Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 8476000, 14 pages
https://doi.org/10.1155/2022/8476000

Research Article

Remote Sensing Image Fusion Algorithm Based on Two-Stream Fusion Network and Residual Channel Attention Mechanism

Mengxing Huang, Shi Liu, Zhenfeng Li, Siling Feng, Di Wu, Yuanyuan Wu, and Feng Shu

School of Information and Communication Engineering, Hainan University, Haikou 570228, China

Correspondence should be addressed to Yuanyuan Wu; wyuanyuan82@163.com and Feng Shu; shufeng0101@163.com

Received 21 August 2021; Accepted 1 December 2021; Published 11 January 2022

Academic Editor: Chunguo Li

A two-stream remote sensing image fusion network (RCAMTFNet) based on the residual channel attention mechanism is proposed by introducing the residual channel attention mechanism (RCAM) in this paper. In the RCAMTFNet, the spatial features of PAN and the spectral features of MS are extracted, respectively, by a two-channel feature extraction layer. Multiresidual connections allow the network to adapt to a deeper network structure without the degradation. The residual channel attention mechanism is introduced to learn the interdependence between channels, and then the correlation features among channels are adapted on the basis of the dependency. In this way, image spatial information and spectral information are extracted exclusively. What is more, pansharpening images are reconstructed across the board. Experiments are conducted on two satellite datasets, GaoFen-2 and WorldView-2. The experimental results show that the proposed algorithm is superior to the algorithms to some existing literature in the comparison of the values of reference evaluation indicators and nonreference evaluation indicators.

1. Introduction

With the widespread application of GIS, the demand for spatial and geographic data in various industries is increasing [1]. However, due to various reasons such as the complexity of the scene, the time and spectral transformation of the input data set, different spatial data standards, and specific spatial object classification and classification systems, there are many differences in the accuracy and form of remote sensing data [2]. Nowadays, there are many remote sensing satellites with different functions on various observation platforms outside the earth. These satellites can provide different spatial, temporal, and spectral images, that is, remote sensing images. Because of the limitation of satellite sensor, remote sensing satellite can only obtain hyperspectral image and high spatial resolution panchromatic image, respectively. In practical applications, both high spectral resolution remote sensing images and high spatial resolution remote sensing images are required, which make remote sensing image fusion with important application value [3].

Remote sensing image fusion is an algorithm that fuses the panchromatic image with high spatial resolution (panchromatic, PAN) and the multispectral image with low spatial resolution (low-resolution multispectral, LMS) into the high-resolution multispectral (high-resolution multispectral, HRMS) images, shortened to pansharpening [4]. High-resolution multispectral images can calculate the reflection spectrum of each pixel on the earth surface to get a variety of information. It can also provide help for subsequent remote sensing scene segmentation, classification, and feature extraction such as the survey of forest resources, the ground feature classification, the precision agriculture, and the weather forecast [5]. However, due to the limitations of the current hardware, it is difficult to obtain high-resolution remote sensing images from a single sensor. Only single-band panchromatic images and multiband multispectral images can be obtained, respectively, in this way. The

information carried by the two images is different, but they are complementary [6, 7]. As the remote sensing image becomes more and more important, the remote sensing image fusion algorithm is constantly improved. How to fuse the spatial and spectral information of panchromatic image and multispectral image as much as possible to improve the fusion effect is a key concern in remote sensing image fusion [8]. Pansharpening plays a significant role in it. So far, many methods of pansharpening have been proposed. They can be divided into three categories roughly.

The first category is component substitution (CS) [9]. In this method, the low resolution multispectral image is transformed into spatial and spectral components. After that, the spatial components are replaced by a high spatial resolution panchromatic image. Finally, the fused image is obtained by the inverse transformation [10]. It is apparent that the spatial characteristics of the resulting image after pansharpening are closely related to the replacement components of the panchromatic image. The popular algorithms are IHS (Intensity Hue Saturation) transformation [11], Principal component analysis (Principal Component Analysis) [12], Brovey transform (Brovey) [13], and so on. The fusion algorithm based on the component substitution has high computational efficiency, but it prones to the spectral distortion [14].

Multiresolution analysis (MRA) is the second category, which aims to extract the spatial information of panchromatic images by multiresolution analysis such as the Laplace pyramid and the wavelet transform. Fused images are obtained by the weighted fusion and the reconstruction of different image multiresolution representation coefficients. Compared with the component substitution, the method based on MRA is unlikely to be affected by the spectral distortion. For example, there is an intensity modulation based on smoothing filter (SFIM) in [15] and the coupled nonnegative matrix factorization (CNMF) for data fusion in [16]. Modulation Transfer Function-Generalized Laplacian Pyramid (MTF_GLP) [17] is also proposed. This kind of method with high computational complexity prones to the space distortion [18].

