

Research Article

Superpixel Spectral Unmixing for Hyperspectral Image Superresolution Using a Coupled Encoder-Decoder Network

Shao-lei Zhang , Guang-yuan Fu , Hong-qiao Wang, and Yu-qing Zhao 

Xi'an Research Institute of Hi-Tech, Hongqing Town, Xi'an, China

Correspondence should be addressed to Shao-lei Zhang; jianjianyuanqu11@163.com

Received 7 June 2020; Revised 7 September 2020; Accepted 24 September 2020; Published 6 October 2020

Academic Editor: Carlos Michel

Copyright © 2020 Shao-lei Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In this paper, we propose a novel hyperspectral image superresolution method based on superpixel spectral unmixing using a coupled encoder-decoder network. The hyperspectral image and multispectral images are fused to generate high-resolution hyperspectral images through the spectral unmixing framework with low-rank constraint. Specifically, the endmember and abundance information is extracted via a coupled encoder-decoder network integrating the priori for unmixing. The coupled network consists of two encoders and one shared decoder, where spectral information is preserved through the encoder. The multispectral image is clustered into superpixels to explore self-similarity, and then, the superpixels are unmixed to obtain an abundance matrix. By imposing a low-rank constraint on the abundance matrix, we further improve the superresolution performance. Experiments on the CAVE and Harvard datasets indicate that our superresolution method outperforms the other compared methods in terms of quantitative evaluation and visual quality.

1. Introduction

With rich spectral and spatial information, hyperspectral images (HSI) have received extensive attention and have been widely used in many fields, especially for remote sensing [1, 2] and medical imaging [3]. High resolution is essential to obtain high performance in many HSI applications, such as spectral unmixing, pixelwise classification, object detection, and object tracking. However, due to hardware limitations, hyperspectral imaging requires long exposure times to ensure a high signal-to-noise ratio, which leads to low spatial resolution [4]. Image superresolution is a promising way to acquire high-resolution images in both spatial and spectral domains.

In recent decades, several excellent methods were proposed for HSI superresolution fusing observed low-resolution HSI and the corresponding high-resolution multispectral image (MSI). There were some HSI superresolution methods based on sparse representation theory that is widely used in computer vision tasks. Akhtar et al. [4] impose a nonnegative

constraint on spectral dictionary and sparse representations to improve HSI superresolution performance. Dong et al. [5] utilized structural sparsity constraint to exploit the clustering-based sparsity. The superpixel-based sparse representation model [6] learns sparse coding upon MSI superpixel to use spatial self-similarity of the MSI. Han et al.'s work [7] combined nonnegative sparse representation, local similarity, and nonlocal similarity for HSI superresolution, which explores self-similarity in superpixel and across the entire image. However, when learning the spectral dictionary, these methods do not take into account the spectral correlation in the HSI causing severe spectral distortion in the reconstructed images.

In order to overcome this problem, spectral unmixing-based superresolution methods were proposed. Yokoya et al. [8] exploited nonnegative property while unmixing HSI and MSI. Wycoff et al. [9] further introduced sparse constraint in extracting the endmember and the abundance matrix. HSI and MSI were jointly unmixed with sparse constraint and nonnegative and abundance sum-

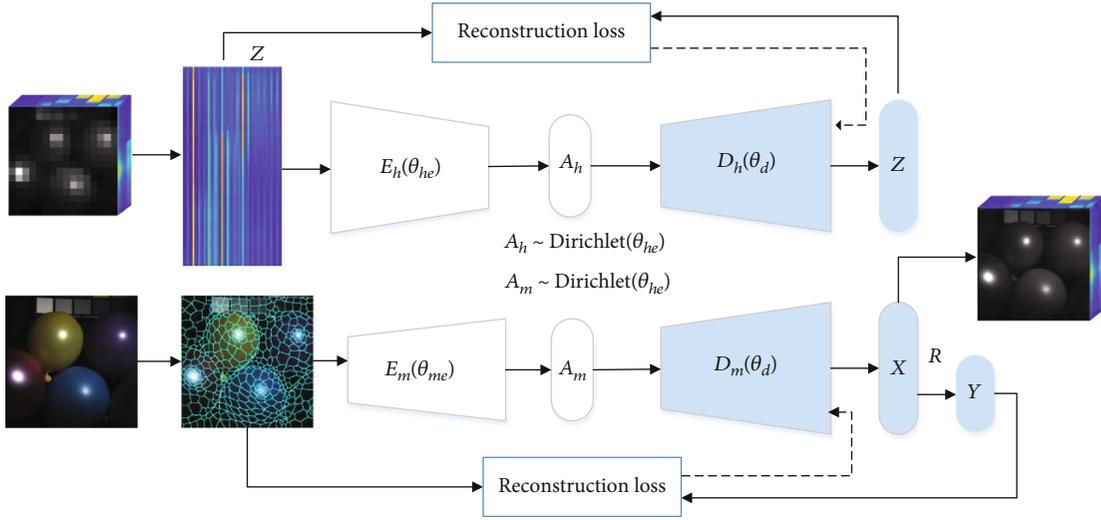


FIGURE 1: The schematic diagram of the proposed approach: (1) cluster the high-resolution MSI into superpixels; (2) unmix the low-resolution HSI and superpixels in MSI to obtain endmember and abundance matrix with useful constrains by a coupled network.

to-one constraints in [10]. Zou and Xia [11] explored spatial structure information in spectra and abundance via mutual distance and graph Laplacian regularization, respectively. These methods showed excellent superresolution performance. However, spectral unmixing-based methods generally assume that the downsampling operation from high-resolution to low-resolution images is known. The assumption is not always often valid.

