Spatial–Spectral Fusion by Combining Deep Learning and Variational Model

Huanfeng Shen, *Senior Member, IEEE*, Menghui Jiang, Jie Li, *Member, IEEE*, Qiangqiang Yuan, *Member, IEEE*, Yanchong Wei, and Liangpei Zhang, *Fellow, IEEE*

***Abstract*— In the field of spatial–spectral fusion, the variational model-based methods and the deep learning (DL)-based methods are state-of-the-art approaches. This paper presents a fusion method that combines the deep neural network with a variational model for the most common case of spatial–spectral fusion: panchromatic (PAN)/multispectral (MS) fusion. Specifically, a deep residual convolutional neural network (CNN) is first trained to learn the gradient features of the high spatial resolution multispectral image (HR-MS). The image observation variational models are then formulated to describe the relationships of the ideal fused image, the observed low spatial resolution multispectral image (LR-MS) image, and the gradient priors learned before. Then, fusion result can then be obtained by solving the fusion variational model. Both quantitative and visual assessments on high-quality images from various sources demonstrate that the proposed fusion method is superior to all the mainstream algorithms included in the comparison, in terms of overall fusion accuracy.**

***Index Terms*— Deep learning (DL), gradient network, model-based optimization, spatial–spectral fusion.**

I. Introduction

SPATIAL–SPECTRAL fusion [1] is an important approach in remote sensing image fusion. It is aimed at obtaining a fused image with both high spatial and high spectral resolutions. Classic spatial–spectral fusion methods include panchromatic (PAN)/multispectral (MS) fusion, PAN/hyperspectral (HS) fusion, and MS/HS fusion. Among them, PAN/MS fusion, which is also called pansharpening, is the most common case, and aims at integrating the geometrical details of the high spatial resolution panchromatic image (HR-PAN) and the rich spectral information of the low spatial resolution multispectral image (LR-MS) to obtain a high spatial resolution multispectral image (HR-MS) [2]. To date, a large number of PAN/MS fusion methods have been proposed, which can be generally divided into four major branches [3]: 1) the component substitution (CS)-based methods; 2) the multiresolution analysis (MRA)-based methods; 3) the variational model-based methods; and 4) the deep learning (DL)-based methods.

The CS-based methods are the simplest and the most widely used methods in PAN/MS fusion [3]. The traditional understanding of this approach is to first project the upsampled LR-MS image into a new space. The component representing the spatial information is then replaced by the HR-PAN image. The fused image is obtained by a final inverse projection. The representative methods in this class include the intensity–hue–saturation (IHS) methods [4], the principal component analysis (PCA) methods [5], the Gram–Schmidt (GS) methods [6], and the adaptive GS (GSA) methods [7]. Unfortunately, the traditional methods of this class can produce significant spectral distortions when rendering the spatial details [8]. The many advanced CS-based pansharpening methods proposed in recent years have changed this situation, and they perform well in both the spatial and spectral fidelity [9]–[12]. For example, the data fitting method proposed by Xu *et al.* [12] was found to be effective in reducing spectral distortion, without blurring the spatial details.

Differing from the CS-based methods, the MRA-based methods produce fewer spectral distortions in the fused image, but they are sensitive to the spatial distortions. In general, this method first separates the high-pass information and low-pass information of images by wavelet transform, Laplacian pyramid, etc., and then injects the spatial structure information extracted from the HR-PAN image into the upsampled LR-MS image to obtain the fused image. Examples of this type of method include the high-pass filter (HPF) [13], decimated wavelet transform (DWT) [14], smoothing filter-based intensity modulation (SFIM) [15], and the generalized Laplacian pyramid with modulation transfer function (MTF) matched filter (MTF-GLP) [16]. Due to the widespread use and high efficiency of the CS-based methods and the MRA-based

Manuscript received October 30, 2018; revised January 30, 2019; accepted February 23, 2019. Date of publication April 11, 2019; date of current version July 22, 2019. This work was supported by the National Natural Science Foundation of China under Grant 41701400 and Grant 61671334. *(Corresponding author: Jie Li.)*

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

M. Jiang is with the School of Resource and Environmental Science, Wuhan University, Wuhan 430079, China (e-mail: 15271848120@163.com).

J. Li is with the School of Geodesy and Geomatics, Wuhan University, Wuhan 430079, China (e-mail: aaronleecool@whu.edu.cn).

Q. Yuan is with the School of Geodesy and Geomatics, Wuhan University, Wuhan 430079, China, and also with the Collaborative Innovation Center of Geospatial Technology, Wuhan University, Wuhan 430079, China (e-mail: qqyuan@sgg.whu.edu.cn).

Y. Wei is with MCFLY Technology, Beijing 100102, China (e-mail: decriphonie@gmail.com).

L. Zhang is with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China, and also with 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 paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2019.2904659

methods, they can be considered as reference state-of-the-art algorithms for pansharpening [8]. However, neither of these approaches set up an explicit relation model between the observed images and the desired images.

The variational model-based were developed after the CS-based methods and the MRA-based methods, and considered as a second-generation models [17].The variational model-based methods regard the fusion process as an ill-posed inverse optimization problem and construct the energy function based on the HR-PAN image, the LR-MS image, and the ideal fused image. The construction is usually based on observation models [18]–[20] and the sparse representation [21], [22]. An iterative optimization algorithm, such as the gradient descent algorithm [23], the conjugate gradient algorithm [24], the split Bregman iteration algorithm [25], or the alternating direction method of multipliers (ADMM) algorithm [26]–[29], is then used to solve the fusion model to obtain the fused image. Due to the solid mathematical foundation and the strict logic system of this kind of methods, the variational model-based methods generally produce precise fusion results. However, these methods are usually very sensitive to the unavoidable inaccuracies of the adopted observational model [30]. The linear simulation from the observed and fusion images is still a limitation and cannot fully describe the potentially complex relationship between the observed and fusion images. For instance, according to the spectral response of recent spaceborne sensors, no obvious relationship exists between the PAN and MS input modalities, and such a relationship is certainly not linear [1]. Using the linear simulation between the desired HR-MS image and the observed HR-PAN image would thus restrict the fusion quality.

With the specific capability of feature extraction and learning [31], DL tends to deliver a promising performance in describing the nonlinear relationship between data, which have been applied to various remote sensing problems [32]–[36]. The DL-based pansharpening methods have been proposed in recent years and can be regarded as another new branch of the PAN/MS fusion methods [3]. In general, the DL-based pansharpening methods train a network model to learn the mapping between the LR-MS image, the HR-PAN image, and the ideal HR-MS image, and then obtain the fused image by inputting the LR-MS image and the HR-PAN image into the learned network model. Masi *et al.* [37] stacked the HR-PAN image with the upsampled LR-MS image to form an input volume and used a convolutional neural network (CNN) to learn the mapping between the input volume and the HR-MS image. Wei *et al.* [38] adopted deep residual learning to learn the mapping. Yuan *et al.* [39] proposed a multiscale CNN for PAN/MS fusion, where each layer was constituted by filters with different sizes for multiscale features. It is well known that the fusion methods of pansharpening make tradeoff between the spatial enhancement and the spectral fidelity. The DL-based methods usually achieve superior spectral fidelity, which means that the spatial enhancement in the fused result is insufficient, that is to say, a more or less fuzzy result will occur.

As discussed above, on the one hand, the nonlinear mapping capacity of the DL-based methods can make up for the deficiency of the variational model-based approach; on the other hand, the variational model-based methods have strong spatial enhancement capabilities, unlike the DL-based methods. It is therefore of interest to combine a DL-based method with a variational model-based fusion method.

In this paper, we propose a fusion mothed that combines a deep residual gradient CNN with a variational model-based fusion method. Specifically, a deep residual gradient CNN is first trained to generate the gradient information of the HR-MS image. The generated gradient priors are then plugged into a variational model-based optimization method, which can simultaneously utilize the precise characteristics of the variational model-based method and the feature learning capacity of DL to obtain a high-quality fusion performance.

The rest of this paper is organized as follows. Section II gives a detailed description of the proposed method. In Section III, the experiments and discussions are presented. The conclusions and future research directions are summarized in Section IV.

II. Proposed Method

A. Variational Model-Based Fusion Framework

In general, the energy function of the variational model-based spatial–spectral fusion methods can be summed up as three terms: 1) the spectral fidelity model; 2) the spatial enhancement model; and 3) the prior model. This can be generally represented as the following expression:

$$E(\boldsymbol{X}) = p_{\text{spectral}}(\boldsymbol{X}, \boldsymbol{Y}) + p_{\text{spatial}}(\boldsymbol{X}, \boldsymbol{Z}) + p_{\text{prior}}(\boldsymbol{X}) \quad (1)$$

where $\boldsymbol{X}$ denotes the ideal fused image, $\boldsymbol{Y}$ denotes the high spectral resolution but low spatial resolution image, and $\boldsymbol{Z}$ denotes the high spatial resolution but low spectral resolution image. More specifically, in PAN/MS fusion, $\boldsymbol{X} \in \mathbb{R}^{\text{MN}\times S}$ represents the desired HR-MS image; M, N, and S are the width, the height, and the band number of the ideal image. $\boldsymbol{Y} \in \mathbb{R}^{\text{mn}\times S}$ denotes the LR-MS image, and $\boldsymbol{Z} \in \mathbb{R}^{\text{MN}\times 1}$ denotes the HR-PAN image. M/m is the spatial resolution ratio of the LR-MS image to the HR-MS image.

