Pan-Sharpening Using an Efficient Bidirectional Pyramid Network

Yongjun Zhang, Chi Liu, Mingwei Sun, and Yangjun Ou

***Abstract*— Pan-sharpening is an important preprocessing step for remote sensing image processing tasks; it fuses a low-resolution multispectral image and a high-resolution (HR) panchromatic (PAN) image to reconstruct a HR multispectral (MS) image. This paper introduces a new end-to-end bidirectional pyramid network for pan-sharpening. The overall structure of the proposed network is a bidirectional pyramid, which permits the network to process MS and PAN images in two separate branches level by level. At each level of the network, spatial details extracted from the PAN image are injected into the upsampled MS image to reconstruct the pan-sharpened image from coarse resolution to fine resolution. Subpixel convolutional layers and the enhanced residual blocks are used to make the network efficient. Comparison of the results obtained with our proposed method and the results using other widely used state-of-the-art approaches confirms that our proposed method outperforms the others in visual appearance and objective indexes.**

***Index Terms*— Bidirectional pyramid network (BDPN), deep learning, image fusion, multilevel, pan-sharpening, remote sensing.**

I. Introduction

DUE to the hardware limitations of sensors, optical remote sensing satellites can only provide a low-resolution (LR) multispectral (LRMS) image and a high-resolution (HR) panchromatic (PAN) image; pan-sharpening refers to the technique of fusing the two to reconstruct a HR multispectral (HRMS). Pan-sharpening is very important for remote sensing image processing tasks and is often used as a preprocessing step for applications such as segmentation, classification, and object detection [1], [2]. In the last few decades, various pan-sharpening algorithms have been proposed to address this problem.

Among the available pan-sharpening techniques, component substitution (CS) methods are powerful approaches which are fast and easy to implement. These approaches usually transform the multispectral (MS) image into a suitable domain in which one of the components I is replaced by the PAN image. Then, the new components are converted back into the original domain using an inverse transformation. The representative algorithms are principal component analysis [3], intensity hue saturation transform [4], [5], and Gram–Schmidt (GS) sharpening [6]. However, the spectral characteristics and spectral range of the MS and PAN images differ from each other; the PAN image and the substituted component I do not generally have the same radiation characteristics. Therefore, the fusion process not only injects spatial details, but also leads to spectral distortions. Due to their efficiency and impressive spatial quality, CS methods are still investigated by researchers, concentrating on improving the spectral quality; strategies such as partial replacement [7], local coefficient calculation [8], and image classification [9] are used to reduce the spectral distortions.

Another class of pan-sharpening algorithms is based on multiresolution analysis (MRA). These approaches inject high-frequency details extracted from the PAN image into the upsampled MS image. The details are obtained through a MRA such as Laplacian pyramid [10], wavelet transform [11], [12], curvelets transform [13], and non-subsampled contourlets transform [14]. In general, MRA-based methods provide fused images with better spectral fidelity than those based on CS. However, spatial distortions may occur because of the aliasing effects and blurring of textures replacement [7].

The model-based optimization (MBO) approaches are another series of methods that have drawn much attention. The main idea of these methods is to build an energy function based on some reasonable assumptions and to minimize the energy to reconstruct the HRMS image [15]. Since reconstructing an HRMS image from an LRMS image is an ill-posed inverse problem, the MBO methods require appropriate regularizations to build an energy function, such as sparsity regularization [16]–[19], variational models [20]–[22], and Markov random fields [23], [24]. MBO methods make a tradeoff between spectral quality and spatial quality and generally achieve satisfying results. However, an appropriate model is challenging to be built, and the time complexity of the MBO methods is much higher than many other algorithms.

The complexity of ground objects and different spectral responses of different sensors make it difficult to formulate the relationship among the LRMS image, the PAN image, and the desired HRMS image. Fortunately, the development of deep learning offers new solutions to the abovementioned problem. The high nonlinearity of the convolutional neural network makes it effective to deal with the pan-sharpening problem.

Manuscript received October 24, 2018; revised January 8, 2019; accepted February 12, 2019. This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFB0505003, and in part by the National Natural Science Foundation of China under Grant 41871368. *(Corresponding author: Yongjun Zhang.)*

The authors are with the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China (e-mail: zhangyj@whu.edu.cn; liuchi@whu.edu.cn; Mingweis@whu.edu.cn; ouyangjun@whu.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/TGRS.2019.2900419

Existing deep learning-based pan-sharpening methods take the idea of single image super-resolution (SISR) as a reference, and network structures such as sparse denoising autoencoders networks [25] and deep residual convolutional networks [26] are often used. However, there is a significant difference between pan-sharpening and SISR: the spatial details of SISR are inferred from the LR image, whereas the pan-sharpening details are extracted from the HR PAN image. Existing deep learning-based approaches [1], [2], [27], [29] ignore this difference and are thus unable to make full use of the high-frequency information in PAN images.

In this paper, we propose a bidirectional pyramid network (BDPN) for pan-sharpening. The proposed network is superior to existing networks due to the following aspects.

1) The MS and the PAN images are processed separately, which enables better spectral preservation and details extraction;

2) Following the general idea of MRA, multilevel details are extracted from the PAN image and injected into the MS image to reconstruct the pan-sharpened image from coarse resolution to fine resolution.

3) The PAN image is downsampled while the MS image is upsampled in the network, which reduces the computation.

4) The use of subpixel convolutional layers and residual blocks (ResBlocks) makes the network more efficient.

The remainder of this paper is organized as follows. Section II briefly reviews SISR and existing deep learning-based pan-sharpening methods. Section III introduces our proposed pan-sharpening network. The experimental results of our proposed algorithm are presented and discussed in Section IV. Finally, our conclusions and future work are discussed in Section V.

II. Related Work

A. Deep Learning-Based SISR

SISR is a technique that reconstructs a HR image from the observed LR image. Due to the substantial loss of information during the transformation from an HR image to an LR image, the reconstruction process is a highly ill-posed problem. Fortunately, the relationship between an HR image and an LR image can be inferred based on the theory that most of the high-frequency components in an image are redundant and can be reconstructed from the low-frequency components. Among the existing SISR methods, deep learning-based methods provide superior performance due to their nonlinearity and have achieved state-of-the-art reconstruction accuracy.

Since Dong *et al.* [30] first proposed a deep learning-based SISR method, various CNN networks have been proposed for SISR. Another significant achievement in the evolution of SISR was the residual network architecture proposed by Xu *et al.* [23]. The input and output of the SISR network are highly similar, and the network learns the sparse residual between the two, so skip connection [31]–[33] and recursive convolution [34] alleviate the burden of carrying identity information in the super-resolution network and allow for a network with more convolutional layers. To reduce the computation time and memory required by the deep network architecture, efficient upsample strategies such as the deconvolution layer [35] and the subpixel convolutional layer [33] are proposed to upsample images in the network.

Of particular relevance to our paper are the works of Lai *et al.* [36]. The pyramid structure, which upsamples the LR image level by level, inspired the design of our pan-sharpening network.

B. Deep Learning-Based Pan-Sharpening

Pan-sharpening is a special form of super-resolution [1]; it also reconstructs an HR image from an LR image, with the difference that the spatial details are mainly learned from a PAN image. Most existing deep learning-based pan-sharpening methods are adapted from the SISR network, and can generally be divided into two groups.

The first group [2], [24] assumes that the relationship between the HR/LR PAN image patches is the same as that between the HR/LR MS image patches. To train a model, the downsampled PAN images and the original PAN images are used as the inputs and outputs, respectively. The trained model is then used to predict the HRMS images from the LRMS images. The PAN images are used only in training and not in the reconstruction, so the spatial quality of the results is unsatisfactory. Moreover, the difference between the PAN images and each band of the MS images, which is illustrated by Li *et al.* [16], is ignored, so the results also suffer from spectral distortion.

The other group of methods [1], [23], [25] takes the MS image and the PAN image as the input and trains an end-to-end network that directly outputs the pan-sharpened image. In the preparation phase, the LRMS image is upsampled to the scale of the PAN image using bicubic interpolation; then, the PAN image is concatenated with the upsampled LRMS image to comprise the five-band input. The output of the CNN is a four-band MS image with the same spatial resolution as the PAN image. However, the simplex CNN structure processes each band of the input with no discrimination, making it difficult to extract the spatial details from the PAN image. In addition, the upsampling operation before the network makes the network computationally complex.

III. Methodology

In this section, we describe the design methodology of the proposed network. As illustrated in Fig. 1, the overall structure of the proposed network is a bidirectional pyramid. Different from the existing pan-sharpening networks, the proposed network processes the MS and PAN images in two separate branches. In the reconstruction branch (the part inside the red dotted area), the LRMS image is upsampled level by level, which effectively suppresses the reconstruction artifacts caused by the bicubic interpolation and dramatically reduces the computational complexity. In the details extraction branch (the part inside the green dotted area), the spatial details of the PAN image are extracted and injected into the corresponding upsampled MS image.

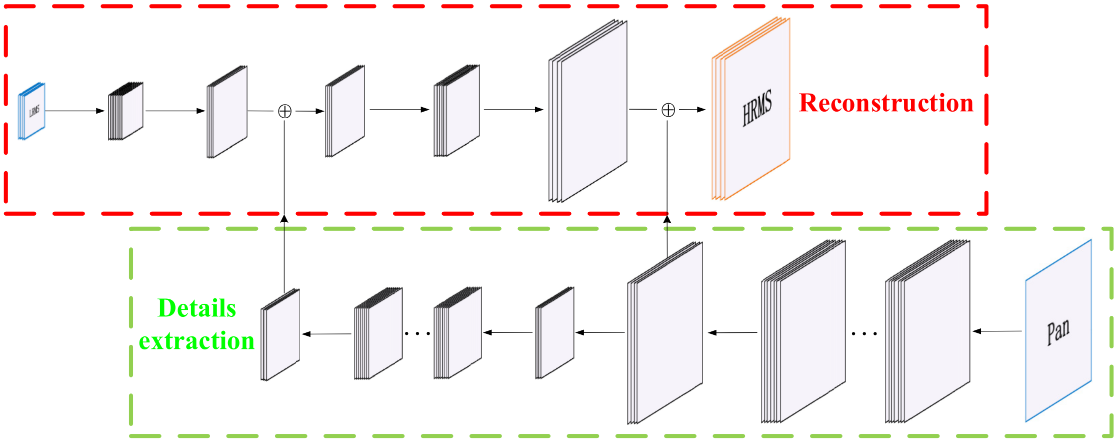


Fig. 1. Structure of the BDPN. The reconstruction branch is inside the red dotted area, the details extraction branch is in the green dotted area, the inputs are shown with a blue border, and the output has an orange border.

