Hyperspectral Pansharpening With Deep Priors

Weiying Xie, Member, IEEE, Jie Lei, Member, IEEE, Yuhang Cui, Yunsong Li, and Qian Du, Fellow, IEEE

Abstract—Hyperspectral (HS) image can describe subtle differences in the spectral signatures of materials, but it has low spatial resolution limited by the existing technical and budget constraints. In this paper, we propose a promising HS pansharpening method with deep priors (HPDP) to fuse a low-resolution (LR) HS image with a high-resolution (HR) panchromatic (PAN) image. Different from the existing methods, we redefine the spectral response function (SRF) based on the larger eigenvalue of structure tensor (ST) matrix for the first time that is more in line with the characteristics of HS imaging. Then, we introduce HFNet to capture deep residual mapping of high frequency across the upsampled HS image and the PAN image in a hand-by-hand manner. Specifically, the learned residual mapping of high frequency is injected into the structural transformed HS images, which are the extracted deep priors served as additional constraint in a Sylvester equation to estimate the final HR HS image. Comparative analyses validate that the proposed HPDP method presents the superior pansharpening performance by ensuring higher quality both in spatial and spectral domains for all types of data sets. In addition, the HFNet is trained in the high-frequency domain based on multispectral (MS) images, which overcomes the sensitivity of deep neural network (DNN) to data sets acquired by different sensors and the difficulty of insufficient training samples for HS pansharpening.

Index Terms—Deep priors, high frequency, hyperspectral (HS) pansharpening, structure tensor (ST), Sylvester equation.

I. INTRODUCTION

Hyperspectral (HS) imaging sensor is capable of simultaneously capturing numerous narrow channels in a certain wavelength range. As a result, HS images can describe the detailed spectrum regarding physical nature of materials, which is the main advantage over traditional images [1], [2].

Manuscript received November 26, 2018; revised March 28, 2019 and May 30, 2019; accepted May 30, 2019. Date of publication June 28, 2019; date of current version May 1, 2020. This work was supported in part by the National Natural Science Foundation of China under Grant 61871359, Grant 61571345, Grant 91538101, Grant 61501346, Grant 61502367, and Grant 61701360, in part by the Young Talent fund of University Association for the Yangtse River Scholar Bonus Schemes under Grant CJT160102, in part by the Special Financial Grant from the China Postdoctoral Science Foundation under Grant 2019T120878, in part by the 111 project under Grant B08038, in part by the Fundamental Research Funds for the Central Universities under Grant JB180104, in part by the Natural Science Basic Research Plan in Shaanxi Province of China under 2019JQ153, Grant 2016QJ6023, and Grant 2016JQ6018, in part by the General Financial Grant from the China Postdoctoral Science Foundation under Grant 2017M620440, in part by the Yangtse Rive Scholar Bonus Schemes under Grant CJT160102, and in part by the Ten Thousand Talent Program. (Corresponding authors: Jie Lei; Yunsong Li.)

W. Xie, J. Lei, Y. Cui, and Y. Li are with the State Key Laboratory of Integrated Service Networks, Xidian University, Xi’an 710071, China (e-mail: wyxie@xidian.edu.cn; jielei@mail.xidian.edu.cn; yuli@mail.xidian.edu.cn).

Q. Du is with the Department of Electronic and Computer Engineering, Mississippi State University, Starkville, MS 39759 USA (e-mail: du@ece.msstate.edu).

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

Digital Object Identifier 10.1109/TNNLS.2019.2920857

Contributions. Exploring a specifically HS pansharpening method based on requirement make them impractical. Hence, it is worthwhile burden and large graphics processing unit (GPU) memory and 31 bands in [35]. When these methods are applied sets with few spectral bands, such as 16 bands in [33]ption, the recent CNNs-based methods were validated on data sets and the HS images in these data sets cover relatively This method was conducted on only two ground-based data sets with high-dimensional HS image by DNN, which enforces the grouped 4-D tensors.

Recently, with the success of deep neural networks (DNNs) in image processing, these techniques, especially convolutional neural networks (CNNs) have been explored for image superresolution (SR). As an important milestone, super-resolution convolutional neural network (SRCNN) [27] model was designed to superresolve gray-scale and RGB images. Thereafter, various DNN architectures have been investigated for this purpose, such as very deep super-resolution network (VDSR) [28], super-resolution generative adversarial network [29], Laplacian pyramid super-resolution network [30], and super-resolution network for multiple degradations [31]. As their superiority in image SR, deep models have been specifically designed to fuse remote sensing images. Most of them were originally proposed for fusing MS with PAN images [32], [33]. The method in [34] was originally proposed for HS pansharpening that directly applied SRCNN model [27]. Furthermore, Dian et al. [35] incorporated the priors leaned by CNN into the HS-MS fusion framework, named as deep hyperspectral image sharpening (DHSIS).

This method was conducted on only two ground-based data sets and the HS images in these data sets cover relatively narrow spectral range (400–700 nm or 420–720 nm). In addition, the recent CNNs-based methods were validated on data sets with few spectral bands, such as 16 bands in [33] and 31 bands in [35]. When these methods are applied to HS with hundreds of bands, the heavy computational burden and large graphics processing unit (GPU) memory requirement make them impractical. Hence, it is worthwhile exploring a specifically HS pansharpening method based on DNN that is applicable to the data sets with wider spectral range.

In this paper, we propose an HS pansharpening method with deep priors (HPDP) learned by CNN, which has the following contributions.

1) Different from the existing methods, we redefine the spectral response function (SRF) based on the larger eigenvalue of structure tensor (ST) matrix for the first time that is more in line with the characteristics of HS imaging.

