

Deep Hyperspectral Image Sharpening

Renwei Dian, *Student Member, IEEE*, Shutao Li[✉], *Senior Member, IEEE*,
Anjing Guo, *Student Member, IEEE*, and Leyuan Fang, *Senior Member, IEEE*

Abstract—Hyperspectral image (HSI) sharpening, which aims at fusing an observable low spatial resolution (LR) HSI (LR-HSI) with a high spatial resolution (HR) multispectral image (HR-MSI) of the same scene to acquire an HR-HSI, has recently attracted much attention. Most of the recent HSI sharpening approaches are based on image priors modeling, which are usually sensitive to the parameters selection and time-consuming. This paper presents a deep HSI sharpening method (named DHSIS) for the fusion of an LR-HSI with an HR-MSI, which directly learns the image priors via deep convolutional neural network-based residual learning. The DHSIS method incorporates the learned deep priors into the LR-HSI and HR-MSI fusion framework. Specifically, we first initialize the HR-HSI from the fusion framework via solving a Sylvester equation. Then, we map the initialized HR-HSI to the reference HR-HSI via deep residual learning to learn the image priors. Finally, the learned image priors are returned to the fusion framework to reconstruct the final HR-HSI. Experimental results demonstrate the superiority of the DHSIS approach over existing state-of-the-art HSI sharpening approaches in terms of reconstruction accuracy and running time.

Index Terms—Deep convolutional neural network (CNN), fusion, hyperspectral imaging, residual learning, super-resolution.

I. INTRODUCTION

HYPERSPECTRAL imaging is a promising imaging modality, which can simultaneously acquire images of the same scene across a number of different wavelengths [1]. Rich spectral characteristics existed in hyperspectral images (HSIs) with hundreds of spectral bands have been demonstrated to improve the performance of many remote sensing [2]–[7] and computer vision [8], [9] applications. Most of the potentials of the HSIs are linked with their high spatial resolution (HR), which enables a precise identification of the materials in the scene under analysis. Due to the limitations of existing imaging sensors, there is the tradeoff between the spatial resolution and the spectral resolution. In specific, HSIs with high spectral resolution can only be acquired with low spatial resolution (LR) by hyperspectral imaging

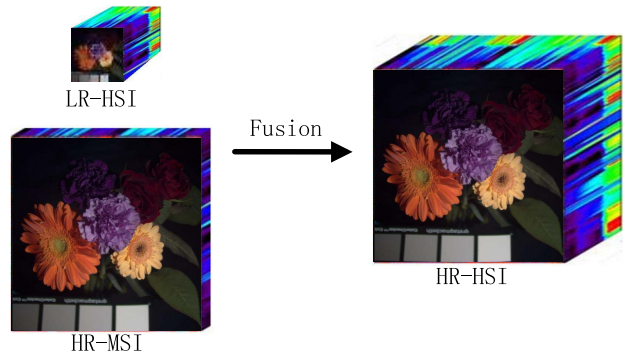


Fig. 1. HSI sharpening: fuse an LR-HSI with an HR-MSI of the same scene to reconstruct an HR-HSI.

sensors, to ensure a high signal-to-noise ratio (SNR) [10]. Compared with hyperspectral imaging sensors, the existing multispectral imaging sensors can capture a multispectral image (MSI) with lower spectral resolution but much higher spatial resolution [10]. Therefore, to acquire an HR-HSI, an LR-HSI is often fused with an HR-MSI of the same scene, as shown in Fig. 1. This procedure is called as HSI sharpening or HSI super-resolution. There is an increasing demand for HSI sharpening due to their use in many applications, such as change detection [11], anomaly detection [12], and objection classification [13]. For example, experiments in [12] demonstrate that anomaly detection on the fused HR-HSI can, indeed, improve the detection accuracy.

Image priors has an important effect on the image processing problems [14]. Most of the recent HSI sharpening methods are based on image priors modeling [10], which formulate the fusion problem as the optimization problem constrained by various HSI priors. Although image prior's modeling-based HSI sharpening methods have produced the state-of-the-art performance, they still have drawbacks. On one hand, different kinds of priors may be appropriate for different kinds of HSIs, and thus, it may be hard to choose what kinds of priors to use for different HSIs. On the other hand, the related optimization may be complex to solve and sensitive to the parameters selections, which may need parallel computing [15], [16] to speed up. In recent years, neural networks have been successfully applied in many automatic control [17]–[20] and computer vision tasks [21]–[25]. Especially, convolutional neural network (CNN) with deep architecture is very effective to exploit image characteristics [23], [26]–[29].

In this paper, we propose a deep HSI sharpening (DHSIS) method for the fusion of an LR-HSI and an HR-MSI of the same scene. Different from the HSI priors modeling-based

Manuscript received November 8, 2017; revised December 29, 2017; accepted January 16, 2018. Date of publication February 20, 2018; date of current version October 16, 2018. This work was supported in part by the National Natural Science Fund of China for International Cooperation and Exchanges under Grant 61520106001, in part by the National Natural Science Fund of China for Distinguished Young Scholars under Grant 61325007, and in part by the Fund of Hunan Province for Science and Technology Plan Project under Grant 2017RS3024. (Corresponding author: Shutao Li.)

The authors are with the College of Electrical and Information Engineering, Hunan University, Changsha 410082, China (e-mail: drw@hnu.edu.cn; shutao_li@hnu.edu.cn; anjing_guo@hnu.edu.cn; fangleyuan@gmail.com).

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

Digital Object Identifier 10.1109/TNNLS.2018.2798162

2162-237X © 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.

See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

HSI sharpening methods, we directly use the priors learned by CNN-based residual learning, which can better model the priors of HR-HSIs. The proposed DHSIS incorporates the priors learned by deep CNN-based residual learning into the fusion framework. The incorporation is reflected in two aspects. On one hand, we first initialize the HR-HSI from the known LR-HSI and HR-MSI by solving a Sylvester equation. Then, to learn the priors of HR-HSI, we utilize the initialized HR-HSI as the input of the CNN to map the residuals between reference HR-HSI and initialized HR-HSI. This initialization can fully utilize the constraints of the fusion framework, thus improving the quality of the input data. Besides, the residual learning can not only speed up the training but also boost the learning performance. On the other hand, the learned priors of HR-HSI are returned to the fusion framework to reconstruct the final estimated HR-HSI, which can further improve the performance.

The remainder of this paper is organized as follows. In Section II, we give the review of the HSI sharpening methods. Section III formulates the problem of the HSI sharpening. The proposed DHSIS approach for HSI sharpening is introduced in Section IV. In Section V, experimental results on two public HSI databases and the corresponding discussions are presented. Conclusions and future research directions are given in Section VI.

II. RELATED WORKS

To acquire an HR-HSI, one class of methods combine an LR MSI with an HR panchromatic (PAN) image (gray image), which is called as the pan sharpening [30]. Pan-sharpening methods mainly include component substitution methods [31], variational methods [32], [33], and deep learning-based methods [34], and so on. Since a single-band PAN image has little spectral resolution, the reconstructed HR-HSIs by these approaches usually contain considerable spectral distortions.

Recently, HSI sharpening, which fuses an LR-HSI with an HR-MSI to acquire an HR-HSI, has been actively investigated. HSI sharpening algorithms can be roughly categorized into two types: matrix factorization-based methods [1], [35]–[41] and tensor factorization methods [42].

Matrix factorization-based HSI sharpening methods assume that the HR-HSI contains a small number of pure spectral signatures [43], and thus, the HR-HSI can be approximated by the spectral basis multiplication by the coefficients. Kawakami *et al.* [35] propose to learn the spectral basis from the LR-HSI with a sparse prior and then conduct sparse coding on the HR-MSI to estimate the coefficients. Furthermore, methods proposed in [36]–[38] use the priors of spectral unmixing to regularize the fusion problem. Instead of estimating spectral basis in advance and keeping it fixed, these methods alternately update spectral basis and coefficients with nonnegative and sum-to-one constraints, which yield a more accurate reconstruction. To better use the prior information of the HR-HSI, approaches proposed in [1], [39], and [40] also take the spatial structures of the HR-HSI into consideration to regularize the fusion problem. For example, Akhtar *et al.* [39] acquire coefficients with the simultaneous greedy pursuit

algorithm for each local patch, which exploits the prior that nearby pixels are much similar to each other. Furthermore, Dong *et al.* [1] propose a nonnegative dictionary-learning algorithm to learn the spectral basis and utilize the structured sparse coding approach to estimate the coefficients. They fully exploit the nonlocal spatial similarities of the HR-HSI, thus achieving the good fusion results.