A. Multilevel Structure

For most satellite sensors, the resolution of the desired pan-sharpened MS image is four times that of the input MS image. Directly upsampling the MS image by four times definitely creates severe reconstruction artifacts. Inspired by the work of Lai *et al.* [36], we propose a multilevel structure to reconstruct the pan-sharpened image from coarse resolution to fine resolution. The MS image is upsampled and the PAN image is downsampled level by level. At each level, the spatial details extracted from the PAN image are injected into the corresponding MS image.

Each details extraction level can be formulated as

$$\mathrm{Pan}_{i+1} = \begin{cases} g^{n_b}(\mathrm{Pan}_i) & \text{if } i = 0 \\ g^{n_b}(\mathrm{maxpooling}(\mathrm{Pan}_i)) & \text{if } i > 0 \end{cases} \tag{1}$$

where i and $i+1$ are the level index, Pan_i and Pan_{i+1} are the input and output of current level, respectively, g is a function denoting the Resblock, n_b is the number of Resblocks in each level, g^{n_b} indicates that the input is processed by n_b Resblocks. For level 0, the details are directly extracted from the Pan image by the Resblocks, while in other levels, the inputs are downsampled and then processed by the Resblocks.

Each level of the reconstruction branch can be formulated as

$$\mathrm{MS}_{i+1} = [f(\mathrm{MS}_i)] \uparrow + \mathrm{Pan}_{N-i} \quad 0 \le i < n \tag{2}$$

where N is the number of levels, for MS and PAN images whose resolutions differ by four times, n is set to 2. i, $i+1$, and $N-i$ are the level index, MS_i is the input MS image of the current reconstruction level, Pan_{N-i} is the output of the corresponding details extraction level, and MS_{i+1} is the output of the current reconstruction level. f and $\uparrow$ are the convolutional operation and upsampling operation in subpixel convolution, respectively.

It should be noted that the parameters of Resblocks and subpixel convolution are not shared, so for each Resblock and subpixel convolution, g and f are different, and they are learned automatically by the network to minimize the difference between the output of each reconstruction level and the corresponding reference image, which is acquired by downsampling the original reference image.

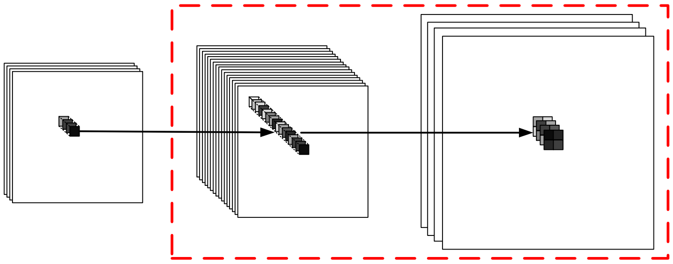


Fig. 2. Single level of the reconstruction branch. The red dotted area indicates the process of subpixel convolution.

B. Reconstruction Branch

The main task of the reconstruction branch is to upsample the input MS image and inject the spatial details extracted from the PAN image without changing the spectral characteristics of the original MS image.

At each level of the reconstruction branch, the four-channel map is fed into a convolutional layer and a 16-channel feature map is generated. The feature map is realigned to an upsampled feature map, which also has four channels but is doubled in size. Then, the spatial details extracted from the corresponding details extraction level are injected into the upsampled MS image. The output of each reconstruction level is an upsampled MS image that has twice the resolution of the input.

The diagram of the subpixel convolutional layer is shown below in Fig. 2, each pixel of the 16-band feature map forms a 1×16 vector, and the vector is realigned into a $2 \times 2 \times 4$ matrix. The subpixel convolution was first proposed by Shi in 2016 [37]. Researches showed that a subpixel convolutional layer with kernel $(o \times r{\times}r, I, k, k)$ had the same effect as a deconvolution layer with kernel $(o, I, k \times r, k \times r)$, where I is the input channels, r is the scale factor, k and $k \times r$ are the kernel width, $o \times r \times r$, and o is the output channels. By preshuffling the training data to match the output of the

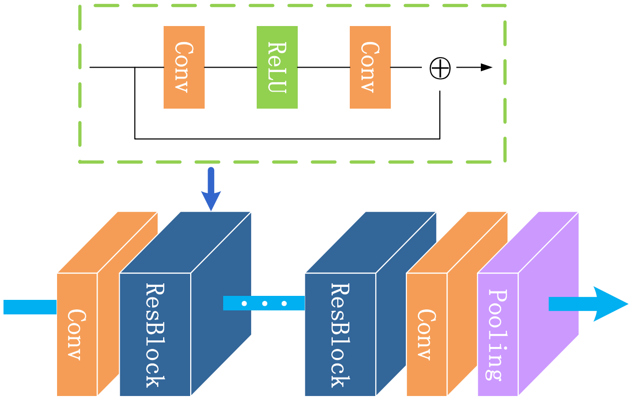


Fig. 3. Single level of the details extraction branch. The green dotted area indicates the structure of a ResBlock.

layer, the subpixel convolutional layer is $\log_2 r^2$ times faster compared to the deconvolution layer in training and r^2 times faster compared to implementations using various forms of upscaling before convolution [38].

C. Detailed Extraction Branch

The residual network (ResNet) solves the gradient vanishing problem and allows for a deeper CNN with more filtering layers to exploit high nonlinearities and extract more representative features. The idea of the ResNet structure is especially suitable for solving the pan-sharpening problem because the input MS image and output pan-sharpened image are highly similar, enabling ResNet to produce a residual image in which most of the values are likely to be zero or very small. Therefore, most of the recently proposed deep learning-based pan-sharpening methods are based on ResNet [26]–[29].

However, the original ResNet was proposed to solve higher level computer vision problems such as image classification and object detection, and applying it directly to pan-sharpening may introduce unnecessary computational expenses. The ResBlock structure that was proposed subsequently [38] is a better choice for pan-sharpening. By removing all the batch normalization layers and the rectified liner unit layers after the shortcut connection of the original residual block, the model convergence accelerated and its size decreased.

The diagram of details extraction branch is shown in Fig. 3. First, a 64-channel feature map is extracted from the PAN image by a convolutional layer. Then, the stacked ResBlocks are used to extract the residual features. An additional convolutional layer is connected to the last ResBlock to convert the feature map into a four-channel details map, which will be injected into the reconstruction branch. At the end of each detail extraction level, the size of the details map is reduced by half by a max-pooling layer.

The number of ResBlocks determines the receptive field of the network. In the case of pan-sharpening, local structures receive more attention than global features, so there is no need for an extra deep network structure. The same number of ResBlocks is used in different levels, which allows for a wider receptive field on the lower resolution feature maps to catch the structural information, and a smaller receptive field on the higher resolution feature maps to concentrate on local spatial details.

D. Loss Function

The network predicts residual images at different levels and produces multiscale output images. The corresponding ground truth is obtained by downsampling the reference image. In this way, loss at each level can be calculated. For a two-level network, the total loss of the model is represented as

$$\text{Loss} = \lambda \text{loss}_1 + (1-\lambda)\text{loss}_2 \tag{3}$$

where loss_1 and loss_2 are the losses of the first level and second level, respectively, and λ weighs the importance between the two losses. At the beginning of training, λ is set to 1 and the total loss is equal to loss_1 so only the reconstruction result of the first level is supervised, allowing for the model to converge quickly. As the training proceeds, λ decreases gradually and the weight of loss_2 becomes progressively heavier. Finally, λ decreases to 0 and the final loss is equal to loss_2, which guarantees the accuracy of the final reconstruction results.

For each reconstruction level, the loss function is based on the relative dimensionless global error in synthesis (ERGAS) index [40], which is an overall assessment of pan-sharpening

$$\text{loss}_i = \sqrt{\frac{1}{B}\sum_{b=1}^{B}(\text{RMSE}_i(b) \times e^{-u_i(b)})^2} \tag{4}$$

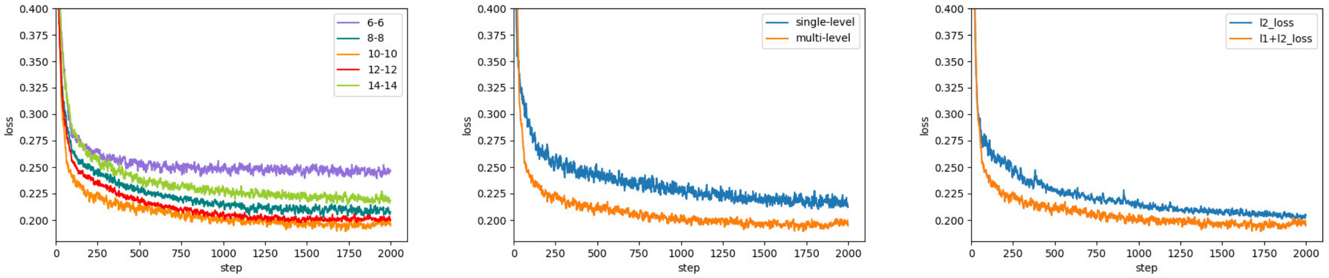
where i is the level index, B is the number of bands labeled with the b index, $\mu_i(b)$ is the mean of the bth band of the current level reference image, and $\text{RMSE}_i(b)$ is the root-mean-square error between current level prediction and the corresponding reference image.

IV. Experiments

Experiments were conducted to evaluate the performance of the proposed network. Trained models with different network parameters were evaluated and compared to select the best one. Then, full-resolution and reduced-resolution experiments were performed. Our best model was compared with other seven existing methods based on visual appearance and the widely accepted objective indexes.

