

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,800

Open access books available

122,000

International authors and editors

135M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)



# Hyperspectral Image Super-Resolution Using Optimization and DCNN-Based Methods

*Xian-Hua Han*

## Abstract

Reconstructing a high-resolution (HR) hyperspectral (HS) image from the observed low-resolution (LR) hyperspectral image or a high-resolution multispectral (RGB) image obtained using the exiting imaging cameras is an important research topic for capturing comprehensive scene information in both spatial and spectral domains. The HR-HS hyperspectral image reconstruction mainly consists of two research strategies: optimization-based and the deep convolutional neural network-based learning methods. The optimization-based approaches estimate HR-HS image via minimizing the reconstruction errors of the available low-resolution hyperspectral and high-resolution multispectral images with different constrained prior knowledge such as representation sparsity, spectral physical properties, spatial smoothness, and so on. Recently, deep convolutional neural network (DCNN) has been applied to resolution enhancement of natural images and is proven to achieve promising performance. This chapter provides a comprehensive description of not only the conventional optimization-based methods but also the recently investigated DCNN-based learning methods for HS image super-resolution, which mainly include spectral reconstruction CNN and spatial and spectral fusion CNN. Experiment results on benchmark datasets have been shown for validating effectiveness of HS image super-resolution in both quantitative values and visual effect.

**Keywords:** hyperspectral imaging, image super-resolution, optimization-based approach, deep convolutional neural network (DCNN), spectral reconstruction, spatial and spectral fusion

## 1. Introduction

Hyperspectral (HS) imaging simultaneously obtains a set of images of the same scene on a large number of narrow-band wavelengths which can effectively describe the spectral distribution for every scene point and provide intrinsic and discriminative spectral information of the scene. The acquired dense spectral bands of data are capable to benefit for numerous applications, including object recognition and segmentation [1–9], medical image analysis [10], and remote sensing [11–15], to name a few. Although with the availability of the abundant spectral information with HS imaging, it generally results in much low spatial resolution

compared with ordinary panchromatic and RGB images since photon collection in HS sensors is performed in a much larger spatial region for guaranteeing sufficiently high signal-to-noise ratio. The low spatial resolution in the HS images leads to high spectral mixing of different materials in a scene and greatly affects the performance of scene analysis and understanding. Therefore, the reconstruction of high-resolution hyperspectral (HR-HS) image using image processing and machine learning techniques has attracted a lot of attention.

Especially in remote sensing field, a low-resolution (LR) multispectral or HS image is usually available accompanying with a HR single-channel panchromatic image, and the fusion of these two images is generally known as the pan-sharpening technique [16–39]. Motivated by the fact that human vision is more sensitive to luminance, traditional pan-sharpening technique mainly concentrated the reliable illumination restoration via substituting the calculated component of the LR-HS image with the HR information of panchromatic image via hue saturation exploring and principle component analysis. However, these simple approaches avoidably cause spectral distortion in the resulting image. Recently, the HS image super-resolution actively investigates the optimization methods for minimizing the reconstruction error of the available LR-HS and HR-MS (HR-RGB) images [16–30], which manifested impressive performance. The basic idea of these optimization-based approaches assumes that the spectrum can be represented as matrix decomposition with different constraints such as representation sparsity, spectral physical properties, spatial context similarity, and composited matrixes, which are iteratively optimized for more accurate approximating the observed images. Recently, the matrix factorization and spectral unmixing [40–43]-based HS image super-resolution, which are mainly motivated by the fact the HS observations can be represented by a linear combination of the reflectance function basis (the spectral signatures of the pure materials) and the weight vector denoting the fractions of the pure materials on the spectral response is assumed sparse, have been actively investigated [16, 17, 27, 28]. A coupled nonnegative matrix factorization (CNMF) by Yokoya et al. [19], inspired by the physical property of nonnegative weights for the linear combination, has been proposed to estimate the HR-HS image from a pair of HR-MS and LR-HS images. Although the CNMF approach provided acceptable spectral recovery performance, its solution is usually not unique [44], which cannot always lead to unsatisfied spectral recovery results. Lanaras et al. [10] proposed to integrate coupled spectral unmixing strategy into HS super-resolution and conducted optimization procedure with the proximal alternating linearized minimization method, which requires the good initial points of the two decomposed reflectance signatures and the fraction vectors for providing impressive results. Furthermore, taking consideration of the physical meaning of the spectral linear combination on the reflectance signatures and the implementation effectiveness, most work generally assumes that the number of the pure materials in the observed scene is smaller than the spectral band number, which is not always satisfied in the real application.

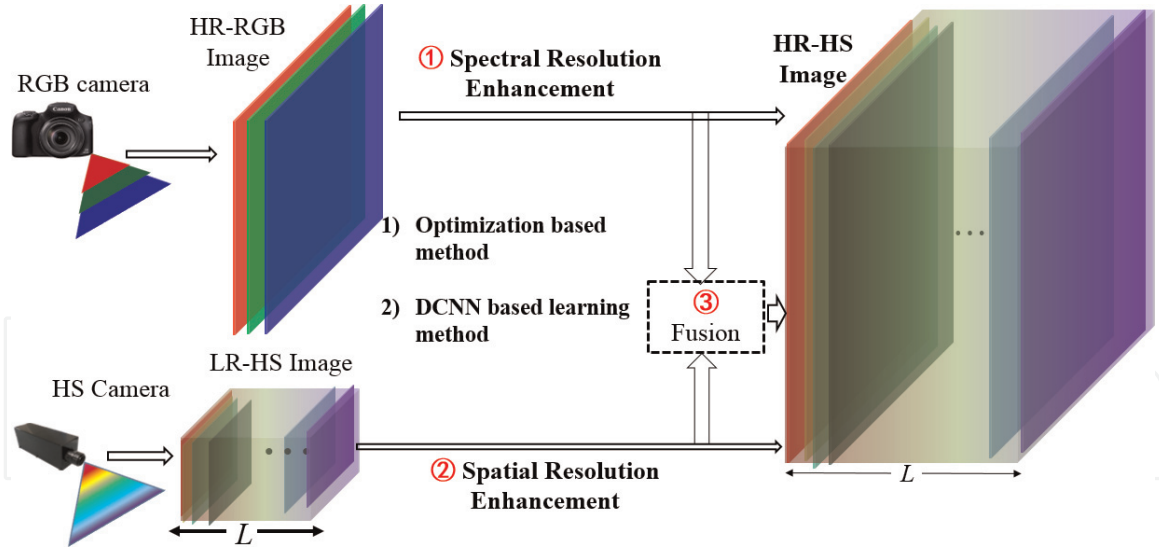
Motivated by the successful applications of the sparse representation on the natural image analysis [14, 15] such as image de-noising, super-resolution, and representation, the sparsity-promoting approaches without considering explicitly the physical meaning constraint on the reflection signature (basis) and thus permitting over-complete basis have widely been applied for HS super-resolution [18, 19]. Inspired by the work in the general RGB image analysis with sparse representation, Grohnfeldt et al. [11] explored a joint sparse representation for HS image super-resolution. Via learning the corresponding HS and MS (RGB) patch dictionaries using the prepared pairs, this work assumed the same sparse coefficients of the corresponding MS and HS patch dictionary, and thus, these can be calculated with only the MS input patch. However the above procedure was

conducted on each individual band, which mainly considered the well reconstruction of the local structure (patch) and completely ignored the spectral correlation between channels. Therefore, several other works [19, 22] investigated the sparse spectral representation via conducting reconstruction of all band spectra instead of the local structure on each individual band. Akhtar et al. [13] explored a sparse spatio-spectral representation via calculating the optimized sparse coefficients of each spectral pixel but assuming the same used atoms for the pixels in a local grid region to integrate the spatial structure. For calculation effectiveness, a generalized simultaneous orthogonal matching pursuit (G-SOMP) was proposed for estimating the sparse coefficients in [22]. Later, the same research group integrated the sparse representation and the Bayesian dictionary learning algorithm for improving the HS image super-resolution performance and manifested its effectiveness. Dong et al. [21] proposed a nonnegative structured sparse representation (NSSR) approach for taking consideration of the spatial structure and then conducted optimization procedure with the alternative direction multiplier method (ADMM) technique. NSSR achieved a large margin on HS image recovery performance compared with the other state-of-the-art approaches. Furthermore, Han et al. [45] proposed to recover the HR-HS output via minimizing the coupled reconstruction error of the available LR-HR and HR-RGB images with the following constraints, (1) the sparse representation with over-complete spectral dictionary in the coupled unmixing strategy [17] and (2) the self-similarity of the sparse spectral representation in the global structures and the local spectra existed in the available HR-RGB image, which further improved the HS image recovery performance in both visual and quality aspects.

Deep convolutional neural networks (CNNs) have recently shown great success in various image processing and computer vision applications. CNN has also been applied to RGB image super-resolution and achieved promising performance. Dong et al. [46] proposed a three-layer CNN architecture (SRCNN), which demonstrates about 0.5–1.5 db improvement and much lower computational cost compared with the popularly used sparse-based methods, and they further extended SRCNN to be capable of directly dealing with the available LR images without mathematical upsampling operation, called as fast SRCNN. Kim et al. [47] exploited a very deep CNN architecture based on VGG-net architecture and concentrated on only estimating the missing high-frequency image (residual image). Ledig et al. integrated two different types of networks, generate network and discriminate network (called as GAN), for estimating much sharper HR image. For applying CNN to HSI SR, Li et al. [48] applied similar structures of SRCNN to super-resolve HSI only from the LR-HS image. These CNN architectures take only the LR image as input, and the expanding factor of resolution enhancement is theoretically limited to be lower than 8 in both height and width. There are also several works exploring CNN-based method with variant backbone architectures to expand the spectral resolution with only HR-RGB image as input [49, 50]. This chapter introduces several research works based on DCNN learning for HS image reconstruction.

On the other hand, regarding to the use of the observed data, the HR-HS image reconstruction can be divided into three research directions: (1) spatial resolution enhancement from hyperspectral imaging, (2) spectral resolution enhancement from RGB imaging, and (3) fusion method based on the observed HR-RGB and low-resolution (LR) HS images of the same scene. Spatial resolution enhancement has popularly been used on single natural image super-resolution [46, 47], and impressive performance has been achieved especially with the deep learning method in the resolution expanding factor from 2 to 4. The deep convolutional neural network (DCNN) has also been adopted for predicting the HR-HS image from a single LR-HS image [48] and validated feasibility of HS image super-resolution for small





**Figure 1.**  
The basic concept of the hyperspectral image super-resolution.

expanding factor. However, the spatial resolution of the available HS image is considerably low compared with the commonly observed RGB image, and then the expanding factor for HR-HS image reconstruction is required to be large enough, for example, more than 10 in horizontal and vertical directions, respectively. Thus, the reconstructed HS image with acceptable quality usually cannot reach the required spatial resolution for different applications. The spectral resolution enhancement for RGB-to-spectrum reconstruction [49, 50] has recently become a hot research line with a single RGB image, which can be lightly collected with a low-price visual sensor. Although the impressive potential of the RGB-spectrum reconstruction is evaluated, there has still large space for performance improving in real applications. Fusing a LR-HS image with the corresponding HR-RGB image to obtain a HR-HS image has shown promising performance [18, 19, 22, 30] compared to spatial and spectral resolution enhancement methods. It is usually solved as an optimization problem with prior knowledge such as sparsity representation and spectral physical properties as constraints, which needs comprehensive analysis of the target scene previously and would be varied scene by scene. Motivated by the amazing performance of the DCNN in natural image super-resolution, Han etc. [51] proposed a spatial and spectral fusion network (SSF-Net) for the HR-HS image reconstruction and validated the better results of the SSF-Net in spite of the simple concatenation of the upsampled LR-HS image and the HR-RGB image. However, the upsampling of the LR-HS image and the simple concatenation cannot effectively integrate the existed spatial structure and spectral property but would lead to computational cost. In addition, precise alignment is needed for the input of LR-HS and HR-RGB images and is extremely difficult due to the large difference of spatial resolution in the LR-HS and HR-RGB images. This chapter introduces several advanced DCNN-based learning methods for hyperspectral image super-resolution and manifests the impressive performance for benchmark datasets. The basic concept of the hyperspectral image super-resolution is shown in **Figure 1**.