Another class of HSI sharpening methods are based on tensor factorization. Dian *et al.* [42] first propose a nonlocal sparse tensor factorization (NLSTF) for the fusion of an LR-HSI with an HR-MSI of the same scene. Different from the matrix factorization-based methods, they consider an HR-HSI as a 3-D tensor. Based on the Tucker decomposition [44], the HR-HSI is approximated by the core tensor multiplying by dictionaries of three modes. To exploit the nonlocal spatial similarities, they first cluster the similar cubes of HR-MSI together. Then, dictionaries of three modes are learned for each group, respectively. Finally, to better model the spatial–spectral correlations of the HSIs, they estimate the core tensor for each HR-HSI cube with a sparse prior on the learned dictionaries.

III. PROBLEM FORMULATION

In this paper, the HR-HSI, the LR-HSI, and the HR-MSI are denoted as the matrices. The first dimension of the matrix stands for the number of spectral bands, and the second dimension of the matrix denotes the number of pixels. The desired HR-HSI is denoted by $\mathbf{X} \in \mathbb{R}^{S \times N}$, where S is the number of spectral bands and N is the number of pixels. $\mathbf{Y} \in \mathbb{R}^{S \times n}$ represents the acquired LR-HSI, where $n = N/d$ is the number of pixels in LR-HSI and d is the decimation factor. With respect to \mathbf{X} , \mathbf{Y} has the same number of spectral bands S and is spatially downsampled. $\mathbf{Z} \in \mathbb{R}^{s \times N}$ represents the HR-MSI of the same scene, where s represents the number of spectral bands. With respect to \mathbf{X} , \mathbf{Z} has the same number of pixels N and is the spectrally downsampled, i.e., $S > s$.

The LR-HSI is the spatially downsampled version of \mathbf{X} , i.e.,

$$\mathbf{Y} = \mathbf{XBS} \quad (1)$$

where $\mathbf{B} \in \mathbb{R}^{N \times N}$ represents a convolution between the point spread function of the sensor and the HR-HSI bands (represented by the rows of \mathbf{X}). The blur matrix \mathbf{B} is a block circulant matrix with circulant blocks, which can be decomposed as

$$\mathbf{B} = \mathbf{FDF}^H \quad (2)$$

where the matrix \mathbf{F} is the discrete-Fourier transform matrix, and its conjugate transpose \mathbf{F}^H stands for the inverse fast Fourier transform operation. The matrix \mathbf{D} is a diagonal matrix, which has eigenvalues of \mathbf{B} . $\mathbf{S} \in \mathbb{R}^{N \times n}$ is the downsampling matrix, whose columns are a subset of the identity matrix, thus selecting a subset of the rows of \mathbf{XB} . This assumption has been widely used in the HSI sharpening literatures [1], [26], [40], [45].

The HR-MSI \mathbf{Z} is the spectrally downsampled version of \mathbf{X} , i.e.,

$$\mathbf{Z} = \mathbf{RX} \quad (3)$$

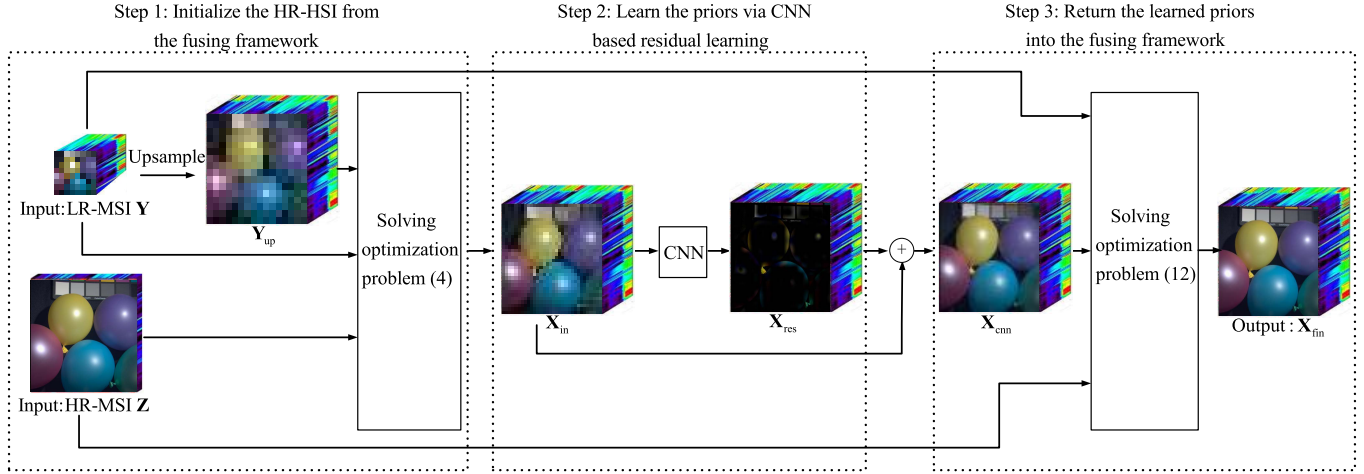


Fig. 2. Scheme of the proposed DHSIS method.

where $\mathbf{R} \in \mathbb{R}^{S \times S}$ is the spectral response matrix of the multispectral imaging sensor, whose rows contain the spectral response of the multispectral sensor. The HR-MSI is often an RGB image.

IV. PROPOSED DHSIS APPROACH

The proposed DHSIS approach incorporates the priors learned by deep CNN-based residual learning into the fusion framework. In specific, the DHSIS method mainly includes three steps: initialize the HR-HSI from the fusion framework, learn the priors of the HR-HSI via deep CNN-based residual learning, and return the learned priors into the fusion framework, as shown in Fig. 2. More details of each step are described in the following contents.

A. Initialize the HR-HSI From the Fusion Framework

In the above-mentioned formulation, the goal is to estimate the HR-HSI \mathbf{X} from \mathbf{Y} and \mathbf{Z} . Based on the imaging models (1) and (3), the fusion problem can be written as

$$\min_{\mathbf{X}} \|\mathbf{Y} - \mathbf{XBS}\|_F^2 + \|\mathbf{Z} - \mathbf{RX}\|_F^2 \quad (4)$$

where $\|\cdot\|_F$ denotes the Frobenius norm. Due to the down-sampling operations \mathbf{R} and \mathbf{S} , the optimization with respect to \mathbf{X} is highly ill-posed and thus has no unique solution. Therefore, we need to use some prior information of the unknown HR-HSI to regularize it and improve the conditioning of the problem. Various priors have been already exploited to regularize the fusion problem, such as low-rank priors [46], sparse priors [35], [41], [42], priors of spectral unmixing [36]–[38], nonlocal spatial similarities [1], and spatial smoothness [40]. For images of different characteristics, we may need to choose different priors for the purpose of good performance. Besides, combining several kinds of priors may achieve better performance, but it makes the model more complex to solve and parameters hard to choose. Deep CNN is very effective to exploit image characteristics, which has been shown the state-of-the-art performance in several image reconstruction problems [24], [26], [47]. Therefore, we intend to use deep

CNN to learn the priors of the HR-HSI. However, there are two inputs, i.e., LR-HSI and HR-MSI, and both of them are of different sizes from the HR-HSI; thus, it is hard to map the two inputs to the HR-HSI directly. To solve this problem, we initialize the HR-HSI \mathbf{X}_{in} from the LR-HSI and the HR-MSI as the input of the CNN. Based on the optimization model (4), we estimate the HR-HSI by solving the following optimization problem:

$$\min_{\mathbf{X}_{in}} \|\mathbf{Y} - \mathbf{X}_{in}\mathbf{BS}\|_F^2 + \|\mathbf{Z} - \mathbf{RX}_{in}\|_F^2 + \eta_1 \|\mathbf{X}_{in} - \mathbf{Y}_{up}\|_F^2 \quad (5)$$

where $\mathbf{Y}_{up} \in \mathbb{R}^{S \times N}$ represents the upsampled LR-HSI produced by the bicubic interpolation algorithm and η_1 is the regularization parameter. In order to minimize (5), we force the derivative of (5) for \mathbf{X}_{in} to be zero, thus creating the following Sylvester equation:

$$\mathbf{H}_1 \mathbf{X}_{in} + \mathbf{X}_{in} \mathbf{H}_2 = \mathbf{H}_3 \quad (6)$$

where

$$\begin{aligned} \mathbf{H}_1 &= \mathbf{R}^T \mathbf{R} + \eta_1 \mathbf{I}_S \\ \mathbf{H}_2 &= (\mathbf{BS})(\mathbf{BS})^T \\ \mathbf{H}_3 &= \mathbf{R}^T \mathbf{Z} + \mathbf{Y}(\mathbf{BS})^T + \eta_1 \mathbf{Y}_{up} \end{aligned} \quad (7)$$

where $\mathbf{I}_S \in \mathbb{R}^{S \times S}$ is the identity matrix.

