HYPERSPECTRAL IMAGE SUPER-RESOLUTION VIA CONVOLUTIONAL NEURAL NETWORK

Shaohui Mei[1], *Xin Yuan*[1], *Jingyu Ji*[1], *Shuai Wan*[1], *Junhui Hou*[2], *Qian Du*[3]

[1] School of Electronics and Information, Northwestern Polytechnical University, Xi'an 710129, China.
[2] Department of Computer Science, City University of Hong Kong, Kowloon, Hong Kong.
[3] Department of Electrical and Computer Engineering, Mississippi State University, MS 39762, USA.

ABSTRACT

Due to the tradeoff between spatial and spectral resolution in remote sensing imaging, hyperspectral images are often acquired with a relative low spatial resolution, which limits their applications in many areas. Inspired by recent achievements in convolutional neural network (CNN) based super resolution (SR), a novel CNN based framework is constructed for SR of hyperspectral images by considering both spatial context and spectral correlation. As a result, the spectral distortion incurred by directly applying traditional SR algorithms to hyperspectral images is alleviated. Experimental results on several benchmark hyperspectral datasets have demonstrated that higher quality of reconstruction and spectral fidelity can be achieved, compared to band-wise manner based algorithms.

Index Terms— Hyperspectral, super-resolution, convolutional neural network, deep learning

1. INTRODUCTION

Hyperspectral remote sensing usually collects solar reflectance information of objects in hundreds of contiguous bands over a certain electromagnetic spectrum. Such technology generally offers images with very high spectral resolution, enabling a fine discrimination of different objects by their spectral signatures. Due to the limitation of sensor technologies, there exists a balance between spatial resolution and spectral resolution. As a result, hyperspectral images are often acquired under a relatively low spatial resolution. For example, the Airborne Visible Infrared Imaging Spectromete (AVIRIS) sensor collects images with 224 spectral bands ranging from 370nm to 2500nm. However, its spatial resolution is only 20m, indicating one pixel corresponds to ground objects of about 400 m^2 area, which limits its applications in many areas. Therefore, it is highly desirable to increase the spatial resolution of hyperspectral images via postprocessing, which can benefit to many applications, such as target recognition, feature classification, and environmental change detection.

Generally, there are several ways to overcome the limitation brought by low spatial resolution in hyperspectral applications. Spectral mixture analysis acquires the fractional abundance of pure ground objects within a pixel [1, 2]. However, the spatial distribution/location of pure ground objects within a pixel requires further estimation. The spatial resolution of hyperspectral images can also be enhanced by fusing it with an auxiliary high-spatial-resolution multispectral or panchromatic image [3, 4]. However, such spatial registered auxiliary sources are rare or sometimes impossible. Therefore, it is necessary to enhance the spatial resolution of hyperspectral images by super-resolution algorithms that require no additional priors. Besides, many algorithms for SR of gray/color images have been proposed for this purpose, such as sparse representation based algorithms [5, 6] and deep learning based algorithms [7, 8]. These algorithms can also been directly applied to the SR of hyperspectral images in a band-by-band manner. However, such implementations ignore the relationship among spectral bands, leading to spectral distortion. The results provided in [9] have confirmed that integrating both spectral and spatial information can contribute SR of hyperspectral images, especially for the noisy cases.

Recently, deep learning has brought in revolutionary achievements in many applications. Many classic problems can be formulated as image transformation tasks [10], where a system receives some input image and transforms it into an output one as desired. SR for natural color images has also been processed by deep learning methods by considering the problem as an image transform task [7, 8]. Inspired by convolutional neural network (CNN) used for single image super-resolution [7, 8], a CNN based hyperspectral image SR framework is proposed by using both the spatial context between neighboring pixels and the spectral correlation in adjacent band images. As a result, the spectral distortion can be alleviated. Finally, experiments on two benchmark datasets are conducted to demonstrate the effectiveness of the proposed CNN for SR of hyperspectral images.

This work is partially supported by National Natural Science Foundation of China (61671383) and Fundamental Research Funds for the Central Universities(3102016ZB012, 3102016ZY019). Email: meish@nwpu.edu.cn.

