

Deep Image Interpolation: A Unified Unsupervised Framework for Pansharpening

Jianhao Gao, Jie Li, Xin Su, Menghui Jiang, Qiangqiang Yuan*

Wuhan University

*Corresponding to: yqiang86@gmail.com

Abstract

Pansharpening, whose aim is to acquire high resolution multispectral data (HRMS) by the fusion of low resolution multispectral data (LRMS) and panchromatic data (PAN), is a specific mission of spatial-spectral fusion in remote sensing field. In recent years, deep learning methods have proved the most feasible methods for pansharpening task. However, these deep learning methods have difficulty in training in an unsupervised manner and become useless when it comes to the condition where no training dataset is available. In this paper, we propose a universal algorithm called deep image interpolation for pansharpening task. The main idea is achieving high-quality fusion results by interpolating two low-quality multispectral images in a deep neural network. We apply it to two conditions: 1) unsupervised training a network when there are enough datasets; 2) directly optimizing the fusion result where no training datasets are available. Simulation and real-data experiments are conducted on various kinds of satellite data. Quantitative and qualitative evaluation results illustrate that the proposed method outperforms traditional pansharpening methods and even catch up with those supervised methods to some extent.

1. Introduction

When acquiring remote sensing data from satellites, a trade-off always exists between spatial and spectral resolution due to the limitation of sensors. In order to relieve the limitation of sensors and make full use of multi-source images with different spatial and spectral resolution, lots of pansharpening methods have been proposed, whose aim is to obtain high resolution multispectral images (HRMS) from panchromatic images (PAN), which have one band with very high spatial resolution, and low resolution multispectral images (LRMS). Examples of HRMS, LRMS and PAN are displayed in Fig. 1. According to [21], traditional pansharpening methods can be divided into three different

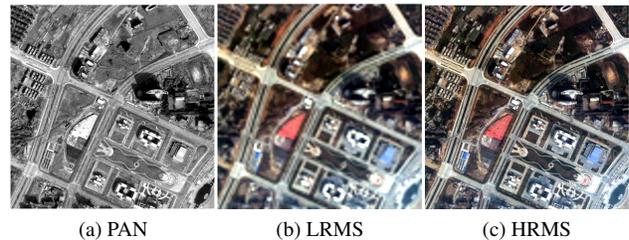


Figure 1. Examples of PAN, LRMS and HRMS.

families: 1) component substitution-based methods (CS-based methods); 2) multiresolution analysis-based methods (MRA-based methods); 3) variational model-based methods (VM-based methods).

CS-based methods are the most classic and fundamental pansharpening methods. They differ from each other in terms of projection algorithms, such as the intensity-hue-saturation (IHS) algorithm [4], the principal component analysis (PCA) algorithm [13], the Gram-Schmidt (GS) algorithm [14] and the Adaptive Gram-Schmidt (GSA) algorithm [3]. CS-based methods can precisely preserve the spatial information from PAN but fail to retain the accurate spectral information. MRA-based methods are another traditional pansharpening methods. Different transformation algorithms lead to different MRA-based methods. The representative transformation algorithms consist of the high-pass filter (HPF) [2], the generalized Laplacian pyramid with modulation transfer function matched filter (HTF-GLP) [5], the decimated wavelet transform (DWT) [24] and the smoothing filter-based intensity modulation (SFIM) [16]. MRA-based methods can better preserve the spectral fidelity in the pansharpening results but they cannot guarantee the spatial accuracy. VM-based methods are a relatively novel kind of pansharpening methods which view the pansharpening task as an ill-posed inverse optimization problem. They take advantage of observation model [32] or sparse representation theory [15] to construct the equations and solve them by iterative optimization algorithms

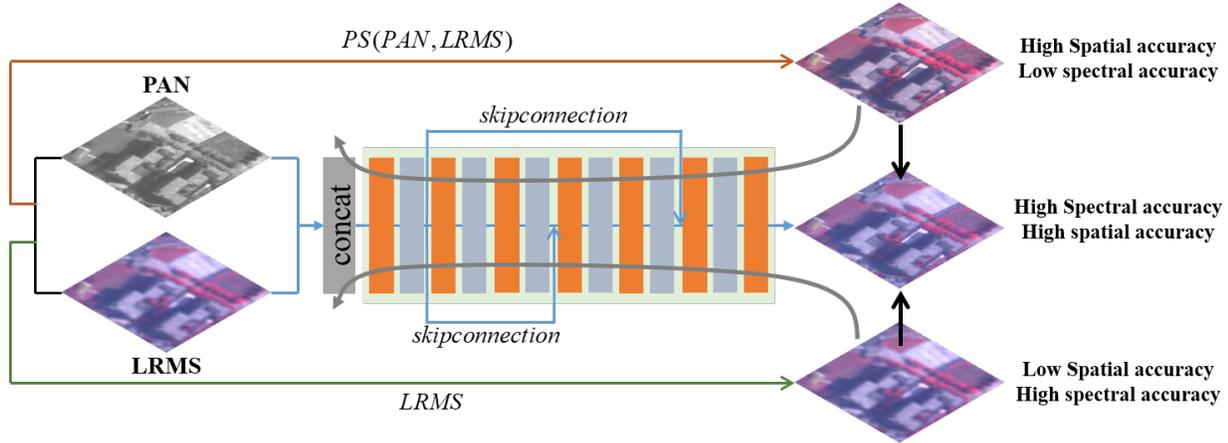


Figure 2. Workflow of the proposed method.

such as alternating direction method of multipliers algorithm (ADMM) [17] or gradient descent algorithm [25]. VM-based methods show their superiority in terms of the balance between spatial and spectral fidelity. Nevertheless, it is difficult to determine the optimal parameters in VM-based methods and the accuracy of fusion results heavily depends on the accuracy of observation model.