If and only if an arbitrary sum of the eigenvalues of \mathbf{H}_1 and \mathbf{H}_2 is not equal to zero, the solution of Sylvester equation (7) has unique solution [48]. \mathbf{H}_1 is the positive matrix, and thus, its eigenvalues are positive values. \mathbf{H}_2 is the semipositive matrix, and thus, its eigenvalues are positive values or zeros. Therefore, an arbitrary sum of the eigenvalues of \mathbf{H}_1 and \mathbf{H}_2 is greater than zero, and the solution of the Sylvester equation is unique. Equation (6) can be solved analytically and efficiently by exploiting the properties of \mathbf{B} and \mathbf{S} , which does not need the iteration for \mathbf{X}_{in} . The approach for solving (6) is summarized in Algorithm 1 and demonstrated in [45].

In Algorithm 1, the blur matrix \mathbf{B} is decomposed as shown in (2). In the decomposition, the matrix $\mathbf{D} \in \mathbb{C}^{N \times N}$ is a

Algorithm 1 Closed-Form Solution of (5) With Respect to \mathbf{X}_{in}

Input: $\mathbf{Y}, \mathbf{Z}, \mathbf{R}, \mathbf{B}, \mathbf{S}, \mathbf{Y}_{\text{up}}, \eta_1$ **Output:** \mathbf{X}_{in} **Initialization:**

$$\mathbf{H}_1 = \mathbf{R}^T \mathbf{R} + \eta_1 \mathbf{I}_S$$

$$\mathbf{H}_2 = (\mathbf{B}\mathbf{S})(\mathbf{B}\mathbf{S})^T$$

$$\mathbf{H}_3 = \mathbf{R}^T \mathbf{Z} + \mathbf{Y}(\mathbf{B}\mathbf{S})^T + \eta_1 \mathbf{Y}_{\text{up}}$$

(a) Eigen-decomposition of \mathbf{B} :

$$\mathbf{B} = \mathbf{F}\mathbf{D}\mathbf{F}^H$$

(b) $\tilde{\mathbf{D}} = \mathbf{D}(\mathbf{1}_d \otimes \mathbf{1}_n)$ (c) Eigen-decomposition of \mathbf{H}_1 :

$$\mathbf{H}_1 = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^{-1}$$

(d) $\tilde{\mathbf{H}}_3 = \mathbf{Q}^{-1}\mathbf{H}_3\mathbf{F}$ (e) Calculate auxiliary matrix $\tilde{\mathbf{X}}_{\text{in}}$ row by row**For** $l = 1$ **to** S **do**

$$\tilde{\mathbf{x}}_l = \lambda_l^{-1}(\tilde{\mathbf{H}}_3)_l - \lambda_l^{-1}(\tilde{\mathbf{H}}_3)_l \tilde{\mathbf{D}}(\lambda_l \mathbf{I}_n + \sum_{t=1}^d \mathbf{D}_t^2) \tilde{\mathbf{D}}^H$$

End(f) Set $\mathbf{X}_{\text{in}} = \mathbf{Q}\tilde{\mathbf{X}}_{\text{in}}\mathbf{F}^H$

diagonal matrix, which can be represented as

$$\mathbf{D} = \begin{bmatrix} \mathbf{D}_1 & 0 & \cdots & 0 \\ 0 & \mathbf{D}_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{D}_d \end{bmatrix} \quad (8)$$

where $\mathbf{D}_i \in \mathbb{C}^{n \times n}$. $\mathbf{\Lambda}$ is the diagonal matrix and has eigenvalues of the matrix \mathbf{H}_1 , which can be written as

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_S \end{bmatrix} \quad (9)$$

The matrix \mathbf{Q} contains eigenvectors of the matrix \mathbf{H}_1 in its columns. $\tilde{\mathbf{x}}_l$ represents l th row of $\tilde{\mathbf{X}}_{\text{in}} = \mathbf{Q}^{-1}\mathbf{X}_{\text{in}}\mathbf{F}$, that is, $\tilde{\mathbf{X}}_{\text{in}} = [\tilde{\mathbf{x}}_1^T, \tilde{\mathbf{x}}_2^T, \dots, \tilde{\mathbf{x}}_S^T]^T$

B. Learn the HR-HSI Priors via CNN-Based Residual Learning

Once the initialized HR-HSI \mathbf{X}_{in} is acquired, we can learn the priors of the HR-HSI via deep CNN-based residual learning. Next, we mainly introduce the input-output, and architecture of the utilized CNN.

1) *Input and Output of the CNN*: In the training process, instead of directly mapping the LR-HSI and the HR-HSI to the HR-HSI, we use the initialized HR-HSI \mathbf{X}_{in} as the input of the CNN, which has two advantages. First, the initialized HR-HSI \mathbf{X}_{in} can be the same size of the output, which is convenient for mapping the residual. Furthermore, we incorporate the two constraints resulting from the imaging models in the process of initializing \mathbf{X}_{in} , which can help the reconstruction.

Instead of using the reference HR-HSI as the output of CNN, we map \mathbf{X}_{in} to the residual HSI \mathbf{X}_{res} , which is the

residuals between the reference HR-HSI and the initialized HR-HSI, defined as

$$\mathbf{X}_{\text{res}} = \mathbf{X} - \mathbf{X}_{\text{in}}. \quad (10)$$

The residual learning can greatly benefit the CNN, as it can not only speed up the training but also boost the leaning performance, which has shown the state-of-the-art performance in image classification [49] and image denoising [24].

2) *Architecture of the CNN*: We adopt a general CNN architecture. Specifically, the CNN has 16 blocks, which consists of the following three types.

- 1) *Conv+ReLU*: In the first block, we use 64 filters of size $3 \times 3 \times S$ to produce 64 feature maps, where S represents the number of spectral bands.
- 2) *Conv+BN+ReLU*: For blocks 2~15, we use 64 filters of size $3 \times 3 \times 64$.
- 3) *Conv*: For the last block, S filters of size $3 \times 3 \times 64$ are exploited to produce the output. In the CNN, batch normalization (BN) [50] is used to accelerate the training process, and rectified linear units (ReLU) [51] are used as activation function.

C. Return the Learned Priors Into the Fusion Framework

In the testing process, we input the initialized testing HR-HSI \mathbf{X}_{in} into the well-trained CNN, thus acquiring the output residual HSI $\tilde{\mathbf{X}}_{\text{res}}$. Therefore, we can obtain the HR-HSI \mathbf{X}_{cnn} estimated by CNN via the following equation:

$$\mathbf{X}_{\text{cnn}} = \mathbf{X}_{\text{in}} + \tilde{\mathbf{X}}_{\text{res}}. \quad (11)$$

Instead of using the HR-HSI estimated via CNN as the final estimation result, \mathbf{X}_{cnn} can be returned to the optimization model (5) to further improve the performance. In this way, the final estimated HR-HSI \mathbf{X}_{fin} can be obtained by solving the following optimization problem:

$$\min_{\mathbf{X}_{\text{fin}}} \|\mathbf{Y} - \mathbf{X}_{\text{fin}}\mathbf{B}\mathbf{S}\|_F^2 + \|\mathbf{Z} - \mathbf{R}\mathbf{X}_{\text{fin}}\|_F^2 + \eta_2 \|\mathbf{X}_{\text{fin}} - \mathbf{X}_{\text{cnn}}\|_F^2 \quad (12)$$

where $\eta_2 > 0$ is the regularization parameter. The problem (12) is similar to (5) and thus can be solved by solving a Sylvester equation, as demonstrated in Algorithm 1.

V. EXPERIMENTS

A. Experimental Database

In this section, experiments are conducted on two public databases, i.e., Columbia computer vision laboratory (CAVE) database [52]¹ and Harvard database [53],² to evaluate the effectiveness of the proposed DHSIS method. The CAVE database [52] consists of 32 indoor HSIs captured by the generalized assorted pixel camera with high quality. The HSIs have 31 spectral bands, and each band has a size of 512×512 . The images of the scenes are acquired at a wavelength interval 10 nm in the range of 400–700 nm. The Harvard database [53] includes 50 HSIs of both indoor and outdoor scenes featuring

¹<http://www.cs.columbia.edu/CAVE/databases/multispectral/>.

²<http://vision.seas.harvard.edu/hyperspec/>.

