IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 8 July 2023, accepted 2 August 2023, date of publication 7 August 2023, date of current version 11 August 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3302324

RESEARCH ARTICLE

Attention Multilayer Perceptron Fusion Network

YUHUA WANG, YUHAO LIAN, MINGJUN YING, AND SHUYU WANG
School of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China
Corresponding author: Yuhua Wang (curley0619@gmail.com)

ABSTRACT With the rise of downstream image tasks, the requirements for the quality of images obtained upstream are becoming higher and higher. In view of the many structural features of remote sensing images, we propose a novel deep neural network architecture for hyperspectral image fusion that integrates attention mechanisms and multi-layer perceptron blocks. The proposed network can capture long-range spatial dependencies between image elements, which is critical for capturing multi-scale features in remote sensing applications. The attention mechanisms selectively focus on important image features while disregarding redundant information, and the multi-layer perceptron blocks can capture multi-scale features by processing image features at different scales. The experimental results demonstrate that the proposed network outperforms other state-of-the-art methods in terms of both objective evaluation metrics and visual quality. The proposed method achieves higher Peak Signal to Noise Ratio and Spatial Consistency and Contrast values compared to other methods while preserving fine details and textures in the fused images. Overall, the proposed network provides an effective and efficient solution for hyperspectral image fusion that can contribute to the development of more accurate and reliable remote sensing applications.

INDEX TERMS Image fusion, pansharpening, multilayer perceptron, attention mechanism.

I. INTRODUCTION

Pansharpening is a data fusion technique that involves merging panchromatic (PAN) and multispectral (MS) images captured simultaneously over the same geographical area [1]. Its primary purpose is to merge the spatial detail from the PAN image (which is not present in the MS image) with the spectral detail from the MS image to produce a high-resolution multispectral image. This process is currently the best way to obtain high-resolution images because it is not feasible to achieve this through a single sensor due to physical limitations. Commercial satellites like IKONOS and GeoEye are capable of capturing both PAN and MS images [2]. Pansharpening is also utilized as a pre-processing step in various remote sensing applications, including change detection, target recognition, and scene interpretation, to improve the images and yield better results.

Pansharpening encompasses two primary approaches [3]: component substitution (CS) and multiresolution analysis(MRA). CS techniques involve using the PAN image to substitute or replace components and include methods like intensity-hue-saturation (IHS) [4], principal component analysis (PCA) [5], and Gram-Schmidt (GS) [6] spectral sharpening. On the other hand, MRA techniques involve enhancing spatial details by resampling the MS bands using a multi-resolution decomposition of the PAN image. Different forms of MRA can be utilized to extract spatial details, including discrete wavelet transform (DWT) [7], undecimated wavelet transform (UDWT) [8], adaptive wavelet transform (ATWT) [9], and Laplacian pyramid (LP) [10].

The increasing prevalence of deep neural networks in image processing has prompted researchers to investigate their potential for solving the image fusion problem. In recent years, various methods have been proposed, including PNN [11], which is based on convolutional neural networks and adopts the architecture previously proposed for image super-resolution. Another notable approach is PanNet [25], which emphasizes the preservation of both spatial and spectral information. These methods represent a departure from traditional techniques for image fusion and offer the potential for significant improvements in image quality and accuracy.

The associate editor coordinating the review of this manuscript and approving it for publication was Gangyi Jiang.

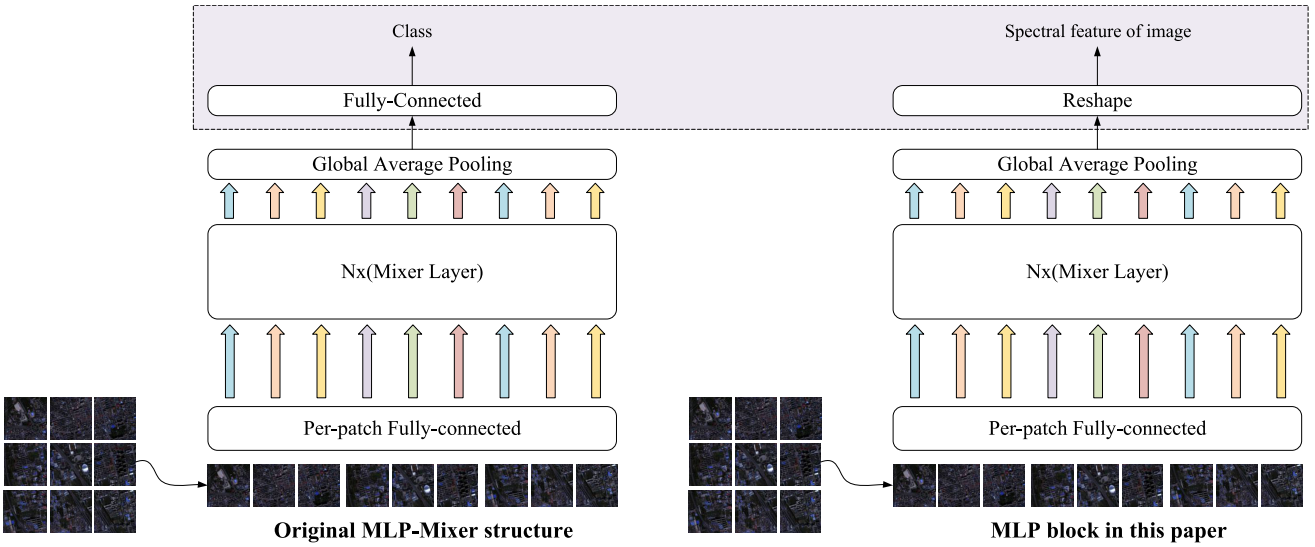

FIGURE 1. Details of the structure of MLP block.