Inspired by the great success of deep learning-based methods in the superresolution recovery of colour images, researchers have attempted to apply the deep learning approach to HSI superresolution. These deep learning-based methods mainly learn the mapping from low-resolution images to high-resolution images in a supervised way. For example, the typical models are the autoencoder, 3D convolution neural network [12, 13], and deep residual networks [14, 15]. These methods assumed that there exists consistent mapping across different image pairs. However, it is not always valid, thus causing severe spectral distortion.

In this paper, the local similarity in high-resolution images is analysed, and the low-rank property is explored in HSI. Then, we propose the superresolution algorithm based on the superpixel spectral unmixing. The coupled encoder-decoder network with sparsity is employed to extract endmembers and abundance information, which naturally integrates meaningful constraints to enhance the unmixing quality. As shown in Figure 1, the shared decoder preserves endmember information for the hyperspectral image. To obtain accurate high-resolution abundance information, we cluster high-resolution images into superpixels, which has adaptive shape and size, and impose a low-rank constraint on the abundance matrix. Finally, we combine the endmember and the abundance matrix to reconstruct the high-resolution HSI. The combination of the coupled encoder-decoder network and low-rank constraint leads to accurate endmember and abundance information so that promising HSI superresolution performance is obtained.

2. Spectral Unmixing-Based HSI Superresolution

The goal of HSI superresolution is to predict high-resolution HSI $\bar{X} \in \mathbb{R}^{W \times H \times B}$, where W and H denote image width and height in spatial dimensions and B indicates the spectral dimension. The spectral unmixing-based method takes HSI $\bar{Z} \in \mathbb{R}^{w \times h \times B}$ and MSI $\bar{Y} \in \mathbb{R}^{W \times H \times b}$ as input. For notational convenience, we represent the three-dimensional hyperspectral image in matrix form. Accordingly, we can get matrix $X \in \mathbb{R}^{B \times N}$, $Y \in \mathbb{R}^{b \times N}$, and $Z \in \mathbb{R}^{B \times n}$ where $N = W \times H$, $n = w \times h$. We can estimate the desired HSI X , as follows:

$$\arg \min_X \|Z - XD\|_F^2 + \|Y - RX\|_F^2, \quad (1)$$

where matrix $D \in \mathbb{R}^{N \times N}$ models the spatial downsampling and blurring operators. $R \in \mathbb{R}^{B \times b}$ is the spectral degradation matrix depending on the hyperspectral sensor. In this paper, we assume that the spectral response function is known.

The naive estimation of the desired HSI according to (1) is computationally complex and inaccurate. Thus, spectral unmixing is used to deal with these problems efficiently. According to the linear spectral mixing model [16], we have $X = EA$ where $E = [e_1, e_2, \dots, e_c]$ is the endmember matrix and c represents the number of spectral bases; $A = [a_1, a_2, \dots, a_N]$ is the abundance matrix. In the linear mixture model, the endmember represents the reflectance spectrum. The abundance is the proportion of endmembers, and thus, they should be nonnegative. Also, each column of abundances should meet sum to one (STO) requirement, i.e., $\sum_1^T = \sum_1^T a_i, i = 1, 2, \dots, n$. The fact that there are several endmembers in one pixel means that the abundance also should be sparse. Therefore, we can get endmembers and abundances according to

$$\arg \min_{E,A} \|Z - EAD\|_F^2 + \|Y - REA\|_F^2 + \lambda_1 \|A\|_0, \quad (2)$$

where $\|A\|_0$ denotes sparse constraint.

3. Method

This section describes the proposed HSI superpixel resolution method in detail. The proposed method is based on the superpixel spectral unmixing and the coupled encoder-decoder network. As shown in Figure 1, in the proposed method, endmembers and abundance are extracted using the coupled encoder-decoder network. In order to take advantage of self-similarity in the high-resolution image, the network extracts abundance from the superpixels of the high-resolution image.

3.1. MSI Superpixel Segmentation and Low-Rank Constraint. The square window cannot effectively represent the complex structure of the image since space information in the natural image is usually not regular. Therefore, the traditional unmixing methods, pixel-based and patch-based methods, cannot fully utilize the spatial information of the image. The high-resolution MSIs are locally self-similar, that is, adjacent pixels in MSI have a similar spectral response. Therefore, we cluster MSI into superpixels to extract the abundance information.