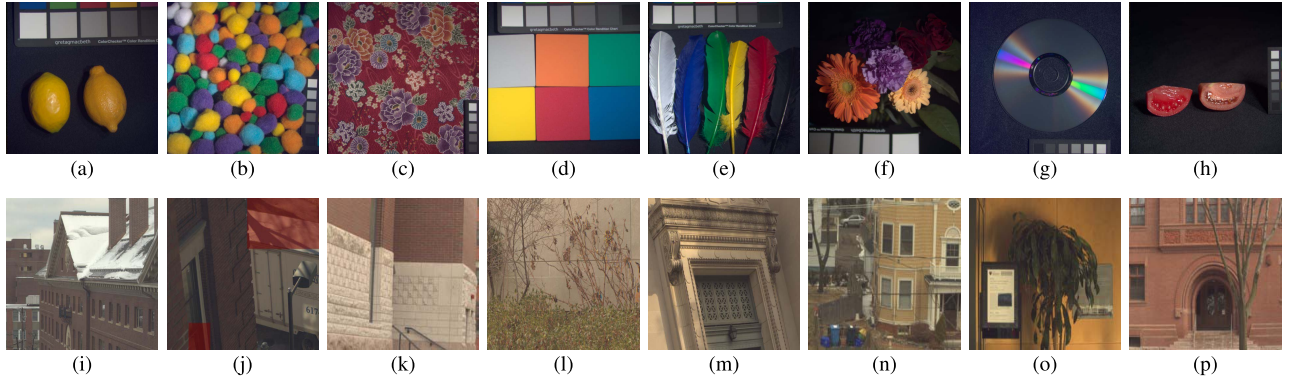


Fig. 3. RGB images from the ground-based databases. The first row: images from the CAVE database [52]. The second row: images from the Harvard database [53]. (a) Lemons. (b) Pompoms. (c) Cloth. (d) Sponges. (e) Feathers. (f) Flowers. (g) CD. (h) Tomatoes. (i) Img 1. (j) Img a2. (k) Img b1. (l) Img b4. (m) Img b5. (n) Img c4. (o) Img d8. (p) Img e6.

a diversity of objects, materials, and scale under daylight illumination. Each HSI has a spatial resolution of 1392×1040 and 31 spectral bands. The HSIs of the scenes are acquired at a wavelength interval 10 nm in the range of 420–720 nm. Both of CAVE database and Harvard database are widely used in HSI sharpening literatures [1], [37], [39], [42], [54]–[56]. Fig. 3 shows some RGB images from the CAVE database and the Harvard database.

The HSIs from two databases are used as ground truths. The LR-HSI is generated by applying an 8×8 Gaussian filter (with a mean of 0 and a standard deviation of 2) and then by downsampling every 8 pixels in both the vertical and horizontal directions for each band of the reference image, i.e., the decimation factor is 8×8 . The HR-MSI (RGB image) \mathbf{Z} of the same scene is stimulated by downsampling \mathbf{X} with the spectral model. The spectral downsampling matrix \mathbf{R} is derived from the response of a Nikon D700 camera.³ The first 20 (about 60%) HSIs of the CAVE database are used for training, and the last 12 HSIs are used for testing. For the Harvard database, the first 30 (about 60%) HSIs are used for training, and the last 20 HSIs are used for testing. Both of \mathbf{X}_{in} and \mathbf{X}_{res} are partitioned into several 32×32 patches in the training process.

B. Compared Methods

We have compared the proposed DHSIS method with three recent state-of-the-art HSI sharpening methods: the coupled spectral unmixing (CSU) [37], the nonnegative-structured sparse representation (NSSR) [1], and NLSTF method [42]. The CSU and NSSR are based on matrix factorization, and the NLSTF is based on tensor factorization.

C. Quantitative Metrics

To evaluate the quality of the reconstructed HSIs, four indexes are used in this paper.

The first index is the peak SNR (PSNR). The PSNR for

HSI is defined as the average PSNR of all bands, e.g.,

$$\text{PSNR}(\mathbf{X}, \hat{\mathbf{X}}) = \frac{1}{S} \sum_{j=1}^S \text{PSNR}(\mathbf{X}^i, \hat{\mathbf{X}}^i) \quad (13)$$

where \mathbf{X}^i and $\hat{\mathbf{X}}^i$ denote the i th band images of ground truth $\mathbf{X} \in \mathbb{R}^{S \times N}$ and estimated HR-HSI $\hat{\mathbf{X}} \in \mathbb{R}^{S \times N}$, respectively. The PSNR measures the similarities between the fused image and the reference image. The higher the PSNR, the better the fusion quality.

The second index is the spectral angle mapper (SAM) [57], which is defined as the angle between the estimated pixel $\hat{\mathbf{x}}_j$ and the ground-truth pixel \mathbf{x}_j , averaged over the whole spatial domain, i.e.,

$$\text{SAM}(\mathbf{X}, \hat{\mathbf{X}}) = \frac{1}{N} \sum_{j=1}^N \arccos \frac{\hat{\mathbf{x}}_j^T \mathbf{x}_j}{\|\hat{\mathbf{x}}_j\|_2 \|\mathbf{x}_j\|_2}. \quad (14)$$

The SAM is given in degrees. It is an indicator of the spectral quality of the estimated image. The smaller SAM, the less spectral distortion.

The third index is the structural similarity index (SSIM) [58], averaged over all bands of \mathbf{X} and $\hat{\mathbf{X}}$, i.e.,

$$\text{SSIM}(\mathbf{X}, \hat{\mathbf{X}}) = \frac{1}{S} \sum_{i=1}^S \text{SSIM}(\hat{\mathbf{X}}^i, \mathbf{X}^i) \quad (15)$$

where $\text{SSIM}(\mathbf{X}^i, \hat{\mathbf{X}}^i)$ is the SSIM computed between the clean band \mathbf{X}^i and the estimated band $\hat{\mathbf{X}}^i$. The SSIM is an indicator of the spatial structures preservation of the estimated image. The higher the SSIM, the better the spatial structures preservation.

The fourth index is the erreur relative globale adimensionnelle de synthèse (ERGAS) [59], defined as

$$\text{ERGAS}(\mathbf{X}, \hat{\mathbf{X}}) = \frac{100}{d} \sqrt{\frac{1}{S} \sum_{i=1}^S \frac{\text{MSE}(\mathbf{X}^i, \hat{\mathbf{X}}^i)}{\mu_{\hat{\mathbf{X}}^i}^2}} \quad (16)$$

where d is the spatial downsampling factor, $\mu_{\hat{\mathbf{X}}^i}$ is the mean value of $\hat{\mathbf{X}}^i$, and $\text{MSE}(\mathbf{X}^i, \hat{\mathbf{X}}^i)$ is the mean square error between \mathbf{X}^i and $\hat{\mathbf{X}}^i$. ERGAS reflects the overall quality of the fused image. A smaller ERGAS value means that the fusion results are better.

³https://www.maxmax.com/spectral_response.htm.

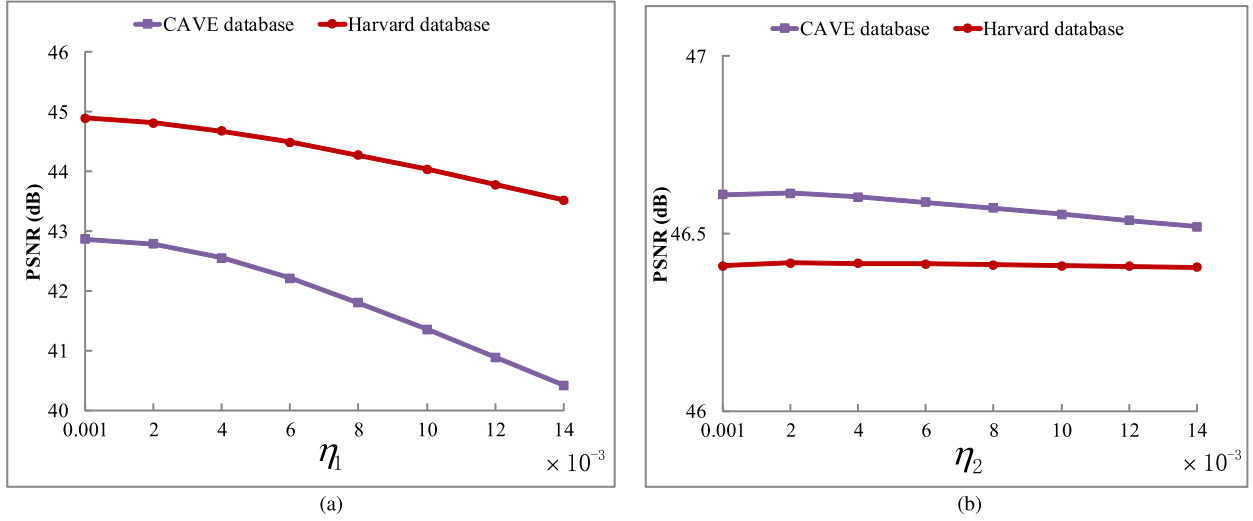


Fig. 4. (a) Average PSNR curves of \mathbf{X}_{in} as functions of η_1 . (b) Average PSNR curves of \mathbf{X}_{fin} as functions of η_2 .