A. Data Set and Model Training

To evaluate the performance of our newly presented architecture, the network was trained on a data set consisting of GF2, IKONOS, QuickBird, and WorldView3 (bands 2, 3, 5, and 7 were selected to comprise the four-band MS image) images. To make the trained model more robust, images covering different areas (including urban, rural, seaside, and mountain areas) are used. The spatial resolution of the PAN images are 1 m (GF2), 1 m (IKONOS), 0.7 m (QuickBird), 0.31 m (WorldView3), respectively. The 4000 image patches were collected in total. For each sensor, 20 image patches were randomly selected for testing, one-fifth of the rest were used for validation, and others for training. Following Wald's protocol spatially degraded images were used as inputs, and the original MS images were used as the reference images.



(a) (b) (c)

Fig. 4. Losses of network with different structures or loss functions. (a) Networks with different numbers of ResBlocks. (b) Networks with and without multilevel structure. (c) Networks with and without multilevel loss function.

The experiments were carried out on a desktop equipped with an NVIDIA GeForce GT 1080Ti GPU. The patch size of the MS images and PAN images for training were 64 × 64 and 256 × 256 pixels, respectively, the batch size was set to 8, the initial learning rate was set to 1×10^{-4}, the learning rate descent factor was set to 0.8 every 100 epochs, and the max iteration was set to 3000. The coefficient λ was decreased by 0.01 every five epochs.

B. Network Exploration

A critical structural parameter of the proposed network is the ResBlocks number, which directly determines the depth of the network. Generally, a network with more convolutional layers extracts features at a higher level and performs better. However, our pan-sharpening experiments show different results. Networks with different numbers of ResBlocks were trained to explore the influence of the parameter. Another structure to be explored is the multilevel structure. By removing the second level of the details extraction branch, the network was simplified to a single-level network. The simplified network was also trained to explore the influence of a multilevel structure. To verify the effectiveness of multilevel loss function, model use only loss of the last construction level was also trained and compared with the combination loss function.

As mentioned earlier, in the proposed structure, losses of different levels were calculated using the same loss function, and their coefficients add up to 1. The networks with different numbers of ResBlocks were trained with the same loss function. For network without multilevel structure or without multilevel loss, their losses can also be treated as combination of multilevel, i.e., middle level loss whose coefficient is 0 and final reconstruction level loss whose coefficient is 1. Especially, after 500 epochs, losses of all the networks contain only loss of the final level. The loss of the validation could be used as a quality index to evaluate the trained model.

As shown in Fig. 4(a), n_b-n_b indicates that there are two levels, and each level has n_b ResBlocks. When the number of ResBlocks is 6-6, the model converges quickly but has the largest final loss because the receptive field of the feature map is too small to catch the local structural information. As the number of ResBlocks increases, the number of parameters increases and the model converges more slowly. When the number of ResBlocks is 10-10, the network achieves its minimum loss and best performance. When the number of ResBlocks continues to increase, the convergence rate declines, and the final loss increases. The reason is that pan-sharpening concentrates on extracting low-level features from the local area, so there is no need for such a wide-ranging receptive field, and the deep network structure makes the training inefficient.

In Fig. 4(b), the superiority of the multilevel structure is verified. Both networks are equipped with 20 ResBlocks, the single-level structure converges slower, and the final loss is heavier. It can be concluded that the multilevel structure speeds up the convergence and improves the performance.

In Fig. 4(c), the loss of network with $l1 + l2$ loss decreases much quicker than the single-level loss one in the beginning 500 epochs, which verifies the effectiveness of taking consideration of middle-level loss. In later epochs, the difference between the two decreases gradually and finally, the losses are very close. It can be concluded that network use only loss of the final level can achieve similar performance to the proposed combination loss, at the cost of more training time.

C. Evaluation Indexes

The quality of the fusion results can be evaluated using two strategies. The first one is conducted on full-resolution image data, which is known as “quality with no reference” (QNR) [41]. The other one is conducted on reduced-resolution image data according to the Wald protocol [42].

The QNR indexes are calculated by exploiting the relationship between the pan-sharpened image and the original MS image and the relationship between the pan-sharpened image and the original PAN image. The QNR indexes are based on the Q index [43], which is defined as

$$Q = \frac{\sigma_{xy}}{\sigma_x \times \sigma_y} \times \frac{2\sigma_x \times \sigma_y}{\sigma_x^2 + \sigma_y^2} \times \frac{2\bar{x} \times \bar{y}}{\bar{x}^2 + \bar{y}^2} \tag{5}$$

where x and y are the test image and the reference image, respectively, $\bar{x}$ and $\bar{y}$ are the means of x and y, respectively, and σ_x and σ_y are the variances of x and y, respectively. Q is comprised of three different factors that account for the correlation, mean bias, and contrast variation of the test spectral bands with respect to their references.

1) Spectral Distortion Index (D_λ) [41]:

$$D_\lambda = \frac{1}{B(B-1)} \sum_{b=1}^{B} \sum_{l=1(b\neq l)}^{B} |Q(x_b, x_l) - Q(\tilde{x}_b, \tilde{x}_l)| \quad (6)$$

where x_b, x_l denote the bth and lth band of the pan-sharpened image, respectively, $\tilde{x}_b, \tilde{x}_l$ denote the bth and lth band of the LRMS image, respectively, and B is the number of bands. The spectral distortion index measures the relative relationship between the interband Q indexes of the fused and original images.

2) Spatial Distortion Index(D_s) [41]:

$$D_s = \frac{1}{B} \sum_{b=1}^{B} |Q(x_b, P) - Q(\tilde{x}_b, \tilde{P})| \quad (7)$$

where P and $\tilde{P}$ are the PAN image and degraded PAN image, respectively. The spatial distortion index measures the interband relationship between the Q indexes of the fused and PAN.

3) QNR:

$$\mathrm{QNR} = (1 - D_\lambda) \times (1 - D_s). \quad (8)$$

The QNR is composed of the spectral distortion index D_λ and the spatial distortion index D_s. The best value of QNR is 1.

The quality indexes with reference are based on the assumption of scale invariance, the MS, and PAN images are degraded at a certain scale and fused to generate the pan-sharpened image, and the original MS image is used as the reference. The following indexes are chosen for reference evaluation.

4) Correlation Coefficient (CC) [44]:

$$\mathrm{CC} = \frac{\sum_{m=1}^{M} (P_m - \bar{p}_m) \times (R_m - \bar{R}_m)}{\sqrt{\sum_{m=1}^{M} (P_m - \bar{p}_m)^2 \times \sum_{m=1}^{M} (R_m - \bar{R}_m)^2}} \quad (9)$$

where m refers to the mth pixel, M is the total number of pixels, R is the reference HR MS image, and P is the pan-sharpened image. $\bar{R}$ and $\bar{P}$ are the mean values of R and P, respectively. CC evaluates the correlation degree between the two.

5) RMSE [45]:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{m=1}^{M} (P_m - R_m)^2}{M}}. \quad (10)$$

The RMSE measures the standard difference between two.

6) ERGAS:

$$\mathrm{ERGAS100} \triangleq \frac{d_h}{d_l} \sqrt{\frac{1}{B} \sum_{b=1}^{B} \left(\frac{\mathrm{RMSE}(b)}{\mu(b)} \right)^2} \quad (11)$$

where $\mathrm{RMSE}(b)$ is the RMSE between the bth fused band and the reference band, dh/dl is the scale ratio between the PAN image and the MS image, $\mu(k)$ is the mean of the bth band, and B is the number of bands. ERGAS accounts for the spatial distortion; the closer the value is to 0, the better the quality of the pan-sharpened MS.

7) Spectral Angle Mapper [46]:

$$\mathrm{SAM(v}, \hat{v}) \triangleq \arccos \left(\frac{\langle v, \hat{v} \rangle}{||v||_2 \times ||\hat{v}||_2} \right) \quad (12)$$

where v and $\hat{v}$ are the pixel vector of the pan-sharpened image and the reference, respectively, and SAM is averaged on the whole image and reflects the spectral distortion between the fused image and the reference image.

8) Universal Image Quality Index (Q4) [47]:

$$Q_4 = \frac{|\sigma_{z_1 z_2}|}{\sigma_{z_1} \cdot \sigma_{z_2}} \times \frac{2\sigma_{z_1} \times \sigma_{z_2}}{\sigma_{z_1}^2 + \sigma_{z_2}^2} \times \frac{2|\bar{z}_1| \times |\bar{z}_2|}{|\bar{z}_1|^2 + |\bar{z}_2|^2} \quad (13)$$

where $z_1 = x_1 + ix_2 + jx_3 + lx_4$, $z_2 = \hat{x}_1 + i\hat{x}_2 + j\hat{x}_3 + l\hat{x}_4$, x_b and $\hat{x}_b$ are the bth band of the fused MS image and reference, respectively. Here, i, j, and l are imaginary units, $\bar{z}$ and σ_z are the mean and variance of variable z, respectively, and $\sigma_{z_1 z_2}$ is the covariance between z_1 and z_2. Q_4 is an improved version of Q for MS images with four spectral bands.

D. Results and Comparison

We compared the performance of the trained model on test data set against seven widely accepted pan-sharpening methods: the GS method [48], guided filter-based pan-sharpening (GFP) method [49], matting model-based pan-sharpening (MMP) method [50], l1/2 gradient prior-based pan-sharpening (L12) method [21], pan-sharpening based on image segmentation(Seg_GLP)[48], deep residual network for pan-sharpening (DRPNN) [28], and target-adaptive CNN-based pan-sharpening (PNN) [26]. The GS method and GFP method are two typical CS-based methods. MMP and L12 are two state-of-the-art MBO methods. (In some papers, MMP is considered a CS-based method; in this paper, MMP is considered an MBO method because it separates the foreground and background based on the matting model.) The Seg_GLP method is an MRA method based on image segmentation. DRPNN and PNN are two recently proposed deep convolutional networks for pan-sharpening. The implementations of the compared algorithms are available online,[1–6] trained models of DRPNN and PNN are also provided by the authors. The default parameters given in their implementations are adopted.