Superpixel was initially proposed by Ren and Malik [17], which groups pixels based on spectral response and other properties. According to the evaluation of 28 state-of-the-art superpixel segmentation algorithms [18], we adopt the efficient topology-preserving segmentation (ETPS) algorithm [19] to obtain superpixels from the MSI. The ETPS formulates the segmentation problem with an objective function similar to k -means clustering, making the obtained superpixels be coherent in appearance and also have a regular shape. The objective function is a Markov energy function, and its definition is as follows:

$$\begin{aligned} E_{\text{mono}}(s, \mu, c) = & \sum_p E_{\text{col}}(s_p, c_{s_p}) + \lambda_{\text{pos}} \sum_p E_{\text{pos}}(s_p, \mu_{s_p}) \\ & + \lambda_b \sum_p \sum_{q \in N_8} E_b(s_p, s_q) + E_{\text{topo}}(s) + E_{\text{size}}(s), \end{aligned} \quad (3)$$

where $c = (c_1, c_2, \dots, c_M)$ and $\mu = (\mu_1, \mu_2, \dots, \mu_M)$ are the set of centres and mean positions for all superpixels. N_8 denotes the eight neighbourhoods of the pixel p . $E_{\text{col}}(s_p, c_{s_p})$ encourages colour homogeneity of each superpixel, $E_{\text{pos}}(s_p, \mu_{s_p})$ encourages that the superpixels should be regular in shape, $E_b(s_p, s_q)$ is used to encourage the superpixel to have small boundary length, and $E_{\text{topo}}(s)$ and $E_{\text{size}}(s)$ promote topological connections and superpixel size, respectively.

Since the pixels within a superpixel have a similar spectral response, the rank of matrix reshaped by the superpixel is low. Therefore, when a hyperspectral image is unmixed in superpixels, the low-rank constraint can be imposed. Also,

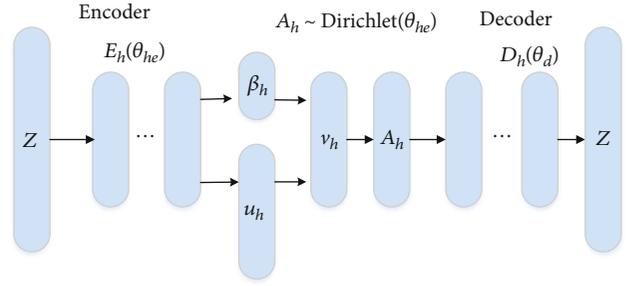


FIGURE 2: The details of the asymmetric stack autoencoder network for HSI.

the subspace where the high-resolution HSI resides is the same as space spanned by the endmember matrix. Based on the fact, we impose the low-rank constraint on the abundance matrix, not on the high-resolution HSI. The optimization model of the coupled network is defined as

$$\arg \min_{E,A} \|Z - EAD\|_F^2 + \|Y - REA\|_F^2 + \lambda_1 \|A\|_0 + \lambda_2 \|A\|_*, \quad (4)$$

with the nonnegative and STO constrains. Here, $\|\bullet\|_*$ denotes matrix nuclear norm, which approximatively replaces low-rank constraint.

3.2. Coupled Encoder-Decoder Network. Motivated by the work in [20], we propose the coupled encoder-decoder network to solve the HSI superresolution problem described in (4). As shown in Figure 2, the basic structure of the coupled network is an asymmetric stack autoencoder network. During the reconstruction process of HSI and MSI, the endmember and abundance information in the image scene are extracted. The autoencoder network consists of an encoder and a decoder. The encoder $E_h(\theta_{he})$ maps the hyperspectral data to the low-dimensional representation layer; the decoder $D_h(\theta_{hd})$ reconstructs the representation layer into hyperspectral data:

$$\begin{aligned} A_h &= f_k \left(W_{he}^k \left(f_{k-1} \left(\dots f_1 \left(Z W_{he}^1 + b_{he}^1 \right) \dots \right) + b_{he}^{k-1} \right) + b_{he}^k \right), \\ \tilde{Z} &= f_k \left(W_d^k \left(f_{k-1} \left(\dots f_1 \left(A_h W_d^1 \right) \dots \right) \right) \right), \end{aligned} \quad (5)$$

where W_{he}^k and W_d^k denote the weights in the k^{th} layer of encoder and decoder, respectively, and f_k denotes the activation function in the k^{th} layer.

Note that when extracting endmember from HSI through the network, the hidden layer A_h should reflect the abundance information, which means A_h needs to meet nonnegative requirements and sum-to-one property. The ReLU function is used as an activation function in the encoder to make sure the hidden layer is nonnegative. Following [21], we make the hidden layer variables obey Dirichlet distribution, which is generated using the stick-breaking process.

Let $\alpha = (\alpha_1, \dots, \alpha_i, \dots, \alpha_n)^T$ represent the hidden layer variable, then α_i is defined as follows:

$$\alpha_i = \begin{cases} v_1, & i = 1, \\ v_i \prod_{j < i} (1 - v_j), & i > 1, \end{cases} \quad (6)$$

$$v_i \sim \left(1 - \left(1 - u^{1/\beta}\right)\right),$$

where v_i is sampled from the inverse transform of the Kumaraswamy distribution. u and β are obtained from the encoder-decoder network.

The bias parameter is not used in the decoder. The identity function is used as the activation function, so the parameters θ_d in the encoder have $\theta_d = W_{d1} W_{d2} \dots W_{dk}$. The weights of the decoder correspond to the endmember. In this way, the deep network can extract the spectral information from HSI, and the hidden layer preserves the spatial information effectively. Similarly, the high-resolution abundance information can be extracted during the reconstruction of the high-resolution MSI.