The first term is the spectral fidelity model, which represents the relationship between the ideal fused image and the LR-MS image. The second term is the spatial enhancement model, which relates the ideal fused image to the HR-PAN image. The last term is the regularization term that imposes constraints on the ideal fused image, where Laplacian priors [40], Huber-Markov priors [18], total variation (TV) priors [41], a nonlocal priors [42], and low-rank priors [43] have been proposed in the many variational model-based PAN/MS fusion methods.

B. Combining Deep Learning and Variational Model

The existing papers that have combined DL and variational model-based optimization have mainly trained a set of fast and effective regularization terms by DL, such as the denoisers [31]. However, the spatial–spectral fusion focuses on mining and integrating the high-resolution information into the desired image, meaning that it is reasonable to incorporate a DL-based spatial enhancement term into the variational model-based methods. In general, the spatial enhancement term is constructed based on two assumptions. The first

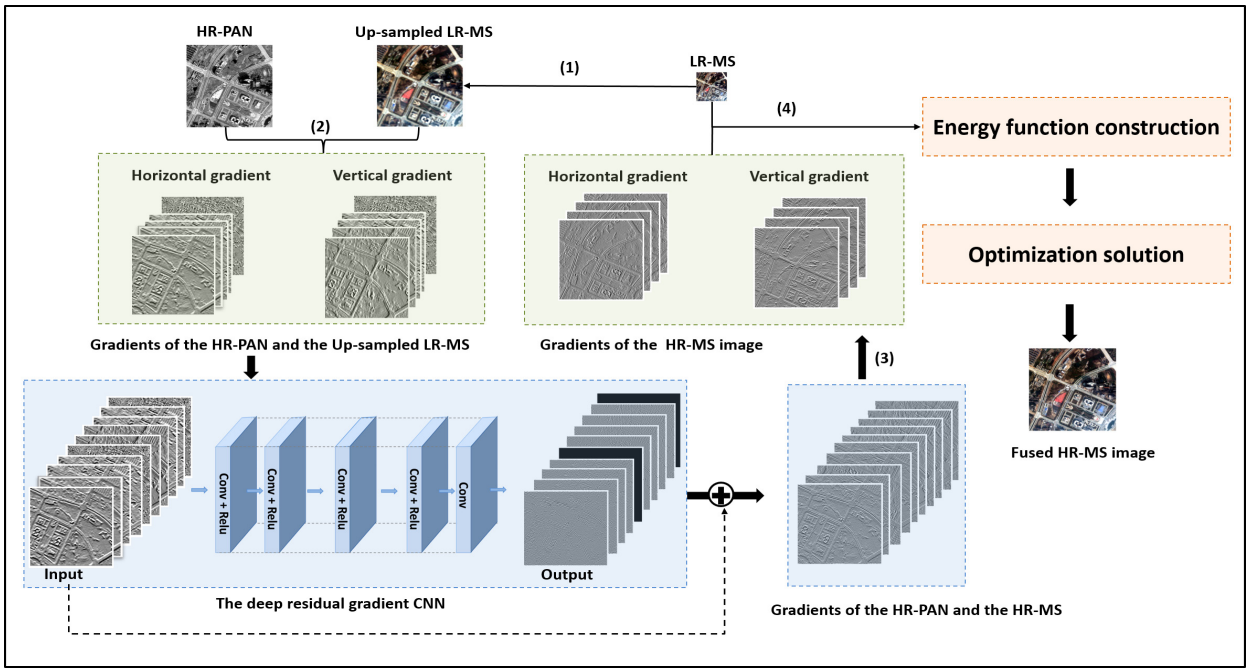


Fig. 1. Flowchart of the proposed method. (1) Upsampling process of the LR-MS image. (2) Stack the upsampled LR-MS image and the HR-PAN image in the spectral dimension first, and then obtain the horizontal and vertical gradients of the stacked volume. (3) Select the gradient of the HR-MS image from the stacked gradients. (4) Construct the fusion variational model with the original LR-MS image and the gradients of the HR-MS image learned from the gradient CNN.

assumes spectral degradation between the HR-MS image and the HR-PAN image, i.e., the HR-PAN image is assumed to be a linear combination of the various bands of the HR-MS image [18], [44]. The second assumes that the spatial structures of the ideal HR-MS image are approximately consistent with the HR-PAN image [30], including gradient features [45], [46] and wavelet coefficients [47]. In the first assumption, a complex nonlinear relationship exists between the HR-PAN image and the spectral bands of the HR-MS image [1], which is unsuitable to express as a linear function. In the second assumption, a consistency constraint fidelity term is usually used to retain the high spatial information such as gradient features. The relationship between the gradients of the HR-PAN image and the HR-MS image can also be nonlinear since the gradients are usually obtained by linear operators on the images. Therefore, nonlinearity can be more suitable to describe the relationship of the images' spatial structures.

With the capability in feature extraction and mapping learning, DL has great potential to model this complex nonlinear relationship [48]. It is therefore of interest to integrate DL into the variational model construction. In this paper, we train the mapping from the gradient of the HR-PAN image and the LR-MS image to that of the HR-MS image via a deep residual CNN, as the gradient features can reflect the spatial structures in images most directly. We then construct a variational model based on the learned gradient guidance. Details are provided in Fig. 1.

1) Gradient Generation Using the Deep Residual CNN: Instead of directly mapping the HR-PAN image and the LR-MS image to the HR-MS image, we stack the gradient of the HR-PAN image and that of the upsampled LR-MS image in the spectral dimension to form the input volume of the proposed network. On the one hand, using the gradient directly makes the goal of the network more clear, which is to obtain accurate spatial structures of the HR-MS image, i.e., the gradient. On the other hand, most pixel values in the gradient image will be very close to zero, and the spatial distribution of the feature maps should be very sparse, which can transfer the gradient descent process to a much smoother hyper-surface of loss to the filtering parameters, as in residual learning [34].

Instead of directly mapping the input volume to the gradient of the ideal HS-MS image, we utilize the popular residual learning strategy [49] and map the input gradient volume to the gradient residual volume. The residual learning strategy was first proposed to solve the degradation of training accuracy with increasing network depth [49]. It is assumed that if the optimal function is closer to an identity mapping, it is easier to learn the residual mapping than the original identity mapping. Different from the residual network [49] that utilizes many residual units, the proposed deep residual CNN uses one residual unit to predict the gradient residual image. Note that, prior to the residual network [49], the strategy of learning the residual image has been utilized in many tasks [50], [51]. Furthermore, when combined with mapping the gradient, the residual learning strategy further improves the sparsity of the network, which can boost the learning performance.

2) Architecture of the Deep Residual Gradient CNN: A general CNN architecture is adopted in this paper (see the deep residual gradient CNN part in Fig. 1). Inspired by the deep residual network for image denoising (DNCNN) [31], this paper presents a specific improvement of the architecture for

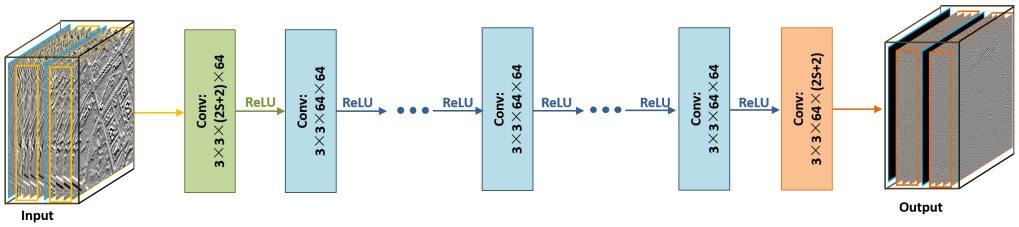


Fig. 2. Architecture of the proposed deep residual gradient CNN network.

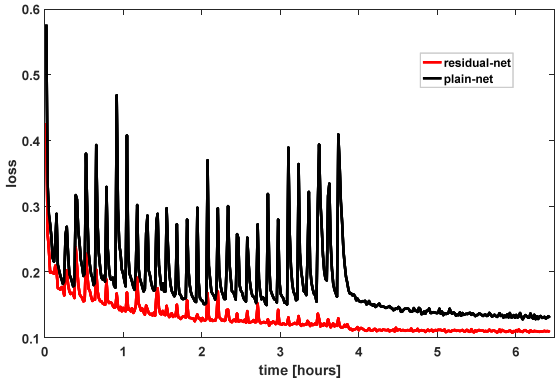


Fig. 3. Loss curves of deep residual gradient CNN and deep plain gradient CNN.

the task of gradient training. Structural details of the network can be seen in Fig. 2.

For better understanding the architecture of the proposed deep residual gradient CNN, two stacks are first introduced. Stack-1 is the stack of the upsampled LR-MS image and the HR-PAN image in the spectral dimension. Stack-2 is the stack of the ideal HR-MS image and the HR-PAN image in the spectral dimension. The input of the proposed network is the gradients of stack-1 in both horizontal and vertical directions. As displayed in the input volume of Fig. 2, the blue boxes represent gradients of the HR-PAN image; the yellow boxes represent gradients of the upsampled LR-MS image.