The methods mentioned earlier treat pansharpening as a straightforward image regression problem and do not take into account the deep features of the images [13]. In contrast, this paper proposes a novel approach that incorporates deep learning techniques to enhance pansharpening. The main contributions and significance of this paper are summarized as follows:

- This paper proposes a panchromatic sharpening method that integrates distal pixel feature extraction into the deep learning framework. Pixels that are not in the 3×3 receptive field are defined here as distant pixels. This receptive field is chosen because traditional convolution usually uses 3×3 as the convolution kernel. Unlike previous methods, this approach fully considers the unique characteristics of remote sensing images and leverages a fully connected layer to extract the relationship between distant modules. By incorporating deep learning techniques and taking into account the spatial relationships between pixels, this method is able to achieve superior results compared to traditional approaches.
- Furthermore, in order to mitigate redundancy in information caused by the fully connected layer, the present study incorporates the fully connected and attention blocks to effectively extract spectral information from images.

The paper follows the outlined structure: The introductory section presents an extensive review of previous studies on remote sensing image fusion. Subsequently, attention is directed towards the examination of fully connected networks and attention mechanisms in relation to the proposed network architecture. The third section outlines the problem formulation, accompanied by a comprehensive exposition of the Attention Multilayer Perceptron Fusion Network (AMPFN) architecture proposed in this study. In the fourth section, the experimental evaluation of the AMPFN model is discussed, encompassing comparative analyses with existing methodologies and a meticulous examination of its constituent elements. Finally, the concluding section summarizes the key findings of the proposed AMPFN model, which have significantly advanced image fusion accuracy, quantitative analysis indices, and visual assessments. Moreover, this section puts forth potential avenues for future research.

II. RELATED WORK

In recent years, numerous techniques for pansharpening have been introduced, with the most popular being component replacement techniques such as IHS [4], PCA [18], and Brovey [19]. While these methods are fast and provide immediate results, they only approximate the spatial resolution of high-resolution multispectral (HRMS) images contained in PAN images and introduce spectral distortion. To address this problem, more advanced techniques, such as adaptive and dependency-based methods [20], have been developed. The multi-resolution method decomposes PAN images and low-resolution multispectral (LRMS) images, usually using wavelets [23] or Laplacian pyramids [24], and then combines them. On the other hand, model-based approaches encode beliefs about the relationship between PAN, HRMS, and LRMS images into regularized objective functions, treating the fusion problem as an image restoration optimization problem [14]. Various algorithms based on these methods have demonstrated outstanding performance.

A. MLP-MIXER

The MLP Block, which is introduced in this article, is mainly inspired by the MLP-Mixer [15] computer vision framework that was recently developed by the Google ViT team. The framework substitutes the traditional convolutional operations found in conventional CNNs and the self-attention

mechanism in Transformers with multilayer perceptron (MLP) operations.The specific structure is shown in Fig.1.

The MLP-Mixer framework consists of three main components: Per-patch Fully-connected, Mixer Layer, and Classifier. The Classifier component [16] uses the standard Global Average Pooling (GAP) method followed by a Fully-connected layer (FC) [17] and Softmax function. However, unlike the convolutional layer, the Fully-connected layer cannot capture information between local regions. To overcome this limitation, MLP-Mixer converts the input image into a 2D table via Per-patch Fully-connected, which enables the fusion of information between local regions in subsequent stages.

The MLP block in this paper is slightly modified on the basis of MLP-mixer, and the fully connected layer used for classification is replaced by the Reshape layer, so as to achieve the purpose of extracting multispectral image features. This paper draws on this idea of replacing traditional convolutions with fully connected blocks and proposes the AMPFN network architecture, which will be described in detail in Section III later.

III. METHODOLOGY

A. LOSS FUNCTION

This section mainly describes the loss function. Let a low-resolution multispectral or noisy input image $M \in \mathbb{R}^{C,w,h}$, and a high-resolution panchromatic image $P \in \mathbb{R}^{c,W,H}$, where c, W, H represent the number of channels, the image width and the height of the image, respectively. In the panchromatic sharpening problem, $w \ll W$, $h \ll H$, and $c \ll C$. $X \in \mathbb{R}^{C,W,H}$ is the output after passing through the network. In this paper, we use the following loss function to solve the optimization problem:

$$Loss = L(M, X, P) \tag{1}$$

The variable L represents a loss function for each task. In this formula, the optimization problem incorporates both multispectral image $\mathbf{M}$ and panchromatic image $\mathbf{P}$. This structure efficiently integrates multiscale spatial details and semantic features.

Equation (1) mainly describes the loss function form of each task in this paper, without specific calculation. For example, the loss functions for spectral preservation and spatial information preservation may differ in detail, but Equation(1) is the main structure in terms of structure.