Due to the limited band number of MSI, it is not ideal for unmixing MSI to generate endmember and abundance information. Since HSI and MSI reflect ground object information in the same scene, the endmember information of MSI and HSI is correlated: i.e., $E_m = RE_h$. Thus, the weights of the decoder are shared for both HSI and MSI, and they are frozen while learning autoencoder network parameters for MSI.

3.3. Network Structure and Implementation Detail. The L_0 regularization in (4) is nonconvex and usually replaced with L_1 regularization. However, the abundance vector should need the STO constraint. Therefore, the commonly used L_1 regularization would not guarantee whether the abundance vector is sparse. As similar to [21], we introduce the generalized Shannon entropy function as the sparsity constraint of the hidden layer. The definition of the function is as follows:

$$H_p(s) = - \sum_{j=1}^N \frac{|s_j|^p}{\|s\|_p^p} \log \frac{|s_j|^p}{\|s\|_p^p}. \quad (7)$$

Because the latent variables of representation layers are nonnegative, we choose $p = 1$ so that it is computationally efficient. Substituting (7) into (4), the optimization function can be rewritten as

$$\arg \min_{E,A} \|Z - EAD\|_F^2 + \|Y - REA\|_F^2 + \lambda_1 H_1(A) + \lambda_2 \|A\|_*. \quad (8)$$

In this paper, we solve the optimization problem via the coupled deep encoder-decoder network. Following the general practice of deep networks, the L2 norm is used on the decoder weights to prevent overfitting. The objective

TABLE 1: The structure of the coupled network.

Coupled network	Encoder			Decoder		
	Layer and node Input layer	Hidden layer	u β v	Layer and node Hidden layer	Out layer	
HSI	31	[31, 10, 10]	10 1 10	[10, 10, 10]		31
MSI	3	[5, 7, 9, 10]	10 1 10	[10]		31

TABLE 2: The details of benchmark datasets.

Dataset	Resolution	Range of wavelength	Bands	Number of scenes
CAVE	512 * 512 pixel	400~700 nm	31	32
Harvard	1392 * 1040 pixels	420~720 nm	31	50

functions of the coupled network are expressed as follows:

$$L(\theta_{he}, \theta_d) = \frac{1}{2} \|Z - \tilde{Z}\|_F^2 + \lambda_1 H_1(A_h) + \mu \|\theta_d\|_F^2, \quad (9)$$

$$L(\theta_{me}) = \frac{1}{2} \|Y - \tilde{Y}\|_F^2 + \lambda_1 H_1(A_m) + \lambda_2 \|A_m\|_*, \quad (10)$$

where λ_1 , λ_2 , and μ are used for balancing the reconstruction error, sparsity constraint, weights loss, and locally low-rank constraint, respectively. \tilde{Z} and \tilde{Y} are reconstructed images, respectively, A_h and A_m are abundance information of HSI and MSI, respectively.

The coupled network consists of two sparse autoencoder networks to extract endmember and abundance information from HSI and MSI, respectively. The network is optimized as follows: first, the autoencoder extracts endmember information from HSI given the objective function in (9); then, the learned decoder weights θ_d are frozen and shared with the decoder of the autoencoder network for MSI. Only the encoder weights θ_{me} of the autoencoder network for MSI are updated according to the objective function in (10).

In the experiments, the number of endmembers is set as ten; that is, the hidden layer of the autoencoding network has ten nodes. The details of the coupled network are shown in Table 1. The autoencoder network for HSI includes three hidden layers. In contrast, the autoencoder network for MSI includes four layers and the nodes for each layer increase from five to ten. Since different HSIs describe different scenes, the changes in ground materials in different scenes may be massive. We extract spectral and spatial information from HSI and the corresponding MSI, respectively.

4. Experimental Results

Two commonly used hyperspectral datasets: CAVE [22] and Harvard [23], are employed to evaluate the proposed method. The details for the two datasets are shown in Table 2. We crop top-left 1024 * 1024 pixels for Harvard in our experiments. For two benchmark datasets, the original images are scaled to [0, 1]. The original image in the dataset is used as ground truth. To simulate low-resolution HSI, we downsample HSI by using the average over 32 * 32 disjoint

TABLE 3: The RMSE and SAM of benchmarked examples.

Methods			CNMF [8]	GSOMP+ [4]	BSR [24]	SNNMF [9]	CSU [10]	NSSR [5]	uSDN [21]	Ours
CAVE	CD	RMSE	7.80	10.55	10.93	7.14	7.71	6.07	4.94	4.56
		SAM	13.10	20.21	9.47	14.12	12.16	5.34	9.88	6.67
	Photo	RMSE	9.51	3.04	2.52	9.21	5.28	2.90	2.32	1.86
		SAM	16.62	11.09	8.54	18.04	10.82	14.60	5.84	5.33
	Spool	RMSE	9.78	5.80	5.91	12.30	8.54	3.06	5.54	4.12
		SAM	14.38	16.01	14.86	11.05	18.10	14.60	13.04	11.64
Harvard	Img1	RMSE	2.32	2.21	1.72	3.01	1.99	3.01	2.0	1.36
		SAM	2.55	2.63	2.26	2.18	2.93	2.23	1.60	1.42
	Imgb5	RMSE	1.92	2.73	1.33	2.49	1.64	2.54	1.15	0.91
		SAM	2.04	3.36	2.25	1.97	2.65	2.10	1.70	1.54

TABLE 4: The average RMSE, PSNR, SAM, and ERGAS over CAVE and Harvard datasets.