2) We propose HFNet to capture deep residual mapping of high frequency across the upsamled HS image and the PAN image. Specifically, the SR using deep learning (SRDL) model is employed to upsample the LR HS image, unlike previous methods with the utilization of bicubic interpolation. As demonstrated in this paper, the SRDL can preserve spatial and spectral information better than bicubic interpolation.

3) The learned residual mapping of high frequency is injected into the structural transformed HS image, which is the extracted deep prior served as additional constraint in a Sylvester equation to estimate the final HR HS image.

4) To overcome the problem of the limited training samples and sensitivity to different sensors in HS pansharpening, we adopt MS images and their high frequency to train SRDL and HFNet, respectively. Experiments conducted on several HS images demonstrate its superiority.

The following part of this paper is organized as follows. Section II describes the proposed HPDP method. Section III presents the experiments and results. Section IV draws the conclusions.

II. PROPOSED HPDP APPROACH

Let $\mathbf{Y} \in \mathbb{R}^{L \times m}$ represent an LR HS image, where $L$ is the number of spectral bands and $m$ is the number of pixels. Let $\mathbf{P} \in \mathbb{R}^{1 \times M}$ represent the HR PAN image. The pansharpened HS image with high spatial resolution is denoted as $\mathbf{X} \in \mathbb{R}^{L \times M}$ with $M$ pixels in spatial domain. Here, $M = dm$, which means the size of $\mathbf{X}$ in spatial domain is $d$ times of $\mathbf{Y}$. Fig. 1 shows the overall flowchart of our proposed HPDP approach.

The traditional methods treat the pansharpening problem as a constrained minimization problem with the following objective function:

$$\min_{\mathbf{X}} f_1(\mathbf{X}, \mathbf{Y}) + f_2(\mathbf{X}, \mathbf{P})$$

(1)

where the first term represents the spectral preservation and the second term enforces the spatial similarity to the HR PAN image. Recently, many researchers adopted some additional priors (such as nonnegative priors [19], sparse priors [36], and low-rank priors [37]) into (1) to obtain a unique optimal solution. Dian et al. [35] introduced the priors leaned by CNN to fuse HR MS image with LR HS image. It illustrates the superiority of priors leaned by CNN in MS-HS fusion framework, but it is complex to solve two Sylvester equations. We propose specific model for fusing PAN image with high-dimensional HS image by DNN, which enforces the constraints in both spatial and spectral domains as follows:

$$\min_{\mathbf{X}} \frac{\|\mathbf{XBS} - \mathbf{Y}\|_F^2}{\text{spectral}} + \frac{\|\mathbf{RX} - \mathbf{P}\|_F^2}{\text{spatial}} + \alpha \frac{\|\mathbf{X}_{hf} - \mathbf{X}\|_F^2}{\text{high-frequency}}$$

(2)
where $B \in \mathbb{R}^{M \times M}$ denotes a blurring operator, and $S \in \mathbb{R}^{M \times m}$ represents a downsampling operation with factor $d$. Inspired by several state-of-the-art methods [16], [18], [35], the blurring operator $B$ can be decomposed as $B = \text{FFD}^H$, where $F$ is the discrete-Fourier transform matrix ($\text{FFD}^H = I_n$) and $D$ is a diagonal matrix containing eigenvalues of $B$.

It is noted that $R \in \mathbb{R}^{1 \times L}$ is the SRF and can be simply defined as $1/L \sum_{l=1}^{L} (\cdot)$ in [16] and [18]. In this way, the term $\ell_2 = \frac{\text{RX} - P}{\ell_2}$ in (2) enforces the average of all the spectral bands in the HR HS image to be close to the PAN image, which means the spectral responses of all the bands are the same and all the bands contribute spatially equal. In general, different bands describe the same scene at various wavelengths, but their qualities vary with bands and spectral responses also vary with wavelength coverage. Therefore, we first redefine the SRF that is more in line with the characteristics of HS imaging in the following part.

### A. SRF Definition

In this section, we redefine SRF $R$ with better approaching capability to the principle of imaging. We assume that $R$ is proportional to the ratio of the information in each band. The problem is how to calculate the ratio of the information in each band. What is refreshing is the utilization of ST. ST is a simple tool to adaptively gain the response function because it can effectively characterize the spectral response of each pixel in a band. More importantly, the SRF $R$ is defined based on the eigenvalue of ST in each band for the first time. First, the ST of the $l$th band in an HS image can be expressed as

$$T^l = \begin{bmatrix} (Y^l)_x^2 & (Y^l)_x(Y^l)_y \\ (Y^l)_x(Y^l)_y & (Y^l)_y^2 \end{bmatrix}$$

(3)

where $T^l$ is the ST of the $l$th band. $(Y^l)_x = (\partial Y^l/\partial x)$ and $(Y^l)_y = (\partial Y^l/\partial y)$ represent the derivatives of the $l$th band image along $x$- and $y$-directions. This ST $T^l$ of the $l$th band can be decomposed as

$$T^l = \begin{bmatrix} \eta^l_1 & \eta^l_2 \\ \eta^l_1 & \eta^l_2 \end{bmatrix} \begin{bmatrix} \lambda^l_1 & 0 \\ 0 & \lambda^l_2 \end{bmatrix}$$