According to the main structure of Equation(1), Equation(2) is proposed for the evaluation of spectral information preservation and spatial information preservation.

$$L = \|\theta_W(P, \uparrow M) + \uparrow M - X\|_F^2 \tag{2}$$

1) SPECTRAL PRESERVATION

In order to better fuse the spectral information, $\mathbf{M}$ is upsampled and skip connections [22] are added to achieve the purpose. We feed $\uparrow\mathbf{M}$ to the network to learn how spatial information in PAN maps to different spectral bands in $\mathbf{X}$. In (2), $\uparrow\mathbf{M}$ represents the MS image after up-sampling, and θ_W represents the transformation through the MLP block, which will be discussed in a later section.

2) STRUCTURAL PRESERVATION

In this paper, PAN images and upsampled LRMS images are fed into MLP block θ_W. In (2), we observe that since $\mathbf{M}$ is low-resolution, which can be viewed as a low-pass spectral content-like network containing $\mathbf{X}$, θ_W can then learn a mapping to fuse the spatial information contained in PAN into $\mathbf{X}$.

B. EVALUATION INDEX

In order to quantitatively validate the results, this study employed both full-reference and no-reference performance metrics. As shown in Table 1, all indicators covered in this article are listed together with references.

1) FULL-REFERENCE

Reference-based standards included Peak Signal-to-Noise Ratio (PSNR), Spectral Angle Mapper (SAM), Spatial Correlation Coefficient (SCC), and the relative dimensionless global error in synthesis (ERGAS).

$$MSE = E[(I - J)]^2 \tag{3}$$

$$RMSE(I, J) = \sqrt{MSE} \tag{4}$$

Here, I represents the pixel vector of the n-band unfused MS image, J represents the predicted vector of n band HRMS. Here, we give priority to spatial/radiation distortion, and in this paper, mean square error (MSE) and root mean square error (RMSE) are introduced to facilitate the definition of subsequent evaluation indexes. The ideal value of both indices is 0, which is realized if and only if $I = J$.

$$PSNR = 10 \cdot log_{10}(\frac{MAX^2}{MSE}) \tag{5}$$

Equation (5) is Peak Signal to Noise ratio, used to measure the ratio between the effective information and noise of the image, can reflect whether the image is distorted or not. Here, MAX is the difference between the maximum and minimum gray values of the ideal reference image (typically 255). The larger the value of PSNR, the better the quality of the fused image.

$$SAM(I_i, J_i) = arccos(\frac{< I_i, J_i >}{\|I_i\|\|J_i\|}) \tag{6}$$

Here, $< \cdot, \cdot >$ denotes the inner product, and $\| \cdot \|$ denotes the magnitude. The SAM global value for the entire image is calculated by taking the average of individual measurements over all pixels. The optimal value for SAM is 0, indicating perfect spectral alignment between the fused image and the reference image.

$$ERGAS = \frac{100}{R}\sqrt{\frac{1}{N}\sum_{i=1}^{n}(\frac{RMSE(I_i, J_i)}{\mu(I_i)})^2} \tag{7}$$

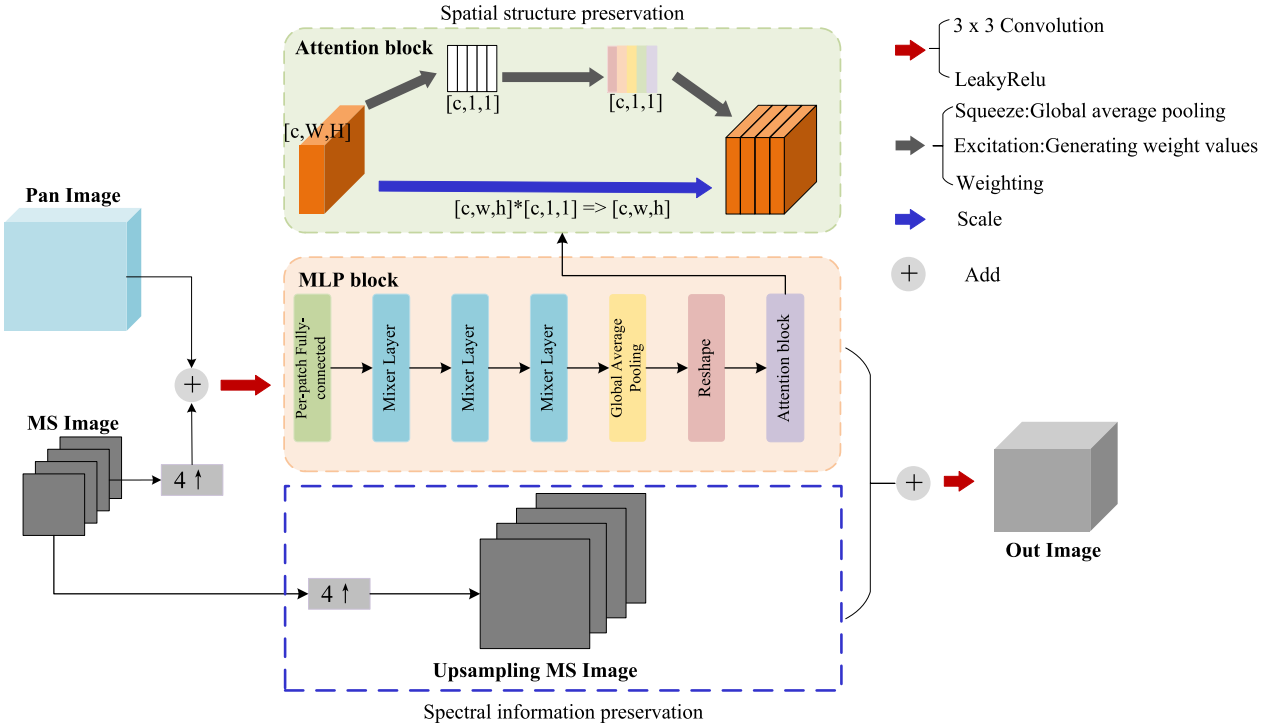

FIGURE 2. Details of the structure of AMPFN.

