A FULL-COLOR SHARPENING METHOD FOR SPECTRAL FEATURE EXTRACTION WITH DYNAMIC SPATIAL FEATURE GUIDANCE

Beiyi WANG[1], Bo WANG[2]*

Remote sensing satellites are generally limited to capturing either high-resolution panchromatic images (PAN) or low-resolution multispectral images (LRMS). To integrate the benefits of both high resolution and multispectral data, this study introduces a novel full-color sharpening technique that employs dynamic spatial feature guidance for spectral feature extraction. This technique utilizes a dynamic convolutional approach between the full-color and low-resolution panchromatic images, effectively guiding the feature extraction process and enhancing the spatial detail in the low-resolution panchromatic image. Additionally, a cross-guided fusion module has been crafted, harnessing the power of cross-attention mechanisms to enrich the fusion of features from different sources, thereby elevating the fusion quality, The proposed method in this study adeptly mitigates the adverse effects of blurred spatial information inherent in low-resolution panchromatic images, culminating in the production of high-resolution panchromatic images that boast remarkable spatial clarity.

Keywords: Remote sensing satellites, panchromatic images, low-resolution multispectral images, full-color sharpening technique, spectral feature extraction, cross-guided fusion module.

1. Introduction

Remote sensing technology detects electromagnetic waves of different wavelengths by reflecting, absorbing, and emitting spectral information of ground objects, in order to obtain remote sensing images [1]. Remote sensing satellites typically can only obtain high-resolution panchromatic images (PAN) or low resolution multispectral images (LRMS). High resolution panchromatic images are single band gray images with high spatial resolution, containing edge details of scenery, but lacking color information of the scenery. Other then, a low-resolution multispectral image is a color image encompassing multiple spectral bands, providing rich spectral information that aids in differentiating the colors of various objects. However, its spatial resolution is low and it cannot clearly

[1] Associate Prof., New Quality Productivity and vocational education Development Institute, Guangdong Polytechnic, Foshan, China, e-mail: 23402891@qq.com
[2] Associate Prof., New Quality Productivity and vocational education Development Institute, Guangdong Polytechnic, Foshan, China, e-mail: hust_wb@126.com

identify edge information of objects. To merge the benefits of two distinct image types, panchromatic sharpening technology has been developed [2].

Initial techniques for panchromatic sharpening encompass component substitution [3-7], multi-resolution analysis [8-12], and variational optimization [13-15]. The advent of deep learning has introduced numerous innovative approaches, notably those utilizing convolutional neural networks (CNNs), which have demonstrated encouraging outcomes. However, achieving more precise spectral and spatial information is still a challenge. GAN-based methods produce high-quality high-resolution multispectral (HRMS) images but suffer from unstable training, complex model and parameter configurations, and the need for extensive training datasets. The transformer-based approach, meanwhile, adapts the Transformer model to image processing. Wang et al. [16] enhanced this by integrating CNN's flexible position encoding into the Transformer framework, effectively extracting both local and global image features and enhancing overall task performance.

The PAN and LRMS images, possessing unique and complementary characteristics, substantially influence the enhancement of spatial details in HRMS images, with the PAN image playing a pivotal role [17]. The ongoing exploration of deep learning in full-color sharpening has made strides in addressing the limitations of traditional methods when dealing with nonlinear tasks. Nonetheless, the LRMS image, while offering rich spectral data, also brings in some spatial blurriness that can degrade the quality of the HRMS image. Consequently, crafting an algorithm capable of reducing or eliminating this blurriness from the LRMS images is of paramount importance.

To address the issue of blurred spatial information inherent in low-resolution multispectral (LRMS) images, our research delves into the development of a full-color sharpening technique that leverages dynamic spatial feature guidance for spectral feature extraction. Recognizing that traditional sharpening approaches often fail to effectively distinguish between spectral and spatial information, leading to compromised image quality, we have devised a new method aimed at overcoming this challenge. We begin by elucidating the fundamental principles of the algorithms under discussion and providing an overview of the datasets employed, laying the groundwork for our subsequent analysis. Following this, we introduce an innovative full-color sharpening approach, specifically engineered to mitigate the interference caused by the blurred spatial details in LRMS images. This method not only enhances the spatial clarity of the images but also preserves the accuracy of the spectral features. The effectiveness of our technique is rigorously tested and confirmed through a series of experiments, thereby solidifying the paper's contribution to the field.

2. Dataset and evaluation criteria

2.1. Dataset selection

In the current field of panchromatic sharpening, common datasets can be divided into private datasets and public datasets, among which public datasets have made important contributions to the development of algorithms. Therefore, this study chose WorldView-3 (WV-3) and Gaofen-2 (GF-2) in the PanCollection dataset for research. Among them, the WV-3 dataset contains 9714 pairs of training data, with LRMS image size of 16 × 16 × 8 and PAN image size of 64 × 64 × 1. The GF-2 dataset contains 19809 pairs of training data, with LRMS image size of 16 × 16 × 4 and PAN image size of 64 × 64 × 1. To assess the performance of the proposed method on both low resolution and full resolution images, two test sets were established for each dataset: one consisting of 20 pairs of 256 × 256 spatial resolution PAN images, and another with 20 pairs of PAN images at 512 × 512 spatial resolution.

The presentation will be straightforward and succinct, with any symbols utilized detailed in a legend if required. International System measurement units will be employed throughout the paper. Descriptions of equipment or setups will not be included in the document.

