Boosting the Accuracy of Multispectral Image Pansharpening by Learning a Deep Residual Network

Yancong Wei, *Student Member, IEEE*, Qiangqiang Yuan, *Member, IEEE*, Huanfeng Shen, *Senior Member, IEEE*, and Liangpei Zhang, *Senior Member, IEEE*

***Abstract*— In the field of multispectral (MS) and panchromatic image fusion (pansharpening), the impressive effectiveness of deep neural networks has recently been employed to overcome the drawbacks of the traditional linear models and boost the fusion accuracy. However, the existing methods are mainly based on simple and flat networks with relatively shallow architectures, which severely limits their performance. In this letter, the concept of residual learning is introduced to form a very deep convolutional neural network to make the full use of the high nonlinearity of the deep learning models. Through both quantitative and visual assessments on a large number of high-quality MS images from various sources, it is confirmed that the proposed model is superior to all the mainstream algorithms included in the comparison, and achieves the highest spatial–spectral unified accuracy.**

***Index Terms*— Convolutional neural network, data fusion, pansharpening, remote sensing, residual learning.**

I. Introduction

PANSHARPENING is a fundamental and significant task in the field of remote-sensing data fusion, in which the spatial details from panchromatic (PAN) images and rich spectral information from multispectral (MS) or hyperspectral (HS) images are fused to yield imagery with a high resolution in both the spatial and spectral domains. In this letter, we focus on the fusion of PAN and MS images.

Traditional pan-sharpening algorithms can be divided into three major branches: 1) component substitution [1], [2]; 2) detail injection [3], [4]; and 3) regularization-model-based methods [5], [6]. In the former two branches, although the spatial details are impressively sharpened, obvious distortions are easily caused in the spectral domain, which severely degrades the quality of the fused image. The branch of regularization-model-based methods describe the whole fusion process as linear functions with strict constraints, which are usually based on prior knowledge or reasonable assumptions of the images included in the fusion process, such as Laplacian prior [5] and sparse coding [6]. Recently, linear regression methods based on geostatistics theory are also proposed for image fusion [7], [8]. However, the performance of such models is limited by their linearity, which cannot accurately describe a fusion process that contains complex transformations in both the spatial and spectral domains. Furthermore, the reliance on prior constraints can also cause severe quality degradation in cases where the prior knowledge does not fit the problem.

In this letter, inspired by the impressive performance of deep learning in the field of computer vision, we propose the deep residual pan-sharpening neural network (DRPNN) to overcome the drawbacks of the previously proposed methods and perform high-quality fusion of PAN and MS images. The prototype of the DRPNN is introduced from a deep residual network for image super-resolution [9], and we make a specific improvement of its architecture to fit it to the task of image pansharpening. Supported by the residual learning architecture, an extremely deep convolutional filtering framework is formed to improve the accuracy of the fusion, while the learning process of the filtering parameters is also guaranteed to converge quickly.

The rest of this letter is organized as follows. Background knowledge about pansharpening and the superiority of deep learning is provided in Section II. A detailed description of the proposed method is given in Section III. The results of the experiments are provided and discussed in Section IV. Finally, the conclusion is drawn in Section V.

II. Background

A. Multispectral Image Pansharpening

We denote the PAN image as $\boldsymbol{g}_{\mathrm{PAN}}$ (size: $H \times W$) and the MS image with S spectral bands as $\boldsymbol{g}_{\mathrm{MS}}$ (size: $H/\mathrm{scale} \times W/\mathrm{scale} \times S$). The aim of pansharpening is to yield an image with high resolution in both spatial and spectral domains, and we denote it as $\boldsymbol{f}_{\mathrm{MS}}$ (size: $H \times W \times S$). The fusion process can be regarded as a guided super-resolution problem, which means that the ill-posed property of the inverse prediction

Manuscript received May 28, 2017; revised July 22, 2017; accepted July 31, 2017. Date of publication August 17, 2017; date of current version September 25, 2017. This work was supported in part by the National Natural Science Foundation of China under Grant 41401383 and Grant 41422108, in part by the Fundamental Research Funds for the Central Universities under Grant 2042017kf0180, and in part by the Natural Science Foundation of Hubei Province under Grant ZRMS2016000241. *(Corresponding author: Qiangqiang Yuan.)*

Y. Wei is with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan 430079, China (e-mail: 770276733@qq.com).

Q. Yuan is with the School of Geodesy and Geomatics, Collaborative Innovation Center of Geospatial Technology, Wuhan University, Wuhan 430079, China (e-mail: yqiang86@gmail.com).