The output of the proposed CNN is the gradient residual image between the gradient of stack-1 and the gradient of stack-2. As displayed in the output volume of Fig. 2, the orange boxes represent the residual image between the gradient of the HR-MS image in stack-2 and the gradient of the upsampled LR-MS image in stack-1. The blue boxes represent residual images between the gradients of the HR-PAN image in stack-2 and stack-1, respectively. The same size between the output and the input makes it convenient to calculate.

To illustrate the superiority of the residual learning strategy, Fig. 3 shows the comparison of the loss curve of the proposed deep residual gradient CNN and the loss curve of deep plain gradient CNN. The architecture and input of the plain network are the same as that of the residual network, but the output is different. The output of the plain network is the gradient image of stack-2 in both horizontal and vertical directions, while the output of the residual network is the gradient residual image. The network models are trained on QuickBird images. As clearly displayed in Fig. 3, the adopted residual learning strategy not only helps to speed up the convergence but also leads to lower training errors.

As shown in Fig. 2, the proposed residual CNN has 17 blocks, which consists of three types.

1) Conv + ReLU, for block 1, 64 filters of size $3 \times 3 \times (2S + 2)$ is first employed to generate 64 feature maps, and then the ReLU is used on the generated feature maps, where S denotes the band number of the original LR-MS image.
2) Conv+ReLU, for blocks 2–16, 64 filters of size $3 \times 3 \times 64$ and the ReLU are utilized.
3) Conv + Loss, for the last block, $(2S + 2)$ filters of size $3 \times 3 \times 64$ is used to obtain the output. The mean square error MSE) is used as the loss function

$$\text{loss} = \frac{1}{2C} \sum_{k=1}^{C} \| f(\boldsymbol{In}_k) - (\boldsymbol{Tr}_k - \boldsymbol{In}_k) \|_F^2 \qquad (2)$$

where $\boldsymbol{In}$ is the input gradient, i.e., the gradient of stack-1 in horizontal and vertical directions. $\boldsymbol{Tr}$ is the ground truth gradient, i.e., the gradient of stack-2 in horizontal and vertical directions. $\boldsymbol{Tr} - \boldsymbol{In}$ is the label data of the proposed network. $f(\cdot)$ is the mapping process, $f(\boldsymbol{In}_k)$ means the output of the CNN, and C represents the training image (patch) pairs in a mini-batch, which is set 128 in this paper.

C. Learned Gradient Guidance-Based Fusion Model

The proposed fusion method combines the deep residual gradient CNN with the variational model-based framework. Specifically, the method includes three main steps. First, a deep residual gradient CNN is trained to obtain the gradient information of the HS-MS image. Variational model-based optimization is then conducted using the LR-MS image, the gradient priors, and the ideal fused image. Finally, an iterative optimization algorithm is used to solve the fusion model. Fig. 1 is the flowchart of the proposed method. As displayed in Fig. 1, a volume is first obtained by stacking the bicubic-upsampled LR-MS image and the LR-PAN image in the spectral dimension. The horizontal and vertical gradients of the stacked volume form the input volume of the deep residual gradient CNN. The output of the deep residual gradient CNN is the gradient residual image. By adding the input and output of the network, the gradient priors of the HR-MS image can be obtained. Next, the energy function of the fusion model is constructed by using the original LR-MS image, the gradient priors of the HR-MS image, and the ideal fused image. Finally,

an iterative optimization is used to solve the energy function and obtain the fused image.

The energy function constructed in the proposed method can be written as

$$\widehat{\boldsymbol{X}} = \arg\min_{\boldsymbol{X}} \frac{1}{2}||\boldsymbol{Y} - \boldsymbol{H}\boldsymbol{X}||_F^2 + \frac{\lambda}{2}\sum_{j=1}^{2}||\nabla_j \boldsymbol{X} - \boldsymbol{G}_j||_F^2 + \frac{\mu}{2}\{||\boldsymbol{Q}\boldsymbol{X}||_F^2\}. \quad (3)$$

In the first term, $\boldsymbol{H} \in \mathbb{R}^{\mathrm{mn}\times\mathrm{MN}}$is the downsampling and blurring matrix. The spectral fidelity model is constructed based on the assumption that the observed LR-MS image can be obtained by blurring, downsampling, and the noise operators performed on the HR-MS image. In the second term, $\nabla_j \in \mathbb{R}^{\mathrm{MN}\times\mathrm{MN}}$ with $j = 1, 2$ means the global first-order finite difference matrix in the horizontal and vertical directions, respectively. $\boldsymbol{G}$ denotes the gradient images of the HR-MS image learned through the deep residual gradient network, $\boldsymbol{G}_1 \in \mathbb{R}^{MN\times S}$ and $\boldsymbol{G}_2 \in \mathbb{R}^{MN\times S}$ represents the horizontal and vertical directions. It is assumed that the gradient of the ideal image is consistent with $\boldsymbol{G}$. The third term is the common Laplacian prior model, where $\boldsymbol{Q} \in \mathbb{R}^{\mathrm{MN}\times\mathrm{MN}}$ indicates the Laplacian matrix. λ and μ are adjustable parameters used to balance the relative contribution of the three terms. The variational model is associated with a quadratic regularized least-squares problem, which can be further written as

$$\boldsymbol{H}^T\boldsymbol{H} + \lambda\sum_{j=1}^{2}\nabla_j^T\nabla_j + \mu\,\boldsymbol{Q}^T\boldsymbol{Q}\boldsymbol{X} = \boldsymbol{H}^T\,\boldsymbol{Y} + \lambda\sum_{j=1}^{2}\nabla_j^T\boldsymbol{G}_j. \quad (4)$$

This function is difficult to solve due to its large dimensions. An iterative optimization algorithm is therefore selected to solve the problem. In this paper, the preconditioned conjugate gradient (PCG) method is selected due to its fast convergence and high efficiency. The solution can be achieved by invoking the PCG function in the MATLAB.

III. Experiments and Discussion

To verify the effectiveness of the proposed method, both simulated and real-data experiments are performed, as described in the following. The proposed method is compared with five mainstream algorithms from different branches: adaptive intensity–hue–saturation (AIHS) fusion method [52], MTF-GLP with high-pass modulation (HPM) injection model algorithm (MTF-GLP-HPM) [53], the two-step sparse coding model (TSSC) [21], the deep residual pansharpening network (DRPNN) [38], and DNCNN [31]-based pansharpening. In order to eliminate the effect of the different network structures, the experimentally designed DNCNN-based pansharpening has a similar network structure to the gradient network in the proposed fusion method. Among the five comparison algorithms, AIHS belongs to the CS-based family, MTF-GLP-HPM belongs to the MRA-based branch, TSSC belongs to the variational model-based methods, and the last two are DL-based methods.

To quantify the accuracy of the fusion results, four indices are used in this paper: the relative dimensionless global error in synthesis (ERGAS), the spectral angle mapper (SAM), the Q metric, and the peak-signal-to-noise-ratio (PSNR).

1) Train Data Sets: Two deep residual gradient CNN models are trained for the QuickBird sensor and the WorldView-2 sensor, respectively. When training the QuickBird gradient network model, three LR-MS images with a size of $5200 \times 4400 \times 4$, $1000 \times 1000 \times 4$, and $3600 \times 1000 \times 4$, and three HR-PAN images with a size of $20\,800 \times 17\,600$, 4000×4000, and $14\,400 \times 4000$, obtained by the QuickBird sensor, are used to obtain 102 400 patches for training and 8960 patches for validation. The size of each patch is $40 \times 40 \times 10$. When training the WorldView-2 gradient CNN model, three LR-MS images with a size of $3500 \times 3500 \times 8$, $2800 \times 3700 \times 8$, and $3200 \times 2800 \times 8$, and three HR-PAN images with a size of $14\,000 \times 14\,000$, $11\,200 \times 14\,800$, and $12\,800 \times 11\,200$, obtained by the WorldView-2 sensor, are used to obtain 51 200 patches for training and 3840 patches for validation. The size of the patches is $40 \times 40 \times 18$, and the batchsize used in both network models is 128.

2) Test Data Sets: Four data sets are employed in the simulated and real-data experiments, as follows. The gray values of each image are all normalized to [0, 1].

1) The first data set is made up of QuickBird images, which are cropped to $250 \times 250 \times 4$ to obtain the LR-MS images and 1000×1000 to obtain the HR-PAN images, respectively. The spatial resolutions of the HR-PAN images and the LR-MS images are 0.61 and 2.44 m, respectively. There are 12 pairs of images with different textures, which are used in the simulated-data experiments.
2) The second data set is made up of WorldView-2 images, which are cropped to $250 \times 250 \times 8$ to obtain the LR-MS images and 1000×1000 to obtain the HR-PAN images, respectively. The spatial resolutions of the HR-PAN image and the LR-MS images are 0.5 and 2 m, respectively. There are six pairs of images with different textures, which are used in the simulated-data experiments.
3) The third data set is a pair of IKONOS images with a size of $200 \times 200 \times 4$ and 800×800. The spatial resolutions of the HR-PAN images and the LR-MS images are 1 and 4 m, respectively. These images are used in the real-data experiments.
4) The fourth data set is a pair of WorldView-2 images with a size of $200 \times 200 \times 8$ and 800×800. The spatial resolutions of the HR-PAN image and the LR-MS image are 0.5 and 2 m, respectively. These images are used in the real-data experiments.