2.2. Quality evaluation criteria for full-color sharpening

We mainly evaluate the correlation between the panchromatic sharpening results generated from full resolution image data and the input source image, and usually use Quality with No Reference (QNR) for quantitative analysis. The greater the QNR, the lesser the total distortion of the image, resulting in superior image quality. The formal definition of QNR is:

$$\mathrm{QNR} = (1 - \mathrm{D}_{\lambda})(1 - \mathrm{D}_{\mathrm{s}}) \tag{1}$$

where D_{λ} and D_{s} represent the spectral and the spatial distortion coefficient, respectively. Each of them has a lower value indicating less image distortion. The detail definition of D_{λ} and D_{s} are described as follows:

$$\mathrm{D}_{\lambda} = \sqrt{\frac{1}{\mathrm{B}(\mathrm{B}-1)} \sum_{b=1}^{B} \sum_{\substack{j=1 \\ j \neq b}}^{B} \left| Q(X_b, X_j) - Q(l_b, l_j) \right|} \tag{2}$$

$$D_s = \sqrt{\frac{1}{B} \sum_{b=1}^{B} \left| Q(X_b, p) - Q(l_b, \bar{p}) \right|} \tag{3}$$

where p and l are the original PAN and LRMS images, respectively. $\bar{p}$ is the downsampled version of the original PAN image, and $Q(\bullet)$ represents the image quality index (IQI) [16], which can be described as follows.

$$Q(Iref, Itest)=\alpha \times PSNR(Iref, Itest)+(1-\alpha) \times SSIM(Iref, Itest) \quad (4)$$

where *Iref* is the reference image, *Itest* is the test image, and α is a weighting coefficient between 0 and 1, used to balance the importance of *PSNR* and *SSIM*. A higher *PSNR* value indicates better image quality, and an *SSIM* value closer to 1 suggests a higher structural similarity between the test image and the reference image.

3. Method

3.1. Overall framework

In this study, we propose a panchromatic sharpening method for spectral feature extraction guided by spatial feature dynamics. Fig. 1 shows the overall network architecture of our method. The principal constituents of the network comprise the PAN Feature Extraction Sub-Network, the LRMS Feature Extraction Sub-Network, the Dynamic Weight Generation Module (DWGM), the Cross-Guided Fusion Module (CGFM), and the Reconstruction Module (RM).

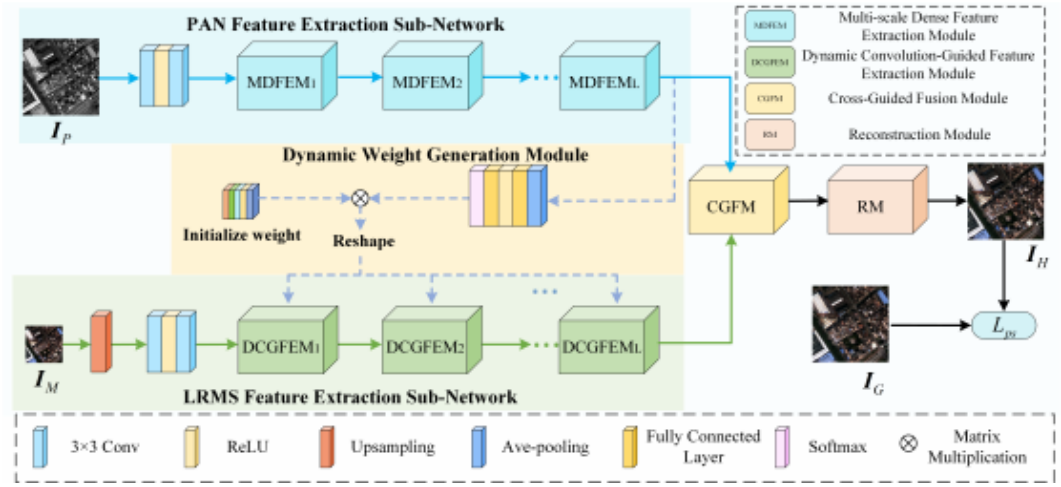

Fig. 1. The overall network architecture of our method

3.2. PAN image feature extraction

The objective of the PAN image feature extraction is to capture the spatial features of the PAN image, which are essential for the subsequent generation of dynamic weights and the reconstruction and restoration of the HRMS image.

Comprising two principal elements, this sub-network includes a shallow feature extraction layer and a series of L cascaded Multi-scale Dense Feature Extraction Modules (MDFEM). Fig. 2 shows the architecture of MDFEM.

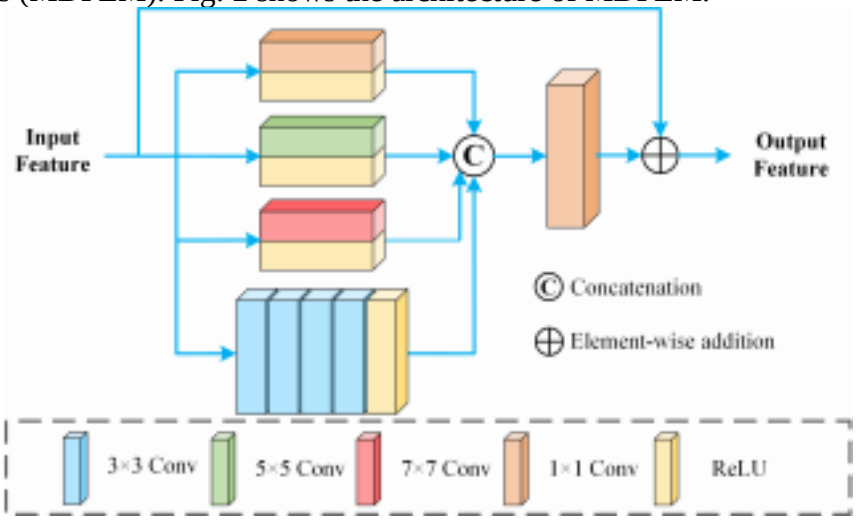

Fig. 2. Network structure of multi-scale dense feature extraction module

For the input PAN image I_p, shallow image features are extracted using the "3×3 convolution ReLU activation function to 3×3 convolution" method:

$$F_p^s = Conv(\mathrm{ReLU}(Conv(I_p, k=3), k=3) \tag{5}$$

where $Conv(\bullet)$ represents the convolution operation, k is the size of the convolution kernel, and F_p^s is the shallow features of the PAN image.

