Regularizing Hyperspectral and Multispectral Image Fusion by CNN Denoiser

Renwei Dian [ID], *Student Member, IEEE*, Shutao Li [ID], *Fellow, IEEE*, and Xudong Kang [ID], *Senior Member, IEEE*

Abstract—**Hyperspectral image (HSI) and multispectral image (MSI) fusion, which fuses a low-spatial-resolution HSI (LR-HSI) with a higher resolution multispectral image (MSI), has become a common scheme to obtain high-resolution HSI (HR-HSI). This article presents a novel HSI and MSI fusion method (called as CNN-Fus), which is based on the subspace representation and convolutional neural network (CNN) denoiser, i.e., a well-trained CNN for gray image denoising. Our method only needs to train the CNN on the more accessible gray images and can be directly used for any HSI and MSI data sets without retraining. First, to exploit the high correlations among the spectral bands, we approximate the desired HR-HSI with the low-dimensional subspace multiplied by the coefficients, which can not only speed up the algorithm but also lead to more accurate recovery. Since the spectral information mainly exists in the LR-HSI, we learn the subspace from it via singular value decomposition. Due to the powerful learning performance and high speed of CNN, we use the well-trained CNN for gray image denoising to regularize the estimation of coefficients. Specifically, we plug the CNN denoiser into the alternating direction method of multipliers (ADMM) algorithm to estimate the coefficients. Experiments demonstrate that our method has superior performance over the state-of-the-art fusion methods.**

Index Terms—**Convolutional neural network (CNN), fusion, hyperspectral imaging, superresolution.**

Renwei Dian is with the College of Electrical and Information Engineering, Hunan University, Changsha 410082, China, also with the Key Laboratory of Visual Perception and Artificial Intelligence of Hunan Province, Hunan University, Changsha 410082, China, and also with the Instituto de Telecomunicações, Instituto Superior Técnico, Universidade de Lisboa, 1049-001 Lisbon, Portugal (e-mail: drw@hnu.edu.cn).

Shutao Li and Xudong Kang are with the College of Electrical and Information Engineering, Hunan University, Changsha 410082, China, and also with the Key Laboratory of Visual Perception and Artificial Intelligence of Hunan Province, Hunan University, Changsha 410082, China (e-mail: shutao_li@hnu.edu.cn; xudong_kang@163.com).

I. INTRODUCTION

HYPERSPECTRAL imaging technique can obtain images of different spectral bands simultaneously. The high spectral resolution of hyperspectral images (HSIs) provide faithful knowledge of the scene and enhance the performance of many applications, such as object classification [1]–[6], anomaly detection [7], and disease diagnosis [8]. However, due to the limited sun irradiance, there is the ineluctable tradeoff among the spatial resolution, spectral resolution, and signal-to-noise-ratio (SNR). In other words, HSI with high spectral resolution suffers from low spatial resolution to guarantee a high SNR [9]. Therefore, spatial resolution enhancement is a basic problem for HSI imaging. Compared with HSIs, multispectral images (MSIs) with lower spectral resolution can be obtained with much higher spatial resolution and SNR [9]. Hence, a possible way to solve the problem is image fusion, which combines a high-spatial-resolution MSI (HR-MSI) with a low-spatial-resolution HSI (LR-HSI) [10], [11]. The fusion technique has been widely used for HSI imaging due to its improved performance in many applications, including object classification [12], anomaly detection [13], and change detection [14]. Experiments performed in [13] demonstrate that the fused HR-HSI can really promote the detection accuracy.

Recently, convolutional neural network (CNN) has been successfully applied to solving the denoising problem for its high speed and good learning performance. Especially, Zhang *et al.* [15] proposed a flexible and fast CNN for denoising, which can adaptively deal with noisy images of different noise levels. Inspired by this work, we propose to use the well-trained CNN for gray image denoising to regularize the HSI-MSI fusion problem. In general, the proposed method has two steps: subspace estimation and coefficients estimation. To exploit the high correlations and redundancies in the spectral mode, the HR-HSI is decomposed as a low-dimensional spectral subspace and corresponding coefficients. Since most of the spectral information exists in the LR-HSI, we learn the spectral subspace from it via the singular value decomposition (SVD). With the spectral subspace known, we propose to plug the CNN denoiser into the ADMM algorithm to estimate the coefficients. In the ADMM iteration, we mainly need to solve two subproblems. One subproblem is the quadratic problem, which can be solved analytically and efficiently. Another subproblem can be seen as the denoising problem, and we solve it efficiently by applying a CNN denoiser for each row of coefficients. Finally, we can acquire the desired

HR-HSI with the coefficients and spectral subspace known. The proposed fusion method mainly has the following advantages.

1) The proposed CNN denoiser method has good generalizability and can flexibly deal with different HSI data sets without retraining.
2) Even though our target is HSI-MSI fusion, we do not need to train the CNN on any HSI and MSI data sets. In fact, the CNN denoiser is trained from more available gray images, and it is very ingenious to use the CNN trained from gray images for HSI recovery.
3) In comparison to the state-of-the-art fusion methods, our method has superior performance.

We arrange the remainder of this article as follows. We review recent works about HSI-MSI fusion in Section II. Our method for fusion of LR-HSI and HR-MSI is proposed in Section III. In Section IV, we present experiments and the corresponding discussions. The conclusion is drawn in Section V.

II. RELATED WORK

HSI and MSI fusion, which combines an LR-HSI with an HR-MSI, has become an effective way to obtain HR-HSI. Loncan *et al.* [16] gave a review of recent state-of-the-art fusion methods. The fusion approaches can be classified into four types: sparse representation-based approaches, low-rank representation based approaches, tensor factorization-based approaches, and deep-learning-based approaches.

Sparse representation-based HSI-MSI approaches assume that each spectral pixel can be sparsely represented by the learned spectral dictionary. Kawakami *et al.* [17] proposed to estimate the overcomplete spectral dictionary from the LR-HSI and then obtained the coefficients by conducting sparse coding for HR-MSI. Akhtar *et al.* [18] first learned the nonnegative dictionary and then estimate coefficients for each patch simultaneously via the simultaneous greedy pursuit approach. Based on the learned nonlocal similarities in the HR-MSI, Dong *et al.* [19] proposed a structured sparse coding approach to estimate the coefficients. Han *et al.* [20] proposed a sparse representation method to learn the local self-similarities in each superpixel and nonlocal self-similarities among patches.

Low-rank representation-based methods exploit the low-rank structure of the HR-HSI and decompose it as the coefficients and low-dimensional spectral basis, which exploits the high correlations among the spectral bands. In this way, the target of fusion is transformed as the estimation of the coefficients and spectral basis. Based on the linear spectral unmixing model, the HSI can be decomposed as endmembers and abundances, which has been used for single HSI super-resolution [21] and fusion [22]–[24]. The unmixing-based fusion methods [22]–[24] treat the spectral basis as endmembers and alternately update the endmembers and coefficients from the HR-MSI and LR-HSI by using the priors of spectral unmixing. The subspace representation model is different from the spectral unmixing model, where the subspace and coefficients are not necessarily nonnegative, and coefficients do not need to satisfy sum-to-one. Besides, the subspace is

often semiunitary. Works [25]–[27] obtain the spectral basis by conducting the SVD or vertex component analysis [28] on the LR-HSI, and they use different regularizer to estimate coefficients based on the maximum *a posteriori* (MAP) [29]. For example, Simoẽs *et al.* [25] made use of total variation regularization to improve spatial piecewise smoothness. Wei *et al.* [27] used a sparse representation-based regularizer to promote the self-similarities of image patches. Zhou *et al.* [30] and Veganzones *et al.* [31] emphasized the HR-HSI is locally low rank, and the subspace and coefficients are estimated for each local region separately.

Tensor factorization-based fusion methods deal with the 3-D HSI from the point of tensor. Work [32] first proposed a nonlocal sparse tensor factorization for HSI-MSI fusion, where they approximate the HR-HSI by dictionaries of three modes and a sparse core tensor based on the Tucker decomposition [33]. To exploit the nonlocal self-similarities, they assume that similar cubes can be sparsely represented by the same dictionaries. Furthermore, Li *et al.* [34] solved the fusion problem by simultaneously conducting sparse Tucker decomposition on the HR-MSI and LR-HSI, where the core tensor and three dictionaries are alternatively updated until convergence. Kanatsoulis *et al.* [35] proposed a canonical polyadic decomposition-based framework for the fusion problem, where the identifiability of the model is guaranteed under mild and realistic conditions. A weighted low-rank tensor recovery model is proposed by Chang *et al.* [36] to solve the fusion problem, which gives the weighted parameters to the elements of core tensor.