(4)

where $\lambda^l_1$ and $\lambda^l_2$ are the nonnegative eigenvalues of the $l$th band, and $\eta^l_1$ and $\eta^l_2$ are their corresponding eigenvectors. Assume $\lambda^l_1 \geq \lambda^l_2$. As shown in Fig. 2, the larger eigenvalue, $\lambda^l_1$, illustrates the response intensity of the $l$th band, and the corresponding eigenvector, $\eta^l_1$, indicates the orientation that maximizes the spectral value fluctuations. It can be observed from Fig. 2 that the relative larger eigenvalue can nearly perfectly illustrate the response intensity of each pixel. If $\lambda^l_1 \geq \lambda^{l+1}_1$, the response intensity of the $l$th band is larger, that is, the $l$th band contains a larger spectral value. Therefore, it should occupy a larger proportion of weighting. Conversely, the $(l \! + \! 1)$th band should occupy a larger proportion of weighting in the case of $\lambda^l_1 < \lambda^{l+1}_1$. When the eigenvalues of all the bands are equal, all the bands have the same weight at this pixel, i.e., $1/L$, which is reduced to the averaging function.
Thus, we define the SRF as

$$R'_l = \frac{\lambda'_1}{\sum^{L}_{l=1}(\lambda'_1)}$$  \hspace{1cm} (5)$$

where $R'_l$ represents the spectral response of the $l$th band.

B. Deep Priors Acquisition

As shown in Fig. 1, there are three main steps to obtain $X_{hs}$: LR HS image upsampling, structural transformation, and high-frequency injection. More details are described as follows.

1) LR HS Image Upsampling: First, we upsample LR HS image by SRDL to the same size of the HR PAN image, which is different from bicubic interpolation in the recent fusion methods based on CNN [33], [35]. This is because the spatial and spectral information of the LR HS image is lost during the process of bicubic interpolation. It can be observed from Fig. 3 that the SRDL can preserve spatial and spectral information better than bicubic interpolation does.

The architecture of the utilized SRDL model is shown in the bottom left corner of Fig. 1. It directly uses the LR HS image as the input (i.e., without a predefined upsampling operator) and, thus, significantly reduces the computational complexity. There are three operations contained in the SRDL process: feature extraction, residual map upscaling, and reconstruction.

1) Feature Extraction: The convolutional layer acts as an extractor of the local conjunctions of features from the previous layer and leaky rectified linear unit (LReLU) is then used as the pointwise nonlinearity. In ReLU, a positive response is kept, while negative responses are suppressed by setting them to zero [38], [39]. This case may cause dead nodes when training. The LReLU overcomes the drawback when negative responses are suppressed by setting them to a small negative slope, such as 0.2 [30]. Followed by [28], the filter size is set as $3 \times 3$, and 64 filters are used to generate 64 feature maps.

2) Residual Map Upscaling: For the last layer of each feature extraction stage, one upsampling layer is utilized to upscale 64 feature maps. Then, the upsampled feature maps are convoluted to generate the residual map that is used to reconstruct the upsampled HS image.

3) Reconstruction: Inspired by [30], a bilinear filter is utilized to upsample the image to the same size of the residual map. Then, the elementwise summation operation is conducted as

$$X'_s = \uparrow Y' + E'$$  \hspace{1cm} (6)$$

where $\uparrow$ aims at upsampling the LR HS image $Y'$ to the same size of the residual map $E'$ at stage $t$. The optimal training model is achieved by minimizing the loss between the reconstructed HS image $X'_s$ and ground truth $X'$ iteratively. Following [30], the Charbonnier
penalty function is utilized as loss function to deal with outliers and improve the reconstruction accuracy

$$\ell_1(\Theta) = \sum_{i=1}^{T} \sqrt{(X' - \uparrow Y' - E)^2 + \epsilon^2}$$

where $T$ is the number of iterations. The empirical value of $\epsilon$ is 0.001 [30].

2) Structural Transformation: The upsampled HS image by super-resolution using deep learning is represented by $X_s$. In our work, the spatial information of $P$ is injected into $X_s$ from two steps: structural transformation and high-frequency injection. Here, we use a GF [40] to transfer the structural information of $P$ into $X_s$ because GF can make the filtering outputs be aligned with boundaries in the guidance image. It is assumed that the output image $X_g$ is a linear transformation of the guidance image $P$ in a local window $\Omega_j$ as

$$X_g(i) = A(j)P(i) + B(j) \quad \forall i \in \Omega_j$$

where $\Omega_j$ is a square window of size $(2r + 1) \times (2r + 1)$ around pixel $j$, $A = \{A_1, \ldots, A_L\}$ and $B = \{B_1, \ldots, B_L\}$ are the unknown coefficient matrices. The optimization problem of minimizing the difference between the input $X_i$ and output $X_g$ to determine the linear coefficients $(A, B)$ becomes

$$\arg\min_{A, B} \sum_{i \in \Omega_j} ((AP + B - X_s)^2 + \mu(A)^2)$$

where $\mu$ is a regularization parameter penalizing a large $A$. The guided filtering output $X_g$ with structural information of PAN image preserved is reconstructed by combining the estimated coefficients with the given PAN image

$$X = \tilde{A}P + \tilde{B}$$

where $\tilde{A}$ and $\tilde{B}$ are the average coefficients of all windows overlapping the current pixel. It is noted that the parameters of the GF are set as the same values for all test images with the filter size $r = 15$, and the blur degree $\mu = 10^{-6}$.