Taking the flaws of the above three kinds of methods into consideration, recent studies [9, 11, 20, 22, 26, 29–31] apply deep learning methods (DL-based methods) to pansharpening task. Compared with traditional methods, DL-based methods have great time-efficiency and achieve results with higher spectral and spatial fidelity. However, two limitations exist in most existing DL-based methods: 1) due to the inexistence of real HRMS, the network has to be trained according to Wald’s protocol in a supervised manner, which means the network cannot utilize the features of original resolution; 2) DL-based methods do not work when there is a small amount of training datasets or no training datasets.

To overcome the limitation of DL-based methods, we propose deep image interpolation, which interpolates two low-quality multispectral images in the network to obtain HRMS, solving the two drawbacks of DL-based mentioned above. One of the low-quality multispectral image should have high spectral fidelity and another should own high spatial accuracy. The proposed method can be used in the 1) unsupervised network training with enough datasets or 2) unsupervised pansharpening result optimization in deep network without datasets. The main process is displayed in Fig. 2. The characteristics and contributions of our work are summarized as follows:

- We propose a unified deep image interpolation framework for unsupervised pansharpening task. The proposed framework need not consider the number, size and resolution of training data, showing the flexibility and transferability of the proposed method.

- We introduce a combination of two simple optimization terms to constrain the spectral fidelity and spatial accuracy of pansharpening results. A simple deep neural network is used to complete the whole process.

- Several simulation and real-data experiments are conducted with data from different satellites. The results of the proposed method outperform many state-of-the-art pansharpening methods, verifying the effectiveness and accuracy of our method.

The whole paper is organized as follow: in Section II, we introduce some related DL-based pansharpening methods; in Section III, main process and two strategies of pansharpening by deep image interpolation are presented elaborately; in Section VI, we demonstrate pansharpening results of simulation and real-data experiments and compare the proposed method with other state-of-the-art methods in quantitative and qualitative manner; some discussions and extra experiments are also presented; in Section V, we draw a conclusion of the whole paper and briefly introduce our future work.

2. Related Works

DL-methods for supervised pansharpening: Generally speaking, DL-based pansharpening methods could be viewed as a variant of DL-based super-resolution [6, 12] methods and are conducted in a supervised manner. For example, [20] took the idea from [6] and firstly introduce convolution neural network to the pansharpening task. They proposed a three-layer network called pansharpening neural network. State-of-the-art fusion results are obtained compared with those traditional methods. [31] further came up with a multiscale and multidepth convolutional neural network, which concatenated feature maps from convolution layers with different kernel sizes and different depths. [11] argues that the residual information between PAN and

HRMS has less information than the residual between PAN and LRMS so they trained the network with the former as output and the latter as input, by which shows a great improvement compared with other DL-based methods.

DL-based methods have great time-efficiency due to its forward propagation and acceleration from GPU. They outperform those traditional methods in terms of spatial and spectral fidelity because of their non-linear fitting ability. However, due to the inexistence of real HRMS, they have to train the model in a supervised manner according to Wald’s protocol which obtains training datasets by downsampling the original data. This behavior may ignore the real image feature in primary resolution.

DL-based methods for unsupervised pansharpening: Up to now, some unsupervised DL-based methods [18, 19, 33] for pansharpening are proposed. They operate in an unsupervised manner by answering the following two questions: 1) how to construct the spectral relationship between output and LRMS; 2) how to construct the spatial relationship between output and PAN. These methods get the same answer for the first question: the output is simply downsampled to the same resolution of LRMS by bicubic algorithm and is then compared with LRMS. For the second question, they attempt to linearly add the channels of output and compare it with PAN. The only difference is the adding ways. For example, [19] directly computed the average of all channels in the output. [18] summed the channels with least square coefficients calculated between down-sampled PAN and LRMS previously. [33] made use of the spectral response function provided previously to sum up the channels of output. Although fine pansharpening results have been achieved in these researches, there are still some potential risks. The first is that the relationship between PAN and HRMS is not simply linear. Such linear summing up may lead to spectral distortion in pansharpening results. The second is that these unsupervised methods cannot be applied when there is a small amount of training datasets and only single pair of PAN and LRMS.

Image optimization in deep neural network without training samples: Considering the condition where no training samples are available, some studies make full use of the strong non-linear fitting ability and inner property of deep neural network to complete many low-level vision tasks, such as [8, 23, 27, 28]. In [28], authors introduce deep image prior to many applications such as image reconstruction, whose experiments show the superiority of their method. Given a sample of texture, [27] can generate a new image with the same type of texture by generative adversarial network. [8] proposed neural style transfer, which extracts the style of one image and injects it to the content of another image to obtain a style transfer result with a vgg network. Given one sample image, [23] could generate various outputs, which has similar content of the given image, with

a generative adversarial network and this algorithm also has great performance in many other applications such as image harmonization.

Although these methods have not touched on the fusion tasks such as pansharpening which focus more on the accuracy of results, they confirm the possibility that deep neural network can also be applied in the condition where no training samples are available. Inspired by these studies, we introduce our deep image interpolation to pansharpening with single pair of PAN and LRMS.

3. Methodology

We give some notations of data which is used in this method for the sake of simplification. $\mathbf{P} \in \mathbb{R}^{W \times H \times 1}$ denotes the PAN data. W and H are respectively the width and height of PAN. $\mathbf{M} \in \mathbb{R}^{w \times h \times C}$ means the LRMS data. The width, height and the number of channels of LRMS are indicated respectively by w and h and C . HRMS is denoted as $\mathbf{R} \in \mathbb{R}^{W \times H \times C}$, which have the same width and height as PAN and the same number of channels as LRMS. We denote the fusion result as $\mathbf{F} \in \mathbb{R}^{W \times H \times C}$. PS denotes a traditional pansharpening method. \downarrow_n denotes the operation of downsampling by n times and \uparrow_n denotes the operation of upsampling by n times. We denote our network as G .