The deep-learning techniques, especially CNNs with deep structure, have been successfully applied to many image processing tasks. In the field of HR-MSI and LR-HSI fusion, Dian *et al.* [37] first used the imaging models to initialize the HR-HSI and then map the initialized HR-HSI to ground truth by the well-trained CNN. In addition, Yang *et al.* [38] proposed the two-branch CNN for the fusion, which makes use of the two-branch CNN to extract the spectral features of each pixel from the LR-HSI and its spatial features from HR-MSI. These methods mainly have two disadvantages. First, they need additional HR-MSI, LR-HSI, and HR-HSI data sets for pretraining, where these data are often not available. What is more, they do not have good generalizability, that is, the CNN trained on one data set cannot directly be applied to another data set of different spectral sizes. Besides, Han *et al.* [39] proposed the multibranch BP neural networks to fuse the LR-HSI and HR-MSI, which clusters the spectral bands for several groups for fusion.

III. PROPOSED CNN-FUS APPROACH

In this section, our CNN-Fus approach for the fusion of HR-MSI and LR-HSI is presented. The proposed CNN-Fus approach has two steps: estimation of subspace and estimation of coefficients.

A. Observation Model

All HR-HSI, HR-MSI, and LR-HSI can be naturally represented by the 3-D data. Here, we unfold the 3-D data with the spectral mode and denote them as the matrices. The HR-HSI
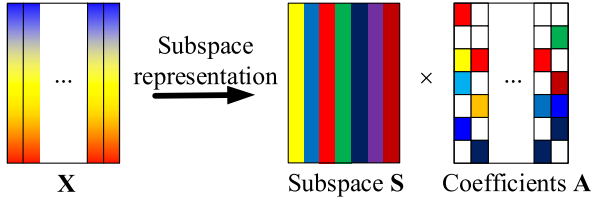
Fig. 1. Subspace representation of the HSI.

is denoted as $\mathbf{X} \in \mathbb{R}^{S \times N}$, where S and $N = W \times H$ are the number of bands and pixels, respectively. W and H are the lengths of the two spatial modes.

The acquired LR-HSI is denoted by $\mathbf{Y} \in \mathbb{R}^{S \times n}$, where n is the number of spectral pixels. Compared with $\mathbf{X}$, $\mathbf{Y}$ is spatially downsampled, and the relationship of them can be modeled as

$$\mathbf{Y} = \mathbf{XBD} + \mathbf{N}_h \qquad (1)$$

where $\mathbf{N}_h$ represents the additive Gaussian noise. $\mathbf{B} \in \mathbb{R}^{N \times N}$ models a blurring operation of hyperspectral camera, and it is a block circulant. Therefore, we can decompose $\mathbf{B}$ as

$$\mathbf{B} = \mathbf{FKF}^H \qquad (2)$$

where $\mathbf{F}$ stands for the discrete Fourier transform (DFT) matrix. The diagonal matrix $\mathbf{K}$ holds the eigenvalues of $\mathbf{B}$ in its diagonal line. $\mathbf{D} \in \mathbb{R}^{N \times n}$ is the spatially subsampling matrix.

$\mathbf{Z} \in \mathbb{R}^{s \times N}$ stands for the HR-MSI. Compared with $\mathbf{X}$, $\mathbf{Z}$ is spectrally downsampled, and the relationship of them is written as

$$\mathbf{Z} = \mathbf{RX} + \mathbf{N}_m \qquad (3)$$

where $\mathbf{R} \in \mathbb{R}^{s \times S}$ denotes the spectral downsampling matrix of MSI sensor and $\mathbf{N}_m$ represents the additive Gaussian noise.

B. Subspace Estimation

Hyperspectral data normally has low-rank structure, and therefore, it lives in a low-dimensional subspace [25], [40]. As shown in Fig. 1, the HR-HSI can be written as

$$\mathbf{X} = \mathbf{SA} + \mathbf{N} \qquad (4)$$

where $\mathbf{S} \in \mathbb{R}^{S \times L}$ and $\mathbf{A} \in \mathbb{R}^{L \times N}$ are the subspace and coefficients, respectively, and $\mathbf{N}$ denotes the additive noise. The subspace representation model mainly has three advantages: 1) it fully exploits high correlations among the spectral bands; 2) small values of L ($L < S$), which reduces the size of spectral mode and therefore makes computationally efficient; and 3) the subspace is semiunitary ($\mathbf{S}^T\mathbf{S} = \mathbf{I}$), and therefore, we have $\mathbf{A} = \mathbf{S}^T\mathbf{X}$. In this way, each row of $\mathbf{A}$ can be linearly expressed by the rows (bands) of $\mathbf{X}$, and rows of $\mathbf{A}$ preserve the spatial structures of the $\mathbf{X}$.

Based on the subspace representation, the target of the fusion is transformed into estimating the subspace $\mathbf{S}$ and coefficients $\mathbf{A}$. Since the spectral information mainly exists in the LR-HSI, we assume that the HR-HSI and LR-HSI share the same spectral subspace, and therefore, the spectral subspace can be estimated from the LR-HSI. We first conduct SVD on the LR-HSI, that is

$$\mathbf{Y} = \mathbf{U}_1 \Sigma_1 \mathbf{V}_1^T \qquad (5)$$

where $\mathbf{U}_1$ and $\mathbf{V}_1$ are semiunitary, and diagonal matrix Σ_1 contains the singular values, which are arranged in nonincreasing order. By only reserving L largest singular values, we give a low-dimensional approximation of $\mathbf{Y}$

$$\hat{\mathbf{Y}} = \hat{\mathbf{U}}_1 \hat{\Sigma}_1 \hat{\mathbf{V}}_1^T \qquad (6)$$

where $\hat{\mathbf{U}}_1 = \mathbf{U}_1(:, 1{:}L)$ and $\hat{\mathbf{V}}_1 = \mathbf{V}_1(:, 1{:}L)$. The subspace $\mathbf{S}$ is equal to

$$\mathbf{S} = \hat{\mathbf{U}}_1 = \mathbf{U}_1(:, 1{:}L). \qquad (7)$$

C. Estimation of Coefficients

With the subspace $\mathbf{S}$ known, we calculate the coefficients via the MAP estimation. By combining (1), (3), and (4), we can obtain the following equation:

$$\underset{\mathbf{A}}{\arg\min} \, ||\mathbf{Y} - \mathbf{SABD}||_F^2 + \alpha ||\mathbf{Z} - \mathbf{RSA}||_F^2 + \lambda\phi(\mathbf{A}) \qquad (8)$$

where $||\cdot||_F$ denotes the Frobenius norm. In (8), $||\mathbf{Y} - \mathbf{SABD}||_F^2 + \alpha ||\mathbf{Z} - \mathbf{RSA}||_F^2$ is the log-likelihood term, and $\lambda\phi(\mathbf{A})$ represents the prior on coefficients $\mathbf{A}$, where λ is the regularization parameter. Formulation (5) is based on the assumption that the noise in HSI and MSI is both independent identically distributed (i.i.d), and the parameter α models different noise variances in the two images. Hence, some prior information is needed to regularize the estimation of $\mathbf{A}$. The coefficients mainly reserve the spatial structures of the HR-HSI. Many handcraft priors have been used for the estimation of the coefficients, including priors of spectral unmixing [22], [23], [41], sparse priors [17], [27], [32], low-rank priors [42], nonlocal spatial similarities [19], and spatial smoothness [25]. Image denoising is a hot research topic in image processing. Many state-of-the-art denoising algorithms have been proposed to solve the problem, such as block-matching and 3-D filtering (BM3D) [43], weighted nuclear norm [44], and CNN denoiser [15], [45]. The recently proposed plug-and-play framework [46] makes it possible that we exploit the state-of-the-art denoising algorithms to solve other image restoration problems. The plug-and-play framework plugs a denoiser into an iterative algorithm, where the denoiser is treated as the proximity operator. Instead of using the handcraft priors, we use the prior learned from the images for the estimation of the coefficients. In specific, inspired by the spirit of plug-and-play, we use the well-trained CNN denoiser to regularize the estimation of coefficients. The CNN denoiser is trained from more available gray images, which can effectively capture the instinct spatial structures of images and remove the noise. The problem (8) is hard to solve directly. The ADMM algorithm can solve (8) by decomposing it as several treatable subproblems. By bringing variable $\mathbf{V}$, we acquire the augmented Lagrangian function

$$L(\mathbf{A}, \mathbf{V}, \mathbf{G}) = ||\mathbf{Y} - \mathbf{SABD}||_F^2 + \alpha ||\mathbf{Z} - \mathbf{RSA}||_F^2$$
$$\times \mu \left\| \mathbf{V} - \mathbf{A} + \frac{\mathbf{G}}{2\mu} \right\|_F^2 + \lambda\phi(\mathbf{V}) \qquad (9)$$
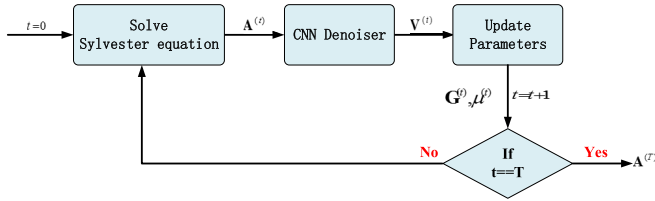
Fig. 2. Flowchart of the coefficients estimation.

where μ is the penalty parameter and $\mathbf{G}$ is the Lagrangian multiplier. Based on the ADMM algorithm, the problem (8) can be transformed as minimizing the augmented Lagrangian function (9). As shown in Fig. 2, the coefficients are estimated by iteratively updating $\mathbf{A}$, $\mathbf{V}$, $\mathbf{G}$, and μ.