Methods			CNMF [8]	GSOMP+ [4]	BSR [24]	SNNMF [9]	CSU [10]	NSSR [5]	uSDN [21]	Ours
CAVE	RMSE	8.98	8.05	4.93	8.45	6.73	4.67	4.12	3.69	
	PSNR	28.72	30.42	35.86	29.33	32.40	36.65	37.48	38.02	
	SAM	11.55	14.47	7.54	9.75	8.17	7.26	6.93	6.53	
	ERGAS	0.79	0.82	0.62	0.73	0.68	0.58	0.51	0.46	
Harvard	RMSE	3.98	4.18	2.47	3.48	3.07	2.29	1.78	1.68	
	PSNR	38.11	37.35	42.36	40.03	41.07	42.63	44.22	45.02	
	SAM	4.87	5.02	3.43	4.25	3.83	3.31	2.85	2.63	
	ERGAS	0.43	0.45	0.31	0.39	0.36	0.28	0.22	0.19	

blocks. The MSIs are synthesized using the spectral response of the Nikon D700 camera integrating the raw data along the spectral dimensions.

In order to evaluate the accuracy of the estimated high-resolution HSI, four widely used quality metrics are used, including root mean square error (RMSE), peak signal-to-noise ratio (PSNR), spectral angle mapper (SAM), and relative dimensionless global error in synthesis (ERGAS). For PSNR, the larger value indicates better performance, while the smaller value indicates better performance for the other three metrics.

4.1. Comparison with State-of-the-Art Methods. The proposed method is compared with existing HSI superresolution methods including CNMF [8], SNNMF [9], GSOMP+ [4], BSR [24], CSU [10], NSSR [5], and uSDN [21]. In the experiment, we use the original code provided by the authors, excluding uSDN. Among them, CNMF, SNNMF, CSU, and NSSR need the downsampling function as a priori, which is generally unknown in practical. Thus, the downsampling operation in the original implementation is replaced by bicubic function, which is different from the method to simulate low-resolution HSI.

Table 3 shows the RMSE and SAM for some examples from CAVE and Harvard datasets. Table 4 shows the average

RMSE, PSNR, SAM, and ERGAS results. Tables 3 and 4 show that our method outperforms all other methods and has a significant improvement on the CAVE dataset. The unmixing-based methods usually have a better SAM score. However, CNMF [8] and SNNMF [9] are not competitive on the CAVE dataset in terms of RMSE because they do not exploit constraints on endmember and abundance matrix. The sparse representation-based approaches usually have better reconstruction errors, i.e., GSOMP+ [4], BSR [24], and NSSR [5] can achieve better RMSE. However, the first two methods have a larger SAM score. The reason is that the spectral correlation in HSI is not considered. The uSDN [21] could extract accurate spectral information and spatial representations via a sparse Dirichlet network, minimizing angle similarity that leads to better RMSE and smaller SAM. Our method further explores the spatial structure information and local low-rank property to estimate abundance matrix, achieving a better reconstruction effect than other methods.

4.2. The Effects of Parameters. There are three parameters in the proposed method: λ_1 , λ_2 , and μ . Here, λ_1 and λ_2 are used to balance the reconstruction error and sparse constraint and low-rank constraint, and μ is used for the decoder weight loss. In the experiments, the parameters are set to $\mu = 1 \times$

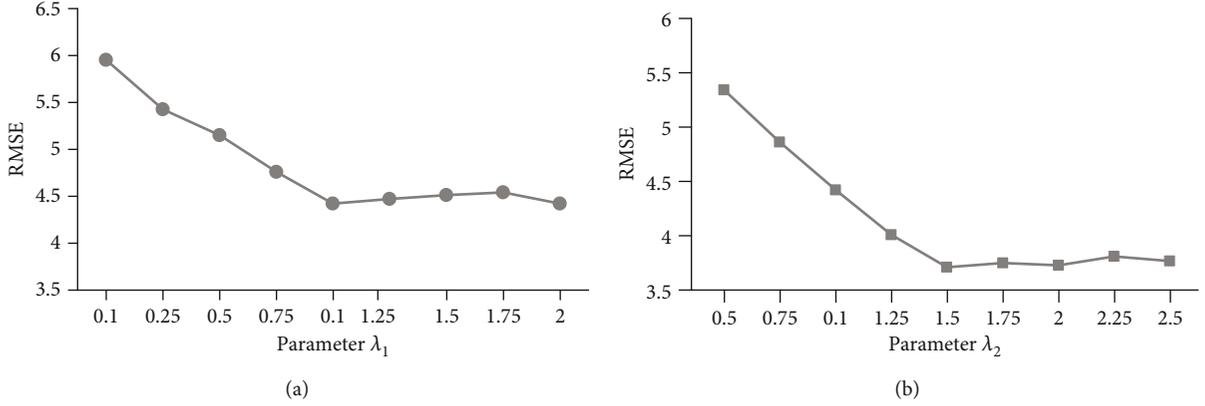


FIGURE 3: The evaluated RMSE with different parameters on the CAVE dataset. (a) The RMSE results according to λ_1 . (b) The RMSE results according to λ_2 .