2. PROPOSED METHOD

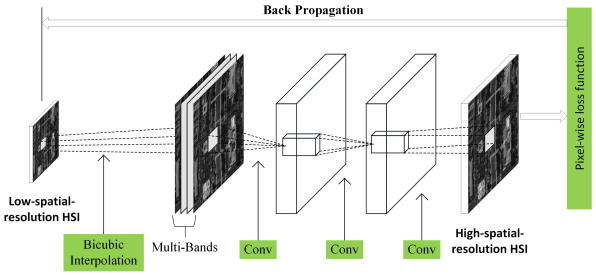


Fig. 1. Proposed CNN based hyperspectral image SR framework.

The framework of our proposed CNN for hyperspectral image SR is shown in Fig. 1. By using bi-cubic interpolation, an initial SR is conducted to generate a low spatial-resolution hyperspectral image. Then spectral neighboring bands are used to improve the performance of initial SR by integrating both spatial context and spectral correlation. Particularly, (3D) convolution operation is used to improve spectral fidelity by exploring the similarity in both spatially neighboring pixels and spectrally neighboring bands. As a result, the spectral distortion in hyperspectral SR problems can be reduced.

Let $\mathbf{X} \in \mathbb{R}^{m\times n\times b}$ be the low spatial-resolution hyperspectral image, where m, n, and b represent the number of rows, columns, and bands of the image, respectively. The aim of SR is to restore a high spatial-resolution hyperspectral image $\mathbf{Y} \in \mathbb{R}^{km\times kn\times b}$, where k is an integer ratio for SR. As shown in Fig. 1, in the proposed framework, a initial high spatial-resolution hyperspectral image $\mathbf{Y1} \in \mathbb{R}^{km\times kn\times b}$ is reconstructed by bi-cubic interpolation. Such initial SR only using spatial context to reconstruct a larger-size hyperspectral images, and spectral distortion may occur. Therefore, a CNN is constructed to restore $\mathbf{Y}$ from $\mathbf{Y1}$ by alleviating spectral distortion integrating both spatial context and spectral correlation.

Assume $\mathbf{B} \in \mathbb{R}^{h\times w\times c}$ is fed into a 3D-convolution neuron in a CNN, the 3D convolution is represented as

$$I = \sum_{p=1}^{h}\sum_{q=1}^{w}\sum_{l=1}^{c} \mathbf{W}_{p,q,l}\mathbf{B}_{p,q,l}, \tag{1}$$

where $\mathbf{W} \in \mathbb{R}^{h\times w\times c}$ is parameterized weights in 3D convolution. Observed from Eq. (1), the information in all the 3 dimensions are implemented by a convolution operation. If we fed an image into the CNN, both spatial and spectral neighbors are used to reconstruct pixels.

As shown in Fig. 1, three convolution layers are used to reduce spectral distortion continuously in the proposed CNN. An activation layer is connected to each of the first two convolutional layer for nonlinear mapping. The 'ReLU' function is used in activation layers since it improves models fitting without extra computational cost or over-fitting risk [11, 12]. Suppose the input of neurons in activation layers is represented as $I_i, i = 1, 2, \ldots$, their output is defined as

$$O_i = \max\{I_i, 0\}, i = 1, 2, \ldots, \tag{2}$$

where $\max\{\cdot\}$ is the 'ReLU' function that outputs the maximum of neuron input and 0.

Table 1. Details of the proposed CNN for SR of hyperspectral images. The input and output are highlighted in bolded.

Layer	Input size	Kernel size	Output size
Conv1	$\mathbf{33}\times\mathbf{33}\times\mathbf{c}$	$9\times 9\times\mathbf{c}\times 64$	$25\times 25\times 64$
Relu1	$25\times 25\times 64$	–	$25\times 25\times 64$
Conv2	$25\times 25\times 64$	$1\times 1\times 64\times 32$	$25\times 25\times 32$
Relu2	$25\times 25\times 32$	–	$25\times 25\times 32$
Conv3	$25\times 25\times 32$	$5\times 5\times 32\times 1$	$\mathbf{21}\times\mathbf{21}\times\mathbf{1}$

Generally, the number of parameters to be optimized is proportional to that of neurons in a CNN. Therefore, the convolution operation is conducted in block-wise to restrict the scale of the CNN as shown in Fig. 1. As a result, less labeled samples would be required for training. The size of convolutional layer and 'ReLU' layer are listed in Table 1. It is observed that the input of the CNN is of the size $33\times 33\times c$ and the size of output is $21\times 21\times 1$, indicating that 33×33 pixels in c continuous neighboring bands is adopted to restore 21×21 pixels in a band. Both spatial extension (using 33×33 pixels to restore 21×21 pixels) and spectral extension (using pixels in c continuous neighboring bands to restore pixels in one band), are used so that the proposed method can fully takes advantages of both spatial context and spectral correlation.