1) Updating of $\mathbf{A}$: In each iteration, $\mathbf{A}$ is updated by minimizing $L(\mathbf{A}, \mathbf{V}, \mathbf{G})$ with regard to it, that is

$$\mathbf{A} \in \underset{\mathbf{A}}{\arg\min}\, L(\mathbf{A}, \mathbf{V}, \mathbf{G})$$
$$= \underset{\mathbf{A}}{\arg\min}\, ||\mathbf{Y} - \mathbf{SABD}||_F^2$$
$$+ \alpha ||\mathbf{Z} - \mathbf{RSA}||_F^2 + \mu \left\|\mathbf{V} - \mathbf{A} + \frac{\mathbf{G}}{2\mu}\right\|_F^2 \quad (10)$$

where problem (10) is strongly convex. Therefore, we force the derivative of (10) with respect to $\mathbf{A}$ to be 0 and acquire the Sylvester equation

$$\mathbf{H}_1\mathbf{A} + \mathbf{A}\mathbf{H}_2 = \mathbf{H}_3. \quad (11)$$

Since the spectral basis $\mathbf{S}$ is acquired by the SVD, it satisfies $\mathbf{S}^T\mathbf{S} = \mathbf{I}_L$, where $\mathbf{I}_L \in \mathbb{R}^{L \times L}$ is the identity matrix. Therefore, we can obtain

$$\mathbf{H}_1 = \alpha(\mathbf{RS})^T\mathbf{RS} + \mu\mathbf{I}_L$$
$$\mathbf{H}_2 = (\mathbf{BD})(\mathbf{BD})^T$$
$$\mathbf{H}_3 = \alpha(\mathbf{RS})^T\mathbf{Z} + \mathbf{S}^T\mathbf{Y}(\mathbf{BD})^T + \mu\left(\mathbf{V} + \frac{\mathbf{G}}{2\mu}\right). \quad (12)$$

Since the matrices $\mathbf{H}_1$ and $\mathbf{H}_2$ are positive, the system matrix of (11) is positive, and therefore, we can apply the conjugate gradient method to solve it. Here, we use a fast method to solve (11) analytically and efficiently [26], [47]. The matrix $\mathbf{H}_1$ is symmetric and positive, and therefore, it can be diagonalized by eigendecomposition, that is

$$\mathbf{H}_1 = \mathbf{Q}_1\Lambda\mathbf{Q}_1^{-1}. \quad (13)$$

where diagonal matrix Λ is written as

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_L \end{bmatrix} \quad (14)$$

and matrix $\mathbf{Q}_1$ is invertible. By multiplying left-hand side of (11) by $\mathbf{Q}_1^{-1}$, we can obtain

$$\Lambda\mathbf{Q}_1^{-1}\mathbf{A} + \mathbf{Q}_1^{-1}\mathbf{A}\mathbf{H}_2 = \mathbf{Q}_1^{-1}\mathbf{H}_3. \quad (15)$$

By multiplying right-hand side of (15) with the DFT matrix $\mathbf{F}$ and combining (2), the following equation is acquired:

$$\Lambda\mathbf{Q}_1^{-1}\mathbf{AF} + \mathbf{Q}_1^{-1}\mathbf{AFKF}^H\mathbf{DD}^H\mathbf{FK}^H = \mathbf{Q}_1^{-1}\mathbf{H}_3\mathbf{F}. \quad (16)$$

Lemma 1 (See [47]): The following equation is satisfied:

$$\mathbf{F}^H\mathbf{DD}^H\mathbf{F} = \frac{1}{d}(\mathbf{1}_d \otimes \mathbf{I}_n)(\mathbf{1}_d^T \otimes \mathbf{I}_n) \quad (17)$$

where $\mathbf{1}_d \in \mathbb{R}^d$ is a vector of ones and $\mathbf{I}_n \in \mathbb{R}^{n \times n}$ is the identity matrix. Here, n is number of pixels of LR-HSI, and $d = (N/n)$.

By combining (16) and (17), we can acquire the following equation:

$$\Lambda\bar{\mathbf{A}} + \bar{\mathbf{A}}\mathbf{M} = \mathbf{C} \quad (18)$$

where $\bar{\mathbf{A}} = \mathbf{Q}_1^{-1}\mathbf{AF}$, $\mathbf{M} = (1/d)\bar{\mathbf{K}}\bar{\mathbf{K}}^H$, $\bar{\mathbf{K}} = \mathbf{K}(\mathbf{1}_d \otimes \mathbf{I}_n)$, and $\mathbf{C} = \mathbf{Q}_1^{-1}\mathbf{H}_3\mathbf{F}$. Equation (18) is the Sylvester equation with regard to $\bar{\mathbf{A}}$. We rewrite the diagonal matrix $\mathbf{K}$ as

$$\mathbf{K} = \begin{bmatrix} \mathbf{K}_1 & 0 & \cdots & 0 \\ 0 & \mathbf{K}_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{K}_d \end{bmatrix} \quad (19)$$

where $\mathbf{K}_i \in \mathbb{C}^{n \times n}$. Therefore, we have $\bar{\mathbf{K}}\bar{\mathbf{K}}^H = \sum_{t=1}^d \mathbf{K}_d^2$. Equation (18) can be solved in an row-by-row manner. We first rewrite $\bar{\mathbf{A}}$ and $\mathbf{C}$ as $\bar{\mathbf{A}} = [\bar{\mathbf{a}}_1, \ldots, \bar{\mathbf{a}}_L]^T$ and $\mathbf{C} = [\mathbf{c}_1, \ldots, \mathbf{c}_L]^T$, respectively, where $\bar{\mathbf{a}}_i$ and $\mathbf{c}_i$ represent ith row of $\bar{\mathbf{A}}$ and $\mathbf{C}$, respectively. In this way, $\bar{\mathbf{A}}$ can be estimated row-by-row manner, that is

$$\lambda_i\bar{\mathbf{a}}_i + \bar{\mathbf{a}}_i\mathbf{M} = \mathbf{c}_i, \quad \text{for } i = 1, \ldots, L. \quad (20)$$

We can obtain

$$\bar{\mathbf{a}}_i = \mathbf{c}_i(\lambda_i\mathbf{I}_n + \mathbf{M})^{-1}, \quad \text{for } i = 1, \ldots, L. \quad (21)$$

By using $\bar{\mathbf{K}}\bar{\mathbf{K}}^H = \sum_{t=1}^d \mathbf{K}_d^2$, $(\lambda_i\mathbf{I}_n + \mathbf{M})^{-1}$ can be computed as

$$(\lambda_i\mathbf{I} + \mathbf{M})^{-1} = \lambda_i^{-1}\mathbf{I}_n - \lambda_i^{-1}\bar{\mathbf{K}}\left(\lambda_i d\mathbf{I}_n + \sum_{t=1}^d \mathbf{K}_d^2\right)^{-1}\bar{\mathbf{K}}^H. \quad (22)$$

Hence, $\bar{\mathbf{a}}_i$ is equivalent as

$$\bar{\mathbf{a}}_i = \lambda_i^{-1}\mathbf{c}_i - \lambda_i^{-1}\mathbf{c}_i\bar{\mathbf{K}}\left(\lambda_i d\mathbf{I}_n + \sum_{t=1}^d \mathbf{K}_d^2\right)^{-1}\bar{\mathbf{K}}^H. \quad (23)$$

After obtaining $\bar{\mathbf{A}}$, the coefficient $\mathbf{A}$ is estimated as

$$\mathbf{A} = \mathbf{Q}_1\bar{\mathbf{A}}\mathbf{F}^H. \quad (24)$$

Algorithm 1 outlines the method for solving (11).

Algorithm 1 Solution for (11) With Respect to $\mathbf{A}$

Input: $\mathbf{H}_1$, $\mathbf{H}_2$, $\mathbf{H}_3$, $\mathbf{S}$, $\mathbf{B}$
1: $\mathbf{B} = \mathbf{FKF}^H$;
2: $\tilde{\mathbf{K}} = \mathbf{K}(\mathbf{1}_d \otimes \mathbf{1}_n)$;
3: $\mathbf{H}_1 = \mathbf{Q}_1\Lambda\mathbf{Q}_1^{-1}$;
4: $\mathbf{C} = \mathbf{Q}_1^{-1}\mathbf{H}_3\mathbf{F}$;
5: **for** $i = 1$ to L **do**
6: $\quad \bar{\mathbf{a}}_l = \lambda_l^{-1}\mathbf{c}_l - \lambda_l^{-1}\mathbf{c}_l\bar{\mathbf{K}}(\lambda_l d\mathbf{I}_n + \sum_{t=1}^d \mathbf{K}_t^2)\bar{\mathbf{K}}^H$;
7: **end for**
8: Set $\mathbf{A} = \mathbf{Q}_1\bar{\mathbf{A}}\mathbf{F}^H$;
Output: $\mathbf{A}$.