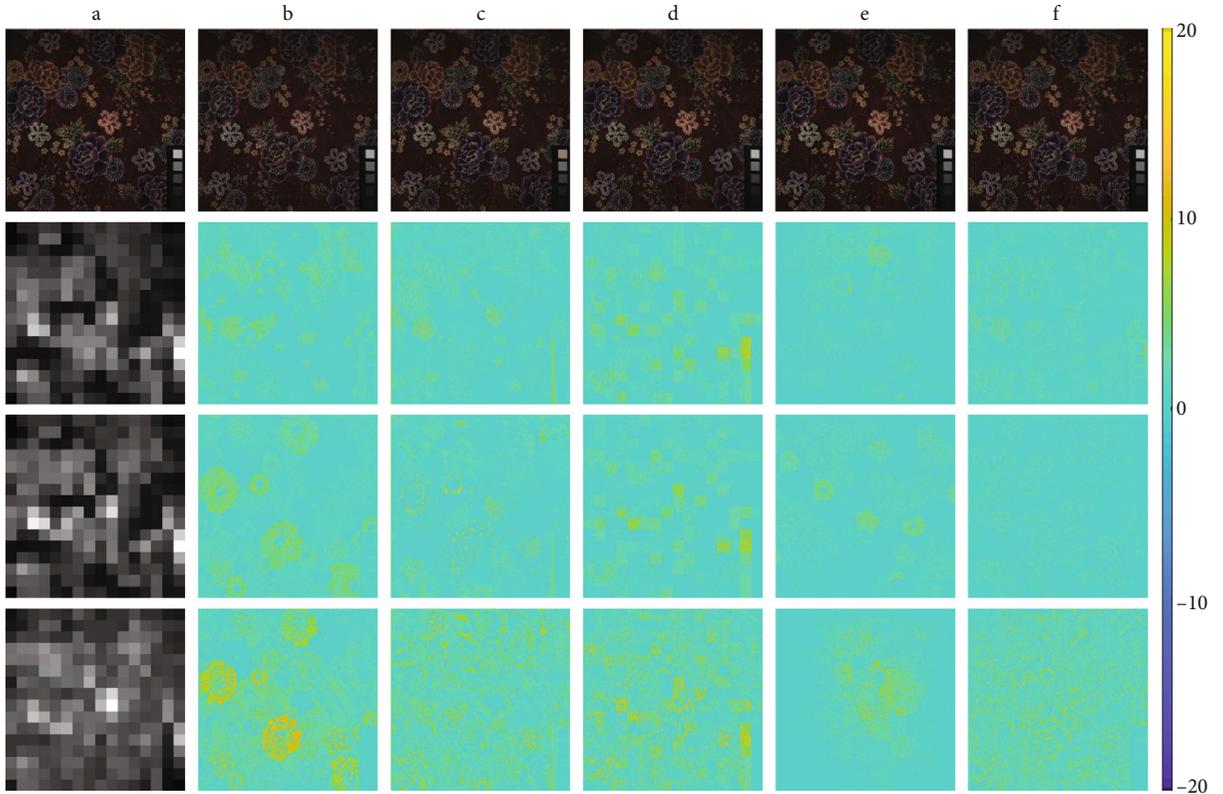


FIGURE 4: The recovered high-resolution pseudocolour images of “cloth” image in the CAVE dataset and the difference image at wavelengths 460, 540, and 620 nm: (a) from top to bottom, GT false colour image, and input HSI at wavelengths 460, 540, and 620 nm; (b–f) from top to bottom, the reconstructed pseudocolour image and difference image by (b) BSR [23], (c) CSU [10], (d) NSSR [5], (e) uSDN [20], and (f) ours.

e^{-6} . To select the optimal parameters, we evaluate the HSI recovery performance according to different λ_1 and λ_2 . We set $\lambda_1 = [0.1, 0.5, 1, 1.5, 2] \times 10^{-6}$ and $\lambda_2 = [0.5, 1, 1.5, 2, 2.5] \times 10^{-1}$, respectively. We first evaluate the RMSE by adjusting λ_1 while fixing the low-rank constrain weight λ_2 . In similar way, we choose the appropriate weight parameter λ_2 . Figure 3 shows the changed RMSE according to varied parameters. Accordingly, optimal values of $\lambda_1 = 1 \times 10^{-6}$ and $\lambda_2 = 1.5 \times 10^{-1}$ are selected and used in our experiments.

4.3. *Qualitative Comparisons.* Figures 4 and 5 show the recovered HSIs and the absolute difference images at wavelengths 460 nm, 540 nm, and 670 nm, for “cloth” from CAVE and “imgb8” from Harvard, respectively. The recovered images from BSR [23], CSU [10], NSSR [5], uSDN [20], and our method are compared according to their performance in Table 4. The results show that the absolute difference by our method (Figures 4(f) and 5(f)) is smaller than other methods, which indicates recovered image by the