H. Shen is with the School of Resource and Environmental Science, Collaborative Innovation Center of Geospatial Technology, Wuhan University, Wuhan 430079, China (e-mail: shenhf@whu.edu.cn).

L. Zhang is with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, the Collaborative Innovation Center of Geospatial Technology, Wuhan University, Wuhan 430079, China (e-mail: zlp62@whu.edu.cn).

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

Digital Object Identifier 10.1109/LGRS.2017.2736020

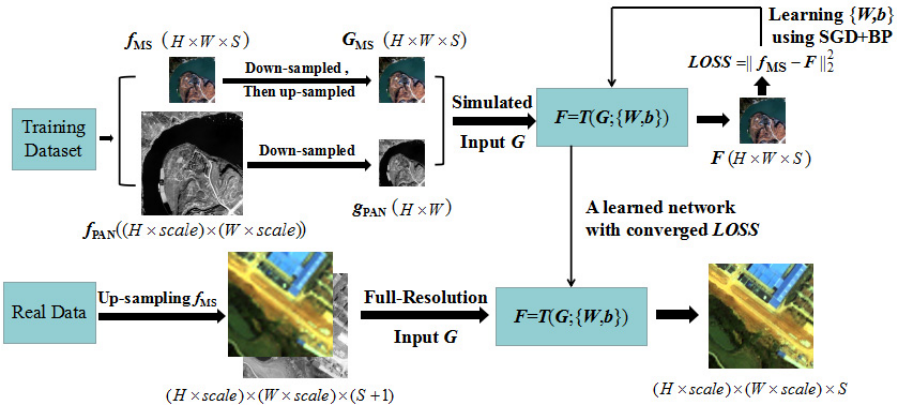

Fig. 1. Framework of pansharpening based on learning a deep neural network $F = T(G; \{W, b\})$.

problem is reduced with $\boldsymbol{g}_{\mathrm{PAN}}$ included, compared to the traditional single-image super-resolution problem [9]. The main difficulty of pansharpening comes from the spectral domain, as the bandwidths covered by the PAN and MS channels are not guaranteed to fully overlap among the various types of sensor, e.g., WorldView-2 (PAN: 400–1040 nm; MS: continuously covering 450–800 nm) and IKONOS (PAN: 526–929 nm; MS: discretely covering 445–853 nm). Thus, to preserve the spectral fidelity of images fused from such observations, the fusion process in the spectral domain is very complex and needs to be simulated using highly nonlinear functions.

B. Pansharpening Based on Deep Learning

To break the limitation of the linear models discussed above, deep neural networks have recently been employed to perform prediction and yield images with state-of-the-art accuracy, relying on the nonlinearity of the mapping process through the deep networks. Similar to the successful applications for the single-image super-resolution problem [9], [10], deep learning models have also been introduced to the field of image fusion [11], [12]. By using bicubic interpolation to coarsely up-sample $\boldsymbol{g}_{\mathrm{MS}}$ to $\boldsymbol{G}_{\mathrm{MS}}(H \times W \times S)$, we obtain an initialized input $\boldsymbol{G} = \{\boldsymbol{G}_{\mathrm{MS}}, \boldsymbol{g}_{\mathrm{PAN}}\}(H \times W \times (S+1))$ for the pan-sharpening task. The high-resolution MS image can be reconstructed by extracting the low-frequency features from $\boldsymbol{G}_{\mathrm{MS}}$ and the high-frequency features from $\boldsymbol{g}_{\mathrm{PAN}}$, and then merging them to form the final estimation. Thus, from the perspective of deep learning, we can summarize the whole process as a filtering function with high nonlinearity. In addition, the requirements can be met well by the nature of the deep neural network, in which multiple linear filtering layers are stacked to form a highly nonlinear transformation, and the optimal allocations for all the parameters can be automatically searched for to minimize the prediction loss between the output of the network $\boldsymbol{F} = \boldsymbol{T}(\boldsymbol{G})$ and the ground truth $\boldsymbol{f}_{\mathrm{MS}}$. The flowchart of learning a deep network for the pan-sharpening process is shown in Fig. 1.