---

2) Updating of $\mathbf{V}$*:* In each iteration, $\mathbf{V}$ is updated by minimizing the Lagrangian function with regard to it, leading to

$$\mathbf{V} \in \underset{\mathbf{V}}{\arg\min}\, L(\mathbf{A}, \mathbf{V}, \mathbf{G})$$
$$= \underset{\mathbf{V}}{\arg\min}\, \mu \|\mathbf{V} - \mathbf{A} + \frac{\mathbf{G}}{2\mu}\|_F^2 + \lambda \phi(\mathbf{V}). \qquad (25)$$

Inspired by the spirit of plug and play, the optimization problem (25) can be seen as the denoising of $\mathbf{A} - (\mathbf{G}/2\mu)$ with a white additive Gaussian noise of variation $\sigma^2 = (\lambda/2\mu)$. Many state-of-the-art algorithms have been proposed to solve the denoising problem, such as BM3D [43], K-SVD denoising method [48], weighted nuclear norm method [49], and deep CNN-based denoising method [15], [45].

Due to the high speed and powerful learning performance of CNN, we take the recently proposed FFDNet [15], a flexible and fast CNN-based method, as the denoising engine. The FFDNet consists of three kinds of operations: 3×3 convolution layer (Conv), batch rectified linear units (ReLUs) [50], and normalization (BN) [51]. In specific, the first layer is "Conv+ReLU," the middle layer is "Conv+BN+ReLU," and the last layer is "Conv," where BN is exploited to speed up the training procedure, and ReLU $\max(0, x)$ is the activation function. Using FFDNet as denoising engine mainly has three advantages. First, FFDNet uses a tunable noise level map as input, and therefore, the FFDNet can flexibly deal with images of different noise levels without retraining. Besides, the FFDNet decomposes the noisy image as four subimages and uses these subimages as the input, which can reduce the number of layers and make the algorithm much faster. Then, the denoised subimages are aggregated to acquire the final denoised image. What is more, we do not need to train the FFDNet on any HSI or MSI data sets and are trained on more available gray images. It is very ingenious to use CNN trained from gray images for HSI recovery.

Since the subspace $\mathbf{E}$ is semiunitary, we have $\mathbf{A} = \mathbf{S}^T\mathbf{X}$, and each row of $\mathbf{A}$ can be linearly expressed by rows of $\mathbf{X}$. Therefore, each row of $\mathbf{A}$ preserves the spatial structures of the HR-HSI. Besides, even though the bands of HR-HSI are highly correlated, the rows of coefficients are much less correlated due to the subspace representation. Based on this conclusion, we apply the well-trained FFDNet to each row of $\mathbf{A} - (\mathbf{G}/2\mu)$ to solve problem (25). In other words, we estimate $\mathbf{V}$ in (25) in a row-by-row manner via the learned map $\mathcal{F}$, that is

$$\mathbf{V}(i, :) = \mathcal{F}\left(\mathbf{H}(i, :), \frac{\lambda}{2\mu}; \Theta\right), \quad \text{for } i =, 1, 2, \dots, L \qquad (26)$$

where Θ is the parameters of the FFDNet, and the variable $\mathbf{H}$ satisfies $\mathbf{H} = \mathbf{A} - (\mathbf{G}/2\mu)$. Since each row of $\mathbf{H}$ is not an image, we need first to scale each row of it to $[0, 1]$ and then reshape it as the matrix $\hat{\mathbf{H}}_i$ of size $W \times H$, where W and H are the dimensions of two spatial modes. The abovementioned operation is represented by $\hat{\mathbf{H}}_i = \text{vec}^{-1}(c_i\mathbf{H}(i, :) + b_i)$. Here, $\text{vec}^{-1}(\cdot)$ denotes the inverse operation of vectorization, and transforms a vector $\mathbf{a} \in \mathbb{R}^{WH}$ into a matrix $\mathbf{B} \in \mathbb{R}^{W \times H}$, where $\mathbf{B} \in \mathbb{R}^{W \times H} = \text{vec}^{-1}(\mathbf{a})$ is equivalent as $\mathbf{B}(i, j) = \mathbf{a}((j - 1) * W + i)$. In this way, the noise level is changed as $c_i\sigma^2$ correspondingly. Finally, we need to scale the denoising results

back. Therefore, the estimation of $\mathbf{V}$ in (26) is transformed as

$$\hat{\mathbf{H}}_i = \text{reshape}(c_i\mathbf{H}(i, :) + b_i, W, H),$$
$$\text{for } i =, 1, 2, \dots, L$$
$$\hat{\mathbf{V}}(i, :) = \text{reshape}\left(\mathcal{F}\left(\hat{\mathbf{H}}_i, \frac{c_i\lambda}{2\mu}; \Theta\right), 1, N\right),$$
$$\text{for } i =, 1, 2, \dots, L$$
$$\mathbf{V}(i, :) = \frac{\hat{\mathbf{V}}(i, :) - b_i}{c_i}, \quad \text{for } i =, 1, 2, \dots, L \qquad (27)$$

where $c_i = (1/(\max(\mathbf{H}(i, :)) - \min(\mathbf{H}(i, :))))$ and $b_i = -c_i * \min(\mathbf{H}(i, :))$.

3) Updating of Lagrangian Multiplier $\mathbf{G}$*:* The Lagrangian multiplier $\mathbf{G}$ is updated via

$$\mathbf{G} = \mathbf{G} + 2\mu(\mathbf{V} - \mathbf{A}). \qquad (28)$$

4) Updating of Penalty Parameter μ*:* The penalty parameter μ has an important effect on the FFDNet denoising process, where the input variance $\sigma^2 = (\lambda/2\mu)$. With the iteration of the algorithm, the image is more close to the clean image, and therefore, the input noise level σ^2 should also be turned down. In this way, we need to increase the value of μ in the iteration. Another advantage of increasing μ is that it can make the algorithm converge. The penalty is updated by

$$\mu = \gamma \mu \qquad (29)$$

where γ is the constant satisfying $\gamma > 1$.

We summarize our method for HSI-MSI fusion in Algorithm 2. The algorithm is stopped when the number of iterations is reached a preset value T, and we set $T = 12$ in the experiments.

Algorithm 2 CNN-Fus-Based HSI-MSI Fusion

Input: $\mathbf{Y}$, $\mathbf{Z}$, λ
1: Estimate the subspace $\mathbf{S}$ via (7);
2: **while** not converged **do**
3:    Update $\mathbf{A}$ via Algorithm 1;
4:    Update $\mathbf{V}$ via equation (27);
5:    Update $\mathbf{G}$ via equation (28);
6:    Update μ via equation (29);
7: **end while**
8: $\mathbf{X} = \mathbf{SA}$;
Output: $\mathbf{X}$.

IV. Experiments

In this section, the experiments of HSI-MSI fusion are conducted on two simulated data sets and one real data set to evaluate the effectiveness of our method. The source code will be available in https://sites.google.com/view/renweidian.

A. Experimental Data Sets

1) Pavia University: This HSI is acquired over the urban area of Pavia University [52]. It has the size of $610 \times 340 \times 115$, where the number of bands is 115. Since some bands have low SNR, these bands are removed and 93 bands are perseverd.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

6

IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS

The HSI is used as a reference image. The LR-HSI is simulated by using a 7×7 Gaussian filter with standard deviation 2 and then by subsampling every 5 pixels in two spatial modes. We simulate the HR-MSI **Z** by filtering **X** with the IKONOS-like reflectance spectral response. The i.i.d Gaussian noise is added to HR-MSI (35 dB) and LR-HSI (30 dB).

2) Cuprite Mine: The Cuprite Mine is acquired by the AVIRIS [53] in Nevada. The HSI has a size of $512 \times 512 \times 224$. The HSI covers the wavelength range of 400–2500 nm with 10-nm interval. Bands 1, 2, 105–115, 150–170, and 223–224 have water absorptions and low SNR, and they are removed. The LR-HSI is simulated by using an 7×7 Gaussian filter with standard deviation 2 and then by subsampling every 4 pixels. Six bands of wavelengths 480, 560, 660, 830, 1650, and 2220 nm are directly selected as the HR-MSI, which correspond to the visible and midinfrared range spectral bands of USGS/NASA Landsat7 satellite. The i.i.d Gaussian noise is added to HR-MSI (35 dB) and LR-HSI (30 dB).

