{\mathbf{H}} - \mathbf{H}_o\|_F^2 \\ &\quad + (1 - \lambda) E \sum_{i=1}^{HW} \left\{ 1 - \frac{(\hat{\mathbf{D}}_i + \tilde{\mathbf{H}}_i)^T \mathbf{H}_{oi}}{\|\hat{\mathbf{D}}_i + \tilde{\mathbf{H}}_i\| \|\mathbf{H}_{oi}\|} \right\} \end{aligned} \quad (10)$$

which tends to be both stable and spectrally predictive, where λ stands for a penalty parameter and weighs the importance between the loss $\ell_1(\theta)$ the loss $\ell_2(\theta)$.

In brief, the proposed HSpNet1 mainly utilizes a hierarchical CNN structure, which can synthesize HS details progressively in a layerwise fashion and comprises three successive subnetworks, and a spectral-fidelity loss to ensure the spectral fidelity during the pansharpening.

C. HSpNet2

Apart from HSpNet1, we develop another spectral-fidelity HS pansharpening CNN, namely, HSpNet2. This network is also composed of three major parts: the feature extraction subnetwork, the feature fusion subnetwork, and the reconstruction subnetwork, as shown in Fig. 3.

The feature extraction subnetwork of HSpNet2 has a similar two-stream structure to that of HSpNet1. As regards its feature fusion subnetwork, it is different from that of HSpNet1. While that of HSpNet1 is flatten (i.e., the output of a convolutional layer is directly fed to the next layer as its input), that of HSpNet2 concatenates the output of a convolutional layer with its input as the input of the next layer, which can be formulated as

$$\begin{aligned} \mathbf{O}_3 &= \mathbf{W}_3 * f([\mathbf{O}_{21}, \mathbf{O}_{22}]) + \mathbf{b}_3 \\ &= \mathbf{W}_3 * f(\mathbf{O}_2) + \mathbf{b}_3 \\ \mathbf{O}_4 &= \mathbf{W}_4 * f([\mathbf{O}_3, \mathbf{O}_2]) + \mathbf{b}_4 \\ \mathbf{O}_5 &= \mathbf{W}_5 * f([\mathbf{O}_4, \mathbf{O}_3, \mathbf{O}_2]) + \mathbf{b}_5 \\ &\vdots \\ \mathbf{O}_7 &= \mathbf{W}_7 * f([\mathbf{O}_6, \mathbf{O}_5, \dots, \mathbf{O}_2]) + \mathbf{b}_7. \end{aligned} \quad (11)$$

This means that, in the fusion subnetwork, the input of a convolutional layer, while being delivered to subsequent layers with local skip connections (as seen in Fig. 3), is collaborated with the layer's output to drive the next layer to fulfill fusion.

In the detail reconstruction subnetwork of HSpNet2, the layer Conv8 serves as strengthening the spectral prediction to yield final HS details, which performed with a formulation similar to (5), and then adding the details to the LRHS image to generate the expected HRHS image. The detail reconstruction subnetwork explicitly contains only one convolutional layer with 1×1 kernels and only one skip connection to perform detail reconstruction. However, before the layer Conv8, details have been band groupwise reconstructed to some degree. This is because, in the fusion subnetwork, inputs of convolution layers are concatenated and delivered to the top layer Conv8, each of which is connected to the detail reconstruction on a group of spectral bands. In fact, the final HS details are synthesized with the collaboration of the fusion subnetwork and the detail reconstruction subnetwork. Similar to HSpNet1, HSpNet2 is a spectral-fidelity network, which comprises three successive subnetworks and is able to synthesize HS details progressively layer by layer. In addition, it can be connected to a spectral-fidelity loss.

To sum up, both HSpNet1 and HSpNet2 assume the decomposability of HS details and accordingly synthesize those details progressively, but they work in different manners. HSpNet1 reconstructs HS details from bottom level to top level, whereas HSpNet2 synthesizes those details in a manner of band groupwise reconstruction.

D. Theoretical Justification

Pansharpening CNNs can be seen as a kind of nonconvex optimization problem, which is usually dealt with by the iterative gradient descent method (where the initialization and the gradient are crucial for the final solution). Lower initial loss of a CNN implies that the initialization of the CNN tends to be closer to the optimal solution, while higher gradients tend to reduce gradient decay so as to help update the CNN parameters during backpropagation [42], [43], both of which are beneficial to getting better gradient descent solution. In the following, we justify our proposed HSpNets theoretically from such perspectives of the initial loss and the gradient.

Proposition 1: Under the loss in (10), the global skip connections in HSpNet1 and HSpNet2 (shown with the purple lines in Figs. 2 and 3) are able to achieve lower initial losses.

Proof: Let consider HSpNet1 first. As shown in (10), the loss $\ell_3(\theta)$ is composed of two losses $\ell_1(\theta)$ and $\ell_2(\theta)$. Here, $\ell_1(\theta)$ and $\ell_2(\theta)$ are considered separately. When the global skip connection is available, according to (8), we have

$$\begin{aligned}\ell_1(\theta) &= E(\|\hat{\mathbf{D}} + \tilde{\mathbf{H}} - \mathbf{H}_o\|_F^2) \\ &= E\{\text{Trace}(\hat{\mathbf{D}}\hat{\mathbf{D}}^T)\} + 2E\{\text{Trace}(\hat{\mathbf{D}}\tilde{\mathbf{H}}^T)\} \\ &\quad - 2E\{\text{Trace}(\hat{\mathbf{D}}\mathbf{H}_o^T)\} + E\{\text{Trace}(\tilde{\mathbf{H}}\tilde{\mathbf{H}}^T)\} \\ &\quad - 2E\{\text{Trace}(\tilde{\mathbf{H}}\mathbf{H}_o^T)\} + E\{\text{Trace}(\mathbf{H}_o\mathbf{H}_o^T)\} \\ &= \text{Trace}\{E(\hat{\mathbf{D}}\hat{\mathbf{D}}^T)\} + \text{Trace}\{E(\tilde{\mathbf{H}}\tilde{\mathbf{H}}^T)\} \\ &\quad - 2\text{Trace}\{E(\tilde{\mathbf{H}}\mathbf{H}_o^T)\} + \text{Trace}\{E(\mathbf{H}_o\mathbf{H}_o^T)\} \quad (12)\end{aligned}$$

where the initializations $E\{\text{Trace}(\hat{\mathbf{D}}\tilde{\mathbf{H}}^T)\} = 0$ and $E\{\text{Trace}(\hat{\mathbf{D}}\mathbf{H}_o^T)\} = 0$ are utilized, which are reasonable under the postulate that the CNN weight metric $\mathbf{W}_i$ and bias metric $\mathbf{b}_i$ are independent and initially follow a zero-mean distribution [30]. If the skip connection is removed, $\ell_1(\theta)$ is degenerated to

$$\begin{aligned}\ell'_1(\theta) &= E(\|\hat{\mathbf{H}} - \mathbf{H}_o\|_F^2) \\ &= E\{\text{Trace}(\hat{\mathbf{H}}\hat{\mathbf{H}}^T) - 2\text{Trace}(\hat{\mathbf{H}}\mathbf{H}_o^T) \\ &\quad + \text{Trace}(\mathbf{H}_o\mathbf{H}_o^T)\} \\ &= \text{Trace}\{E(\hat{\mathbf{H}}\hat{\mathbf{H}}^T)\} + \text{Trace}\{E(\mathbf{H}_o\mathbf{H}_o^T)\}. \quad (13)\end{aligned}$$

Then, we have

$$\begin{aligned}\ell_1(\theta) - \ell'_1(\theta) &= \text{Trace}\{E(\hat{\mathbf{D}}\hat{\mathbf{D}}^T)\} + \text{Trace}\{E(\tilde{\mathbf{H}}\tilde{\mathbf{H}}^T)\} \\ &\quad - 2\text{Trace}\{E(\tilde{\mathbf{H}}\mathbf{H}_o^T)\} - \text{Trace}\{E(\hat{\mathbf{H}}\hat{\mathbf{H}}^T)\} \\ &= \text{Trace}\{E(\tilde{\mathbf{H}}\tilde{\mathbf{H}}^T)\} - 2\text{Trace}\{E(\tilde{\mathbf{H}}(\mathbf{H}_o^T - \tilde{\mathbf{H}}^T + \hat{\mathbf{H}}^T))\} \\ &= -\text{Trace}\{E(\tilde{\mathbf{H}}\tilde{\mathbf{H}}^T)\} - 2\text{Trace}\{E(\tilde{\mathbf{H}}(\mathbf{H}_o^T - \tilde{\mathbf{H}}^T))\} \\ &= -\text{Trace}\{E(\tilde{\mathbf{H}}(\tilde{\mathbf{H}} + 2\hat{\mathbf{D}})^T)\}. \quad (14)\end{aligned}$$

Considering the term $\text{Trace}\{E(\tilde{\mathbf{H}}(\tilde{\mathbf{H}} + 2\hat{\mathbf{D}})^T)\}$, we obtain

$$\begin{aligned}\text{Trace}\{E((\tilde{\mathbf{H}} + 2\hat{\mathbf{D}})\tilde{\mathbf{H}}^T)\} &= \text{Trace}\{E((\tilde{\mathbf{H}} + \hat{\mathbf{D}})(\tilde{\mathbf{H}} + \hat{\mathbf{D}})^T)\} - \text{Trace}\{E(\hat{\mathbf{D}}\hat{\mathbf{D}}^T)\}. \quad (15)\end{aligned}$$

It is well known that the energy of details in an image generally comprises a very small fraction of the total energy, i.e.,

$$\|\tilde{\mathbf{H}}\|_F^2 \gg \|\hat{\mathbf{D}}\|_F^2. \quad (16)$$

That is, for most of the entries of $\tilde{\mathbf{H}}$ and $\hat{\mathbf{D}}$, there is $|h_{ij}| \gg |d_{ij}|$, where h_{ij} and d_{ij} are entries of $\tilde{\mathbf{H}}$ and $\hat{\mathbf{D}}$, respectively. Therefore, $(h_{ij} + d_{ij})^2 > d_{ij}^2$. Consequently, we have

$$\text{Trace}\{(\tilde{\mathbf{H}} + 2\hat{\mathbf{D}})\tilde{\mathbf{H}}^T\} > 0. \quad (17)$$

Therefore, each pair of connected entries of $\tilde{\mathbf{H}} + 2\hat{\mathbf{D}}$ and $\tilde{\mathbf{H}}$ tend to have the same sign (both positive or negative). According to

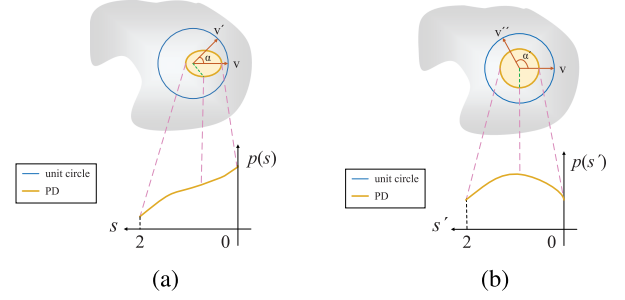


Fig. 4. PDFs of spectral-fidelity losses for (a) the case with the global skip connection and (b) the case without the skip connection.