III. Methodology

A. Convolutional Neural Networks

CNNs are one of the most impressive branches of deep learning models in the field of computer vision. In an end-to-end image restoration task, a CNN built with multiple stacked convolutional layers is expressed as $f \approx F = \boldsymbol{CNN}(G)$ to estimate the high-quality image $\boldsymbol{f}$ from the degraded observation $\boldsymbol{G}$. To train a randomly initialized CNN for pansharpening, down-sampled MS and PAN images are input as $\boldsymbol{G}$, and then an image $\boldsymbol{F}$ with the same size as the original MS image $\boldsymbol{f}_{\mathrm{MS}}$ is produced [12]. In a CNN with L layers, the fusion process is performed via forward passing

$$\boldsymbol{F}_0 = \boldsymbol{G},\ \boldsymbol{F}_l = \max(0, \boldsymbol{W}_l \circ \boldsymbol{F}_{l-1} + \boldsymbol{b}_l),\quad l = 1, \ldots, L-1 \tag{1}$$

$$\boldsymbol{F} = \boldsymbol{W}_L \circ \boldsymbol{F}_{L-1} + \boldsymbol{b}_L \tag{2}$$

where $\boldsymbol{W}$ stands for the 3-D convolutional filters and $\boldsymbol{b}$ represents the bias vectors. With stochastic gradient descent and backpropagation (BP), all the parameters $\{\boldsymbol{W}, \boldsymbol{b}\}$ in the network can be iteratively learned to reach an optimal allocation. The learning process can be summarized as follows:

$$\delta\boldsymbol{\theta}^t = \{\delta\boldsymbol{W}^t, \delta\boldsymbol{b}^t\} = \left\{\frac{\partial\left(||\boldsymbol{f}_{\mathrm{MS}} - \boldsymbol{F}^t||_2^2\right)}{\partial \boldsymbol{W}^t}, \frac{\partial\left(||\boldsymbol{f}_{\mathrm{MS}} - \boldsymbol{F}^t||_2^2\right)}{\partial \boldsymbol{b}^t}\right\} \tag{3}$$

$$\boldsymbol{\theta}^{t+1} = \boldsymbol{\theta}^t + \Delta\boldsymbol{\theta}^t = \boldsymbol{\theta}^t + \mu \cdot \Delta\boldsymbol{\theta}^{t-1} - \varepsilon \cdot \delta\boldsymbol{\theta}^t. \tag{4}$$

B. Deep Residual Learning for Pansharpening

It has been noted that a deeper CNN with more filtering layers tends to extract more abstract and representative features, and thus higher prediction accuracy can be expected. However, due to the gradient vanishing problem, the gradients of the prediction loss to parameters in the shallow layers cannot be smoothly passed via BP [9], [10], [12], which prevents the deep network from being fully learned.

Deep residual learning [13] is an advanced method for solving this problem, in which the transformation $f \approx \boldsymbol{CNN}(\boldsymbol{G})$ is replaced with $f\text{-}G \approx \boldsymbol{RES}(\boldsymbol{G})$ by setting a skip connection between the disconnected layers. It is reasonable to assume that most pixel values in the residual image $\boldsymbol{f}\text{-}\boldsymbol{G}$ are very close to zero, and the spatial distribution of the residual features should be very sparse. Thus, searching for an allocation that is very close to the optimal for $\{\boldsymbol{W}, \boldsymbol{b}\}$ becomes much faster and easier, which allows us to add more layers to the network and boost its performance. However, in the pan-sharpening task, it should be noted that the size of the final output $\boldsymbol{F}(H \times W \times S)$