3) High-Frequency Injection: In order to further inject high frequency into the filtering output $X_g$, we propose HFNet to extract the residual map of high frequency $X_{res}$ between each band of $X_i$ and $P$. The architecture of the proposed HFNet is shown in the bottom right of Fig. 1. The high-frequency information of $X_i$ and $P$ are first extracted by a high-pass filtering and then the residual map of high frequency $X_{res}$ between each band of $X_i$ and $P$ is obtained by a trained DNN model. The architecture of HFNet is the same as in the feature extraction process of SRDL without residual map upsampling and reconstruction processes. Different from SRDL, the high-frequency information of PAN and upsampled HS images are input into HFNet with the residual map of high frequency between these two images as output, and the upsampled HS image is obtained by SRDL. Thus, the loss function of the proposed HFNet is defined as

$$\ell_2(\Theta) = \| f_r(H(P), H(X_i)) + X_r - X_s \|_2^2$$

where $H(P)$ and $H(X_i)$ represent the high-frequency information of $P$ and $X_i$, respectively. In the testing process, we input the high frequency of $X_i$ and $P$ into the well-trained HFNet; thus, the residual map of high frequency $X_{res}$ is acquired. Subsequently, this residual map learned via DNN is injected into the structural transformation result by the following equation:

$$X_{hf} = X_{res} + X_g.$$  

C. Solving the Optimized Model

Through the aforementioned steps, $R$ and $X_{hf}$, are specifically defined or learned by our proposed method. Only $X$ is unknown in (2). In order to minimize (2), we force the derivative of (2) for $X$ to be zero, and it comes to a Sylvester equation

$$C_1X + XC_2 = C$$

where $C_1 = R^T R + \alpha I_L$, $C_2 = (BS)(BS)^T$ and $C = R^T P + Y(BS)^T + \alpha X_{hf}$. Here, $I_L$ is an identity matrix. It is known that (13) has a unique solution if and only if the arbitrary sum of eigenvalues of $C_1$ and $C_2$ is not equal to zero [41]. $C_1$ is a positive matrix, and thus, the eigenvalues of $C_1$ are positive values. $C_2$ is a semipositive number and its eigenvalues are semipositive values. Therefore, the solution of the Sylvester equation is unique. The approach for solving (13) is summarized in Algorithm 1.

### Algorithm 1 Solving Sylvester Equation With Respect to $X$

**Input:** $P, Y, B, S, R, X_{hf}, \alpha$

(a) Eigen-decomposition of $B$: $B = FDF^H$
(b) $D = D(I_d \otimes I_m)$
(c) Eigen-decomposition of $C_1$: $C_1 = QAQ^{-1}$
(d) $C = Q^{-1}CF$
(e) Compute auxiliary matrix $X$ band by band:

<table>
<thead>
<tr>
<th>For $l = 1$ to $L$ do</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_l = \lambda_l^{-1}C_l - \lambda_l^{-1}C_lD(\lambda_l I_m + \sum_{i=1}^d D_i^2)^{-1}D^H$</td>
</tr>
<tr>
<td>End</td>
</tr>
</tbody>
</table>

**Input:** $X$

### III. EXPERIMENTAL RESULTS AND ANALYSIS

A. Experimental Setup

To validate the effectiveness of the proposed method, we perform both remote sensing and ground-based data experiments and compare the experimental results both qualitatively and quantitatively. Several public data sets are used, i.e., Moffett Field data set, University of Pavia data set, Chikusei data set [42] and Columbia computer vision laboratory (CAVE) data set [43]. It is worth noting that these data sets are captured by different sensors, which can illustrate the generalization capability of the compared methods.

1) Moffett Field Data Set: Moffett field data set is a standard data product acquired by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS). This data set contains 224 bands in the spectral range of 400–2500 nm. The water absorption and noise-corrupted bands are removed, and 176 bands are used for experimentation. The PAN image has $300 \times 180$ pixels...
with a spatial resolution of 50 m. The test HS image has 75 \times 45 pixels in the spatial domain with a spatial resolution of 200 m.

2) University of Pavia Data Set: A fractional of the University of Pavia data\textsuperscript{1} is used for the experiments. This image was acquired by the Reflective Optics System Imaging Spectrometer (ROSI) over the city of Pavia, Italy. This data have 115 bands ranging from 430 to 860 nm. For this data, 12 noisy bands are removed and the remaining 103 bands are used for the experiment. The PAN image has 144 \times 144 pixels with a spatial resolution of 1.3 m. The test HS image has 36 \times 36 pixels in the spatial domain with a spatial resolution of 5.2 m.

3) Chikusei Data Set: Chikusei data set\textsuperscript{2} was captured by Headwall’s Hyperspec Visible and Near-Infrared, series C (VNIR-C) imaging sensor over Chikusei, Ibaraki, Japan, on 29 July 2014. It contains 128 bands in the spectral range of 363 to 1018 nm. The PAN image has 600 \times 600 pixels with a spatial resolution of 2.5 m. The LR HS image has 150 \times 150 pixels with a spatial resolution of 10 m.

4) CAVE Data Set: The CAVE data set contains 32 real-world scenes captured by the generalized assorted pixel camera with high quality in the wavelength range 400–700 nm with 10 nm a band. The hyperspectral images (HSIs) have 31 spectral bands, and each band has a size of 512 \times 512. The LR HS image is generated by applying an 8 \times 8 Gaussian filter (with a mean of 0 and a standard deviation of 2), and then by downsampling with the scale factor as 8 \times 8.

The HR PAN image of the same scene is stimulated by downsampling the reference image with the spectral model. The spectral downsampling matrix is redefined in Section II-A. We extract 10000 PAN-HS patch pairs of size 32 \times 32 from the Worldview2 satellite as training data. The MatConvNet toolbox [44] was used to train the proposed model. The momentum parameter was set to 0.9 and the weight decay to 1e-4. The learning rate was initialized to 1e-5 for all layers and decreased.