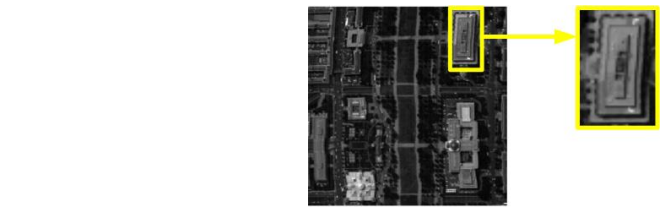
In the proposed CNN for SR of hyperspectral images, each time an image tensor $\mathbf{B} \in \mathbb{R}^{h\times w\times c}$ is fed to the CNN to restore $\mathbf{y}_B \in \mathbb{R}^{kh\times kw}$ as accurate as possible. Let $\hat{\mathbf{y}}_B \in \mathbb{R}^{kh\times kw}$ represent its actual restoration results by the CNN. In this paper, the mean squared error (MSE) between actual output $\hat{\mathbf{y}}_B$ and ideal output $\mathbf{y}_B$ is chosen as the cost function to optimize parameters in the network, which is

$$\begin{aligned} MSE(\hat{\mathbf{y}}_B, \mathbf{y}_B) &= \frac{1}{k^2hw}\sum_{i=1}^{k^2hw}\sqrt{(y_i-\hat{y}_i)^2}, \\ \mathbf{y}_B &= [y_1, y_2, \ldots, y_{k^2hw}], \\ \hat{\mathbf{y}}_B &= [\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_{k^2hw}]. \end{aligned} \tag{3}$$

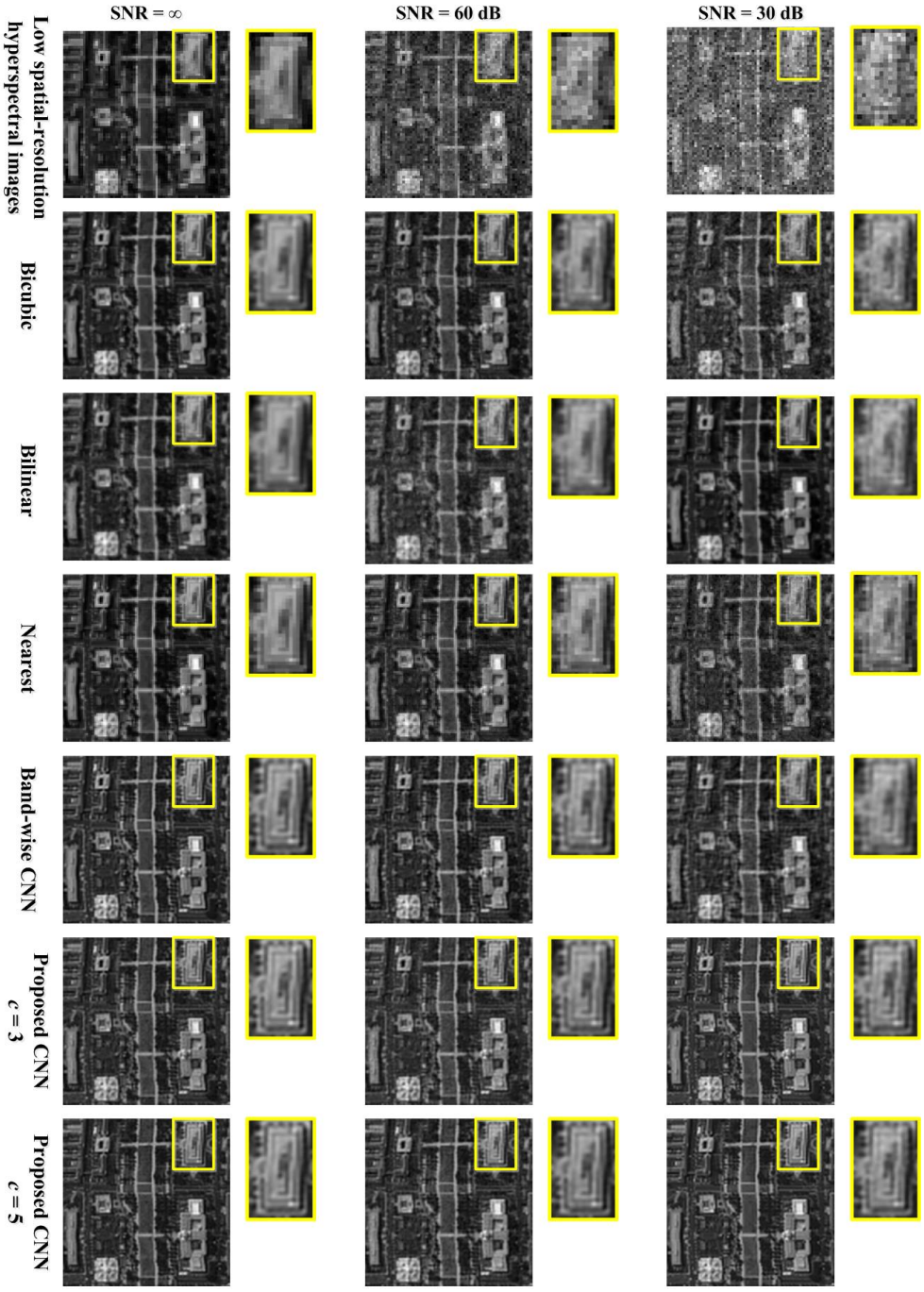
The stochastic gradient descent(SGD) is used to train all the parameters in the CNN.