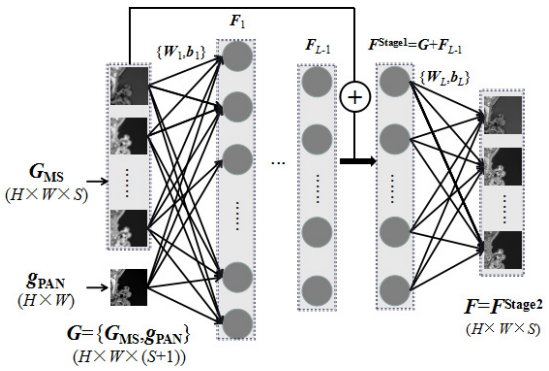

Fig. 2. Flowchart of passing low-resolution MS and PAN images through the DRPNN.

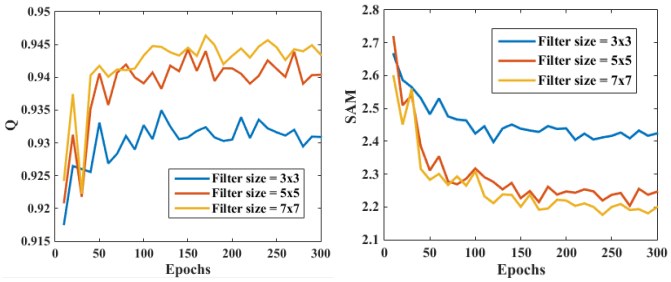

Fig. 3. Q and SAM values of the DRPNN with different filter sizes. The quantitative assessments were undertaken on the data set of 160 QuickBird images.

is not the same as the size of the input $\boldsymbol{G}(H \times W \times (S+1))$; thus, instead of directly predicting the residual image as in [9] and [14], the process through the DRPNN with L layers is divided into two stages.

Stage 1: The first to the $(L-1)$th layers are stacked under a skip connection to estimate the residual between $\boldsymbol{G}$ and $\boldsymbol{F}^{\text{Stage 1}}$ (size: $H \times W \times (S+1)$). The convolutional filtering process in each layer is the same as described in (1). The residual output from the $(L-1)$th layer is then added to $\boldsymbol{G}$ to yield $\boldsymbol{F}^{\text{Stage 1}}$

$$F^{\text{Stage 1}} = G + F_{L-1}. \tag{5}$$

Stage 2: The Lth layer of the DRPNN is set to reduce the spectral dimensionality from $(S+1)$ bands to S bands via the last 3-D convolutional filtering process in the model, yielding a final estimation $\boldsymbol{F}^{\text{Stage 2}}$ $(H \times W \times S)$

$$\boldsymbol{F} = \boldsymbol{F}^{\text{Stage 2}} = \boldsymbol{W}_L \circ \boldsymbol{F}^{\text{Stage 1}} + \boldsymbol{b}_L. \tag{6}$$

The complete architecture of the DRPNN is illustrated in Fig. 2.

IV. Experiments and Discussion

Data Sets: For the training and simulated experiments with the DRPNN, two data sets were collected from QuickBird (51 648 training patches and 160 test patches) and WorldView-2 images (59 840 training patches and 80 test patches). Two networks were then separately learned for MS images with different values of S: one was set with $S = 4$ (QuickBird), and the other was set with $S = 8$ (WorldView-2). In the real-data experiments, for the network with $S = 4$, we collected another data set from a group of high-quality IKONOS images, while the other network was tested on a smaller part of the WorldView-2 data set to yield full-resolution MS images. The huge number of tested images in our collected data sets makes the quantitative results much more convincing.

A. Hyperparameters and Training

The DRPNN proposed in this letter contains $L = 11$ layers, the lth layer of which is filled with C_l groups of filters $\boldsymbol{W}_{l,k}(h_l \times w_l \times C_{l-1})$, where $k = 1, \ldots, C_l$, and one bias vector $\boldsymbol{b}_l(1 \times C_l)$. According to the aimed task, $C_0 = S+1$ and $C_{11} = S$ should first be provided and, similar to the prototype [9], C_l for the other layers is empirically set to 64. For the spatial size $h_l \times w_l$ of the filters, we compared the performances of several values, as shown in Fig. 3. From the results of the comparison, the filter size was set to 7×7 through the whole network.

The training process of each network costs 300 epochs. As the classic momentum algorithm [15] was also applied to accelerate the decrease of the prediction loss, the learning rate ε in (4) was initialized as 0.05 for the first ten layers with the skip connection, and as 0.005 for the last layer, while the momentum μ was fixed at 0.95 for the whole network. After every 60 epochs, the learning rate was multiplied by a descent factor $\gamma = 0.5$. The implementation of the CNN was supported by two deep learning frameworks: Caffe [16] for training and MatConvNet [17] for testing.

B. Quantitative Assessments

The MS and PAN images were down-sampled to simulate a relatively degenerated input $\boldsymbol{G}$. Four metrics—the Q metric, the relative dimensionless global error in synthesis, the spectral angle mapper (SAM), and the spatial correlation coefficient (SCC)—were employed to quantify the accuracy in the spatial and spectral domains, with the original MS image as the ground truth. The performance of the DRPNN was compared with five algorithms from different branches: Gram–Schmidt (GS) pansharpening [1], the generalized Laplacian pyramid with modulation transfer function matched filter (MTF-GLP) [3], smoothing filter-based intensity modulation (SFIM) [4], the adaptive wavelet luminance proportion (AWLP) method [18], and the two-step sparse coding model (TSSC) [6]. In addition to these traditional algorithms, the PNN[12], a flat CNN without residual learning and skip connection, was also included in the comparison. The quantitative results are listed in Table I, where the comparison of the metrics indicates that the DRPNN yields images with the best spatial–spectral unified accuracy.

C. Visual Assessment

Considering that the numeric metrics are mainly employed to quantify the overall accuracy of a predicted image, visual inspection is also required to find any noticeable distortion, which may not be shown in the quantitative assessment. For each value of band number S, one group of simulated fusion results are selected to be displayed as true-color images

TABLE I
QUANTITATIVE RESULTS OF THE SIMULATED EXPERIMENTS

Tested images	Method	Q (↑)	ERGAS (↓)	SAM (↓)	SCC (↑)
Sensor: QuickBird Size: 250×250×4 Total number: 160	GS	0.8305	4.5014	4.0227	0.6090
	MTF-GLP	0.8227	4.4409	3.7893	0.5839
	SFIM	0.8264	5.1491	3.7708	0.5708
	ISTS	0.8498	3.8677	3.6579	0.6157
	TSSC	0.8488	3.9773	3.7154	0.5909
	PNN	0.9206	2.7110	2.6405	0.7951
	DRPNN	**0.9437**	**2.1916**	**2.1936**	**0.8458**
Sensor: WorldView-2 Size: 250×250×8 Total number: 80	GS	0.8606	4.8395	6.1412	0.8190
	MTF-GLP	0.8788	4.3748	5.7698	0.8136
	SFIM	0.8756	4.3230	5.7579	0.8281
	ISTS	0.8720	4.4353	5.8933	0.8126
	TSSC	0.8951	3.9735	5.8269	0.8116
	PNN	0.9389	3.0695	4.4757	0.8674
	DRPNN	**0.9458**	**2.8913**	**4.1998**	**0.8766**

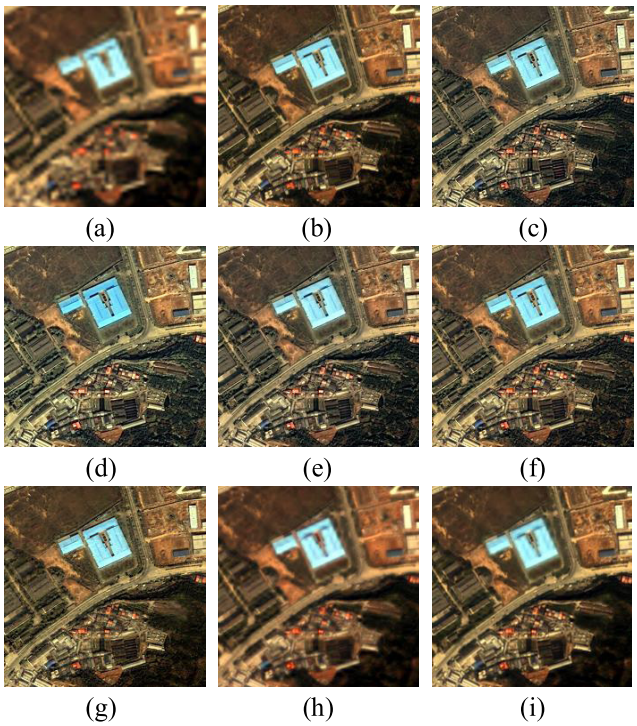

Fig. 4. Simulated fusion results from the QuickBird imagery, Yichang, 2015. (a) Low-resolution MS image simulated by down-sampling. (b) Ground truth. (c) SFIM. (d) GS. (e) MTF-GLP. (f) AWLP. (g) TSSC. (h) PNN. (i) DRPNN.

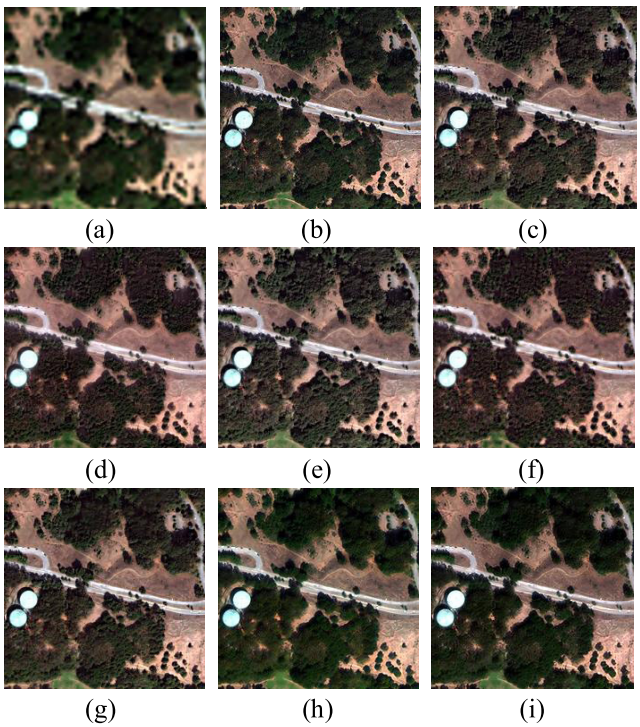

Fig. 5. Simulated fusion results from the WorldView-2 imagery, San Francisco, 2011. (a) Low-resolution MS image simulated by down-sampling. (b) Ground truth. (c) SFIM. (d) GS. (e) MTF-GLP. (f) AWLP. (g) TSSC. (h) PNN. (i) DRPNN.

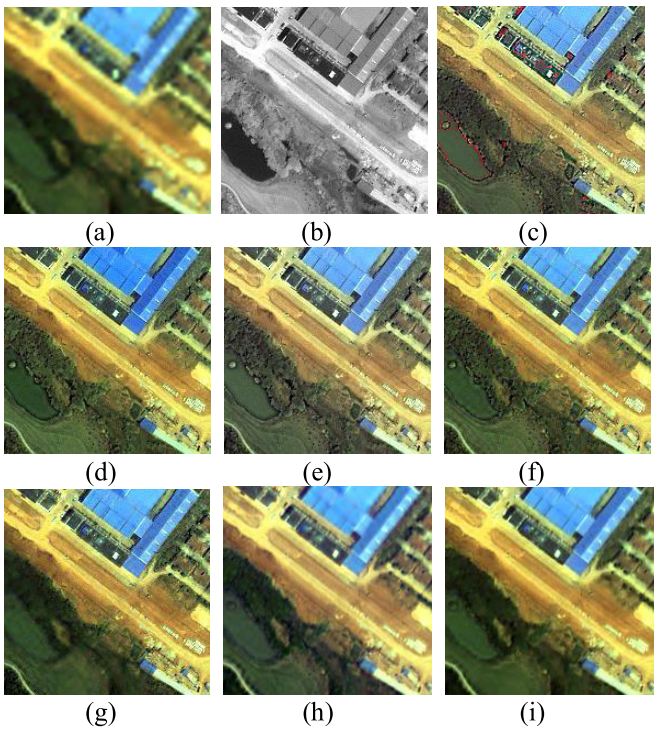

Fig. 6. Full-resolution fusion results from the IKONOS imagery, Wuhan. (a) Up-sampled MS image. (b) PAN. (c) SFIM. (d) GS. (e) MTF-GLP. (f) AWLP. (g) TSSC. (h) PNN. (i) DRPNN.

in Figs. 4 and 5, while some of the results from the real-data experiments are shown in Figs. 6 and 7. By comparing the results, it can be observed that by methods that are not based on deep learning models, sharpened spatial features are achieved, but with severe spectral distortions, such as the industrial area shown in Fig. 4(c)–(g) and the textures of the urban vegetation shown in Fig. 5(c)–(g). Meanwhile, results of the two deep learning-based models are the closest to the ground truth, both in the fusion of the spatial details and in the preservation of spectral fidelity, even for some specific

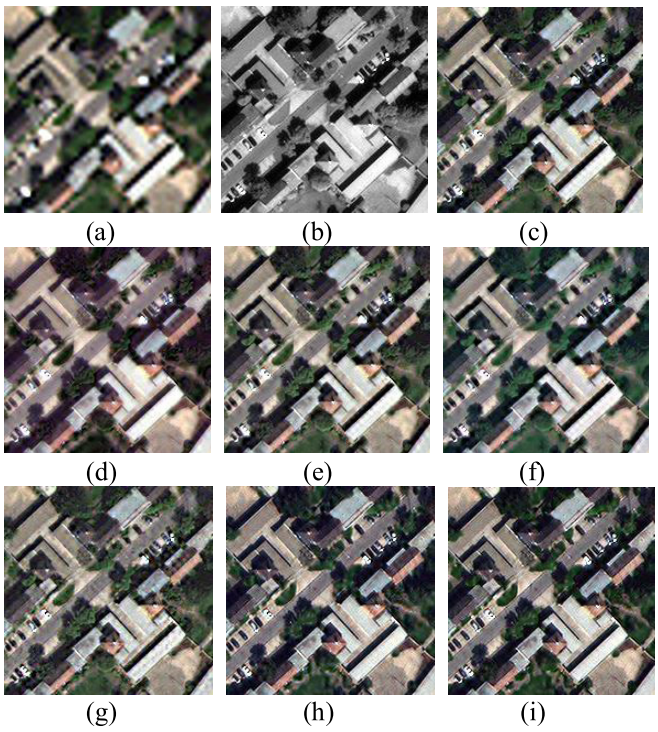

Fig. 7. Full-resolution fusion results from the WorldView-2 imagery, San Francisco, 2011. (a) Up-sampled MS image. (b) PAN. (c) SFIM. (d) GS. (e) MTF-GLP. (f) AWLP. (g) TSSC. (h) PNN. (i) DRPNN.

spectral curves that can be easily distorted, such as the bare soil in the top middle of Fig. 5(h) and (i).

For the two CNN-based methods, the high-quality results are difficult to tell apart, but by investigating the preservation of ground objects with small sizes, it can be confirmed that the deeper architecture of the DRPNN helps to more appropriately sharpen the small edges, such as the edges of the industrial buildings in the bottom left of Fig. 4(h) and (i). The same tendency is also seen in the full-resolution results of the real-data experiments displayed in Figs. 6 and 7.

V. Conclusion

In this letter, we have proposed the DRPNN to perform high-quality fusion of MS and PAN images. The DRPNN employs the high nonlinearity of a CNN to achieve a better performance. Furthermore, to make adequate use of the advantages of deep learning, the residual learning architecture is applied to allow the network to go deeper and boost its performance. The superiority of the proposed network was supported by the results of experiments on a large number of images covering various complex ground scenes.

The successful implementation of the DRPNN has motivated us to apply the framework to further studies in the field of multisource remote-sensing data fusion. Inspired by other recently published works based on deep learning, our strategy of learning a deep residual network for feature representation is expected to be further expanded for other tasks of image restoration and high-level content interpretation, including spatial–temporal unified fusion [19], multidomain unified quality improvement [20], [21], aerial scene classification [22], and saliency detection [23], [24].

References

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

[2] J. Choi, K. Yu, and Y. Kim, "A new adaptive component-substitution-based satellite image fusion by using partial replacement," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 1, pp. 295–309, Jan. 2011.

[3] B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, and M. Selva, "MTF-tailored multiscale fusion of high-resolution MS and pan imagery," *Photogramm. Eng. Remote Sens.*, vol. 72, no. 5, pp. 591–596, May 2006.

[4] J. G. Liu, "Smoothing filter-based intensity modulation: A spectral preserve image fusion technique for improving spatial details," *Int. J. Remote Sens.*, vol. 21, no. 18, pp. 3461–3472, Dec. 2000.

[5] H. Shen, X. Meng, and L. Zhang, "An integrated framework for the spatio–temporal–spectral fusion of remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7135–7148, Dec. 2016.

[6] 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. 1792–1805, May 2014.

[7] Q. M. Wang, W. Z. Shi, Z. B. Li, and P. M. Atkinson, "Fusion of sentinel-2 images," *Remote Sens. Environ.*, vol. 187, pp. 241–252, Dec. 2016.

[8] Q. Wang, W. Shi, and P. M. Atkinson, "Area-to-point regression kriging for pan-sharpening," *ISPRS J. Photogramm. Remote Sens.*, vol. 114, pp. 151–165, Apr. 2016.

[9] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 1646–1654.

[10] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2016.

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

[12] G. Masi, D. Cozzolino, L. Verdoliva, and G. Scarpa, "Pansharpening by convolutional neural networks," *Remote Sens.*, vol. 8, no. 7, p. 594, Jul. 2016.

[13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778.

[14] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, Jul. 2017.

[15] B. T. Polyak, "Some methods of speeding up the convergence of iteration methods," *USSR Comput. Math. Math. Phys.*, vol. 4, no. 5, pp. 1–17, 1964.

[16] Y. Jia *et al.*, "Caffe: Convolutional architecture for fast feature embedding," in *Proc. 22nd ACM Int. Conf. Multimedia*, 2014, pp. 675–678.

[17] *MatConvNet: CNNs for MATLAB*. [Online]. Available: http://www.vlfeat.org/matconvnet

[18] X. Otazu, M. Gonzalez-Audicana, O. Fors, and J. Nunez, "Introduction of sensor spectral response into image fusion methods. Application to wavelet-based methods," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 10, pp. 2376–2385, Oct. 2005.

[19] P. H. Wu, H. F. Shen, L. P. Zhang, and F. M. Göttsche, "Integrated fusion of multi-scale polar-orbiting and geostationary satellite observations for the mapping of high spatial and temporal resolution land surface temperature," *Remote Sens. Environ.*, vol. 156, pp. 169–181, Jan. 2015.

[20] Q. Yuan, L. Zhang, H. Shen, and P. Li, "Adaptive multiple-frame image super-resolution based on U-curve," *IEEE Trans. Image Process.*, vol. 19, no. 12, pp. 3157–3170, Dec. 2010.

[21] J. Li, Q. Yuan, H. Shen, and L. Zhang, "Noise removal from hyperspectral image with joint spectral–spatial distributed sparse representation," *IEEE Trans. Geosci Remote Sens.*, vol. 54, no. 9, pp. 5425–5439, Sep. 2016.

[22] F. Hu, G.-S. Xia, Z. Wang, X. Huang, and L. Zhang, "Unsupervised feature learning via spectral clustering of multidimensional patches for remotely sensed scene classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 5, pp. 2015–2030, May 2015.

[23] Q. Wang, J. Lin, and Y. Yuan, "Salient band selection for hyperspectral image classification via manifold ranking," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 6, pp. 1279–1289, Jun. 2016.

[24] Q. Wang, Y. Yuan, and P. Yan, "Visual saliency by selective contrast," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 7, pp. 1150–1155, Jul. 2013.