

Research Article

Remote Sensing Image Fusion Method Based on Progressive Cascaded Deep Residual Network

Jing Li ¹, Hongwu Hu,¹ Lilan Lei,¹ and Jin Li²

¹Yichun University, Yichun, 336000 Jiangxi, China

²Southwest Medical University, Luzhou, 646000 Sichuan, China

Correspondence should be addressed to Jing Li; 205150@jxycu.edu.cn

Received 7 July 2022; Revised 2 August 2022; Accepted 3 August 2022; Published 22 February 2023

Academic Editor: Kalidoss Rajakani

Copyright © 2023 Jing Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the rapid development of deep learning in recent years, it has shown excellent performance in various image and video processing tasks. In addition, it also has a great role in promoting the spatio-temporal fusion of remote sensing images. The reconstructed image can give people a good visual experience. The invention relates to a remote sensing image fusion method based on a progressive cascade deep residual network and provides an end-to-end progressive cascade deep residual network model for remote sensing image fusion. The use of the MSE loss function may cause oversmoothing of the fused image, so a new joint loss function is defined to capture finer spatial information to improve the spatial resolution of the fused image. Resize-convolution is used to replace the transposed convolution to eliminate the checkerboard effect in the fused image caused by the transposed convolution. Through the experiments on the remote sensing image fusion simulation and real datasets of multiple satellites, the data results of the proposed algorithm are more than 5.25% better than those of the comparative algorithm in the average quantification. The calculation time and system resource occupation are also reduced, which has important theoretical significance and application value in the field of artificial intelligence and image processing. It will play a certain role in promoting the theoretical research and application of remote sensing image fusion.

1. Introduction

High time resolution of remote sensing image means that the time interval between two adjacent shots is short when the same target is taken repeatedly. Similarly, high time resolution is very important for dynamic monitoring. Spectral resolution refers to the ability of remote sensing satellite to distinguish the spectral details of ground objects. It is the minimum wavelength interval that can be identified when the satellite sensor receives the Pope emitted by the ground object box. Hyperspectral information can identify ground objects with subtle spectral differences. Multiangle information provides important information for the inversion of atmospheric or surface parameters (such as atmospheric aerosol thickness and vegetation leaf area index) [1]. Due to the limitation of imaging technology and data capacity of sensors, the existing remote sensing satellite sensors cannot capture a single remote sensing image with high spectral resolution and high spatial resolution at the same time. To

preserve the spectral information and spatial information of ground targets at the same time, for the same scene, most remote sensing satellites have photographed MS (multispectral) and Pan (panchromatic) [2]. Among them, MS has higher spectral resolution but lower spatial resolution, while Pan has the opposite. To comprehensively utilize the spectral information and spatial information contained in remote sensing images to realize the analysis and understanding of remote sensing images, it is necessary to perform pansharpening on MS and Pan to obtain HRMS (high-resolution multispectral) with high resolution.

The spatio-temporal fusion method based on the convolutional neural network proposed by Cui and others [3] has greatly improved the performance of fusion compared with other methods. The number of neural network layers used by the author is less (3 hidden layers), and for MODIS-Landsat images with large spatial scale differences and from different satellite sensors, shallow convolutional neural networks are difficult to accurately learn the nonlinear mapping

relationship between them. At present, how to deal with the spatial difference between the two types of MODIS-Landsat and how to determine the optimal number of layers and filters of the deep convolutional neural network are still urgent problems to be solved in the follow-up research. Roy et al. [4] used the residual network to train a deep convolutional neural network to complete the work of superresolution reconstruction, which also has some inspiration for the subsequent research on spatio-temporal fusion of remote sensing images. Therefore, the application of a deep convolutional neural network in spatio-temporal fusion of remote sensing images still needs to be explored [5].

Most of them focus on traditional algorithms, including the fusion algorithm based on component replacement, the fusion algorithm based on multiresolution analysis, and the fusion algorithms based on variational optimization [6]. The modeling of these algorithms is complex, and it is difficult to achieve a good balance between spectral information preservation and spatial detail enhancement in the fused image, and the performance is limited by the introduction of prior assumptions. To overcome the limitations of traditional algorithms, a few studies have introduced convolutional neural networks, which have been widely used in the field of computer vision in recent years, into the task of remote sensing image fusion. These algorithms have achieved a better compromise between spectral information preservation and spatial detail enhancement, effectively improved the quality of fused images, and demonstrated the superiority of convolutional neural network for remote sensing image fusion tasks [7]. However, in these studies, there are mainly the following shortcomings: to solve the problem that the size of MS and Pan is inconsistent and cannot be fused, the MS is upsampled to the size of Pan to extract features, which introduces inaccurate information and increases unnecessary computational overhead. When reconstructing the fused image, the use of multilevel features is not sufficient. There is a loss of information in the fusion results. In a shallow network, features are extracted by a simple cascade of nonlinear transformation layers, and the context information obtained is limited [8].