Let F_{P-FEM}^{i-1} ($F_{P-FEM}^{0} = F_p^s$) be the input of the i-th (i=1, 2, 3... L) MDFEM, and F_{P-FEM}^{i} be the output. F_{P-FEM}^{i-1} extract multi-scale features through 1×1, 5×5, and 7×7 convolutions, output features through ReLU activation function are:

$$F_{convk}^{i} = \mathrm{ReLU}(Conv(F_{P-FEM}^{i-1}, k)), \quad k = 1, 5, 7 \tag{6}$$

For dense spatial features, we extract them through four consecutive 3×3 convolutions and output the features using the ReLU activation function:

$$F_{conv3}^{i} = \mathrm{ReLU}(Conv^4(F_{P-FEM}^{i-1}, k=3)) \tag{7}$$

where $Conv^4(\bullet)$ represents four consecutive convolutional layers. In order to aggregate multi-scale features and dense features, we concatenate multiple sets of features together using feature concatenation, perform dimensionality reduction through 1×1 convolution, and finally add them element by element to the input features using skip connections to obtain the output feature of the i-th MDFEM:

$$F_{P-FEM}^{i} = Conv([F_{conv1}^{i}, F_{conv5}^{i} F_{conv7}^{i} F_{conv3}^{i}], k=1) + F_{P-FEM}^{i-1} \tag{8}$$

where [•] represents the splicing operation.

3.3. Dynamic weight generation

In this study, we designed a method for applying dynamic convolution to feature enhancement of cross modal images. Firstly, a set of weights $w_0 \in R^{K \times C_{in} \times C_{out} \times k^2}$ is initialized, where K is the number of control weights, C_{in}, C_{out}, and k represent the number of input channels, the number of output channels, and the size of the convolution kernel, respectively. For extracting complete PAN image features F_{P-FEM}^{L}, we convert it into a guiding weight $w_P \in R^{1 \times K}$, which includes an average pooling layer followed by a fully connected layer featuring two non-shared parameters, a ReLU activation function, and softmax. The formula is as follows:

$$w_p = \text{softmax}(F(\text{ReLU}(FC(Avepool(F_{P-FEM}^{L}))))) \tag{9}$$

where $FC(\bullet)$ represents a fully connected layer.

Perform matrix multiplication operation between w_p and w_0, and perform Reshape operation to obtain the required dynamic convolution kernel weights $w_d \in R^{K \times C_{in} \times C_{out} \times k^2}$:

$$w_d = \text{Reshape}(w_p \times w_0) \tag{10}$$

3.4. Feature extraction of LRMS image

The goal of extracting features from the LRMS image is to isolate its spectral characteristics, while also leveraging the convolutional kernel weights provided by the DWGM to dynamically guide the feature extraction process. This guidance is crucial for minimizing the interference caused by blurred spatial information. The sub-network is structured around two core components: a shallow feature extraction layer and a series of L cascaded Dynamic Convolution-Guided Feature Extraction Modules (DCGFEM). Fig. 3 shows the pipeline of the DCGFEM.

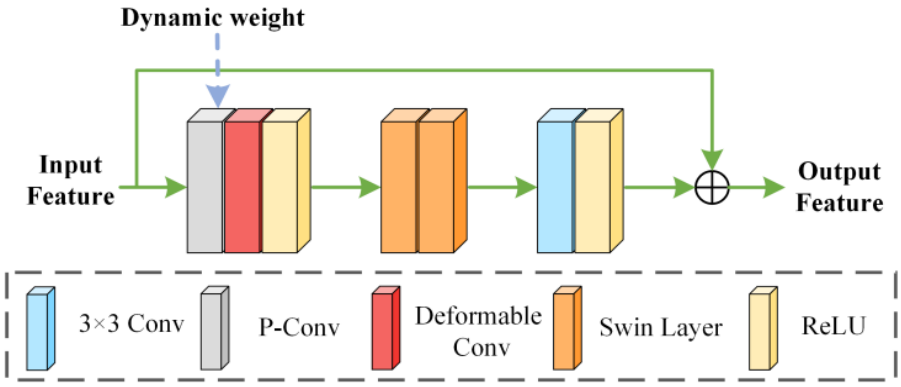

Fig. 3. The pipeline of dynamic convolution guided feature extraction module

The method of extracting shallow features is the same as that of the PAN image feature extraction. In the DCGFEM, a combination of CNN and Transformer is used for feature extraction. Specifically, for the input LRMS image, we firstly extract the shallow image features. Let F_M^S be the shallow features of LRMS images, the detail definition can be described as follows:

$$F_M^S = Conv(\mathrm{ReLU}(Conv(I_M \uparrow_4, k=3)), k=3) \tag{12}$$

where $\uparrow_4$ represents a upscaling process with a scaling ratio of 4. In DCGFEM, we define the dynamic convolution operation of applying the dynamic weight w_d output by DWGM as "P-Conv". In order to better extract local features of the image guided by "P-Conv", we introduce a deformable convolution operation after "P-Conv". Deformable convolution can achieve more precise modeling of local regions of the target by introducing learnable deformation parameters into the convolution kernel, allowing the network to learn the features of the target with greater precision. Let $F_{M-FEM-1}^i$ be the intermediate feature of the i-th DCGFEM, then the detail process is:

$$F_{M-FEM-1}^i = \mathrm{ReLU}(DeConv(P-Conv(F_M^S, k=3), k=3)) \tag{13}$$

where $DeConv(\bullet)$ represents deformable convolution operation. To extract global spectral information, we adopted Swin Transformer [18].