(15) and (17), we can obtain

$$\ell_1(\theta) - \ell'_1(\theta) < 0. \quad (18)$$

That is

$$\ell_1(\theta) < \ell'_1(\theta). \quad (19)$$

Now, let us turn around to the loss $\ell_2(\theta)$ in (10). When the global skip connection is available, we have

$$\begin{aligned}\ell_{2i}(\theta) &= E\left\{1 - \frac{(\hat{\mathbf{D}}_i + \tilde{\mathbf{H}}_i)^T \mathbf{H}_{oi}}{\|\hat{\mathbf{D}}_i + \tilde{\mathbf{H}}_i\| \|\mathbf{H}_{oi}\|}\right\} \\ &= \frac{1}{2} E\left\{\frac{(\hat{\mathbf{D}}_i + \tilde{\mathbf{H}}_i)}{\|\hat{\mathbf{D}}_i + \tilde{\mathbf{H}}_i\|} - \frac{\mathbf{H}_{oi}}{\|\mathbf{H}_{oi}\|}\right\} \\ &= \frac{1}{2} E\left\{\|(\hat{\mathbf{D}}_i + \tilde{\mathbf{H}}_i)^{[n]} - \mathbf{H}_{oi}^{[n]}\|\right\} \quad (20)\end{aligned}$$

for the i th pixel, where $(\cdot)^{[n]}$ denotes the normalization of a vector. If there is no global skip connection available, the loss for the i th pixel is

$$\begin{aligned}\ell'_{2i}(\theta) &= E\left\{1 - \frac{(\hat{\mathbf{H}}_i)^T \mathbf{H}_{oi}}{\|\hat{\mathbf{H}}_i\| \|\mathbf{H}_{oi}\|}\right\} \\ &= \frac{1}{2} E\left\{\|\hat{\mathbf{H}}_i^{[n]} - \mathbf{H}_{oi}^{[n]}\|\right\}. \quad (21)\end{aligned}$$

For convenience, let $s = \|(\hat{\mathbf{D}}_i + \tilde{\mathbf{H}}_i)^{[n]} - \mathbf{H}_{oi}^{[n]}\|$ and $s' = \|\hat{\mathbf{H}}_i^{[n]} - \mathbf{H}_{oi}^{[n]}\|$. Then, we have

$$\ell_{2i}(\theta) = \frac{1}{2} E(s) = \frac{1}{2} \int_0^2 s p(s) ds \quad (22)$$

and

$$\ell'_{2i}(\theta) = \frac{1}{2} E(s') = \frac{1}{2} \int_0^2 s' p(s') ds'. \quad (23)$$

As $\tilde{\mathbf{H}}$ is an LRMS variant of $\mathbf{H}_o$, while $\hat{\mathbf{H}}$ and $\hat{\mathbf{D}}$ are randomized by the (randomly initialized) CNN kernel, $(\hat{\mathbf{D}}_i + \tilde{\mathbf{H}}_i)$ is significantly more correlated with $\mathbf{H}$ than $\hat{\mathbf{H}}_i$, which implies that $(\hat{\mathbf{D}}_i + \tilde{\mathbf{H}}_i)^{[n]}$ tends to be distributed closer to $\mathbf{H}_{oi}^{[n]}$ than $\hat{\mathbf{H}}_i^{[n]}$ does. This means that the angle between $(\hat{\mathbf{D}}_i + \tilde{\mathbf{H}}_i)^{[n]}$ and $\mathbf{H}_{oi}^{[n]}$ is more closely distributed around zero than the angle between $\hat{\mathbf{H}}_i^{[n]}$ and $\mathbf{H}_{oi}^{[n]}$ [as shown in the upper part of Fig. 4(a) and (b)], causing $\|(\hat{\mathbf{D}}_i + \tilde{\mathbf{H}}_i)^{[n]} - \mathbf{H}_{oi}^{[n]}\|$ (i.e., s) to have a

probability density (PD) closer to zero than $\|\hat{\mathbf{H}}_i^{[n]} - \mathbf{H}_{oi}^{[n]}\|$ (i.e., s') [as shown in the lower part of Fig. 4(a) and (b)]. Therefore, according to (22) and (23), we can obtain

$$\ell_{2i}(\boldsymbol{\theta}) < \ell'_{2i}(\boldsymbol{\theta}) \quad (24)$$

and accordingly

$$\ell_2(\boldsymbol{\theta}) < \ell'_2(\boldsymbol{\theta}). \quad (25)$$

Combining (19) with (25), we obtain $\{\ell_1(\boldsymbol{\theta}) + \ell_2(\boldsymbol{\theta})\} < \{\ell'_1(\boldsymbol{\theta}) + \ell'_2(\boldsymbol{\theta})\}$. That is

$$\ell(\boldsymbol{\theta}) < \ell'(\boldsymbol{\theta}). \quad (26)$$

The inequality in (26) means that the global skip connection introduced by our HSpNet1 can lead to a lower loss in the initial step of its propagation process, which implies that the network tends to achieve better convergence. The justification of HSpNet2 can be built in a similar way. ■

Proposition 2: The local skip connections in HSpNet1 and HSpNet2 (shown with green lines in Figs. 2 and 3) are able to reduce the gradient decay of their bottom layers.

Proof: Consider HSpNet2 first. The skip connections in HSpNet2 end in the form of a concatenation, rather than an elementwise summation. According to the structure of HSpNet2 (see Fig. 3), we have

$$\begin{aligned} \mathbf{O}_3 &= \mathbf{W}_3 * f([\mathbf{O}_{21}, \mathbf{O}_{22}]) + \mathbf{b}_3 \\ \mathbf{O}_4 &= \mathbf{W}_4 * f([\mathbf{O}_3, \mathbf{O}_{21}, \mathbf{O}_{22}]) + \mathbf{b}_4 \\ \mathbf{O}_5 &= \mathbf{W}_5 * f([\mathbf{O}_4, \mathbf{O}_3, \mathbf{O}_{21}, \mathbf{O}_{22}]) + \mathbf{b}_5 \\ &\vdots \\ \mathbf{O}_8 &= \mathbf{W}_8 * f([\mathbf{O}_7, \mathbf{O}_6, \dots, \mathbf{O}_{21}, \mathbf{O}_{22}]) + \mathbf{b}_8. \end{aligned} \quad (27)$$

Consider the layer Conv5. Its output $\mathbf{O}_5$ contains the output of Conv4 $\mathbf{O}_4$ and its input $[\mathbf{O}_3, \mathbf{O}_{21}, \mathbf{O}_{22}]$. Therefore, $\mathbf{O}_{21}$ and $\mathbf{O}_{22}$ directly work on $\mathbf{O}_5$. $\mathbf{O}_5$ indicated in (27) can be rewritten as

$$\begin{aligned} \mathbf{O}_5 &= \mathbf{W}_{51} * f(\mathbf{O}_4) + \mathbf{W}_{52} * f(\mathbf{O}_3) + \mathbf{W}_{53} * f(\mathbf{O}_{21}) \\ &\quad + \mathbf{W}_{54} * f(\mathbf{O}_{22}) + \mathbf{b}_5. \end{aligned} \quad (28)$$

Considering the gradient of the layer Conv2_1, we have

$$\begin{aligned} &\frac{\partial \ell(\boldsymbol{\theta})}{\partial \mathbf{W}_{21}} \\ &= \frac{\partial \ell(\boldsymbol{\theta})}{\partial \mathbf{O}_{21}} \frac{\partial \mathbf{O}_{21}}{\partial \mathbf{W}_{21}} \\ &= \frac{\partial \ell(\boldsymbol{\theta})}{\partial \mathbf{O}_5} \frac{\partial \mathbf{O}_5}{\partial \mathbf{O}_{21}} \frac{\partial \mathbf{O}_{21}}{\partial \mathbf{W}_{21}} \\ &= \frac{\partial \ell(\boldsymbol{\theta})}{\partial \mathbf{O}_5} \left(\frac{\partial \mathbf{O}_5}{\partial \mathbf{O}_4} \frac{\partial \mathbf{O}_4}{\partial \mathbf{O}_{21}} + \frac{\partial \mathbf{O}_5}{\partial \mathbf{O}_3} \frac{\partial \mathbf{O}_3}{\partial \mathbf{O}_{21}} + \frac{\partial \mathbf{O}_5}{\partial \mathbf{O}_{21}} \right) \frac{\partial \mathbf{O}_{21}}{\partial \mathbf{W}_{21}} \\ &= \frac{\partial \ell(\boldsymbol{\theta})}{\partial \mathbf{O}_5} \left[\mathbf{W}_{51} f'(\mathbf{O}_4) \frac{\partial \mathbf{O}_4}{\partial \mathbf{O}_{21}} + \mathbf{W}_{52} f'(\mathbf{O}_3) \frac{\partial \mathbf{O}_3}{\partial \mathbf{O}_{21}} \right. \\ &\quad \left. + \mathbf{W}_{53} f'(\mathbf{O}_{21}) \right] \frac{\partial \mathbf{O}_{21}}{\partial \mathbf{W}_{21}}. \end{aligned} \quad (29)$$