In remote sensing image fusion, the MS image with the low spatial resolution is directly upsampled by 4 times to the PAN image size and then fused, which will lead to incomplete feature extraction. The main reason is that it is relatively difficult to learn nonlinear feature mapping in high-dimensional space, which will cause the loss of some accurate high-frequency details [9]. To solve this problem, this paper proposes a remote sensing image fusion method based on the Progressive Cascade Deep Residual Network (PCDRN), which completes image fusion step by step through two residual subnetworks from coarse to fine. Firstly, utilize a first residual subnetwork to carry out feature extraction in low-dimensional feature space to obtain a preliminary fusion result, then carry out upsampling on the preliminary fusion result, and send the upsampled result into a second residual subnetwork. Finally, the second residual subnetwork is used to further extract more fine feature information in the high-dimensional feature space to obtain the final MS image with high spatial resolution. In addition, a

joint loss function is proposed to guide the training and learning of the network to solve the problem that the MSE loss function causes the fused image to be too smooth. To eliminate the checkerboard effect caused by the use of transposed convolution, a resize-convolution method is proposed to replace the transposed convolution to perform the upsampling operation in the network.

The main contributions of this paper include the following three points:

- (1) A new end-to-end progressive cascaded deep residual network is proposed to gradually generate fused images with high spatial resolution and high spectral resolution
- (2) A joint loss function is proposed to help the network extract more fine feature information and realize the preservation of spatial information
- (3) Resize-convolution method is proposed to replace the transposed convolution to upsample the feature map in the network

2. Related Work

2.1. Remote Sensing Image Data. With the launch of a large number of earth remote sensing satellite sensors, the amount of remote sensing data acquired has increased dramatically, and the newly launched satellite sensors are developing in the direction of high spatial, high temporal, high spectral, and multiangle resolution data acquisition capabilities [10], such as China's high-resolution earth observation system [11]. However, due to the limitations of existing remote sensing satellite sensors in terms of hardware technology and launch cost, there is a limitation of mutual checks and balances between resolutions in the satellite remote sensing systems. For example, to improve the spatial resolution of satellite sensors, it is necessary to reduce the spectral resolution, temporal resolution, and angular resolution [12]. Secondly, the observation data obtained by satellite sensors launched in the early stage also have the problem of compromise between various resolutions. Therefore, whether in the further mining of massive historical satellite remote sensing image information or in the full use of newly acquired remote sensing images in the future, the contradiction between remote sensing image resolution is a realistic problem that cannot be ignored [13].

2.2. Image Fusion Algorithm. In remote sensing image fusion, the MSE loss function is often used to guide the training and learning of the CNN model. However, because the MSE loss function is difficult to capture the difference in high-frequency details between the fused image and the reference image, it is easy to lose high-frequency details such as texture, resulting in an oversmooth image and poor perceptual quality [14]. The calculation method of the joint loss function L_{Fusion} is as shown

$$L^{\text{Fusion}} = \alpha \cdot L_{\text{NMSE}}^{\text{Fusion}} + (1 - \alpha) \cdot L_{\text{NUIQI}}^{\text{Fusion}}, \quad (1)$$

where $L_{\text{NMSE}}^{\text{Fusion}}$ denotes the normalized MSE loss; $L_{\text{NUIQI}}^{\text{Fusion}}$ denotes the normalized UIQI loss; and α denotes the weight coefficient. See formulas (2) and (3) for the definitions of $L_{\text{NMSE}}^{\text{Fusion}}$ and $L_{\text{NUIQI}}^{\text{Fusion}}$, respectively:

$$L_{\text{NMSE}}^{\text{Fusion}} = \beta \cdot \left(\frac{1}{n} \sum_{i=1}^n \|F(x^{(i)}, y^{(i)}) - R^{(i)}\|^2 \right), \quad (2)$$

$$L_{\text{NUIQI}}^{\text{Fusion}} = \gamma \cdot \left(1 - \frac{1}{n} \sum_{i=1}^n \text{UIQI}(F(x^{(i)}, y^{(i)}), R^{(i)}) \right), \quad (3)$$

where β and γ denote the normalization coefficients; n denotes the number of training images; i denotes the i th image ($i \in [1, n]$); $x^{(i)}$ denotes the i th MS image with low spatial resolution; $y^{(i)}$ denotes the i th PAN image; $F(x^{(i)}, y^{(i)})$ denotes the i th fused image; and $R^{(i)}$ denotes the i th reference image [15].

3. Progressive Remote Sensing Image Fusion Method

3.1. Deep Residual Network Architecture. In the process of image fusion, because the size of the original MS image is not consistent with that of the PAN image, it is easy to cause incomplete feature extraction if the MS image is directly upsampled by four times the size of the PAN image and then fused. The main reason is that it is relatively difficult for the network to learn nonlinear feature mapping in high-dimensional feature space, which easily leads to the loss of some accurate high-frequency details [16]. To solve this problem, a remote sensing image fusion method based on the progressive cascaded deep residual network is proposed in this paper. This method uses two cascaded residual subnetworks ResNet and ResNet2 to gradually extract feature details from PAN and MS images in a coarse-to-fine manner to better learn the nonlinear feature mapping between source images (MS and PAN images) and MS images with high spatial resolution. The ResNet1 and ResNet2 subnetworks are mainly composed of one 3×3 convolution layer, one ReLU activation function, and five residual blocks, respectively, and their overall framework is shown in Figure 1.