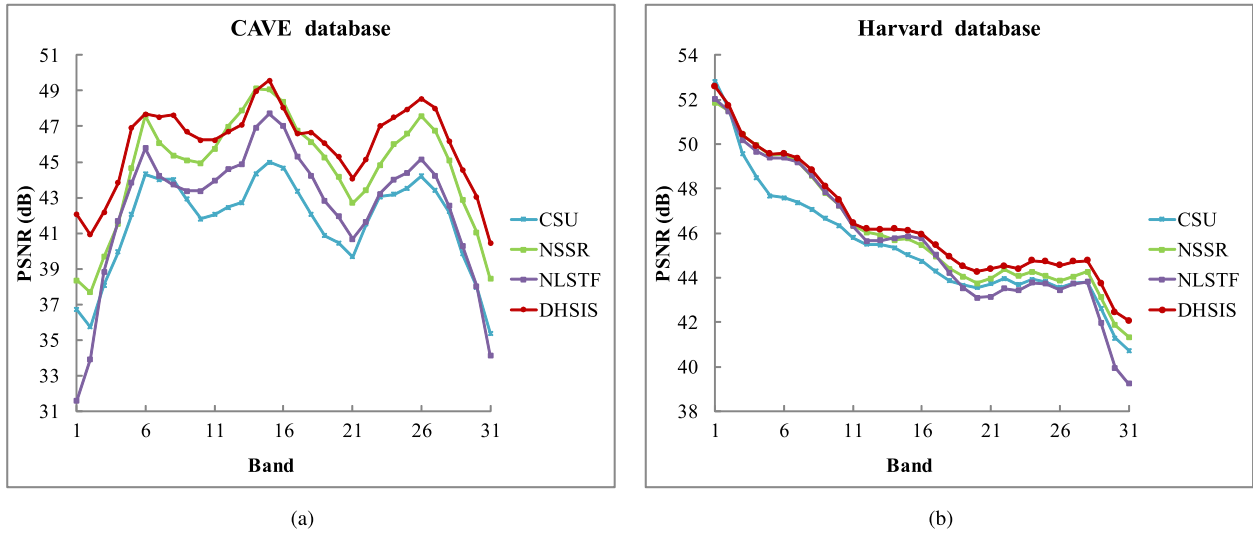


Fig. 5. Average PSNR curves as functions of the spectral bands for the test method. (a) CAFE database. (b) Harvard database.

D. Parameters Selection

In the DHSIS method, two parameters η_1 and η_2 need to be set. The regularization parameters η_1 are used for the initialization of the HR-HSI \mathbf{X}_{in} in the optimization problem (5). The regularization parameters η_2 are used for the optimization of the final estimated HR-HSI \mathbf{X}_{cnn} in the optimization problem (12). Therefore, η_1 and η_2 are highly related to the quality of \mathbf{X}_{in} and \mathbf{X}_{fin} , respectively. Besides, the related two optimization problems (5) and (12) are in a similar form, which lead to solving a Sylvester equation.

Fig. 4(a) and (b) shows the average PSNR of \mathbf{X}_{in} and \mathbf{X}_{cnn} as functions of η_1 and η_2 , respectively. To ensure that the optimization problems (5) and (12) have unique solution, η_1 and η_2 must satisfy $\eta_1 > 0$ and $\eta_2 > 0$. As can be seen from Fig. 4(a), the PSNR of \mathbf{X}_{in} decreases obviously on both the CAFE database and the Harvard database, as η_1 increases. However, it can be seen from Fig. 4(b), the PSNR does not change distinctly on both the CAFE database and the Harvard

database, as η_2 varies. The reason may be that the differences between \mathbf{Y}_{up} and ground truth \mathbf{X} are much bigger than the differences between \mathbf{X}_{cnn} and ground truth \mathbf{X} . Therefore, it is appropriate to select η_1 and η_2 as a very small positive number for the optimization problems (5) and (12). We set $\eta_1 = 5 \times 10^{-4}$ and $\eta_2 = 5 \times 10^{-4}$ for both the CAFE database and the Harvard database.

E. Experimental Results

Table I shows the average objective results of testing methods on 12 recovered HR-HSIs (testing data) of the CAFE database. The best results are marked in bold for clarity. As can be seen from Table I, the proposed DHSIS method performs consistently better than the other compared methods. Specifically, the advantage of DHSIS is obvious in terms of PSNR, SAM, SSIM, and ERGAS. In order to further compare the performance of the test methods in each spectral band, Fig. 5(a) shows the average PSNR curves as functions of the

TABLE I

AVERAGE QUANTITATIVE RESULTS (PSNR, SAM, SSIM, ERGAS, AND RUNNING TIME) OF THE TEST METHODS ON THE CAVE DATABASE [52]

Method	CAVE database [52]				
	PSNR	SAM	SSIM	ERGAS	Time (s)
Best Values	$+\infty$	0	1	0	0
CSU [37]	41.66	6.29	0.9817	1.190	516
NSSR [1]	44.70	4.31	0.9854	0.827	153
NLSTF [42]	42.70	6.47	0.9729	1.101	368
DHSIS	46.62	3.77	0.9912	0.640	9

TABLE II

AVERAGE QUANTITATIVE RESULTS (PSNR, SAM, SSIM, ERGAS, AND RUNNING TIME) OF THE TEST METHODS ON THE HARVARD DATABASE [53]

Method	Harvard database [53]				
	PSNR	SAM	SSIM	ERGAS	Time (s)
Best Values	$+\infty$	0	1	0	0
CSU [37]	45.40	3.74	0.9816	1.355	3641
NSSR [1]	46.04	3.60	0.9817	1.253	934
NLSTF [42]	45.67	3.86	0.9792	1.281	2058
DHSIS	46.42	3.47	0.9827	1.212	51

wavelengths of the spectral bands over the CAVE database for the test methods. As can be seen from Fig. 5(a), the DHSIS performs the best in the most of spectral bands among the test methods. Besides, all the test methods perform well in the middle spectral bands and relatively bad in the first and second bands. The reason is that there is a blur in these two bands [56]. For visual comparison, we show the reconstructed HR-HSIs at 12th and 31st bands and corresponding error images of the competing method for the test image *Peppers* (an HSI in the CAVE database) in Fig. 6. The error images reflect the differences between estimated images and ground truths. Three meaningful regions of the image are marked for easy visual comparison. It can be seen from the three marked meaningful regions and error images that the HR-HSIs reconstructed by CSU, NSSR, and NLSTF have blur details, stripe artifacts, and grid artifacts, respectively. The proposed DHSIS approach performs the best in reconstructing the detailed structures.

The average PSNR, SAM, SSIM, and ERGAS of the testing methods on 20 recovered HR-HSIs (testing data) of the Harvard database are reported in Table II. The proposed DHSIS method also outperforms other competing methods, as it has the lowest SAM and ERGAS and has the biggest PSNR and SSIM. Fig. 5(b) shows the PSNR curves as functions of the wavelengths of the spectral bands over the Harvard database for the test methods. As can be seen from Fig. 5(b), the DHSIS also performs the best in most spectral bands among the test methods. In Fig. 7, the reconstruction results and the corresponding error images of testing approaches for image *e0* at 30th and 31st bands in the Harvard database are shown. One meaningful region of each reconstructed HR-HSI is marked. As can be observed from Fig. 7, the HR-HSIs reconstructed by CSU, NSSR, and NLSTF have obvious blur and artifacts in the marked region. The DHSIS can provide the clearest detail in the marked region.

TABLE III

AVERAGE QUANTITATIVE RESULTS (PSNR, SAM, SSIM, AND ERGAS) OF \mathbf{Y}_{up} , \mathbf{X}_{in} , \mathbf{X}_{cnn} , AND \mathbf{X}_{fin} ON THE CAVE DATABASE [52]

Method	CAVE database [52]			
	PSNR	SAM	SSIM	ERGAS
Best Values	$+\infty$	0	1	0
\mathbf{Y}_{up}	25.89	7.62	0.8026	6.171
\mathbf{X}_{in}	42.87	4.81	0.9834	1.027
\mathbf{X}_{cnn}	45.97	3.92	0.9912	0.684
\mathbf{X}_{fin}	46.62	3.77	0.9912	0.640

TABLE IV

AVERAGE QUANTITATIVE RESULTS (PSNR, SAM, SSIM, AND ERGAS) OF \mathbf{Y}_{up} , \mathbf{X}_{in} , \mathbf{X}_{cnn} , AND \mathbf{X}_{fin} ON THE HARVARD DATABASE [53]

Method	Harvard database [53]			
	PSNR	SAM	SSIM	ERGAS
Best Values	$+\infty$	0	1	0
\mathbf{Y}_{up}	32.20	4.55	0.8316	4.273
\mathbf{X}_{in}	44.87	3.97	0.9789	1.779
\mathbf{X}_{cnn}	46.33	3.50	0.9825	1.246
\mathbf{X}_{fin}	46.42	3.47	0.9827	1.212

F. Running Time

The proposed DHSIS method mainly has three steps, i.e., initialize the HR-HSI from the fusion framework, learn the priors via deep CNN-based residual learning, and return the learned priors into the fusion framework. The first and third steps solve Sylvester equations of the same form, which has unique solution, and can be solved analytically and efficiently without any iteration. The second step is CNN-based residual learning, which is completely feedforward, and can be processed efficiently. However, other compared methods need to solve complex optimization problems. All the experiments are conducted by using the computer with Intel i7-6700 3.4-GHz CPU and 16-GB random access memory. The running time of test methods on the CAVE database and the Harvard database is shown in Tables I and II. As can be seen from Tables I and II, the proposed DHSIS is several times faster than the other compared methods.