B. Compared Methods

Three recent state-of-the-art fusion approaches are used for comparison, which includes the nonnegative structured sparse representation (NSSR) [19],[1] coupled spectral unmixing (CSU) [23],[2] coupled sparse tensor factorization (CSTF) [34],[3] coupled nonnegative matrix factorization (CNMF), and Bayesian sparse fusion method (Fuse-S). The NSSR, CSU, and CSTF belong to the sparse representation-based methods, low-rank representation-based methods, and tensor factorization-based methods, respectively. We tune the parameters of the compared methods for the best performance. For the NSSR, the number of atoms in spectral dictionary K, nonlocal similarities regularization parameter η_1, and sparsity regularization parameter η_2 are set as $K = 75$, $\eta_1 = 10^{-4}$, and $\eta_2 = 10^{-4}$, respectively. For the CSU, the number of iterations is set as 1500. For the CSTF, the parameters are set as $n_w = 500$ and $n_h = 500$ and $n_w = 500$ and $n_h = 300$ for Cuprite Mine and Pavia University, respectively. The other parameters of CSTF are set as $n_s = 15$, $\beta = 0.01$, and $\lambda = 10^{-5}$. All codes of compared methods are publicly available.

C. Quantitative Metrics

We use four quantitative metrics to measure the quality of the recovered HR-HSIs.

1) PSNR: The first quantitative metric is the peak signal-to-noise ratio (PSNR). The PSNR that is extended for HSI is by computing average PSNR of all bands.

2) SAM: The spectral angle mapper (SAM) [54] is the average angle between the estimated and referenced spectral pixels.

3) UIQI: The universal image quality index (UIQI) [55] for HSI is calculated on a sliding window and then is averaged on all bands and all windows. For two windows **a** and **b**, the

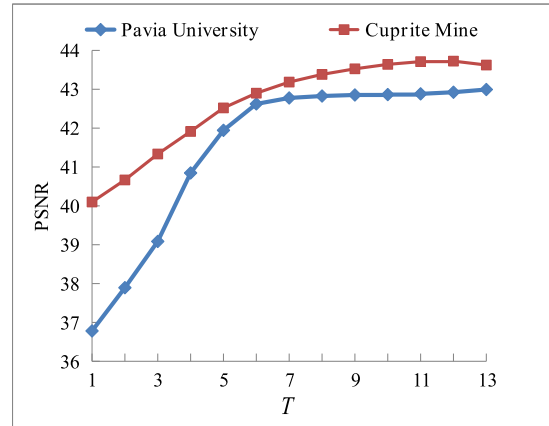[1]http://see.xidian.edu.cn/faculty/wsdong
[2]https://github.com/lanha/SupResPALM
[3]https://sites.google.com/view/renweidian

Fig. 3. Effect of parameter T on our method.

UIQI is defined as

$$\text{UIQI}(\mathbf{a}, \mathbf{b}) = \frac{4\mu_{\mathbf{a}}\mu_{\mathbf{b}}}{\mu_{\mathbf{a}}^2 + \mu_{\mathbf{b}}^2} \frac{\sigma_{\mathbf{a},\mathbf{b}}^2}{\sigma_{\mathbf{a}}^2 + \sigma_{\mathbf{b}}^2} \qquad (30)$$

where $\mu_{\mathbf{a}}$ and $\sigma_{\mathbf{a}}$ denote the mean value and standard variance of **a**, respectively, and $\sigma_{\mathbf{a},\mathbf{b}}^2$ stands for the covariances between **a** and **b**.

4) SSIM: The structural similarity index (SSIM) [56] is used to measure the structural similarities of the gray image. The SSIM is extended to evaluate the qualities of HSI by averaging on all spectral bands.

D. Parameters Selection

The proposed method has three important parameters, i.e., the maximum number of iterations T, the dimensional of the spectral subspace L, and the regularization parameter λ.

The estimation of the coefficients **A** is the iterative process. The maximum number of iterations T is an important stop criterion of our method. In the Tth iteration, we obtain the coefficients $\mathbf{A}^{(T)}$ and estimated HR-HSI $\mathbf{Z}^{(T)} = \mathbf{S}\mathbf{A}^{(T)}$. To show the results of the intermediate steps, Fig. 3 shows that the PSNR of estimated HR-HSI varies from T. We can see from Fig. 3 that the PSNR for Pavia University has a sharp rise when T varies from 1 to 7, and then, it keeps relatively stable. The PSNR for Cuprite Mine rises as T changes from 1 to 10 and then does not have obvious change. Therefore, we set the maximum number of iterations as 12 for our method.
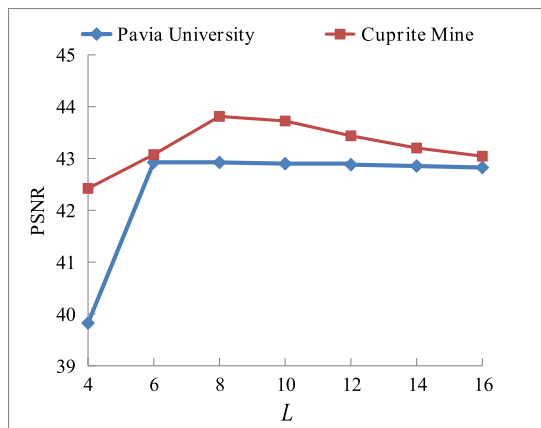
The parameter L controls the dimension of the spectral basis, which can highly influence the final result. To test the effect of L, we plot the PSNR as a function of L in Fig. 4. It can be seen from Fig. 4 that the PSNR for Pavia University has a sharp increase as L varies from 4 to 6, and then, the PSNR keeps relatively stable. The PSNR for Cuprite Mine increases when L ranges from 4 to 8, and then, it decreases when L is bigger than 8. Therefore, we set the dimension of spectral basis $L = 8$ for the best performance, which indicates that only eight atoms are enough to represent the spectral information, and the spectral vectors really live in a low-dimensional subspace.

Since the noise level is $\sigma^2 = (\lambda/2\mu)$ in each iteration, the parameter λ highly influences the performance of

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

DIAN *et al.*: REGULARIZING HSI AND MSI FUSION BY CNN DENOISER

7

Fig. 4. Effect of parameter L on our method.

Fig. 5. Effect of parameter λ on our method.

CNN denoiser. To discuss the influence of λ, we plot the PSNR for Pavia University and Cuprite Mine as a function of λ in Fig. 5. The parameter $\lambda = 0$ means that the CNN denoiser is not used for the estimation of the coefficients. We can see from Fig. 5 that the PSNR for Pavia University and Cuprite simultaneously has an obvious increase of λ that grows from 0 to 8×10^{-4}, which indicates that our method highly relies on the estimation of the coefficients, and the CNN denoiser really helps the estimation of the coefficients. The PSNR for Pavia University and Cuprite decreases as λ is bigger than 1.2×10^{-3}. Therefore, we set $\lambda = 1 \times 10^{-3}$ for both Pavia University and Cuprite Mine.

E. Experimental Results

1) Experimental Results on Simulated Data Fsuion: The quality matrices of the HR-HSIs reconstructed by compared approaches on Pavia University are reported in Table I. We highlight the best results in bold for clarity. From Table I, we observe that the proposed method and Fuse-S consistently outperform the other testing approaches in terms of the quality metrics. The superiority of our method mainly comes from the low-dimensional subspace representation and CNN denoiser, where the low-dimensional subspace representation can effectively model the correlations among the spectral bands, and the

TABLE I
QUANTITATIVE METRICS OF THE COMPARED APPROACHES ON PAVIA UNIVERSITY [52]

Method	Pavia University [52]			
	PSNR	SAM	UIQI	SSIM
Best Values	$+\infty$	0	1	1
NSSR [19]	39.455	3.520	0.979	0.968
CSU [23]	40.607	2.671	0.986	0.980
CSTF [34]	41.468	2.554	0.988	0.979
CNMF [22]	42.554	2.379	0.990	0.984
Fuse-S [27]	42.994	2.284	0.991	0.986
CNN-Fus	**43.017**	**2.235**	**0.992**	**0.987**

TABLE II
QUANTITATIVE METRICS OF THE COMPARED APPROACHES ON CUPRITE MINE [53]

Method	Cuprite Mine [53]			
	PSNR	SAM	UIQI	SSIM
Best Values	$+\infty$	0	1	0
NSSR [19]	38.912	1.730	0.866	0.929
CSU [23]	42.279	1.283	0.929	0.960
CSTF [34]	41.038	1.262	0.915	0.950
CNMF [22]	42.200	1.279	0.924	0.959
Fuse-S [27]	43.694	1.153	0.942	0.968
CNN-Fus	**44.008**	**1.144**	**0.944**	**0.970**

CNN denoiser can well depict the spatial prior of the HSI. The reconstructed Pavia University at the 20th and 50th bands by the CSTF, CNMF, Fuse-S, and CNN-Fus and corresponding error images are shown in Fig. 6. The images reconstructed by NSSR and CSU are not shown since they perform relatively worse in this data set. From the recovered HR-HSIs, all testing methods perform well in recovering the details compared with the observed LR-HSI, which indicates the effectiveness of these methods. The error images reflect the differences between the estimated images and ground truths. From the error images, we can see that the HR-HSIs recovered by our method have fewer errors and are more close to the ground truths.