Note that, all the test data sets are spatially disjoint from the training and validation patches used in the network training. For example, three images used to train the QuickBird network model are of Nanchang, Yichang, and Shenzhen, while the test data set of QuickBird images is of Wuhan.

3) Parameter Setting: In (3), λ and μ are adjustable parameters used to balance the relative contribution of the three terms. Fig. 4 analyzes the sensitivity of the two adjustable parameters of the proposed method. Specifically, Fig. 4(a)–(d), respectively, shows the sensitivity analysis of λ and μ on the

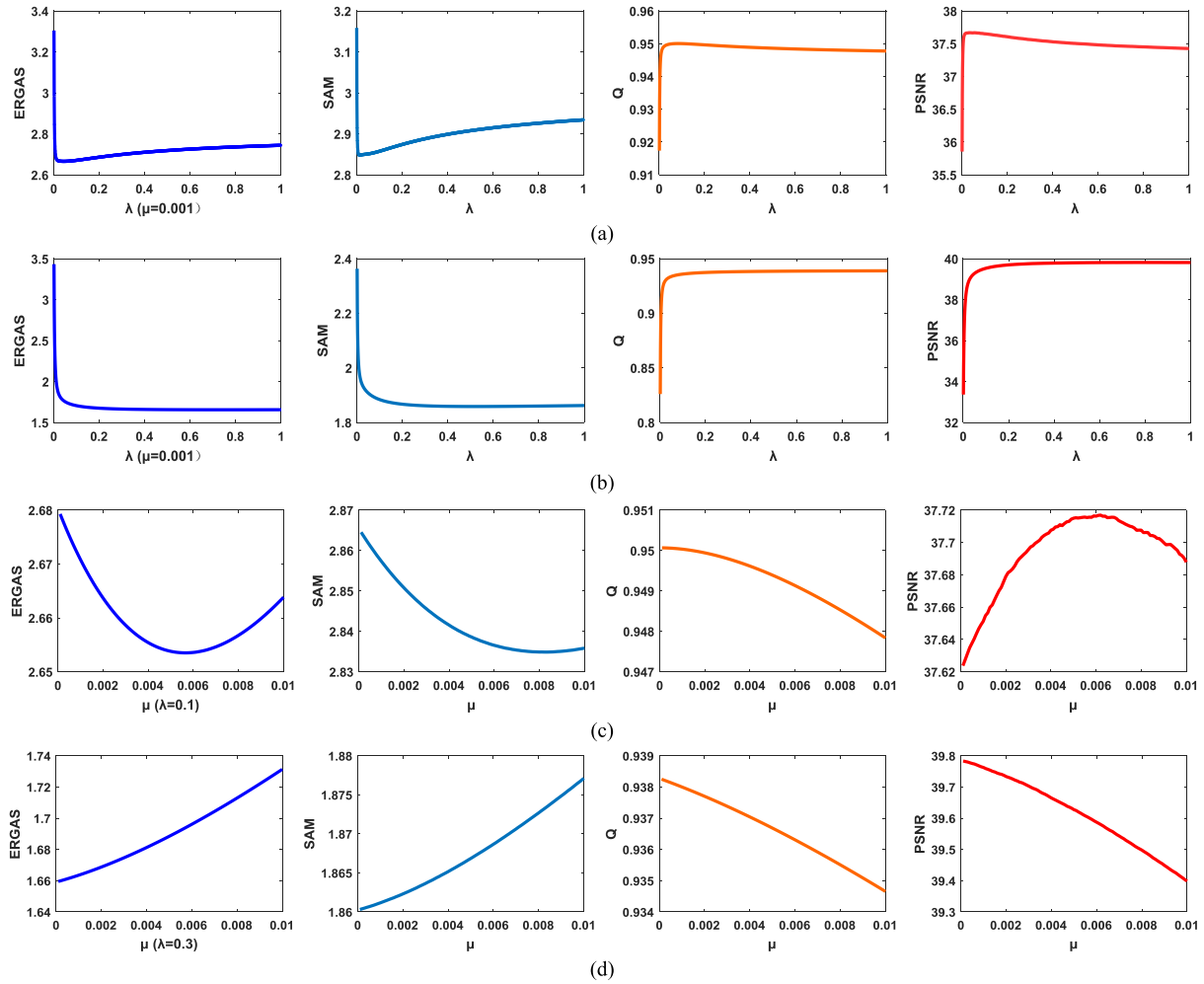


Fig. 4. Analysis of the adjustable parameters λ and μ of the proposed method. (a) Analysis of λ on the simulated experiment of QuickBird images. (b) Analysis of λ on the simulated experiment of WorldView-2 images. (c) Analysis of μ on the simulated experiment of QuickBird images. (d) Analysis of μ on the simulated experiment of WorldView-2 images.

simulated experiment of QuickBird images and WorldView-2 images, where λ is plotted at 0.01 equidistance in [0.01–1] and μ is plotted at 0.0001 equidistance in [0.0001–0.01]. As displayed in Fig. 4(a) and (b), when the parameter λ is in the range of 0.1–0.3, we can obtain favorable indicators in both the QuickBird experiment and the WorldView-2 experiment. Combining with visual inspections, we set $\lambda = 0.1$ for the QuickBird data set and $\lambda = 0.3$ for the WorldView-2 data set. As displayed in Fig. 4(c) and (d), in the range of 0–0.01, the parameter μ is relatively stable, and only causes the minor variation on all the evaluation indicators for both the QuickBird experiment and the WorldView-2 experiment. Accordingly, combining with visual inspections, we set $\lambda = 0.1$ and $\mu = 0.001$ for the simulated QuickBird experiments and the real-data IKONOS experiment, $\lambda = 0.3$ and $\mu = 0.001$ for both the simulated and real-data WorldView-2 experiments.

A. Simulated Experiments

In the simulated experiments, we first downsample the HR-PAN image to obtain the low-resolution PAN image (LR-PAN) and downsample the LR-MS image to obtain the lower resolution MS image. Then, we fuse the LR-PAN image and the lower resolution MS image to obtain the fused image. The original LR-MS image acts as a reference to evaluate the fused image both qualitatively and quantitatively.

The first series of simulated experiments are performed on the QuickBird images. Table I shows the simulated QuickBird experimental results with the average of 12 groups, where the best performance for each quality index is marked in bold and the second-best performance for each quality index is underlined. In these experiments, the SAM value of the proposed method is the second-best, and all the other indicators are higher than those of the other methods, which demonstrate that the proposed method can provide a better tradeoff between spectral information fidelity and spatial detail enhancement. In the last column of Table I, the running times of all fusion methods are listed, which represent the total time of running all groups of images in simulated experiments. Like the two DL-based method in the comparison, the gradient network of the proposed method runs fast, and the selected PCG iterative optimization algorithm efficiently solve the constructed energy function, which leads to the acceptable running time of the proposed method.

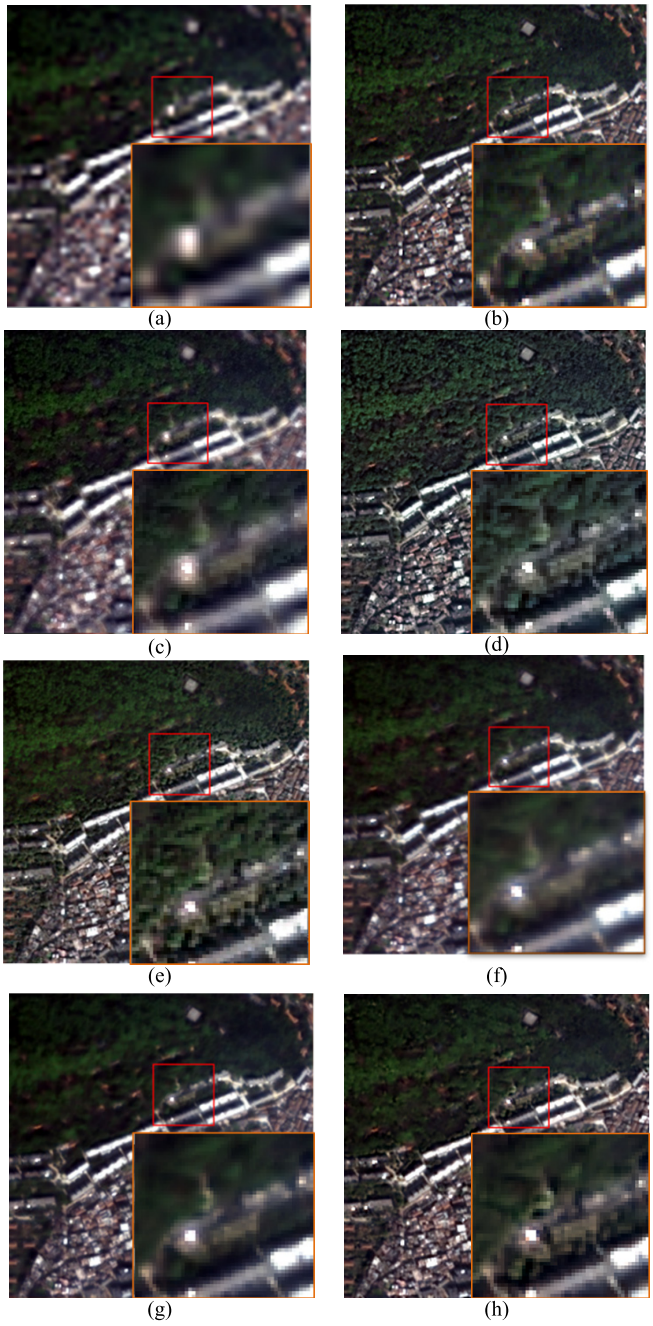


Fig. 5. Simulated fusion results for the QuickBird image. (a) Low-resolution MS image simulated by downsampling. (b) Ground truth. (c) AIHS. (d) MTF-GLP-HPM. (e) TSSC. (f) DNCNN. (g) DRPNN. (h) Proposed method.