Fig. 5 shows an example of the full-resolution experiment performed on the GF2 image. For visualization, all the images were rendered by ArcGIS Desktop [52] with default parameters; for the MS images, the red, green, and blue bands were chosen for display. Fig. 5(a) and (b) shows the original MS and PAN images, respectively, Fig. 5(c)–(k) shows the pan-sharpened MS images obtained by different methods. The two CS-based methods work well for enhancing the spatial details but produce severe spectral distortions, especially in the vegetation region. The MMP, L12, Seg_GLP, PNN, and

[1] https://github.com/mustafateke/Pansharp

[2] https://github.com/sjtrny/FuseBox

[3] http://smartdsp.xmu.edu.cn/PansharpeningStructure.html

[4] http://openremotesensing.net/kb/codes/pansharpening

[5] https://github.com/Decri/DRPNN-Deep-Residual-Pan-sharpening-Neural-Network

[6] https://github.com/sergiovitale/pansharpening-cnn

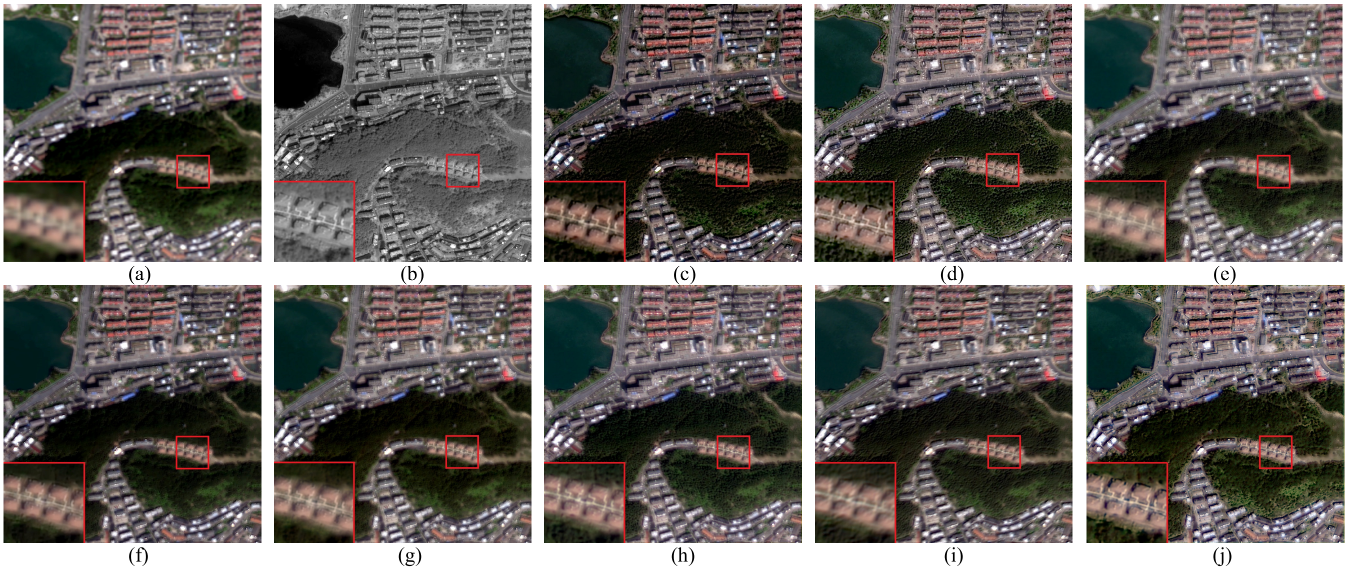


Fig. 5. Comparison of pan-sharpening results obtained by different methods (GF2 image). (a) LR MS image. (b) PAN image. (c)–(j) Pan-sharpening results of GS, GFP, MMP, L12, Seg_GLP, DRPNN, PNN, and the proposed method.

TABLE I
Objective Performance of the Pan-Sharpening Methods in Fig. 5

Method	GS	GFP	MMP	L12	Seg_GLP	DRPNN	PNN	Proposed
D_λ	0.1528	0.1951	0.0696	0.0354	0.0466	0.0261	**0.0248**	0.0566
Ds	0.0498	0.0847	0.0864	0.1256	0.0783	0.0783	0.1421	**0.0443**
QNR	0.8050	0.7367	0.8500	0.8434	0.8787	0.8976	0.8366	**0.9016**

DRPNN methods preserve the spectral information well but produce different levels of blurring artifacts. The proposed method produces pan-sharpened images with the best visual quality. Table I shows the objective performance of different methods. The PNN method achieves the best D_λ and the proposed method achieves the best Ds and QNR.

Fig. 6 shows another GF2 experiment performed on the downsampled version. Generally, all the pan-sharpened images are of good spectral quality. Regarding spatial quality, GS and GFP show the best performance, followed by the proposed method and PNN. The holes in the roof are hard to recognize in the Seg_GLP, MMP, L12, and DRPNN results, which suggests that these methods suffer from blurring artifacts in the object boundaries. Table II shows the objective performance of different methods. The proposed method shows the best pan-sharpening results based on most of the image quality indexes, i.e., CC, Q4, and RMSE.

Fig. 7 illustrates the full-resolution experiment performed on the IKONOS image. The L12, PNN, and DRPNN results are blurry and single trees are not recognizable in the enlarged view. GS, GFP, and MMP preserve the spatial details well in most regions but suffer from artifacts in local areas such as vegetation in the center, which is due to the complex textures and numerous spatial details in vegetation regions. Seg_GLP suffers slight spectral distortions in vegetation regions. Regarding visual effects, the proposed method produces the highest quality image. The objective evaluation in Table III is consistent with the visual comparison, the proposed method gets the best D_λ, D_s, and QNR, indicating that the proposed method performs the best in both enhancing spatial details and preserving the spectral information.

Fig. 8 shows the results of the downsampled experiment on the IKONOS image. The GS, GFP, and Seg_GLP methods produce artifacts in the vegetation regions. Two model-based methods, MMP and L12, fail to restore the spatial details properly; they suffer from artifacts in the texture-complex regions and blur in weakly textured regions. The result of the PNN, DRPNN, and proposed methods are more similar to the reference image. The objective quality indexes comparison is shown in Table IV which indicates that the proposed method performs the best for the ERGAS, CC, and RMSE indexes

Fig. 9 shows an experiment performed on a QuickBird seaside image. It can be seen that the GS, GFP, and PNN results suffer from spectral distortion, as the color of the vegetation is different from that in the original MS image. The results produced by MMP, L12, Seg_GLP, and DRPNN are blurry and the houses in the enlarged view are not clear. The Seg_GLP result also suffers from severe artifacts in the vegetation regions. Only the proposed method produces satisfactory results. The objective evaluation in Table V shows that the GS and DWFT method get the best D_s and D_λ, respectively, and the proposed method gets the best QNR.

Fig. 10 shows the downsampled experiment performed on the QuickBird images. In this experiment, the results produced

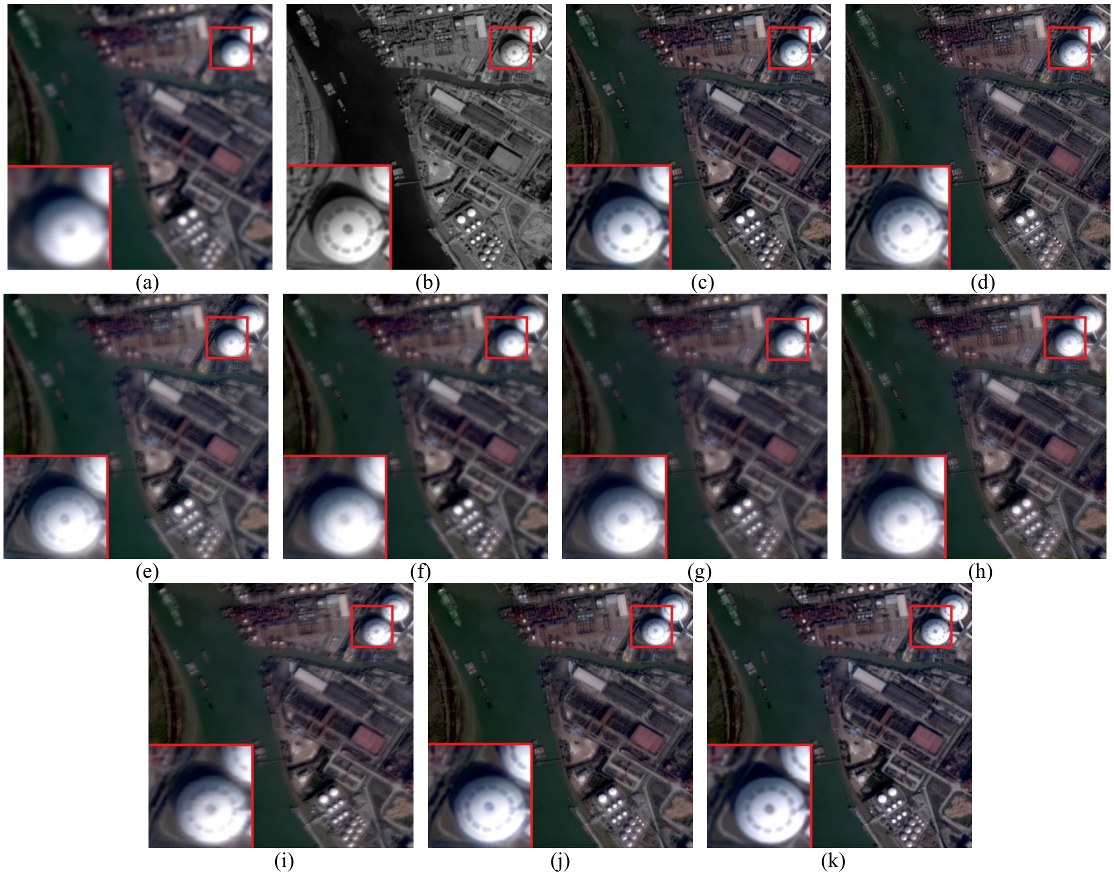


Fig. 6. Comparison of pan-sharpening results obtained by different methods (downsampled GF2 image). (a) LR MS image. (b) PAN image. (c)–(j) Pan-sharpening results of GS, GFP, MMP, L12, Seg_GLP, DRPNN, PNN, and the proposed method. (k) Reference image

TABLE II
OBJECTIVE PERFORMANCE OF THE PAN-SHARPENING METHODS IN FIG. 6

Method	GS	GFP	MMP	L12	Seg_GLP	DRPNN	PNN	Proposed
ERGAS	2.2353	2.3166	2.1091	**1.9270**	2.0116	1.9347	2.1062	2.0896
SAM	1.0209	1.0782	0.9573	**0.8739**	0.9617	1.0828	1.3288	1.1461
CC	0.9631	0.9582	0.9670	0.9721	0.9696	0.9737	0.9734	**0.9739**
Q4	0.7227	0.6799	0.7134	0.7250	0.7110	0.7225	0.7449	**0.7506**
RMSE	20.4982	21.2825	19.4364	17.2065	18.5474	16.9153	17.1711	**16.8674**

TABLE III
OBJECTIVE PERFORMANCE OF THE PAN-SHARPENING METHODS IN FIG. 7

Method	GS	GFP	MMP	L12	Seg_GLP	DRPNN	PNN	Proposed
D_λ	0.1596	0.2214	0.1751	0.0393	0.1375	0.0573	0.0450	**0.0387**
Ds	0.1842	0.1352	0.1300	0.0949	0.1049	0.0752	0.0452	**0.0340**
QNR	0.6856	0.6733	0.7177	0.8695	0.7720	0.8718	0.9118	**0.9286**

by GS and Seg_GLP are blurry and the road boundaries in the enlarged view are not clear; the GS result also suffers from spectral distortion in the vegetation regions. Similar to previous experiments, the GFP method suffers from artifacts in texturally complex regions and spectral distortion in bare land areas. The MMP result suffers from slight spectral distortion as indicated by the abnormal color of vegetation regions in the enlarged view. L12, PNN, DRPNN, and the proposed methods perform well in producing clear pan-sharpened images. The objective evaluation results in Table VI show that the proposed method performs best regarding ERGAS and Q_4, and DRPNN performs best regarding SAM, CC, and RMSE.