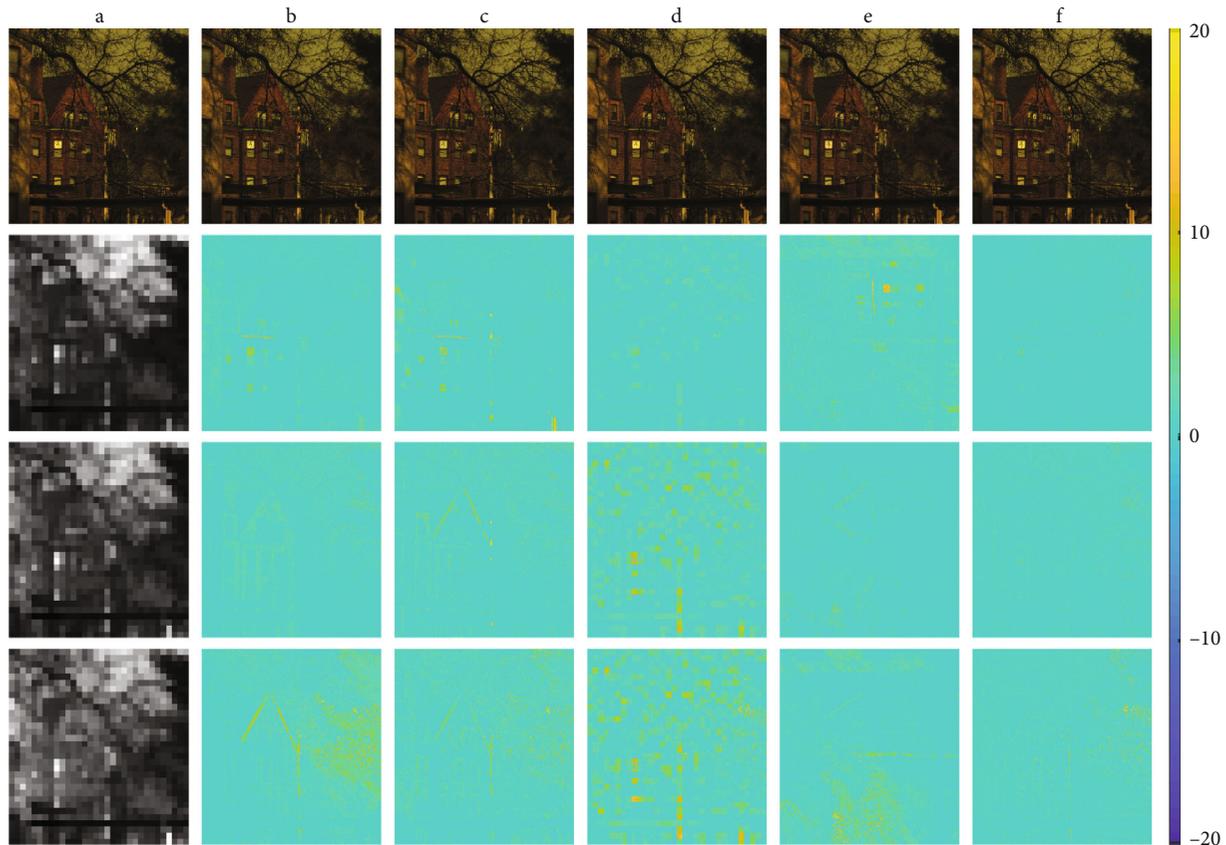


FIGURE 5: The recovered high-resolution pseudocolour images of “imgb8” image in the Harvard dataset and the difference image at wavelengths 460, 540, and 620 nm: (a) from top to bottom, GT false colour image, and input HSI at wavelengths 460, 540, and 620 nm; (b–f) from top to bottom, the reconstructed colour image and difference image by (b) BSR [23], (c) CSU [10], (d) NSSR [5], (e) uSDN [20], and (f) ours.

proposed method is more similar with the original image. Compared with uSDN [20] (Figures 4(e) and 5(e)), our method adopts superpixel segmentation and low-rank constraint. The smaller difference indicates better performance of superresolution performance, which means the effectiveness of our improvements.

5. Conclusion

In this paper, we propose the HSI superresolution method based on superpixel spectral unmixing with a deep network. The method combines spectral unmixing with superpixel to simultaneously exploit spectral and spatial information in the image. Due to the powerful representation capabilities of deep networks, we can preserve more accurate spectral information. Experiments on two widely used HSI datasets show that our method achieved outstanding performance over other methods. The average RMSE and SAM on CAVE dataset reach 3.69 and 6.53, respectively. In future work, we will explore spectral correlation in HSI for superresolution.

Data Availability

CAVE dataset is available from the website: <https://www.cs.columbia.edu/CAVE/databases/multispectral/>. Harvard data-

set is available from the website: <http://vision.seas.harvard.edu/hyperspec/index.html>.

Conflicts of Interest

The authors declare no conflict of interests.

Acknowledgments

We would like to thank Prof. F. Yasuma for CAVE dataset and Prof. Ayan Chakrabarti for Harvard dataset. This work is jointly supported by the National Natural Science Foundation for Young Scientists of China (Grant No: 61403397, 61503389, and 61202332) and the Natural Science Foundation of Shaanxi Province, China (Grant No: 2015JM6313).