B. Compared Methods
To achieve comprehensive comparison, several representative methods covering the complete categories of pansharpening are utilized: SFIM, GSA, MTF-GLP, MTF-GLP-HPM, GFPCA, Bayesian Sparse, HySure, FUSE, Lanaras’s, CNMF, and DHSIS. With respect to the CNN-based approach, we compare the proposed method with DHSIS, which was originally proposed to achieve HS-MS images fusion. In order to make a fair comparison, we transfer DHSIS method to achieve HS-PAN images fusion with the same input of other compared methods. These methods were implemented

\textsuperscript{1}http://www.ehu.es/ccwintco/uploads/e/e3/Pavia.mat
\textsuperscript{2}http://naotoyokoya.com/Download.html
via the publicly released codes with the default parameter settings. Because the DHSIS method only open the residual map of CAVE data set, thus, we compare this method on CAVE data set.

C. Quantitative Metrics

Four widely used metrics are adopted for performance evaluation, including cross correlation (CC) [45], spectral angle mapper (SAM) [46], root-mean-squared error (RMSE), and erreur relative global adimensionnelle de synthèse (ERGAS) [47]. CC is a spatial measure and its optimal value of CC is 1. SAM evaluates the spectral distortion between the reference HS image and the fused one. RMSE and ERGAS measure the global quality of the pansharpened HS image. The best value of SAM, RMSE, and ERGAS is 0.

D. Parameter Setting

In the proposed HPDP model, $\alpha$ is used to optimize the estimated HR HS image in the optimization problem (2), which controls the quantity of the estimated HR HS image. Besides spatial resolution enhancement, spectral preservation is also important in the process of pansharpening. Fig. 4 shows the SAM curves as functions of $\alpha$ for four test data sets. It can be seen that the best SAM values are achieved when $\alpha = 4 \times 10^{-3}$, $2 \times 10^{-3}$, $1 \times 10^{-3}$, and $3 \times 10^{-3}$ for Moffett field, University of Pavia, Chikusei and CAVE data sets, respectively. For Moffett field data set, the SAM value significantly decreased from $\alpha = 2 \times 10^{-3}$ to $3 \times 10^{-3}$. In addition, the SAM value is similar when $\alpha = 3 \times 10^{-3}$ and $4 \times 10^{-3}$. Considering the trend of the SAM value for all data sets, we choose $\alpha = 3 \times 10^{-3}$ in the following experiments. This default parameter setting cannot obtain the best performances for all sample images, but it can provide stable and acceptable performances for most of the sample images.

E. Mid-Level Representation

Some feature maps for the Moffett field data set extracted by the SRDL model at different layers are shown in Fig. 5. The SRDL model can upsample the LR HS image to the same size of the PAN image, which can preserve spectral information better than bicubic interpolation. Example feature maps of different layers extracted by the proposed HFNet are shown in Fig. 6. Obviously, HFNet can accurately extract high-frequency residual map between the PAN and the upsampled HS images.

F. Experimental Results

The visual comparison of the reconstructed results obtained by different methods for the Moffett field data is displayed in Fig. 7. It can be seen that the results obtained by the GSA, MTF-GLP, MTF-GLP-HPM, GFPCA, and Bayesian Sparse methods look blurry since the effective spatial information is not sufficiently injected. The edges in the reconstructed images obtained by the SFIM and FUSE are too sharp in some areas. The Lanaras’s and CNMF methods generate spectral distortion because there is much higher chromatic aberration compared to the ground truth in some areas. The halo artifacts and the blurring problems can be eliminated by the proposed method. In particular, the color of the reconstructed HS image is close to that of the input one. All of the performance metrics for each method on Moffett field image are reported in Table I. It apparently indicates the proposed HPDP method outperforms all the other methods. To further clarify this point, Fig. 8 shows the spectral difference vectors of different methods at four randomly selected locations. It can be observed that the proposed HPDP method attains the best approximation of
the intrinsic spectral patterns of the reference HS image, which fully complies with our quantitative evaluation in Table I.

For the University of Pavia data set, the false color results of the pansharpened images obtained by different methods are shown in Fig. 9. Obviously, most methods achieve blurry results in some areas such as edges of metal sheets, which illustrates that the spatial information of the PAN image is not sufficiently injected. In particular, the pansharpened
Visual results obtained by different methods on the University of Pavia data set. (a) Ground truth. (b) SFIM. (c) GSA. (d) MTF-GLP. (e) MTF-GLP-HPM. (f) GFPCA. (g) Bayesian Sparse. (h) HySure. (i) FUSE. (j) Lanaras’s. (k) CNMF. (l) HPDP. Note that the false color image is chosen for clear visualization (red: 80, green: 40, and blue: 20).

TABLE I

<table>
<thead>
<tr>
<th>Methods</th>
<th>CC</th>
<th>SAM</th>
<th>RMSE</th>
<th>ERGAS</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SFIM</td>
<td>0.9591</td>
<td>7.9166</td>
<td>0.0279</td>
<td>5.5586</td>
<td>0.5288</td>
</tr>
<tr>
<td>GSA</td>
<td>0.9711</td>
<td>7.4896</td>
<td>0.0237</td>
<td>4.7728</td>
<td>1.3017</td>
</tr>
<tr>
<td>MTF-GLP</td>
<td>0.9681</td>
<td>7.0855</td>
<td>0.0252</td>
<td>5.0155</td>
<td>21.601</td>
</tr>
<tr>
<td>MTF-GLP-HPM</td>
<td>0.9695</td>
<td>7.6444</td>
<td>0.0243</td>
<td>5.6066</td>
<td>2.6049</td>
</tr>
<tr>
<td>GFPCA</td>
<td>0.9408</td>
<td>9.2672</td>
<td>0.0364</td>
<td>7.1852</td>
<td>2.3028</td>
</tr>
<tr>
<td>Bayesian Sparse</td>
<td>0.9554</td>
<td>8.5311</td>
<td>0.0294</td>
<td>6.0525</td>
<td>130.34</td>
</tr>
<tr>
<td>HySure</td>
<td>0.9569</td>
<td>8.6395</td>
<td>0.0275</td>
<td>5.6386</td>
<td>73.625</td>
</tr>
<tr>
<td>FUSE</td>
<td>0.9644</td>
<td>7.3740</td>
<td>0.0269</td>
<td>5.3954</td>
<td>1.4610</td>
</tr>
<tr>
<td>Lanaras’s</td>
<td>0.9725</td>
<td>6.6227</td>
<td>0.0220</td>
<td>4.5102</td>
<td>13.262</td>
</tr>
<tr>
<td>CNMF</td>
<td>0.9780</td>
<td>6.2843</td>
<td>0.0209</td>
<td>4.2364</td>
<td>6.9308</td>
</tr>
<tr>
<td>HPDP</td>
<td>0.9875</td>
<td>5.0499</td>
<td>0.0152</td>
<td>3.1291</td>
<td>4.1824</td>
</tr>
</tbody>
</table>