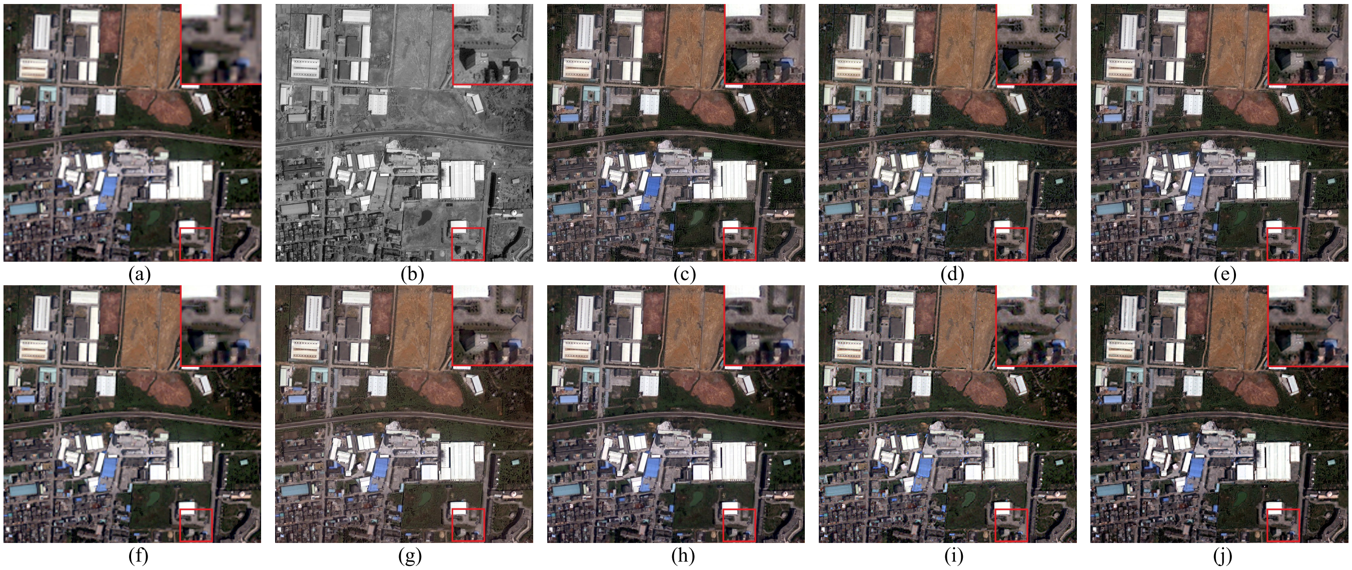


Fig. 7. Comparison of pan-sharpening results obtained by different methods (IKONOS image). (a) LR MS image. (b) Pan image. (c)–(j) Pan-sharpening results of GS, GFP, MMP, L12, Seg_GLP, DRPNN, PNN, and the proposed method.

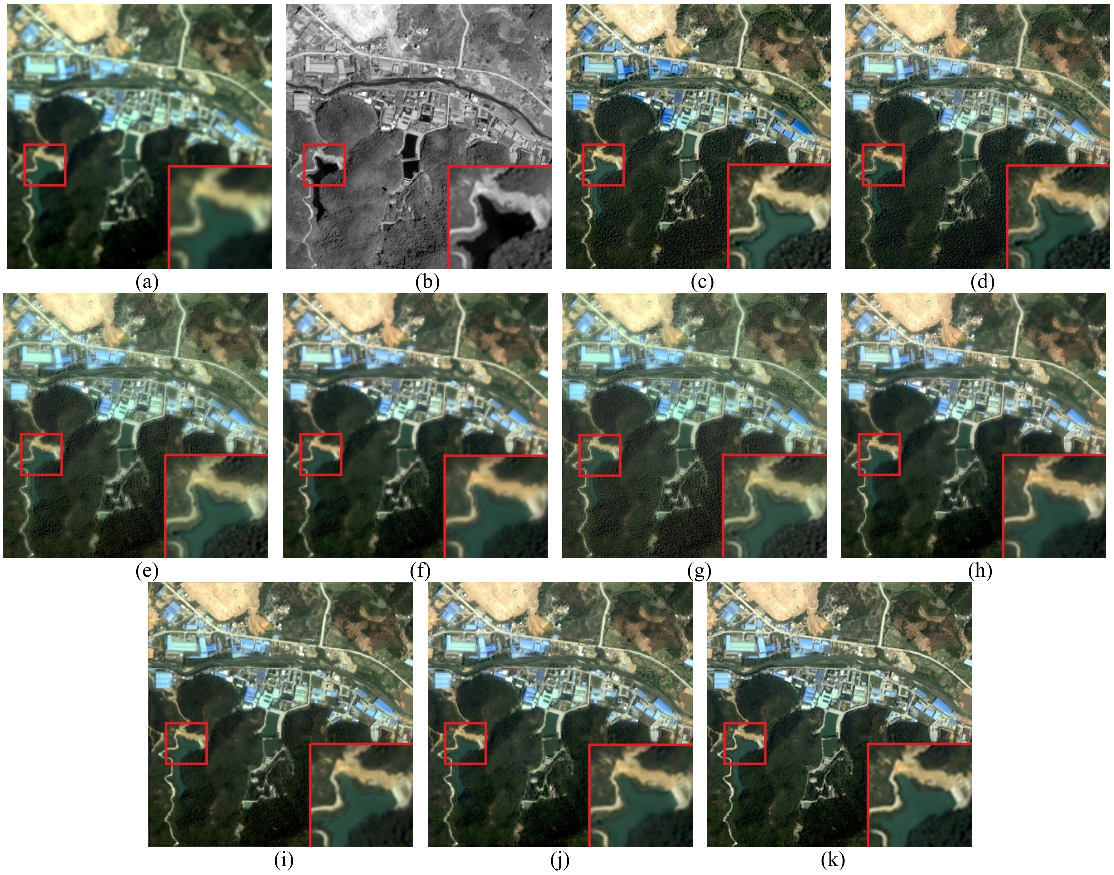


Fig. 8. Comparison of pan-sharpening results obtained by different methods (downsampled IKONOS image). (a) LR MS image. (b) PAN image. (c)–(j) Pan-sharpening results of GS, GFP, MMP, L12, Seg_GLP, DRPNN, PNN, and the proposed method. (k) Reference image.

The results of full-resolution WorldView3 data are shown in Fig. 11; the spatial details in the GS result are well reconstructed but apparent spectral distortion can be found in areas such as roads, vegetation, roof, and water. The GFP result also suffers from spectral distortion in vegetation regions. The L12, and DRPNN methods fail to reconstruct the roof details. The MMP, Seg_GLP, PNN, and proposed methods produce results with good spectral and spatial quality. The objective

TABLE IV
OBJECTIVE PERFORMANCE OF THE PAN-SHARPENING METHODS IN FIG. 8

Method	GS	GFP	MMP	L12	Seg_GLP	DRPNN	PNN	Proposed
ERGAS	2.6283	2.4671	2.5525	2.4257	2.4886	2.4158	2.2888	**2.1567**
SAM	3.3600	3.3923	3.2557	2.9053	3.3191	2.5955	**2.2342**	2.7022
CC	0.9581	0.9616	0.9617	0.9631	0.9614	0.9652	0.9659	**0.9698**
Q4	**0.8599**	0.8444	0.7863	0.7854	0.8048	0.8460	0.8567	0.8443
RMSE	46.9970	41.7538	43.5609	42.0916	42.6467	40.5339	38.8833	**37.7224**

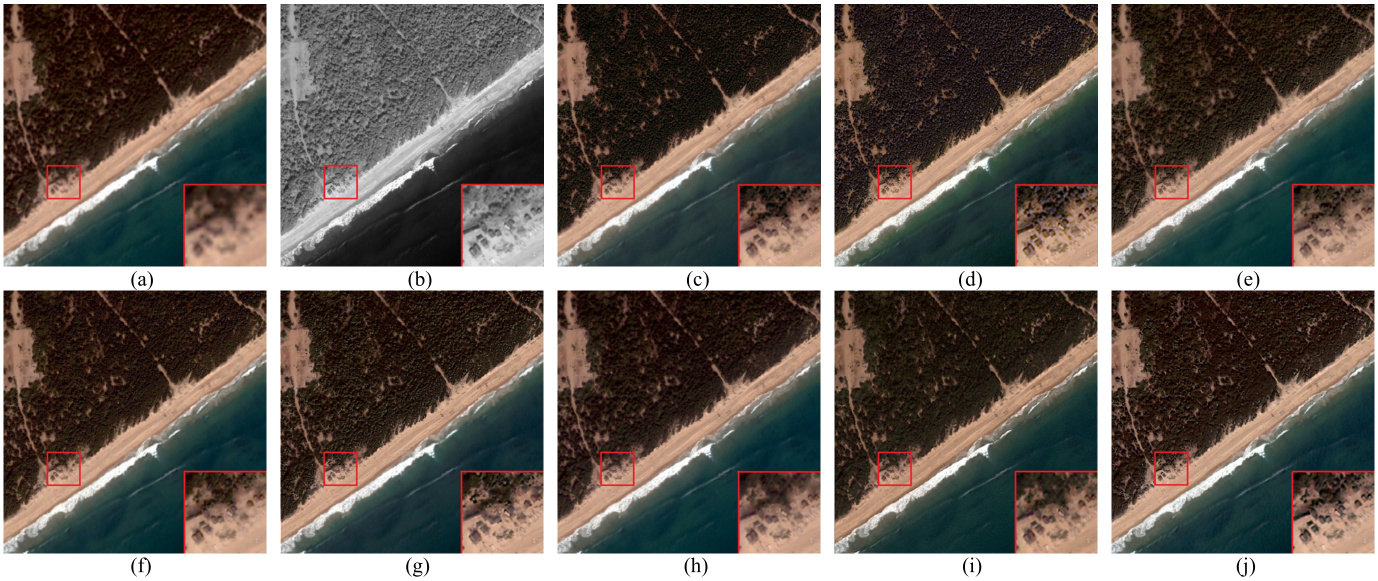

(a) (b) (c) (d) (e)
(f) (g) (h) (i) (j)