G. Effectiveness of Three Steps in DHSIS

As mentioned in Section IV, the proposed DHSIS method has three steps, i.e., initialize the HR-HSI from the fusion framework, learn the priors via deep CNN-based residual learning, and return the learned priors into the fusion framework. Here, we illustrate the effectiveness of the three steps from the perspective of objective experimental results, respectively. Average quantitative results of \mathbf{Y}_{up} , \mathbf{X}_{in} , \mathbf{X}_{cnn} , and \mathbf{Y}_{fin} on the CAVE database and the Harvard database are shown in Tables III and IV, respectively. In the first step, based on the fusion framework, we acquire the initialized HR-HSI \mathbf{X}_{in} from \mathbf{Y}_{up} produced by bicubic interpolation on LR-HSI. As can be seen from Tables III and IV, \mathbf{X}_{in} has much better quantitative results on both the CAVE database and the Harvard database compared with \mathbf{Y}_{up} , which demonstrates that the initialization can obviously improve the qualities of input data of CNN. The fusion result produced by CNN \mathbf{X}_{cnn}

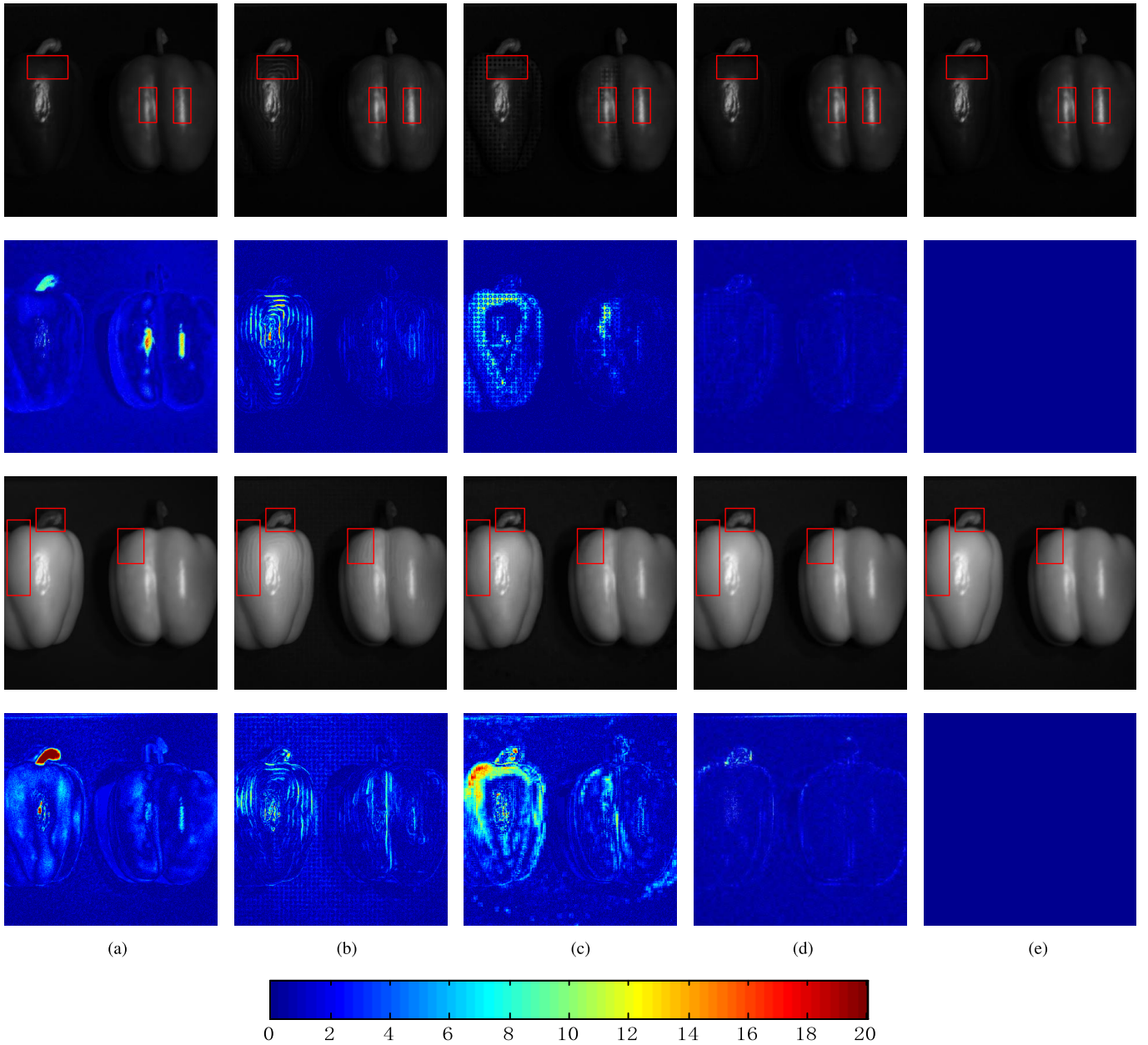


Fig. 6. First row: reconstructed images of *Peppers* (an HSI in the CAVE database) at 12th band. Second row: corresponding error images of the competing approaches for the image *Peppers* at 12th band. Third row: reconstructed images of *Peppers* at 31st band. Fourth row: corresponding error images of the competing approaches for the image *Peppers* at 31st band. (a) CSU [37]. (b) NSSR [1]. (c) NLSTF [42]. (d) DHSIS. (e) Ground truth.

also has obviously better quantitative indexes than the input of CNN \mathbf{X}_{in} , which illustrates that the priors leaned by deep CNN-based residual learning are really effective for the fusion. In the third step, \mathbf{X}_{cnn} is returned to the fusion framework to reconstruct the final HR-HSI \mathbf{X}_{fin} . \mathbf{X}_{fin} also has a raise in terms of quantitative results compared with \mathbf{X}_{cnn} . The raise is relatively obvious in the CAVE database and is sight in the Harvard database. Therefore, the final optimization with learned priors is also helpful for the reconstruction.

H. Effectiveness Residual Learning

We use the CNN-based residual learning to learn the priors of HR-HSI. In specific, we map the input \mathbf{X}_{in} into the residual

TABLE V
AVERAGE QUANTITATIVE RESULTS (PSNR, SAM, SSIM, AND ERGAS) OF NONRESIDUAL LEARNING (\mathbf{X}_{NFIN}) AND RESIDUAL LEARNING (\mathbf{X}_{FIN}) ON THE CAVE DATABASE [52]

Method	CAVE database [52]			
	PSNR	SAM	SSIM	ERGAS
Best Values	$+\infty$	0	1	0
\mathbf{X}_{nfin}	45.27	4.09	0.9873	0.738
\mathbf{X}_{fin}	46.62	3.77	0.9912	0.640

image \mathbf{X}_{res} , which can boost the learning performance. In order to demonstrate the effectiveness of the residual learning. We conduct the experiments of nonresidual learning on the