3.1. Deep Image Interpolation

All existing unsupervised deep learning methods aim to establish a linear relationship between the output of network and \mathbf{P} to constrain the spatial information of the output. However, the exact relationship between \mathbf{P} and \mathbf{R} is far from linear and a spatial constraint based on linear relationship will lead to severe spectral distortion. In the proposed method, we first diffuse the spatial information from \mathbf{P} to all bands of \mathbf{M} by a traditional method to constrain the spatial information of fusion results. Compared with solely \mathbf{P} , traditional pansharpening results are close to ground truth in terms of spectral information which makes them have less effect on the spectral accuracy of network outputs. Afterward, deep image interpolation serves to acquire pansharpening results by interpolating the results of the traditional method and \mathbf{M} in the image level.

First, we make use of a traditional method PS to acquire a multispectral image $\mathbf{R}_0 \in \mathbb{R}^{W \times H \times C}$ with high spatial fidelity but low spectral accuracy, which is described in Equation 1:

$$\mathbf{R}_0 = PS(\mathbf{P}, \mathbf{M}) \quad (1)$$

A lot of methods can serve as PS , such as SFIM [16], MTF-GLP [5] and BDSF [7]. Here we select the BDSF method as PS due to its better ability for spatial information maintenance. Then we upsample the LRMS to the same resolution of PAN and concatenate it with PAN as the input

of the network to obtain the output \mathbf{F} :

$$\mathbf{F} = G(\mathbf{P}, \mathbf{M}) \quad (2)$$

The spatial optimization term is constructed between \mathbf{R}_0 and \mathbf{F} for spatial information accuracy, which is displayed in Equation 3:

$$\mathcal{L}_{spatial} = \frac{1}{W \times H \times C} \|\mathbf{R}_0 - \mathbf{F}\|_1^1 \quad (3)$$

The spectral optimization term is constructed with \mathbf{M} and \mathbf{F} in order to guarantee the spectral accuracy of pansharpening result, which is illustrated in Equation 4.

$$\mathcal{L}_{spectral} = \frac{1}{W \times H \times C} \|\mathbf{M} \uparrow_n - \mathbf{F} \downarrow_n \uparrow_n\|_1^1 \quad (4)$$

The upsampling operation in Equation 4 is added to augment the spatial information of \mathbf{M} and the downsampled \mathbf{F} . The two terms are added by weight and the total optimization term is obtained:

$$\mathcal{L}_{total} = \mathcal{L}_{spatial} + \lambda \mathcal{L}_{spectral} \quad (5)$$

The final fusion result is obtained until the optimization term reach the minimum.

3.2. Network structure

The proposed network structure is shown in Figure 2. The whole network adopts a simple 7-layer convolution neural network structure. The first six layers contain a convolution operation and a ReLU activation operation. According to previous studies, gradient vanishing may occur when it comes to a deep network structure with more than three layers. In order to avoid this gradient vanishing phenomenon, we make use of concatenation operation within the network, which has been proven effective in dealing with such problem. The output of the first layer and third layer are concatenated as input of the fourth layer. The output of the first layer and fifth layer are also concatenated as input of the sixth layer. The seventh layer has just a simple convolution operation whose output is our expected fusion result.

3.3. Strategies for two conditions

The proposed framework can be used to process the following two tasks: 1) unsupervised network training with large datasets; 2) unsupervised pansharpening result optimization without training samples.

For unsupervised network training with large datasets, the aim is to obtain the trained network. Then the trained network can be further applied to acquire pansharpening results of other images except training datasets. The optimization term is used to obtain this trained network which is explained in Equation 6.

$$G^* = \arg \min \mathcal{L}_{total}(\mathbf{P}, \mathbf{M}, \mathbf{F}). \quad (6)$$

When there is a small amount of training datasets or only one pair of images for pansharpening, the proposed unsupervised network will not be trained well. Under this condition, we aim to obtain the suitable network output but not the trained network. So the optimization term is used to directly optimize the pansharpening result which is mentioned in Equation 2.

$$\mathbf{F}^* = \arg \min \mathcal{L}_{total}(\mathbf{P}, \mathbf{M}, \mathbf{F}). \quad (7)$$

For unsupervised network training with large datasets, the network acquires the knowledge of pansharpening during the training process. So it can be retained for further application. In the unsupervised pansharpening result optimization, the whole process is an overfitting process compared with the unsupervised network training. It cannot be applied in other data so we abandon the network after optimization but retain the fusion result. For the sake of simplification, we name the algorithm used in unsupervised network training as DIIA and the algorithm used in unsupervised image optimizing as DIIB.

4. Experiments

In order to testify the effectiveness of the proposed method, both simulation and real-data experiments are conducted in this section. In Section 4.1, we introduce the experiment settings including datasets, comparison methods, evaluation methods and optimization details. Qualitative and qualitative evaluation of simulation experiments are presented in Section 4.2. In section 4.3, results of real-data experiments are displayed to verify the superiority of the proposed methods. We show some other relevant experiments in Section 4.4.