In Figure 1, the red line is the MS image workflow. It can be seen from Figure 1 that the PCDRN proposed in this paper is mainly implemented in the following three stages:

- (1) The main purpose of the first stage is to adjust the MS image with low spatial resolution and the PAN image with a high spatial resolution to the same size and send the results of direct splicing of the two images into the network. First, the length and width of the low spatial resolution MS image are doubled by the nearest neighbor interpolation method; then, the length and width of the corresponding PAN image are downsampled to 1/2 of the original; finally, the upsampled MS image and the downsampled PAN image are spliced in the channel direction to generate a 5-band input tensor [17]

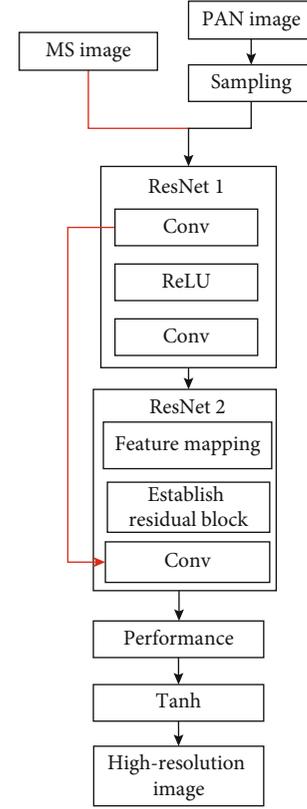


FIGURE 1: Overall architecture based on progressive cascaded deep residual network.

- (2) The main purpose of the second stage is to complete the preliminary integration of MS and PAN images and get the preliminary fusion image. Firstly, the 5-band tensor obtained from the previous stage is input into the first residual subnetwork ResNet1 to obtain the preliminary feature information, and then, the feature map obtained from the MS image which is upsampled twice through a convolution layer 3×3 and the feature map obtained through ResNet1 are summed according to elements. Secondly, a convolution layer of 1×1 is used to reduce the dimension of the summed result to obtain a 4-channel feature map. Finally, the feature map is sent to the Tanh function for a calculation to obtain a fused image twice the size of the initial MS image [18]
- (3) The main function of the third stage is to extract finer spatial details in higher dimensions by using the ResNet2 on the basis of the second stage. Firstly, after the preliminary fused image obtained in the previous stage is upsampled by two times, it is spliced with the PAN image, and the spliced result is sent to the second residual subnetwork ResNet2. Then, as the operation mode of the previous stage, a 3×3 convolution layer is used to calculate the MS image after two times of upsampling, and the obtained feature map and the output result of

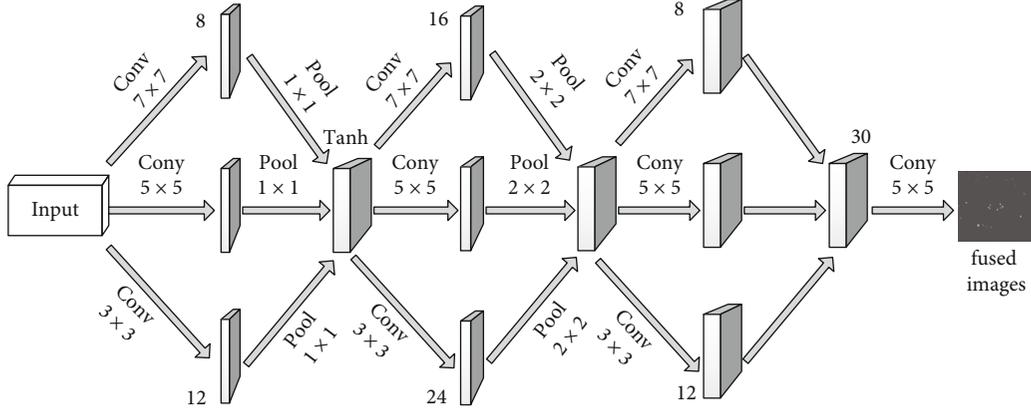


FIGURE 2: Flow of convolution algorithm.

Input: MS image (ms), PAN image (pan),

Training period (Max_train_epoch)

Output: Normalization Coefficients β and γ

1. $c \leftarrow 0.00001$; num $\leftarrow 0$; $L_{MSE}(0) \leftarrow 0$; $L_{UIQI}(0) \leftarrow 0$; //Initialize

For $j = 1$ to max_train_epoch do

2. Input ms(j) and pan(j), through the $L_{MSB}(j) = 1/n \sum_{i=1}^n \|F(x^{(i)}, y^{(i)}) - R^{(i)}\|^2$ and $L_{UIQI}(j) = 1 - 1/n \sum_{i=1}^n UIQI(F(x^{(i)}, y^{(i)}), R^{(i)})$, than $L_{MSB}(j)$ and $L_{UIQI}(j)$

If(Convergence)

3. if $\nabla L_{MSE}(\text{num}) = |L_{MSE}(j) - L_{MSE}(j-1)|$, than $\nabla L_{MSE}(\text{num})$;

4. if $\nabla L_{UIQI}(\text{num}) = |L_{UIQI}(j) - L_{UIQI}(j-1)|$, than $\nabla L_{UIQI}(\text{num})$;

5. num = num + 1

6. End if

7. End for

8. if $\nabla L_{MSE}(k)$, $k = 1, 2, \dots, \text{num}$, than $\overline{\nabla L_{MSE}}$;

9. if $\nabla L_{UIQI}(k)$, $k = 1, 2, \dots, \text{num}$, than $\overline{\nabla L_{UIQI}}$;

Calculate the coefficients β and γ by $c/\overline{\nabla L_{MSE}}$ and $c/\overline{\nabla L_{UIQI}}$, respectively.

ALGORITHM 1: Joint Loss Function Normalization Algorithm.

ResNet2 are summed by elements. Secondly, a 1×1 convolution layer is used to reduce the summation result from 64 channels to 4 channels. Finally, a Tanh function is used to normalize it, and finally, the MS image with the high spatial resolution is obtained [19]. The convolution algorithm flow is shown in Figure 2

3.2. Joint Loss Function. The UIQI index, one of the structural distortion metrics commonly used in fusion image quality evaluation, can effectively capture the difference between the reference image and the fusion image [20]. In this paper, the MSE loss function and UIQI index are combined to design a new joint loss function to guide the training and learning of the CNN model, which is conducive to improving the performance of the fusion method in maintaining spatial details.

To make the order of magnitude of the two losses of MSE and UIQI roughly the same, the normalization algorithm is introduced, and its specific process is shown in Algorithm 1.

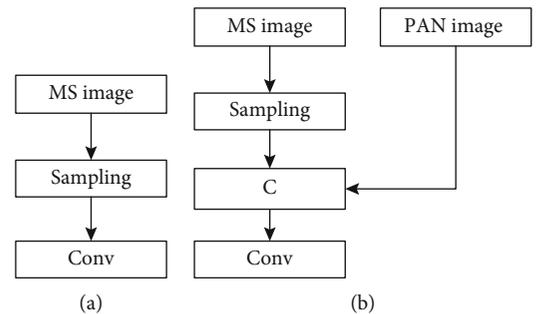


FIGURE 3: Schematic diagram of resize-convolution method.

3.3. Resize-Convolution Method. Transpose convolution is a common upsampling method in CNN. This method does not need to be defined in advance, and its weights can be learned automatically, but it is easy to produce uneven overlap, resulting in the checkerboard effect of the fused image [21]. To solve this problem, in the PCDRN method proposed in this paper, the resize-convolution method is used to replace the transposed convolution to achieve the

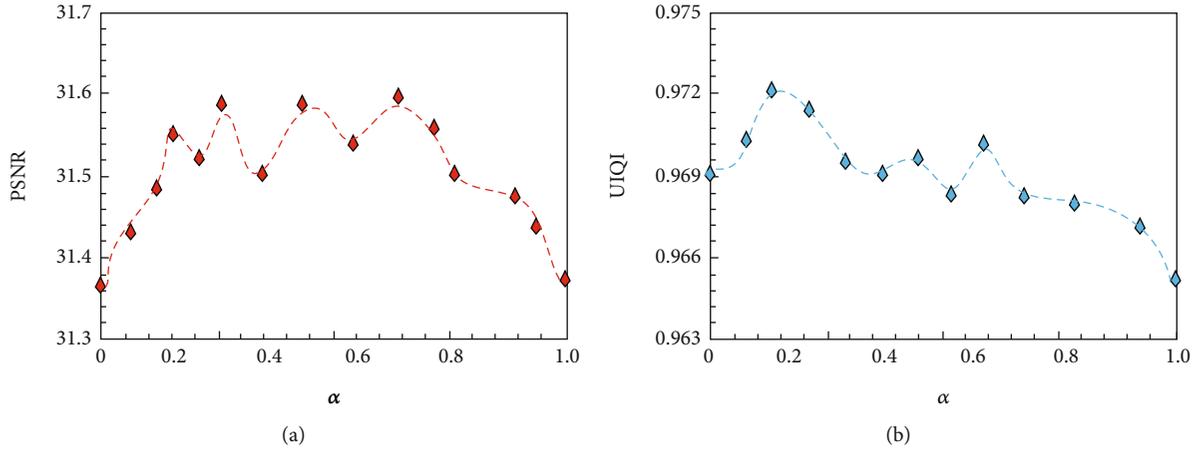


FIGURE 4: Average quality index.

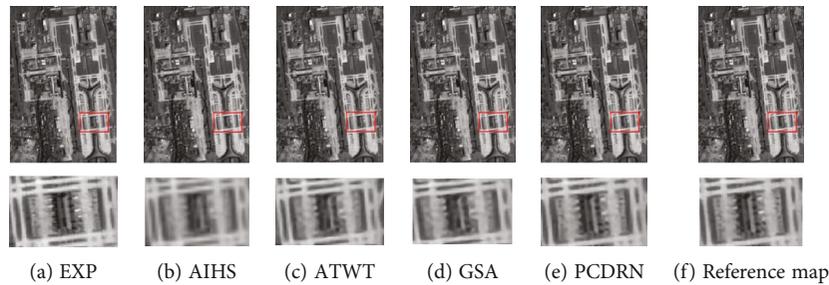


FIGURE 5: Experimental results of simulation data.

upsampling function of the image. According to the different input images, the method is divided into single input and double input, as shown in Figure 3.

(1) Resize-convolution method with single input

The only input to this method is the MS image. It consists of two steps: first, the nearest neighbor interpolation method is used to improve the resolution of the MS image, and then, a 3×3 convolution layer is used to calculate it.

(2) Resize-convolution method with two inputs

The input of the method is both PAN and MS images. It includes three steps: firstly, the nearest neighbor interpolation method is used to improve the resolution of the MS image; secondly, the MS image after upsampling is spliced with the PAN image; finally, the spliced result is sent to a 3×3 convolution layer for calculation.

4. Experimental Results and Analysis

4.1. Experimental Setup. To verify the performance of the PCDRN method proposed in this paper, the data of PI Pléiades and IKONOS satellites are selected for experiments.

Simulation experiment dataset: 5238 and 5760 groups of simulation data are selected from the satellite dataset as the training dataset. 100 and 108 groups of simulation data are selected as the simulation test dataset, where each group of

simulation data consists of a degraded MS image, a degraded PAN image and a reference image [22].

4.2. Parameter Setting. The experiment is carried out with 180 sets of PI Pléiades simulation data.

Figure 4 shows the average quality index (PSNR and UIQI) results of PCDRN under different α values.

In Figure 4, when α in formula (1) is set to 0.1, the performance of the PCDRN method proposed in this paper is the best.

4.3. Analysis of Experimental Results. To verify the effectiveness of the PCDRN method proposed in this paper, this paper uses the simulation data of PI Pléiades and IKONOS satellite datasets as experimental datasets and uses subjective and objective indicators to evaluate the experimental results of the four existing fusion methods and the PCDRN method proposed in this paper [23].

(1) Fusion experiment based on PI Pléiades simulation data

In this paper, image fusion experiments are carried out on PI Pléiades simulation data. Figure 5 shows a set of fusion results based on the PI Pléiades simulation data.

In Figure 5, the EXP method results in an upsampled image that, although of good spectral quality, shows severe spatial distortion. It can be seen from the enlarged area that the fusion results of ATWT have serious spectral distortion, while the fusion methods of AIHS, ATWT, and GSA all

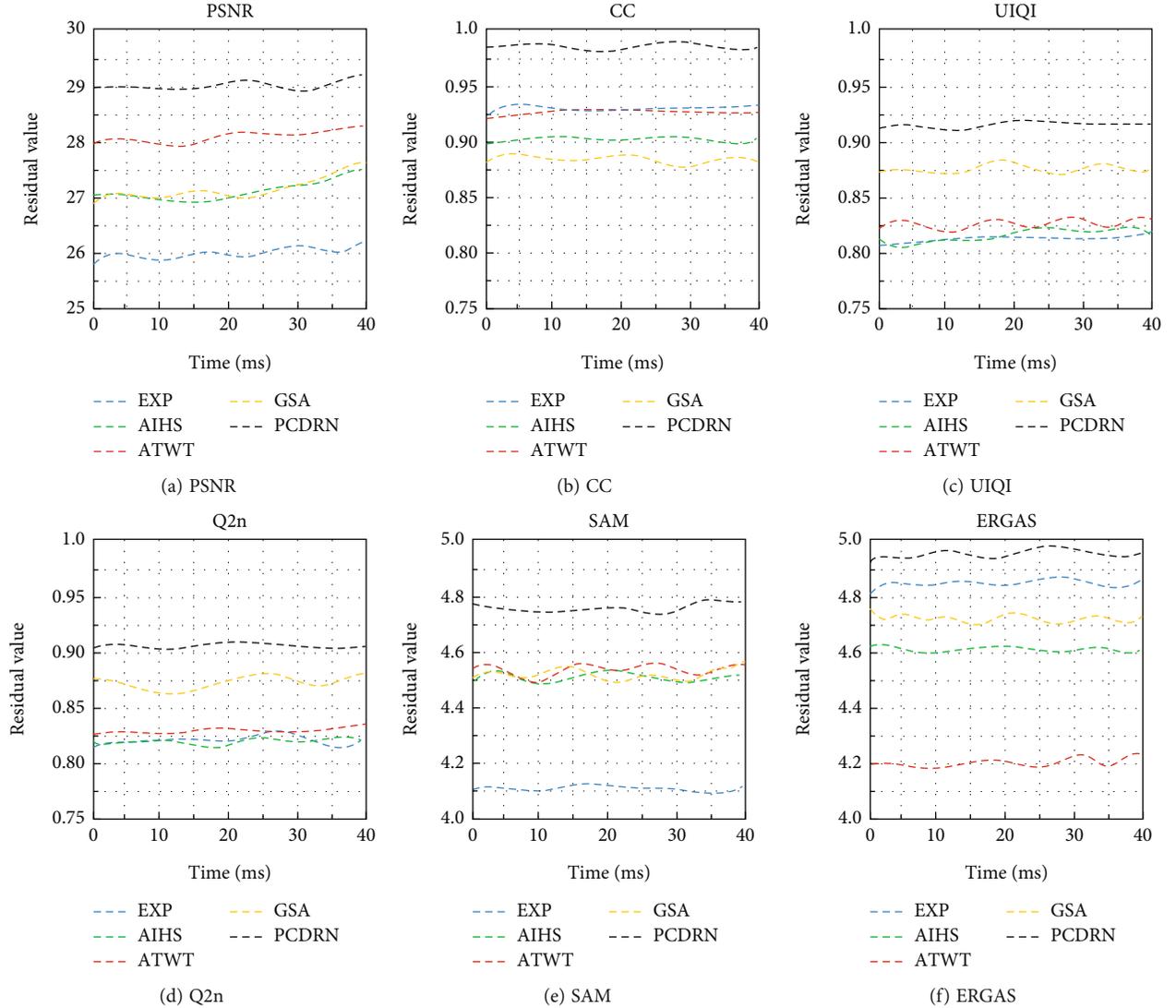


FIGURE 6: Comparison of residual data of fusion results.