## 2. Problem formulation of HS image super-resolution

The goal of HS image super-resolution is to recover a HR-HS image  $\mathbf{Z}' \in \mathbb{R}^{W \times H \times L}$ , where  $L$  denotes the spectral band number and  $W$  and  $H$  denote the image width and height, respectively, from a HR-MS image  $\mathbf{Y}' \in \mathbb{R}^{W \times H \times l}$  ( $l \ll L$ )

and a LR-HS image  $\mathbf{X}' \in \mathbb{R}^{w \times h \times L}$  ( $w \ll W, h \ll H$ ). The common used HR-MS image in the HS image SR scenario is generally a RGB image with  $l = 3$  spectral bands. The matrix forms of  $\mathbf{Z}'$ ,  $\mathbf{X}'$ , and  $\mathbf{Y}'$  are denoted as  $\mathbf{Z} \in \mathbb{R}^{L \times N}$  ( $N = W \times H$ ),  $\mathbf{X} \in \mathbb{R}^{L \times M}$  ( $M = w \times h$ ), and  $\mathbf{Y} \in \mathbb{R}^{3 \times N}$ , respectively. Both  $\mathbf{X}$  (LR-HS) and  $\mathbf{Y}$  (HR-RGB) can be expressed as a linear transformation from  $\mathbf{Z}$  (the desired HS image) as:

$$\mathbf{X} = \mathbf{Z}\mathbf{D}, \mathbf{Y} = \mathbf{R}\mathbf{Z} \quad (1)$$

where  $\mathbf{D} \in \mathbb{R}^{N \times M}$  is the decimation matrix, which blurs and down-samples the HR-HS image to form the LR-HS image, and  $\mathbf{R} \in \mathbb{R}^{3 \times L}$  represents the RGB camera spectral response functions that maps the HR-HS image to the HR-RGB image. With the given  $\mathbf{X}$  and  $\mathbf{Y}$ ,  $\mathbf{Z}$  can be estimated by minimizing the following reconstruction error:

$$\hat{\mathbf{Z}} = \text{arrmin} \|\mathbf{X} - \mathbf{Z}\mathbf{D}\|_F^2 + \|\mathbf{Y} - \mathbf{R}\mathbf{Z}\|_F^2 \quad (2)$$

where  $\|\cdot\|_F$  denotes the Frobenius norm. Via minimizing the reconstruction errors of the observed LR-HSI,  $\mathbf{X}$ , and the HR-RGB image,  $\mathbf{Y}$ , in Eq. (2), we attempt to recover the HR-HSI,  $\mathbf{Z}$ . The intuitive way to solve Eq. (2) is to adopt an optimization-based strategy to minimize Eq. (2) for providing an estimation of the HR-HSI,  $\mathbf{Z}$ . This chapter firstly explores the alternative back-projection (ABP) algorithm to iteratively update the HR-HSI,  $\mathbf{Z}$ , aiming at minimizing Eq. (2). Back-projection [12] is well-known as the efficient iterative procedure to minimize the reconstruction error. Since the back-projection requires an initial estimation for updating the next  $\mathbf{Z}_t$ , we simply upsample the LR-HS image  $\mathbf{X}$  as the initial state,  $\mathbf{Z}_0 = \text{Up}(\mathbf{X})$ . The alternative update for  $\mathbf{Z}_t$  at the  $t$ -th step is formulated as:

$$\begin{aligned} \mathbf{Z}'_t &= \mathbf{Z}_{t-1} + \lambda_1 \mathbf{R}^{-1} * (\mathbf{Y} - \mathbf{R}\mathbf{Z}_{t-1}) \\ \mathbf{Z}_t &= \mathbf{Z}'_t + \lambda_2 \mathbf{D} * (\mathbf{D}^T * \mathbf{D})^{-1} (\mathbf{X} - \mathbf{Z}'_t \mathbf{D}) \end{aligned} \quad (3)$$

where  $(\cdot)^T$  denotes the transpose operation of a matrix and  $(\cdot)^{-1}$  represents the inverse operation of a matrix.  $\lambda_1$  and  $\lambda_2$  denote the hyper-parameters for controlling the updating weights. After the predefined number of alternative iterations, it is prospected to obtain an estimated HR-HSI,  $\mathbf{Z}$ , for well reconstructing the observed LR-HSI,  $\mathbf{X}$ , and HR-RGB image,  $\mathbf{Y}$ .

Since the number of the unknowns ( $N^*L$ ) is much larger than the number of available measurements ( $M^*L + 3^*N$ ), the above optimization problem is highly ill-posed, and proper regularization terms are required to narrow the solution space and ensure stable estimation. A widely adopted constraint is that each pixel spectral  $\mathbf{z}_n \in \mathbb{R}^L$  of  $\mathbf{Z}$  lies in a low-dimensional space, and it can be decomposed as [30]:

$$\mathbf{z}_n = \sum_{k=1}^K \mathbf{b}_k \alpha_{k,n} \text{ subject to: } b_{i,k} \geq 0, \alpha_{k,n} \geq 0, \sum_{k=1}^K \alpha_{k,n} = 1 \quad (4)$$

where  $\mathbf{B} \in \mathbb{R}^{L \times K} = [\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_K]$  is the set of all spectral signatures ( $\mathbf{b}_k$ , also called as the  $k$ -th endmember) of  $K$  distinct materials.  $\alpha_n$  represents the fractional abundance of all  $K$  materials for the  $n$ -th pixel. Taking consideration of the physical property on the spectral reflectance, the elements in the spectral signatures and the fractional abundance are nonnegative as shown in the first and second constraint terms of Eq. (4), and the summation of abundance vector for each pixel is one.

According to  $\mathbf{Y} = \mathbf{RZ}$ , each pixel  $\mathbf{y}_n \in \mathbb{R}^3$  in the HR-RGB image can be decomposed as:

$$\mathbf{y}_n = \mathbf{Rz}_n = \mathbf{RB}\boldsymbol{\alpha}_n = \hat{\mathbf{B}}\boldsymbol{\alpha}_n \quad (5)$$

where  $\hat{\mathbf{B}}$  denotes the RGB spectral dictionary obtained via transforming the HS dictionary  $\mathbf{B}$  with camera spectral function  $\mathbf{R}$ . With a corresponding set of the previously learned spectral dictionaries,  $\hat{\mathbf{B}}$  and  $\mathbf{B}$ , the sparse fractional vector  $\boldsymbol{\alpha}_n$  is able to be estimated from the HR-RGB pixel  $\mathbf{y}_n$  only.

The matrix representation forms of Eqs. (4) and (5) can be formulated as:

$$\mathbf{Z} = \mathbf{BA}, \mathbf{Y} = \hat{\mathbf{B}}\mathbf{A} \quad (6)$$

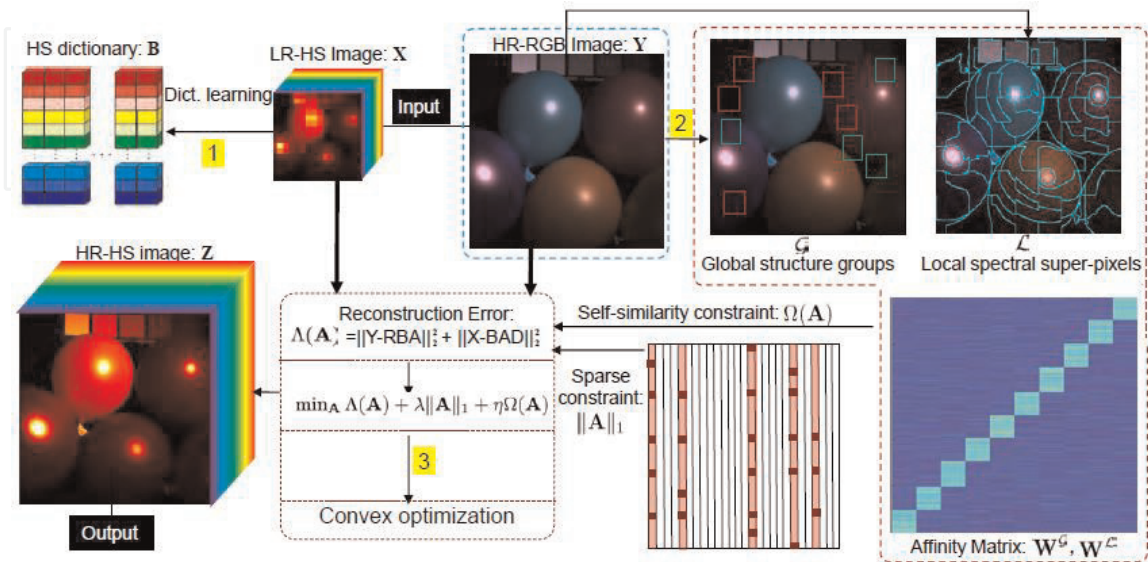
where  $\mathbf{A} = [\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \dots, \boldsymbol{\alpha}_N] \in \mathbb{R}_+^{K \times N}$  is a nonnegative sparse coefficient matrix. Substituting Eq. (4) into Eq. (2), we obtain the nonnegative constraints on both  $\mathbf{B}$  and  $\hat{\mathbf{B}}\mathbf{A}$ , which are applied in the same manner as in Eq. (2). Unless otherwise noted, the nonnegative constraint is imposed on both dictionary and sparse matrix in the following deductions:

$$\{\mathbf{B}^*, \mathbf{A}^*\} = \underset{\mathbf{B}, \mathbf{A}}{\operatorname{argmin}} \|\mathbf{X} - \mathbf{B}\mathbf{A}\|_F^2 + \|\mathbf{Y} - \hat{\mathbf{B}}\mathbf{A}\|_F^2 \quad (7)$$

The goal of Eq. (7) is to solve both spectral dictionary  $\mathbf{B}$  and coefficient matrix  $\mathbf{A}$  with proper regularization terms to achieve stable and accurate solution.