As displayed in (29), the gradient of $\mathbf{W}_{21}$ is composed of $\frac{\partial \ell(\boldsymbol{\theta})}{\partial \mathbf{O}_5}$ and $\frac{\partial \mathbf{O}_5}{\partial \mathbf{W}_{21}}$. Moreover, the latter one is aggregated with three units, i.e., $\frac{\partial \mathbf{O}_5}{\partial \mathbf{O}_4} \frac{\partial \mathbf{O}_4}{\partial \mathbf{O}_{21}}$, $\frac{\partial \mathbf{O}_5}{\partial \mathbf{O}_3} \frac{\partial \mathbf{O}_3}{\partial \mathbf{O}_{21}}$, and $\frac{\partial \mathbf{O}_5}{\partial \mathbf{O}_{21}}$. Herein, $\frac{\partial \mathbf{O}_3}{\partial \mathbf{O}_{21}}$ can be easily solved and $\frac{\partial \mathbf{O}_4}{\partial \mathbf{O}_{21}}$ can be computed as

$$\begin{aligned} \frac{\partial \mathbf{O}_4}{\partial \mathbf{O}_{21}} &= \frac{\partial \mathbf{O}_4}{\partial \mathbf{O}_3} \frac{\partial \mathbf{O}_3}{\partial \mathbf{O}_{21}} + \frac{\partial \mathbf{O}_4}{\partial \mathbf{O}_{21}} \\ &= \mathbf{W}_{41} f'(\mathbf{O}_3) \frac{\partial \mathbf{O}_3}{\partial \mathbf{O}_{21}} + \mathbf{W}_{42} f'(\mathbf{O}_{21}). \end{aligned} \quad (30)$$

By combining (29) with (30), we find that the derivative from $\mathbf{O}_5$ to $\mathbf{O}_{21}$ consists of four units, corresponding to four paths, from $\mathbf{O}_{21}$ to $\mathbf{O}_5$, respectively. Similarly, when considering the direct derivative from $\mathbf{O}_8$ to $\mathbf{O}_{21}$, there are totally five skip connections. Thus, $\mathbf{O}_{21}$ is able to convey information to $\mathbf{O}_8$ through 2^5 paths. In other words, the gradient of $\mathbf{O}_{21}$ is summed by 2^5 units. This makes the gradient of our network tend to higher values than the gradient of a network with less skip connections, which subsequently enables the gradient of $\mathbf{O}_{11}$ to be higher. Therefore, the local skip connections in our HSpNet2 can significantly alleviate the gradient decay of its bottom layers. Similarly, we can easily have the justification of HSpNet1. ■

III. EXPERIMENTAL RESULTS

A. Experimental Setup

In this section, we test our newly proposed HS pansharpening method on four datasets. Table I summarizes their properties.

- 1) The Washington DC Mall dataset was gathered by the Hyperspectral Digital Imagery Collection Experiment (HYDICE) sensor, which collects data in the spectral range from 0.4 to 2.4 μm using 210 bands. After discarding the spectral bands with low signal-to-noise ratio (SNR), 191 bands were retained for subsequent analyses.
- 2) The Moffett Field dataset was acquired by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor, which collects data in the spectral range from 0.4 to 2.5 μm using 224 bands (176 bands were retained after discarding low SNR bands).
- 3) The Pavia University dataset was collected by the Reflective Optics System Imaging sensor (ROSIS) which records data in the spectral range from 0.4 to 0.9 μm using 115 bands (103 bands were retained).
- 4) The Halls Creek dataset was collected by the Earth Observing-1 (EO-1) Hyperion satellite, which gathers data in the spectral range from 0.4 to 2.5 μm using 242 bands (171 bands were retained).

For the purpose of comparison, four state-of-the-art HS pansharpening methods were considered in our experiments, including SFIM [19], CNMF [22], GFPCA [8], BayeSparse [24], and HyperPNN [34]. Additionally, we also applied representative deep learning-based pansharpening methods to the HS case, including PNN [29] and DRPNN [31].

At this point, it is important to note that quantitative assessment of pansharpening with HS images is a complicated task, due to a lack of references in reality [4]. In order to evaluate

TABLE I
PROPERTIES OF THE DATASETS

Datasets	Dimensions	Spectral Range	Bands Number	Instrument
Washington DC Mall	PAN 1200×300	0.4–2.4 μm	210	HYDICE
	HS 240×60			
Moffett Field	PAN 395×185	0.4–2.5 μm	224	AVIRIS
	HS 79×37			
Pavia University	PAN 610×340	0.4–0.9 μm	115	ROSIS
	HS 122×68			
Halls Creek	PAN 3483×567	0.4–2.5 μm	242	EO-1
	HS 1161×189			

different approaches versus a common reference, a feasible strategy is to utilize semisynthetic HS and PAN images. In this case, a real HS image serves as reference. The HS and PAN images to be processed are obtained by degrading the references using Wald’s protocol [44]. Generally, an LRHS image can be attained by Gaussian blurring and downsampling, while the PAN image can be simulated by averaging the visible spectral bands of the reference. In addition, the LRHS image is preinterpolated to the same spatial size as the PAN image using a polynomial kernel [45].

Supported by reference images, several quantitative assessments can be conducted to intercompare different pansharpening methods. In this article, we consider four of the most widely used metrics, namely: cross-correlation (CC), SAM, root mean squared error (RMSE), and *Erreur Relative Globale Adimensionnelle de Synthèse* (ERGAS) [4]. CC, as a spatial index, characterizes the geometric distortion by the average CC for each single-band image. SAM, as a spectral index, measures the spectral shape preservation by calculating the angle of each pixel across different bands. RMSE and ERGAS, as global indexes, measure the l_2 norm between the estimated and reference images (the latter is composed of a sum of RMSE values).

B. Parameter Setting

As mentioned before, our newly proposed model contains eight layers in total. The settings are shown in Figs. 2 and 3. Specifically, B denotes the total number of bands in the HS images. The adopted activation function, rectified linear unit [46], is used in each convolutional layer. The weight in loss function λ was set to the same value for all the training procedures: $\lambda = 0.9$ (a detailed analysis of the impact of parameter λ was also conducted). Furthermore, the training samples are divided into 11-by-11 patches, with stride 5, and 64 patches are regarded as a mini-batch for stochastic gradient descent. The training process terminates at 4000 epochs. The basic learning rate is initially set to 0.0001, and it is adaptively updated based on Adam’s optimizer [47]. We employed the *Tensorflow* framework to train the proposed HSpEtnets on a PC with 16-GB RAM, an Intel core i5 8400 CPU, and an NVIDIA GeForce GTX 1060 GPU.

C. Simulated Image Experiments

The HYDICE Washington DC Mall, AVIRIS Moffett Field, and ROSIS Pavia University are all simulated datasets. For these datasets, the PAN image and the LRHS image are simulated as mentioned before. The original HS image serves as a

TABLE II
QUALITY METRICS OBTAINED FOR DIFFERENT HS PANSHARPENING METHODS APPLIED TO THE HYDICE WASHINGTON DC MALL DATASET

	CC	SAM	RMSE	ERGAS
Ground-truth	1	0	0	0
Interpolation	0.811	7.664	789.893	6.742
SFIM	0.916	7.260	576.606	4.785
GFPCA	0.806	9.785	869.484	7.542
CNMF	0.886	8.953	639.456	5.637
BayeSparse	0.930	6.689	475.723	4.390
PNN	0.930	5.376	396.549	4.517
DRPNN	0.944	<u>4.857</u>	388.755	<u>4.006</u>
HyperPNN1	0.923	5.247	420.351	5.660
HyperPNN2	0.923	5.286	444.015	4.993
HSpEtnet1	0.953	4.518	374.260	3.667
HSpEtnet2	0.945	4.518	384.406	4.029

high-resolution reference, aimed at evaluating the performance of the fused results.

1) *HYDICE Washington DC Mall Dataset*: This dataset covers an urban area in Washington DC Mall. The size of the degraded HS image is 240×60 pixels, and that of the PAN image is 1200×300 . We choose a subscene with 200×200 pixels for testing and another part of 1200×100 pixels for training.

Table II shows a quantitative assessment of different methods on the HYDICE Washington DC Mall dataset. The best results are shown in bold, and the second - best results are underlined. As shown in Table II, CNN-based methods outperform the conventional ones. Among them, HSpEtnet1 exhibits a large performance gain with respect to all other methods, regardless of the considered metric, while HSpEtnet2 takes the role of the closest challenger.

In addition to the aforementioned numeric assessments, visual inspection is also important to identify artifacts and distortions that the quantitative analysis may fail to reveal. Fig. 5(b)–(l) presents the HS pansharpening results yielded by different methods. Among them, the pansharpened images produced by our HSpEtnets look visually very similar to the ground-truth, with comparable spatial resolution and without noticeable artifacts or spectral distortions, such as the area depicted in the enlarged box. GFPCA suffers from strong spatial blurring. SFIM and CNMF perform slightly better in spatial terms, but still worse in spectral preservation. BayeSparse, PNN, DRPNN, and HyperPNNs are characterized by a subtle diffused blurring.