4.1. Experiment settings

Datasets: We use two training datasets from two different satellites for DIIA in unsupervised training task. The first training dataset is Gaofen-2 dataset used in [21], which makes use of multispectral (LRMS) and panchromatic (PAN) data from GanFen-2 satellite. The spatial resolutions of LRMS and PAN are respectively 4m and 1m. LRMS has a total of four bands which are respectively R, G, B and NIR bands. The second training dataset is QuickBird dataset used in [20], which uses the LRMS and PAN from QuickBird satellite. The spatial resolutions of LRMS and PAN are respectively 2.44m and 0.61m. LRMS of QuickBird satellite has also four bands which are R, G, B, NIR. Two test datasets from Gaofen-2 and QuickBird are provided for both DIIA in unsupervised training task and DIIB for pansharpening results optimization task. They are obtained in different areas from the training datasets. The

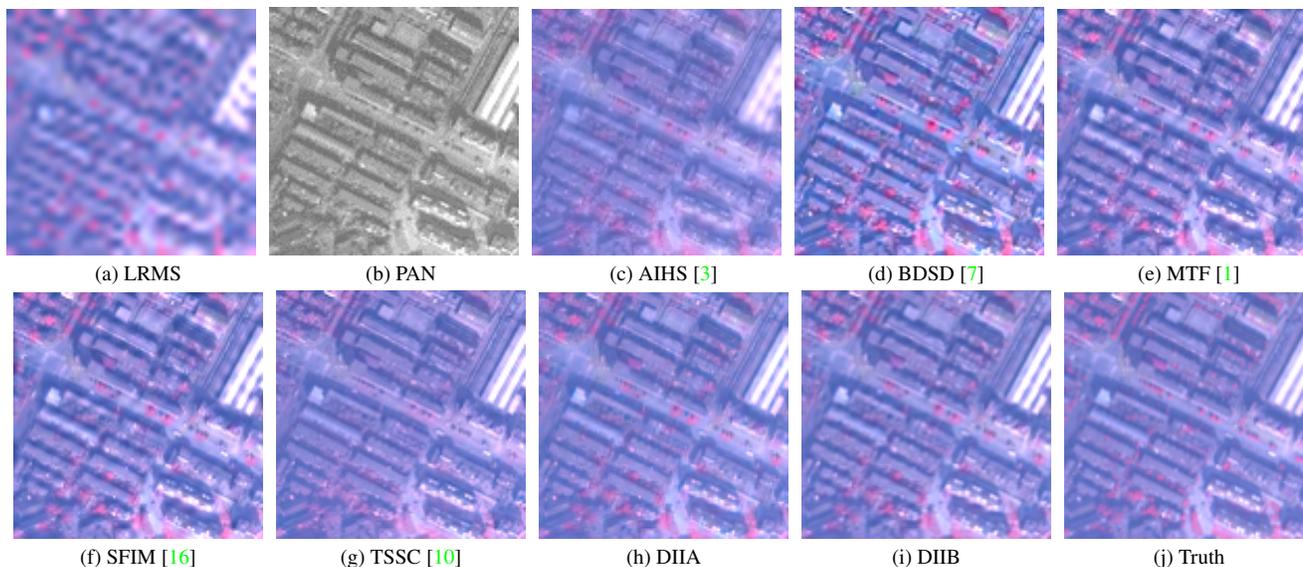


Figure 3. Simulation experiment results of QuickBird data.

Gaofen-2 test dataset contains 20 pairs of LRMS and PAN. The QuickBird test dataset has 12 pairs of LRMS and PAN.

Two simulation experiments are conducted with QuickBird and Gaofen-2 datasets. We follow the Wald’s protocol and downsample the PAN and LRMS in both training and testing datasets by four times to obtain new PAN and LRMS. The original LRMS is viewed as ground truth in the simulation experiment. Specifically, for the unsupervised training in DIIA, we obtain the training datasets by cropping images into small patches with a size of 80×80 with a step of 40 pixels. The final QuickBird simulation training dataset has 15139 patches and the Gaofen simulation training dataset has 8777 patches. A real experiment is conducted on Gaofen-2 dataset. We make training datasets for unsupervised training by directly cropping image patches from original PAN and LRMS. The images of Gaofen-2 data are cropped into 80×80 with a step of 80 and we finally obtain 35108 patches.

Comparison methods: We choose five conventional but useful pansharpening methods from different classes as comparison methods. They are respectively Smoothing Filter-based Intensity Modulation method (SFIM [16]), Adaptive Intensity Hue Saturation method (AIHS [3]) which belong to CS-based class, MTF-GLP with High-Pass Modulation injection method [1], Band-Dependent Spatial Detail method (BDSD [7]) which pertain to MRA-based class, and Two-Step Sparse Coding method (TSSC [10]) which is a VM-based method. All the methods mentioned above are state-of-the-art pansharpening methods in their classes.

Evaluation methods: Six indexes are used to evaluate the veracity of pansharpening results. They are spectral

Method	ERGAS	PSNR	Q	SAM	spCC	SSIM
MTF [1]	3.088	35.124	0.828	2.640	0.810	0.962
BDSD [7]	2.942	35.664	0.813	3.352	0.788	0.956
AIHS [3]	2.435	37.095	0.826	2.578	0.832	0.967
SFIM [16]	2.815	35.907	0.833	2.497	0.805	0.964
TSSC [10]	2.369	37.431	0.854	2.665	0.809	0.971
DIIA	<u>1.962</u>	<u>38.993</u>	<u>0.895</u>	<u>2.179</u>	<u>0.857</u>	<u>0.981</u>
DIIB	1.620	40.483	0.918	2.063	0.905	0.983

Table 1. Quantitative evaluation of QuickBird dataset.

angle mapper (SAM), respectively relative dimensionless global error in synthesis (ERGAS), peak-signal-to-noise-ratio (PSNR), similarity structure index (SSIM), Q index (Q) and spatial correlation coefficient (spCC). The first three indices are used to evaluate the spectral information of the fusion results and the last three indexes can judge their accuracy of spatial information.

Optimization environment and details: We conduct the whole experiment with Pytorch1.0 under the environment of Ubuntu16.04. One 2080Ti GPU is utilized to run the program. We make use of Adam optimizer to optimize the whole network. Learning rate is set as 0.00001 and the training epoch is set as 16.