### 3. Self-similarity constrained sparse representation for HS image super-resolution

The complete pipeline of self-constrained sparse representation for HS image super-resolution is illustrated in **Figure 2**. The main contribution of this method is to propose a nonnegative sparse representation coupled with self-similarity



**Figure 2.** Schematics of self-similarity constrained sparse representation for HS image super-resolution: (1) learn the HS dictionary  $\mathbf{B}$  from the input LR-HS image  $\mathbf{X}$ , (2) explore self-similarity of the global-structure and local-spectral, (3) convex optimization of the objective function with sparse and self-similarity constraints on the sparse matrix  $\mathbf{A}$  for estimating the required HR-HS image.

constraint to regularize the solution of Eq. (7). Denoting  $\Lambda(\mathbf{B}, \mathbf{A}) = \|\mathbf{X} - \mathbf{B}\mathbf{A}\|_F^2 + \|\mathbf{Y} - \hat{\mathbf{B}}\mathbf{A}\|_F^2$ , two additional terms are added to Eq. (7) as:

$$\{\mathbf{B}^*, \mathbf{A}^*\} = \text{arrmin}_{\mathbf{B}, \mathbf{A}} \Lambda(\mathbf{B}, \mathbf{A}) + \lambda \|\mathbf{A}\|_1 + \eta \Omega(\mathbf{A}) \quad (8)$$

where  $\|\mathbf{A}\|_1$  denotes the sparse constrained term on the coefficient matrix and  $\Omega(\mathbf{A})$  represents the self-similarity regularized term.  $\lambda$  and  $\eta$  are the hyper-parameters, for controlling the contribution of the two constrained terms. Our study solves Eq. (8) with the following three steps: (1) online learning the HS dictionary from the input LR-HS image, (2) exploring the self-similarity properties of the global-structure and local-spectral self-similarity from the input HR-RGB image, and (3) conducting the convex optimization with the previously learned HS dictionary and the extracted self-similarity for estimating the HR-HS image. Next, we will describe the details of the above procedures in the following three subsections.

### 3.1 Online HS dictionary learning

Since different materials would have very large variety of the HS reflectance, learning a common HS dictionary for various scenes with different materials would lead to considerable spectral distortion. In order to obtain a set of adaptive HS dictionary for well reconstructing the pixel spectra, this study conducts the learning procedure directly using the observed LR-HS image  $\mathbf{X}$  in an online manner. The objective function to build the HS dictionary for representing the pixel spectra is formulated as follows:

$$\{\mathbf{B}^*, \hat{\mathbf{A}}^*\} = \text{argmin}_{\mathbf{B}, \mathbf{A}} \|\mathbf{X} - \mathbf{B}\hat{\mathbf{A}}\|_F^2 + \lambda \|\hat{\mathbf{A}}\|_1 \quad (9)$$

where  $\hat{\mathbf{A}}$  is the sparse matrix for the pixels in the LR-HS image. In our study, we also impose the nonnegative constraints on both sparse matrix  $\hat{\mathbf{A}}$  and spectral dictionary  $\mathbf{B}$ , and thus, the existing dictionary learning method such as K-SVD cannot be applied for our optimization problem. We follow the optimization algorithm [21] and adopt ADMM technique to transform the constrained dictionary learning problem into an unconstrained version. The unconstrained dictionary learning problem is then solved with alternative optimization algorithm. After obtaining the HS dictionary  $\mathbf{B}^*$  via optimizing Eq. (9) with the observed LR-HS image, we would only optimize  $\mathbf{A}$  to solve Eq. (8) via fixing  $\mathbf{B}^*$ .

### 3.2 Extraction of self-similarity constraint

The regularization term  $\Omega(\mathbf{A})$  in Eq. (8) is formulated with two types of self-similarities, which are extracted from the HR-RGB image (see **Figure 2** for illustration):

- **Global-structure self-similarity:** Since pixels with similar spatial structure, which are represented as the concatenated RGB spectra within a local square windows, share similar hyperspectral information, thus the sparse vectors for reconstructing the hyper-spectra of these pixels would also be similar; this applies for both nearby patches and nonlocal patches in the whole image plane, and we name these as global-structure self-similarity.



- **Local-spectral self-similarity:** Since pixels in a local region have the same material with RGB values in the HR-RGB image, the sparse vector for different HR pixels is similar in a local region (superpixel). Note the superpixel is usually not a square patch.

The global-structure self-similarity is represented by global-structure groups  $\mathcal{G} = \{\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_P\}$  (in total  $P$  groups), which are obtained by clustering all similar patches (spatial structure) in the HR-RGB image with  $K$ -means;  $\mathbf{g}_p$  (each  $\mathbf{g}_p$  may have different length) is a vector consisting of the pixel indices in the  $p$ -th group. The local-spectral self-similarity is formulated as the superpixels  $\mathcal{L} = \{\mathbf{l}_1, \mathbf{l}_2, \dots, \mathbf{l}_Q\}$  (in total  $Q$  superpixels) obtained via SLIC superpixel segmentation method;  $\mathbf{l}_q$  is also a vector composed with the pixel indices in the  $q$ -th superpixel. Since the pixels in the same global-structure group have similar spectral-spatial structure, we calculate the sparse vector of any pixel in a given group by a weighted average of the sparse matrix for all pixels in this group. Similarly, the sparse vector of a pixel can also be approximated by a weighted average of the sparse matrix for all pixels in the same local-spectral superpixel. With both self-similarity constraints, the sparse vector for the  $n$ -th pixel can be formulated as:

$$\alpha_n = \gamma \sum_{i \in \mathbf{g}_p} w_{n,i}^{\mathcal{G}} \alpha_i + (1 - \gamma) \sum_{j \in \mathbf{l}_q} w_{n,j}^{\mathcal{L}} \alpha_j \quad (10)$$

with  $n \in \mathbf{g}_p \wedge n \in \mathbf{l}_q$

where  $w_{n,i}^{\mathcal{G}}$  is the global-structure weight for the  $n$ -th sparse vector  $\alpha_n$ ; it adjusts and merges the contribution of the  $i$ -th sparse vector  $\alpha_i$  belonging to the same global-structure group. Analogously,  $w_{n,j}^{\mathcal{L}}$  weights the  $j$ -th sparse vector  $\alpha_j$  belonging to the same local-spectral superpixel. And  $\gamma$  is a parameter for balancing the contribution between the global-structure and local-spectral self-similarity.

To be more specific,  $w_{n,i}^{\mathcal{G}}$  ( $0 < w_{n,i}^{\mathcal{G}} < 1$  and  $\sum_i w_{n,i}^{\mathcal{G}} = 1$ ) measures the similarity between the RGB intensities of patches  $\mathbf{p}_n$  and  $\mathbf{p}_i$  centered around the  $n$ -th and  $i$ -th pixels. Each patch is a set of pixels in a  $R \times R$  window, so each  $\mathbf{p}$  is a  $3R^2$ -dimensional ( $R \times R \times \text{RGB}$ ) vector. It is a decreasing function of the Euclidean distance between the spatial RGB values as:

$$w_{n,i}^{\mathcal{G}} = \begin{cases} \frac{1}{z_n^{\mathcal{G}}} \exp \left( -\frac{\|\mathbf{p}_i - \mathbf{p}_n\|^2}{h^{\mathcal{G}}} \right), & (n, i) \in \mathbf{g}_p, \forall p \\ 0, & \text{others} \end{cases} \quad (11)$$

where  $z_n^{\mathcal{G}}$  is a normalization factor defined as  $z_n^{\mathcal{G}} = \sum_{i \in \mathbf{g}_p} \exp \left( -\frac{\|\mathbf{p}_i - \mathbf{p}_n\|^2}{h^{\mathcal{G}}} \right)$  to guarantee and ensure that  $\sum_{i \in \mathbf{g}_p} w_{n,i}^{\mathcal{G}} = 1$  and  $h^{\mathcal{G}}$  are a smoothing kernel for  $3R^2$ -dimensional vectors. The local-spectral weight  $w_{n,j}^{\mathcal{L}}$  is defined in the exactly same format but with  $\mathbf{p}_n$  and  $\mathbf{p}_i$  being the RGB values of the  $n$ -th and  $i$ -th pixels (so each  $\mathbf{p}$  is a three-dimensional vector here) and a smoothing kernel  $h^{\mathcal{L}}$  for three-dimensional vectors.

We then build affinity matrices  $\mathbf{W}^{\mathcal{G}} \in \mathbb{R}^{N \times N}$  and  $\mathbf{W}^{\mathcal{L}} \in \mathbb{R}^{N \times N}$ , whose element encodes the pairwise similarity calculated using Eq. (11). Finally, the regularization term constrained by two types of self-similarities is represented as:

$$\Omega(\mathbf{A}) = \|\mathbf{A} - \gamma \mathbf{W}^{\mathcal{G}} \mathbf{A} - (1 - \gamma) \mathbf{W}^{\mathcal{L}} \mathbf{A}\|_F^2 \quad (12)$$

With the self-similarity constraints of the global-structure and local-spectral, the sparse representation will be more robust and prospected to be similar for the locations in the same clustered global group and local superpixel. Given the HS dictionary  $\mathbf{B}^*$  pre-learned using Eq. (9) and the regularization term with self-similarity in Eq. (12), Eq. (8), is convex and can be efficiently solved by optimization algorithm. We apply the ADMM technique to solve Eq. (8), and please refer to [45] for detail optimization procedure.

### 3.3 Experimental results

We evaluate the self-similarity constrained sparse representation method using two publicly released hyperspectral imaging databases: the CAVE and Harvard datasets. The CAVE dataset includes 32 indoor images consisting of paintings, toys, food, and so on, which are captured under controlled illumination. The Harvard dataset has 50 indoor and outdoor images captured under daylight illumination. The image size in the CAVE dataset is  $512 \times 512$  pixels, and 31 spectral bands of 10 nm wide, which covers the visible spectrum from 400 to 700 nm. The image size in the Harvard dataset is  $1392 \times 1040$  pixels, and 31 spectral bands of width 10 nm, basically covering the visible spectrum from 420 to 720 nm. In our experiments, we extract the top left  $1024 \times 1024$  pixels as the understudying HR images. We take the original images in the datasets as ground-truth  $\mathbf{Z}$  and resize them by a factor of 32 to create  $16 \times 16$  images in the CAVE dataset and  $32 \times 32$  images in the Harvard dataset, which is implemented by averaging over  $32 \times 32$  pixel blocks as done in [10, 21]. The observed HR-RGB images  $\mathbf{Y}$  are generated by multiplying the spectral channels of the ground-truth image with the spectral response  $\mathbf{R}$  of a Nikon D700 camera. We evaluate the recovery performance of the estimated HS images using four quantitative metrics including root-mean-square error (RMSE), peak-signal-to-noise ratio (PSNR), spectral angle mapper (SAM) [9], and relative dimensionless global error in synthesis (ERGAS) [34]. The quantitative metric, SAM [9], gives the spectral distortion degree of the pixel spectrum in the estimated HR-HS image with the corresponding one in the ground-truth HR-HS image. We calculate the overall SAM metric of one understudying by averaging the SAMs computed from all pixels. The value of SAM is expressed in degrees and thus normalized into the range  $(-90, 90)$ . The smaller the absolute value of SAM, the less the spectral distortion is. The ERGAS [34] calculates the average amount of the relative difference error, where the absolute difference error is normalized by intensity mean in each band. The smaller the ERGAS, the smaller the relative difference error is.

#### 3.3.1 Compare results with the state-of-the-art methods

Firstly, we manifest the compared recovery performance of the HR-HS images with our proposed method (including the online dictionary learning procedure and self-similarity constraints) and the state-of-the-art HS image SR methods including matrix factorization (MF) method [18], coupled nonnegative matrix factorization method [19], sparse nonnegative matrix factorization (SNNMF) method [20], generalization of simultaneous orthogonal matching pursuit method [13], Bayesian sparse representation (BSR) method [9], couple spectral unmixing (CSU) method [10], and nonnegative structured sparse representation method [21]. **Table 1** manifests the average RMSE, PSNR, SAM, and ERGAS results of the 32 images in the CAVE dataset [32], while **Table 2** shows the average results of the 50 images from the Harvard dataset [33].