A group of simulated fusion results is selected to be displayed as true-color images in Fig. 5. By comparing the results, it can be observed that AIHS performs poorly in both spatial detail enhancement and spectral fidelity, as shown in Fig. 5(c). For the MTF-GLP-HPM and TSSC methods, sharpened spatial features are achieved, but with severe spectral distortion, such as the vegetation area shown in Fig. 5(d) and (e). For the two CNN-based methods, they both show a good performance in spectral fidelity but are poor in spatial texture information enhancement, as can be seen in the zoomed area

TABLE I

QUANTITATIVE RESULTS FOR THE SIMULATED QUICKBIRD IMAGES (12 GROUPS)

Method\index	ERGAS(↓)	SAM(↓)	Q(↑)	PSNR(↑)	Time(s)
AIHS	3.5005	3.5989	0.8815	33.3684	3.3
MTF-GLP-HPM	3.4599	3.2430	0.9065	33.6465	3.2
TSSC	3.1377	3.1602	0.9158	34.4225	43.7
DNCNN	3.0136	2.9402	0.9144	34.7869	**1.8**
DRPNN	2.7863	**2.6903**	0.9265	35.4484	**1.8**
Proposed	**2.4598**	2.7124	**0.9441**	**36.4841**	6.5

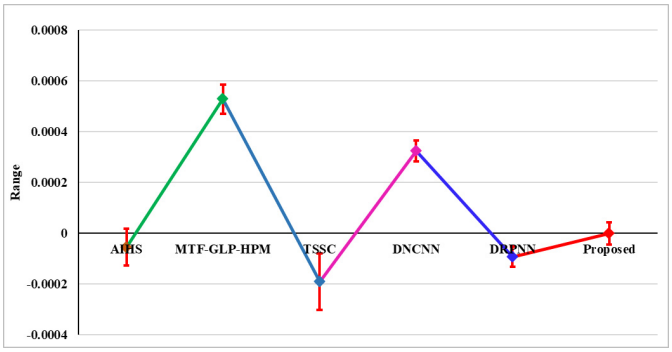


Fig. 6. Point plot of the difference images between the various fusion results and the ground truth in Fig. 5.

in Fig. 5(f) and (g). Overall, DRPNN performs better than the DNCNN-based pansharpening. The proposed method combines the respective advantages of the variational model-based method and the DL-based method, and the fused results are the closest to the ground truth, both in the fusion of the spatial details and in the preservation of the spectral fidelity [see Fig. 5(h)].

Fig. 6 shows the point plot of the difference images between the various fusion results and the ground truth in Fig. 5. Six points of different colors denote the mean values of six difference images. The closer the mean value is to zero, the better the fusion result is. The vertical lines at each point represent the error line of each difference image, which show the confidence interval of each difference image. As shown in Fig. 6, the proposed method shows the best fusion result.

To further analyze the fusion results, Fig. 7 shows the horizontal profiles of the digital values of the various fusion results and the ground truth in Fig. 5. The vertical axis named DN stands for the digital values, and the horizontal axis represents the column number. The closer the profile of the fusion result is to the profile of the ground truth, the better the fusion result is. In Fig. 7, the proposed method outperforms the five mainstream contrastive algorithms in bands and is closest to the ground truth, which illustrates the superior fidelity performance of the proposed method.

The second series of simulated experiments are performed on the WorldView-2 images. Table II shows the simulated WorldView-2 experimental results with the average of six groups, where the best performance for each quality index is marked in bold, and the second-best performance for each quality index is underlined. In these experiments, the Q value

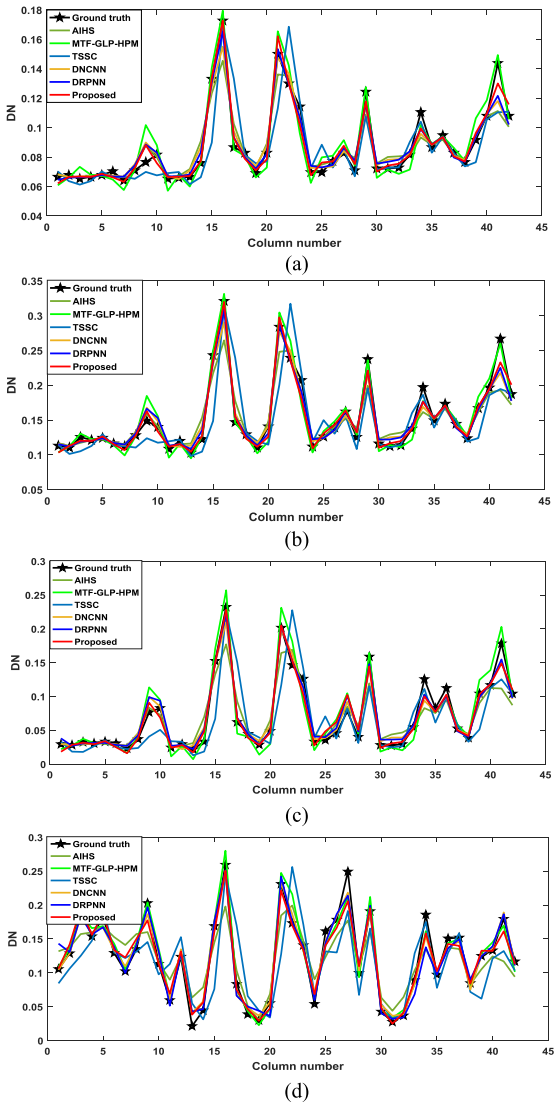


Fig. 7. Horizontal profiles of the digital values of the 124th line of the various fusion results and the ground truth in Fig. 5. (a)–(d) Bands 1, 2, 3, and 4.

TABLE II

QUANTITATIVE RESULTS FOR THE SIMULATED WORLDVIEW-2 IMAGES (SIX GROUPS)

Method\index	ERGAS(↓)	SAM(↓)	Q(↑)	PSNR(↑)	Time(s)
AIHS	2.9039	3.5143	0.8674	32.5207	2.2
MTF-GLP-HPM	2.5411	3.2740	0.9156	33.9657	2.9
TSSC	2.0803	3.0405	0.9276	35.7345	37.9
DNCNN	2.0757	2.9492	0.9303	35.7703	**1.3**
DRPNN	1.9990	2.8259	**0.9384**	35.9148	1.4
Proposed	**1.8615**	**2.7028**	0.9381	**36.6352**	6.2