The quality matrices of the HR-HSIs reconstructed by compared approaches on Cuprite Mine are shown in Table II. Our method still delivers the best results among the testing approaches, and the Fuse-S takes the second place. In Fig. 7, the fusion results and the corresponding error images of CSU, CNMF, Fuse-S, and CNN-Fus for Cuprite Mine at the 50th and 90th bands are shown. The images reconstructed by NSSR and CSTF are not shown since they perform relatively worse in this data set. As we can observe from Fig. 7, the HR-HSIs reconstructed by our method and Fuse-S still have fewer errors and higher PSNR.

2) Experimental Results on Real Data Fusion: To further assess the performance of our method, we have tested the compared methods on real LR-HSI and HR-MSI fusion. The LR-HSI is captured by the Hyperion sensor loaded on the Earth Observing-1 satellite. The LR-HSI has the spatial resolution of 30 m and 220 spectral bands in the spectral range of 400–2500 nm. After removing the bands of low SNR, 89 bands are retained. An area of spatial size 100×100 is used
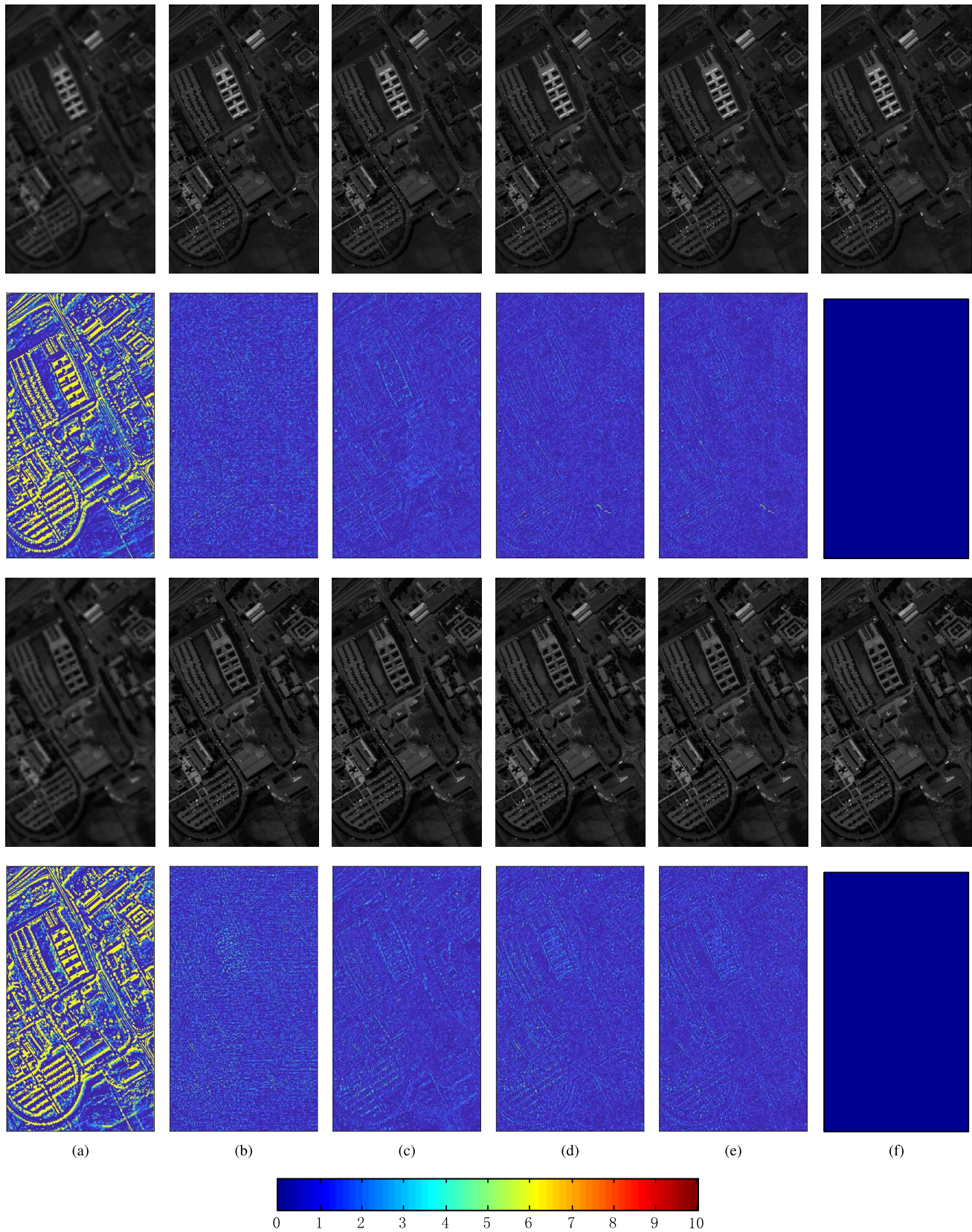
Fig. 6. First and second rows show the fusion results and corresponding error images by the CSTF (42.525 dB), CNMF (43.308 dB), Fuse-S (44.694 dB), and CNN-Fus (44.806 dB) at the 20th band of Pavia University. The third and fourth rows show the fusion results and corresponding error images by the CSTF (41.361 dB), CNMF (43.355 dB), Fuse-S (42.945 dB), and CNN-Fus (44.192 dB) at the 50th band of Pavia University. (a) LR-HSI. (b) CSTF [34]. (c) CNMF [22]. (d) Fuse-S [27]. (e) CNN-Fus. (f) Ground truth.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

DIAN *et al.*: REGULARIZING HSI AND MSI FUSION BY CNN DENOISER

9

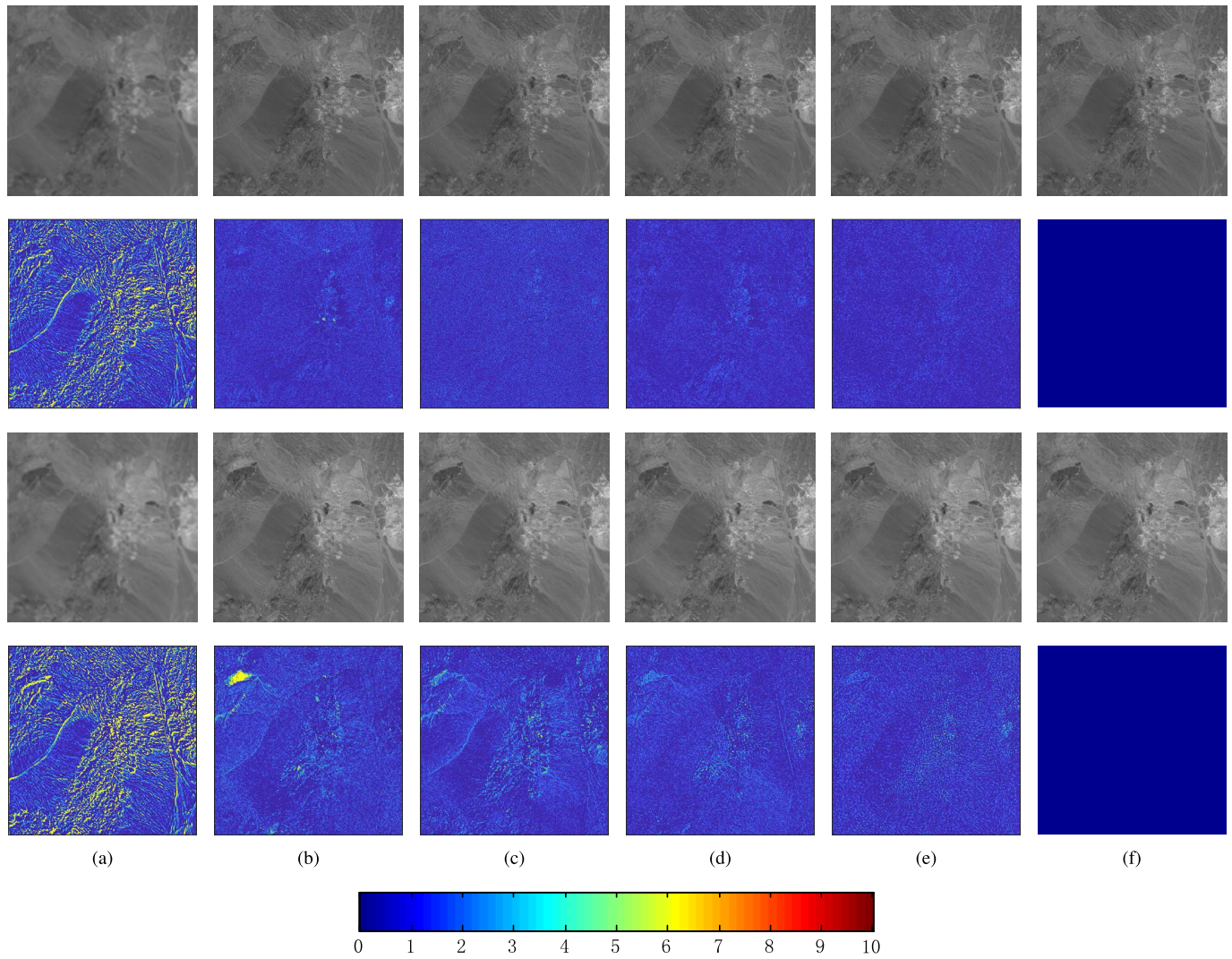(a)   (b)   (c)   (d)   (e)   (f)