4.2. Simulation experiments

QuickBird results: The first simulation experiment is conducted with QuickBird dataset. The six quantitative evalu-

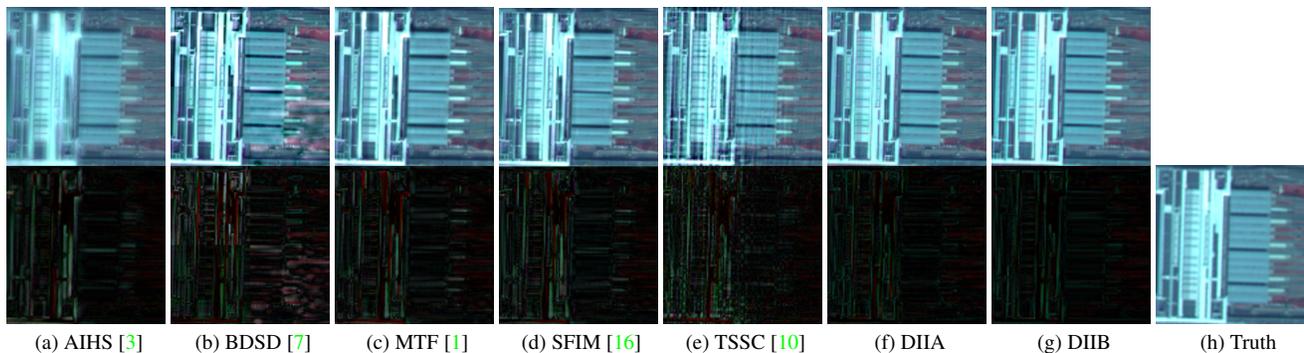


Figure 4. Simulation experiment results and residual maps of Gaofen-2 data.

Method	ERGAS	PSNR	Q	SAM	spCC	SSIM
MTF [1]	3.599	26.338	0.873	2.695	0.790	0.868
BDSB [7]	3.725	25.959	0.854	3.365	0.744	0.794
AIHS [3]	2.776	28.645	0.886	2.580	<u>0.845</u>	0.889
SFIM [16]	3.348	26.955	0.875	2.626	0.778	0.858
TSSC [10]	3.188	27.553	0.886	3.320	0.747	0.860
DIIA	<u>2.471</u>	<u>29.491</u>	<u>0.917</u>	<u>2.447</u>	0.801	<u>0.903</u>
DIIB	2.015	31.575	0.949	2.191	0.897	0.934

Table 2. Quantitative evaluation of Gaofen-2 dataset.

ation results of 7 methods are listed in Table 1. We mark the highest score of each index in bold and the second highest score with underline. We can find out from the table that compared with traditional methods, the two proposed methods, DIIA and DIIB can obtain fusion results with high spectral and spatial accuracy, outperforming the rest by a large extent. Another issue worth mentioning from Table. 1 is that results of single data pansharpening by DIIB outperform results of unsupervised training by DIIA slightly although they make use of the same auxiliary data. This phenomenon illustrates a trade-off of time and accuracy in deep image interpolation.

We select a representative pansharpening result and display them in Fig. 3 here for visual evaluation. In this result, some traditional methods such as SFIM and MTF-GLP-HRM inject some disruptive spatial information from LRMS into the fusion result. AIHS cannot even restore some buildings. BDSB and TSSC perform greatly in recovering this ground object. However, overall result of BDSB is brighter than the ground truth image and result of TSSC is darker, which show the weakness of BDSB and TSSC in restoring spectral information. The proposed two methods can perfectly restore the building and obtain accurate spectral information overall. Specifically, DIIB slightly outper-

forms DIIA.

Gaofen-2 results: Table 2 lists the six evaluation indexes of seven methods in gaofen-2 testing dataset. Again we mark the highest score in each index in bold and the second highest with underline. Our deep image interpolation outperforms the traditional fusion methods by a large extent again. We select a representative scene and display its fusion results in Fig. 4. We also display the absolute error maps between the fusion result of each method and ground truth. We can find from Figure. 4 that there is much residual information left in the absolute error maps of BDSB, SFIM, MTF-GLP-HPM. There is less residual information in absolute error maps of AIHS results and TSSC results. However, they cannot well recover the spatial information in the fusion results which is vividly shown in Fig. 4. Our two interpolation methods, DIIA and DIIB, have less residual information compared with the above five methods, which shows the superiority of our deep image interpolation.

4.3. Real-data experiments

We apply the trained network into the real data to testify the efficacy of the proposed method. One prominent pansharpening result from Gaofen-2 test dataset is selected to display in Fig. 5. We magnify and display an object which has strong light reflecting, making it hard to recover. TSSC and AIHS perform badly in injecting the spatial information into fusion results in terms of buildings. SFIM, MTF-FLP-HPM and BDSB cannot guarantee the spectral fidelity of vegetation. Our two methods, DIIA and DIIB, can take both of the spectral fidelity and spatial accuracy into account, which make it outperform the comparison methods.