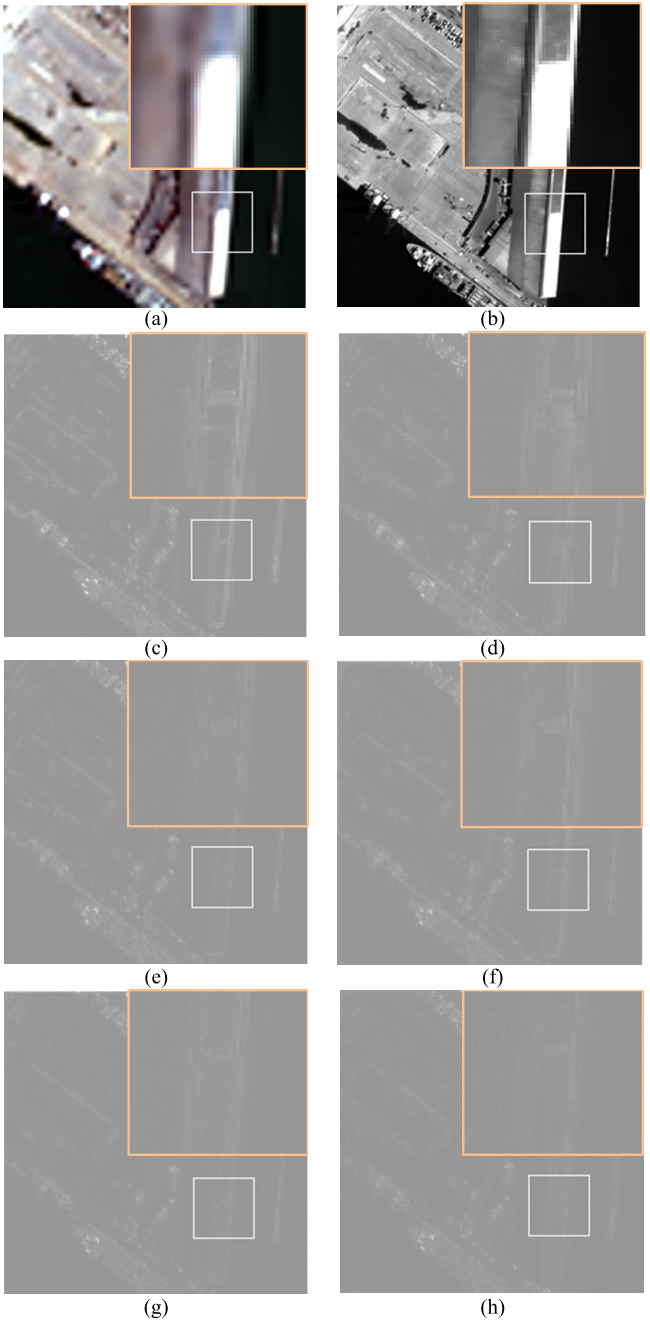


Fig. 8. Absolute difference images between the various fusion results and the ground truth. (a) LR-MS image. (b) HR-PAN image. (c) Difference image of AIHS. (d) Difference image of MTF-GLP-HPM. (e) Difference image of TSSC. (f) Difference image of DNCNN. (g) Difference image of DRPNN. (h) Difference image of the proposed method.

of the proposed method is very close to the best result, but the other indicators are all higher than those of the other methods.

A group of simulated fusion results is selected to be displayed in Fig. 8. Fig. 8 shows the absolute difference images, which is composed of the maximum values of the absolute difference image between the various fusions and the ground truth in the spectral dimension. The less the spatial information can be seen in the difference image, the better the fusion result is. For instance, the spatial structure of the ground truth is clearly visible in the AIHS difference image and MTF-GLP-HPM difference image, as seen in the zoomed area in Fig. 8(c) and (d). This means that there is still some high-resolution spatial information in the HR-PAN image that is not integrated into the fused image, and the AIHS and MTF-GLP-HPM methods are poor in spatial enhancement. The clearly reduced spatial information in the zoomed area in Fig. 8(e)–(h) shows the better spatial enhancement of TSSC, DNCNN, DRPNN, and the proposed method. Moreover, the

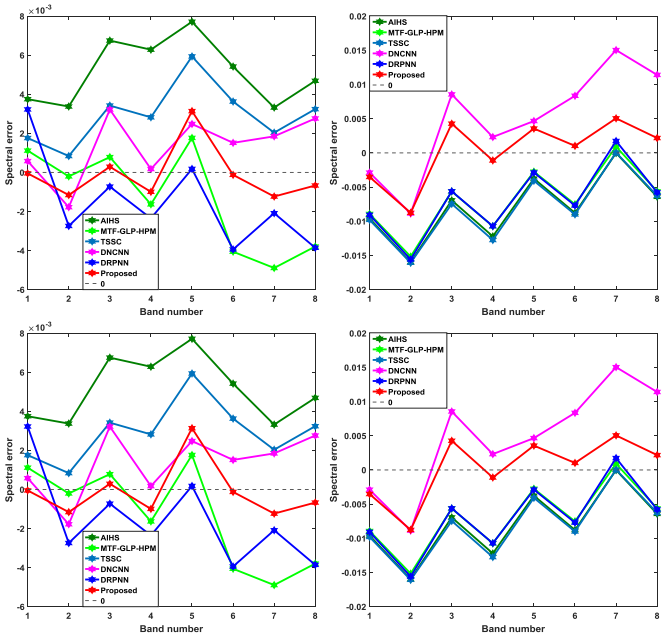


Fig. 9. Spectral error curves of four random pixels of difference images between the fusion results and the ground truth in Fig. 8.

almost invisible spatial structure in the proposed difference image in Fig. 8(h) indicates the superior spatial enhancement of the proposed method.

Fig. 9 shows the comparison of the spectral error curves of four random pixels of the difference images between the fusion results and the ground truth. In Fig. 9, the vertical axis named spectral error stands for the error values between the fusion results and the ground truth in each band, and the horizontal axis represents the band number. Similar to Fig. 6, the closer the error value is to zero, the better the fusion result is. It can be seen in Fig. 9 that the spectral error curve of the proposed method is closest to zero, which means that the proposed method has the best spectral fidelity. When combined with Fig. 8, it confirms the favorable performance of the proposed method in both spatial detail enhancement and spectral fidelity.

B. Real-Data Experiments

To further verify the effectiveness of the proposed method, two real data sets are employed in the real-data experiments. In the real-data experiments, we fuse the HR-PAN image and the LR-MS image directly to obtain the fused image.

Fig. 10 shows the results for the IKONOS images fused by the various mainstream methods. Note that, due to the lack of IKONOS images for training a network model, the three DL-based methods do not train a network model for the IKONOS sensor, and the network model trained for the QuickBird sensor is used to fuse the IKONOS images. Sharpened spatial features with severe spectral distortion are achieved by AIHS, MTF-GLP-HPM, and TSSC methods, as can be seen in the zoomed area in Fig. 10(c)–(e). DNCNN and DRPNN show a good performance in spectral fidelity but are poor in spatial texture information enhancement, as can be seen in the zoomed area in Fig. 10(f) and (g). Overall, the proposed method shows a superior performance, in both the fusion of the

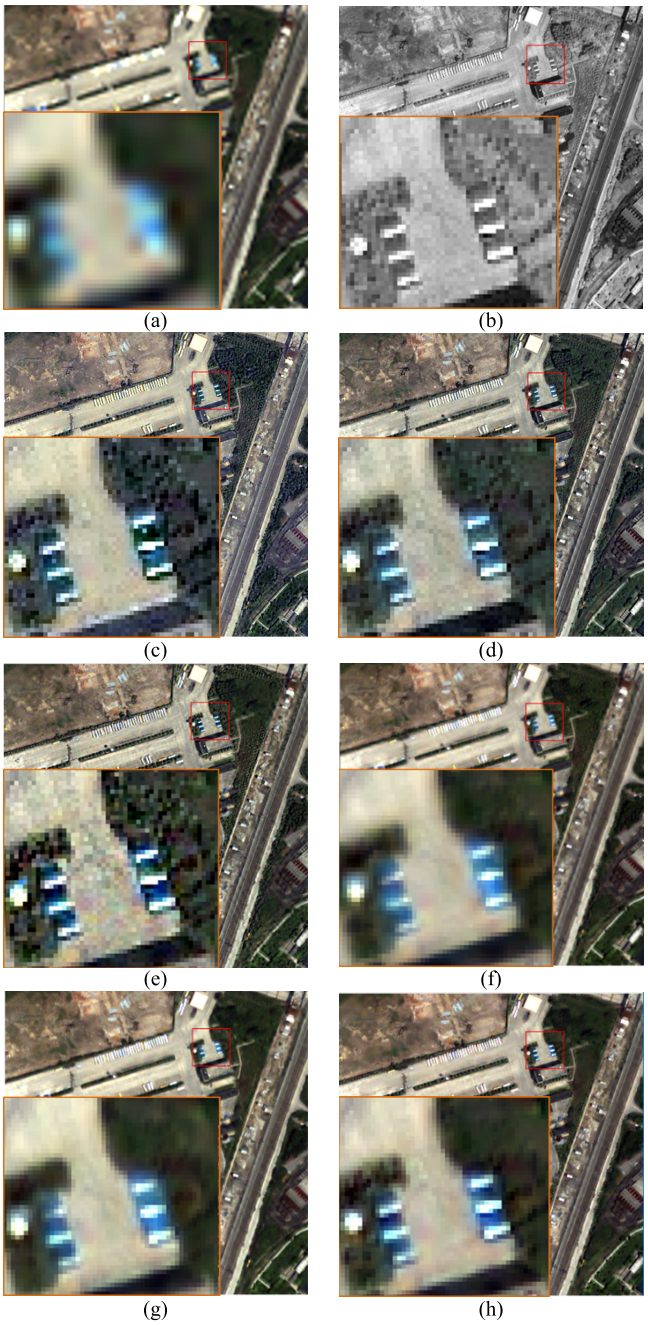


Fig. 10. Real-data fusion results of the IKONOS image. (a) LR-MS image. (b) HR-PAN image. (c) AIHS. (d) MTF-GLP-HPM. (e) TSSC. (f) DNCNN. (g) DRPNN. (h) Proposed method.

spatial details and in the preservation of the spectral fidelity, as can be seen in Fig. 10(h). Fig. 11 displays the real-data fusion results for the WorldView-2 images, which shows the same tendency as the simulated WorldView-2 experiments.

