Remote Sensing Pan-Sharpening via Cross-Spectral–Spatial Fusion Network

Yu Wang, Zhenfeng Shao, Tao Lu, *Member, IEEE*, Jiaming Wang, Gui Cheng, Xiaolong Zuo, and Chaoya Dang

***Abstract*— Pan-sharpening is a technique used to create high-resolution multispectral (HRMS) images by merging low-resolution multispectral (LRMS) images with corresponding high-resolution panchromatic (PAN) images. Despite achieving state-of-the-art performance, existing panchromatic sharpening networks based on deep-learning (DL) methods suffer from spectral distortion and insufficient spatial texture enhancement. To address this challenge, this letter introduces a novel cross-spectral–spatial fusion network (CSSFN) for pan-sharpening remote sensing images. The network utilizes a cross-spectral–spatial attention block (CSSAB) to extract both the spectral information of the MS branch and the spatial information of the PAN branch. The spectral and spatial feature representations of remote sensing images are then progressively enhanced, which improves the fusion process and generates multispectral images with high spatial resolution. Our network outperforms other pan-sharpening methods on two publicly available datasets, as demonstrated by extensive experiments, yielding state-of-the-art results.**

***Index Terms*— Deep neural networks, pan-sharpening, remote sensing, residual fusion network.**

I. Introduction

HIGH-RESOLUTION multispectral (HRMS) remote sensing images play a critical role in geographic information management and are widely used in various fields, including change detection [1], [2], scene classification [3], [4], environmental pollution monitoring [5], and land surveying [6]. However, due to the limitations of imaging equipment, satellites often capture low-resolution multispectral (LRMS) images and high-resolution panchromatic (PAN) images. Compared to MS images, PAN images have the characteristics of high spatial resolution and a single band. To obtain HRMS images, researchers employ pan-sharpening technology to merge the spectral information of LRMS images with the spatial information of PAN images. As a result, the pan-sharpening algorithm is a crucial fusion technology for remote sensing images in practical applications.

Remote sensing pan-sharpening methods can be broadly classified into four categories: component substitution (CS), multiresolution analysis (MRA), variational optimization (VO), and deep-learning (DL) methods. Among these, CS-based methods transfer the LRMS image to a specific domain, replace the spatial information component of the LRMS with that of the PAN image, and then transform it back to the HRMS image space to obtain the final fused result. Common CS-based methods include intensity-hue-saturation [7], principal component analysis [8], adaptive component replacement methods [9], and others. MRA-based methods perform multiscale decomposition on PAN images to extract spatial details and incorporate them into MS images. Common representative methods are based on wavelet transform [10] and discrete wavelet transform [11]. VO-based methods rely on known prior information and apply appropriate algorithms to enforce reasonable constraints. Common methods include sparse representation-based methods [12] and slice smoothness-based methods [13], among others.

With the significant advancements in DL technology, remote sensing pansharpening methods have made substantial progress in the realm of DL [14], [15], [16]. However, existing DL-based remote sensing pan-sharpening methods face the challenge of insufficient feature extraction, primarily due to the following two aspects. On the one hand, remote sensing images exhibit complex terrains and environmental changes, thus demanding powerful feature extraction capabilities to capture this complexity. DL networks are capable of automatically learning nonlinear features but require sufficient depth and complexity to achieve comprehensive feature extraction capabilities. On the other hand, remote sensing pan-sharpening tasks require networks to simultaneously process various types of information, such as high-resolution spectral information and LRMS information. Inappropriate network architecture

Manuscript received 17 October 2023; accepted 27 November 2023. Date of publication 30 November 2023; date of current version 19 December 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 42090012; in part by the Guangxi Science and Technology Plan Project under Grant Guike 2021AB30019; in part by the Hubei Province Key Research and Development Project under Grant 2022BAA048; in part by the Sichuan Province Key Research and Development Project under Grant 2022YFN0031, Grant 2023YFN0022, and Grant 2023YFS0381; in part by the Zhuhai Industry-University-Research Cooperation Project under Grant ZH22017001210098PWC; in part by the Shanxi Provincial Science and Technology Major Special Project under Grant 202201150401020; in part by the Guangxi Key Laboratory of Spatial Information and Surveying and Mapping Fund Project under Grant 21-238-21-01; and in part by the Hubei Provincial Natural Science Foundation of China under Grant 2023AFB158. *(Corresponding author: Zhenfeng Shao.)*

Yu Wang is with the State Key Laboratory for Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China, and also with the School of General Aviation, Jingchu University of Technology, Jingmen 448000, China (e-mail: wangyu949374585@gmail.com).

Zhenfeng Shao, Gui Cheng, Xiaolong Zuo, and Chaoya Dang are with the State Key Laboratory for Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China (e-mail: shaozhenfeng@whu.edu.cn; chenggui@whu.edu.cn; zuoxiaolong@whu.edu.cn; dangchaoya@whu.edu.cn).

Tao Lu and Jiaming Wang are with the Hubei Key Laboratory of Intelligent Robot, Wuhan Institute of Technology, Wuhan 430200, China (e-mail: lutxyl@gmail.com; wjmecho@whu.edu.cn).

Digital Object Identifier 10.1109/LGRS.2023.3337844

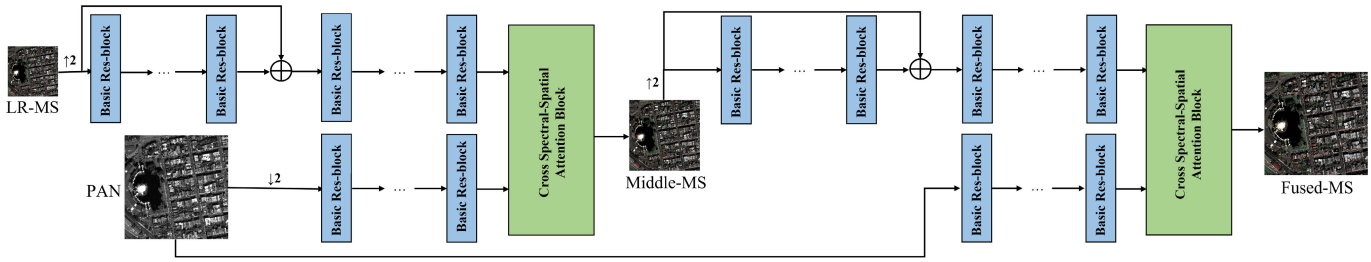

Fig. 1. Overall structure of the proposed CSSFN, where ⊕ denotes element-wise sum.

TABLE I

COMPARISON OF OBJECTIVE EVALUATION METRICS OF FUSED RESULTS ON THE WORDVIEW II DATASET. THE TOP PERFORMANCE IS DENOTED BY RED FONTS, FOLLOWED BY THE SECOND BEST REPRESENTED BY BLUE, WHILE THE THIRD BEST PERFORMANCE IS INDICATED BY A HORIZONTAL LINE

Algorithms	PSNR/dB ↑	SSIM ↑	SAM ↓	ERGAS ↓	SCC ↑	Q ↑	D_λ ↓	D_s ↓	QNR ↑
GSA	20.03	0.6677	0.2538	17.6188	0.8188	0.6342	0.0382	0.1969	0.7733
MTF-GLP-HPM	17.56	0.5713	0.2037	27.0295	0.7268	0.5293	0.0922	0.1345	0.7847
SRPPNN	22.27	0.5841	0.1270	20.0058	0.7158	0.4668	0.0370	0.1105	0.8572
MSDCNN	25.45	0.8314	0.0597	10.1998	0.9041	0.7456	0.0208	0.0705	0.9101
GPPNN	24.81	0.8147	0.0622	11.0969	0.8858	0.7201	0.0209	0.0776	0.9031
HMCNN	22.33	0.5845	0.1189	17.9181	0.6898	0.4562	0.0272	0.1399	0.8371
LAG	25.26	0.8206	0.0694	10.5425	0.8966	0.7270	0.0186	0.0736	0.9092
Band-Aware	24.92	0.8092	0.0761	11.1488	0.8841	0.7119	0.0187	0.0721	0.9106
Ours	25.56	0.8265	0.0714	10.1883	0.9053	0.7352	0.0166	0.0786	0.9062

design may result in information loss or inadequate feature extraction. Therefore, network architecture design needs to thoroughly consider how to effectively integrate and utilize different types of information.

To address these issues, this letter proposes a novel cross-spectral–spatial fusion network for pan-sharpening. This network leverages deeper multilayer residual modules to fully extract remote sensing features. Additionally, for high-resolution spectral information and LRMS information, we employ corresponding spatial attention blocks and spectral attention blocks on different branches of remote sensing data. Such a network design can more effectively integrate and utilize different types of information to resolve the aforementioned feature extraction limitations. Furthermore, inspired by super-resolution algorithms in low-level vision tasks [17], [18], our overall network adopts a progressive learning approach. This learning method decomposes the complex high-upscaling task into multiple steps, significantly reducing the learning difficulty, and achieving improved performance.

In summary, this letter makes the following contributions.

1) We propose a novel cross-spectral–spatial fusion network (CSSFN) for pan-sharpening remote sensing images. This network mainly adopts a dual-branch feature extraction method and a progressive learning approach to fuse multilevel features to obtain the final fused image.
2) We design a novel cross-spectral–spatial attention block (CSSAB) that can better extract the spectral feature information of LRMS images and the spatial feature information of PAN images. The superiority of the method proposed in this letter is effectively verified on the public dataset from two aspects: subjective visual quality and objective evaluation indicators.

II. PROPOSED METHOD

A. Network Architecture

As shown in Fig. 1, the proposed CSSFN network is composed of a feature extraction branch and a feature fusion module. The feature extraction of this network is divided into two branches, MS and PAN. The feature fusion module is composed of a CSSAB. The entire network model of CSSFN is divided into multilevel learning stages. Specifically, in the first-level learning stage, the MS branch inputs the LRMS image I_{LRMS} obtained by downsampling the original MS image I_{MS}. Then, it passes through the feature extraction branch composed of basic residual blocks $\mathcal{F}_{\text{res}}(.)$, whose feature extraction can be expressed as

$$\mathcal{M}_0 = \mathcal{F}_{\text{res}}\left(I_{\text{LRMS}\uparrow 2}\right) \tag{1}$$

where $\mathcal{M}_0$ represents the rough remote sensing features of $I_{\text{LRMS}\uparrow 2}$, and $I_{\text{LRMS}\uparrow 2}$ is obtained by upsampling I_{LRMS} through bicubic interpolation. Then, MS remote sensing image features are extracted through 11 basic residual blocks and added

TABLE II

COMPARISON OF OBJECTIVE EVALUATION METRICS OF FUSED RESULTS ON THE QUICKBIRD DATASET. THE TOP PERFORMANCE IS DENOTED BY RED FONTS, FOLLOWED BY THE SECOND BEST REPRESENTED BY BLUE, WHILE THE THIRD BEST PERFORMANCE IS INDICATED BY A HORIZONTAL LINE

Algorithms	PSNR/dB ↑	SSIM ↑	SAM ↓	ERGAS ↓	SCC ↑	Q ↑	D_λ ↓	D_s ↓	QNR ↑
GSA	25.58	0.7791	0.1068	6.4901	0.9136	0.7094	0.0224	0.1673	0.8137
MTF-GLP-HPM	22.25	0.7328	0.1069	10.902	0.8626	0.6502	0.0655	0.1163	0.8238
SRPPNN	24.68	0.7954	0.1242	6.0816	0.9573	0.7566	0.0578	0.1582	0.7926
MSDCNN	27.13	0.8748	0.0878	4.3263	0.9764	0.8520	0.0360	0.1609	0.8086
GPPNN	26.29	0.8704	0.0893	4.7485	0.9720	0.8484	0.0399	0.1597	0.8065
HMCNN	26.95	0.8690	0.0871	4.4449	0.9745	0.8435	0.0479	0.1612	0.7979
LAG	27.12	0.8758	0.0831	4.3764	0.9757	0.8511	0.0569	0.1675	0.7841
Band-Aware	27.18	0.8741	0.0862	4.3282	0.9759	0.8504	0.0513	0.1616	0.7945
Ours	27.34	0.8810	0.0801	4.3085	0.9766	0.8597	0.0468	0.1549	0.8049

TABLE III

COMPARISON OF ABLATION EXPERIMENT INDICATORS ON THE WORDVIEW II DATASET. SPE-AB AND SPA-AB REPRESENT THE SPECTRAL ATTENTION BLOCK AND THE SPATIAL ATTENTION BLOCK, RESPECTIVELY

Spe-AB	Spa-AB	PSNR/dB ↑	SSIM ↑	SAM ↓	ERGAS ↓	SCC ↑	Q ↑	D_λ ↓	D_s ↓	QNR ↑
✓	×	22.47	0.6000	0.1465	18.2557	0.7462	0.4800	0.0408	0.0727	0.8895
×	✓	23.02	0.6117	0.1415	18.468	0.7608	0.4879	0.0473	0.0781	0.8786
✓	✓	25.56	0.8265	0.0714	10.1883	0.9053	0.7352	0.0166	0.0786	0.9062

to their features through residual connections. The feature learning process can be expressed as

$$\mathcal{M}_{\mathrm{LRMS}\uparrow 2} = \mathcal{F}_{\mathrm{res}}\left(I_{\mathrm{LRMS}\uparrow 2} + M_0\right) \tag{2}$$

where $\mathcal{M}_{\mathrm{LRMS}\uparrow 2}$ represents the fine remote sensing features of $I_{\mathrm{LRMS}\uparrow 2}$. Similar to the MS branch, the PAN branch inputs PAN images. First, the PAN image is downsampled to obtain the $I_{\mathrm{PAN}\downarrow 2}$ image, and then fine PAN image features are extracted through 11 basic residual blocks. Finally, the image features extracted by the two branches are fed into the CSSAB to achieve 2× multispectral remote sensing image reconstruction

$$\mathcal{M}_{\mathrm{PAN}\downarrow 2} = \mathcal{F}_{\mathrm{res}}\left(I_{\mathrm{PAN}\downarrow 2}\right) \tag{3}$$

$$\mathcal{M}_{M-\mathrm{MS}} = \mathcal{F}_{\mathrm{cssab}}\left(\mathcal{M}_{\mathrm{LRMS}\uparrow 2}, \mathcal{M}_{\mathrm{PAN}\downarrow 2}\right) \tag{4}$$

where $\mathcal{M}_{\mathrm{PAN}\downarrow 2}$ represents the remote sensing features of $I_{\mathrm{PAN}\downarrow 2}$, $\mathcal{F}_{\mathrm{cssab}}(.)$ represents the feature fusion process of the CSSAB, and $\mathcal{M}_{M-\mathrm{MS}}$ represents the middle multispectral remote sensing feature.

To achieve 4× remote sensing image reconstruction of I_{LRMS} images, the second learning stage of the CSSFN network involves further upsampling of the multispectral remote sensing images reconstructed in the first learning stage. The feature learning process is similar to the previous stage, and the reconstruction process can be expressed as

$$\mathcal{M}_1 = \mathcal{F}_{\mathrm{res}}\left(M_{M-\mathrm{MS}\uparrow 2}\right) \tag{5}$$

$$\mathcal{M}_{F-\mathrm{MS}} = \mathcal{F}_{\mathrm{cssab}}\left(\mathcal{F}_{\mathrm{res}}\left(M_{M-\mathrm{MS}\uparrow 2} + \mathcal{M}_1\right), \mathcal{M}_{\mathrm{PAN}}\right) \tag{6}$$

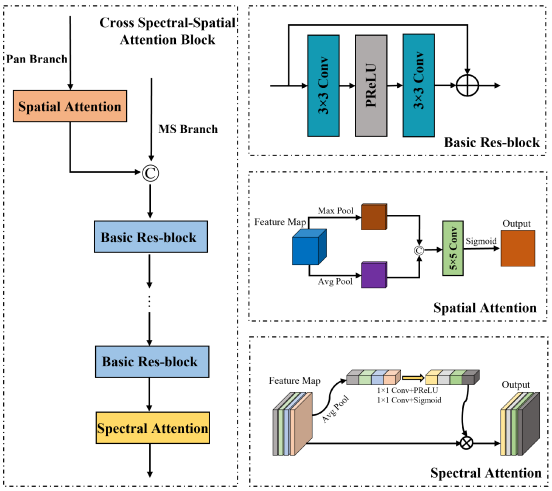

Fig. 2. Proposed CSSAB, ⓒ and ⊗ denote concatenate and element-wise product, respectively.

among them, $\mathcal{M}_1$ and $\mathcal{M}_{F-MS}$, respectively, represent the image features of $M_{M-\mathrm{MS}\uparrow 2}$ and the final 4× fused MS image features.

B. Cross-Spectral–Spatial Attention Block

The spatial information injection method in pan-sharpening is one of the fusion methods of PAN image and MS image, which mainly injects the extracted spatial remote sensing information into the LRMS image as losslessly as possible. Motivated by Wang et al. [19], we design a novel CSSAB to

better fuse the LRMS images and the PAN images. It should be noted that the difference from [19] is that the CSSAB designed by this network can better avoid spectral distortion and enhance spatial remote sensing information.

The CSSAB structure proposed in this letter is shown in Fig. 2. First, the PAN remote sensing image features extracted by the PAN branch are input into the spatial attention block to extract the unique spatial remote sensing information of the PAN image. Then, the MS remote sensing image feature map extracted by the MS branch is concatenated, and the middle multispectral remote sensing feature map is obtained through the spectral attention block. The fusion process can be expressed as

$$\mathcal{M}_{M-\mathrm{MS}} = \mathcal{F}_{\mathrm{spe}}\left(\mathcal{F}_{\mathrm{res}}\left(\mathrm{concat}(\mathcal{F}_{\mathrm{spa}}(\mathcal{M}_{\mathrm{PAN}\downarrow 2}), \mathcal{M}_{\mathrm{LRMS}\uparrow 2})\right)\right) \tag{7}$$

where $\mathcal{F}_{\mathrm{spe}}(.)$ and $\mathcal{F}_{\mathrm{spa}}(.)$ denote the spectral attention block and spatial attention block, respectively. Finally, the proposed CSSFN model uses CSSAB in the fusion stage of two scales to achieve $4\times$ fusion results.

III. Experimental Results

A. Datasets and Experimental Settings

To verify the effectiveness of our network, we conducted extensive experiments on WordView II and QuickBird datasets. Both datasets contain four bands of red, green, blue, and near-infrared. We block the HRMS images and PAN images of the two datasets and cropped them to $256 \times 256 \times 4$ pixels and $256 \times 256 \times 1$ pixels, respectively. Then, the HRMS images are downsampled by bicubic interpolation to obtain the LRMS images, whose pixels are $64 \times 64 \times 4$.

Our network is implemented on an Ubuntu 18.04 system with two GeForce RTX 2080TI GPUs. This network uses the Adam optimizer to optimize the model. We set the initial learning rate to 1e−4. The training process of the entire network is set to 2000 epochs, and the final test uses the best model saved during the training process for testing. To evaluate the superiority of the fused image, we use the objective evaluation index of the reference image and the objective evaluation index of the nonreference image. The objective evaluation index of the reference image includes PSNR, SSIM, SAM, ERGAS, SCC, Q, while the objective evaluation index of the nonreference image includes D_λ, D_s, QNR.

B. Comparisons With Other Methods

To validate the superiority of our network, we compared it to eight pan-sharpening fusion algorithms on two different datasets: GSA [20], MTF-GLP-HPM [21], SRPPNN [22], MSDCNN [15], GPPNN [23], HMCNN [19], LAG [24], and Band-Aware [25]. Tables I and II present the objective evaluation metrics for the two datasets, while subjective reconstruction results are displayed in Figs. 3 and 4. From Tables I and II, we can see that our CSSFN model has only achieved partial improvements on the WordView II dataset. However, on the QuickBird dataset, most of the evaluation indicators have outperformed other algorithms. By combining the experimental results of both datasets, we can confirm that our CSSFN model is more general and can guarantee superior performance in pan-sharpening. This implies that the CSSFN model can better preserve spatial detail information and spectral feature information during the fusion process while avoiding spectral distortion.

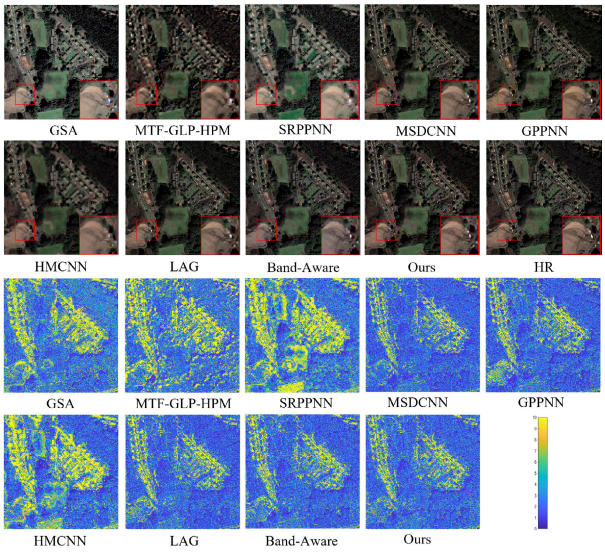

Fig. 3. Proposed CSSFN compares the subjective results and error results of eight pan-sharpening methods on the WordView II dataset. The larger the value, the worse the fusion effect.

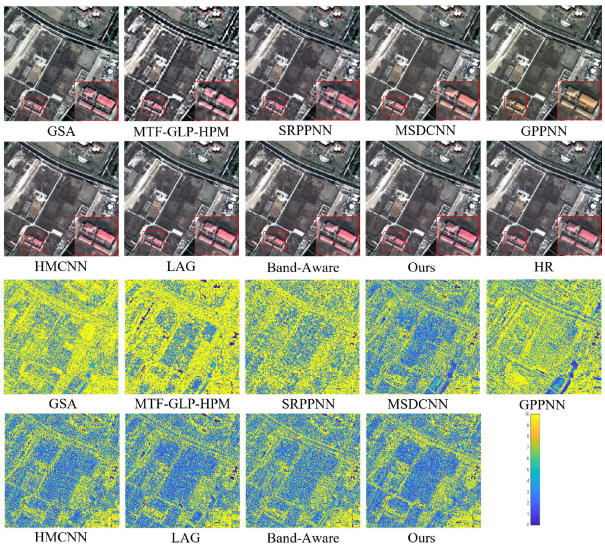

Fig. 4. Proposed CSSFN compares the subjective results and error results of eight pan-sharpening methods on the QuickBird dataset. The larger the value, the worse the fusion effect.

To provide a more intuitive comparison of the experimental results of different model fusions, we selected a test sample from both the WordView II and QuickBird datasets and arranged them to display the pan-sharpening results. To emphasize the details of the subjective results, we calculated a graph of the mean squared error (MSE) between the reconstructed result and the actual value. A smaller MSE value indicates a result closer to the actual value. The experimental results show that our model also exhibits certain advantages in subjective comparison.

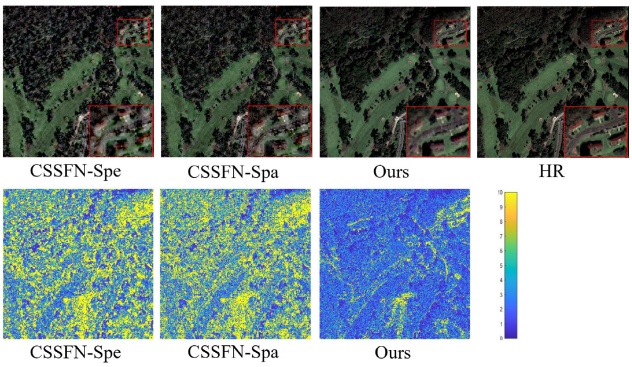

Fig. 5. Comparison of subjective results of different combinations of ablation experiments on the WordView II dataset.

C. Ablation Studies

This network proposes a new CSSAB to extract spectral feature information and spatial detail information more effectively. To verify the efficacy of the model, we conducted ablation experiments by combining different components, and the results are presented in Table III. As shown in Table III, retaining only the spatial attention block resulted in a decrease in each evaluation index, indicating that the spectral attention block improves the pan-sharpening ability of the model by effectively expressing the spectral feature information. Similarly, keeping only the spectral attention block led to a decrease in the objective evaluation indicators, demonstrating the contribution of the spatial attention block to the extraction of spatial feature information. However, when both the spectral attention block and spatial attention block are incorporated into the network, the objective evaluation metrics improve. The comparison of subjective experimental results is shown in Fig. 5, and it is obvious from Fig. 5 that the effectiveness of our final model can be drawn. Therefore, the results of the ablation experiments indicate that these two modules are not mutually exclusive in the CSSFN network, and their combination leads to superior pan-sharpening performance.

IV. Conclusion

In this letter, we introduce a novel network for pan-sharpening remote sensing images, named the CSSFN. This network constructs CSSAB with high feature preservation by placing spectral attention blocks and spatial attention blocks on separate branches of remote sensing data. Moreover, we employ a multilevel progressive learning approach to further enhance the performance of remote sensing image pan-sharpening. Our experimental results on WordView II and QuickBird datasets show that CSSFN outperforms existing state-of-the-art pan-sharpening methods. Due to limitations in real remote sensing data fusion within this network, we will steer future research toward an unsupervised direction to better handle unlabeled real remote sensing data.

References

[1] T. Bai et al., "Deep learning for change detection in remote sensing: A review," *Geo-Spatial Inf. Sci.*, vol. 26, no. 3, pp. 262–288, 2022.

[2] Q. Zhu, X. Guo, Z. Li, and D. Li, "A review of multi-class change detection for satellite remote sensing imagery," *Geo-Spatial Inf. Sci.*, pp. 1–15, 2022, doi: 10.1080/10095020.2022.2128902.

[3] L. Zhou, Z. Shao, S. Wang, and X. Huang, "Deep learning-based local climate zone classification using Sentinel-1 SAR and Sentinel-2 multispectral imagery," *Geo-spatial Inf. Sci.*, vol. 25, no. 3, pp. 383–398, Jul. 2022.

[4] N. Ammour, Y. Bazi, H. Alhichri, and N. Alajlan, "Continual learning approach for remote sensing scene classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[5] Q. Zhuang et al., "Isolating the direct and indirect impacts of urbanization on vegetation carbon sequestration capacity in a large oasis city: Evidence from Urumqi, China," *Geo-spatial Inf. Sci.*, vol. 26, no. 3, pp. 379–391, Jul. 2023.

[6] R. Li, S. Zheng, C. Duan, L. Wang, and C. Zhang, "Land cover classification from remote sensing images based on multi-scale fully convolutional network," *Geo-Spatial Inf. Sci.*, vol. 25, no. 2, pp. 278–294, Apr. 2022.

[7] T.-M. Tu, S.-C. Su, H.-C. Shyu, and P. S. Huang, "A new look at IHS-like image fusion methods," *Inf. Fusion*, vol. 2, no. 3, pp. 177–186, Sep. 2001.

[8] V. P. Shah, N. H. Younan, and R. L. King, "An efficient pan-sharpening method via a combined adaptive PCA approach and contourlets," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1323–1335, May 2008.

[9] 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.

[10] R. L. King and J. Wang, "A wavelet based algorithm for pan sharpening Landsat 7 imagery," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2001, pp. 849–851.

[11] Y. Zhang, "A new merging method and its spectral and spatial effects," *Int. J. Remote Sens.*, vol. 20, no. 10, pp. 2003–2014, Jan. 1999.

[12] 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.

[13] L.-J. Deng, M. Feng, and X.-C. Tai, "The fusion of panchromatic and multispectral remote sensing images via tensor-based sparse modeling and hyper-Laplacian prior," *Inf. Fusion*, vol. 52, pp. 76–89, Dec. 2019.

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

[15] 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.

[16] X. Liu, Q. Liu, and Y. Wang, "Remote sensing image fusion based on two-stream fusion network," *Inf. Fusion*, vol. 55, pp. 1–15, Mar. 2020.

[17] Y. Wang, Z. Shao, T. Lu, C. Wu, and J. Wang, "Remote sensing image super-resolution via multiscale enhancement network," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, pp. 1–5, 2023.

[18] Y. Wang et al., "A lightweight distillation CNN-transformer architecture for remote sensing image super-resolution," *Int. J. Digit. Earth*, vol. 16, no. 1, pp. 3560–3579, Oct. 2023.

[19] J. Wang, Z. Shao, X. Huang, T. Lu, R. Zhang, and J. Ma, "Pan-sharpening via high-pass modification convolutional neural network," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2021, pp. 1714–1718.

[20] 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.

[21] G. Vivone, R. Restaino, M. D. Mura, G. Licciardi, and J. Chanussot, "Contrast and error-based fusion schemes for multispectral image pansharpening," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 5, pp. 930–934, May 2014.

[22] J. Cai and B. Huang, "Super-resolution-guided progressive pansharpening based on a deep convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 6, pp. 5206–5220, Jun. 2021.

[23] S. Xu, J. Zhang, Z. Zhao, K. Sun, J. Liu, and C. Zhang, "Deep gradient projection networks for pan-sharpening," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 1366–1375.

[24] Z.-R. Jin, T.-J. Zhang, T.-X. Jiang, G. Vivone, and L.-J. Deng, "LAGConv: Local-context adaptive convolution kernels with global harmonic bias for pansharpening," in *Proc. AAAI Conf. Artif. Intell.*, 2022, vol. 36, no. 1, pp. 1113–1121.

[25] M. Zhou, K. Yan, X. Fu, A. Liu, and C. Xie, "PAN-guided band-aware multi-spectral feature enhancement for pan-sharpening," *IEEE Trans. Comput. Imag.*, vol. 9, pp. 238–249, 2023.