4.4. Relevant experiments

Selection of spatial constraint: The ideal auxiliary image except LRMS should have high spatial accuracy in the proposed methods. BDSB [7] method is no doubt the most suitable method in our framework. Other popular traditional methods, such as SFIM [16] and MTF [1], sometimes inject the inaccurate spectral information to the fusion re-

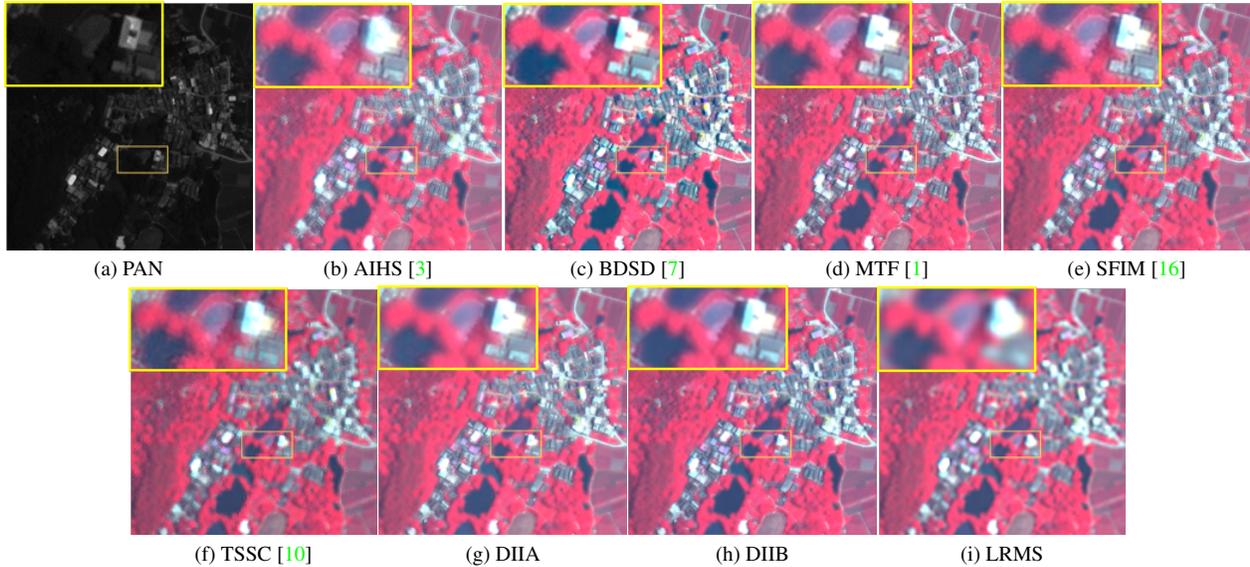


Figure 5. Real data experiment result of Gaofen-2 data.

Method	ERGAS	PSNR	Q	SAM	spCC	SSIM
MTF [1]	3.09	35.12	0.83	2.64	0.81	0.96
with MTF	1.60	40.42	0.91	<u>2.04</u>	<u>0.90</u>	0.99
SFIM [16]	2.81	35.91	0.83	2.50	0.81	0.96
with SFIM	1.62	40.65	<u>0.91</u>	1.99	0.889	0.98
BDSD [7]	2.94	35.66	0.81	3.35	0.79	0.96
with BDSD	<u>1.62</u>	40.48	0.92	2.06	0.91	<u>0.98</u>

Table 3. Spatial constraint analysis.

Method	ERGAS	PSNR	Q	SAM	spCC	SSIM
PanGAN [19]	2.11	37.90	0.88	2.54	0.87	0.98
PNN [20]	1.98	38.91	0.88	2.15	0.87	0.98
PanNet [30]	<u>1.78</u>	<u>39.84</u>	<u>0.90</u>	1.93	<u>0.89</u>	<u>0.98</u>
DIIA	1.96	38.99	0.90	2.18	0.86	0.98
DIIB	1.62	40.48	0.92	<u>2.06</u>	0.91	0.98

Table 4. Comparison with supervised and unsupervised DL-based methods.

sults which is vividly shown in Fig. 2, but BDSD [7] does not. To quantitatively illustrate this idea, we adopt the proposed DIIB and compare some representative methods with BDSD [7] method in the QuickBird dataset. The evaluation results are listed in Table 3. These results show the validity of choosing BDSD method.

Supervised and unsupervised deep learning methods: In order to further verify the effectiveness of the proposed

method, we compare the two proposed methods in our framework with two supervised network training methods, which are PNN [20] and PanNet [30] respectively, and one state-of-the-art unsupervised training method PanGAN [19]. The experiment is conducted under the same settings with the QuickBird dataset. Quantitative evaluation results of the proposed two unsupervised pansharpening methods and three comparison methods are listed in Table 4. As expected, the unsupervised training does not perform as well as the model created with supervised training. However, we find that DIIB outperforms all supervised and unsupervised training methods in terms of all indexes in QuickBird dataset.

In Fig. 6, we present the simulation experiment results and their absolute error map from ground truth in Fig. 6. In Fig. 7, we displays the real data experiment results. Results of PanGAN have sharp spatial information but bad spectral accuracy. Results of PNN and PanNet suffer from both spectral distortion and spatial inaccuracy. Result of the proposed DIIA shares similar quality with those of PNN and PanNet. However, the proposed DIIB outperforms all other methods in terms of spectral and spatial fidelity, again showing the superiority of the proposed framework. We impute the failure of PanGAN to the inaccurate spatial constraint constructed linearly between PAN and the output of network. The above results also confirm that the spatial constraint used in the proposed method is the main reason why the proposed unsupervised training, DIIA, can catch up with the supervised training. As for the phenomenon that supervised network training cannot catch up with DIIB, we attribute the reason to the network capacity. In DIIB, the capacity of network is enough for the fusion of only one