TABLE 1: Objective evaluation of fusion results (1).

Method	PSNR	CC	UIQI	Q2n	SAM	ERGAS
EXP	26.554	0.947	0.842	0.824	4.142	4.897
AIHS	27.472	0.925	0.905	0.823	4.535	4.625
ATWT	28.014	0.953	0.892	0.834	4.554	4.253
GSA	27.055	0.894	0.915	0.874	4.564	4.778
PCDRN	29.525	0.997	0.989	0.914	4.785	4.922

produce different degrees of spectral distortion. Compared with the above methods, the color of the fused image of the PCDRN method proposed in this paper is more natural and closest to the reference image.

In addition, to judge the advantages and disadvantages of each fusion method more intuitively and conveniently, this paper uses residual data for comparison. The residual data is obtained by calculating the absolute value of the dif-

ference between the reference image and the fused image [24]. Figure 6 shows the experimental results and the residual data of the reference image.

In Figure 6, among all the methods, the PCDRN method proposed in this paper has the least residual information in the residual data.

To further evaluate the performance of various fusion methods accurately, this paper also uses objective indicators to evaluate. Table 1 shows the corresponding quantitative evaluation results of the fusion results in Figure 6.

It can be seen from Table 1 that the PCDRN method proposed in this paper is superior to the other four fusion methods in all objective indicators.

(2) Fusion experiment based on IKONOS simulation data

In this paper, image fusion experiments are carried out on the IKONOS simulation data. Figure 7 shows a set of fusion results obtained using the four contrast fusion methods and

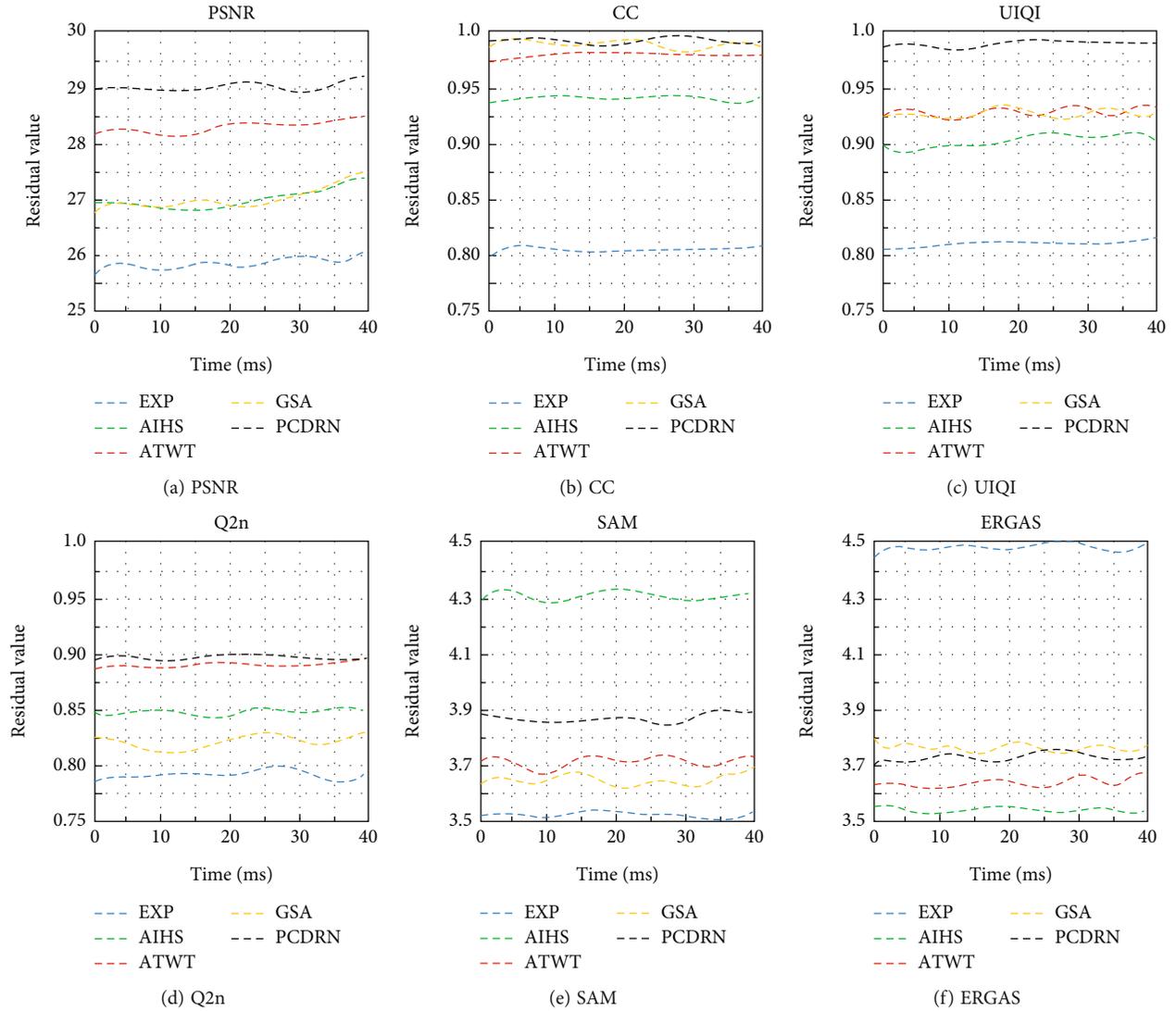


FIGURE 7: Experimental results of simulation data.