TABLE 1. Full-reference and no-reference performance metrics.

Full Reference	**PSNR**	Peak Signal-to-Noise Ratio [27]
	SAM	Spectral Angle Mapper [28]
	SCC	Spatial Correlation Coefficient [29]
	ERGAS	*Erreur Relative Globale Adimensionnelle de Synthèse* [30]
No Reference	$\mathbf{D}_\lambda$	Spectral component of QNR
	$\mathbf{D}_s$	Spatial component of QNR
	QNR	Quality with no-Reference index [13]

Equation (7) is *Erreur Relative Globale Adimensionnelle de Synthèse*(ERGAS). This is a more plausible global index, where RMSE is defined in (4), μ denotes the average of the image (average over pixels), and since ERGAS is composed of the sum of RMSE values, its optimal value is 0.

2) NO-REFERENCE

In the context of full-color sharpening, there exist various criteria for assessing the method's performance on the original spatial resolution image in the absence of reference data. These standards include the spectral quality index (D_λ), the spatial quality index (D_S), and the no-reference spectral and spatial joint quality index (QNR).

$$D_\lambda = \sqrt[p]{\frac{1}{N(N-1)}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}|d_{i,j}(MS,\widehat{MS})|^p} \tag{8}$$

where $d_{i,j}(MS,\widehat{MS}) = Q(MS_i, MS_j) - Q(\widehat{MS_i},\widehat{MS_j})$.

Equation (8) aims to generate a synthetic image with the same spectral characteristics as the original MS image. Therefore, the relationship between the mass bands must be preserved during the enhancement process. The Q-index is used to calculate the difference between band pairs, and the parameter p is usually set to one.

$$D_s = \sqrt[q]{\frac{1}{N}\sum_{i=1}^{N}|Q(\widehat{MS_i},P) - Q(MS_i,P_{LP})|^q} \tag{9}$$

Equation (9), P_{LP} is a low-resolution PAN image of the same scale as the MS image, and q is usually set to 1. From a practical point of view, perfect alignment between the interpolated MS image and the PAN image should be guaranteed to avoid

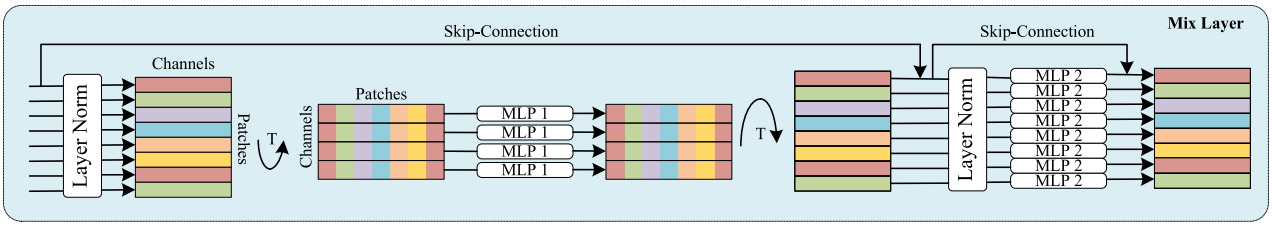

FIGURE 3. Detailed structure of the Mix Layer.

the loss of meaning of this quality metric.

$$QNR = (1 - D_\lambda)^\alpha (1 - D_s)^\beta \tag{10}$$

Equation (10) integrates (8) and (9) to quantify spectral and spatial distortions, respectively, weighted by coefficients α and β. The higher the QNR index, the better the quality of the fused product. The maximum theoretical value of this index is 1 when both D_λ and D_s are 0.

C. AMPFN ARCHITECTURE

Through the study of PanNet network architecture, this study carried out further research and construction on the basis of PanNet's success. Like other pan-sharpening methods, our deep network is designed to preserve both spectral and spatial information. Considering the potential correlation between remote blocks in remote sensing images, we introduce a multi-layer perceptron (MLP) module to extract remote feature relationships. Remote block here refers to pixel points that are not in the 3×3 receptive domain. To prevent the introduction of redundant information, the attention module is used in the MLP module to extract important information. We discuss these topics in more detail below.

The overall architecture is shown in Fig. 2, which mainly includes the network structure of two multi-layer perceptron (MLP) blocks with attention mechanism. At the same time, the attention mechanism is used on the channel dimension. Before applying the squeeze-excitation (SE [21]) attention mechanism, the feature map assigns equal importance to each channel. The SE block then assigns different weights to different channels, as shown in different colors. This results in different importance for each feature channel, allowing the neural network to prioritize channels with higher weight values.