Fig. 9. Comparison of pan-sharpening results obtained by different methods (QuickBird image). (a) LR MS image. (b) PAN image. (c)–(j) Pan-sharpening results of GS, GFP, MMP, L12, Seg_GLP, DRPNN, PNN, and the proposed method.

TABLE V
OBJECTIVE PERFORMANCE OF THE PAN-SHARPENING METHODS IN FIG. 9

Method	GS	GFP	MMP	L12	Seg_GLP	DRPNN	PNN	Propose
D_λ	0.0720	0.1920	0.0332	**0.0187**	0.0618	0.0409	0.0896	0.0528
Ds	**0.0527**	0.1096	0.0854	0.1166	0.0672	0.1191	0.0875	0.0661
QNR	0.8791	0.7194	0.8842	0.8669	0.8752	0.8449	0.8307	**0.8846**

TABLE VI
OBJECTIVE PERFORMANCE OF THE PAN-SHARPENING METHODS IN FIG. 10

Method	GS	GFP	MMP	L12	Seg_GLP	DRPNN	PNN	Proposed
ERGAS	2.3965	2.5142	2.3048	2.0223	2.1805	2.0124	2.1819	**2.0087**
SAM	2.6423	2.9468	2.3916	2.3129	2.1493	**1.9903**	2.0446	2.0673
CC	0.9540	0.9466	0.9608	0.9658	0.9646	**0.9691**	0.9633	0.9668
Q4	0.8574	0.8656	0.8339	0.8519	0.8687	0.8759	0.8733	**0.8824**
RMSE	32.1591	27.7098	27.7389	25.6286	25.3577	**24.376**	25.2516	24.4343

evaluation in Table VII is consistent with the visual assessment that the proposed method performs best regarding Ds and QNR.

Fig. 12 shows experiment performed on the downsampled WorldView3 data. The GS and GFP method results suffer color distortion in vegetation regions, and the GFP method also

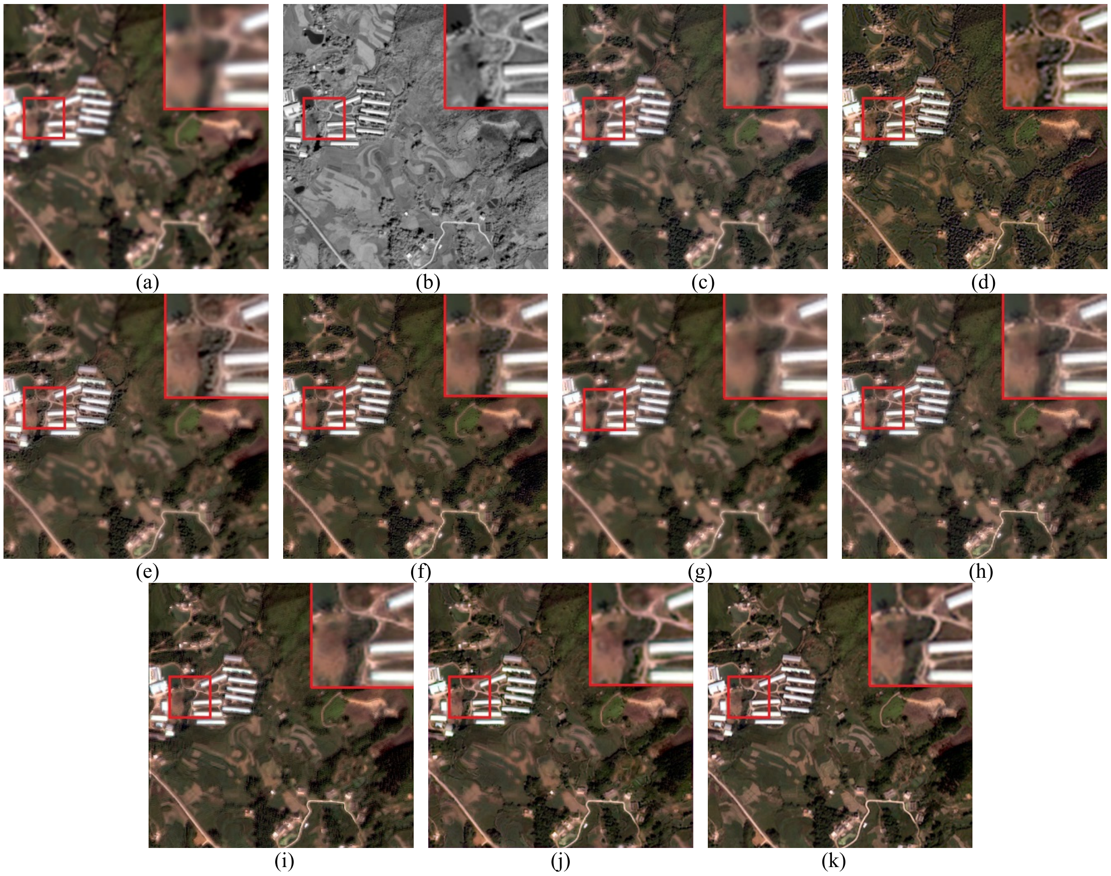


Fig. 10. Comparison of the pan-sharpening results obtained by different methods (downsampled QuickBird image). (a) LR MS image. (b) PAN image. (c)–(j) Pan-sharpening results of GS, GFP, MMP, L12, Seg_GLP, DRPNN, PNN, and the proposed method. (k) Reference image.

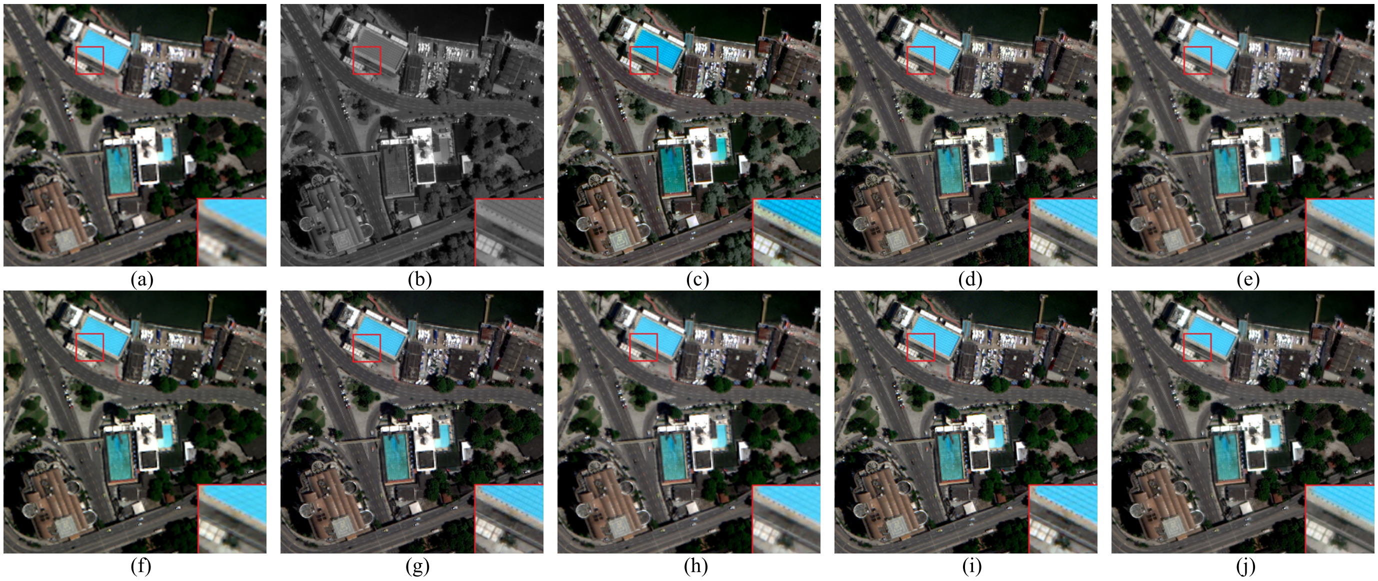
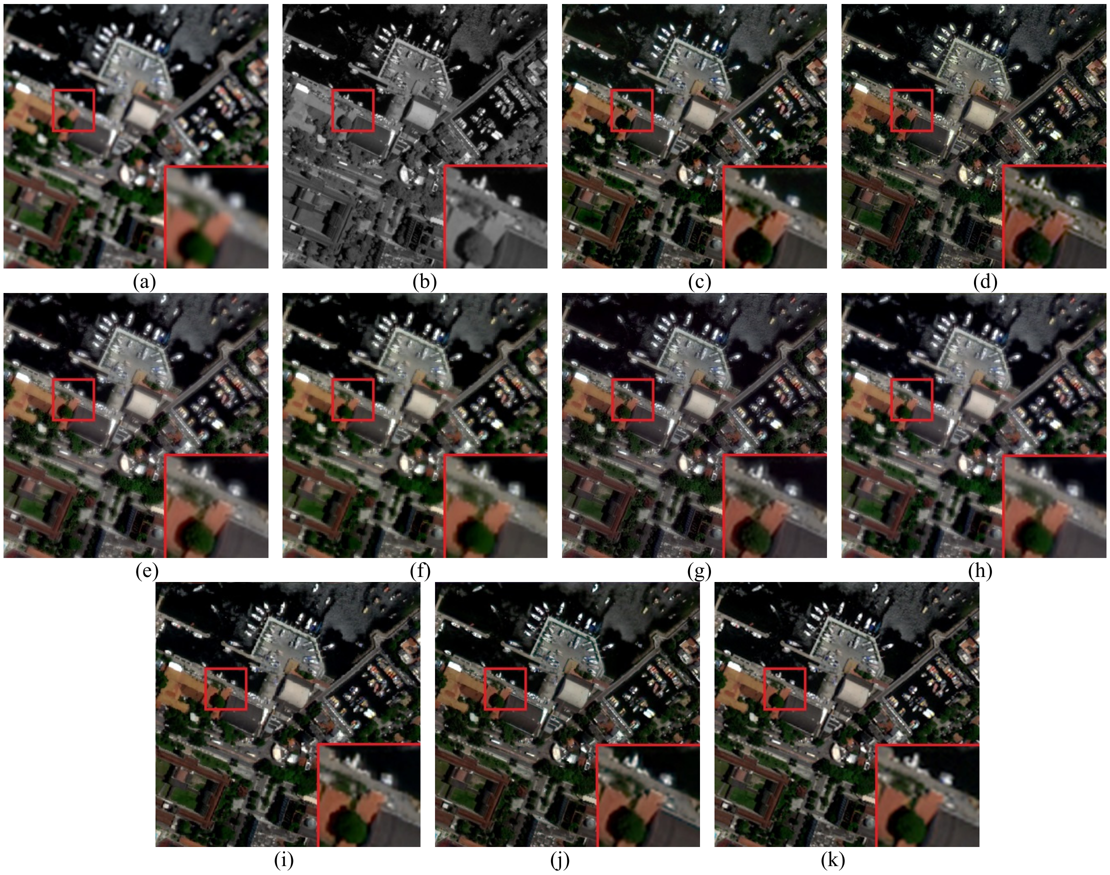