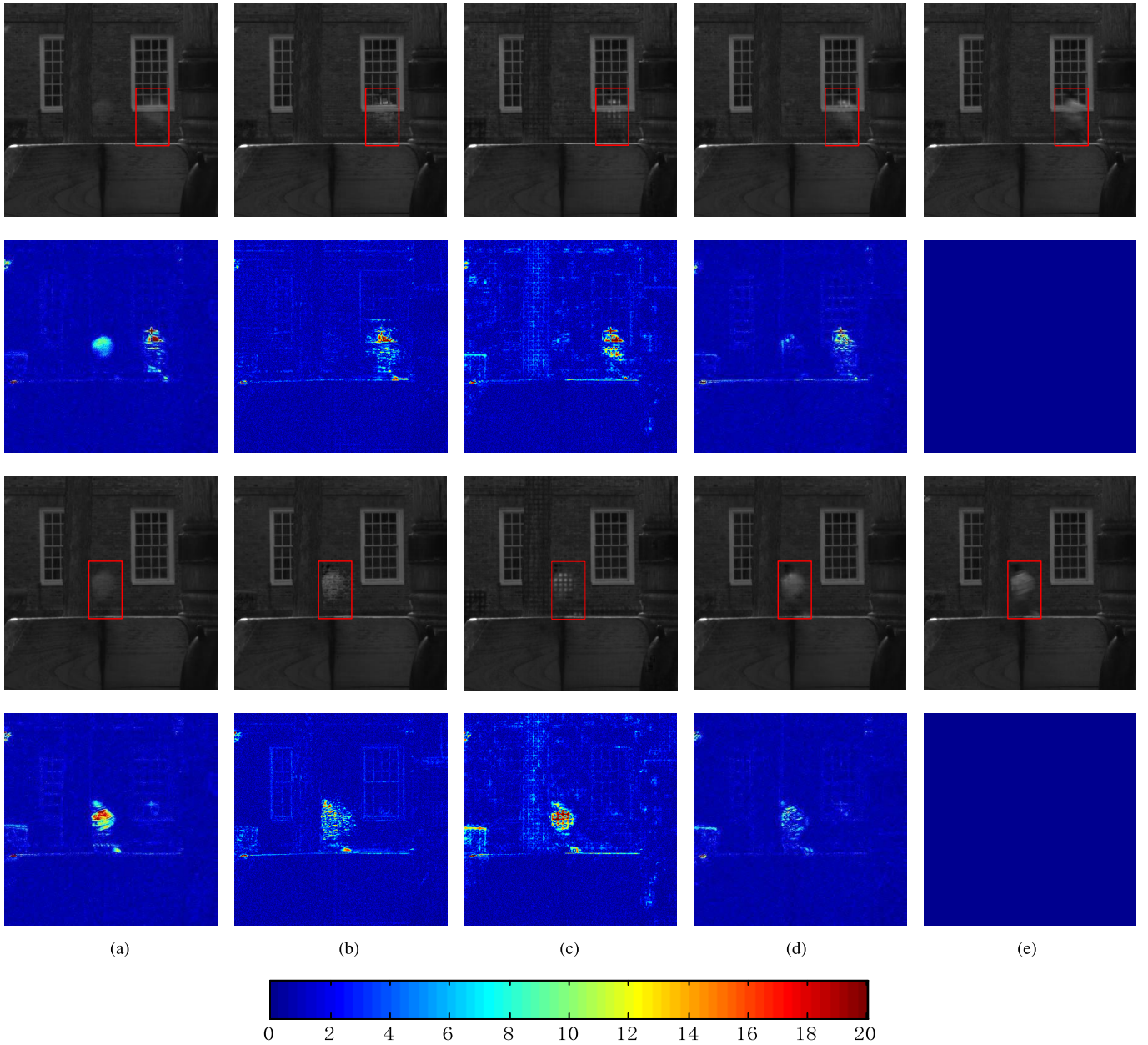


Fig. 7. First row: reconstructed images of $e0$ (an HSI in the Harvard database) at 30th band. Second row: corresponding error images of the competing approaches for the image $e0$ at 30th band. Third row: reconstructed images of $e0$ at 31st band in the Harvard database. Fourth row: corresponding error images of the competing approaches for the image $e0$ at 31st band. (a) CSU [37]. (b) NSSR [1]. (c) NLSTF [42]. (d) DHSIS. (e) Ground truth.

CAVE database, which directly maps \mathbf{X}_{in} into the ground truth image \mathbf{X} . The average quality results of residual learning and nonresidual learning on the CAVE database are shown in Table V, where \mathbf{X}_{nfin} and \mathbf{X}_{fin} denote the fusion results of nonresidual learning and residual learning, respectively. As can be seen from Table V, the performance of residual learning is better than that of nonresidual learning.

VI. CONCLUSION

In this paper, we present a DHSIS method, which fuses an LR-HSI with an HR-MSI counterpart to reconstruct an HR-HSI. In the proposed DHSIS method, instead of using traditional image priors, we utilize the priors learned by deep CNN-based residual learning to regularize the fusion problem.

Specifically, in the CNN training process, we first initialize the HR-HSI from the fusion framework as the input of CNN, which can not only help the reconstruction but also makes it convenient for the late residual learning process. Besides, the learned priors are also returned to the fusion framework to further improve the performance. Experimental results on two public HSI databases demonstrate the effectiveness of the proposed DSHIS method.

In the future works, the proposed DHSIS method can be extended in two directions. On one hand, we just utilize a simple and generally used CNN to learn the priors, and therefore, it is necessary to discover the CNN with different architectures, which may improve the performance. On the other hand, the traditional image priors, such as nonlocal

spatial similarities, sparse priors, and low-rank priors, can also be incorporated into the DHSIS method, which may further improve the performance.

REFERENCES

- [1] W. Dong *et al.*, "Hyperspectral image super-resolution via non-negative structured sparse representation," *IEEE Trans. Image Process.*, vol. 25, no. 5, pp. 2337–2352, May 2016.
- [2] J. M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. M. Nasrabadi, and J. Chanussot, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 2, pp. 6–36, Jun. 2013.
- [3] N. Akhtar and A. Mian, "Nonparametric, coupled, Bayesian, dictionary, and classifier learning for hyperspectral classification," *IEEE Trans. Neural Netw. Learn. Syst.*, to be published, doi: [10.1109/TNNLS.2017.2742528](https://doi.org/10.1109/TNNLS.2017.2742528).
- [4] Y. Gao, X. Wang, Y. Cheng, and Z. J. Wang, "Dimensionality reduction for hyperspectral data based on class-aware tensor neighborhood graph and patch alignment," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 8, pp. 1582–1593, Aug. 2015.
- [5] Q. Wang, J. Lin, and Y. Yuan, "Salient band selection for hyperspectral image classification via manifold ranking," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 6, pp. 1279–1289, Jun. 2016.
- [6] P. Zhong and R. Wang, "Jointly learning the hybrid CRF and MLR model for simultaneous denoising and classification of hyperspectral imagery," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 7, pp. 1319–1334, Jul. 2014.
- [7] M. J. Mendenhall and E. Merenyi, "Relevance-based feature extraction for hyperspectral images," *IEEE Trans. Neural Netw.*, vol. 19, no. 4, pp. 658–672, Apr. 2008.
- [8] X. Lin, Y. Liu, J. Wu, and Q. Dai, "Spatial-spectral encoded compressive hyperspectral imaging," *ACM Trans. Graph.*, vol. 33, no. 6, pp. 1–11, Nov. 2014.
- [9] Y. Gao, R. Ji, P. Cui, Q. Dai, and G. Hua, "Hyperspectral image classification through bilayer graph-based learning," *IEEE Trans. Image Process.*, vol. 23, no. 7, pp. 2769–2778, Jul. 2014.
- [10] N. Yokoya, C. Grohnfeldt, and J. Chanussot, "Hyperspectral and multispectral data fusion: A comparative review of the recent literature," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 2, pp. 29–56, Jun. 2017.
- [11] V. Ferraris, N. Dobigeon, Q. Wei, and M. Chabert, "Robust fusion of multiband images with different spatial and spectral resolutions for change detection," *IEEE Trans. Comput. Imag.*, vol. 3, no. 2, pp. 175–186, Jun. 2017.
- [12] Y. Qu, H. Qi, B. Ayhan, C. Kwan, and R. Kidd, "DOES multispectral/hyperspectral pansharpening improve the performance of anomaly detection?" in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2017, pp. 6130–6133.
- [13] L. Gómez-Chova, D. Tuia, G. Moser, and G. Camps-Valls, "Multimodal classification of remote sensing images: A review and future directions," *Proc. IEEE*, vol. 103, no. 9, pp. 1560–1584, Sep. 2015.
- [14] C. Yan, H. Xie, S. Liu, J. Yin, Y. Zhang, and Q. Dai, "Effective Uyghur language text detection in complex background images for traffic prompt identification," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 1, pp. 220–229, Jan. 2018.
- [15] C. Yan *et al.*, "A highly parallel framework for HEVC coding unit partitioning tree decision on many-core processors," *IEEE Signal Process. Lett.*, vol. 21, no. 5, pp. 573–576, May 2014.
- [16] C. Yan *et al.*, "Efficient parallel framework for HEVC motion estimation on many-core processors," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 12, pp. 2077–2089, Dec. 2014.
- [17] L. Zhou, Q. Zhu, Z. Wang, W. Zhou, and H. Su, "Adaptive exponential synchronization of multislave time-delayed recurrent neural networks with Lévy noise and regime switching," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 12, pp. 2885–2898, Dec. 2017.
- [18] R. Li and J. Cao, "Finite-time stability analysis for Markovian jump memristive neural networks with partly unknown transition probabilities," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 12, pp. 2924–2935, Dec. 2017.
- [19] Y. Sheng, Y. Shen, and M. Zhu, "Delay-dependent global exponential stability for delayed recurrent neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 12, pp. 2974–2984, Dec. 2017.
- [20] Y. Wang, Y. Xia, P. Zhou, and D. Duan, "A new result on H_∞ state estimation of delayed static neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 12, pp. 3096–3101, Dec. 2017.
- [21] T. Chen, L. Lin, L. Liu, X. Luo, and X. Li, "DISC: Deep image saliency computing via progressive representation learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 6, pp. 1135–1149, Jun. 2016.
- [22] H. Cecotti, M. P. Eckstein, and B. Giesbrecht, "Single-trial classification of event-related potentials in rapid serial visual presentation tasks using supervised spatial filtering," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 11, pp. 2030–2042, Nov. 2014.
- [23] C. Yan, H. Xie, D. Yang, J. Yin, Y. Zhang, and Q. Dai, "Supervised hash coding with deep neural network for environment perception of intelligent vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 1, pp. 284–295, Jan. 2018.
- [24] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, Jul. 2017.
- [25] M. Gong, J. Zhao, J. Liu, Q. Miao, and L. Jiao, "Change detection in synthetic aperture radar images based on deep neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 1, pp. 125–138, Jan. 2016.
- [26] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2016.
- [27] W. Samek, A. Binder, G. Montavon, S. Lapuschkin, and K.-R. Müller, "Evaluating the visualization of what a deep neural network has learned," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 11, pp. 2660–2673, Nov. 2017.
- [28] A. Dundar, J. Jin, B. Martini, and E. Culurciello, "Embedded streaming deep neural networks accelerator with applications," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 7, pp. 1572–1583, Jul. 2017.
- [29] J. Li, X. Mei, D. Prokhorov, and D. Tao, "Deep neural network for structural prediction and lane detection in traffic scene," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 3, pp. 690–703, Mar. 2017.
- [30] L. Alparone, L. Wald, J. Chanussot, C. Thomas, P. Gamba, and L. M. Bruce, "Comparison of pansharpening algorithms: Outcome of the 2006 GRS-S data-fusion contest," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3012–3021, Oct. 2007.
- [31] Q. Du, N. H. Younan, R. King, and V. P. Shah, "On the performance evaluation of pan-sharpening techniques," *IEEE Geosci. Remote Sens. Lett.*, vol. 4, no. 4, pp. 518–522, Oct. 2007.
- [32] S. Li and B. Yang, "A new pan-sharpening method using a compressed sensing technique," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 2, pp. 738–746, Feb. 2011.
- [33] S. Yang, K. Zhang, and M. Wang, "Learning low-rank decomposition for pan-sharpening with spatial-spectral offsets," *IEEE Trans. Neural Netw. Learn. Syst.*, to be published, doi: [10.1109/TNNLS.2017.2736011](https://doi.org/10.1109/TNNLS.2017.2736011).
- [34] J. Yang, X. Fu, Y. Hu, Y. Huang, X. Ding, and J. Paisley, "PanNet: A deep network architecture for pan-sharpening," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 5449–5457.
- [35] R. Kawakami, J. Wright, Y.-W. Tai, Y. Matsushita, M. Ben-Ezra, and K. Ikeuchi, "High-resolution hyperspectral imaging via matrix factorization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2011, pp. 2329–2336.
- [36] 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.
- [37] C. Lanaras, E. Baltsavias, and K. Schindler, "Hyperspectral super-resolution by coupled spectral unmixing," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 3586–3594.
- [38] Q. Wei, J. Bioucas-Dias, N. Dobigeon, J.-Y. Tourneret, M. Chen, and S. Godsill, "Multiband image fusion based on spectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7236–7249, Dec. 2016.
- [39] N. Akhtar, F. Shafait, and A. Mian, "Sparse spatio-spectral representation for hyperspectral image super-resolution," in *Proc. Eur. Conf. Comput. Vis.*, Sep. 2014, pp. 63–78.
- [40] M. Simoes, J. Bioucas-Dias, L. B. Almeida, and J. Chanussot, "A convex formulation for hyperspectral image superresolution via subspace-based regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 6, pp. 3373–3388, Jun. 2015.
- [41] Q. Wei, J. Bioucas-Dias, N. Dobigeon, and J. Y. Tourneret, "Hyperspectral and multispectral image fusion based on a sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 7, pp. 3658–3668, Jul. 2015.
- [42] R. Dian, L. Fang, and S. Li, "Hyperspectral image super-resolution via non-local sparse tensor factorization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 3862–3871.