3. EXPERIMENTS

Two hyperspectral dasetsets are employed in this experiment. The first is acquired by HYDICE sensor over Washington



(a)



(b)

Fig. 2. Band 30 in experimental results of different algorithms on Washington DC Mall hyperspectral dataset: (a). The ground-truth high spatial-resolution hyperspectral image. (b). SR results different algorithms.

Table 2. Quantitative comparison of different methods for SR of hyperspectral images.

			Bicubic	Bilinear	Nearest	Band-wise CNN	Proposed CNN	
							$c = 3$	$c = 5$
WashingtonDC	SNR = ∞	PSNR	47.988	47.039	47.19	48.50	**49.639**	49.472
		SSIM	0.989	0.986	0.987	0.992	0.992	**0.993**
		SAM	0.457	0.518	0.485	0.389	**0.374**	0.375
	SNR=60 dB	PSNR	41.225	41.472	40.274	40.882	42.583	**43.394**
		SSIM	0.986	0.985	0.984	0.987	0.990	**0.991**
		SAM	0.563	0.586	0.616	0.557	0.476	**0.450**
	SNR=30 dB	PSNR	38.005	38.623	36.984	38.507	**41.616**	40.913
		SSIM	0.979	0.98	0.974	0.980	0.987	**0.988**
		SAM	0.763	0.728	0.854	0.726	0.566	**0.555**
Pavia	SNR = ∞	PSNR	31.1	29.909	29.983	33.675	33.933	**33.958**
		SSIM	0.93	0.915	0.921	0.952	**0.954**	**0.954**
		SAM	4.591	5.009	4.786	4.076	**4.009**	4.039
	SNR=60 dB	PSNR	31.096	29.907	29.979	33.659	33.952	**34.014**
		SSIM	0.93	0.915	0.921	0.951	**0.954**	**0.954**
		SAM	4.606	5.019	4.805	4.117	**4.036**	4.050
	SNR=30 dB	PSNR	31.082	29.901	29.966	33.595	33.907	**33.929**
		SSIM	0.9301	0.914	0.920	0.950	**0.953**	**0.953**
		SAM	4.648	5.046	4.86	4.235	**4.079**	4.091

DC Mall, comprising 1280 lines, 307 columns, and 191 bands. The second dataset is acquired by the ROSIS sensor over an urban area of Pavia, Italy. It contains 1096×715 pixels with 102 bands after removing 12 water absorption bands. All the datasets are scaled to [0,1] in the experiments. These two datasets are used as ground-truth high spatial hyperspectral images to train and evaluate the performance of the proposed SR framework. In order to simulate low spatial-resolution hyperspectral images, these two images are firstly down-sampled by a factor of 3. In addition, additive Gaussian white noise is added.

The proposed CNN based framework is designed, trained, and tested based on the Caffe framework [13], in which the size of all strides is set as 1. We choose to restore a 200×200 sub-regions from the two datasets to test the performance of our proposed CNN, while the remaining pixels in the two datasets are used for training. For the SGD based training, the momentum is set as 0.9, the weight decay is chosen as 0.0005, and the base learning rate is 0.0001. The number of continuous bands to restore a single band is set as $c = 3$ or 5.