The spectral reflectivity plays an important role in HSI applications, including classification, object detection, and unmixing [4], due to the physical properties of different ground objects. To further evaluate the spectral preservation capability

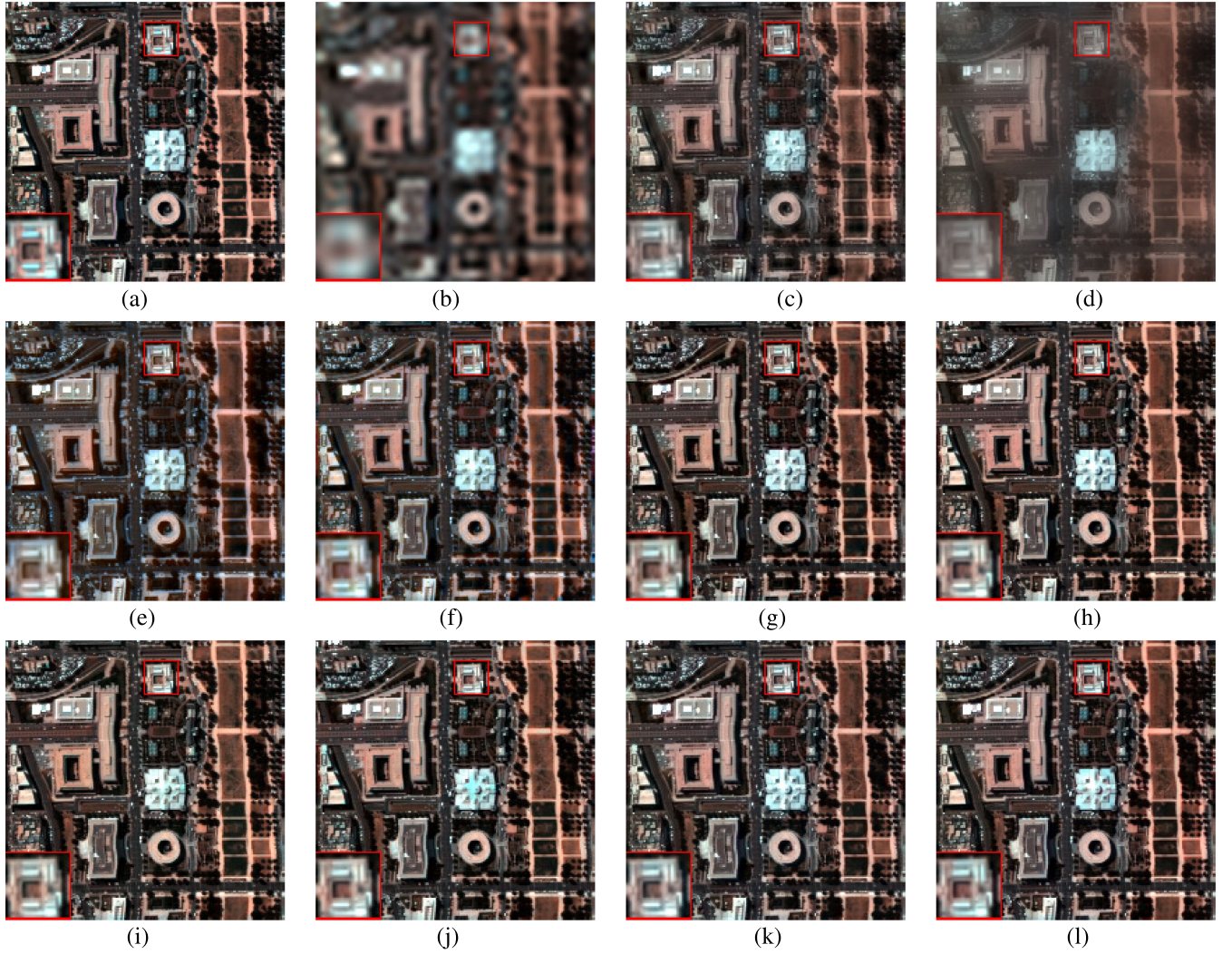


Fig. 5. HS pansharpening results for the HYDICE Washington DC Mall dataset (displayed in RGB). (a) Ground-truth. (b) Interpolation. (c) SFIM. (d) GFPCA. (e) CNMF. (f) BayeSparse. (g) PNN. (h) DRPNN. (i) HyperPNN1. (j) HyperPNN2. (k) HSpNet1. (l) HSpNet2

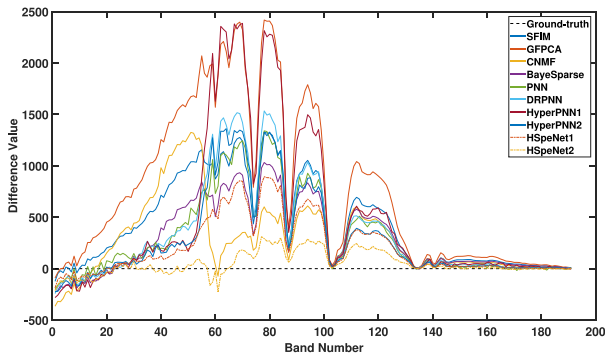


Fig. 6. Difference between the ground-truth spectrum of the pixel with spatial coordinates (160,48) and the HS pansharpening results in the same pixel for the HYDICE Washington DC Mall dataset.

of different methods, Fig. 6 shows the spectral difference curves between the ground-truth spectrum and the HS pansharpening results in a selected random pixel of the HYDICE Washington DC Mall dataset. As displayed in Fig. 6(a), our HSpNets

improves other methods in terms of spectral preservation, with a spectral difference value closer to zero.

2) *AVIRIS Moffett Field Dataset*: This dataset covers a mixed urban/rural area in Moffett Field, California. The size of the degraded HS image is 79×37 pixels, with 100-m resolution, and the size of the PAN image is 395×185 pixels, with 20-m resolution. We choose a subscene with 150×150 pixels for testing and another part of 245×185 pixels for training.

Table III shows the quantitative results obtained for different methods when applied to the AVIRIS Moffett Field dataset. As we can see in the table, our HSpNets outperform all other methods, while CNN-based methods and BayeSparse are the closest followers. Relatively, GFPCA lags behind others and even performs worse than the bicubic-interpolated method.

Fig. 7(b)–(l) presents the HS pansharpening images generated by various methods for the AVIRIS Moffett Field dataset. Our HSpNet1 and HSpNet2 also achieve excellent HS performance on this dataset. GFPCA suffers from significant diffused blurring. SFIM, CNMF, and BayeSparse are characterized by some spectral distortions, as the enlarged box reveals. PNN,

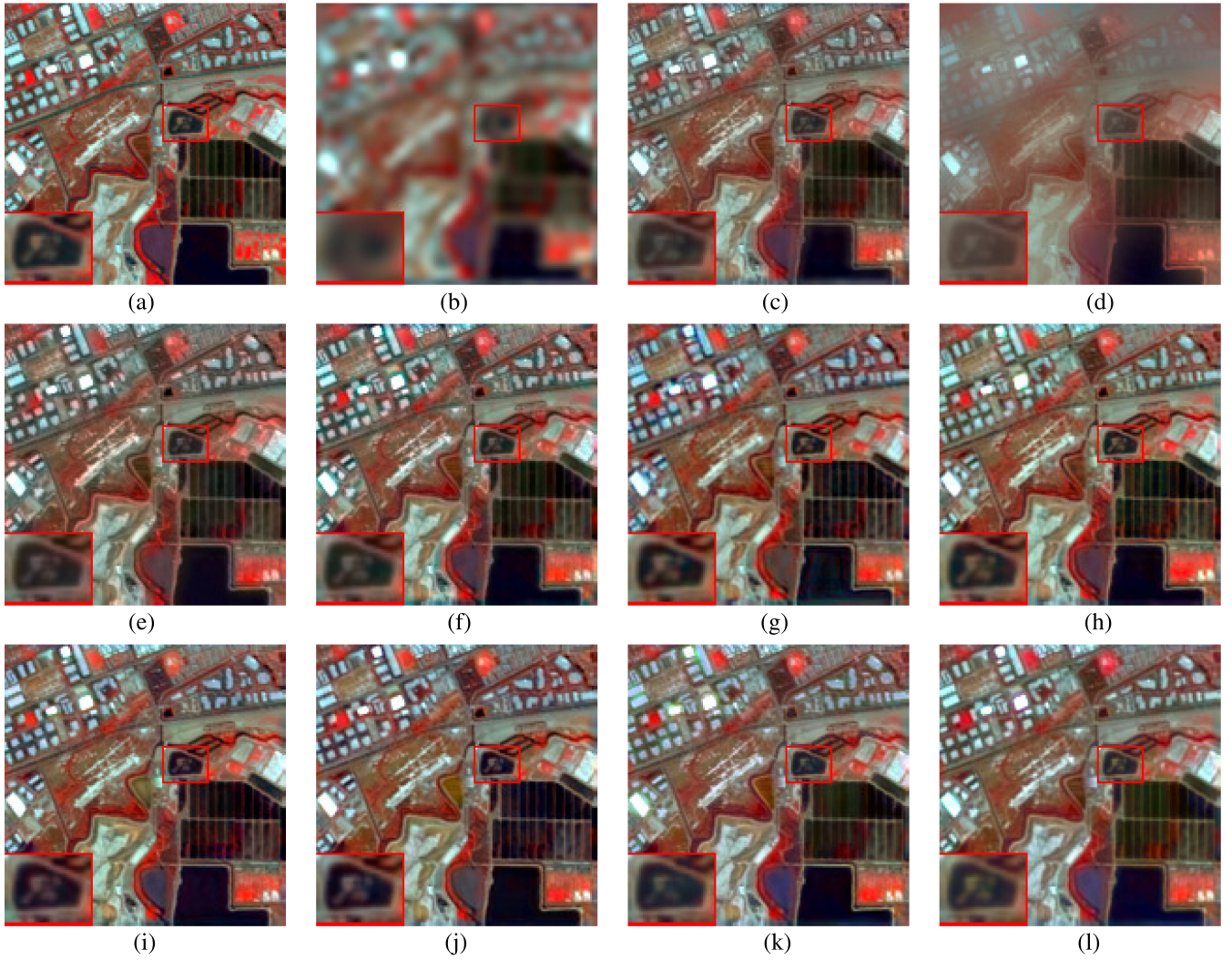


Fig. 7. HS pansharpening results for the AVIRIS Moffett Field dataset (displayed in RGB). (a) Ground-truth. (b) Interpolation. (c) SFIM. (d) GFPCA. (e) CNMF. (f) BayeSparse. (g) PNN. (h) DRPNN. (i) HyperPNN1. (j) HyperPNN2. (k) HSpNet1. (l) HSpNet2.