It can be seen from **Tables 1** and **2** that our approach obtains the best recovery performance for all quantitative metrics, and the performance improvement on the

	MF [18]	CNMF [19]	SNMF [20]	GSOMP [13]	BSR [9]	CSU [10]	NNSR [21]	Our
RMSE	3.03 ± 0.97	2.93 ± 1.30	3.26 ± 1.57	6.47 ± 2.53	3.13 ± 1.57	3.0 ± 1.40	2.21 ± 1.19	<b>2.17 ± 1.08</b>
PSNR	39.37 ± 3.76	39.53 ± 3.55	38.73 ± 3.79	32.48 ± 3.08	39.16 ± 3.91	39.50 ± 3.63	42.26 ± 4.11	<b>42.28 ± 3.86</b>
SAM	6.12 ± 2.17	5.48 ± 1.62	6.50 ± 2.32	14.19 ± 5.42	6.75 ± 2.37	5.8 ± 2.21	4.33 ± 1.37	<b>3.98 ± 1.27</b>
ERGAS	0.40 ± 0.22	0.39 ± 0.21	0.44 ± 0.23	0.77 ± 0.32	0.37 ± 0.22	0.41 ± 0.27	0.30 ± 0.18	<b>0.28 ± 0.18</b>

**Table 1.**  
Quantitative comparison results of the self-similarity constrained sparse representation with the state-of-the-art methods on the CAVE dataset.

	MF [18]	CNMF [19]	SNMF [20]	GSOMP [13]	BSR [9]	CSU [10]	NNSR [21]	Our
RMSE	1.96 ± 0.97	2.08 ± 1.34	2.20 ± 0.94	4.08 ± 3.55	2.10 ± 1.60	1.7 ± 1.24	1.76 ± 0.79	<b>1.64 ± 1.20</b>
PSNR	43.19 ± 3.87	43.00 ± 4.44	42.03 ± 3.61	38.02 ± 5.71	43.11 ± 4.59	43.40 ± 4.10	44.00 ± 3.63	<b>45.20 ± 4.56</b>
SAM	2.93 ± 1.06	2.91 ± 1.18	3.17 ± 1.07	4.99 ± 2.99	2.93 ± 1.33	2.9 ± 1.05	2.64 ± 0.86	<b>2.63 ± 0.97</b>
ERGAS	0.23 ± 0.14	0.23 ± 0.11	0.26 ± 0.27	0.41 ± 0.24	0.24 ± 0.15	0.24 ± 0.20	0.21 ± 0.12	<b>0.16 ± 0.15</b>

**Table 2.**  
Quantitative comparison results of the self-similarity constrained sparse representation with the state-of-the-art methods on the Harvard dataset.

CAVE dataset is more significant than on the Harvard dataset. The NNSR method [21] has the closest performance to ours, and both methods show relatively larger advantage over other methods. In addition, our method shows the best improvement on SAM values over NNSR [21]. This is because for SAM, a slight spectral distortion of the pixels with small magnitudes affects its value greatly. Thus, we can conclude that our proposed approach not only robustly recovers the HS image but also suppresses the noise and artifacts, especially for those pixels with small spectral magnitudes, due to the imposed constraints of the global-structure and local-spectral self-similarities.

3.3.2 Compared results without self-similarity constraints

One of the key differences of our method from existing ones (such as MF [18]) is the two types of imposed self-similarities formulated by the regularized term,  $\Omega(\mathbf{A})$  in Eq. (8). Without the  $\Omega(\mathbf{A})$  term, Eq. (8) can still be solved by an optimization method such as the ADMM. In addition, we can also adopt either global or local self-similarity separately, i.e., by taking only the  $\mathbf{W}^{\mathcal{G}}$  or  $\mathbf{W}^{\mathcal{L}}$  terms in Eq. (12). We conduct such experiments under the same experimental conditions, and the same quantitative metrics as in **Tables 1** and **2** for both datasets are shown in **Table 3**.

	CAVE dataset			Harvard dataset		
	Without both	Local simil. only	Global simil. only	Without both	Local simil. only	Global simil. only
RMSE	2.81 ± 1.42	2.25 ± 1.15	2.32 ± 1.20	1.83 ± 1.30	1.66 ± 1.20	1.88 ± 1.32
PSNR	40.05 ± 3.97	42.00 ± 3.91	41.78 ± 4.05	44.16 ± 4.39	45.01 ± 4.51	44.02 ± 4.56
SAM	5.46 ± 1.89	4.24 ± 1.36	4.59 ± 1.46	2.86 ± 1.06	2.69 ± 1.00	2.99 ± 1.09
ERGAS	0.37 ± 0.20	0.30 ± 0.18	0.31 ± 0.19	0.23 ± 0.16	0.19 ± 0.15	0.18 ± 0.16

**Table 3.**  
Results without local, global, and both similarities on the CAVE and Harvard datasets.

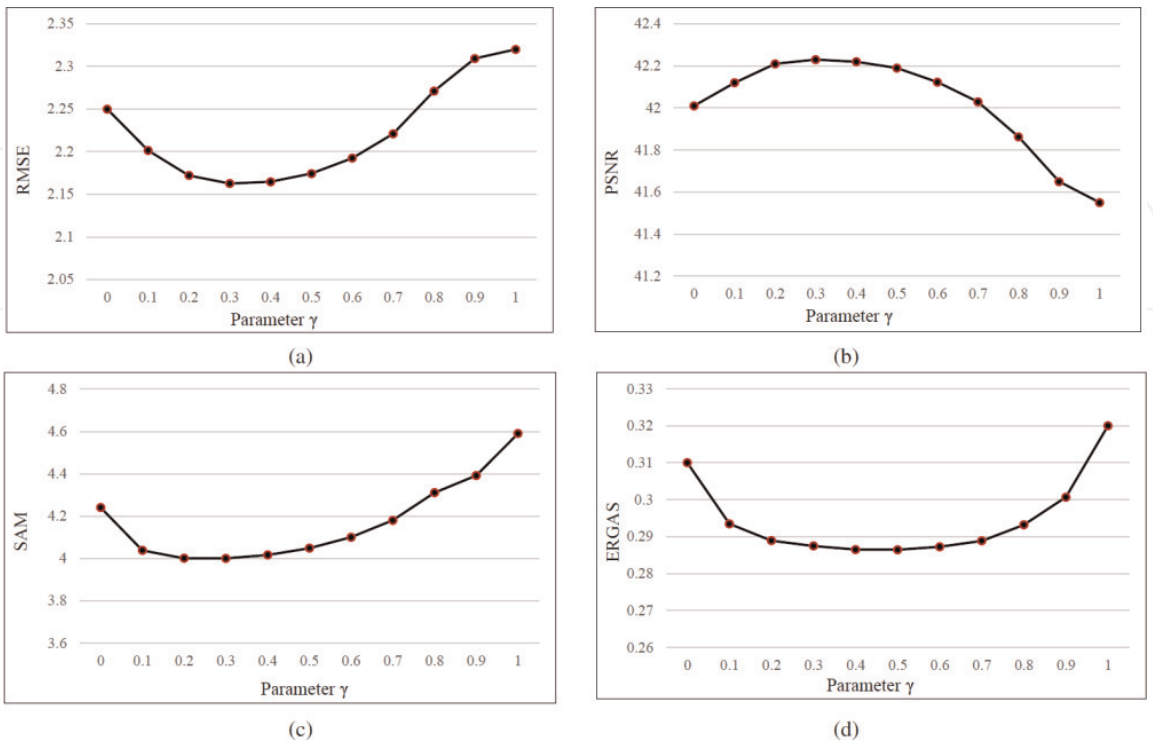
Taking local self-similarity only into consideration significantly improves the results on both datasets for all quantitative metrics which shows relatively larger contribution than considering global self-similarity only, but integrating global self-similarity as our complete approach could further improve the results.

3.3.3 Evaluation results by changing parameter  $\gamma$

In addition, we evaluate the HR-HS image recovery performance via changing the parameter  $\gamma$  for adjusting the contribution of global-structure and local-spectral self-similarity. For CAVE dataset, the parameter  $\gamma$  is changed from 0 (local-spectral self-similarity only) to 1 (global-structure self-similarity only) with interval 0.1, and apply the same measure metrics for manifesting the contribution of the global and local self-similarity. **Figure 3** (a)–(d) gives the curves of the quantitative measures, RMSE, PSNR, SAM, and ERGAS, respectively, which manifests that  $\gamma = 0.3$  gives the best performances. For Harvard dataset, we also conducted experiments with the parameter  $\gamma$ , 0, 0.1, 0.2, ..., and the curves of the quantitative measures, RMSE, PSNR, SAM, and ERGAS, are given in **Figure 4**.

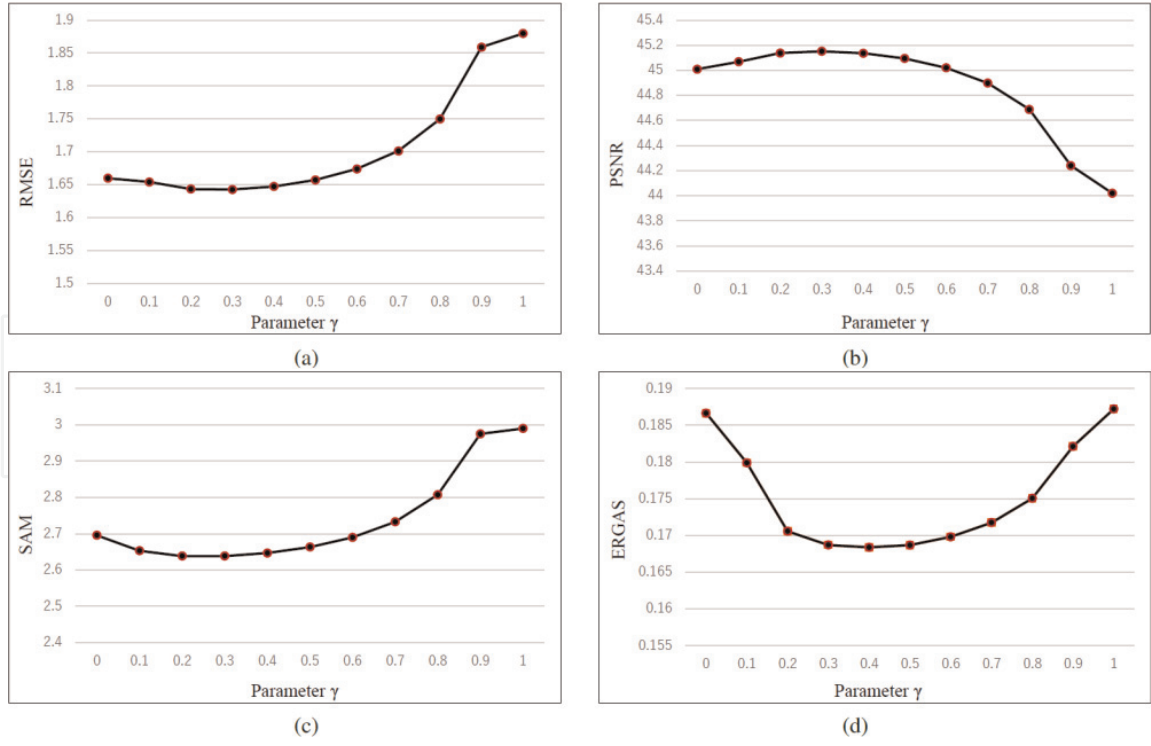
3.3.4 Visual quality comparison

**Figures 5** and **6** manifest the recovered HS images and the difference images with respect to the ground-truth, which includes one example from the CAVE and Harvard dataset, respectively. Since including our method, the CSU [10] and NNSR [21] methods provide the impressive performance compared with all other evaluated methods as shown in **Tables 1** and **2**, we only give the compared results of our method, the CSU [10] and NNSR [21] methods for checking the differences in visual quality. It is obvious that the recovered HS images by our approach have smaller absolute difference magnitude for most pixels than the result by the CSU and NNSR method. It is also worth noting that when self-similarity is not applied,