References

- [1] J. Xia, N. Falco, J. A. Benediktsson, P. du, and J. Chanussot, “Hyperspectral image classification with rotation random forest via KPCA,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 4, pp. 1601–1609, 2017.
- [2] X. Cao, F. Zhou, L. Xu, D. Meng, Z. Xu, and J. Paisley, “Hyperspectral image classification with Markov random fields and a

- convolutional neural network,” *IEEE Transactions on Image Processing*, vol. 27, no. 5, pp. 2354–2367, 2018.
- [3] G. Lu and B. Fei, “Medical hyperspectral imaging: a review,” *Journal of Biomedical Optics*, vol. 19, no. 1, article 010901, 2014.
- [4] N. Akhtar, F. Shafait, and A. Mian, “Sparse spatio-spectral representation for hyperspectral image super-resolution,” in *Computer Vision – ECCV 2014. ECCV 2014*, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds., vol. 8695 of Lecture Notes in Computer Science, pp. 63–78, Springer, Cham, 2014.
- [5] W. Dong, F. Fu, G. Shi et al., “Hyperspectral image super-resolution via non-negative structured sparse representation,” *IEEE Transactions on Image Processing*, vol. 25, no. 5, pp. 2337–2352, 2016.
- [6] L. Fang, H. Zhuo, and S. Li, “Super-resolution of hyperspectral image via superpixel-based sparse representation,” *Neurocomputing*, vol. 273, pp. 171–177, 2017.
- [7] X.-H. Han, B. Shi, and Y. Zheng, “Self-similarity constrained sparse representation for hyperspectral image super-resolution,” *IEEE Transactions on Image Processing*, vol. 27, no. 11, pp. 5625–5637, 2018.
- [8] N. Yokoya, T. Yairi, and A. Iwasaki, “Coupled nonnegative matrix factorization unmixing for hyperspectral and multi-spectral data fusion,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 2, pp. 528–537, 2012.
- [9] E. Wycoff, T.-H. Chan, K. Jia, W. K. Ma, and Y. Ma, “A non-negative sparse promoting algorithm for high resolution hyperspectral imaging,” in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 1409–1413, Vancouver, BC, Canada, 2013.
- [10] C. Lanaras, E. Baltsavias, and K. Schindler, “Hyperspectral super-resolution by coupled spectral unmixing,” in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 3586–3594, Santiago, Chile, 2015.
- [11] C. Zou and Y. Xia, “Hyperspectral image superresolution based on double regularization unmixing,” *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 7, pp. 1022–1026, 2017.
- [12] F. Palsson, J. R. Sveinsson, and M. O. Ulfarsson, “Multispectral and hyperspectral image fusion using a 3-D-convolutional neural network,” *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 5, pp. 639–643, 2017.
- [13] S. Mei, X. Yuan, J. Ji, Y. Zhang, S. Wan, and Q. du, “Hyperspectral image spatial super-resolution via 3D full convolutional neural network,” *Remote Sensing*, vol. 9, no. 11, p. 1139, 2017.
- [14] C. Wang, Y. Liu, X. Bai, W. Tang, P. Lei, and J. Zhou, “Deep residual convolutional neural network for hyperspectral image super-resolution,” in *Image and Graphics. ICIG 2017*, Y. Zhao, X. Kong, and D. Taubman, Eds., vol. 10668 of Lecture Notes in Computer Science, pp. 370–380, Springer, Cham, 2017.
- [15] Z. Shi, C. Chen, Z. Xiong, D. Liu, Z.-J. Zha, and F. Wu, “Deep residual attention network for spectral image super-resolution,” in *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, pp. 214–229, Munich, Germany, 2018.
- [16] J. M. Bioucas-Dias, A. Plaza, N. Dobigeon et al., “Hyperspectral unmixing overview: geometrical, statistical, and sparse regression-based approaches,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, no. 2, pp. 354–379, 2012.
- [17] X. Ren and J. Malik, “Learning a classification model for segmentation,” in *Proceedings Ninth IEEE International Conference on Computer Vision*, Nice, France, 2003.
- [18] D. Stutz, A. Hermans, and B. Leibe, “Superpixels: an evaluation of the state-of-the-art,” *Computer Vision and Image Understanding*, vol. 166, pp. 1–27, 2018.
- [19] J. Yao, M. Boben, S. Fidler, and R. Urtasun, “Real-time coarse-to-fine topologically preserving segmentation,” in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2947–2955, Boston, MA, USA, 2015.
- [20] R. Guo, W. Wang, and H. Qi, “Hyperspectral image unmixing using autoencoder cascade,” in *2015 7th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, pp. 1–4, Tokyo, Japan, 2015.
- [21] Y. Qu, H. Qi, and C. Kwan, “Unsupervised sparse Dirichlet-Net for hyperspectral image super-resolution,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2511–2520, Salt Lake City, UT, U.S., 2018.
- [22] F. Yasuma, T. Mitsunaga, D. Iso, and S. K. Nayar, “Generalized assorted pixel camera: postcapture control of resolution, dynamic range, and spectrum,” *IEEE Transactions on Image Processing*, vol. 19, no. 9, pp. 2241–2253, 2010.
- [23] A. Chakrabarti and T. Zickler, “Statistics of real-world hyperspectral images,” in *CVPR 2011*, pp. 193–200, Providence, RI, USA, 2011.
- [24] N. Akhtar, F. Shafait, and A. Mian, “Bayesian sparse representation for hyperspectral image super resolution,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3631–3640, Boston, MA, USA, 2015.