TABLE III
QUALITY METRICS OBTAINED FOR DIFFERENT HS PANSHARPENING METHODS
APPLIED TO THE AVIRIS MOFFETT FIELD DATASET

	CC	SAM	RMSE	ERGAS
Ground-truth	1	0	0	0
Interpolation	0.876	6.069	420.754	4.819
SFIM	0.948	5.777	288.334	3.174
GFPCA	0.849	7.967	487.498	5.494
CNMF	0.919	7.401	364.119	3.914
BayeSparse	0.952	5.748	285.425	2.978
PNN	0.952	6.485	275.459	3.094
DRPNN	0.953	5.810	284.734	3.010
HyperPNN1	0.952	6.446	279.061	3.085
HyperPNN2	0.953	6.423	275.822	3.160
HSpNet1	0.955	5.441	<u>271.826</u>	<u>2.893</u>
HSpNet2	0.957	<u>5.630</u>	263.517	2.876

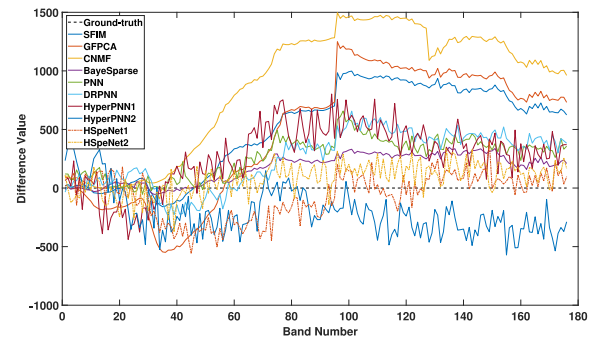


Fig. 8. Difference between the ground-truth spectrum of the pixel with spatial coordinates (83,75) and the HS pansharpening results in the same pixel for the AVIRIS Moffett Field dataset.

DRPNN, and HyperPNNs perform well in terms of spatial enhancement, but poorly in terms of spectral preservation. To validate the effectiveness of spectral preservation after HS pansharpening, Fig. 8 shows the spectral difference curves between

the ground-truth spectrum and the HS pansharpening results in a selected random pixel of the AVIRIS Moffett Field dataset. As we can see, our HSpNets are equipped with stronger spectral preservation ability than other methods.

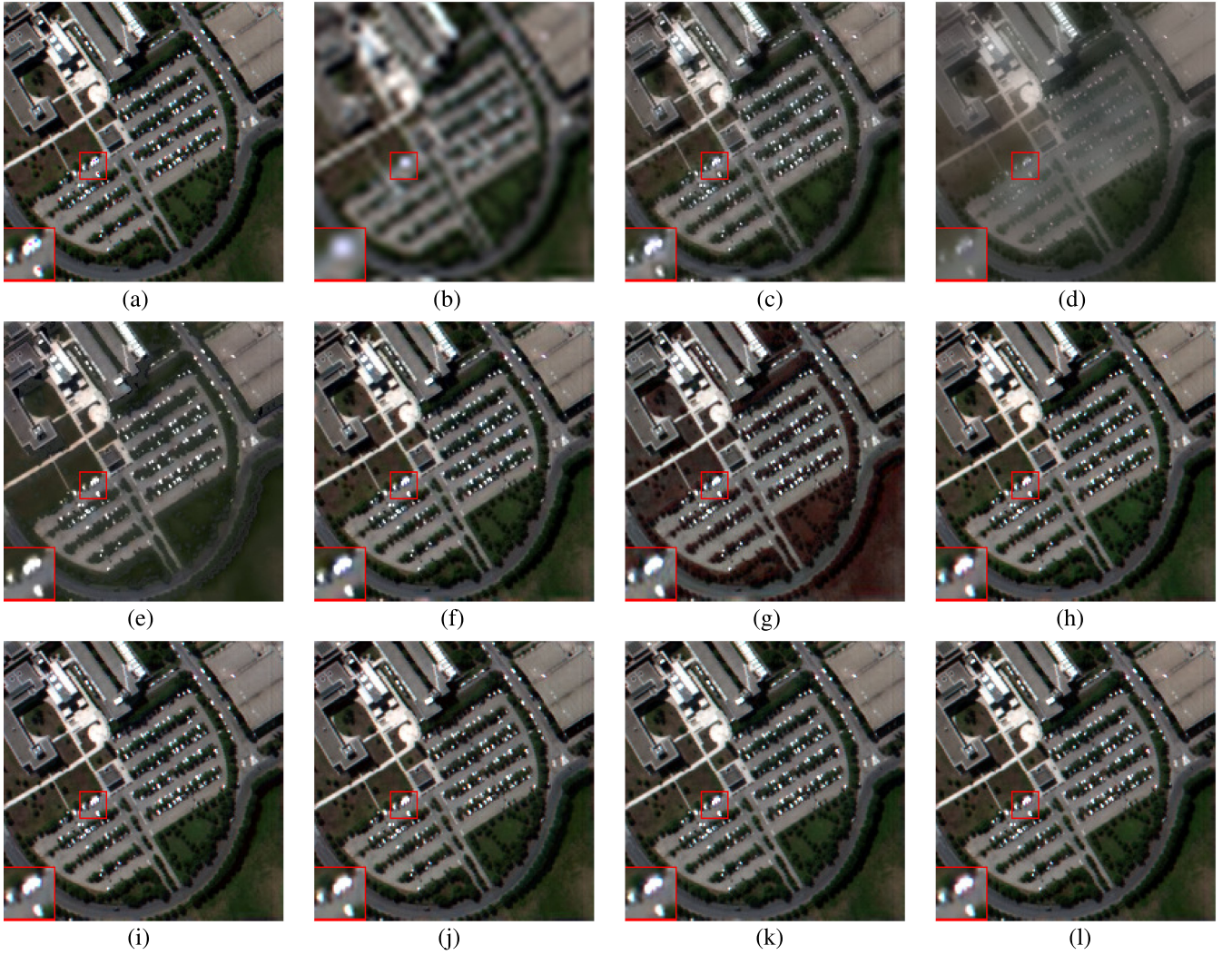


Fig. 9. HS pansharpening results for the ROSIS Pavia University dataset (displayed in RGB). (a) Ground-truth. (b) Interpolation. (c) SFIM. (d) GFPCA. (e) CNMF. (f) BayeSparse. (g) PNN. (h) DRPNN. (i) HyperPNN1. (j) HyperPNN2. (k) HSpNet1. (l) HSpNet2.

TABLE IV
QUALITY METRICS OBTAINED FOR DIFFERENT HS PANSHARPENING METHODS
APPLIED TO THE ROSIS PAVIA UNIVERSITY DATASET

	CC	SAM	RMSE	ERGAS
Ground-truth	1	0	0	0
Interpolation	0.825	7.262	469.247	7.123
SFIM	0.902	6.620	375.320	4.812
GFPCA	0.752	9.462	568.794	7.880
CNMF	0.820	8.470	494.840	5.806
BayeSparse	0.895	7.350	419.122	4.700
PNN	0.950	6.341	273.002	3.166
DRPNN	0.967	4.514	233.502	2.589
HyperPNN1	0.968	4.612	216.056	2.609
HyperPNN2	0.968	4.521	214.049	2.580
HSpNet1	0.970	4.196	207.949	2.525
HSpNet2	0.971	4.116	203.693	2.433

3) *ROSIS Pavia University Dataset*: This dataset covers an urban area in Pavia, northern Italy. The size of the degraded HS image is 122×68 pixels, and the size of the PAN image is

610×340 pixels. We choose a subscene with 200×200 pixels for testing and another part of 610×140 pixels for training.

A quantitative comparison for different methods with the ROSIS Pavia University dataset is displayed in Table IV. Our HSpNets obtain the best performance with these scene for all considered metrics. GFPCA and CNMF lag behind other methods, while SFIM and BayeSparse achieve relatively better performance among traditional methods.

Fig. 9(b)–(l) shows the pansharpened HS images by various methods for the ROSIS Pavia University dataset. The HS images given by HSpNets are the most similar to the ground truth, both in terms of spatial detail and spectral fidelity. SFIM performs well in terms of spectral preservation, but poorly in terms of spatial restoration. Both CNMF and GFPCA suffer from diffused blurring. BayeSparse incurs in subtle spectral distortion, as shown in the area in the first enlarged box. PNN, DRPNN, and HyperPNNs provide images with relatively high quality, but still suffer from some spectral distortion, such as the edge of the structure in the enlarged box.

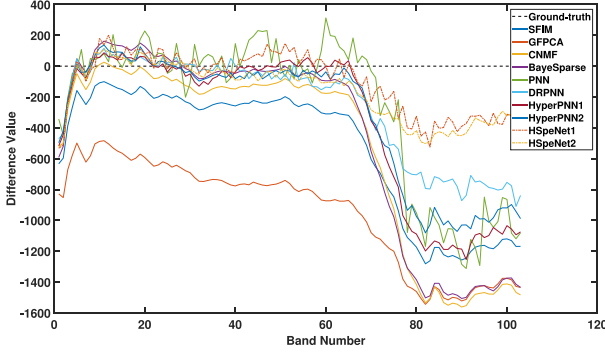


Fig. 10. Difference between the ground-truth spectrum of the pixel with spatial coordinates (102,116) and the HS pansharpening results in the same pixel for the ROSIS Pavia University dataset.

To further evaluate the spectral preservation capability of different methods, Fig. 10 shows the spectral difference curves of a random pixel in the ROSIS Pavia University dataset. Apparently, our HSpNeTs provide lower spectral differences, especially for the second half of the spectral range.