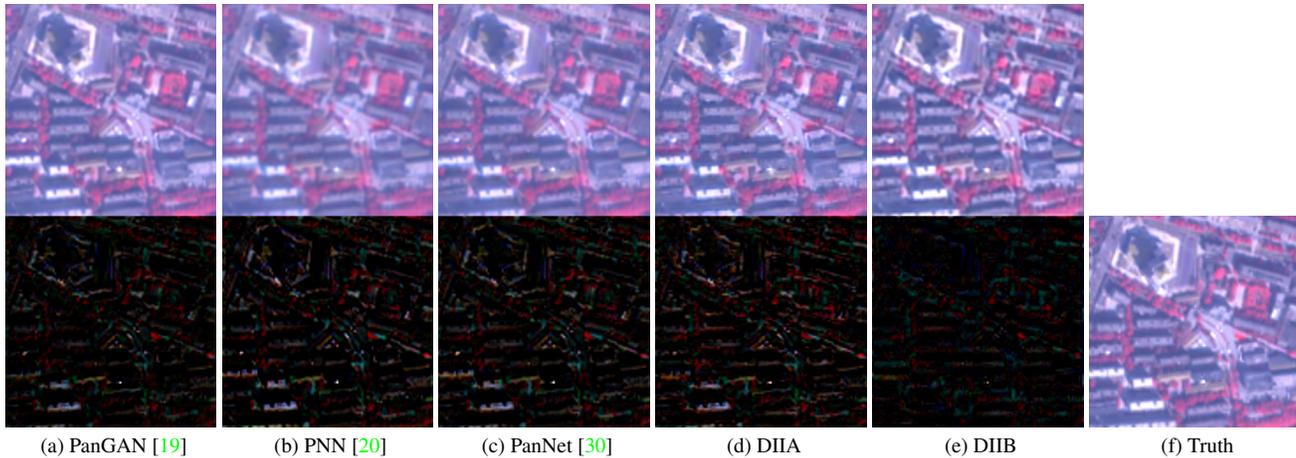


Figure 6. Simulation results and residual maps of supervised and unsupervised training.

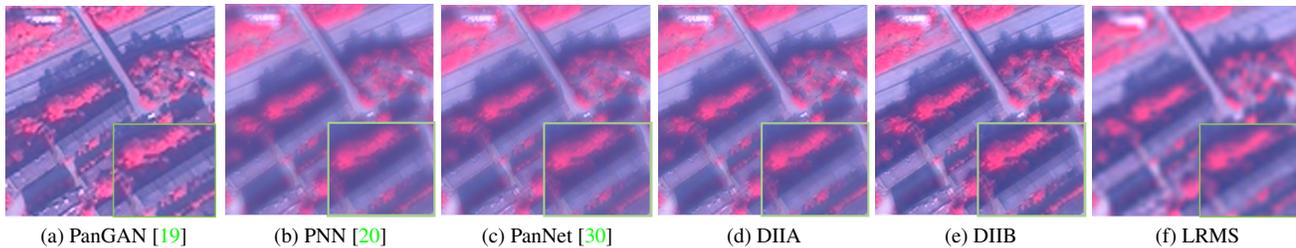


Figure 7. Real data results of supervised and unsupervised training.

image. While in PNN and PanNet, the networks have to consider all images in the training dataset so they cannot precisely deal with only one image.

5. Conclusions

In this paper, we propose an idea called deep image interpolation based on deep neural network for pansharpening task. The method obtains high-quality pansharpening results by interpolating two low-quality data during the optimization process of the network. Two optimization terms made up from LRMS and BSD result are proposed to guarantee the spectral and spatial fidelity of fusion result. Our model can be applied in both unsupervised training with large datasets and unsupervised pansharpening result optimization without training process. Many simulation and real-data experiments are conducted and testify the superiority and convenience of the proposed method. We also compare the proposed method with the supervised method. The gap exists but is not a big one. In our future work, the first improvement we want to make is to find out the better and more precise optimization term. Another improvement is the network structure. The network used in this paper is designed optionally and far from the optimal. We will try the Network Architecture Search (NAS) technic for

better solution. Furthermore, the proposed method can be applied into other spatial-spectral data fusion work, such as panchromatic-hyperspectral data fusion, multispectral-hyperspectral data fusion, even in other tasks.

References

- [1] B Aiazzi, L Alparone, S Baronti, A Garzelli, and M Selva. An mtf-based spectral distortion minimizing model for pansharpening of very high resolution multispectral images of urban areas. In *2003 2nd GRSS/ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas*, pages 90–94. IEEE, 2003. 5, 6, 7
- [2] B Aiazzi, L Alparone, S Baronti, A Garzelli, and M Selva. Mtf-tailored multiscale fusion of high-resolution ms and pan imagery. *Photogrammetric Engineering & Remote Sensing*, 72(5):591–596, 2006. 1
- [3] Bruno Aiazzi, Stefano Baronti, and Massimo Selva. Improving component substitution pansharpening through multivariate regression of ms + pan data. *IEEE Transactions on Geoscience and Remote Sensing*, 45(10):3230–3239, 2007. 1, 5, 6, 7
- [4] WJOSEPH CARPER, THOMASM LILLESAND, and RALPHW KIEFER. The use of intensity-hue-saturation transformations for merging spot panchromatic and multispectral image data. *Photogrammetric Engineering and remote sensing*, 56(4):459–467, 1990. 1