TABLE II

<table>
<thead>
<tr>
<th>Methods</th>
<th>CC</th>
<th>SAM</th>
<th>RMSE</th>
<th>ERGAS</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SFIM</td>
<td>0.8910</td>
<td>8.9459</td>
<td>0.0483</td>
<td>6.7779</td>
<td>0.2034</td>
</tr>
<tr>
<td>GSA</td>
<td>0.9070</td>
<td>9.0159</td>
<td>0.0447</td>
<td>6.2299</td>
<td>0.2876</td>
</tr>
<tr>
<td>MTF-GLP</td>
<td>0.8987</td>
<td>8.9253</td>
<td>0.0463</td>
<td>6.4268</td>
<td>5.0809</td>
</tr>
<tr>
<td>MTF-GLP-HPM</td>
<td>0.9270</td>
<td>8.0359</td>
<td>0.0393</td>
<td>5.7509</td>
<td>1.0595</td>
</tr>
<tr>
<td>GFPCA</td>
<td>0.8140</td>
<td>11.1132</td>
<td>0.0676</td>
<td>9.3797</td>
<td>1.2543</td>
</tr>
<tr>
<td>Bayesian Sparse</td>
<td>0.8769</td>
<td>10.7767</td>
<td>0.0534</td>
<td>7.2383</td>
<td>70.119</td>
</tr>
<tr>
<td>HySure</td>
<td>0.8686</td>
<td>11.7496</td>
<td>0.0532</td>
<td>7.2068</td>
<td>24.458</td>
</tr>
<tr>
<td>FUSE</td>
<td>0.8827</td>
<td>9.1694</td>
<td>0.0496</td>
<td>6.9570</td>
<td>0.2889</td>
</tr>
<tr>
<td>Lanaras’s</td>
<td>0.9081</td>
<td>8.5209</td>
<td>0.0448</td>
<td>6.3251</td>
<td>6.3850</td>
</tr>
<tr>
<td>CNMF</td>
<td>0.9005</td>
<td>7.9677</td>
<td>0.0463</td>
<td>6.6354</td>
<td>1.9045</td>
</tr>
<tr>
<td>HPDP</td>
<td>0.9404</td>
<td>7.7923</td>
<td>0.0350</td>
<td>5.0030</td>
<td>0.7188</td>
</tr>
</tbody>
</table>

Experiments have also been conducted on the Chikusei data set acquired by VNIR-C that is different from the previous data sets. The false color image of the estimated HR HS images is shown in Fig. 11. It can be seen that the structural information is better reconstructed by the proposed HPDP method. For this data set, the MTF-GLP-HPM and HySure methods introduce redundant information. The objective results of the HPDP method and the other methods are given in Table III. It can be observed that the proposed HPDP method is still superior with respect to all the objective measures.
method is closer to the baseline, which means that our HPDP can preserve the spectral information better for the Chikusei data set.

Furthermore, Table IV lists the objective assessment of the compared methods including the recent proposed DHSIS method on the CAVE data set. The authors who proposed DHSIS method only conduct experiments on the ground-based data sets such as CAVE data set and only open the residual map of an HSI from CAVE data set. Therefore, the DHSIS method is only compared in this section. It can be seen that the proposed HPDP method can still achieve the best values for all the metrics. In order to clearly show the effectiveness of the proposed method, the absolute differences maps between the fusion results and the reference one are shown in Fig. 13. Obviously, the difference map obtained by the proposed method achieves the smallest value difference, which means the pansharpened result obtained by the proposed method is the closest to the reference image.

In addition, we measure the running time of each comparative method on all the testing images and the results recorded
Fig. 11. Visual results obtained by different methods on the Chikusei data set. (a) Ground truth. (b) SFIM. (c) GSA. (d) MTF-GLP. (e) MTF-GLP-HPM. (f) GFPCA. (g) Bayesian Sparse. (h) HySure. (i) FUSE. (j) Lanaras’s. (k) CNMF. (l) HPDP. Note that the false color image is chosen for clear visualization (red: 80, green: 40, and blue: 20).

Fig. 12. Spectral reflectance difference vectors at four randomly selected locations on Chikusei data set.

in seconds are given in Tables I–IV. All the experiments are performed in the TensorFlow 1.3.0 and MATLAB (R2014a) environment on a server: Intel(R) Xeon(R) CPU E5-2650 v4 at 2.20GHz and NVIDIA Tesla K80 GPU. In comparison to the traditional methods such as SFIM, GSA, and GFPCA, the proposed HPDP method is more time-consuming but illustrates the improvement in objective performance (CC, RMSE, SAM, and ERGAS). However, the implementation of the proposed method is faster than that of the recent pansharpening methods based on DNN, taking about 2.8185 s for the CAVE data sets, where the size of the input image is 512 × 512 × 31. The proposed HPDP method consists of three procedures, i.e., SRF definition, deep priors acquisition, and solving the optimized model. When the model is trained, the deep priors can be efficiently acquired. In addition, the solution of the Sylvester equation is unique and the optimized model can be solved
without any iteration. Furthermore, only one parameter needs to be tuned as described in Section III-D.