In conclusion, HSpNet1 and HSpNet2 achieve inconsistent performance on the above datasets, which can be mainly attributed to the scene characteristics and the network architectures. For the Washington DC Mall dataset, there are large amount of spatial details in the testing scene, e.g., dots and lines. Bottom layers in a deep CNN, e.g., the first or the second layer, are sensitive to such simple features, while our HSpNet1 does pass bottom-layer features directly to top layers with summation mode (as shown in Fig. 2). Therefore, our HSpNet1 is able to well mine features for the Washington DC Mall dataset; thus, it yields excellent performance on this dataset. For Moffett Field and Pavia University datasets, homogeneous areas are dominant in the testing scene. Within each area, spectra diversity is relatively low, i.e., spectra of pixels are relatively similar. Our HSpNet2 delivers the input of a convolutional layer to subsequent layers with local skip connections (as seen in Fig. 3), aiming at driving the next layer to fulfill fusion. As a result, various receptive fields are aggregated to generate the final output, which are beneficial to the extraction of area feature. Thus, our HSpNet2 is able to achieve better performance on Moffett Field and Pavia University datasets.

D. Real Image Experiments

The Hyperion Halls Creek dataset is utilized to evaluate the performance of the proposed method in a real case study. In this case, the spatial resolution of the PAN image is 10 m, and that of the HS image is 30 m. Two kinds of experiments are carried out on the Hyperion Halls Creek dataset: reduced-resolution experiment and full-resolution experiment.

In the reduced-resolution experiment, the HS and PAN images are downsampled simultaneously and then fused to yield HRHS images. Similar to the simulated-data experiment, the original HS images serve as reference images to evaluate the performance of different methods. The dimension of the test PAN image is 120×120 pixels, and the size of the experimental HS image is 40×40 pixels.

TABLE V
QUALITY METRICS OBTAINED FOR DIFFERENT HS PANSHARPENING METHODS APPLIED TO THE HYPERION HALLS CREEK DATASET

	CC	SAM	RMSE	ERGAS
Ground-truth	1	0	0	0
Interpolation	0.884	1.585	93.427	2.744
SFIM	0.827	1.932	118.202	5.588
GFPCA	0.821	2.274	124.551	3.426
CNMF	0.842	1.875	103.742	3.156
BayeSparse	0.790	2.266	128.792	3.665
PNN	0.864	3.188	125.878	5.533
DRPNN	0.889	2.005	92.899	3.583
HyperPNN1	0.894	2.099	93.911	3.162
HyperPNN2	0.894	2.084	90.471	3.522
HSpNet1	0.914	1.571	86.154	2.449
HSpNet2	0.919	1.794	84.826	2.733

As listed in Table V, our HSpNeTs achieve the best performance with respect to the four quality metrics in the reduced-resolution Hyperion Halls Creek dataset experiment. The visual appearance of the images obtained in this experiment is shown in Fig. 11. Although the pansharpened HS image obtained by interpolation obtains a relatively high score in quality metrics, a visual assessment of the result reveals severe spatial blurring. Moreover, PNN suffers from significant spectral distortion. BayeSparse, DRPNN, and HyperPNNs exhibit similar ability to preserve spatial details according to visual appearance, but their overall image spatial quality is not as good as that of our HSpNeTs. To further analyze spectral distortion, spectral difference curves are drawn in Fig. 12 for a randomly selected pixel in the Hyperion Halls Creek image. Overall, HSpNeTs exhibit excellent spectral preservation ability.

In the full-resolution experiment, the original HS and PAN images serve as the inputs to yield a real HRHS image. In this experiment, there is no reference image, so visual inspection is the only way to evaluate different methods. The pansharpened HS images generated by the considered methods are shown in Fig. 13. Interpolation and GFPCA are both dominated by significant spatial blurring. SFIM and BayeSparse suffer from ringing artifacts. PNN also suffers from spectral distortion, as in the reduced-resolution experiment. Generally, our HSpNeTs perform better than other methods in this experiment also.

To provide a more comprehensive evaluation and comparison, runtimes for all the CNN-based methods in comparison are listed in Table VI. It is indicated that our HSpNet1 and HSpNet2 are relatively low time-consuming.

E. Further Discussions

1) *Penalty Parameter λ* : As described in Section II-B4, λ balances the importance between MSE loss and cosine similarity. The former measures the absolute error in pixel-by-pixel fashion, while the latter measures the spectral similarity of each pixel. Different values of λ result in totally different results. To further explore the influence of λ in the results of our HSpNeTs, Fig. 14 shows the obtained quantitative metrics (with different values of λ) on the AVIRIS Moffett Field dataset.

Fig. 14 reveals that CC, RMSE, and ERGAS increase with λ , which means that the CNN focuses on reducing MSE loss during the optimization process. These metrics are based on a

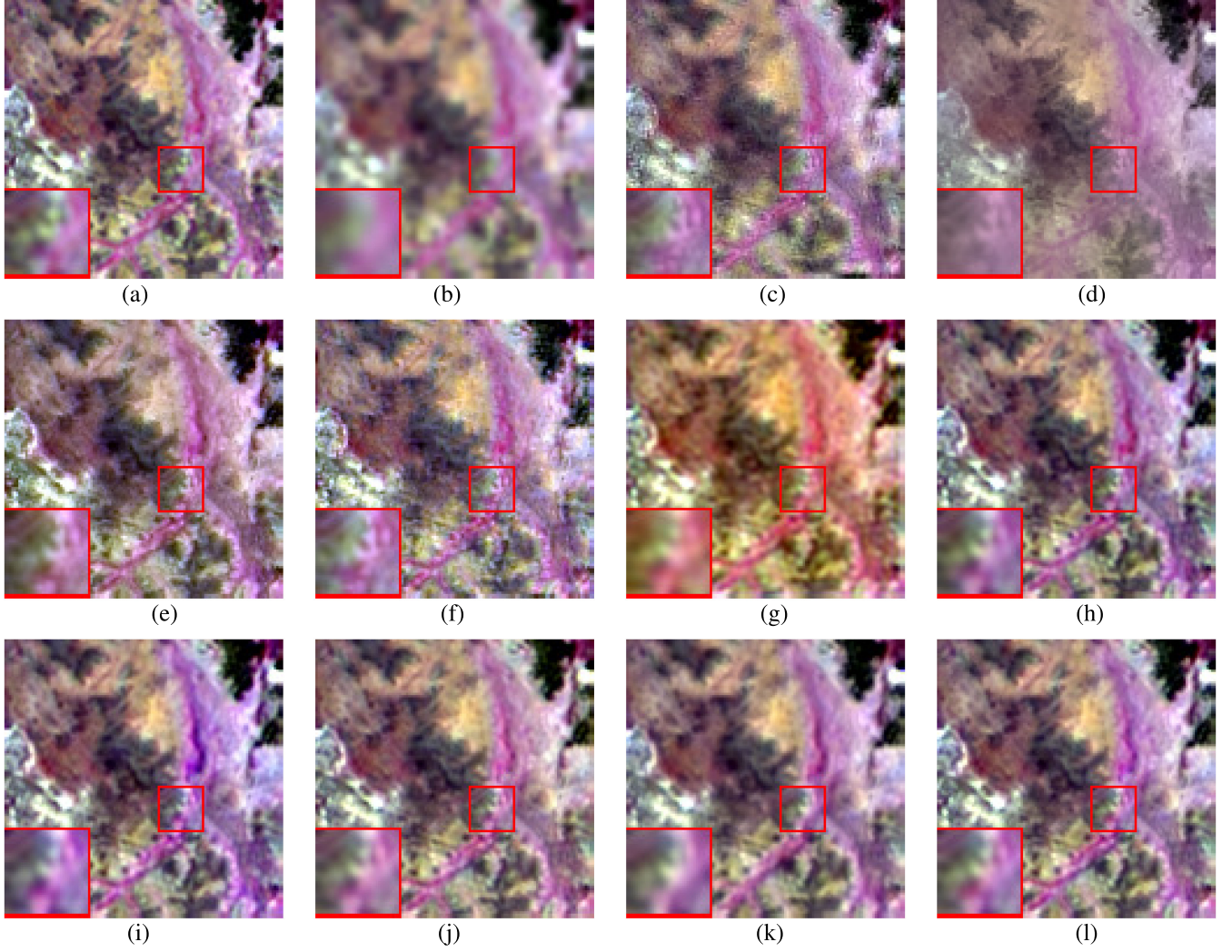


Fig. 11. HS pansharpening results for the reduced-resolution Hyperion Halls Creek dataset (displayed in RGB). (a) Ground-truth. (b) Interpolation. (c) SFIM. (d) GFPCA. (e) CNMF. (f) BayeSparse. (g) PNN. (h) DRPNN. (i) HyperPNN1. (j) HyperPNN2. (k) HSpNet1. (l) HSpNet2.

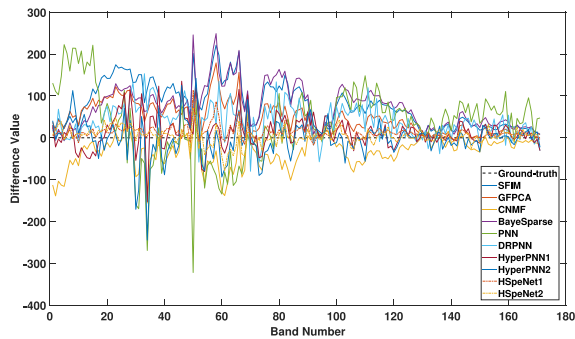


Fig. 12. Difference between the ground-truth spectrum of the pixel with spatial coordinates (73,107) and the HS pansharpening results in the same pixel for the Hyperion Halls Creek dataset.

criterion that measures differences by considering each single pixel separately. In this case, MSE is the optimal choice for the loss function. Additionally, with the decrease of λ SAM also tends to decrease, which can be explained by the fact that SAM

TABLE VI
AVERAGE RUNTIME ON DIFFERENT DATASETS

Method	Training Time	Testing Time(s)
PNN	10h 38min	0.075
DRPNN	6h 40min	0.066
HyperPNN1	5h 43min	0.054
HyperPNN2	5h 43min	0.054
HSpeNet1	6h 20min	0.055
HSpeNet2	6h 35min	0.056

and cosine similarity both evaluate the spectral similarity in pixel-by-pixel fashion. To balance the performance for different criteria, we set $\lambda = 0.9$ in our experiments.