C. Further Discussion

As described in Section II-B, the spatial enhancement model in the variational model-based fusion methods is generally constructed based on two assumptions: the image prior and the spatial structure prior. The proposed model combines the deep residual gradient CNN with a variational model,

Fig. 11. Real-data fusion results of the WorldView-2 images. (a) LR-MS image. (b) HR-PAN image. (c) AIHS. (d) MTF-GLP-HPM. (e) TSSC. (f) DNCNN. (g) DRPNN. (h) Proposed method.

which is based on the second assumption. To explore this more comprehensively, a combination based on the first assumption is constructed, as shown in (5). In (5), $\widehat{\boldsymbol{X}} \in \mathbb{R}^{\mathrm{MN}\times S}$ is the image prior learned from the neural network. To compare the effects of the two variational models, simulated experiments are carried out on the QuickBird images and WorldView-2 images

$$\widehat{\boldsymbol{X}} = \arg\min_{\boldsymbol{X}} \frac{1}{2}||\boldsymbol{Y} - \boldsymbol{H}\boldsymbol{X}||_F^2 + \frac{\lambda}{2}\sum_{j=1}^{2}||\boldsymbol{X} - \widehat{\boldsymbol{X}}||_F^2 + \frac{\mu}{2}\{||\boldsymbol{Q}\boldsymbol{X}||_F^2\}. \quad (5)$$

TABLE III
QUANTITATIVE RESULTS OF DIFFERENT MODELS

Sensor	Method\index	ERGAS(↓)	SAM(↓)	Q(↑)	PSNR(↑)
QB 12 groups	DNCNN_image	2.9107	2.9791	0.9232	34.8554
	DRPNN_image	2.7646	2.7723	0.9312	35.4675
	Proposed	**2.4598**	**2.7124**	**0.9441**	**36.4841**
WV-2 6 groups	DNCNN_image	2.3385	3.3084	0.9165	34.7308
	DRPNN_image	2.0202	2.8169	0.9370	35.8640
	Proposed	**1.8615**	**2.7028**	**0.9381**	**36.6352**

TABLE IV
QUANTITATIVE RESULTS OF DIFFERENT STRATEGIES OF OBTAINING THE GRADIENT PRIOR

Sensor	Method\index	ERGAS(↓)	SAM(↓)	Q(↑)	PSNR(↑)
QB 12 groups	DNCNN_gra	2.7657	2.8333	0.9309	35.4573
	DRPNN_gra	2.6831	**2.7024**	0.9339	35.7351
	Proposed	**2.4598**	2.7124	**0.9441**	**36.4841**
WV-2 6 groups	DNCNN_gra	2.1657	3.2530	0.9219	35.5314
	DRPNN_gra	1.9206	2.7557	**0.9415**	36.4459
	Proposed	**1.8615**	**2.7028**	0.9381	**36.6352**

Table III shows the quantitative results of different variational models, where DNCNN_image represents that the image prior $\widehat{\boldsymbol{X}}$ in (5) is the fusion result of DNCNN, and DRPNN_image represents that the $\widehat{\boldsymbol{X}}$ in (5) is the fusion result of DRPNN. The upper part of Table III shows the simulated QuickBird experimental results with the average of 12 groups, and the lower part of Table III shows the simulated WorldView-2 experimental results with the average of six groups. In both parts, the best performance for each quality index is marked in bold and the second-best performance for each quality index is underlined. As shown in Table III, all the indices for the proposed method are the best in both the QuickBird and WorldView-2 experiments. This demonstrates that in the spatial enhancement terms, the gradient prior, which directly represents the spatial information, can obtain better results than the image prior.

To further illustrate the superiority of the proposed method, we design simulated experiments to compare the effect of the gradient priors obtained by different networks. Table IV lists the quantitative results of different strategies of obtaining the gradient prior, where DNCNN_gra and DRPNN_gra denote strategies that add the gradient of the fusion results obtained by DNCNN and DRPNN to the proposed constructed variational model, i.e., (3), respectively. Table IV has the same settings as Table III. It can be seen that most of the indices for the proposed method are the best, and the others are very close to the best. The reason for this could be that in the proposed gradient network, the purpose of the network is to obtain accurate HR-MS image gradient information. A lot of nongradient information existing in the image fusion network restricts the training of the image gradient, while the gradient network better preserves the gradient of the images.

IV. CONCLUSION

In this paper, we have proposed a PAN/MS fusion method that incorporates a deep residual gradient CNN into a

variational model-based framework. In the proposed method, we first train a gradient network to obtain the gradient of the HR-MS image and then utilize the learned gradient prior to construct a fusion variational model. Experiments on QuickBird, WorldView-2, and IKONOS data sets confirmed the effectiveness of the proposed method.

In our future works, the proposed method will be extended in two directions. On the one hand, as we just utilize a simple CNN to learn the gradient information, it may be possible to discover CNNs with different architectures, which can improve the performance. On the other hand, it will be of great significance to think about other ways to integrate DL into the variational model-based methods, to fully utilize the characteristics of the variational model-based methods and the feature learning capability of DL.

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] Y. Zhang, "Understanding image fusion," *Photogram. Eng. Remote Sens*, vol. 70, no. 6, pp. 657–661, 2004.

[3] X. Meng, H. Shen, H. Li, L. Zhang, and R. Fu, "Review of the pansharpening methods for remote sensing images based on the idea of meta-analysis: Practical discussion and challenges," *Inf. Fusion*, vol. 46, pp. 102–113, Jun. 2018.

[4] W. Carper, T. Lillesand, and R. Kiefer, "The use of intensity-hue-saturation transformations for merging SPOT panchromatic and multispectral image data," *Photogramm. Eng. Remote Sens.*, vol. 56, no. 4, pp. 459–467, Apr. 2004.

[5] 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.

[6] C. A. Laben and B. V. Brower, "Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening," U.S. Patent 6011875, Jan. 4, 2000.

[7] 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.

[8] G. Vivone *et al.*, "A critical comparison among pansharpening algorithms," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2565–2586, May 2015.

[9] L. Loncan *et al.*, "Hyperspectral pansharpening: A review," *IEEE Trans. Geosci. Remote Sens.*, vol. 3, no. 3, pp. 27–46, Sep. 2015.

[10] 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.

[11] X. Meng, J. Li, and H. Shen, "Pansharpening with a guided filter based on three-layer decomposition," *Sensors*, vol. 16, no. 7, pp. 1068–1084, Jul. 2016.

[12] Q. Xu, B. Li, Y. Zhang, and L. Ding, "High-fidelity component substitution pansharpening by the fitting of substitution data," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 7380–7392, Nov. 2014.

[13] P. S. Chavez, Jr., S. C. Sides, and J. A. Anderson, "Comparison of three different methods to merge multiresolution and multispectral data: Landsat TM and SPOT panchromatic," *Photogramm. Eng. Remote Sens.*, vol. 57, no. 3, pp. 295–303, 1991.

[14] 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.

[15] L. Wald and T. Ranchin, "Liu' Smoothing filter-based intensity modulation: A spectral preserve image fusion technique for improving spatial details,"' *Int. J. Remote Sens.*, vol. 23, no. 3, pp. 593–597, Jan. 2002.

[16] 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.

[17] M. Guo, H. Zhang, J. Li, L. Zhang, and H. Shen, "An online coupled dictionary learning approach for remote sensing image fusion," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 4, pp. 1284–1294, Apr. 2014.

[18] L. Zhang, H. Shen, W. Gong, and H. Zhang, "Adjustable model-based fusion method for multispectral and panchromatic images," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 42, no. 6, pp. 1693–1704, Dec. 2012.

[19] P. Liu, L. Xiao, J. Zhang, and B. Naz, "Spatial-Hessian-feature-guided variational model for pan-sharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 4, pp. 2235–2253, Apr. 2016.

[20] Z. Li, H. Shen, H. Li, G. Xia, and L. Zhang, "Multi-feature combined cloud and cloud shadow detection in GaoFen-1 wide field of view imagery," *Remote Sens. Environ.*, vol. 191, pp. 342–358, Mar. 2017.

[21] 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.

[22] 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, Jun. 2016.

[23] H. Shen, "Integrated fusion method for multiple temporal-spatial-spectral images," in *Proc. 22nd Congr. Int. Soc. Photogramm. Remote Sens. (ISPRS)*, Melbourne, VIC, Australia, 2012, pp. 407–410.

[24] 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.

[25] 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.

[26] Q. Wei, "Bayesian fusion of multi-band images: A powerful tool for super-resolution," Inst. Nat. Polytechn. Toulouse, Labège, France, 2015.

[27] X. Liu, H. Shen, Q. Yuan, X. Lu, and C. Zhou, "A universal destriping framework combining 1-D and 2-D variational optimization methods," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 808–822, Feb. 2018.

[28] X. Liu, X. Lu, H. Shen, Q. Yuan, Y. Jiao, and L. Zhang, "Stripe noise separation and removal in remote sensing images by consideration of the global sparsity and local variational properties," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 5, pp. 3049–3060, May 2016.

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

[30] A. Garzelli, "A review of image fusion algorithms based on the super-resolution paradigm," *Remote Sens.*, vol. 8, no. 10, pp. 797–816, Sep. 2016.

[31] K. Zhang, W. Zuo, and S. Gu, "Learning deep CNN denoiser prior for image restoration," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 3929–3938.

[32] Q. Zhang, Q. Yuan, J. Li, Z. Yang, and X. Ma, "Learning a dilated residual network for SAR image despeckling," *Remote Sens.*, vol. 10, no. 2, p. 196, Jan. 2018.