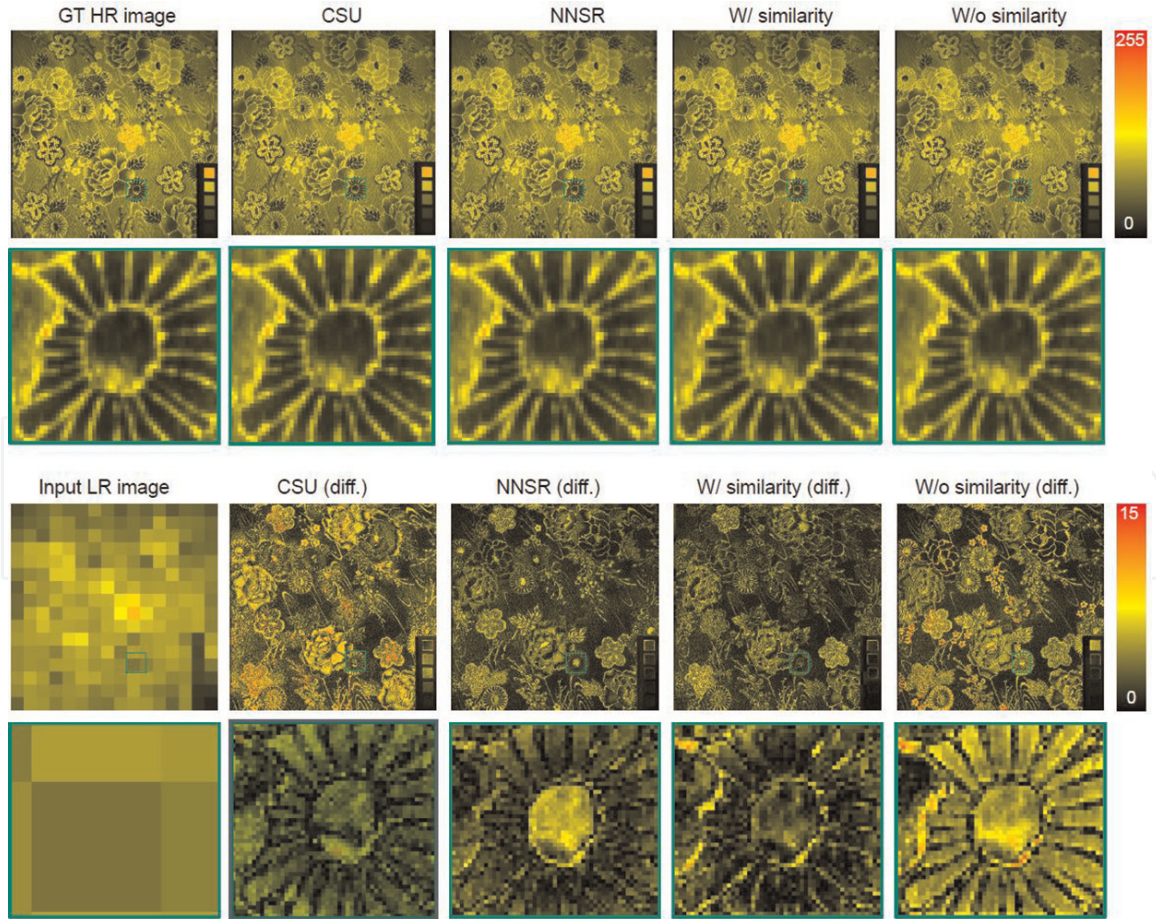


**Figure 3.**  
The evaluated performances with different values of the parameter  $\gamma$  on CAVE dataset. (a) RMSE, (b) PSNR, (c) SAM, and (d) ERGAS.

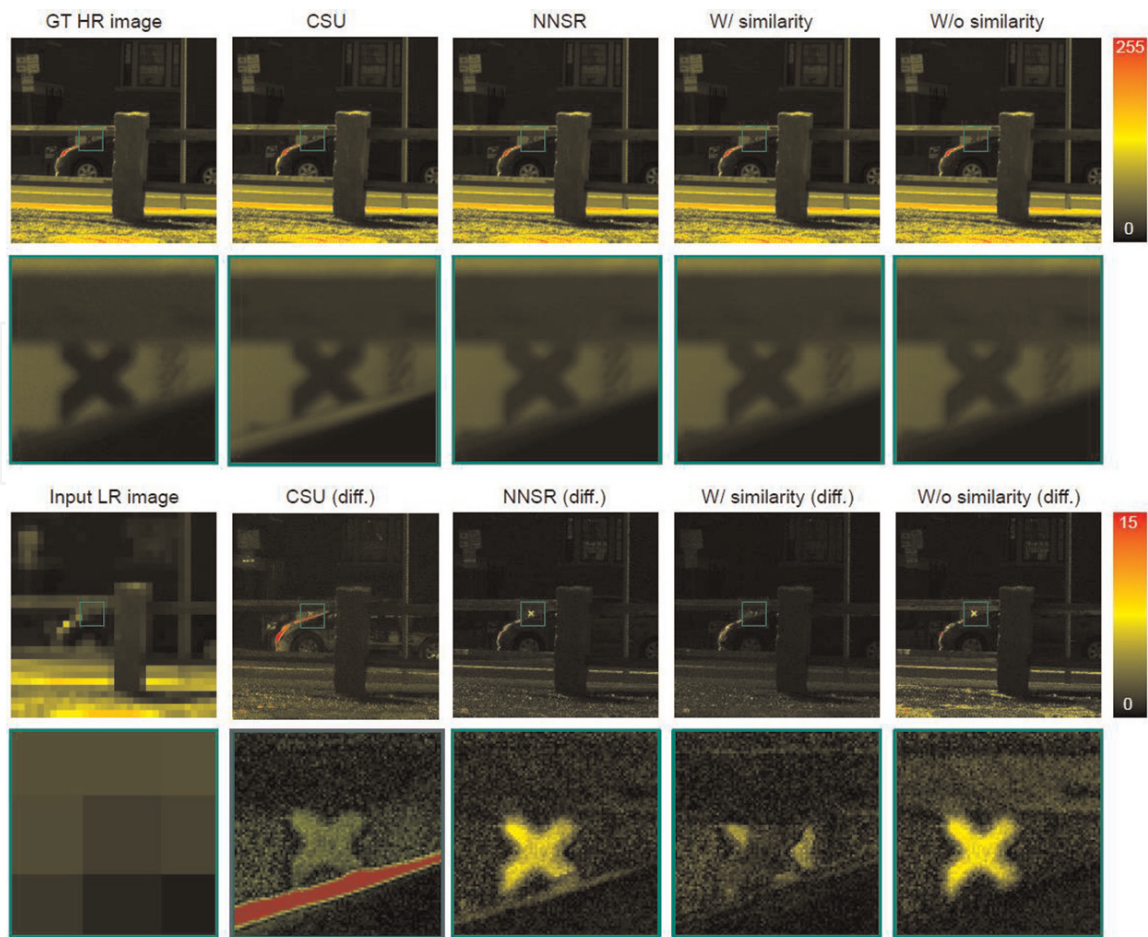




**Figure 4.** The evaluated performances with different values of the parameter  $\gamma$  on Harvard dataset. (a) RMSE, (b) PSNR, (c) SAM, and (d) ERGAS.



**Figure 5.** The visualized results of the recovered HR images from the "cloth" image in the CAVE dataset. The first column shows the ground-truth HR image and the input LR image, respectively. The second to fifth columns show results from CSU [10], NNSR [21], and our method with and without self-similarity, where the upper part provides the recovered images and the lower part gives the absolute difference maps w.r.t. ground-truth. Close-up views are provided below each full resolution image.



**Figure 6.**  
 The visualized results of the recovered HR images from the “imgf1” image in the Harvard dataset. The first column shows the ground-truth HR image and the input LR image, respectively. The second to fifth columns show results from CSU [10], NNSR [21], and our method with and without self-similarity, with the upper part showing the recovered images and the lower part showing the absolute difference maps w.r.t. ground-truth. Close-up views are provided below each full resolution image.

our results manifest quite similar appearance to those from the NNSR method [21], which also reflects the effectiveness of imposing the self-similarity constraint.

#### 4. DCNN-based HS image super-resolution

Motivated by the success for image super-resolution and simply formulation, our previous work explored a simple DCNN-based HS image super-resolution method following the similar CNN structure as in [46], which mainly consists of three convolutional layers and was explained as three operations for the mapping process from LR images to HR images. This explanation follows the schematic concept in sparse coding-based SR: patch extraction, representation learning, nonlinear mapping, and reconstruction. Patch extraction obtains the overlapping patches from the input image and represents each patch as a high-dimensional vector. The convolution layers in CNN are used as feature learning and act as a nonlinear function, which maps a high-dimensional vector (conceptually the patch representation) to another high-dimensional vector (the feature map in the middle-layer of CNN). Reconstruction process combines the mapped CNN features into the final HR image. The above CNN architecture for Y-component recovery of natural image SR adopts the spatial filters in three convolutional layers with sizes  $9 \times 9$ ,  $1 \times 1$ , and  $5 \times 5$ . Since HSI SR attempts to recover high resolution in not only spatial but also spectral domain, which has been proven that the spectral response is more important