G. Component Analysis

This section analyzes the effects of four major processing components, i.e., upsampling with SRDL, structural transformation by GF and high-frequency injection. Let $X_s$, $X_g$, and $X_{hf}$ represent the generation of the three main operations, respectively. In addition, we analyze the influence of the proposed SRF on the performance of pansharpening. Let $X - R_o$ represent the pansharpened HS image with the original SRF. As listed in Table V, CC that reflects the geometric distortion is a spatial index and it shows an increasing trend, which means spatial information of the HR PAN image is effectively injected into the LR HS image in each process, similar to RMSE and ERGAS. In the final step, the deep priors are returned to a Sylvester equation to optimize the result in both spatial and spectral domains. The gradual improvement of the four metrics is relatively obvious, which means the learned deep priors can enforce the final HR output close to the reference one. Furthermore, the objective measurements of $X$ are all better than that of $X - R_o$, which means the proposed SRF has positive influence on the performance of pansharpening.

H. Experiments on Real Data

The Hyperion data set is used to evaluate the feasibility of our proposed HPDP method on real data. The PAN image is of size $300 \times 300$ and the HS image is of size $100 \times 100 \times 171$. In the LR HS image upsampling step, we modify the SRDL model when the upscale is 3 according to VDSR [28]. The original HS and PAN images as well as the pansharpened results are displayed in Fig. 14. It can be

<table>
<thead>
<tr>
<th>Components</th>
<th>CC</th>
<th>SAM</th>
<th>RMSE</th>
<th>ERGAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_s$</td>
<td>0.8813</td>
<td>6.6967</td>
<td>0.0486</td>
<td>7.9977</td>
</tr>
<tr>
<td>$X_g$</td>
<td>0.9089</td>
<td>6.6067</td>
<td>0.0399</td>
<td>6.8512</td>
</tr>
<tr>
<td>$X_{hf}$</td>
<td>0.9215</td>
<td>6.7644</td>
<td>0.0374</td>
<td>6.4154</td>
</tr>
<tr>
<td>$X - R_o$</td>
<td>0.9675</td>
<td>4.9159</td>
<td>0.0214</td>
<td>3.8210</td>
</tr>
<tr>
<td>$X$</td>
<td>0.9702</td>
<td>4.7752</td>
<td>0.0179</td>
<td>3.7744</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Methods</th>
<th>CC</th>
<th>SAM</th>
<th>RMSE</th>
<th>ERGAS</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSA</td>
<td>0.9423</td>
<td>3.1844</td>
<td>0.0133</td>
<td>5.4557</td>
<td>0.4494</td>
</tr>
<tr>
<td>MTF-GLP</td>
<td>0.9404</td>
<td>3.2072</td>
<td>0.0139</td>
<td>5.6610</td>
<td>12.703</td>
</tr>
<tr>
<td>MTF-GLP-HPM</td>
<td>0.9203</td>
<td>3.2946</td>
<td>0.0175</td>
<td>6.5924</td>
<td>1.2976</td>
</tr>
<tr>
<td>GFPCA</td>
<td>0.8502</td>
<td>4.8784</td>
<td>0.0227</td>
<td>8.5260</td>
<td>2.1984</td>
</tr>
<tr>
<td>Bayesian Sparse</td>
<td>0.8426</td>
<td>5.0698</td>
<td>0.0258</td>
<td>9.7499</td>
<td>45.457</td>
</tr>
<tr>
<td>HySure</td>
<td>0.9335</td>
<td>3.1888</td>
<td>0.0158</td>
<td>6.1959</td>
<td>14.168</td>
</tr>
<tr>
<td>FUSE</td>
<td>0.9355</td>
<td>3.2174</td>
<td>0.0015</td>
<td>5.8582</td>
<td>3.8837</td>
</tr>
<tr>
<td>Lanaras’s</td>
<td>0.9291</td>
<td>3.8169</td>
<td>0.0148</td>
<td>6.0985</td>
<td>8.2970</td>
</tr>
<tr>
<td>CNMF</td>
<td>0.9489</td>
<td>3.6940</td>
<td>0.0126</td>
<td>5.2001</td>
<td>2.9935</td>
</tr>
<tr>
<td>HPDP</td>
<td>0.9590</td>
<td>3.1571</td>
<td>0.0119</td>
<td>4.9473</td>
<td>1.3884</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Methods</th>
<th>CC</th>
<th>SAM</th>
<th>RMSE</th>
<th>ERGAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>SFIM</td>
<td>0.8943</td>
<td>9.0297</td>
<td>0.0424</td>
<td>7.4522</td>
</tr>
<tr>
<td>GSA</td>
<td>0.9132</td>
<td>8.7315</td>
<td>0.0379</td>
<td>6.7460</td>
</tr>
<tr>
<td>MTF-GLP</td>
<td>0.9060</td>
<td>8.1285</td>
<td>0.0399</td>
<td>6.9089</td>
</tr>
<tr>
<td>MTF-GLP-HPM</td>
<td>0.8583</td>
<td>10.855</td>
<td>0.0482</td>
<td>11.954</td>
</tr>
<tr>
<td>GFPCA</td>
<td>0.8619</td>
<td>8.8053</td>
<td>0.0499</td>
<td>8.8350</td>
</tr>
<tr>
<td>Bayesian Sparse</td>
<td>0.8959</td>
<td>8.9032</td>
<td>0.0401</td>
<td>8.0829</td>
</tr>
<tr>
<td>HySure</td>
<td>0.8638</td>
<td>11.369</td>
<td>0.0457</td>
<td>9.0121</td>
</tr>
<tr>
<td>FUSE</td>
<td>0.8945</td>
<td>9.0429</td>
<td>0.0419</td>
<td>7.4213</td>
</tr>
<tr>
<td>Lanaras’s</td>
<td>0.9017</td>
<td>7.7808</td>
<td>0.0375</td>
<td>7.0677</td>
</tr>
<tr>
<td>CNMF</td>
<td>0.9217</td>
<td>6.9237</td>
<td>0.0348</td>
<td>6.2893</td>
</tr>
<tr>
<td>HPDP</td>
<td>0.9356</td>
<td>6.6566</td>
<td>0.0327</td>
<td>5.8819</td>
</tr>
</tbody>
</table>
seen that the pansharpened result by our method is capable of properly enhancing spatial resolution while finely preserving the structure underlying the HS image.