Fig. 11. Comparison of the pan-sharpening results obtained by different methods (WorldView3 image). (a) LR MS image. (b) PAN image. (c)–(j) pan-sharpening results of GS, GFP, MMP, L12, Seg_GLP, DRPNN, PNN, and the proposed method

suffers from artifacts in texture-complex regions. The MMP, L12, and DRPNN method results are blurry. The results of the Seg_GLP, PNN, and the proposed methods look more similar to the ground truth. The objective evaluation in Table VIII shows that the proposed method performs the best regarding most indexes including ERGAS, CC, and RMSE. This example shows that the proposed method also performs well on WorldView3 images.

TABLE VII
OBJECTIVE PERFORMANCE OF THE PAN-SHARPENING METHODS IN FIG. 11

Method	GS	GFP	MMP	L12	Seg_GLP	DRPNN	PNN	Proposed
D_λ	0.0525	0.1063	0.0307	0.0693	0.0462	0.0642	0.0855	**0.0290**
Ds	0.0190	0.0337	0.0957	0.1882	0.0256	0.1250	0.0760	**0.0179**
QNR	0.9295	0.8636	0.8765	0.7555	0.9292	0.8188	0.8450	**0.9536**



(a) (b) (c) (d)
(e) (f) (g) (h)
(i) (j) (k)

Fig. 12. Comparison of pan-sharpening results obtained by different methods (downsampled WorldView3 image). (a) LR MS image. (b) PAN image. (c)–(j) Pan-sharpening results of GS, GFP, MMP, L12, Seg_GLP, DRPNN, PNN, and the proposed method. (k) Reference image.

TABLE VIII
OBJECTIVE PERFORMANCE OF THE PAN-SHARPENING METHODS IN FIG. 12

Method	GS	GFP	MMP	L12	Seg_GLP	DRPNN	PNN	Proposed
ERGAS	4.1077	4.8126	4.1898	4.1067	3.9496	4.4064	4.0012	**3.9111**
SAM	3.2959	3.1599	2.9630	2.2933	2.6335	2.4630	**2.1680**	3.1247
CC	0.9623	0.9433	0.9653	0.9630	0.9589	0.9592	0.9650	**0.9667**
Q4	**0.8607**	0.8283	0.8369	0.8402	0.8476	0.8226	0.8498	0.8586
RMSE	45.9789	52.3946	46.4166	49.0402	43.6620	48.3781	45.1543	**43.5251**

Objective evaluation of the performance on the test data set (80 image patches, 20 patches for each sensor) is shown in Table IX, and the running time on CPU is shown in Table X. It should be noted that PNN is fine-tuned on the test data set for 50 epochs, which is a default parameter provided by the authors in the source code. The running time of PNN also includes the time for fine-tuning. It can be seen that the proposed method performs the best for most indexes with reference. As for full resolution indexes, the PNN method performs the best for GF2 and IKONOS images, and the proposed method perform the best for QuickBird and WorldView3 images. As for time cost, for 100 × 100-pixel

TABLE IX
OBJECTIVE PERFORMANCE OF THE PAN-SHARPENING METHODS ON TEST DATA SET

sensor	method	quality with reference					quality with no reference		
		Cc	Ergas	Q4	RMSE	SAM	Dl	Ds	QNR
GF2	GS	0.9184	2.1210	0.8202	22.3545	1.8130	0.1304	0.0482	0.8306
	GFP	0.9327	1.7642	0.7888	18.5908	1.5956	0.1933	0.0966	0.7338
	MMP	0.9576	1.4950	0.8031	15.6883	1.3928	0.1149	0.0428	0.8479
	L12	0.9598	1.4497	0.7976	15.0998	1.3077	**0.0561**	0.0798	0.8680
	Seg_GLP	0.9588	1.4419	0.8317	14.9989	1.3476	0.1137	0.0403	0.8539
	DRPNN	0.9428	1.6504	0.7620	17.1544	1.6584	0.0696	0.0876	0.8492
	PNN	0.9588	1.4216	0.8420	14.9768	1.3612	0.0668	**0.0396**	**0.8964**
	BDPN	**0.9693**	**1.2493**	**0.8757**	**13.3310**	**1.3067**	0.0643	0.0645	0.8753
ikonos	GS	0.9306	2.6077	0.7421	45.7783	3.5291	0.1628	0.0688	0.7814
	GFP	0.9240	2.2487	0.7465	40.7870	3.3022	0.2017	0.1615	0.6707
	MMP	0.9400	2.2380	0.6823	40.6271	2.9361	0.1082	0.0617	0.8381
	L12	0.9294	2.4037	0.6116	44.8779	3.0838	0.0652	0.0706	0.8703
	Seg_GLP	0.9434	2.0729	0.7651	37.1087	2.6932	0.1190	0.0684	0.8222
	DRPNN	0.9380	2.1604	0.7108	38.8512	2.5552	0.0632	0.0640	0.8768
	PNN	0.9212	4.4292	0.6944	66.2860	2.8828	**0.0540**	**0.0412**	**0.9072**
	BDPN	**0.9521**	**1.9294**	**0.8518**	**33.1365**	**2.5533**	0.0734	0.0441	0.8857
quickbird	GS	0.9184	2.1763	0.7593	27.1841	2.6086	0.0745	0.0661	0.8671
	GFP	0.7889	2.5345	0.7283	29.4479	3.2919	0.1512	0.2137	0.6672
	MMP	0.9214	1.9551	0.7157	26.0510	2.4005	0.0470	0.0425	0.9130
	L12	0.8509	2.1924	0.6563	29.7305	3.0609	0.1084	0.1415	0.6563
	Seg_GLP	0.9250	1.8978	0.7960	24.4431	2.2360	0.0501	0.0395	0.9136
	DRPNN	**0.9320**	1.6416	0.8400	20.5072	1.9640	0.0396	**0.0360**	0.9256
	PNN	0.9032	3.0864	0.7316	34.7976	2.3944	0.0332	0.0464	0.9219
	BDPN	0.9314	**1.5339**	**0.9120**	**16.7969**	**1.7905**	**0.0273**	0.0483	**0.9257**
Worldview3	GS	0.9383	2.8005	0.7620	59.9301	3.0433	0.0525	**0.0260**	0.9229
	GFP	0.9248	2.8015	0.7787	57.4865	3.4979	0.1252	0.0626	0.8205
	MMP	0.9423	2.6680	0.7282	57.5161	3.1524	0.0680	0.0451	0.8899
	L12	0.9420	2.5818	0.7315	55.8698	2.8545	**0.0325**	0.0782	0.8916
	Seg_GLP	0.9458	2.4013	0.7801	52.0450	2.8826	0.0560	0.0297	0.9164
	DRPNN	0.9324	2.7604	0.7068	58.2578	2.9944	0.0732	0.1208	0.8168
	PNN	0.9040	5.2608	0.7956	114.55	2.9720	0.0354	0.0508	0.9156
	BDPN	**0.9632**	**1.9394**	**0.8589**	**44.0779**	**2.6418**	0.0351	0.0399	**0.9264**

TABLE X
CPU TIME OF THE PAN-SHARPENING METHODS ON TEST DATA SET

Method	GS	GFP	MMP	L12	Seg_GLP	DRPNN	PNN	Proposed
Time(s)	2.5874	16.6390	37.8930	271.0849	8.7848	369.6774	558.8031	154.1261

image at MS resolution, each image takes less than 2 s, which is faster than DRPNN and PNN, and acceptable compared to traditional methods. When running on GPU, the proposed method finishes the test data set in 5 s, which verifies the efficiency and practicableness of the proposed method.

V. Conclusion

In this paper, an end-to-end deep learning network called BDPN is introduced to solve the pan-sharpening problem in producing high spatial and spectral resolution images. BDPN automatically extracts multilevel spatial details from a PAN image and injects them into an upsampled MS image, producing a pan-sharpened image with excellent spectral and spatial quality. Our proposed method is compared with several widely accepted pan-sharpening methods, and results on images from different sensors verify the superiority of the proposed method. However, limited by the structure, the proposed network can only be used to process MS and PAN images whose resolutions differ by 2^n times. Our future work will explore on extending the network to MS and PAN images with any level of scaling factors.