1) MLP BLOCK

The lower part in Fig. 2 uses MLP blocks to extract the relationship features between distant pixels of the image and prevents the redundancy of information through the attention mechanism. The Mixer Layer and MLP structure used are shown in Fig. 3 and Fig. 4 respectively. The specific structure is shown in Table 2.

2) ATTENTION MECHANISM

The attention mechanism shown in Fig. 5 uses another new neural network to obtain the importance of each channel of

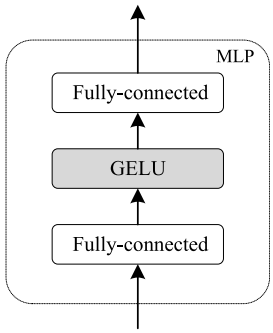

FIGURE 4. Detailed structure of the MLP.

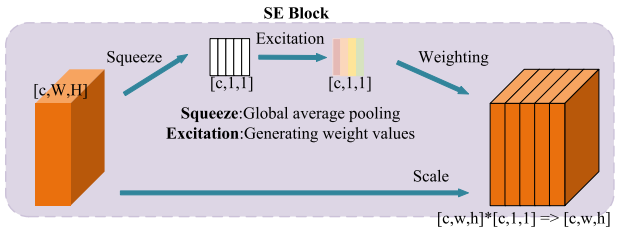

FIGURE 5. Details of the structure of attention block.

the feature map by means of automatic learning, and then assigns a weight value to each feature using this importance, so that the neural network can focus on certain feature channels. The channels of feature maps that are useful for the current task are promoted, and the feature channels that are not useful for the current task are suppressed.

Here, Squeeze compreszes the two-dimensional features $(H * W)$ of each channel to 1 real number by global average pooling, and the feature map is transformed from $[c, w, h] =>$ $[c, 1, 1]$. excitation generates a weight value for each feature channel. In this paper, two fully connected layers are used to build the correlation between channels, and the number of output weight values is the same as the number of input feature maps. Finally, Scale is used to weight the previously obtained normalized weights to the features of each channel. Multiplication is used in this paper, multiplying the weight coefficients channel by channel. $[c, w, h] * [c, 1, 1] => [c, w, h]$.

IV. EXPERIMENTS

We conducted several experiments using data acquired from the IKONOS satellite, which has a PAN resolution

TABLE 2. The detailed structure of the MLP block.

Module part	Layer (type)	Output Shape	Param
Per-Patch Fully connected processing	input7 (InputLayer)	[(None, 64, 64, 16)]	0
	projector1 (Conv2D)	(None, 16, 16, 256)	65792
	reshape4 (Reshape)	(None, 256, 256)	0
Mixer Layer	mixer block10 (MixerBlock)	(None, 256, 256)	50784
	mixer block 11 (MixerBlock)	(None, 256, 256)	50784
	mixer block 12 (MixerBlock)	(None, 256, 256)	50784
	mixer block 13 (MixerBlock)	(None, 256, 256)	50784
	mixer block 14 (MixerBlock)	(None, 256, 256)	50784
Reshape layer	flatten 2 (Flatten)	(None, 65536)	0
	pre head layer norm1 (LayerNormalization)	(None, 65536)	131072
	reshape 5 (Reshape)	(None, 64, 64, 16)	0

of 4m. To minimize the objective function in equation 1, we employed the adaptive momentum stochastic optimization method Adam. For our experiments, we extracted PAN/LRMS/HRMS patch pairs of size 64×64. We randomly divided these patches into training and validation sets, with a ratio of 90% and 10%, respectively. We compared the performance of six commonly used pan-sharpening methods, namely IHS, Brovey, GSA, MTF-GLP-HPM, PanNet, and PNN. Multiple parameter settings were tested for each method, and the optimal configuration was selected based on performance metrics.

A. DATASET

IKONOS is a high resolution, optical remote sensing commercial satellite, including land, agriculture, forestry, environmental protection and other departments to provide HD satellite images. The resolution of panchromatic images and multispectral images used in this paper is 4m and 1m, respectively. Here, four different image scenes are selected for experiments, including agricultural, urban, forest, or a mixture of these scenes. In subsequent comparison and ablation experiments, this data set was used in this paper for experimental verification.

B. IMPLEMENTATION DETAILS

Here we set validation ratio $= 0.1$, batch size $= 32$, epochs $= 50$, Adam optimizer, learning rate $= 9 \times 10^{-4}$, training image size $= 64$, and reconstruction size $= 320$. All parameters obtained are the most effective values obtained after multiple parameter tuning, and will not be described here.

C. RESULTS

1) QUANTITATIVE INDEX TEST

In Fig.6, this paper shows the PSNR changes of the proposed AMPFN and the current more efficient network during training. We can see that all training methods rise rapidly at the beginning of training and tend to be stable after a period of time. The PSNR of AMPFN is obviously higher than that of the other two network structures, but it is not enough to only use PSNR to evaluate our results. In this paper, other reference and non-reference indicators mentioned in Section III-B are used to further refine the analysis of image indicators.