Four strategies are chosen as baseline for comparison: Bicubic SR, Bilinear SR, SR by nearest neighbor, and CNN based SR for natural color images in [8] which is applied in band-wise. The experimental results of all these algorithms for SR of Washington DC Mall image is shown in Fig. 2. As enlarged in the yellow rectangular, the proposed CNN based algorithms provide slightly better visual results. Three metrics are further used to quantitatively evaluate the performance of SR of hyperspectral images: the mean peak signal-to-noise ratio (MPSNR) index, the mean structural similarity (MSSIM) index, and the average of spectral angle mapper (SAM). MPSNR and MSSIM are used to evaluate the performance of reconstruction, while SAM measures the spectral fidelity of the results. Generally, higher MPSNR value indicates lower error compared with the original image, higher MSSIM value means a better visual quality, and lower MSA value implies less spectral distortion. The qualitative comparison results of all the considered algorithms are listed in Table 2. It is also observed that: 1) the proposed CNN obviously outperforms the other algorithms no matter the number of neighboring band images is selected. 2) compared with other algorithms, the more noisy the dataset is, the better the proposed CNN performs, indicating that the proposed CNN is robust to noise. 3) the proposed CNN based algorithm provides much better spectral fidelity (i.e., smaller SAM) compared with band-wise CNN. This is because both spatial context in neighboring areas and spectral correlation in neighboring band images are considered in the proposed method.

4. CONCLUSION

In this paper, a CNN based framework is proposed for SR of hyperspectral images. Compared with traditional SR algorithms in color images, the proposed CNN based algorithm reconstructs a high spatial-resolution hyperspectral image by integrating both spatial context in neighboring areas and spectral correlation in neighboring band images. As a result, spectral distortion can be alleviated. Experimental results on two simulated datasets demonstrated that the proposed CNN based SR algorithm obviously outperforms existing methods, and yields high spectral fidelity.

5. REFERENCES

[1] S. Mei, M. He, Z. Wang, and D. Feng, "Unsupervised spectral mixture analysis of highly mixed data with hopfield neural network," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 6, pp. 1922–1935, 2014.

[2] S. Mei, Q. Du, and M. He, "Equivalent-sparse unmixing through spatial and spectral constrained endmember selection from an image-derived spectral library," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 6, pp. 2665–2675, 2015.

[3] Y. Zhang, A. Duijster, and P. Scheunders, "A bayesian restoration approach for hyperspectral images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 9, pp. 3453–3462, 2012.

[4] M. Bendoumi, M. He, and S. Mei, "Hyperspectral image resolution enhancement using high-resolution multispectral image based on spectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 52, pp. 6574–6583, 2014.

[5] J. Yang, J Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation.," *IEEE Transactions on Image Processing*, vol. 19, no. 11, pp. 2861, 2010.

[6] X. Gao, K. Zhang, D. Tao, and X. Li, "Image super-resolution with sparse neighbor embedding," *IEEE Transactions on Image Processing*, vol. 21, no. 7, pp. 3194, 2012.

[7] W. Huang, L. Xiao, Z. Wei, H. Liu, and S. Tang, "A new pan-sharpening method with deep neural networks," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 5, pp. 1037–1041, May 2015.

[8] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 2, pp. 295–307, Feb 2016.

[9] J. Li, Q. Yuan, H. Shen, X. Meng, and L. Zhang, "Hyperspectral image super-resolution by spectral mixture analysis and spatial-spectral group sparsity," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 9, pp. 1250–1254, Sept 2016.

[10] J. Johnson, A. Alahi, and F. Li, "Perceptual losses for real-time style transfer and super-resolution," in *European Conference on Computer Vision*. Springer, 2016, pp. 694–711.

[11] K. Jarrett, K. Kavukcuoglu, M. Ranzato, and Y. LeCun, "What is the best multi-stage architecture for object recognition?," in *IEEE 12th International Conference on Computer Vision*, Sept 2009, pp. 2146–2153.

[12] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," *Journal of Machine Learning Research*, vol. 15, 2011.

[13] Y. Jia, E. Shelhamer, J. Donahue, and et al., "Caffe: Convolutional architecture for fast feature embedding," *arXiv preprint arXiv:1408.5093*, 2014.