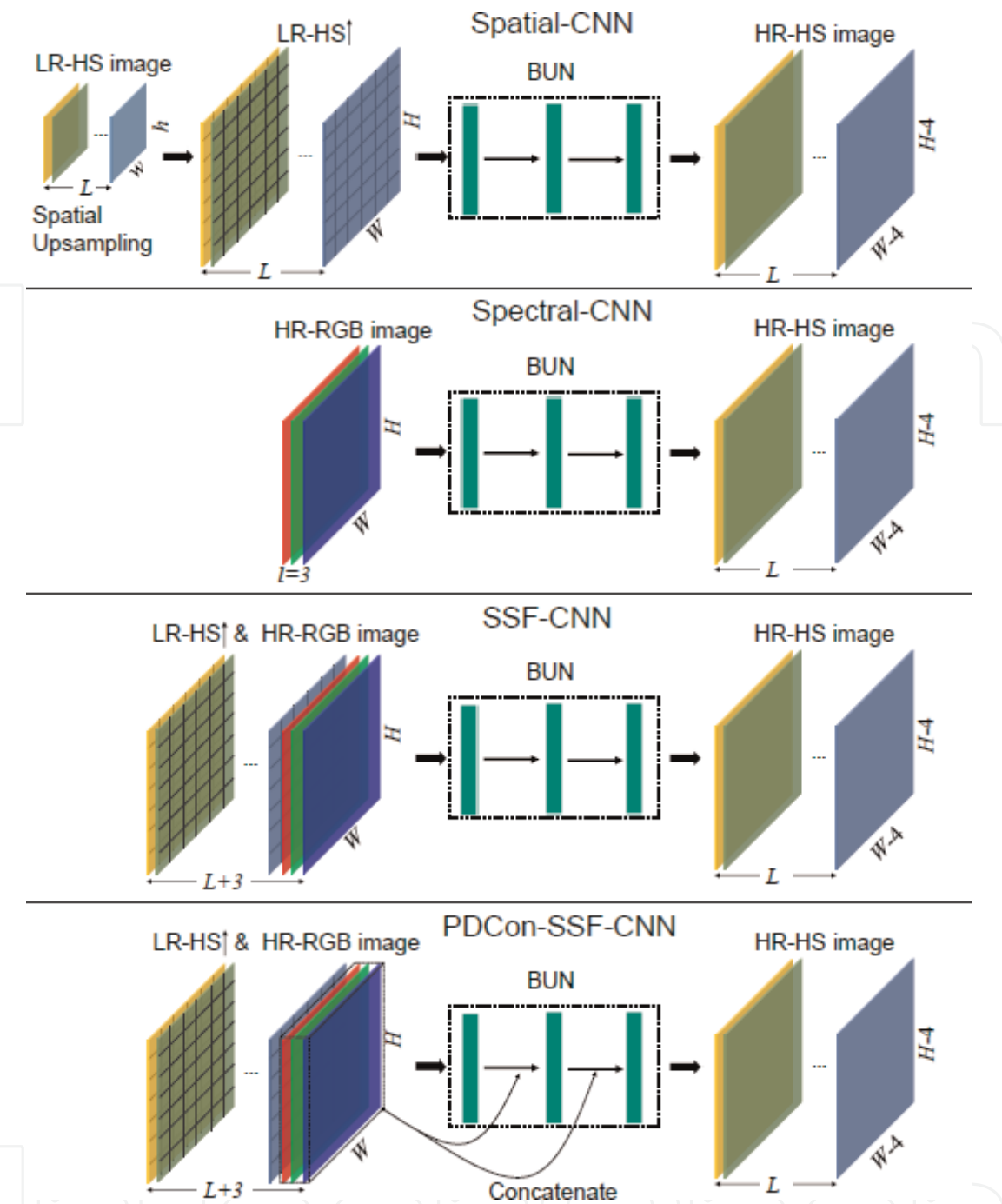
in HIS SR, we set the spatial filter sizes as  $3 \times 3$ ,  $3 \times 3$ , and  $5 \times 5$  with full connection in spectral domain from either one of the available LR-HS and HR-RGB images or the concatenated LR-HS and HR-RGB cubic data.

The intuitive way to apply the above baseline architecture of CNN for HSI SR is to learn the HR-HS image,  $Z$  directly from the available LR-HS image  $X$ , called as spatial CNN. Another research line exploits CNN architecture for learning HSI SR  $Z$  from the available HR-MS (RGB) image  $X$ , named as spectral CNN. However, spatial CNN and spectral CNN take only one domain data of the available LR-HS and HR-MS images,  $X$  or  $Y$  as input, and completely exclude the other domain data. Therefore, this chapter introduces a spatial and spectral fusion architecture of CNN, named as SSF-CNN for recovering the HR-HS image. Recent CNN work incorporates shorter connections between layers for more accurate and efficient training of substantially deeper architectures such as ResNets and Highway Networks, or exploits concatenation between different layer for information and feature reuse such as Densenet, which manifest considerable improvements in different applications. In the scenario of our HSI SR application, since the available HR-RGB image has the same high spatial resolution and the expanding factor (about 10 from 3 to 31) in spectral domain is much smaller than those in spatial domain (32 times from 16/32 to 512/1024 in horizontal and vertical directions, respectively), we concatenate the available HR-RGB image (a part data of the input: Partial) to the outputs of the Conv and RELU blocks (Densely) in the CNN structure for transferring the available maximum spatial information, and name this new CNN architecture as PDCon-SSF. The schematic structures of the spatial CNN, spectral CNN, SSF-CNN and PDCon-SSF are shown in **Figure 7**.

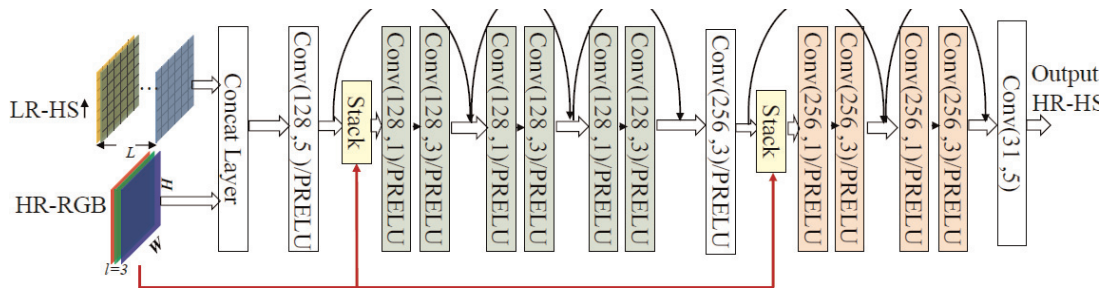
Recently, we also investigated a residual network architecture for HS image super-resolution. The residual network takes the concatenated cubic data of both available HR-RGB and upsampled LR-HS images as input, and simultaneously maintains spectral attribute in LR-HS image and spatial context in HR-RGB image to estimate a more robust HS-HS image. Taking consideration of the characteristic in HS image super-resolution, we modified the ResNet architecture, which is originally proposed for solving higher-level computer vision problems such as image classification and detection, via removing unnecessary modules to simplify the network architecture for this low-level vision problem. Furthermore, as evidenced in pansharpening research that the estimated HR-HS image should have similar spatial structure information with HR-RGB image, we utilize the input RGB image to guide the spatial structure of the learned feature maps in our proposed ResNet. We firstly upsample the LR-HS image to the same size with the HR-RGB image, and stack them together with a “Concat” layer in our method. Multiple residual layer modules with alternately conjuncted spectral and spatial reconstruction layers, which are implemented with convolutional kernel size 1 and  $n$  ( $n > 1$ ), are used for effectively investigating the nonlinear spectral mapping and spatial structure. Our constructed ResNet architecture consists of 5 residual blocks and each block includes a set of the conjuncted spectral and spatial reconstruction layers as shown in **Figure 8**. In **Figure 8**, the first 3 residual blocks have 128 feature maps, and the last 2 residual blocks are with 256 feature maps. The output of the  $m$ -th residual block is expressed as:

$$F_m = \text{Spat}^3(\text{Spec}^1(F_{m-1})) + F_{m-1} \quad (13)$$

where  $\text{Spec}^1(\cdot)$  denotes the spectral reconstruction layer with convolutional kernel size 1, and  $\text{Spat}^3(\cdot)$  denotes the spatial reconstruction layer with convolutional kernel size 3.  $F_{m-1}$  is the input of the residual block. Furthermore, considering the HR spatial structure in the observed HR-RGB image, we use the HR-RGB image to guide the spatial structure of the learned feature maps in the



**Figure 7.** The network architectures of four different types of CNNs. The top row denotes the baseline upsampling network, and the bottom rows are the architectures of spatial CNN, spectral CNN, and the SSF-CNN, respectively.



**Figure 8.** The ResNet architecture for the residual component reconstruction.



residual blocks, which is modeled by stacking the input HR-RGB image and the input feature map  $F_{m-1}$ . Thus, with the added guidance connection, the output of a residual block is modified as:

$$F_m = \text{Spat}^3(\text{Spec}^1(\text{stack}(F_{m-1}, Y))) + F_{m-1} \quad (14)$$

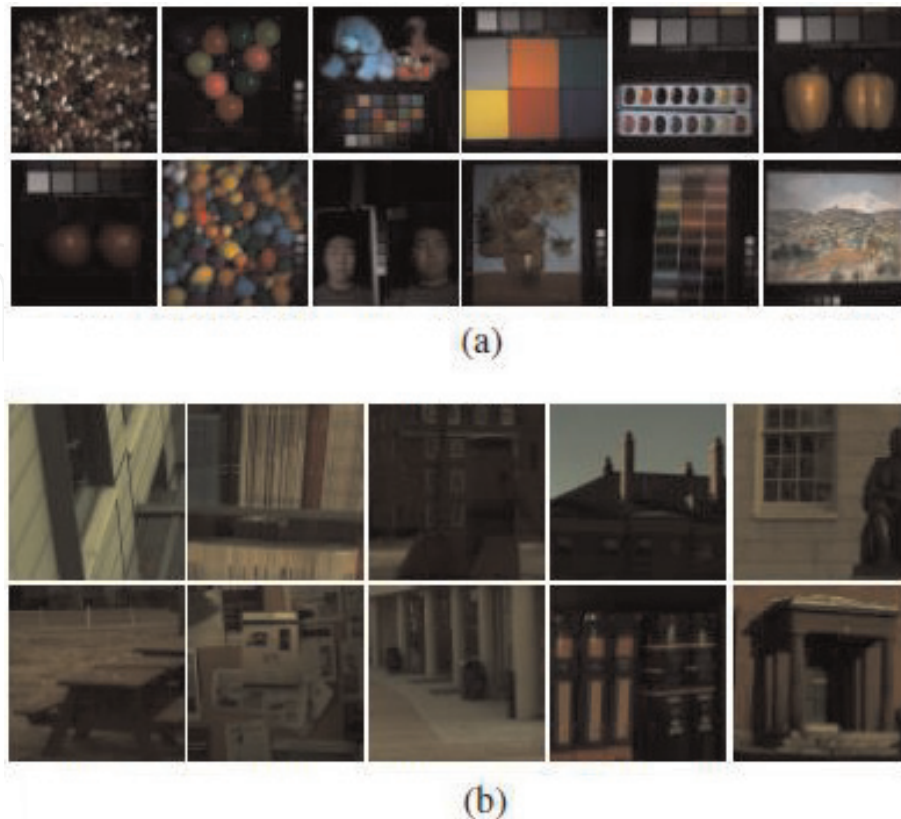
The guidance connections of the HR-RGB image are shown in dot lines in **Figure 7**. Our ResNet-based HR-HS image recovery model is trained by minimizing the Mean Square Error (MSE) between the estimated HR-HS image and the ground-truth  $Z$ .

#### 4.1 Experimental results

We also validate the performance of the HR image reconstruction with the DCNN-based method using CAVE and Harvard datasets. We have randomly selected 20 HSIs from CAVE database to train CNN model, and the remainder is used for validation of the performance of the proposed CNN method. For Harvard database, 10 HSIs have been randomly selected for CNN model training, and the remainder 40 HSIs are as test for validation. **Figure 9** manifests the HR-RGB images of the test samples from CAVE database and several test samples from Harvard databases.