Fig. 7.   First and second rows show the fusion results and the corresponding error images by the CSU (44.933 dB), CNMF (43.887 dB), Fuse-S (46.395 dB), and CNN-Fus (46.942 dB) at the 50th band of Cuprite Mine. The third and fourth rows show the fusion results and the corresponding error images by the CSU (40.904 dB), CNMF (42.043 dB), Fuse-S (43.208 dB), and CNN-Fus (43.147 dB) at the 90th band of Cuprite Mine. (a) LR-HSI. (b) CSU [23]. (c) CNMF [22]. (d) Fuse-S [27]. (e) CNN-Fus. (f) Ground truth.

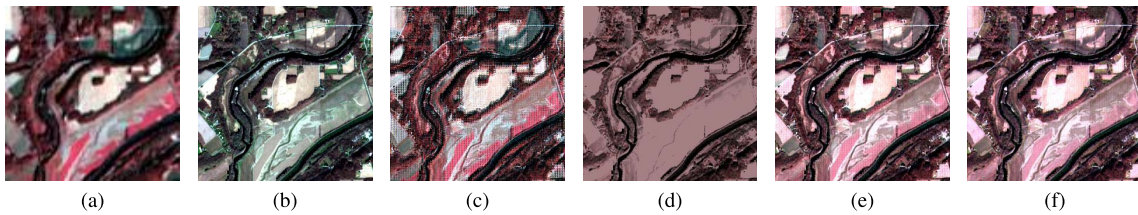(a)   (b)   (c)   (d)   (e)   (f)

Fig. 8.   Hyperion LR-HSI, Sentinel-2A HR-MSI, and false-color images consisting of 16th, 5th, and 2nd bands of the recovered HR-HSIs. (a) Hyperion LR-HSI. (b) Sentinel-2A HR-MSI. (c) NSSR [19]. (d) CSU [23]. (e) CSTF [34]. (f) CNN-Fus.

for experiments. The HR-MSI is taken by the Sentinel-2A satellite. It has 13 spectral bands, and we use the four bands with a 10-m spatial resolution for the fusion. The central wavelengths of the four bands are 490, 560, 665, and 842 nm. The spatial size of the HR-MSI is 300×300. We use the method proposed by Simoẽs *et al.* [25] to estimate $\mathbf{B}$ and $\mathbf{R}$ simultaneously. From the observation models $\mathbf{X}_{(3)} = \mathbf{Z}_{(3)}\mathbf{BS}$ and $\mathbf{Y}_{(3)} = \mathbf{RZ}_{(3)}$, we get the equation $\mathbf{RX}_{(3)} = \mathbf{Y}_{(3)}\mathbf{BS}$.

Therefore, $\mathbf{B}$ and $\mathbf{S}$ are estimated by solving the following equation $\min_{\mathbf{B},\mathbf{S}} ||\mathbf{RX}_{(3)} - \mathbf{Y}_{(3)}\mathbf{BS}||_F^2 + \lambda_b \phi_b(\mathbf{B}) + \lambda_r \phi_r(\mathbf{R})$, where $\lambda_b \phi_b(\mathbf{B})$ and $\lambda_r \phi_r(\mathbf{R})$ are the regularization term on $\mathbf{B}$ and $\mathbf{S}$, respectively. To smooth the blur matrix $\mathbf{B}$, $\phi_b(\mathbf{B})$ is set as $\phi_b(\mathbf{B}) = ||\mathbf{D}_h B||_F^2 + ||\mathbf{D}_v B||_F^2$, where $\mathbf{D}_v$ and $\mathbf{D}_v$ calculate the horizontal and vertical differences of $\mathbf{B}$. Since we only need to smooth the spectral response matrix along the vertical mode (row), $\phi_r(\mathbf{R})$ is set as $\phi_b(\mathbf{B}) = ||\mathbf{D}_v B||_F^2$.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

10

IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS

TABLE III
RUNNING TIME IN SECONDS OF THE TESTING METHODS

Method	Dataset		
	Pavia University	Cuprite Mine	Hyperion
NSSR [19]	217	481	101
CSU [23]	455	505	**10**
CSTF [34]	419	759	219
CNMF [22]	115	**207**	45
Fuse-S [27]	662	770	319
CNN-Fus	**110**	518	42

We first estimate **R** with the strong spatial blur and then estimate **B** with known **R**. Fig. 8 shows the false-color image consisting of 16th, 5th, and 2nd bands of recovered HR-HSIs. The recovered images by CSTF and CNMF are not shown since they perform worse in this data set. As shown in Fig. 8, all testing approaches can obviously improve the spatial resolution of the observable LR-HSI, and the fusion results of CNN-Fus have much fewer flaws. The CSTF applies to the case that the blur is decomposable in two spatial dimensions, and the estimated blur dose not satisfy the case. Hence, the fusion results of CSTF have obvious artifacts.

F. Computational Efficiency

To compare the computational efficiency of the testing methods, the running time of testing approaches on the Pavia University, Cuprite Mine, and Hyperion data sets is reported in Table III. All experiments are conducted at MATLAB R2018b and computer equipped with 8-GB random access memory and Intel Core-i5-9300H CPU with 2.4 GHz. As can be seen from Table III, the CNN-Fus has obvious speed advantage on Pavia University. The speed advantage of our method mainly comes from the subspace representation, which can largely reduce the size of HSI data. Besides, the CNMF is the fastest method on Cuprite Mine.

V. CONCLUSION

We propose a new HSI-MSI fusion method, which is based on the subspace representation and CNN denoiser. First, to exploit the high correlations among the spectral bands, we approximate the desired HR-HSI with the low-dimensional subspace multiplied by the coefficients, which can not only speed up the algorithm but also more accurate recovery. Since the LR-HSI preserves most of the spectral information, the subspace is learned from it via SVD. Due to the powerful learning performance and high speed of CNN, we use the well-trained CNN to regularize the estimation of coefficients. Specifically, we propose to plug the CNN denoiser into the ADMM iteration to estimate the coefficients. Experiments on both simulated and real data fusion demonstrate the superiority of the proposed approach over the existing state-of-the-art HSI-MSI fusion approaches.

The proposed subspace representation and CNN denoiser-based framework can be easily used for other HSI restoration tasks, such as HSI denoising and compressed sensing, and is expected to obtain good performance.

REFERENCES

[1] N. Akhtar and A. Mian, "Nonparametric coupled Bayesian dictionary and classifier learning for hyperspectral classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 9, pp. 4038–4050, Sep. 2018.

[2] 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.

[3] 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.

[4] J. Peng, L. Li, and Y. Y. Tang, "Maximum likelihood estimation-based joint sparse representation for the classification of hyperspectral remote sensing images," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 6, pp. 1790–1802, Jun. 2019, doi: 10.1109/TNNLS.2018.2874432.

[5] L. Zhang, L. Zhang, and B. Du, "Deep learning for remote sensing data: A technical tutorial on the state of the art," *IEEE Geosci. Remote Sens. Mag.*, vol. 4, no. 2, pp. 22–40, Jun. 2016.

[6] L. Zhang, L. Zhang, B. Du, J. You, and D. Tao, "Hyperspectral image unsupervised classification by robust manifold matrix factorization," *Inf. Sci.*, vol. 485, pp. 154–169, Jun. 2019.

[7] X. Kang, X. Zhang, S. Li, K. Li, J. Li, and J. A. Benediktsson, "Hyperspectral anomaly detection with attribute and edge-preserving filters," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 10, pp. 5600–5611, Oct. 2017.

[8] H. Akbari, Y. Kosugi, K. Kojima, and N. Tanaka, "Detection and analysis of the intestinal ischemia using visible and invisible hyperspectral imaging," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 8, pp. 2011–2017, Aug. 2010.

[9] 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.

[10] C. Kwan, J. Choi, S. Chan, J. Zhou, and B. Budavari, "A super-resolution and fusion approach to enhancing hyperspectral images," *Remote Sens.*, vol. 10, no. 9, p. 1416, 2018.