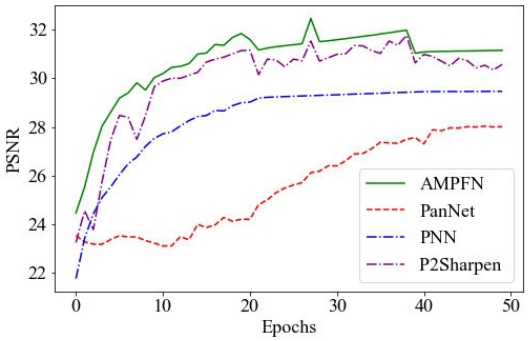

FIGURE 6. Variation of PSNR for each network structure.

In Table 3, where MGH stands for MTF-GLP-HPM. It presents the results of the test images in terms of quantitative metrics, indicating that AMPFEN outperforms existing methods across all metrics. It also shows that MTF-GLP-HPM produces the minimum value of D_λ, indicating that it provides the most similar reconstructed spectrum to the LR-MS image's spectrum. Meanwhile, all other indicator values are optimal for AMPFEN, implying its superiority in terms of spatial details over other methods, and its strong performance in spectral preservation. Moreover, AMPFEN's superiority in the QNR index over other methods suggests its ability to balance the trade-off between spectral resolution and spatial resolution effectively, resulting in a superior QNR value.

TABLE 3. Quantitative evaluation results of pansharpening in different image scenes. ↓ indicates the lower the better, ↑ indicates the higher the better, and bold indicates the optimal value.

	IHS	Brovery	GSA	MGH	PanNet	PNN	P2Sharpen	AMPFN
PSNR↑	25.4087	25.2994	26.4543	26.8807	28.4330	28.5427	30.2741	**31.2875**
SAM↓	0.1796	0.1451	0.212	0.1993	0.1473	0.1419	0.1134	**0.1106**
SCC↑	0.8831	0.8914	0.9189	0.928	0.9483	0.9554	0.9603	**0.9654**
ERGAS↓	7.8865	7.9374	6.885	6.4427	5.1673	4.8567	4.5520	**4.4947**
D_λ↓	0.1047	**0.0606**	0.1604	0.0734	0.0784	0.1063	0.0744	0.0751
D_s↓	0.1564	0.1485	0.1374	0.0889	0.0757	0.0795	0.0748	**0.0722**
QNR↑	0.7553	0.7999	0.7243	0.8442	0.8519	0.8226	0.8504	**0.8581**

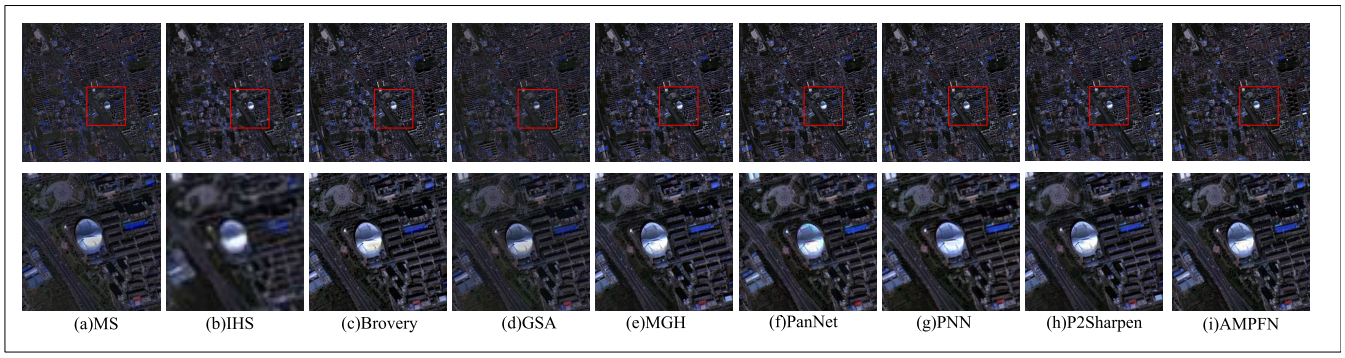

FIGURE 7. Image fusion result. First row: RGB image of the pansharpening MS image. Second row: magnified RGB image.

2) VISUAL INSPECTION

Additionally, visually, the HRMS images reconstructed by AMPFEN exhibit superior spectral and spatial information quality. Fig. 7 displays the RGB image of the fused HR-MS image, with detailed image comparisons provided after zooming in. Our analysis reveals that while PanNet, and PNN obtained clearer edges in the fused image, they also suffered from significant spectral distortion, which led to discrepancies in color and reference images. In contrast, AMPFEN retains the spectral information while also generating spatial details that are similar to the reference data. This reduces the error of the fused image.

3) COMPUTATION ANALYSIS

In order to comprehensively analyze the performance of the network proposed in this paper, we also compare the computational complexity on the three deep learning networks. As shown in Table 4, AMPFN greatly outperforms the other two network architectures in terms of computation and memory application. At this point, it can be proved that the AMPFN architecture proposed in this paper is an efficient remote sensing image network with comprehensive improvement.

D. ABLATION EXPERIMENT

In order to investigate the role of certain modules within the AMPFN framework, a series of ablation experiments were conducted in this study. The experiments involved 3 different configurations, and the results are presented in Table 5.

TABLE 4. Computation complexity of different methods.