[33] Q. Zhang, Q. Yuan, C. Zeng, X. Li, and Y. Wei, "Missing data reconstruction in remote sensing image with a unified spatial-temporal-spectral deep convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4274–4288, Mar. 2018.

[34] Q. Yuan, Q. Zhang, J. Li, H. Shen, and L. Zhang, "Hyperspectral image denoising employing a spatial-spectral deep residual convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 1205–1218, Sep. 2018.

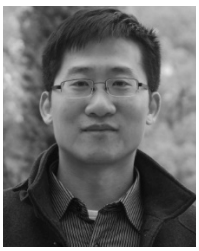
[35] H. Shen, T. Li, Q. Yuan, and L. Zhang, "Estimating regional ground-level $PM_{2.5}$ directly from satellite top-of-atmosphere reflectance using deep belief networks," *J. Geophys. Res., Atmos.*, vol. 123, no. 24, pp. 13875–13886, 2018.

[36] T. Li, H. Shen, Q. Yuan, X. Zhang, and L. Zhang, "Estimating ground-level $PM_{2.5}$ by fusing satellite and station observations: A geo-intelligent deep learning approach," *Geophys. Res. Lett.*, vol. 44, no. 23, pp. 11985–11993, 2017.

[37] G. Masi, D. Cozzolino, L. Verdoliva, and G. Scarpa, "Pansharpening by convolutional neural networks," *Remote Sens.*, vol. 8, no. 7, pp. 594–615, Jul. 2016.

[38] 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.

[39] 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.
[40] R. Molina, M. Vega, J. Mateos, and A. K. Katsaggelos, "Variational posterior distribution approximation in Bayesian super resolution reconstruction of multispectral images," *Appl. Comput. Harmon. Anal.*, vol. 24, no. 2, pp. 251–267, 2008.
[41] 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.
[42] J. Duran, A. Buades, B. Coll, and C. Sbert, "A nonlocal variational model for pansharpening image fusion," *SIAM J. Imag. Sci.*, vol. 7, no. 2, pp. 761–796, 2014.
[43] P. Liu, L. Xiao, and T. Li, "A variational pan-sharpening method based on spatial fractional-order geometry and spectral–spatial low-rank priors," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 3, pp. 1788–1802, Mar. 2018.
[44] Z. Li and H. Leung, "Fusion of multispectral and panchromatic images using a restoration-based method," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 5, pp. 1482–1491, May 2009.
[45] C. Chen, Y. Li, W. Liu, and J. Huang, "Image fusion with local spectral consistency and dynamic gradient sparsity," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Columbus, OH, USA, Jun. 2014, pp. 2760–2765.
[46] X. Meng, H. Shen, Q. Yuan, H. Li, and L. Zhang, "An integrated fusion framework for joint information reconstruction and resolution enhancement," *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. 42, pp. 831–835, Sep. 2017.
[47] M. Moeller, T. Wittman, and A. L. Bertozzi, "Variational wavelet pansharpening," CAM Rep., 2008, pp. 8–81.
[48] J. Yang, Y. Q. Zhao, and J. C. Chan, "Hyperspectral and multispectral image fusion via deep two-branches convolutional neural network," *Remote Sens.*, vol. 10, no. 5, pp. 800–822, May 2018.
[49] 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.
[50] R. Timofte, V. De Smet, and L. V. Gool, "A+: Adjusted anchored neighborhood regression for fast super-resolution," in *Proc. Asian Conf. Comput. Vis.* Cham, Switzerland: Springer, 2014, pp. 111–126.
[51] D. Kiku, Y. Monno, M. Tanaka, and M. Okutomi, "Residual interpolation for color image demosaicking," in *Proc. Int. Conf. Image Process.*, Sep. 2013, pp. 2304–2308.
[52] 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.
[53] B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, and M. Selva, "An MTF-based spectral distortion minimizing model for pan-sharpening of very high resolution multispectral images of urban areas," in *Proc. 2nd GRSS/ISPRS Joint Workshop Remote Sens. Data Fusion URBAN Areas*, May 2003, pp. 90–94.



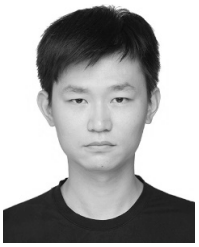
Huanfeng Shen (M'10–SM'13) received the B.S. degree in surveying and mapping engineering and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2002 and 2007, respectively.

In 2007, he joined the School of Resource and Environmental Sciences, Wuhan University, where he is currently a Luojia Distinguished Professor. He has been supported by several talent programs, such as the Youth Talent Support Program of China in 2015, the China National Science Fund for Excellent Young Scholars in 2014, and the New Century Excellent Talents by the Ministry of Education of China in 2011. He has authored over 100 research papers. His research interests include image quality improvement, remote sensing mapping and application, data fusion and assimilation, and regional and global environmental changes.

Dr. Shen is currently a member of the Editorial Board of the *Journal of Applied Remote Sensing*.

Menghui Jiang received the B.S. degree in geographical science from Wuhan University, Wuhan, China, in 2017, where she is currently pursuing the Ph.D. degree with the School of Resource and Environmental Sciences.

Her research interests include image data fusion, quality improvement, remote sensing image processing, and deep learning.



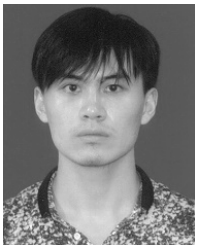
Jie Li (M'16) received the B.S. degree in sciences and techniques of remote sensing and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2011 and 2016, respectively.

He is currently a Lecturer with the School of Geodesy and Geomatics, Wuhan University. His research interests include image quality improvement, image super-resolution reconstruction, data fusion, remote sensing image processing, sparse representation, and deep learning.

Qiangqiang Yuan (M'13) received the B.S. degree in surveying and mapping engineering and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2006 and 2012, respectively.

In 2012, he joined the School of Geodesy and Geomatics, Wuhan University, where he is currently an Associate Professor. He has authored more than 50 research papers, including more than 30 peer-reviewed articles in international journals such as the IEEE TRANSACTIONS ON IMAGE PROCESSING and the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. His research interests include image reconstruction, remote sensing image processing and application, and data fusion.

Dr. Yuan was a recipient of the Top-Ten Academic Star of Wuhan University in 2011. He received the Hong Kong Scholar Award from the Society of Hong Kong Scholars and the China National Postdoctoral Council in 2014. He has frequently served as a referee for more than 20 international journals for remote sensing and image processing.



Yanchong Wei received the B.E. degree in geodesy and geomatics engineering and the M.E. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2015 and 2018, respectively.

He is currently a Remote Sensing Data Analyst with MCFLY Technology, Beijing. His research interests include degraded information reconstruction for remote sensed images, data fusion, deep learning, and computer vision.

Liangpei Zhang (M'06–SM'08–F'19) received the B.S. degree in physics from Hunan Normal University, Changsha, China, in 1982, the M.S. degree in optics from the Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an, China, in 1988, and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 1998.

He is currently the Head of the Remote Sensing Division, State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing (LIESMARS), Wuhan University. He is also a "Chang-Jiang Scholar" Chair Professor appointed by the Ministry of Education of China. He is currently a Principal Scientist for the China State Key Basic Research Project (2011–2016) appointed by the Ministry of National Science and Technology of China to lead the Remote Sensing Program in China. He has authored more than 500 research papers and five books. He holds 15 patents. His research interests include hyperspectral remote sensing, high-resolution remote sensing, image processing, and artificial intelligence.

Dr. Zhang is a fellow of the Institution of Engineering and Technology (IET) and an Executive Member (Board of Governor) of the China National Committee of International Geosphere–Biosphere Program and the China Society of Image and Graphics. He received the best reviewer awards from the IEEE Geoscience and Remote Sensing Society (GRSS) for his service to the IEEE JOURNAL OF SELECTED TOPICS IN EARTH OBSERVATIONS AND APPLIED REMOTE SENSING (JSTARS) in 2012 and IEEE GEOSCIENCE AND REMOTE SENSING LETTERS (GRSL) in 2014. His research teams won the top three prizes of the IEEE GRSS 2014 Data Fusion Contest, and his students have been selected as the winners or finalists of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS) Student Paper Contest in recent years. He was a recipient of the 2010 Best Paper Boeing Award and the 2013 Best Paper ERDAS Award from the American Society of Photogrammetry and Remote Sensing (ASPRS). He was the General Chair of the 4th IEEE GRSS Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS) and the Guest Editor of JSTARS. He is the Founding Chair of the IEEE GRSS Wuhan Chapter. He regularly serves as the Co-Chair of the series SPIE conferences on multispectral image processing and pattern recognition, conference on Asia remote sensing, and many other conferences. He edits several conference proceedings, issues, and geoinformatics symposiums. He also serves as an Associate Editor for the *International Journal of Ambient Computing and Intelligence*, the *International Journal of Image and Graphics*, the *International Journal of Digital Multimedia Broadcasting*, the *Journal of Geo-Spatial Information Science*, and the *Journal of Remote Sensing*, and the Guest Editor of the *Journal of Applied Remote Sensing*and the *Journal of Sensors.* He is currently serving as an Associate Editor for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING.