A New Pan-Sharpening Method With Deep Neural Networks

Wei Huang, *Student Member, IEEE*, Liang Xiao, *Member, IEEE*, Zhihui Wei, Hongyi Liu, and Songze Tang

***Abstract*—A deep neural network (DNN)-based new pan-sharpening method for the remote sensing image fusion problem is proposed in this letter. Research on representation learning suggests that the DNN can effectively model complex relationships between variables via the composition of several levels of nonlinearity. Inspired by this observation, a modified sparse denoising autoencoder (MSDA) algorithm is proposed to train the relationship between high-resolution (HR) and low-resolution (LR) image patches, which can be represented by the DNN. The HR/LR image patches only sample from the HR/LR panchromatic (PAN) images at hand, respectively, without requiring other training images. By connecting a series of MSDAs, we obtain a stacked MSDA (S-MSDA), which can effectively pretrain the DNN. Moreover, in order to better train the DNN, the entire DNN is again trained by a back-propagation algorithm after pretraining. Finally, assuming that the relationship between HR/LR multispectral (MS) image patches is the same as that between HR/LR PAN image patches, the HR MS image will be reconstructed from the observed LR MS image using the trained DNN. Comparative experimental results with several quality assessment indexes show that the proposed method outperforms other pan-sharpening methods in terms of visual perception and numerical measures.**

***Index Terms*—Deep neural networks (DNNs), multispectral (MS) image, panchromatic (PAN) image, pan-sharpening.**

I. Introduction

EARTH observation satellites such as IKONOS and QuickBird provide two different types of images: a panchromatic (PAN) image with high spatial and low spectral resolutions and a multispectral (MS) image with high spectral and low spatial resolutions. Due to technological limitations of current satellite sensor, it is very difficult to acquire a high spatial resolution MS image directly. As a postprocessing method, pan sharpening can be employed to produce a high spatial resolution MS image by fusing the information of the PAN and MS images. The fusing process has become a key preprocessing step in many remote sensing applications such as feature detection and land-cover classification [1].

During the last decades, various pan-sharpening methods have been proposed to address the problem of remote sensing image fusion. Initial methods are the case of the component-substitution methods, which mainly include the intensity–hue–saturation (IHS) [2], [3], the principal component analysis (PCA) [4], and the Gram–Schmidt (GS) transform [5]-based methods. These methods can achieve high spatial resolution but severe spectral distortion. On the contrary, multi-resolution-analysis-based wavelet transform methods [6] can preserve good spectral information. The most famous one among them is the $\grave{a}$ trous wavelet transform (ATWT) method [7], which is simple but robust. However, these methods may suffer from significant spatial distortion.

Compressive sensing (CS)-based pan-sharpening methods have gained general acceptance in recent years. Li and Yang [8] first proposed a CS-based pan-sharpening method, which has achieved a great success. The shortcoming of the method is that it needs plenty of high-resolution (HR) MS training images, which may be nonavailable. To deal with this problem, Jiang *et al.* [9] constructed a joint dictionary from upsampled low-resolution (LR) MS and HR PAN training images. However, they still need to collect numerous LR MS and HR PAN image pairs as the training set. Li *et al.* [10] proposed a pan-sharpening method over learned dictionary without a training set. However, the three dictionaries for PAN and HR/LR MS images must be constructed, which will lead to expensive computation. Zhu and Bamler [11] proposed a new pan-sharpening method named SparseFI, which explores the sparse representation of MS image patches in a dictionary trained only from the PAN image without training images. These methods have difficulty in choosing dictionary atoms when the structural information is weak or lost. To overcoming this problem, a two-step sparse coding pan-sharpening method is proposed [12]. Because the CS-based methods assumed that a sparse signal can be represented as a linear combination of a few atoms in an overcomplete dictionary, they only share a shallow linear structure.

Recent research has shown that the nonlinear deep neural networks (DNNs) have significant great representational power for complex structures and have obtained superior performance in the field of image processing. For example, the great success had been achieved in image denoising and blind inpainting by combining sparse coding and DNN pretrained with denoising autoencoders (DAs) [13]. Based on [13], Agostinelli *et al.* [14] presented the adaptive multicolumn stacked sparse DA (SDA) method, which can achieve state-of-the-art denoising performance with a single system on a variety of different noise types.

Manuscript received September 25, 2014; revised November 4, 2014; accepted November 21, 2014. This work was supported in part by the National Natural Science Foundation of China under Grant 11431015, Grant 61301215, Grant 61101194, and Grant 61171165; by the National Scientific Equipment Developing Project of China under Grant 2012YQ050250; and by the Jiangsu Provincial Postdoctoral Research Funding Plan of China under Grant 1301025C.

W. Huang, L. Xiao, Z. Wei, and S. Tang are with the School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China (e-mail: hnhw235@163.com; xiaoliang@mail.njust.edu.cn; gswei@mail.njust.edu.cn; ts198708@163.com).

H. Liu is with the School of Science, Nanjing University of Science and Technology, Nanjing 210094, China (e-mail: hyliu@njust.edu.cn).

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

Digital Object Identifier 10.1109/LGRS.2014.2376034

In view of this, a DNN-based new pan-sharpening method is proposed for remote sensing image fusion problem. To the best of our knowledge, the proposed method is the first work in this field. The main contribution of this letter is twofold. First, different from the CS-based methods aforementioned, we model the relationship between the HR/LR image patch pairs as a nonlinear mapping defined by the feedforward functions. Second, we propose a DNN learning framework for pan sharpening with pretraining and fine-tuning stages. In the pretraining stage, a stacked modified SDA (S-MSDA) is presented to train the feedforward functions of each layer in turn. Meanwhile, in the fine-tuning stage, the entire DNN is trained again using a back-propagation algorithm.

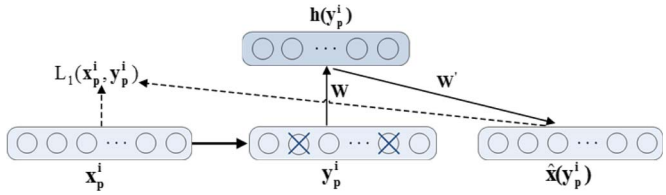

Fig. 1. Modified SDAs' (MSDAs) architecture. The HR/LR image patches $\mathbf{x_p^i}$ and $\mathbf{y_p^i}$ can be seen as the clean data and corrupted data, respectively. MSDA aims to map $\mathbf{y_p^i}$ to hidden layer $\mathbf{h}(\mathbf{y_p^i})$ via encoder $\mathbf{W}$ and reconstruct the HR image patch $\hat{\mathbf{x}}(\mathbf{y_p^i})$ by minimizing the cost function $\mathbf{L_1}(\mathbf{x_p^i}, \mathbf{y_p^i})$ via decoder $\mathbf{W}'$. Note that symbols ◯ and ⊗ denote the clean data and corrupted data, respectively.

II. Pan-Sharpening Method With DNN

Pan-sharpening aims to reconstruct the HR MS image by improving the spatial resolution of the LR MS image while preserving the spectral information with the help of the HR PAN image. The proposed DNN-based method can effectively reconstruct HR MS image with high spatial resolution and less spectral distortions. It mainly consists of three operations, namely, patch extraction for generating training set, DNN training using the MSDA algorithm and S-MSDA architecture, and HR MS image reconstruction.

In this letter, given are the HR PAN image X_{pan} and LR MS image Z_{ms}. First, we upsample the LR MS image Z_{ms} to the size of X_{pan}, resulting in the coregistered LR MS image Y_{ms}, and we normalize X_{pan} and each band of Y_{ms} to the range [0,1]. Moreover, the LR PAN image Y_{pan} is obtained by a linear combination of each band of the LR MS image Y_{ms}. The HR image patches $\{\mathbf{x_p^i}\}_{i=1}^N$ and the corresponding LR image patches $\{\mathbf{y_p^i}\}_{i=1}^N$ are extracted from X_{pan} and Y_{pan}, respectively, generating the training set $\{\mathbf{x_p^i}, \mathbf{y_p^i}\}_{i=1}^N$, where N is the number of training image patches. Then, the DNN is trained using the MSDA algorithm and S-MSDA architecture with the help of the training set $\{\mathbf{x_p^i}, \mathbf{y_p^i}\}_{i=1}^N$. In the reconstruction stage, each band of the LR MS image Y_P is divided into LR MS image patches $\{\mathbf{y_k^j}\}$ with overlapping, where $\mathbf{k}$ represents the kth band, and $\mathbf{j}$ stands for the jth image patch. The HR MS image patches $\{\hat{\mathbf{x}}_\mathbf{k}^\mathbf{j}\}$ are obtained by feedforward of the test LR MS image patch $\{\mathbf{y_k^j}\}$ using the trained DNN. Finally, the sharpened MS image $\hat{X}_{\mathrm{ms}}$ is reconstructed by the overlapping image patches in all individual bands.

A. MSDA Algorithm

Denoising autoencoder (DA) algorithm is usually employed as a layerwise greedy unsupervised pretraining principle to train the DNN, avoiding the problem of "diffusion of gradients" caused by randomly initializing the DNN [15]. Combining sparse coding and DA, an SDA-based approach has gained a great success in image denoising and blind inpainting, in which the DNN is trained to reconstruct a clean "repaired" input from a corrupted version of it [13]. For pan sharpening, this letter proposes an MSDA algorithm for pretraining each layer of the DNN, which represents the relationship between HR image patches $\{\mathbf{x_p^i}\}_{i=1}^N$ (as clean data) and the corresponding LR image patches $\{\mathbf{y_p^i}\}_{i=1}^N$ (as corrupted data), as illustrated in Fig. 1.

The pretraining of MSDA consists in finding a value of parameter vector Θ via minimizing reconstruction error of the cost function. To be specific, let image patch pairs $\{\mathbf{x_p^i}, \mathbf{y_p^i}\}_{i=1}^N$ be training example, parameter vector Θ is trained according to Fig. 1. For each LR image patch $\mathbf{y_p^i}$, the output of the trained DNN will produce an HR image patch $\hat{\mathbf{x}}(\mathbf{y_p^i})$, which will get as close as possible to the corresponding HR PAN image patch $\mathbf{x_p^i}$.

More formally, let $\{\mathbf{x_p^i}\}_{i=1}^N$ be the clean data and $\{\mathbf{y_p^i}\}_{i=1}^N$ be the corrupted data; the feedforward functions of MSDA, including the encoder and the decoder, are defined as follows:

$$\mathbf{h}(\mathbf{y_p^i}) = s(\mathbf{W}\mathbf{y_p^i} + b) \tag{1}$$

$$\hat{\mathbf{x}}(\mathbf{y_p^i}) = s(\mathbf{W}'\mathbf{h}(\mathbf{y_p^i}) + b') \tag{2}$$

where $s(x) = (1 + \exp(-x))^{-1}$ is the sigmoid activation function, $\mathbf{W}$ and b are the encoding weights and biases, $\mathbf{W}'$ and b' are the decoding weights and biases,[1] $\mathbf{h}(\mathbf{y_p^i})$ is the hidden layer's activation,[2] and $\hat{\mathbf{x}}(\mathbf{y_p^i})$ is the reconstruction of the input, which is approximations of $\mathbf{x_p^i}$.

Then, parameter vector $\Theta = \{\mathbf{W}, \mathbf{W}', b, b'\}$ is trained by minimizing the following cost function:

$$L_1\left(\{\mathbf{x_p^i}, \mathbf{y_p^i}\}_{i=1}^N; \Theta\right) = \frac{1}{N}\sum_{i=1}^{N}\left\|\mathbf{x_p^i} - \hat{\mathbf{x}}(\mathbf{y_p^i})\right\|_2^2 + \frac{\lambda}{2}\left(\|\mathbf{W}\|_F^2 + \|\mathbf{W}'\|_F^2\right) + \beta\mathrm{KL}(\hat{\rho} \parallel \rho) \tag{3}$$

where λ and β are balancing parameters determined by cross validation, the second term is a weight decay term, and the third term $\mathrm{KL}(\hat{\rho} \parallel \rho)$ is a sparsity term, which is defined as

$$\mathrm{KL}(\hat{\rho} \parallel \rho) = \rho \log \frac{\rho}{\hat{\rho}} + (1-\rho)\log\frac{1-\rho}{1-\hat{\rho}} \tag{4}$$

[1]For rendering the parameterizations identical, we define $\mathbf{W}' = \mathbf{W}^T$.

[2]The dimension of the hidden layers $\mathbf{d_h}$ can be either smaller or larger than the input dimension $\mathbf{d_{x_p^i}}$. If $\mathbf{d_h} < \mathbf{d_{x_p^i}}$, it can be seen as a dimensionality reduction technique like PCA. If $\mathbf{d_h} > \mathbf{d_{x_p^i}}$, it can be seen as overcomplete representations like sparse coding.

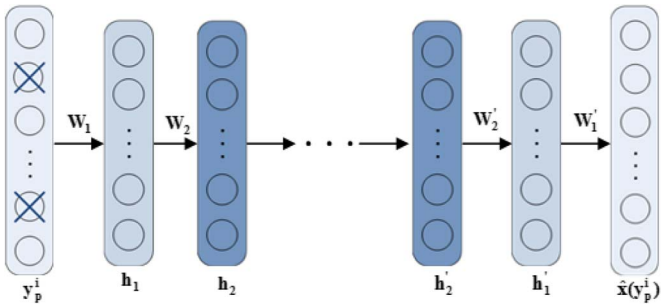

Fig. 2. S-MSDAs' architecture.

where $\hat{\rho} = (1/N)\sum_{i=1}^{N} \mathbf{h}(\mathbf{y_P^i})$ is the average activation of the hidden layer, and ρ is a sparsity parameter, typically a small value close to zero. $\mathrm{KL}(\hat{\rho} \parallel \rho)$ is the Kullback–Leibler divergence, which forces $\hat{\rho}$ to be close to ρ. Hence, the activations of the hidden units must mostly be near zero and achieve sparsity. In this letter, the back-propagation algorithm is employed to obtain the parameter vector Θ according to the loss rule in (3).

B. S-MSDA to Construct Deep Architecture

In order to construct a deep network architecture, a series of MSDAs are connected to yield S-MSDAs. The S-MSDA is a greedy layerwise approach to pretrain all the parameter vectors of the DNN, which works by training each layer in turn. After training an MSDA, the successive layer is trained by using the hidden activation values $\mathbf{h}(\mathbf{y_p^i})$ and $\mathbf{h}(\mathbf{x_p^i})$ of the previous layer as the input data of this layer. The S-MSDA architecture (also named as DNN architecture) is shown in Fig. 2.

To produce better pan-sharpened results, after the pretraining stage, the entire DNN is trained again by the back-propagation algorithm in the fine-tuning stage, minimizing the following objective:

$$L_2\left(\{\mathbf{x_p^i}, \mathbf{y_p^i}\}_{i=1}^{N}; \Theta\right) = \frac{1}{N}\sum_{i=1}^{N}\left\|\mathbf{x_p^i} - \hat{\mathbf{x}}\left(\mathbf{y_p^i}\right)\right\|_2^2 + \frac{\lambda}{2}\sum_{l=1}^{L}\|\mathbf{W}_l\|_F^2 \quad (5)$$

where L is the number of stacked MSDAs, and $\mathbf{W}_l$ represents the parameter of the lth layer in the DNN. The sparsity term is removed for this step because the sparsity has been already utilized in the pretraining stage [13].

C. HR MS Image Reconstruction

In this letter, we assume that the relationship between HR/LR PAN image patches is the same as that between HR/LR MS image patches. After training the DNN, the trained DNN will reconstruct the HR MS image from the observed LR MS image well. For each LR MS image patch $\mathbf{y_k^j}$, the trained DNN will be used to reconstruct the corresponding HR MS image patch. Specifically, let $\mathbf{y_k^j}$ be the input data of the trained DNN, the sharpened HR MS image patches $\hat{\mathbf{x}}_\mathbf{k}^\mathbf{j}$ will be obtained according to the feedforward functions in (1) and (2). Then, the sharpened MS image $\hat{X}_{\mathrm{ms}}$ can be reconstructed by averaging the overlapping image patches $\hat{\mathbf{x}}_\mathbf{k}^\mathbf{j}$ in all individual bands. Finally, to better reconstruct the spatial details of sharpened MS image, a residual compensation method [16] is employed to enhance the high-frequency spatial details of each band of the reconstructed HR MS image individually.

III. Experimental Results and Analysis

Wald *et al.* [17] have shown that any synthetic image should be as identical as possible to the image that the corresponding sensor would observe with the highest resolution. According to this synthesis property, the original images are filtered by a 7×7 Gaussian kernel with standard deviation 1 and downsampled by a factor of 4. We conduct the experiments on the degraded images, and the original images are seen as the reference images. In this letter, the IKONOS and QuickBird data sets are employed to test the performance of the proposed method, and the following five quality metrics are used to quantitatively assess the quality of the results: correlation coefficient (CC), root-mean-squared error (RMSE), *erreur relative global adimensionnelle de synthèse* (ERGAS), spectral angle mapper (SAM), and Q4. The $\mathrm{CC_{AVG}}$ and $\mathrm{RMSE_{AVG}}$ represent the average values of the CC and RMSE among all bands of the MS images, respectively.

For the proposed method, the S-MSDA constructs a deep architecture consisting of L hidden layers. Here, in order to reduce computation load, L is set to 2 for all the data sets. The dimension of hidden layers is set five times the dimension of the input data. We evaluate the hyperparameter combinations on different data sets and report the best result. The proposed method is compared with three popular traditional methods, i.e., GS method [5] (implemented with ENVI 4.5), ATWT method [7], adaptive IHS method [3], and one state-of-the-art SparseFI method [11].

A. Experiments on IKONOS Data Set

The IKONOS satellite collects a PAN image with 1-m resolution and an MS image with 4-m resolution and four channels (i.e., red, green, blue, and near infrared). The scene is captured over Sichuan, China, in May 2008. In order to quantitatively assess the quality of fusion results, we filter and downsample by a factor of 4 the PAN and MS images to obtain a PAN image with 4-m resolution and an MS image with 16-m resolution. Then, pan-sharpening methods are used to fuse the degraded PAN and MS images to yield the HR MS image with 4-m resolution. Finally, the fused HR MS image is compared with the original HR MS image. In this experiment, both sizes of PAN and upsampled LR MS images are 600×600. In the training stage, we randomly sample 200 000 image patch pairs $\{\mathbf{x_p^i}, \mathbf{y_p^i}\}$ with the same size of 7×7. In the reconstruction stage, each band of the upsampled LR MS image is divided into image patches $\mathbf{y_k^j}$ (of size 7×7) with overlapping.

The visual results of five different methods are shown in Fig. 3(c)–(g). Fig. 3(a) and (b) shows the resampled color LR MS image consisting of three channels (i.e., red, green, and blue) and PAN image, respectively. Fig. 3(h) shows an original color HR MS image. We can obviously see that Fig. 3(c) reconstructed by the GS method has good spatial information but severe color distortion. The fused image using the ATWT

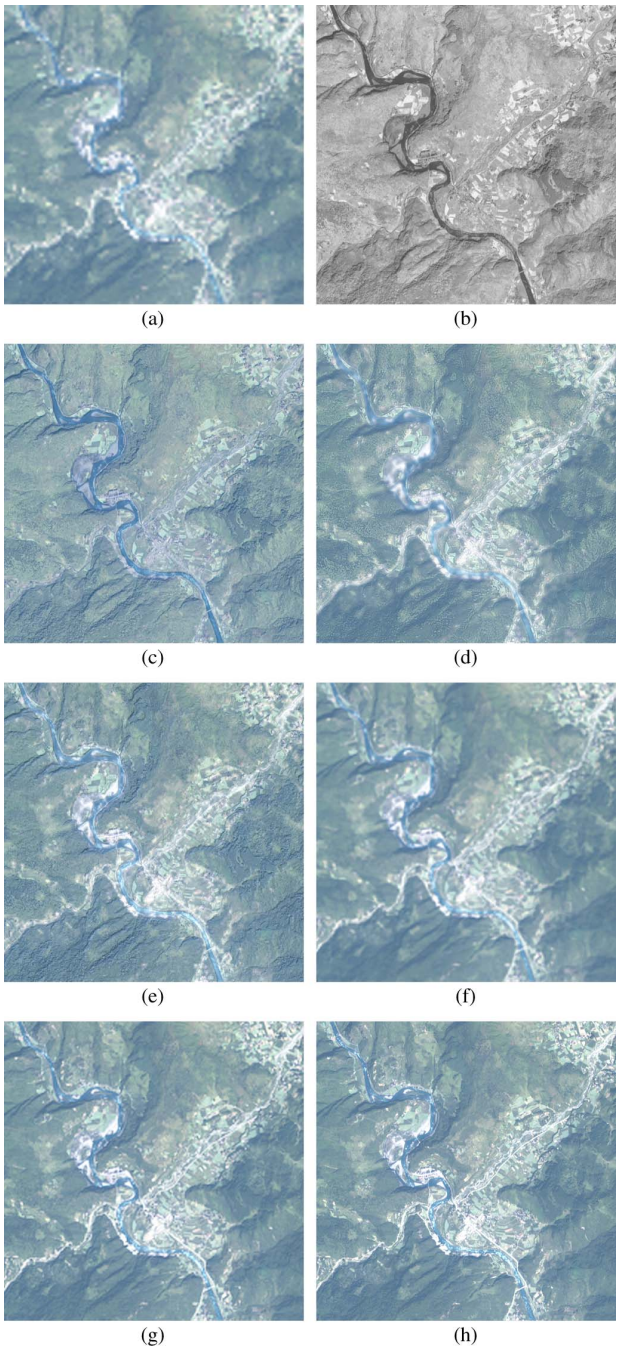

Fig. 3. IKONOS images and experimental results by different methods. (a) Resampled LR MS image. (b) PAN image. (c) GS method. (d) ATWT method. (e) Adaptive IHS method. (f) SparseFI method. (g) Proposed method. (h) Original MS image.

method in Fig. 3(d) has good color information, which reflects good performance in preserving the spectral information. However, the fused image suffers from significant spatial distortions. Fig. 3(e) reconstructed by the adaptive IHS method gains better performance on a tradeoff between the spatial and spectral information than Fig. 3(c) and (d). However, compared with the original MS image, Fig. 3(e) has the great difference in spatial details. The linear sparse representation-based SparseFI method, as shown in Fig. 3(f), preserves the spectral properties better than the above methods, but it shows poor spatial results in fine detail. On the whole, the proposed method shown in Fig. 3(g) not only shows finer spatial details but also preserves the spectral information best of all the methods.

TABLE I
COMPARISON OF THE PROPOSED METHOD WITH OTHER METHODS ON IKONOS DATA SHOWN IN FIG. 3

	GS	ATWT	Adaptive IHS	SparseFI	Proposed
CC_{AVG}	0.7924	0.8819	0.8828	0.9096	**0.9230**
$RMSE_{AVG}$	0.0061	0.0037	0.0035	0.0033	**0.0028**
ERGAS	3.1983	2.4858	2.3943	2.2793	**2.1150**
SAM	0.0653	0.0520	0.0558	0.0456	**0.0431**
Q4	0.7627	0.8211	0.8248	0.7918	**0.8395**

The qualitative assessments of fused images in Fig. 3 are shown in Table I, in which the best results for each criterion are labeled in bold, and the second best are underlined. In Table I, we can see that the proposed method performs best.

B. Experiments on QuickBird Data Set

To further verify the effectiveness of the proposed method, we move on to conduct the experiments on the QuickBird satellite data set. The QuickBird satellite provides a PAN image with 0.7-m resolution and an MS image with 2.8-m resolution and four channels (i.e., red, green, blue, and near-infrared). The scene captures over North Island, New Zealand, in August 2012. In order to quantitatively assess the quality of the fusion results, we filter and downsample by a factor of 4 the PAN and MS images to obtain a PAN image with 2.8-m resolution and an MS image with 11.2-m resolution. Then, pan-sharpening methods are used to fuse the degraded PAN and MS images to yield the HR MS image with 2.8-m resolution. Finally, the fused HR MS image is compared with the original HR MS images. In this experiment, the size of PAN image and the number of image patch pairs are the same as those on the IKONOS data set.

Fig. 4(a) shows a resampled color LR MS image. Fig. 4(b) gives the corresponding HR PAN image. The original HR MS image is shown in Fig. 4(h). As shown in Fig. 4(c)–(g), they are reconstructed by GS method, ATWT method, adaptive IHS method, SparseFI method, and the proposed method, respectively. Compared with the results in Fig. 3, the similar conclusions can be drawn from Fig. 4.

The qualitative assessments of fused images in Fig. 4 are shown in Table II, in which the best results for each criterion are labeled in bold, and the second best are underlined. Similar to Table I, the proposed method performs best in all quality indexes. Therefore, the proposed method produces the best fused results and performs the robustness for both the IKONOS and QuickBird data sets.

IV. CONCLUSION

In this letter, we have proposed a DNN-based new pan-sharpening method for remote sensing image fusion problem. An MSDA algorithm is presented to train the relationship between the HR/LR image patches. Then, an S-MSDA architecture that is a greedy layerwise approach to pretrain all the parameter vectors of the DNN is obtained by connecting

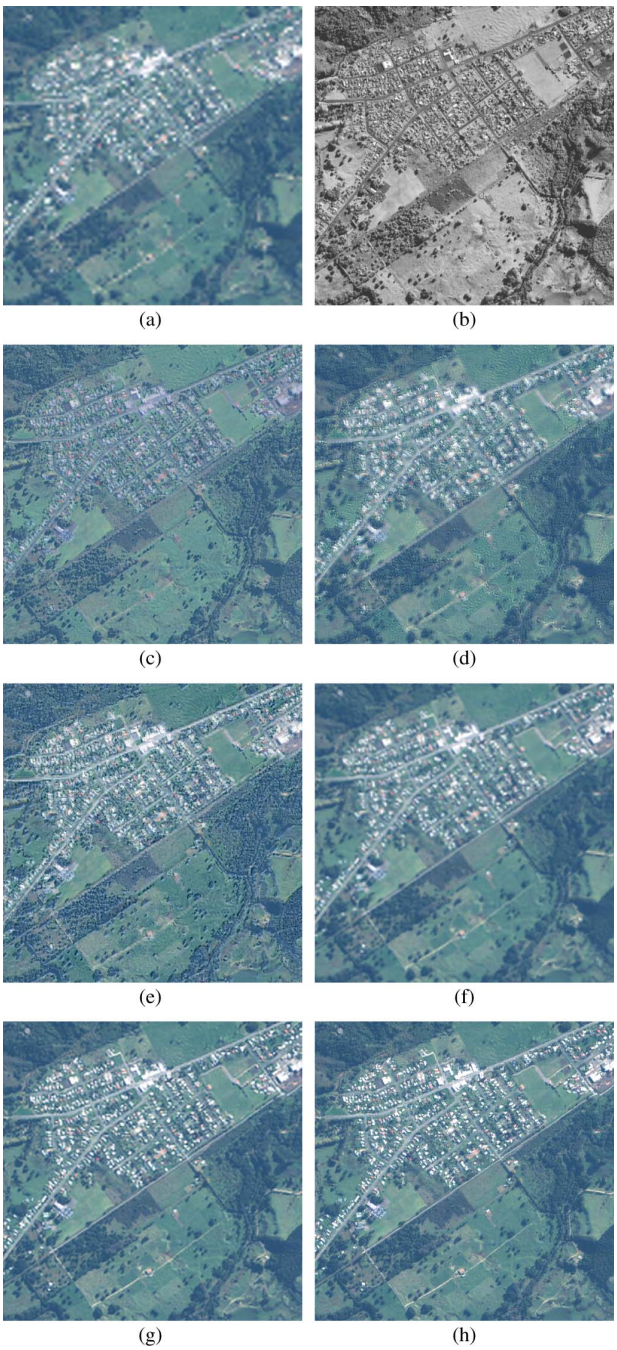

Fig. 4. QuickBird images and experimental results by different methods. (a) Resampled LR MS image. (b) PAN image. (c) GS method. (d) ATWT method. (e) Adaptive IHS method. (f) SparseFI method. (g) Proposed method. (h) Original MS image.

TABLE II
COMPARISON OF THE PROPOSED METHOD WITH OTHER METHODS ON QUICKBIRD DATA SHOWN IN FIG. 4

	GS	ATWT	Adaptive IHS	SparseFI	Proposed
CC_{AVG}	0.8148	0.8412	0.8273	0.8792	**0.8904**
$RMSE_{AVG}$	0.0079	0.0070	0.0070	0.0064	**0.0059**
ERGAS	4.5807	4.3198	4.3114	4.0230	**3.8141**
SAM	0.0984	0.0914	0.1063	0.0833	**0.0780**
Q4	0.8128	0.8223	0.8239	0.7995	**0.8376**

a series of MSDA architecture. In the fine-tuning stages, the parameter vectors of the DNN are trained again by a back-propagation algorithm. After training, the DNN will be able to reconstruct the HR MS image from the observed LR MS image. The experimental results have demonstrated that the proposed method can achieve better performance than other traditional and state-of-the-art pan-sharpening methods.

However, the proposed method takes more time than traditional methods. Compared with the SparseFI method, the proposed method is much faster in the reconstruction stage because it only needs forward propagation according to the feedforward functions without solving any optimization problem, whereas the training stage of the proposed method requires much more time. In our future work, offline training and the many-core-based parallel algorithms will be adopted to speed up the training process of the DNN.

REFERENCES

[1] C. Thomas, T. Ranchin, L. Wald, and J. Chanussot, "Synthesis of multispectral images to high spatial resolution: A critical review of fusion methods based on remote sensing physics," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1301–1312, May 2008.

[2] T.-M. Tu, S.-C. Su, H.-C. Shyu, and P. S. Huang, "A new look at IHS-like image fusion methods," *Inf. Fusion*, vol. 2, no. 3, pp. 177–186, Sep. 2001.

[3] S. Rahmani, M. Strait, D. Merkurjev, M. Moeller, and T. Wittman, "An adaptive IHS pan-sharpening method," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 4, pp. 746–750, Oct. 2010.

[4] V. P. Shah, N. H. Younan, and R. L. King, "An efficient pan-sharpening method via a combined adaptive PCA approach and contourlets," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1323–1335, May 2008.

[5] B. V. Brower and C. A. Laben, "Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening," U.S. Patent 6 011 875, Jan. 4, 2000.

[6] K. Amolins, Y. Zhang, and P. Dare, "Wavelet based image fusion techniques: An introduction, review and comparison," *ISPRS J. Photogramm. Remote Sens.*, vol. 62, no. 4, pp. 249–263, Sep. 2007.

[7] J. Nunez *et al.*, "Multiresolution-based image fusion with additive wavelet decomposition," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1204–1211, May 1999.

[8] S. Li and B. Yang, "A new pan-sharpening method using a compressed sensing technique," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 2, pp. 738–746, Feb. 2011.

[9] C. Jiang, H. Zhang, H. Shen, and L. Zhang, "A practical compressed sensing-based pan-sharpening method," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 4, pp. 629–633, Jul. 2012.

[10] S. Li, H. Yin, and L. Fang, "Remote sensing image fusion via sparse representations over learned dictionaries," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 9, pp. 4779–4789, Sep. 2013.

[11] X. X. Zhu and R. Bamler, "A sparse image fusion algorithm with application to pan-sharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 2827–2836, May 2013.

[12] C. Jiang, H. Zhang, H. Shen, and L. Zhang, "Two-step sparse coding for the pan-sharpening of remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 5, pp. 1939–1404, May 2014.

[13] J. Xie, L. Xu, and E. Chen, "Image denoising and inpainting with deep neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 350–358.

[14] F. Agostinelli, M. R. Anderson, and H. Lee, "Robust image denoising with multi-column deep neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2013, pp. 1493–1501.

[15] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *J. Mach. Learn. Res.*, vol. 11, pp. 3371–3408, 2010.

[16] S. Yang, M. Wang, Y. Chen, and Y. Sun, "Single-image super-resolution reconstruction via learned geometric dictionaries and clustered sparse coding," *IEEE Trans. Image Process.*, vol. 21, no. 9, pp. 4016–4028, Sep. 2012.

[17] L. Wald *et al.*, "Fusion of satellite images of different spatial resolutions: assessing the quality of resulting images," *Photogramm. Eng. Remote Sens.*, vol. 63, no. 6, pp. 691–699, Jun. 1997.