##### 4.1.1 Compared results of different CNN models

As we introduced above, the CNN-based method can be used for recovering the HR-HS image from either of the available LR-HS, HR-RGB images or the concatenated cubic data of the LR-HS, HR-RGB images, which are named as spatial CNN, spectral CNN, Spatial and spectral Fusion CNN (SSF-CNN) and an extended



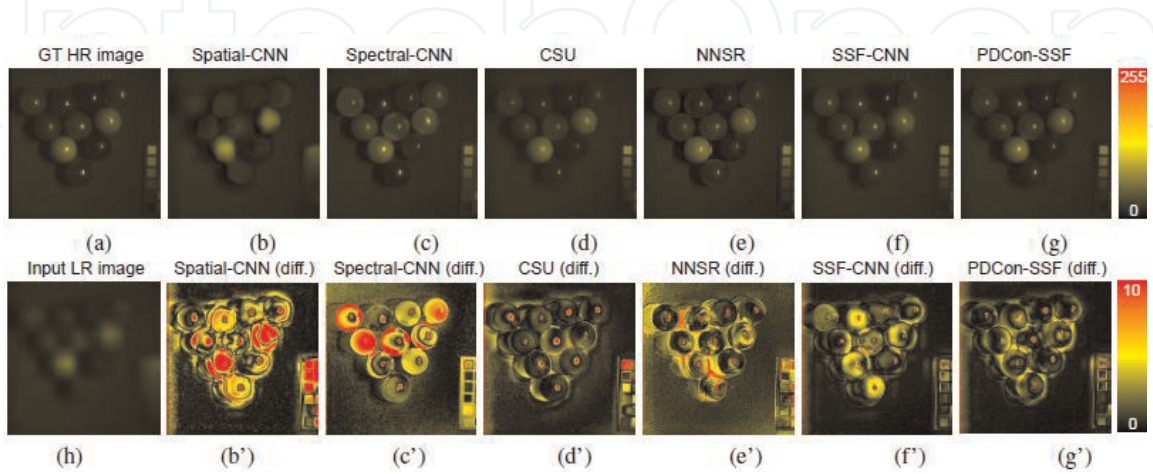
**Figure 9.**  
The HR-RGB images of test samples from CAVE and Harvard databases.

version of SSF-CNN, PDCon-SSF. The baseline network is a three-layer convolution architecture. For CAVE database, we randomly select 20 images for learning the different types of CNN models, and save the CNN model parameters after 0.5 and 1 million iterations. The remainder 12 images in CAVE database are used for evaluating the recovering performance of different CNN models. The average and the standard deviation of RMSE, PSNR, SAM, and ERGAS of the 12 test images in CAVE database are shown in **Table 4**, which manifests much better results of the spectral CNN than spatial CNN due to the smaller expanding factor in spectral domain (about 10 from 3 to 32) than spatial domain (32 from 16 to 512 for horizontal and vertical directions, respectively) and significant performance improvement using SSF-CNN and PDCon-SSF-CNN models. One recovered HS image example and the corresponding residual images with the ground-truth HR images from CAVE database are visualized in **Figure 10** using different CNN models.

From **Table 4** and **Figure 10**, it can be seen that the SSF-based CNN models provide significant performance improvement compared with the spatial CNN and the spectral CNN, and thus for Harvard database, we only train the SSF-CNN and PDCon-SSF models with 1 million iterations using 10 randomly selected 10 images, and the remainder 40 images are used for evaluation. In addition, in order to validate the generation of the learned CNN model, we predict the HR-HS image of the Harvard test samples according to the parameters of the learned SSF-CNN and

CNN Models	Spatial-CNN	Spectral-CNN	SSF-CNN	PDCon-SSF
0.5 million iterations				
RMSE	22.27±8.95	6.29±2.96	2.68±0.87	2.52±0.83
PSNR	21.66±3.58	32.89±3.60	40.09±3.25	40.59±3.24
SAM	19.53±11.01	9.60±3.56	5.04±1.47	4.95±1.54
ERGAS	2.37±0.83	0.60±0.21	0.28±0.10	0.27±0.11
1 million iterations				
RMSE	24.49±9.35	5.73±2.20	2.20±0.73	2.18±0.75
PSNR	20.97±3.53	33.63±3.70	41.85±3.48	41.93±3.57
SAM	17.61±6.30	8.40±2.37	4.39±1.42	4.38±1.39
ERGAS	2.54±0.89	0.57±0.37	0.23±0.11	0.22±0.10

**Table 4.**  
The average and standard deviation of RMSE, PSNR, SAM, and ERGAS using different CNN models of three-layer architecture under 0.5 and 1 million iteration training on CAVE database.



**Figure 10.**  
The “superballs” image example from the CAVE database. The first row shows the ground-truth HR image and the recovered images by spatial CNN, spectral CNN, CSU [22], NNSR [12], and the proposed spatial and spectral CNN architectures, SSF-CNN and PDCon-SSF-CNN, respectively. The second row gives the input LR image, the absolute difference images between the ground-truth image, and the recovered HR-HS images in the first row.

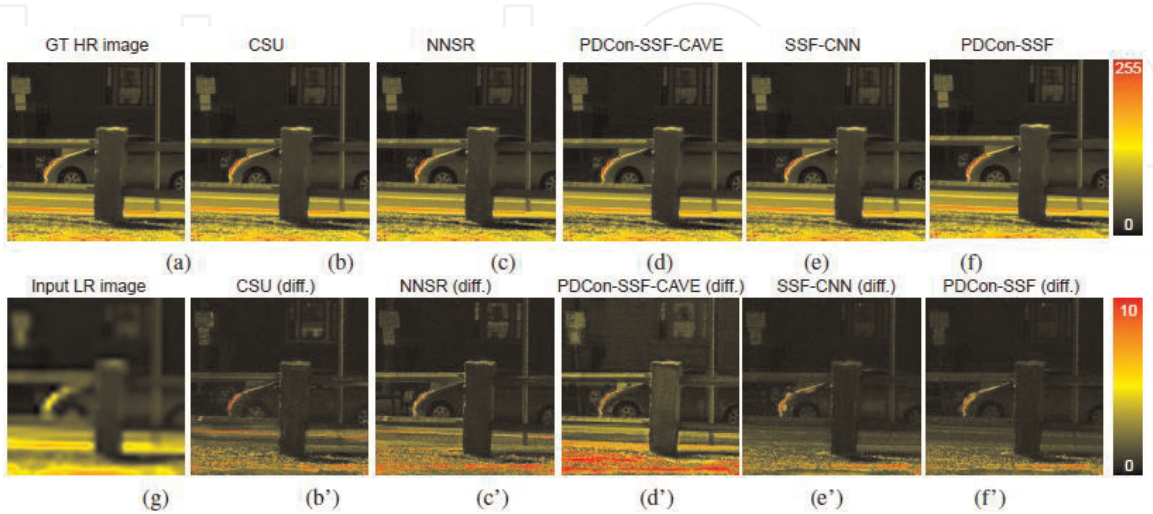
PDCon-SSF-CNN with the CAVE training samples. The average and the standard deviation of RMSE, PSNR, SAM, and ERGAS of the 40 test images in Harvard database are shown in **Table 5**, which shows that the learned SSF-CNN and PDCon-SSF models even with the training samples from CAVE database can provide reasonable recovery performance and the quantitative measures can further be improved using the learned SSF-CNN and PDCon-SSF models even with 10 training images only. One recovered HS image example and its corresponding residual images with the ground-truth HR image from Harvard database are visualized in **Figure 11** using the learned SSF and PDCon-SSF-CNN models with the CAVE and Harvard training samples, respectively.

4.1.2 Compared results of different baseline CNN architectures

As mentioned above, we also investigated a residual network architecture for HS image super-resolution, which has different baseline CNN architecture with the SSF-CNN. Under the same experimental results, we implemented the DCNN-based HS image reconstruction using three-layer CNN and the ResNet architecture with five residual blocks. The compared quantitative results are shown in **Table 6** for both CAVE and Harvard datasets. One recovered HS image example and the corresponding residual images with the ground-truth HR image from CAVE database are visualized in **Figure 12** using the ResNet-RGB, SSF-Net, and the ResNet-based fusion models.

CNN Models	SSF-CNN-CAVE	PDCon-SSF-CAVE	SSF-CNN	PDCon-CNN
RMSE	4.23±2.36	3.97±2.08	1.76±1.00	<b>1.74±0.95</b>
PSNR	36.88±4.81	37.31±5.58	44.31±4.28	<b>44.33±4.18</b>
SAM	5.43±1.28	5.36±1.24	3.05±1.12	<b>3.03±1.15</b>
ERGAS	1.79±0.63	1.37±0.521	0.22±0.18	<b>0.20±0.17</b>

**Table 5.** The average and standard deviation of RMSE, PSNR, SAM, and ERGAS of the test samples of Harvard database using different CNN models, where “SSF-CNN-CAVE” and “PDCon-SSF-CAVE” denote the learned CNN models using the training images from CAVE database.



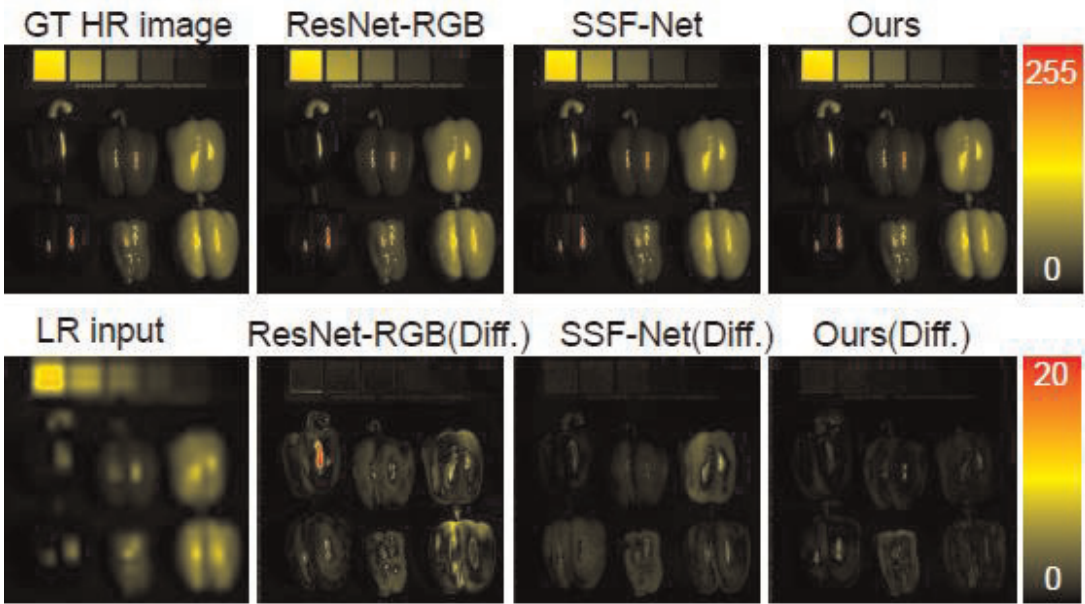
**Figure 11.** An image example from the Harvard database. The first row shows the ground-truth HR image and the recovered images by CSU [22], NNSR [12], and the proposed PDCon-SSF-CNN using CAVE training images, SSF-CNN, and PDCon-SSF-CNN using Harvard training images, respectively. The second row gives the input LR image, the absolute difference images between the ground-truth image, and the recovered HR-HS images in the first row.