[11] J. Zhou, C. Kwan, and B. Budavari, "Hyperspectral image super-resolution: A hybrid color mapping approach," *J. Appl. Remote Sens.*, vol. 10, no. 3, Sep. 2016, Art. no. 035024.

[12] L. Gomez-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.

[13] 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. (IGARSS)*, Jul. 2017, pp. 6130–6133.

[14] 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.

[15] K. Zhang, W. Zuo, and L. Zhang, "FFDNet: Toward a fast and flexible solution for CNN-based image denoising," *IEEE Trans. Image Process.*, vol. 27, no. 9, pp. 4608–4622, Sep. 2018.

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

[17] R. Kawakami, Y. Matsushita, J. Wright, M. Ben-Ezra, Y.-W. Tai, and K. Ikeuchi, "High-resolution hyperspectral imaging via matrix factorization," in *Proc. CVPR*, Jun. 2011, pp. 2329–2336.

[18] 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.

[19] 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.

[20] X.-H. Han, B. Shi, and Y. Zheng, "Self-similarity constrained sparse representation for hyperspectral image super-resolution," *IEEE Trans. Image Process.*, vol. 27, no. 11, pp. 5625–5637, Nov. 2018.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

DIAN *et al.*: REGULARIZING HSI AND MSI FUSION BY CNN DENOISER

11

[21] H. Irmak, G. B. Akar, and S. E. Yuksel, "A MAP-based approach for hyperspectral imagery super-resolution," *IEEE Trans. Image Process.*, vol. 27, no. 6, pp. 2942–2951, Jun. 2018.

[22] N. Yokoya, T. Yairi, and A. Iwasaki, "Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 2, pp. 528–537, Feb. 2012.

[23] C. Lanaras, E. Baltsavias, and K. Schindler, "Hyperspectral super-resolution by coupled spectral unmixing," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 3586–3594.

[24] M. A. Bendoumi, M. He, and S. Mei, "Hyperspectral image resolution enhancement using high-resolution multispectral image based on spectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 10, pp. 6574–6583, Oct. 2014.

[25] 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.

[26] Q. Wei, N. Dobigeon, and J.-Y. Tourneret, "Fast fusion of multi-band images based on solving a Sylvester equation," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4109–4121, Nov. 2015.

[27] 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.

[28] J. M. P. Nascimento and J. M. B. Dias, "Vertex component analysis: A fast algorithm to unmix hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 898–910, Apr. 2005.

[29] R. C. Hardie, M. T. Eismann, and G. L. Wilson, "MAP estimation for hyperspectral image resolution enhancement using an auxiliary sensor," *IEEE Trans. Image Process.*, vol. 13, no. 9, pp. 1174–1184, Sep. 2004.

[30] Y. Zhou, L. Feng, C. Hou, and S.-Y. Kung, "Hyperspectral and multispectral image fusion based on local low rank and coupled spectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 10, pp. 5997–6009, Oct. 2017.

[31] M. A. Veganzones, M. Simoes, G. Licciardi, N. Yokoya, J. M. Bioucas-Dias, and J. Chanussot, "Hyperspectral super-resolution of locally low rank images from complementary multisource data," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 274–288, Jan. 2016.

[32] 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. (CVPR)*, Jul. 2017, pp. 3862–3871.

[33] L. R. Tucker, "Some mathematical notes on three-mode factor analysis," *Psychometrika*, vol. 31, no. 3, pp. 279–311, Sep. 1966.

[34] S. Li, R. Dian, L. Fang, and J. M. Bioucas-Dias, "Fusing hyperspectral and multispectral images via coupled sparse tensor factorization," *IEEE Trans. Image Process.*, vol. 27, no. 8, pp. 4118–4130, Aug. 2018.

[35] C. I. Kanatsoulis, X. Fu, N. D. Sidiropoulos, and W.-K. Ma, "Hyperspectral super-resolution: A coupled tensor factorization approach," *IEEE Trans. Signal Process.*, vol. 66, no. 24, pp. 6503–6517, Dec. 2018.

[36] Y. Chang, L. Yan, H. Fang, S. Zhong, and Z. Zhang, "Weighted low-rank tensor recovery for hyperspectral image restoration," 2017, *arXiv:1709.00192*. [Online]. Available: http://arxiv.org/abs/1709.00192

[37] R. Dian, S. Li, A. Guo, and L. Fang, "Deep hyperspectral image sharpening," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 11, pp. 5345–5355, Nov. 2018.

[38] J. Yang, Y.-Q. Zhao, and J. Chan, "Hyperspectral and multispectral image fusion via deep two-branches convolutional neural network," *Remote Sens.*, vol. 10, no. 5, p. 800, 2018.

[39] X. Han, J. Yu, J. Luo, and W. Sun, "Hyperspectral and multispectral image fusion using cluster-based multi-branch BP neural networks," *Remote Sens.*, vol. 11, no. 10, p. 1173, 2019.

[40] L. Zhuang and J. M. Bioucas-Dias, "Fast hyperspectral image denoising and inpainting based on low-rank and sparse representations," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 3, pp. 730–742, Mar. 2018.

[41] Q. Wei, J. Bioucas-Dias, N. Dobigeon, J.-Y. Tourneret, M. Chen, and S. Godsill, "Multi-band image fusion based on spectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7236–7249, Dec. 2016.

[42] 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.

[43] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.

[44] S. Gu, Q. Xie, D. Meng, W. Zuo, X. Feng, and L. Zhang, "Weighted nuclear norm minimization and its applications to low level vision," *Int. J. Comput. Vis.*, vol. 121, no. 2, pp. 183–208, Jan. 2017.

[45] 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.

[46] S. V. Venkatakrishnan, C. A. Bouman, and B. Wohlberg, "Plug-and-play priors for model based reconstruction," in *Proc. IEEE Global Conf. Signal Inf. Process.*, Dec. 2013, pp. 945–948.

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

[48] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006.

[49] S. Gu, L. Zhang, W. Zuo, and X. Feng, "Weighted nuclear norm minimization with application to image denoising," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 2862–2869.

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

[51] 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.

[52] F. Dell'Acqua, P. Gamba, A. Ferrari, J. A. Palmason, J. A. Benediktsson, and K. Arnason, "Exploiting spectral and spatial information in hyperspectral urban data with high resolution," *IEEE Geosci. Remote Sens. Lett.*, vol. 1, no. 4, pp. 322–326, Oct. 2004.

[53] R. O. Green *et al.*, "Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (AVIRIS)," *Remote Sens. Environ.*, vol. 65, no. 3, pp. 227–248, Sep. 1998.

[54] R. H. Yuhas, A. F. Goetz, and J. W. Boardman, "Discrimination among semi-arid landscape endmembers using the spectral angle mapper (SAM) algorithm," *JPL Airborne Geosci. Workshop*, vol. 1, 1992, pp. 147–149.

[55] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.*, vol. 9, no. 3, pp. 81–84, Mar. 2002.

[56] 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.

Renwei Dian (Student Member, IEEE) 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.

From November 2017 to November 2018, he was a Visiting Ph.D. Student with the University of Lisbon, Lisbon, Portugal, supported by the China Scholarship Council. His research interests include hyperspectral image super-resolution, image fusion, tensor decomposition, and deep learning. More information can be found in his homepage https://sites.google.com/view/renweidian/.

Shutao Li (Fellow, IEEE) received the B.S., M.S., and Ph.D. degrees from Hunan University, Changsha, China, in 1995, 1997, and 2001, respectively.

In 2001, he joined the College of Electrical and Information Engineering, Hunan University. From May 2001 to October 2001, he was a Research Associate with the Department of Computer Science, The Hong Kong University of Science and Technology, Hong Kong. From November 2002 to November 2003, he was a Post-Doctoral Fellow with the Royal Holloway, University of London, Egham, U.K. From April 2005 to June 2005, he was a Visiting Professor with the Department of Computer Science, The Hong Kong University of Science and Technology. He is currently a Full Professor with the College of Electrical and Information Engineering, Hunan University. He has authored or coauthored over 200 refereed articles. His current research interests include image processing, pattern recognition, and artificial intelligence.

Dr. Li is a member of the Editorial Board of the Information Fusion and the Sensing and Imaging. He received two 2nd-Grade State Scientific and Technological Progress Awards of China in 2004 and 2006. He is an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING and the IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT.

Xudong Kang (Senior Member, IEEE) received the B.Sc. degree from Northeast University, Shenyang, China, in 2007, and the Ph.D. degree from Hunan University, Changsha, China, in 2015.

In 2015, he joined the College of Electrical Engineering, Hunan University. His research interests include hyperspectral feature extraction, image classification, image fusion, and anomaly detection.

Dr. Kang received the National Nature Science Award of China (Second Class and Rank as third) and the Second Prize in the Student Paper Competition at the International Geoscience and Remote Sensing Symposium (IGARSS) 2014. He was also selected as the Best Reviewer for the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS and the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. He was an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING from 2018 to 2019. He has been an Associate Editor of the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS since 2018.