Methods	Computation/FLOPS	Memory access/byte
PNN	658.39M	0.3068M
PanNet	629.31M	0.2961M
P2Sharpen	775.42M	0.3578M
AMPFN	**563.78M**	**0.2641M**

1) ATTENTION BLOCK

In configuration I and II, we investigated the impact of integrating SE blocks with attention mechanisms to enhance the network's performance. The results presented in Table 5 demonstrate a significant improvement in the performance of various indicators when attention blocks are included.

2) MLP BLOCK

The inclusion of MLP blocks in the network architecture improves the ability to extract long-range point relationships, which is particularly relevant for remote sensing images. In the first and second experiments, we have demonstrated that the addition of MLP blocks leads to significant improvements in various performance indicators, while the removal of

TABLE 5. The results of ablation experiments.

Configuration	MLP	Attention	PSNR↑	SCC↑	SAM↓	ERGAS↓
I	×	×	29.9282	0.9603	0.1296	4.7315
II	×	√	30.1919	0.9617	0.1194	4.7202
III	√	×	31.0280	0.9635	0.1121	4.5100
AMPFN	√	√	**31.2875**	**0.9654**	**0.1106**	**4.4947**

this module results in a deterioration of these indicators. This suggests that the MLP block plays a crucial role in capturing distant point relationships.

All in all, the addition of attention block and MLP block structure can reduce the amount of parameters while paying attention to extract the relationship between distant pixels of the image, and obtain more accurate high-resolution remote sensing images.

V. CONCLUSION

This paper proposes an attention multi-layer perceptron fusion network based on the two goals of previous deep model pansharpening: Spectral and spatial information preservation introduces a multi-layer perceptron with attention mechanism, which allows the network to focus on the relationship between distant pixels of remote sensing images. At the same time, the attention mechanism is used to deal with the redundant information that may be brought by the fully connected layer, reducing the amount of parameters used in deep learning training. Experimental results show that AMPFN achieves better image reconstruction than PNN and PanNet, and achieves excellent results in both quantitative index test and visual test.

ACKNOWLEDGMENT

The authors would like to express their deepest gratitude to their thesis advisor Shuyue Luo, for their guidance, support, and patience throughout the research process. Their insightful feedback and expertise have been invaluable in shaping this thesis. They would also like to thank their colleagues Yuhao Lian, for their helpful discussions, encouragement, and assistance in various stages of their research. Their insights and ideas have greatly contributed to the development of this thesis. Their gratitude also goes to their family and friends for their unwavering support, encouragement, and love. Their emotional support has helped them to stay motivated and focused during the difficult times of their research. Lastly, they would also like to thank the participants of their study, without whom this research would not have been possible. Their willingness to participate and share their experiences and insights have been crucial to the success of this study. Thank you all for your invaluable support and contributions.

REFERENCES

[1] Y. Wang, G. Liu, L. Wei, L. Yang, and L. Xu, "A method to improve full-resolution remote sensing pansharpening image quality assessment via feature combination," *Signal Process.*, vol. 208, Jul. 2023, Art. no. 108975.

[2] X. Sun, "Automatic extraction of digital elevation from IKONOS in-track stereo images," 2021, doi: 10.32920/ryerson.14666019.v1.

[3] H. Lu, Y. Yang, S. Huang, X. Chen, H. Su, and W. Tu, "Intensity mixture and band-adaptive detail fusion for pansharpening," *Pattern Recognit.*, vol. 139, Jul. 2023, Art. no. 109434.

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

[5] 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.

[6] C. A. Laben and B. V. Brower, "Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening," U.S. Patent 09/069 232, Aug. 8, 2023.

[7] G. P. Nason and B. W. Silverman, "The stationary wavelet transform and some statistical applications," in *Wavelets and Statistics*, vol. 103, A. Antoniadis and G. Oppenheim, Eds. New York, NY, USA: Springer-Verlag, 1995, pp. 281–299.

[8] M. J. Shensa, "The discrete wavelet transform: Wedding the a trous and Mallat algorithms," *IEEE Trans. Signal Process.*, vol. 40, no. 10, pp. 2464–2482, Oct. 1992.

[9] P. Burt and E. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Trans. Commun.*, vol. COM-31, no. 4, pp. 532–540, Apr. 1983.

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

[11] G. Masi, D. Cozzolino, L. Verdoliva, and G. Scarpa, "Pansharpening by convolutional neural networks," *Remote Sens.*, vol. 8, no. 7, p. 594, Jul. 2016, doi: 10.3390/rs8070594.

[12] H. Zhang, H. Wang, X. Tian, and J. Ma, "P2Sharpen: A progressive pansharpening network with deep spectral transformation," *Inf. Fusion*, vol. 91, pp. 103–122, Mar. 2023.

[13] L. Alparone, B. Aiazzi, S. Baronti, A. Garzelli, F. Nencini, and M. Selva, "Multispectral and panchromatic data fusion assessment without reference," *Photogramm. Eng. Remote Sens.*, vol. 74, no. 2, pp. 193–200, Feb. 2008.

[14] Z. Zhu, X. Cao, M. Zhou, J. Huang, and D. Meng, "Probability-based global cross-modal upsampling for pansharpening," 2023, *arXiv:2303.13659*.

[15] S.-H. Kong, S. Cho, and E. Kim, "GPS first path detection network based on MLP-mixers," *IEEE Trans. Wireless Commun.*, vol. 21, no. 9, pp. 7764–7777, Sep. 2022.