2) *PAN Subtraction*: To evaluate the advantage of concatenating the difference between PAN and its stream output, in this experiment, we evaluate the performance of our HSpNets with and without such concatenation in the AVIRIS Moffett Field dataset. Here, without difference means that the PAN stream output is directly concatenated with the output from the HS. The obtained results are displayed in Table VII.

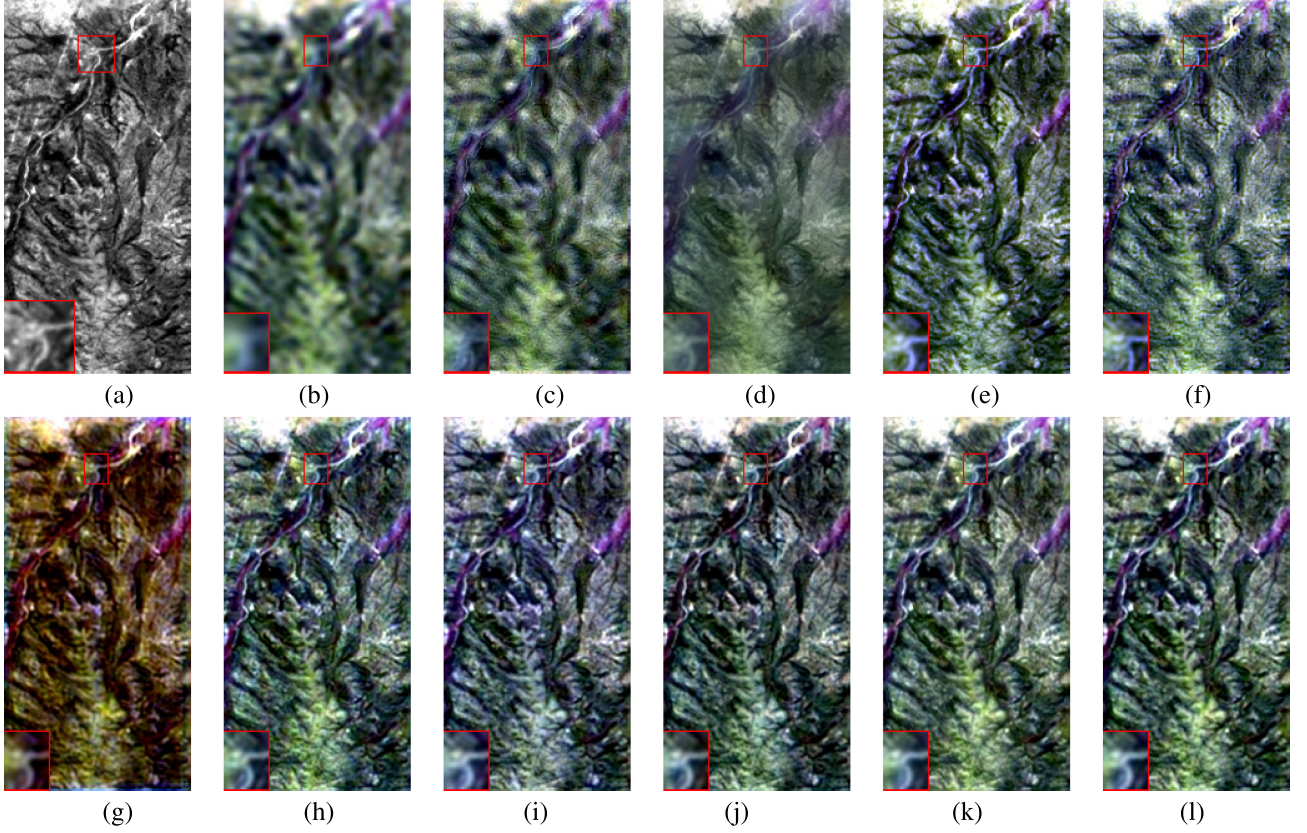


Fig. 13. HS pansharpening results for the full-resolution Hyperion Halls Creek dataset (displayed in RGB). (a) PAN. (b) Interpolation. (c) SFIM. (d) GFPCA. (e) CNMF. (f) BayeSparse. (g) PNN. (h) DRPNN. (i) HyperPNN1. (j) HyperPNN2. (k) HSpNet1. (l) HSpNet2.

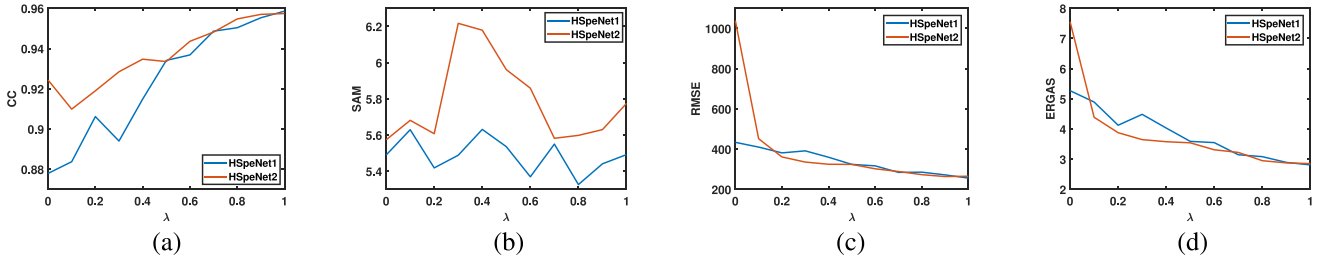


Fig. 14. (a)–(d) HS pansharpening results obtained with different values of λ for the AVIRIS Moffett Field dataset.

TABLE VII
QUALITY METRICS OBTAINED BY HSPeNET WITH AND WITHOUT
CONCATENATING THE DIFFERENCE BETWEEN PAN AND ITS STREAM
OUTPUT IN THE AVIRIS MOFFETT FIELD DATASET

		CC	SAM	RMSE	ERGAS
HSpNet1	Without	0.951	5.522	282.311	3.000
	With	0.955	5.441	271.826	2.893
HSpNet2	Without	0.952	5.666	278.696	3.034
	With	0.957	5.630	263.517	2.876

As we can see in Table VII, our proposed HSpNets (with concatenation of the difference between PAN and its stream output) achieve much better performances than without such concatenation. Specifically, the improvement in terms of global

indexes, i.e., RMSE and ERGAS, is much more significant. In other words, concatenating with PAN details tends to boost the performance of HS pansharpening.

To further explore the impact of concatenating or not the difference with PAN and its stream output, we visualize the feature maps obtained by HSpNet1 with and without such concatenation in the AVIRIS Moffett Field dataset (see Fig. 15). With the concatenation, the obtained feature maps are more likely to reveal the structure of the PAN image, such as the first, second, fourth, and sixth feature maps displayed in Fig. 15(a). Without the concatenation, the obtained feature maps exhibit significant blurring, such as the first, second, fifth, and eleventh feature maps displayed in Fig. 15(b). In contrast, compared with the feature maps in Fig. 7(a), it is easy to see that most of the

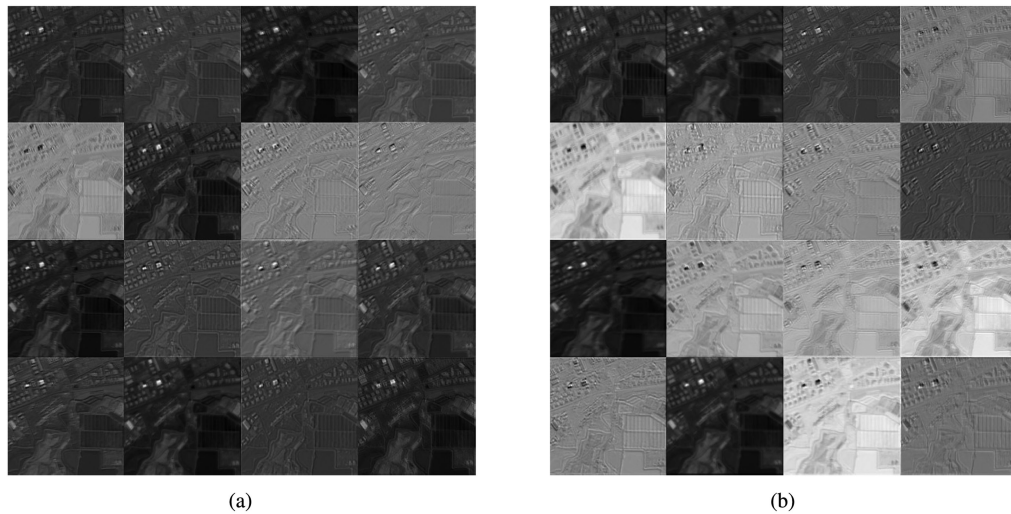


Fig. 15. Feature maps obtained by HSpNet1 (a) with and (b) without concatenating the difference between PAN and its stream output in the AVIRIS Moffett Field dataset.

feature maps in Fig. 15(b) suffer from spectral distortion, for instance in the top leftmost area. We can, therefore, conclude that the concatenation of the difference with PAN and its stream output in our proposed method yields significant enhancements, both in spatial and spectral terms.

IV. CONCLUSION

In this work, we have focused on CNNs specifically for HS pansharpening and hence designed two HSpNets, i.e., HSpNet1 and HSpNet2. HS pansharpening is faced with two special challenges: 1) wide spectral range gap between an HS images and the connected PAN images and 2) having to recover spatial details in many continuous spectral bands. To deal with the situation and make the pansharpened HS image well consistent with the true spectra, a pansharpening CNN could utilize the decomposability of HS details and accordingly synthesize these details progressively to relief the difficulty to reconstruct a full-resolution HS image in one step; meanwhile, a spectral-fidelity loss could be employed to further strengthen the spectral synthesis. Essentially, our HSpNets are designed by strictly following the above intuitions, rather than directly inspired by existing CNN structures used in other research fields like computer vision, which may be an often adopted trick to build CNNs for HS image processing. Thus, our HSpNets are innovative. Although not very complicated, such two pansharpening networks can achieve excellent performance and have clear theoretical interpretations.