- [5] Pats Chavez, Stuart C Sides, Jeffrey A Anderson, et al. Comparison of three different methods to merge multiresolution and multispectral data- landsat tm and spot panchromatic. *Photogrammetric Engineering and remote sensing*, 57(3):295–303, 1991. [1](#), [3](#)
- [6] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 38(2):295–307, 2015. [2](#)
- [7] Andrea Garzelli, Filippo Nencini, and Luca Capobianco. Optimal mmse pan sharpening of very high resolution multispectral images. *IEEE Transactions on Geoscience and Remote Sensing*, 46(1):228–236, 2007. [3](#), [5](#), [6](#), [7](#)
- [8] Leon A Gatys, Alexander S Ecker, and Matthias Bethge. Image style transfer using convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2414–2423, 2016. [3](#)
- [9] Wei Huang, Liang Xiao, Zhihui Wei, Hongyi Liu, and Songze Tang. A new pan-sharpening method with deep neural networks. *IEEE Geoscience and Remote Sensing Letters*, 12(5):1037–1041, 2015. [2](#)
- [10] Cheng Jiang, Hongyan Zhang, Huanfeng Shen, and Liangpei Zhang. Two-step sparse coding for the pan-sharpening of remote sensing images. *IEEE journal of selected topics in applied earth observations and remote sensing*, 7(5):1792–1805, 2013. [5](#), [6](#), [7](#)
- [11] Menghui Jiang, Huanfeng Shen, Jie Li, Qiangqiang Yuan, and Liangpei Zhang. A differential information residual convolutional neural network for pansharpening. *ISPRS Journal of Photogrammetry and Remote Sensing*, 163:257–271, 2020. [2](#)
- [12] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *European conference on computer vision*, pages 694–711. Springer, 2016. [2](#)
- [13] P Kwarteng and A Chavez. Extracting spectral contrast in landsat thematic mapper image data using selective principal component analysis. *Photogramm. Eng. Remote Sens*, 55(1):339–348, 1989. [1](#)
- [14] Craig A Laben and Bernard V Brower. Process for enhancing the spatial resolution of multispectral imagery using pansharpening, Jan. 4 2000. US Patent 6,011,875. [1](#)
- [15] Jie Li, Qiangqiang Yuan, Huanfeng Shen, and Liangpei Zhang. Noise removal from hyperspectral image with joint spectral-spatial distributed sparse representation. *IEEE Transactions on Geoscience and Remote Sensing*, 54(9):5425–5439, 2016. [1](#)
- [16] JG Liu. Smoothing filter-based intensity modulation: A spectral preserve image fusion technique for improving spatial details. *International Journal of Remote Sensing*, 21(18):3461–3472, 2000. [1](#), [3](#), [5](#), [6](#), [7](#)
- [17] Xinxin Liu, Huanfeng Shen, Qiangqiang Yuan, Xiliang Lu, and Chunping Zhou. A universal destriping framework combining 1-d and 2-d variational optimization methods. *IEEE Transactions on Geoscience and Remote Sensing*, 56(2):808–822, 2017. [2](#)
- [18] Shuyue Luo, Shangbo Zhou, Yong Feng, and Jiangan Xie. Pansharpening via unsupervised convolutional neural networks. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13:4295–4310, 2020. [3](#)
- [19] Jiayi Ma, Wei Yu, Chen Chen, Pengwei Liang, Xiaojie Guo, and Junjun Jiang. Pan-gan: An unsupervised pan-sharpening method for remote sensing image fusion. *Information Fusion*, 62:110–120, 2020. [3](#), [7](#), [8](#)
- [20] Giuseppe Masi, Davide Cozzolino, Luisa Verdoliva, and Giuseppe Scarpa. Pansharpening by convolutional neural networks. *Remote Sensing*, 8(7):594, 2016. [2](#), [7](#), [8](#)
- [21] Xiangchao Meng, Huanfeng Shen, Huifang Li, Liangpei Zhang, and Randi Fu. Review of the pansharpening methods for remote sensing images based on the idea of meta-analysis: Practical discussion and challenges. *Information Fusion*, 46:102–113, 2019. [1](#)
- [22] Giuseppe Scarpa, Sergio Vitale, and Davide Cozzolino. Target-adaptive cnn-based pansharpening. *IEEE Transactions on Geoscience and Remote Sensing*, 56(9):5443–5457, 2018. [2](#)
- [23] Tamar Rott Shaham, Tali Dekel, and Tomer Michaeli. Singan: Learning a generative model from a single natural image. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4570–4580, 2019. [3](#)
- [24] Hamid Reza Shahdoosti and Nayereh Javaheri. Pansharpening of clustered ms and pan images considering mixed pixels. *IEEE Geoscience and Remote Sensing Letters*, 14(6):826–830, 2017. [1](#)
- [25] Huanfeng Shen. Integrated fusion method for multiple temporal-spatial-spectral images. *Proc. 22nd Congr. ISPRS*, pages 407–410, 2012. [2](#)
- [26] Huanfeng Shen, Menghui Jiang, Jie Li, Qiangqiang Yuan, Yanchong Wei, and Liangpei Zhang. Spatial-spectral fusion by combining deep learning and variational model. *IEEE Transactions on Geoscience and Remote Sensing*, 57(8):6169–6181, 2019. [2](#)
- [27] Assaf Shocher, Shai Bagon, Phillip Isola, and Michal Irani. Ingan: Capturing and retargeting the” dna” of a natural image. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4492–4501, 2019. [3](#)
- [28] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 9446–9454, 2018. [3](#)
- [29] Yancong Wei, Qiangqiang Yuan, Huanfeng Shen, and Liangpei Zhang. Boosting the accuracy of multispectral image pansharpening by learning a deep residual network. *IEEE Geoscience and Remote Sensing Letters*, 14(10):1795–1799, 2017. [2](#)
- [30] Junfeng Yang, Xueyang Fu, Yuwen Hu, Yue Huang, Xinghao Ding, and John Paisley. Pannet: A deep network architecture for pan-sharpening. In *Proceedings of the IEEE international conference on computer vision*, pages 5449–5457, 2017. [2](#), [7](#), [8](#)
- [31] Qiangqiang Yuan, Yancong Wei, Xiangchao Meng, Huanfeng Shen, and Liangpei Zhang. A multiscale and multidepth convolutional neural network for remote sensing imagery

- pan-sharpening. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(3):978–989, 2018. 2
- [32] Liangpei Zhang, Huanfeng Shen, Wei Gong, and Hongyan Zhang. Adjustable model-based fusion method for multispectral and panchromatic images. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 42(6):1693–1704, 2012. 1
- [33] Changsheng Zhou, Jianshe Zhang, Junmin Liu, Chunxia Zhang, Rongrong Fei, and Shuang Xu. PercepPan: Towards unsupervised pan-sharpening based on perceptual loss. *Remote Sensing*, 12(14):2318, 2020. 3