References

[1] G. Masi, "Pansharpening by convolutional neural networks," *Remote Sens.*, vol. 8, no. 7, p. 594, 2016.

[2] 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.

[3] P. S. Chavez, Jr., and A. Y. Kwarteng, "Extracting spectral contrast in landsat thematic mapper image data using selective principal component analysis," *Photogramm. Eng. Remote Sens.*, vol. 55, no. 3, pp. 339–348, 1989.

[4] 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.

[5] T.-M. Tu, P. S. Huang, C.-L. Hung, and C.-P. Chang, "A fast intensity-hue-saturation fusion technique with spectral adjustment for IKONOS imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 1, no. 4, pp. 309–312, Oct. 2004.

[6] 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.

[7] 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.

[8] J. Liu, Y. Hui, and P. Zan, "Locally linear detail injection for pansharpening," *IEEE Access*, vol. 5, pp. 9728–9738, 2017.

[9] H. R. Shahdoosti and N. Javaheri, "Pansharpening of clustered MS and Pan images considering mixed pixels," *IEEE Trans. Geosci. Remote. Lett.*, vol. 14, no. 6, pp. 826–830, Jun. 2017.

[10] B. Aiazzi, L. Alparone, S. Baronti, and A. Garzelli, "Context-driven fusion of high spatial and spectral resolution images based on oversampled multiresolution analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 10, pp. 2300–2312, Oct. 2002.

[11] J. Nunez, X. Otazu, O. Fors, A. Prades, V. Pala, and R. Arbiol, "Multiresolution-based image fusion with additive wavelet decomposition," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1204–1211, May 1999.

[12] S. Li, J. T. Kwok, and Y. Wang, "Using the discrete wavelet frame transform to merge landsat TM and SPOT panchromatic images," *Inf. Fusion*, vol. 3, no. 1, pp. 17–23, 2002.

[13] F. Nencini, A. Garzelli, S. Baronti, and L. Alparone, "Remote sensing image fusion using the curvelet transform," *Inf. Fusion*, vol. 8, no. 2, pp. 143–156, 2007.

[14] 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.

[15] F. Fang, F. Li, C. Shen, and G. Zhang, "A variational approach for pan-sharpening," *IEEE Trans. Image Process.*, vol. 22, no. 7, pp. 2822–2834, Jul. 2013.

[16] 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.

[17] 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.

[18] M. Ghahremani and H. Ghassemian, "A compressed-sensing-based pan-sharpening method for spectral distortion reduction," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 4, pp. 2194–2206, Apr. 2016.

[19] H. R. Shahdoosti and A. Mehrabi, "Multimodal image fusion using sparse representation classification in tetrolet domain," *Digit. Signal Process.*, vol. 79, pp. 9–22, Aug. 2018.

[20] F. Palsson, J. R. Sveinsson, and M. O. Ulfarsson, "A new pansharpening algorithm based on total variation," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 1, pp. 318–322, Jan. 2014.

[21] D. Zeng, Y. Hu, Y. Huang, Z. Xu, and X. Ding, "Pan-sharpening with structural consistency and $\ell_{1/2}$ gradient prior," *Remote Sens. Lett.*, vol. 7, no. 12, pp. 1170–1179, 2016.

[22] M. Joshi and A. Jalobeanu, "MAP estimation for multiresolution fusion in remotely sensed images using an IGMRF prior model," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 3, pp. 1245–1255, Mar. 2010.

[23] M. Xu, H. Chen, and P. K. Varshney, "An image fusion approach based on Markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 12, pp. 5116–5127, Dec. 2011.

[24] J. Xie, L. Xu, and E. Chen, "Image denoising and inpainting with deep neural networks," *Adv. Neural Inf. Process. Syst.*, 2012, pp. 1–9.

[25] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.

[26] G. Scarpa, S. Vitale, and D. Cozzolino, "Target-adaptive CNN-based pansharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 9, pp. 5443–5457, Sep. 2018.

[27] A. Azarang and H. Ghassemian, "A new pansharpening method using multi resolution analysis framework and deep neural networks," in *Proc. 3rd Int. Conf. Pattern Recognit. Image Anal. (IPRIA)*, Apr. 2017, pp. 1–6.

[28] Y. Wei, Q. Yuan, H. Shen, and L. Zhang, "Boosting the accuracy of multispectral image pansharpening by learning a deep residual network," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1795–1799, Oct. 2017.

[29] Q. Yuan, Y. Wei, X. Meng, H. Shen, and L. Zhang, "A multiscale and multidepth convolutional neural network for remote sensing imagery pan-sharpening," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 3, pp. 978–989, Mar. 2018.

[30] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2014, pp. 184–199.

[31] 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.*, Jun. 2016, pp. 1646–1654.

[32] C. Ledig *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. CVPR.*, 2017, vol. 2, no. 3, pp. 4681–4690.

[33] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR) Workshops*, Jul. 2017, vol. 1, no. 2, pp. 136–144.

[34] J. Kim, J. K. Lee, and K. M. Lee, "Deeply-recursive convolutional network for image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 1637–1645.

[35] C. Dong, C. C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," in *Proc. Eur. Conf. Comput. Vis.*, Cham, Switzerland: Springer, 2016, pp. 391–407.

[36] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Deep laplacian pyramid networks for fast and accurate super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, vol. 2, no. 3, pp. 624–632.

[37] W. Shi *et al.*, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 1874–1883.

[38] W. Shi *et al.* (2016). "Is the deconvolution layer the same as a convolutional layer?" [Online]. Available: https://arxiv.org/abs/1609.07009

[39] S. Nah, T. H. Kim, and K. M. Lee, "Deep multi-scale convolutional neural network for dynamic scene deblurring," in *Proc. CVPR*, 2017, vol. 1, no. 2, pp. 3883–3891.

[40] L. Wald, *Data Fusion: Definitions and Architectures—Fusion of Images of Different Spatial Resolutions*. Paris, France: Les Presses de l' École des Mines, 2002.

[41] L. Alparone, B. Aiazzi, S. Baronti, A. Garzelli, F. Nencini, and M. Selva, "Multispectral and panchromatic data fusion assessment without reference," *Photogramm. Eng. Remote Sens.*, vol. 74, no. 2, pp. 193–200, Feb. 2008.

[42] L. Wald, T. Ranchin, and M. Mangolini, "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, 1997.

[43] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.*, vol. 9, no. 3, pp. 81–84, Mar. 2002.

[44] F. Palsson, J. R. Sveinsson, J. A. Benediktsson, and H. Aanaes, "Classification of pansharpened urban satellite images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 1, pp. 281–297, Feb. 2012.

[45] T. Ranchin and L. Wald, "Fusion of high spatial and spectral resolution images: The ARSIS concept and its implementation," *Photogramm. Eng. Remote Sens.*, vol. 66, no. 1, pp. 49–61, Jan. 2000.

[46] R. H. Yuhas, A. F. H. Goetz, and J. W. Boardman,"Discrimination among semi-arid landscape endmembers using the spectral angle mapper (SAM) algorithm," in *Proc. Ann. JPL Airborne Geosci. Workshop*, vol. 1, R. O. Green, Ed. 1992, pp. 147–149.

[47] L. Alparone, S. Baronti, A. Garzelli, and F. Nencini, "A global quality measurement of pan-sharpened multispectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 1, no. 4, pp. 313–317, Oct. 2004.

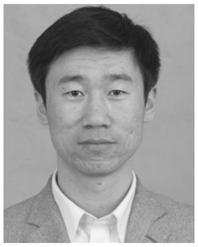
[48] B. Aiazzi, S. Baronti, and M. Selva, "Improving component substitution pansharpening through multivariate regression of MS+Pan data," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3230–3239, Oct. 2007.

[49] J. Liu and S. Liang, "Pan-sharpening using a guided filter," *Int. J. Remote Sens.*, vol. 37, no. 8, pp. 1777–1800, 2016.

[50] X. Kang, S. Li, and J. A. Benediktsson, "Pansharpening with matting model," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 5088–5099, Aug. 2014.

[51] R. Restaino, M. D. Mura, G. Vivone, and J. Chanussot, "Context-adaptive pansharpening based on image segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 753–766, Feb. 2017.

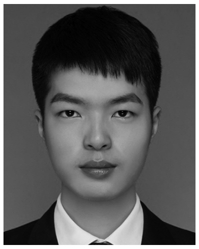
[52] Desktop, ESRI ArcGIS, *Release 10*. Redlands, CA, USA: Environ. Syst. Res. Inst., vol. 437, 2011, p. 438.



Yongjun Zhang was born in 1975. He received the B.S., M.S., and Ph.D. degrees from Wuhan University (WHU), Wuhan, China, in 1997, 2000, and 2002, respectively.

He is currently a Professor of photogrammetry and remote sensing with the School of Remote Sensing and Information Engineering, WHU. His research interests include aerospace and low-attitude photogrammetry, image matching, combined block adjustment with multisource data sets, integration of LiDAR point clouds and images, and 3-D city reconstruction.

Dr. Zhang is the Winner of the second-class National Science and Technology Progress Award in 2017. He has been supported by the Changjiang Scholars Program from the Ministry of Education of China in 2017, the China National Science Fund for Excellent Young Scholars in 2013, and the New Century Excellent Talents in University from the Ministry of Education of China in 2007.



Chi Liu was born in 1994. He received the B.S. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2015, where he is currently pursuing the Ph.D. degree in photogrammetry and remote sensing.

He is involved in high-spatial-resolution remote sensing image processing, including pan-sharpening, color balancing for remote sensing imagery, and change detection.

Mingwei Sun was born in 1982. He received the B.S. degree from the Wuhan University of Technology, Wuhan, China, in 2004, and the Ph.D. degree from Wuhan University, Wuhan, in 2009.

He is currently an Associate Professor of photogrammetry and remote sensing with the School of Remote Sensing and Information Engineering, Wuhan University. His research interests include industrial and aerial photogrammetry, 3-D reconstruction of cultural relics and buildings, automatic orthoimagery mosaicking, true orthophoto production, and parallel computing.

Yangjun Ou was born in 1994. She received the B.S. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2016, where she is currently pursuing the Ph.D. degree in photogrammetry and remote sensing.

Her research interests include image retrieval, computer vision, and pattern recognition.