[16] Y. Kang, Y. Xu, X. Wang, B. Pu, X. Yang, Y. Rao, and J. Chen, "HN-PPISP: A hybrid network based on MLP-mixer for protein–protein interaction site prediction," *Briefings Bioinf.*, vol. 24, no. 1, p. bbac480, Jan. 2023.

[17] C. Jameson, B. Basyildiz, D. Moore, K. Clark, and Z. Gong, "Time optimal quantum state transfer in a fully-connected quantum computer," 2023, *arXiv:2303.04804*.

[18] V. K. Shettigara, "A generalized component substitution technique for spatial enhancement of multispectral images using a higher resolution data set," *Photogram. Eng. Remote Sens.*, vol. 58, no. 5, pp. 561–567, May 1992.

[19] A. R. Gillespie, A. B. Kahle, and R. E. Walker, "Color enhancement of highly correlated images. II. Channel ratio and 'chromaticity' transformation techniques," *Remote Sens. Environ.*, vol. 22, no. 3, pp. 343–365, Aug. 1987.

[20] Q. Dong, "Research on adaptive enhancement of robot vision image based on multi-scale filter," *Int. J. Image Data Fusion*, pp. 1–17, 2022, doi: 10.1080/19479832.2022.2149630.

[21] G. Xu, Y. Xu, S. Zhang, and X. Xiem, "SFRNet: Feature extraction-fusion steganalysis network based on squeeze-and-excitation block and RepVgg block," *Secur. Commun. Netw.*, vol. 2021, p. 111, Jul. 2021, doi: 10.1155/2021/3676720.

[22] J. Di, S. Ma, J. Lian, and G. Wang, "A U-Net network model for medical image segmentation based on improved skip connections," in *Proc. 14th Int. Conf. Measuring Technol. Mechatronics Autom. (ICMTMA)*, Jan. 2022, pp. 298–302.

[23] L. Wang, G. Liu, Q. Qin, and Q. Zhang, "Fusion algorithm of high spatial and spectral resolution images based on contourlet transform," *Proc. SPIE*, vol. 6787, Nov. 2007, Art. no. 67870N.

[24] B. Aiazzi, L. Alparone, S. Baronti, and A. Garzelli, "Context-driven fusion of high spatial and spectral resolution images based on oversampled multiresolution analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 10, pp. 2300–2312, Oct. 2002, doi: 10.1109/tgrs.2002.803623.
[25] J. Yang, X. Fu, Y. Hu, Y. Huang, X. Ding, and J. Paisley, "PanNet: A deep network architecture for pan-sharpening," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 1753–1761.
[26] S. Aggarwal and H. Agarwal, "Ridge regression for PSNR of restored images by recursive median filter," in *Proc. Int. Conf. Cogn. Intell. Comput.* Singapore: Springer, 2023, pp. 469–481.
[27] R. H. Yuhas, A. F. H. Goetz, and J. W. Boardman, "Discrimination among semi-arid landscape endmembers using the spectral AngleMapper (SAM) algorithm," in *Proc. Summaries 3rd Annu. JPL Airborne Geosci. Workshop*, Pasadena, CA, USA, 1992, pp. 147–149.
[28] J. Zhou, D. L. Civco, and J. A. Silander, "A wavelet transform method to merge Landsat TM and SPOT panchromatic data," *Int. J. Remote Sens.*, vol. 19, no. 4, pp. 743–757, Jan. 1998.
[29] L. Wald, *Data Fusion: Definitions and Architectures: Fusion of Images of Different Spatial Resolutions*. Paris, France: Presses des Mines, 2002.

YUHUA WANG is currently pursuing the degree in data science and big data with the School of Computer Science and Technology, Chongqing University of Posts and Telecommunications. Through her studies and research, she has gained expertise in statistical analysis, machine learning algorithms, and data visualization techniques. Her research work has focused on developing efficient and accurate algorithms for recognizing patterns in time series data, which have potential applications in various fields, such as finance, healthcare, and environmental monitoring. Her research contributions have been published in peer-reviewed journals and presented at international conferences, demonstrating her commitment to advancing the field of data science and artificial intelligence. Her research interests include recognition, time series data processing, and artificial intelligence.

YUHAO LIAN is currently pursuing the degree in communication engineering with the School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications. Through his research, he has proposed efficient detection algorithms for OFDM-IM systems and improved the communication efficiency of OFDM-IM by combining it with the WHT. He has strong interdisciplinary skills and has applied zero-phase filters to time-series models to improve their performance, making outstanding contributions to the fields of communication and artificial intelligence. His research interests include B5G/6G wireless communication, optical communication, machine learning, and deep learning.

MINGJUN YING received the B.E. degree in communication engineering from the Chongqing University of Posts and Telecommunications, Chongqing, China, in 2023. He is currently pursuing the Ph.D. degree in electrical engineering with the NYU WIRELESS Research Center, New York University Tandon School of Engineering, Brooklyn, NY, USA, under the supervision of Prof. Theodore S. Rappaport. His research interests include the application of consumption factor theory, mmWave and terahertz channel modeling, reconfigurable intelligent surfaces, robust resource allocation, and backscatter communication.

SHUYU WANG is currently pursuing the B.E. degree with the School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing, China. Her research interests include resource allocation, reconfigurable intelligent surfaces, cell-free communication, and image fusion.