Finally, the features of the i-th DCGFEM, denoted as F_{M-FEM}^i, are obtained through a 3×3 convolution and ReLU activation function. The detail process is:

$$F_{M-FEM}^i = \mathrm{ReLU}(Conv(Swin^2(F_{M-FEM-1}^i), k=3)) \tag{14}$$

where $\mathrm{Swin}^2(\bullet)$ represents two cascaded Swin layers, and each Swin layer contains two cascaded Swin Transformers.

Fig. 4 shows the structure of a single Swin Transformer. The network architecture primarily consists of normalization layers, multi-head self-attention mechanisms, and multi-layer perceptrons.

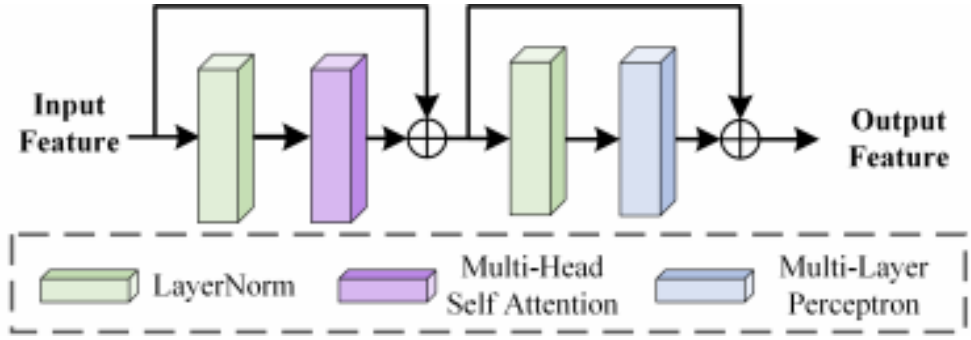

Fig. 4. The structure of a single Swin Transformer

3.5. Cross guidance fusion

In order to achieve effective fusion of PAN and LRMS image features, we designed CGFM. Fig. 5 shows the network structure of CGFM. This module mainly enhances two types of image features through cross attention.

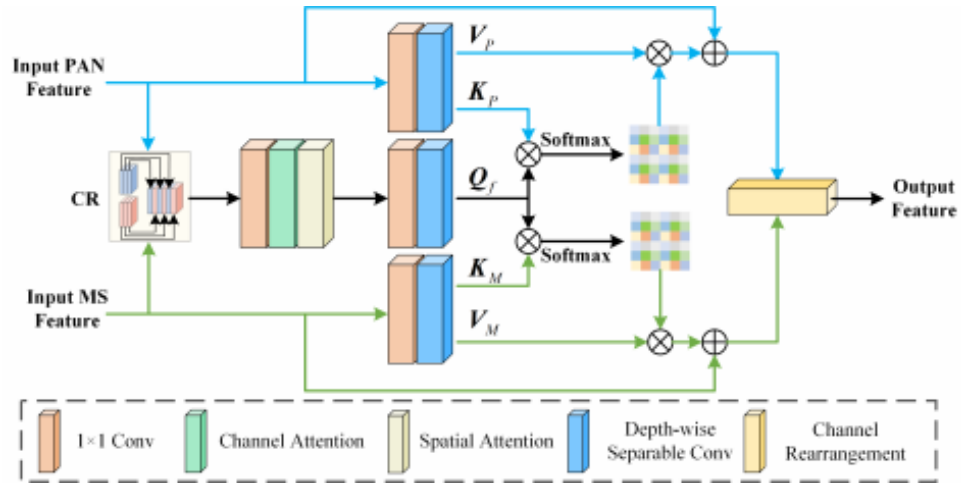

Fig. 5. Network structure of cross-guided fusion module

Specifically, for PAN image features F^{L}_{P-FEM} and LRMS image features F^{L}_{M-FEM}, we first perform a channel rearrangement (CR) operation to cross arrange and combine the two image features in the channel dimension, generating a preliminary fusion feature F^{1}_{fuse}, which contains comprehensive information from

both modalities. Meanwhile, we aggregate the feature information of F_{fuse}^{1} in the channel dimension through a 1 × 1 convolution to enhance the comprehensive features of different modal features, and then highlight the spectral and the spatial information in F_{fuse}^{1} through cascaded channel attention [19] and spatial attention [20] operations. The detail process can be described as follows:

$$F_{fuse}^{2} = SA(CA(Conv(F_{fuse}^{1}, k=1))) \tag{15}$$

where $SA(\bullet)$ and $CA(\bullet)$ represent spatial and channel attention operation, respectively. The purpose of this operation is to obtain Q, K, and V related to attention operations. The formulas are expressed as follows:

$$K_P, V_P = DWConv(Conv(F_{P-FEM}^{L}, k=1), k=3) \tag{16}$$

$$K_M, V_M = DWConv(Conv(F_{M-FEM}^{L}, k=1), k=3) \tag{17}$$

$$Q_f = DWConv(Conv(F_{fuse}^{2}, k=1), k=3) \tag{18}$$

where $DWConv(\bullet)$ represents depthwise separable convolution operations, the purpose of which is to use the generated query feature Q_f containing comprehensive information to perform cross attention operations with [K_P, V_P], [K_M, V_M]. The formulas are expressed as follows:

$$F_{P-fuse} = \text{softmax}\{\frac{Q_{f'}(K_p)^T}{t}\}V_p + F_{P-FEM}^{L} \tag{19}$$

$$F_{M-fuse} = \text{softmax}\{\frac{Q_{f'}(K_M)^T}{t}\}V_M + F_{M-FEM}^{L} \tag{20}$$

where $(\bullet)^T$ represents the matrix transpose operation, and t is the control coefficient. F_{P-fuse} and F_{M-fuse} represent the enhanced features of PAN and LRMS images. Finally, generate the output features of the module through another CR operation. The detail process can be described as follows:

$$F_{Fin} = CR(F_{P-fuse}, F_{M-fuse}) \tag{21}$$

3.6. Reconstruction module

The reconstruction module mainly achieves the reconstruction of HRMS images through 5 cascaded residual blocks and a 1 × 1 convolution, as shown in Fig. 6.

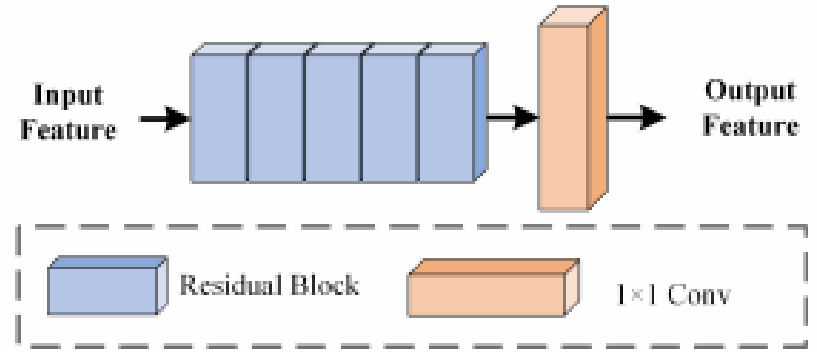

Fig. 6. The network structure of reconstruction module

The detail process can be described as follows:

$$I_H = Conv(RBs^5(F_{Fin}), k = 1) \tag{22}$$

where $RBs^5(\bullet)$ represents five cascaded residual blocks.

4. Experiments

In this section, we initially outline the experimental setup, delineate the assessment metrics, and identify the comparative methods. Subsequently, we showcase the comparative outcomes against extant techniques, utilizing various datasets.

4.1. Experimental setup

In this study, we use the WV-3 and GF-2 from PanCollection as the experimental datasets. Fig. 7 shows test samples from the WV-3 and GF-2 datasets.

To substantiate the preeminence of our methodology, we juxtaposed it against conventional approaches alongside several deep learning-oriented techniques. The conventional methods encompass BT-H, BDSD-PC, MTF-GLP-HPM-R, MTF-GLP-FS, and TV. Meanwhile, the deep learning methods involve PNN, PanNet, MSDCNN, DiCNN, BDPN, FusionNet, LAGConv, and S2DBPN (Spatial Dual Back Projection Network). All traditional methods were tested and validated on the Matlab based ToolBox proposed by Vivone et al [21].

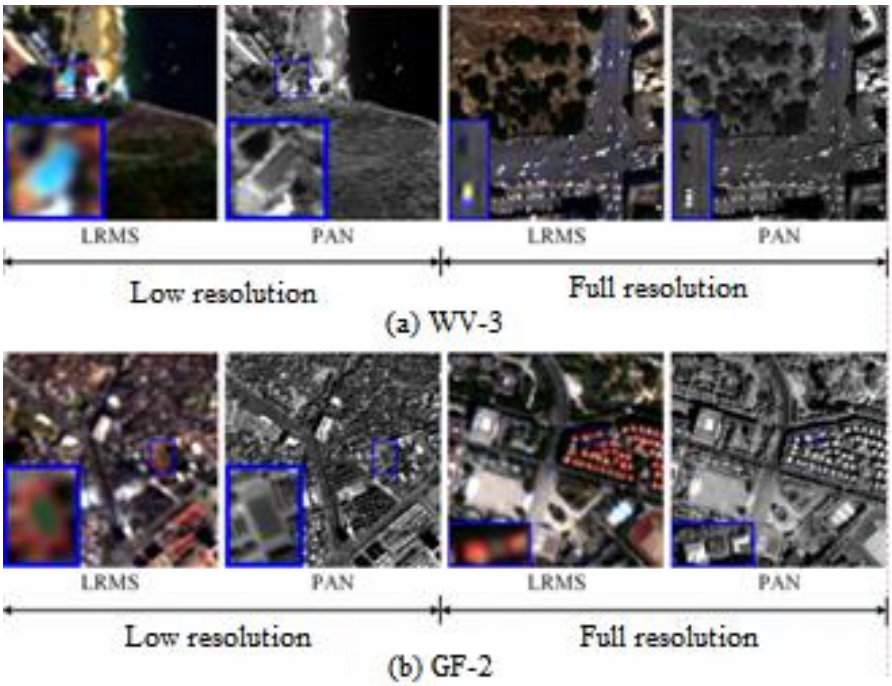

Fig. 7. Test samples from different datasets

All deep learning based methods followed the original author's parameter settings and were retrained and tested on the WV-3 and GF-2 datasets. In addition, we use QNR, D_s, and D_λ as the evaluation criteria for full resolution scale image data.

The experimental method was completed within a unified code framework based on Pytorch. Throughout the training phase, the Adam optimizer was utilized to adjust the network parameters, with the learning rate fixed at 0.0002 and a batch size established at 16. The complete network model underwent training on a GeForce RTX 3090 GPU for a total of 150 epochs.