The HR HS image of the real data set is not available. For quantitative assessment, we degrade the original HS and PAN images, and then fuse them. Table VI lists the objective measurements between the fused results and the original HS image. It can be seen that the proposed method shows the best objective performance, i.e., ranking as the first for the CC, RMSE, SAM, and ERGAS, which further illustrates the superior performance of the proposed method in spatial SR and spectral preservation. In addition, the proposed method takes 1.3884 s for this real data.

1. Experiments on Noisy Input

Because some methods perform denoising implicitly while others do not, additive Gaussian noise in the simulated input images has a significant influence on the pansharpening quality [6]. Therefore, additive Gaussian noise is added to two input images for all test remote sensing data sets. Similar to [6], a denoising method in [48] is utilized to enhance the reliability of quantitative evaluation. We have provided a direct illustration of the false color images created by the pansharpened Moffett field via various methods when SNR is 20 dB. As seen from the pansharpened image in Fig. 15, our proposed HPDP method can reconstruct a more eye-catching
HS image from the noisy input. Furthermore, Table VII reports the average performance of the compared methods on all remote sensing data sets when SNR is 20 dB. The best results of CC, SAM, ERGAS, and RMSE indexes are provided by the proposed method, which demonstrates that HPDP can provide the best pansharpening performance for the noisy PAN and HS images.

IV. CONCLUSION

In this paper, we propose a powerful HS pansharpening method based on deep priors, termed as HPDP, to enhance the spatial resolution of an LR HS image with the help of an HR PAN image. Based on the analysis of the larger eigenvalue of ST matrix in each spectral band, we first define an effective SRF in accordance with the principle of imaging. Furthermore, we propose HFNet to learn deep priors, and then return the deep priors into a Sylvester equation to optimize the final HR HS image. To better preserve spectral information, we exploit SRDL rather than the traditional bicubic interpolation to upsample the LR HS image. Different from the recent CNN-based methods for HS pansharpening, we train the samples in high-frequency domain, which improves the generalization of the proposed deep model. The obtained results systematically outperform the competitors, providing experimental evidence of the effectiveness of the proposed HPDP method.

REFERENCES


Weliying Xie (M’18) received the B.S. degree in electronic information science and technology from the University of Jinan, Jinan, China, in 2011, the M.S. degree in communication and information systems from Lanzhou University, Lanzhou, China, in 2014, and the Ph.D. degree in communication and information systems from Xidian University, Xi’an, China, in 2017.

She is currently a Post-Doctoral Researcher and a Teacher with Xidian University. Her current research interests include neural networks, deep learning, hyperspectral image processing, and high-performance computing.

Jie Lei (M’19) received the M.S. degree in telecommunication and information systems and the Ph.D. degree in signal and information processing from Xidian University, Xi’an, China, in 2006 and 2010, respectively.

From 2014 to 2015, he has been a Visiting Scholar with the Department of Computer Science, University of California at Los Angeles, Los Angeles, CA, USA. He is currently an Associate Professor with the School of Telecommunications Engineering, Xidian University, where he is a member of the Image Coding and Processing Center, State Key Laboratory of Integrated Services Networks. His current research interests include image and video processing, computer vision, and customized computing for Big Data applications.

Yuhang Cui received the B.E. degree in telecommunication engineering from Xidian University, Xi’an, China, in 2018, where he is currently pursuing the M.S. degree with the Image Coding and Processing Center, State Key Laboratory of Integrated Service Network.

His current research interests include image processing, hyperspectral image superresolution and deep learning.

Yunsong Li received the M.S. degree in telecommunication and information systems and the Ph.D. degree in signal and information processing from Xidian University, Xi’an, China, in 1999 and 2002, respectively.

In 1999, he joined the School of Telecommunications Engineering, Xidian University, where he is currently a Professor and also the Director of the Image Coding and Processing Center, State Key Laboratory of Integrated Service Networks. His current research interests include image and video processing and high-performance computing.

Qian Du (S’98–M’00–SM’05–F’18) received the Ph.D. degree in electrical engineering from the University of Maryland, Baltimore, MD, USA, in 2000. He is currently the Bobby Shackouls Professor with the Department of Electrical and Computer Engineering, Mississippi State University, Starkville, MS, USA. She is also an Adjunct Professor with the College of Surveying and Geo-informatics, Tongji University, Shanghai, China. Her current research interests include hyperspectral remote sensing image analysis and applications, pattern classification, data compression, and neural networks.

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