(a)			
Inputs	ResNet-RGB	SSF-Net [20]	Ours
RMSE	4.77±2.44	2.20±0.73	<b>1.93±1.13</b>
PSNR	35.51±4.23	41.85±4.48	<b>43.57±4.64</b>
SAM	8.32±2.67	4.39±1.42	<b>3.58±1.40</b>
SSIM	0.977±0.02	0.986±0.01	<b>0.991±0.01</b>

(b)			
Inputs	ResNet-RGB	SSF-Net [20]	Ours
RMSE	3.00±2.26	1.94±1.23	<b>1.83±1.01</b>
PSNR	40.40±5.77	43.56±4.88	<b>44.05±4.85</b>
SAM	3.83±1.54	3.14±0.97	<b>2.37±0.65</b>
SSIM	0.980±0.01	0.988±0.01	<b>0.984±0.01</b>

**Table 6.**  
The compared average and standard deviation of RMSE, PSNR, SAM, and SSIM using the ResNet-RGB, SSF-Net [51], and the ResNet-based fusion methods on both CAVE and Harvard databases.



**Figure 12.**  
The visualized results of the recovered HR images from an example image in CAVE dataset.

5. Conclusions

This chapter introduced recently research on HS image super-resolution. We firstly described the problem formulation for HS image super-resolution and provided the mathematical model between the observed HR-RGB, LR-HS images, and the required HR-HS image. Then we gave the detail description for an optimization-based method: self-similarity constrained sparse representation and the recently proposed DCNN-based method. Experimental results validated that the recently proposed HR image super-resolution methods manifest promising performance on benchmark datasets.



IntechOpen

IntechOpen

### **Author details**

Xian-Hua Han  
Graduate School of Science and Technology for Innovation, Yamaguchi University,  
Yamaguchi, Japan

\*Address all correspondence to: [hanxhua@yamaguchi-u.ac.jp](mailto:hanxhua@yamaguchi-u.ac.jp)

### **IntechOpen**

---

© 2019 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

## References

- [1] Fauvel M, Tarabalka Y, Benediktsson J, Chanussot J, Tilton J. Advances in spectral-spatial classification of hyperspectral images. *Proceedings of the IEEE*. 2013;**101**(3): 652-675
- [2] Uzair M, Mahmood A, Mian A. Hyperspectral face recognition using 3d-dct and partial least squares. *BMVC*. 2013:57.1-57.10
- [3] Zhang D, Zuo W, Yue F. A comparative study of palmprint recognition algorithm. *ACM Computing Surveys*. 2012;**44**(1):2:1-2:37
- [4] Nguyen H, Benerjee A, Chellappa R. Tracking via object reflectance using a hyperspectral video camera. *CVPRW*. 2010:44-51
- [5] Tarabalka Y, Chanussot J, Benediktsson J. Segmentation and classification of hyperspectral images using minimum spanning forest grown from automatically selected markers. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*. 2010;**40**(5): 1267-1279
- [6] Zhou Y, Chang H, Barner K, Spellman P, Parvin B. Classification of histology sections via multispectral convolutional sparse coding. *CVPR*. 2014:3081-3088
- [7] Bioucas-Dias J, Plaza A, Camps-Valls G, Scheunders P, Nasrabadi NM, Chanussot J. Hyperspectral remote sensing data analysis and future challenges. *IEEE Geoscience and Remote Sensing Magazine*. 2013; **1**(2):6-36
- [8] Akhtar N, Shafait F, A M. Sungp: A greedy sparse approximation algorithm for hyperspectral unmixing. *ICPR*. 2014: 3726-3731
- [9] Wei Q, Bioucas-Dias J, Dobigeon N, Tourerret J. Hyperspectral and multispectral image fusion based on a sparse representation. *IEEE Transactions on Geoscience and Remote Sensing*. 2015;**53**(7):3658-3668
- [10] Lanaras C, Baltsavias E, Schindler K. Hyperspectral superresolution by coupled spectral unmixing. *ICCV*. 2015: 3586-3595
- [11] Grohnfeldt C, Zhu XX, Bamler R. Jointly sparse fusion of hyperspectral and multispectral imagery. *IGARSS*. 2013:4090-4093
- [12] Akhtar N, Shafait F, Mian A. Bayesian sparse representation for hyperspectral image super resolution. *CVPR*. 2015:3631-3640
- [13] Akhtar N, Shafait F, Mian A. Sparse spatio-spectral representation for hyperspectral image super-resolution. *ECCV*. 2014:63-78
- [14] Elad M, Aharon M. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image Processing*. 2006;**15**(12): 3736-3745
- [15] Tropp JA, Gilbert AC. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Transactions on Information Theory*. 2007;**53**(12):4655-4666
- [16] Donoho DL, Tsaig Y, Drori I, Starck J-L. Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit. *IEEE Transactions on Information Theory*. 2012;**58**(2):1094-1121
- [17] Wright JA, Yang AY, Ganesh A, Sastry SS, Ma Y. Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2009;**31**(2): 210-227

- [18] Kawakami R, Wright J, Tai Y-W, Matsushita Y, Ben-Ezra M, Ikeuchi K. High-resolution hyperspectral imaging via matrix factorization. *CVPR*. 2011: 2329-2336
- [19] Yokoya N, Yairi T, Iwasaki A. Coupled nonnegative matrix factorization for hyperspectral and multispectral data fusion. *IEEE Transactions on Geoscience and Remote Sensing*. 2012;**50**(2):528-537
- [20] Wcoff E, Chan T, Jia K, Ma W, Ma Y. A non-negative sparse promoting algorithm for high resolution hyperspectral imaging. *ICASSP*. 2013: 1409-1413
- [21] Dong W, Fu F, Shi G, Cao X, Wu J, Li G, et al. Hyperspectral image super-resolution via non-negative structured sparse representation. *IEEE Transactions on Image Processing*. 2016; **25**(3):2337-2352
- [22] Huang B, song H, Cui H, Peng J, Xu Z. Spatial and spectral image fusion using sparse matrix factorization. *IEEE Transactions on Geoscience and Remote Sensing*. 2014;**52**(3):1693-1704
- [23] Chavez P, Sides S, Anderson J. Comparison of three different methods to merge multiresolution and multispectral data: Landsat tm and spot panchromatic. *Photogrammetric Engineering and Remote Sensing*. 1991; **30**(7):1779-1804
- [24] Haydn R, Dalke G, Henkel J, Bare J. Application of the ihs color transform to the processing of multisensor data and image enhancement. *International Symposium on Remote Sensing of Environment*. 1982:559-616
- [25] Aiazzi B, Baronti S, Lotti F, Selva M. A comparison between global and context-adaptive pansharpening of multispectral images. *IEEE Geoscience and Remote Sensing Letters*. 2009;**6**(2): 302-306
- [26] Minghelli-Roman A, Polidori L, Mathieu-Blanc S, Loubersac L, Cauneau F. Spatial resolution improvement by merging meris-etm images for coastal water monitoring. *IEEE Geoscience and Remote Sensing Letters*. 2006;**3**(2):227-231
- [27] Zurita-Milla R, Clevers J, Schaepman ME. Unmixing-based landsat tm and meris fr data fusion. *IEEE Geoscience and Remote Sensing Letters*. 2008;**5**(3):453-457
- [28] Cetin M, Musaoglu N. Merfing hyperspectral and panchromatic image data: Qualitative and quantitative analysis. *International Journal of Remote Sensing*. 2009;**30**(7):1779-1804
- [29] Lee DD, Seung SH. Algorithms for non-negative matrix factorization. *NIPS*. 2001:556-562
- [30] Bioucas-Dias JM, Plaza A, Dobigeon N, Parente M, Du Q, Gader P, et al. Hyperspectral unmixing overview: Geometrical, statistical and sparse regression-based approaches. *IEEE Journal of Selected Topics in Applied Earth Observation and Remote Sensing*. 2012;**5**(2):354-379
- [31] Achanta R, Shaji A, Smith K, Lucchi A, Fua P, Susstrunk S. Slic superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2012;**34**(11):2274-2282
- [32] Yasuma F, Mitsunaga T, Iso D, Nayar S. Generalized assorted pixel camera: Post-capture control of resolution, dynamic range and spectrum. *IEEE Transactions on Image Processing*. 2010;**19**(9):2241-2253
- [33] Chakrabarti A, Zickler T. Statistics of real-world hyperspectral images. *CVPR*. 2011:193-200
- [34] Wald L. Quality of high resolution synthesised images: Is there a simple

criterion? Proceedings of Fusion Earth Data. 2000:99-103

[35] Minghelli-Roman A, Polidori L, Mathieu-Blanc S, Loubersac L, Cauneau F. Spatial resolution improvement by merging meris-etm images for coastal water monitoring. IEEE Geoscience and Remote Sensing Letters. 2006;3(2):227-231

[36] Zurita-Milla R, Clevers J, Schaepman ME. Unmixing-based landsat tm and meris fr data fusion. IEEE Geoscience and Remote Sensing Letters. 2008;5(3):453-457

[37] Duran J, Buades A, Sbert C, Blanchet G. A Survey of Pansharpening Methods with A New Band-Decoupled Variational Model. CoRR, vol. abs/1606.05703; 2016

[38] Kidiyo K, Miloud CE-M, Nasreddine T. Recent trends in satellite image pan-sharpening techniques. In: 1st International Conference on Electrical, Electronic and Computing Engineering. 2014

[39] Cetin M, Musaoglu N. Merging hyperspectral and panchromatic image data: Qualitative and quantitative analysis. International Journal of Remote Sensing. 2009;30(7):1779-1804

[40] Halimi A, Bioucas-Dias J, Dobigeon N, Buller G, McLaughlin S. Fast hyperspectral unmixing in presence of nonlinearity or mismodelling effects. Transactions on Computational Imaging. 2017;3(2):146-159

[41] Sigurdsson J, Ulfarsson M, Sveinsson J, Bioucas-Dias J. Sparse distributed multitemporal hyperspectral unmixing. IEEE Transactions on Geoscience and Remote Sensing. 2017; 55(11):6069-6084

[42] Wei Q, Bioucas-Dias J, Dobigeon N, Tourneret J-Y, Chen M, Godsill SS. Multi-band image fusion based on spectral unmixing. IEEE Transactions on Image Processing. 2016;54(12):7236-7249

[43] Fu X, Ma W-K, Bioucas-Dias J, Chan T-H. Semiblind hyperspectral unmixing in the presence of spectral library mismatches. IEEE Transactions on Geoscience and Remote Sensing. 2016;54(9):5171-5184

[44] Lee DD, Seung SH. Algorithms for non-negative matrix factorization. NIPS. 2001:556-562

[45] Han X-H, Shi B, Zheng Y. Self-similarity constrained sparse representation for hyperspectral image superresolution. IEEE Transactions on Image Processing. 2018;27(11):5625-5637

[46] Dong C, Loy CC, He KM, Tang XO. Image super-resolution using deep convolutional networks. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI). 2015; 38(2):295-307

[47] Kim J, Lee JK, Lee KM. Accurate image super-resolution using very deep convolutional networks. IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016: 1646-1654

[48] Li YS, Hua J, Zhao X, Xie WY, Li JJ. Hyperspectral image super-resolution using deep convolutional neural network. Neurocomputing. 2017;266: 29-41

[49] Alvarez-Gila A, van de Weijer J, Garrote E. Adversarial networks for spatial contextaware spectral image reconstruction from rgb. In: IEEE International Conference on Computer Vision Workshop (ICCVW 2017). 2017

[50] Galliani S, Lanaras C, Marmanis D, Baltsavias E, and Schindler K. Learned Spectral Super-Resolution," arXiv preprint arXiv:1703.09470; 2017

[51] Han X-H, Shi B, Zheng Y. SSF-CNN: Spatial and spectral fusion with cnn for hyperspectral image super-resolution. ICIP. 2018:2506-2510