TABLE 2: Objective evaluation of fusion results (2).

Method	PSNR	CC	UIQI	Q2n	SAM	ERGAS
EXP	24.242	0.815	0.812	0.784	4.041	4.447
AIHS	27.275	0.947	0.914	0.850	4.327	3.523
ATWT	28.126	0.985	0.932	0.866	3.737	3.665
GSA	27.674	0.987	0.928	0.829	3.615	3.757
PCDRN	29.557	0.993	0.992	0.895	3.823	3.725

the PCDRN method proposed in this paper on the IKONOS simulation dataset.

The image obtained by the EXP method has obvious blur. The fusion images of AIHS and GSA have obvious color distortion. From the image enlargement area, other traditional fusion methods have serious spectral distortion. Through subjective observation, the PCDRN method proposed in this paper is visually superior to other fusion methods.

Table 2 shows the objective evaluation results of the respective fused image results in Figure 7.

TABLE 3: Average quantitative results of simulation data.

Satellite	Method	PSNR	cc	UIQI	Q2n	SAM	ERGAS
Pleiades	EXP	26.942	0.971	0.924	0.805	3.113	3.955
	AIHS	28.257	0.927	0.931	0.812	3.442	3.452
	ATWT	28.724	0.905	0.914	0.817	3.338	3.224
	GSA	27.979	0.935	0.915	0.823	3.525	3.506
	PCDRN	28.924	0.934	0.993	0.895	3.786	3.878
IKONOS	EXP	26.042	0.991	0.887	0.719	3.423	4.117
	AIHS	27.507	0.912	0.924	0.808	3.847	3.741
	ATWT	28.127	0.914	0.916	0.837	4.127	3.511
	GSA	27.047	0.921	0.935	0.831	4.624	3.958
	PCDRN	27.922	0.992	0.995	0.894	4.897	4.055

In Table 2, compared with other fusion methods, the PCDRN method proposed in this paper obtains the optimal values in all objective evaluation indexes.

- (3) Averaged quantitative experiments based on simulation dataset

To further verify the effectiveness of the PCDRN proposed in this paper, the average quantitative experiments are conducted on the Pl Pléiades and IKONOS simulation datasets, and the four fusion methods in the previous experiments are compared with the PCDRN method proposed in this paper. The average quantitative results of the experiments are shown in Table 3.

In Table 3, the PCDRN method proposed in this paper achieves the optimal values in most of the evaluation indexes.

To sum up, the experiments based on Pl Pléiades and IKONOS simulation datasets can conclude that the performance of the PCDRN method proposed in this paper is better than other fusion methods.

5. Conclusion

In this paper, a new end-to-end progressive cascaded deep residual network is proposed, which is used for remote sensing image fusion for CNN. The main work includes the following.

The PCDRN method proposed in this paper first uses two cascaded residual subnetworks to extract effective feature information in high-dimensional and low-dimensional feature spaces step by step in a coarse-to-fine manner. To avoid the oversmoothing problem of the fused image, a joint loss function is designed to prevent the loss of high-quality details and obtain high-quality fusion results.

The resize-convolution method is used to replace the transposed convolution for upsampling to eliminate the checkerboard effect in the fused image.

Through a large number of experiments on the simulated data and real data of three satellites, Pl Pléiades, IKONOS, and WorldView-3 (4 bands), it is shown that the PCDRN method proposed in this paper is superior to other advanced remote sensing image fusion methods in both subjective and objective evaluation. The improvement is about 0.001, 0.015, 0.3, and 0.8% on the standard test dataset, respectively. The experimental results also show that this method can reconstruct the details of the image better.

To make up for the large spatial resolution gap between the two datasets used in the spatio-temporal fusion of remote sensing images, a two-stage CGAN network model is used to establish the corresponding relationship between MODIS and Landsat. But this training method will have the cumulative error of the two network models to eliminate the impact of this cumulative error. It is hoped that the following work can directly establish an end-to-end network between MODIS and Landsat to replace the current method of learning the correspondence between two data in a two-stage way and automatically learn this complex mapping relationship through an end-to-end network.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Authors' Contributions

The authors of the manuscript "Remote Sensing Image Fusion Method Based on Progressive Cascaded Deep Residual Network" declare the following contributions to the creation of the manuscript: Jing Li is responsible for the conceptualization, resources, methodology, and writing; Hongwu Hu for the supervision and project administration; Lilan Lei for the original draft and writing—review and editing; and Jin Li for the resources and review.