4.2. Comparative experimental results

Fig. 8 shows the outcomes of various methodologies applied to the WV-3 full resolution test dataset.

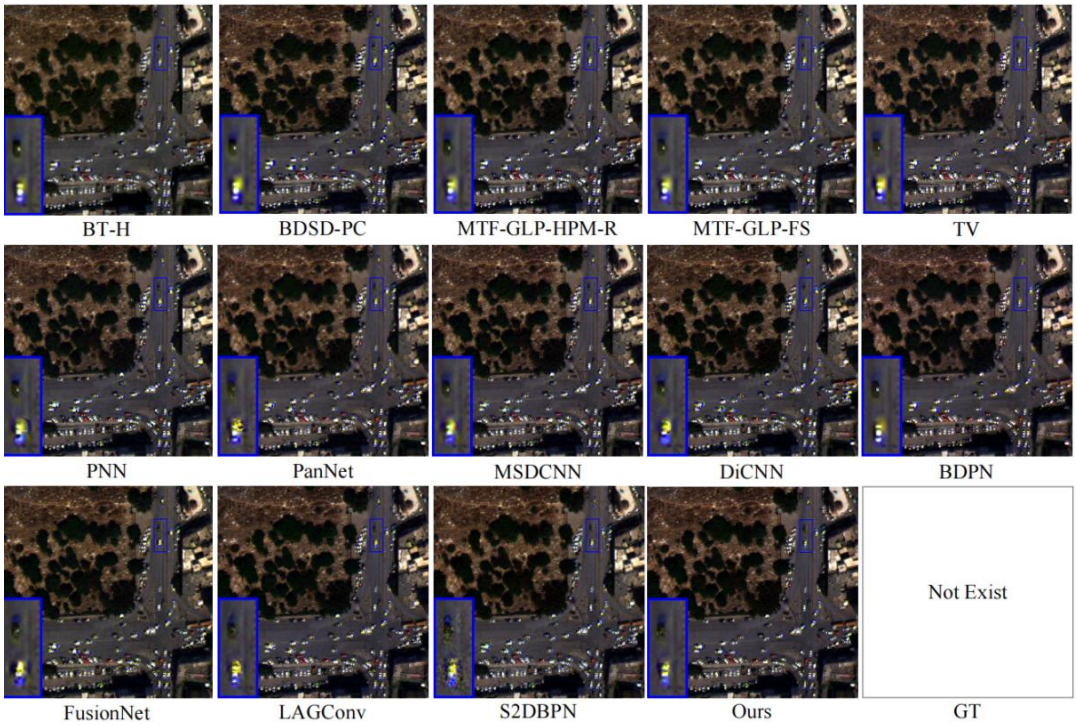

Fig. 8. The outcomes of various methodologies applied to the WV-3 full resolution test dataset

Table 1 shows the performance metrics of various methods evaluated on the WV-3 test set.

Table 1

Performance metrics of various methods evaluated on the WV-3 test set

Methods	QNR	D_s	D_λ
BT-H[6]	0.7568	0.1214	0.1493
BDSD-PC[7]	0.7901	0.0862	0.1464
MTF-GLP-HPM-R[10]	0.7870	0.1006	0.1370
MTF-GLP-FS[11]	0.7833	0.1031	0.1390
TV[14]	0.7992	0.0943	0.1323
PNN[22]	0.8650	<u>0.0840</u>	0.0582
PanNet[23]	0.8332	0.0852	0.0979
MSDCNN[24]	0.8509	**0.0838**	0.0762
DiCNN[25]	0.8437	0.0960	0.0714
BDPN[26]	0.7845	0.1067	0.1280
FusionNet[27]	0.8514	0.0984	0.0594
LAGConv[28]	**0.8576**	0.0930	0.0595
S2DBPN[29]	0.8502	0.0948	0.0676
Ours	<u>0.8565</u>	0.0953	**0.0567**

Note: Bold values in each column of the table indicate optimal values, while underlined values indicate suboptimal values.

From the analysis of Fig. 8 and Table 1, it becomes evident that our approach surpasses all competing methods numerically on the low-resolution test set, showcasing commendable performance. On the full-resolution test set, our method ranks second-best and best in terms of the QNR and D_s metrics, respectively. When contrasted with alternative methods, ours demonstrates superior overall performance, signifying that the technique presented herein yields outstanding results in panchromatic sharpening tasks across varying scales.

Fig. 9 shows the outcomes of various methods tested on the GF-2 full resolution dataset.