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] Z. Wang, R. Zhu, K. Fukui, and J. Xue, "Matched shrunken cone detector (MSCD): Bayesian derivations and case studies for hyperspectral target detection," *IEEE Trans. Image Process.*, vol. 26, no. 11, pp. 5447–5461, Nov. 2017.
- [3] Y. Yuan, D. Ma, and Q. Wang, "Hyperspectral anomaly detection by graph pixel selection," *IEEE Trans. Cybern.*, vol. 46, no. 12, pp. 3123–3134, Dec. 2016.
- [4] L. Loncan *et al.*, "Hyperspectral pansharpening: A review," *IEEE Geosci. Remote Sens. Mag.*, vol. 3, no. 3, pp. 27–46, Sep. 2015.
- [5] 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.
- [6] Y. Zhang, "Understanding image fusion," *Photogramm. Eng. Remote Sens.*, vol. 70, no. 6, pp. 657–661, 2004.
- [7] V. K. Shettigara, "A generalized component substitution technique for spatial enhancement of multispectral images using a higher resolution data set," *Photogramm. Eng. Remote Sens.*, vol. 58, pp. 561–567, 1992.
- [8] 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.
- [9] V. P. Shah, N. H. Younan, and R. L. King, "An efficient pan-sharpening method via a combined adaptive PCA approach and contourlets," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1323–1335, May 2008.
- [10] W. Carper, T. Lillesand, and R. Kiefer, "The use of intensity-hue-saturation transformations for merging SPOT panchromatic and multispectral image data," *Photogramm. Eng. Remote Sens.*, vol. 56, no. 4, pp. 459–467, 1990.
- [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] C. A. Laben and B. V. Brower, "Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening," 2000, US Patent 6,011,875.
- [14] B. Aiazzi, S. Baronti, and M. Selva, "Improving component substitution pansharpening through multivariate regression of MS + Pan data," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3230–3239, Oct. 2007.
- [15] 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.
- [16] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674–693, Jul. 1989.

- [17] G. P. Nason and B. W. Silverman, "The stationary wavelet transform and some statistical applications," in *Wavelets and Statistics*. New York, NY, USA: Springer, 1995, pp. 281–299.
- [18] M. J. Shensa, "The discrete wavelet transform: Wedding the a trous and Mallat algorithms," *IEEE Trans. Signal Process.*, vol. 40, no. 10, pp. 2464–2482, Oct. 1992.
- [19] J. 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.
- [20] B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, and M. Selva, "MTF-tailored multiscale fusion of high-resolution MS and PAN imagery," *Photogram. Eng. Remote Sens.*, vol. 72, no. 5, pp. 591–596, 2006.
- [21] 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.
- [22] 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.
- [23] Q. Wei, N. Dobigeon, and J.-Y. Tourneret, "Fast fusion of multi-band images based on solving a Sylvester equation," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4109–4121, Nov. 2015.
- [24] Q. Wei, N. Dobigeon, and J.-Y. Tourneret, "Bayesian fusion of multi-band images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 6, pp. 1117–1127, Sep. 2015.
- [25] 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.
- [26] C. Dong, 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.
- [27] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 1646–1654.
- [28] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 2472–2481.
- [29] G. Masi, D. Cozzolino, L. Verdoliva, and G. Scarpa, "Pansharpening by convolutional neural networks," *Remote Sens.*, vol. 8, no. 7, p. 594, 2016.
- [30] L. He *et al.*, "Pansharpening via detail injection based convolutional neural networks," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 4, pp. 1188–1204, Apr. 2019.
- [31] 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.
- [32] 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.
- [33] 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.
- [34] L. He, J. Zhu, J. Li, A. Plaza, J. Chanussot, and B. Li, "HyperPNN: Hyperspectral pansharpening via spectrally predictive convolutional neural networks," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 8, pp. 3092–3100, Aug. 2019.
- [35] W. Xie, J. Lei, Y. Cui, Y. Li, and Q. Du, "Hyperspectral pansharpening with deep priors," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 5, pp. 1529–1543, May 2020.
- [36] G. He, J. Zhong, J. Lei, Y. Li, and W. Xie, "Hyperspectral pansharpening based on spectral constrained adversarial autoencoder," *Remote Sens.*, vol. 11, no. 22, p. 2691, 2019.
- [37] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
- [38] 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.
- [39] R. A. Borsoi, T. Imbiriba, and J. C. M. Bermudez, "Deep generative endmember modeling: An application to unsupervised spectral unmixing," *IEEE Trans. Comput. Imag.*, vol. 6, pp. 374–384, 2020.
- [40] T. Imbiriba, R. A. Borsoi, and J. C. M. Bermudez, "Low-rank tensor modeling for hyperspectral unmixing accounting for spectral variability," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 3, pp. 1833–1842, Mar. 2020.
- [41] F. Khajehrayeni and H. Ghassemian, "Hyperspectral unmixing using deep convolutional autoencoders in a supervised scenario," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 567–576, 2020.
- [42] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE Trans. Neural Netw.*, vol. 5, no. 2, pp. 157–166, Mar. 1994.
- [43] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. Int. Conf. Artif. Intell. Statist.*, 2010, pp. 249–256.
- [44] L. Wald, T. Ranchin, and M. Mangolini, "Fusion of satellite images of different spatial resolutions: Assessing the quality of resulting images," *Photogram. Eng. Remote Sens.*, vol. 63, no. 6, pp. 691–699, 1997.
- [45] B. Aiazzi, L. Alparone, S. Baronti, and A. Garzelli, "Context-driven fusion of high spatial and spectral resolution images based on oversampled multiresolution analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 10, pp. 2300–2312, Oct. 2002.
- [46] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. Int. Conf. Mach. Learn.*, 2010, pp. 807–814.
- [47] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Representations*, 2015, pp. 1–41.



Lin He (Member, IEEE) 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, 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, photograph and biography not available at the time of publication.



Deyu Meng (Member, IEEE) received the B.Sc. degree in information science, M.Sc. degree in applied mathematics, and Ph.D. degree in computer science from Xi'an Jiaotong University, Xi'an, China, in 2001, 2004, and 2008, respectively.

He is currently a Professor with the School of Mathematics and Statistics, Xi'an Jiaotong University, and an Adjunct Professor with the Faculty of Information Technology, The Macau University of Science and Technology, Macau, China. His current research interests include model-based deep learning,

variational networks, and meta-learning.

Dr. Meng is an Associate Editor for the IEEE TRANSACTIONS ON PATTERN RECOGNITION AND MACHINE INTELLIGENCE, *Science China Information Sciences*, and the *Frontiers of Computer Science*.



Jocelyn Chanussot (Fellow, IEEE) 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 in electrical engineering from the Université de Savoie, Annecy, France, in 1998.

Since 1999, he has been with Grenoble INP, where he is currently a Professor of Signal and Image Processing. He has been a Visiting Scholar with Stanford University, Stanford, CA, USA; the KTH Royal Institute of Technology, Stockholm, Sweden; and the

National University of Singapore, Singapore. Since 2013, he has been an Adjunct Professor with the University of Iceland, Reykjavik, Iceland. From 2015 to 2017, he was a Visiting Professor with the University of California at Los Angeles, Los Angeles, CA, USA. He holds the AXA Chair in remote sensing and is also an Adjunct Professor with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China. His research interests include image analysis, hyperspectral remote sensing, data fusion, machine learning, and artificial intelligence.

Dr. Chanussot was a member of the Machine Learning for Signal Processing Technical Committee of the IEEE Signal Processing Society from 2006 to 2008 and the Program Chair of the IEEE International Workshop on Machine Learning for Signal Processing in 2009. He is a member of the Institut Universitaire de France from 2012 to 2017 and a Highly Cited Researcher (Clarivate Analytics/Thomson Reuters) from 2018 to 2019. He was the Vice President of the IEEE Geoscience and Remote Sensing Society, in charge of meetings and symposia, from 2017 to 2019. He was the General Chair of the first IEEE GRSS Workshop on Hyperspectral Image and Signal Processing, Evolution in Remote Sensing. He was the Chair and the Co-Chair of the GRS Data Fusion Technical Committee from 2009 to 2011 and 2005 to 2008, respectively. He is the Founding President of the IEEE Geoscience and Remote Sensing French Chapter from 2007 to 2010, which received the 2010 IEEE GRS-S Chapter Excellence Award. He has received multiple outstanding paper awards. He is an Associate Editor for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE TRANSACTIONS ON IMAGE PROCESSING, and the PROCEEDINGS OF THE IEEE. He was the Editor-in-Chief of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING from 2011 to 2015. In 2014, he has served as a Guest Editor for the *IEEE Signal Processing Magazine*.



Antonio Plaza (Fellow, IEEE) received the M.Sc. and Ph.D. degrees in computer engineering from the University of Extremadura, Cáceres, Spain, in 1999 and 2002, respectively.

He is currently a Full Professor and the Head of the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, University of Extremadura. He has authored more than 600 publications, including 293 JCR journal papers (more than 170 in IEEE journals), 24 book chapters, and more than 300 peer-reviewed conference proceeding papers.

He has guest edited ten special issues on hyperspectral remote sensing for different journals. His main research interests include hyperspectral data processing and parallel computing of remote sensing data.

Prof. Plaza is a recipient of the recognition of Best Reviewers of the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS (in 2009) and the recognition of Best Reviewer of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING (in 2010), for which he served as an Associate Editor from 2007 to 2012. He is also an Associate Editor for the IEEE ACCESS. He was a member of the Editorial Board of the IEEE GEOSCIENCE AND REMOTE SENSING NEWSLETTER from 2011 to 2012 and the *IEEE Geoscience and Remote Sensing Magazine*, in 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 is 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 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 has reviewed more than 500 manuscripts for more than 50 different journals. He served as the Editor-in-Chief of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING from 2013 to 2017. He is included in the 2018 and 2019 Highly Cited Researchers List (Clarivate Analytics).