References

- [1] L. Tan, W. Liu, L. Li, and Z. Pan, "A fast computational approach for illusory contour reconstruction," *Multimedia Tools and Applications*, vol. 78, no. 8, pp. 10449–10472, 2019.
- [2] Q. Ma and D. Kong, "A new variational model for joint restoration and segmentation based on the Mumford-Shah model," *Journal of Visual Communication & Image Representation*, vol. 53, pp. 224–234, 2018.
- [3] X. Cui, K. Zheng, L. Gao, B. Zhang, D. Yang, and J. Ren, "Multiscale spatial-spectral convolutional network with image-based framework for hyperspectral imagery classification," *Remote Sensing*, vol. 11, no. 19, p. 2220, 2019.
- [4] R. Roy, T. Chakraborti, and A. S. Chowdhury, "A deep learning-shape driven level set synergism for pulmonary nodule segmentation," *Pattern Recognition Letters*, vol. 123, pp. 31–38, 2019.
- [5] G. Li, Z. Liu, R. Shi, and W. Wei, "Constrained fixation point based segmentation via deep neural network," *Neurocomputing*, vol. 368, pp. 180–187, 2019.
- [6] H. Song, Q. Liu, G. Wang, R. Hang, and B. Huang, "Spatiotemporal satellite image fusion using deep convolutional neural networks," *IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing*, vol. 11, no. 3, pp. 821–829, 2018.
- [7] C. Ledig, L. Theis, F. Huszár et al., "Photo-realistic single image super-resolution using a generative adversarial network," in *Proceeding GS of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4681–4690, Honolulu, HI, USA, 2017.
- [8] X. Zhang, H. Song, K. Zhang, J. Qiao, and Q. Liu, "Single image super-resolution with enhanced Laplacian pyramid network via conditional generative adversarial learning," *Neurocomputing*, vol. 398, pp. 531–538, 2020.
- [9] W. Lai, J. Huang, N. Ahuja, and M. H. Yang, "Fast and accurate image super-resolution with deep Laplacian pyramid networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 11, pp. 2599–2613, 2018.
- [10] D. Bretreger, I. Y. Yeo, J. Quijano, J. Awad, G. Hancock, and G. Willgoose, "Monitoring irrigation water use over paddock scales using climate data and landsat observations," *Agricultural Water Management*, vol. 221, pp. 175–191, 2019.
- [11] Z. Hou, R. E. McRoberts, G. Ståhl, P. Packalen, J. A. Greenberg, and Q. Xu, "How much can natural resource inventory benefit

- from finer resolution auxiliary data?," *Remote Sensing of Environment*, vol. 209, pp. 31–40, 2018.
- [12] Y. Fu, J. Li, Q. Weng et al., "Characterizing the spatial pattern of annual urban growth by using time series Landsat imagery," *Science of the Total Environment*, vol. 666, pp. 274–284, 2019.
- [13] M. Meroni, D. Fasbender, F. Rembold, C. Atzberger, and A. Klisch, "Near real-time vegetation anomaly detection with MODIS NDVI: timeliness vs. accuracy and effect of anomaly computation options," *Remote Sensing of Environment*, vol. 221, pp. 508–521, 2019.
- [14] A. Habib, B. Chen, B. Khalid et al., "Estimation and inter-comparison of dust aerosols based on MODIS, MISR and AERONET retrievals over Asian desert regions," *Journal of Environmental Sciences*, vol. 76, pp. 154–166, 2019.
- [15] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceeding GS of the IEEE Conference on Computer Vision and Ppattern Recognition*, pp. 7132–7141, Salt Lake City, UT, USA, 2018.
- [16] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Proceedings of the AAAI conference on artificial intelligence (Vol. 31, No. 1)*, San Francisco, California, USA, 2017.
- [17] J. Dai, H. Qi, Y. Xiong et al., "Deformable convolutional networks," in *Proceeding GS of the IEEE International Conference on Computer Vision*, pp. 764–773, Venice, Italy, 2017.
- [18] L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 834–848, 2017.
- [19] Q. Dai, B. Luo, L. Wang, and Z. Tu, "Context adaptive panchromatic band simulation and detail injection for image pancharpening," *Journal of Applied Remote Sensing*, vol. 12, no. 1, pp. 1–19, 2018.
- [20] Y. Pan, X. Li, A. Gao, L. Li, S. Mei, and S. Yue, "A new pansharpening method with multi-scale structure perception," in *Proceeding GS of 2018 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 8046–8049, Valencia, Spain, 2018.
- [21] J. Jiao, X. Gong, L. Wu, and X. Meng, "Spectral-driven pansharpening using adaptive image segmentation to reduce color distortion," in *Proceedings of 2019 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 3133–3136, Yokohama, Japan, 2019.
- [22] P. Liu, "Joint spectral and spatial consistency priors for variational pansharpening," *IEEE Access*, vol. 7, pp. 174847–174858, 2019.
- [23] G. Khademi and H. Ghassemian, "A variational pan sharpening algorithm based on total variation and primal-dual optimization," in *Proceeding GS of 4th International Conference on Pattern Recognition and Image Analysis (IPRIA)*, pp. 64–69, Tehran, Iran, 2019.
- [24] X. Meng, H. Shen, Q. Yuan, H. Li, and L. Zhang, "An integrated fusion framework for joint information reconstruction and resolution enhancement," *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, vol. XLII-2/W7, no. 2, pp. 831–835, 2017.