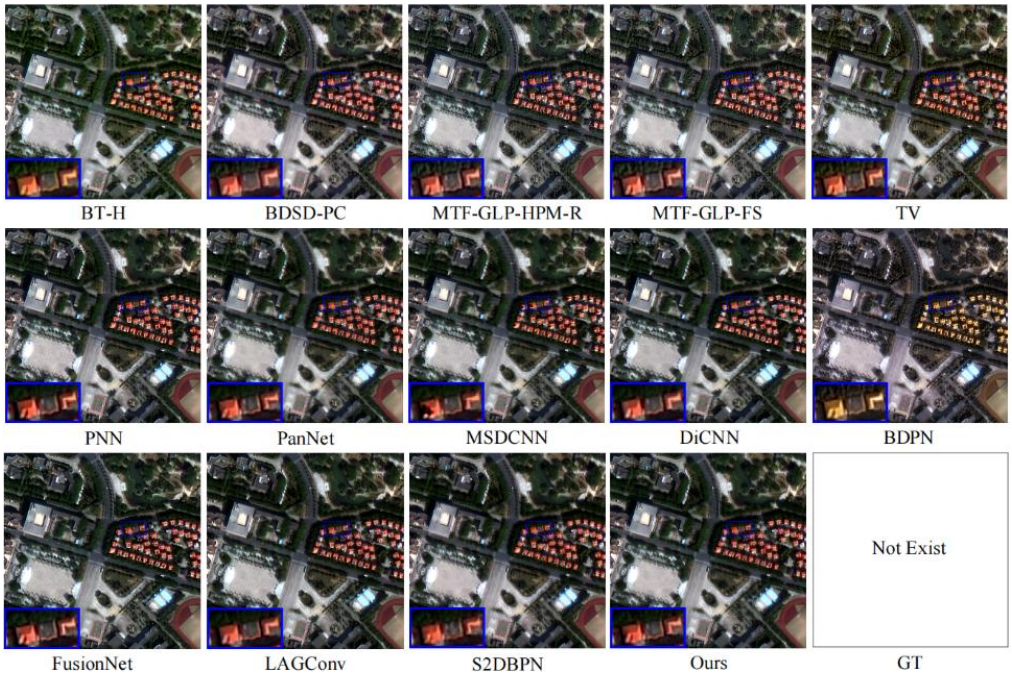

Fig. 9. The outcomes of various methods tested on the GF-2 full resolution dataset

As depicted in Fig. 9, BT-H fails to align with the original image regarding spectral information. Meanwhile, BDSD-PC, MTF-GLP-HPM-R, MTF-GLP-FS, and TV have varying degrees of loss in preserving spectral information and are also not ideal in preserving spatial details. In deep learning based methods, MSDCNN generated incorrect color blocks, BDPN showed severe spectral distortion, and S2DBPN exhibited color confusion in the overall visual effect. Compared to other methods, our approach shows the best visual performance.

Table 2 shows the quantitative comparison results on the GF-2 dataset.

Table 2

Quantitative comparison results on the GF-2 dataset

Methods	QNR	D_s	D_λ
BT-H[6]	0.7568	0.1214	0.1493
BDSD-PC[7]	0.7901	0.0862	0.1464
MTF-GLP-HPM-R[10]	0.7870	0.1006	0.1370
MTF-GLP-FS[11]	0.7833	0.1031	0.1390
TV[14]	0.7992	0.0943	0.1323
PNN[22]	0.8650	<u>0.0840</u>	0.0582
PanNet[23]	0.8332	0.0852	0.0979
MSDCNN[24]	0.8509	**0.0838**	0.0762
DiCNN[25]	0.8437	0.0960	0.0714
BDPN[26]	0.7845	0.1067	0.1280
FusionNet[27]	0.8514	0.0984	0.0594
LAGConv[28]	**0.8576**	0.0930	0.0595
S2DBPN[29]	0.8502	0.0948	0.0676
Ours	<u>0.8565</u>	0.0953	**0.0567**

Note: Bold values in each column of the table indicate optimal values, while underlined values indicate suboptimal values.

As evidenced in Table 2, our method attained the highest scores across all four assessment metrics for the low resolution test set. Regarding the full resolution test set, our method is slightly lower than S2DBPN in terms of QNR, D_s, and D_λ, but the difference is not significant and is at a suboptimal level. This indicates that our method has demonstrated strong competitiveness on two different scale test sets, and also validates the superiority of our proposed method.

5. Conclusions

In conclusion, our research introduces a novel approach for panchromatic image sharpening that innovatively integrates spatial features into the dynamic guidance of spectral feature extraction. The method dynamically aligns panchromatic and low-resolution counterparts, thereby guiding the feature extraction and significantly enhancing the spatial detail of the latter. Additionally, a cross-guided fusion module, powered by a cross-attention mechanism, has been developed to refine the feature fusion across different modalities, leading to a more comprehensive and higher quality fusion. The empirical evidence from our experiments confirms that our proposed method not only mitigates the adverse effects of blurred spatial information on the sharpening of low-resolution panchromatic images but also successfully produces high-resolution images with superior spatial clarity. This advancement holds promise for applications requiring the enhancement of low-resolution multispectral imagery.

Acknowledgments

This work was supported by the High End Low-carbon Intelligent Manufacturing Technology Innovation Team for Textile and Clothing.

R E F E R E N C E S

[1]. *Y. Xing, S. Yang, Z. Feng, et al.* "Dual-collaborative fusion model for multispectral and panchromatic image fusion". IEEE Transactions on Geoscience and Remote Sensing, **vol. 60**, 2020, pp. 1-15.

[2]. *Y. Xing, M. Wang, S. Yang, et al*. "Pansharpening with multiscale geometric support tensor machine". IEEE Transactions on Geoscience and Remote Sensing, **vol. 56**, no. 5, 2018, pp. 2503-2517.

[3]. *A. Ma, N. Yu, Z. Zheng, et al.* "A supervised progressive growing generative adversarial network for remote sensing image scene classification". IEEE Transactions on Geoscience and Remote Sensing, **vol. 60**, 2022, pp. 1-18.

[4]. *Y. Shen, D. Liu, F. Zhang, et al*. "Fast and accurate multi-class geospatial object detection with large-size remote sensing imagery using CNN and Truncated NMS" ISPRS Journal of Photogrammetry and Remote Sensing, **vol. 191**, 2022, pp. 235- 249.

[5]. *E.L. Bullock, C. E. Woodcock, P. Olofsson*. "Monitoring tropical forest degradation using spectral unmixing and Landsat time series analysis". Remote sensing of Environment, **vol. 238**, 2020, pp. 110968.