- [43] M.-D. Iordache, J. Bioucas-Dias, and A. Plaza, "Sparse unmixing of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 2014–2039, Jun. 2011.
- [44] L. R. Tucker, "Some mathematical notes on three-mode factor analysis," *Psychometrika*, vol. 31, no. 3, pp. 279–311, Sep. 1996.
- [45] Q. Wei, N. Dobigeon, and J.-Y. Tournet, "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.
- [46] K. Zhang, M. Wang, and S. Yang, "Multispectral and hyperspectral image fusion based on group spectral embedding and low-rank factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 3, pp. 1363–1371, Mar. 2017.
- [47] L. Xu, J. S. S. Ren, C. Liu, and J. Jia, "Deep convolutional neural network for image deconvolution," in *Proc. Adv. Neural Inf. Process. Syst.*, Dec. 2014, pp. 1790–1798.
- [48] R. H. Bartels and G. W. Stewart, "Solution of the matrix equation $AX + XB = C$ [F4]," *Commun. ACM*, vol. 15, no. 9, pp. 820–826, 1972.
- [49] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.
- [50] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, Jul. 2015, pp. 448–456.
- [51] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. Int. Conf. Mach. Learn.*, Jun. 2010, pp. 807–814.
- [52] F. Yasuma, T. Mitsunaga, D. Iso, and S. K. Nayar, "Generalized assorted pixel camera: Postcapture control of resolution, dynamic range, and spectrum," *IEEE Trans. Image Process.*, vol. 19, no. 9, pp. 2241–2253, Sep. 2010.
- [53] A. Chakrabarti and T. Zickler, "Statistics of real-world hyperspectral images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2011, pp. 193–200.
- [54] H. Kwon and Y.-W. Tai, "RGB-guided hyperspectral image upsampling," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 307–315.
- [55] S. Rahmani, M. Strait, D. Merkurjev, M. Moeller, and T. Wittman, "An adaptive IHS pan-sharpening method," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 4, pp. 746–750, Oct. 2010.
- [56] N. Akhtar, F. Shafait, and A. Mian, "Hierarchical beta process with Gaussian process prior for hyperspectral image super resolution," in *Proc. Eur. Conf. Comput. Vis.*, Oct. 2016, pp. 103–120.
- [57] R. H. Yuhas, A. F. H. Goetz, and J. W. Boardman, "Discrimination among semi-arid landscape endmembers using the spectral angle mapper (SAM) algorithm," in *Proc. JPL Airborne Geosci. Workshop*, vol. 1, 1992, pp. 147–149.
- [58] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
- [59] L. Wald, "Quality of high resolution synthesised images: Is there a simple criterion?" in *Proc. Int. Conf. Fusion Earth Data*, Jan. 2000, pp. 99–103.



Renwei Dian (S'16) received the B.S. degree from the Wuhan University of Science and Technology, Wuhan, China, in 2015. He is currently pursuing the Ph.D. degree with the Laboratory of Vision and Image Processing, Hunan University, Changsha, China.

Since 2017, he has been a Visiting Ph.D. Student with the University of Lisbon, Lisbon, Portugal, supported by the China Scholarship Council. His current research interests include hyperspectral image super-resolution, image fusion, tensor decomposition, and deep learning.



Shutao Li (M'07–SM'15) received the B.S., M.S., and Ph.D. degrees from Hunan University, Changsha, China, in 1995, 1997, and 2001, respectively.

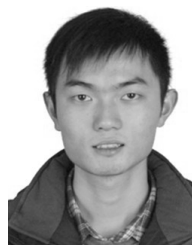
He was a Research Associate with the Department of Computer Science, The Hong Kong University of Science and Technology, Hong Kong, in 2011. From 2002 to 2003, he was a Post-Doctoral Fellow with the Royal Holloway College, University of London, London, U.K., with Prof. J. Shawe-Taylor. In 2005, he visited the Department of Computer Science, The Hong Kong University of Science and Technology, as a Visiting Professor. He joined the College of Electrical and Information Engineering, Hunan University, in 2001, where he is currently a Full Professor with the College of Electrical and Information Engineering. He has authored or co-authored over 160 refereed papers. His current research interests include compressive sensing, sparse representation, image processing, and pattern recognition.

Dr. Li was a recipient of two Second-Grade National Awards at the Science and Technology Progress of China in 2004 and 2006. He is currently an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING and the IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT, and a member of the Editorial Board of *Information Fusion and Sensing and Imaging*.



Anjing Guo (S'17) received the B.S. degree from Hunan University, Changsha, China, in 2016, where he is currently pursuing the Ph.D. degree with the Laboratory of Vision and Image Processing.

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



Leyuan Fang (S'10–M'14–SM'17) received the B.S. and Ph.D. degrees from the College of Electrical and Information Engineering, Hunan University, Changsha, China, in 2008 and 2015, respectively.

From 2011 to 2012, he was a Visiting Ph.D. Student with the Department of Ophthalmology, Duke University, Durham, NC, USA, supported by the China Scholarship Council. Since 2017, he has been an Associate Professor with the College of Electrical and Information Engineering, Hunan University. His current research interests include

sparse representation and multiresolution analysis in remote sensing and medical image processing.

Dr. Fang received the Scholarship Award for Excellent Doctoral Student granted by the Chinese Ministry of Education in 2011.