[6]. *S. Lolli, L. Alparone, A. Garzelli, et al*. "Haze correction for contrast-based multispectral pansharpening". IEEE Geoscience and Remote Sensing Letters, **vol. 14**, no. 12, 2017, pp. 2255-2259.

[7]. *G. Vivone*. "Robust band-dependent spatial-detail approaches for panchromatic sharpening". IEEE transactions on Geoscience and Remote Sensing, **vol. 57**, no. 9, 2019, pp. 6421-6433.

[8]. *G. Scarpa, S. Vitale, D. Cozzolino*. "Target-adaptive CNN-based pansharpening". IEEE Transactions on Geoscience and Remote Sensing, **vol. 56**, no. 9, 2018, pp. 5443-5457.

[9]. *K. Zhang, A. Wang, F. Zhang, et al*. "Spatial and spectral extraction network with adaptive feature fusion for pansharpening". IEEE Transactions on Geoscience and Remote Sensing, **vol. 60**, 2022, pp. 1-14.

[10]. *G. Vivone, R. Restaino, J. Chanussot*. "A regression-based high-pass modulation pansharpening approach". IEEE Transactions on Geoscience and Remote Sensing, **vol. 56**, no. 2, 2017, pp. 984-996.

[11]. *G. Vivone, R. Restaino, J. Chanussot*. "Full scale regression-based injection coefficients for panchromatic sharpening". IEEE Transactions on Image Processing, **vol. 27**, no. 7, 2018, pp. 3418-3431.

[12]. *M. Gong, J. Ma, H. Xu, et al*. "D2TNet: A ConvLSTM network with dual-direction transfer for pan-sharpening". IEEE Transactions on Geoscience and Remote Sensing, **vol. 60**, 2022, pp. 1-14.

[13]. *F. Fang, F. Li, C. Shen, et al*. "A variational approach for pan-sharpening". IEEE Transactions on Image Processing, **vol. 22**, no. 7, 2013, pp. 2822-2834.

[14]. *F. Palsson, J. R. Sveinsson, M. O. Ulfarsson*. "A new pansharpening algorithm based on total variation". IEEE Geoscience and Remote Sensing Letters, **vol. 11**, no. 1, 2013, pp. 318-322.

[15]. *P. Liu, L. Xiao, J. Zhang, et al*. "Spatial-Hessian-feature-guided variational model for pan-sharpening". IEEE Transactions on Geoscience and Remote Sensing, **vol. 54**, no. 4, 2015, pp. 2235-2253.

[16]. *X. He, K. Yan, J. Zhang, et al*. "Multi-Scale Dual-Domain Guidance Network for Pan-sharpening". IEEE Transactions on Geoscience and Remote Sensing, vol. 61, 2023, pp.1-13.
[17]. *L. Tang, J. Yuan, J. Ma*. "Image fusion in the loop of high-level vision tasks: A semantic-aware real-time infrared and visible image fusion network". Information Fusion, **vol. 82**, 2022, pp. 28-42.
[18]. *Z. Liu, Y. Lin, Y. Cao, et al*. "Swin transformer: Hierarchical vision transformer using shifted windows". in Proceedings of the IEEE/CVF international conference on computer vision. 2021, pp. 10012-10022.
[19]. *J. Hu, L. Shen, G. Sun*. "Squeeze-and-excitation networks". in Proceedings of the IEEE conference on computer vision and pattern recognition. 2018, pp. 7132-7141.
[20]. *M. Jaderberg, K. Simonyan, A. Zisserman*. "Spatial transformer networks". Advances in neural information processing systems, 2015, pp. 28.
[21]. *G. Vivone, M. Dalla Mura, A. Garzelli, et al.* "A benchmarking protocol for pansharpening: Dataset, preprocessing, and quality assessment". IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, **vol. 14**, 2021, pp. 6102-6118.
[22]. *G. Masi, D. Cozzolino, L. Verdoliva, et al*. "Pansharpening by convolutional neural networks". Remote Sensing, **vol. 8**, no. 7, 2016, pp. 594.
[23]. *J. Yang, X. Fu, Y. Hu, et al*. "PanNet: A deep network architecture for pan- sharpening". in Proceedings of the IEEE international conference on computer vision. 2017: 5449-5457.
[24]. *Q. Yuan, Y. Wei, X. Meng, et al*. "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, **vol. 11**, no. 3, 2018, pp. 978-989.
[25]. *L. He, Y. Rao, J. Li, et al*. "Pansharpening via detail injection based convolutional neural networks". IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, **vol. 12**, no. 4, 2019, pp. 1188-1204.
[26]. *Y. Zhang, C. Liu, M. Sun, et al*. "Pan-sharpening using an efficient bidirectional pyramid network". IEEE Transactions on Geoscience and Remote Sensing, **vol. 57**, no. 8, 2019, pp. 5549-5563.
[27]. *L. Deng, G. Vivone, C. Jin, et al.* "Detail injection-based deep convolutional neural networks for pansharpening". IEEE Transactions on Geoscience and Remote Sensing, **vol. 59**, no. 8, 2020, pp. 6995-7010.
[28]. *Z. Jin, T. Zhang, T. Jiang T X, et al*. "LAGConv: Local-context adaptive convolution kernels with global harmonic bias for pansharpening". in Proceedings of the AAAI Conference on Artificial Intelligence. **vol. 36**, no, 1, 2022, pp. 1113-1121.
[29]. *K. Zhang, A. Wang, F. Zhang, et al*. "Spatial-spectral dual back-projection network for pansharpening". IEEE Transactions on Geoscience and Remote Sensing, **vol. 61**, 2023, pp. 1-16.