The third category is based on the method of deep learning, in which the low-resolution multispectral image is regarded as the input of the deep network [19]. Moreover, the high-resolution image is outputed by learning the end-to-end mapping of the low-resolution and high-resolution image quality [20]. The performance of this kind of fusion method is improved by constructing more reasonable loss function, processing image residual and using deeper frame structure [21–24]. Traditional methods, whether in spectral domain or spatial domain, have great losses. In order to overcome the above shortcomings, the method based on the deep learning has a typical advantage. It can reduce the difficulty of training the network by initializing parameters layer by layer and finish the initialization layer by layer by means of the unsupervised learning. The deep learning can simulate the hierarchical structure of the visual perception system. After the establishment of a machine learning model with multiple hidden layers, more useful features can be obtained from the training of a large number of data [25]. For example, in [26], the three-layer CNN structure based on convolution neural network is used for the remote sensing image fusion, and the remote sensing image fusion algorithm PNN (Pansharpening by CNN) based on CNN is proposed, which significantly promotes the performance of remote sensing image fusion algorithm. In [27], the deep residual network is used to increase the accuracy of pansharpening of multispectral images (deep residual pansharpening neural network, DPRNN). There is also a deep network architecture (deep network architecture for pansharpening, PanNet) for pansharpening [28]. To get a ampler extraction of features, a two-channel CNN to extract image features of PAN and LMS, respectively, is proposed. The residual join is also adopted to enhance the ability of feature learning and the results of image fusion in [29]. The generative adversarial network (GAN) is proposed in [30]. By inputting PAN and MS images, the estimated MS images share the same distribution with the reference HRMS images. It can extract and transmit spectral and spatial features, respectively, and measure the differences between distributions effectively. According to the structural characteristics of multiscale information, a fusion rule based on the fuzzy logic is proposed to fuse low-order components, which is able to effectively fuse the high frequency components of PAN and MS images [31].

However, there are some problems in the deep learning algorithm above mentioned during the fusion. The information contained in the pan image and MS image is different. There is a lot of high-frequency and texture information in pan image while MS images are abundant in spectral information and low-frequency components. It is hard to learn the relationship between them directly with high redundancy [32]. A larger convolution kernel will lead to a more complicated network model and more network parameters in the training process. Thus the test will be more difficult. In the process of the convolution, each convolution operator has only one local receptive field. It cannot make full use of the context information; so, the obtained features also lack the context information [33].

In recent years, good results have been achieved since the channel attention mechanism is applied to the superresolution of the image by [34–36] In [37], a hybrid attention mechanism applied in pansharpening is proposed to alleviate the spectral distortion and improve the spatial resolution. A channel spatial attention residual block (CSAResBlocks) is proposed in [38], which can make full use of the relationship between the channel and the space of the feature graph at the same time. The accuracy of pansharpening is able to be further increased by stimulating more related features and suppressing the less useful features in some spectral channels and spatial positions. In addition, the attention mechanism module is added to the Tri-UNet in [39] to make the network extract multilevel features and reduce the loss of details in the downsampling process in the meantime.

In order to solve the above problems, a two-stream fusion network based on the residual channel attention mechanism is proposed in this paper. During the fusion, because there is no high-resolution multispectral reference image, the fusion image is degraded by design space degradation, and then

the fusion network is trained. In order to obtain a better fusion effect, the method also adds a channel attention mechanism, which effectively improves the fusion performance and better retains the spectral characteristics. The main contributions of this paper are as follows.

(1) A novel remote sensing image fusion network is proposed for pansharpening. This is a deep end-to-end structure of the network that is designed to generate remote sensing images with high spatial resolution

(2) The model of two-stream fusion network combined with the residual channel attention mechanism is constructed. The detail features of PAN images and the spectral features of MS images can be extracted, respectively, to generate deeper features

(3) A residual channel attention group is designed to model the interdependence between feature channels. It can adaptively adjust the feature weight of each channel and get useful information as much as possible. Multiple residual connections allow a large number of shallow information to realize jump connections, simplifying the flow of information

The rest of this article is organized as follows. In Section 2, we introduced residual learning and channel attention mechanisms. Section 3 introduces our network structure and explains in detail. Section 4 gives experiments and comparisons with other methods. The conclusion of the article is in Section 5.

2. Related Work

2.1. Residual Learning for Pansharpening. The higher precision can be obtained based on the fully extract features by the deep network with more convolution layers and hidden layers. However, the deep network always cause a lot of problems such as the gradient disappears, the gradient explodes and so on. The residual learning is regarded as the useful technique, in which the output is expressed as a linear superposition of a nonlinear transformation of input and the spatial distribution of residual features will be very sparse [40, 41]. As a result, it becomes easier to find the optimal weights and biases (W, b) while the network is allowed to add more hidden layers and show a better performance [42]. In the pansharpening process, the residual learning can be divided into the following two stages:

Stage 1: the residual output F_{Stage1} can be calculated to predict the residual under the residual connection according to the input M as follows:

$$\begin{aligned} &F_0 = M, F_l = \max\left(0, W_l \times F_{l-1} + b_l\right), l = 1, \cdots, L-1, \\ &F_{\text{Stage1}} = M + F_{L-1}. \end{aligned} \tag{1}$$

W_l is the weight of the l layer, and b_l is the bias of the l layer. F_{Stage1} denotes the residual output of the layer $L-1$.

Stage 2: the L layer of the residual network is established, and the band of $N+1$ is restored to N bands by the final convolution operation of the network. Thus, the final result F_{Stage2} can be computed by

$$F_{\text{Stage2}} = W_L \times F_{\text{Stage1}} + b_L. \tag{2}$$

2.2. Channel Attention Mechanism. Employing the convolutional neural network, for any picture initially represented by three channels (R, G, B), a new signal can be generated in each channel after passing through different convolution kernels. For example, 64 kernels convolution is chosen for each channel of the image feature, and then it will generate a matrix of new 64 channels $(H, W, 64)$, where H, W represents the height and width of the picture feature, respectively. The characteristics of each channel actually represent the components of the picture on different convolution kernels, in which the convolution of the convolution kernel is similar to the Fourier transform of the signal. Thus, the information of one channel of this feature can be decomposed into signal components on 64 convolution kernels [43]. Since each signal can be decomposed into components on the kernel function, the new 64 channels should involve the key information more or less. If the weight parameter is introduced for each channel to show the correlation with the key information, it is clear that the greater weight parameter always indicates the higher correlation. Thus, we should pay more attention to the channels with greater weight parameters.

Based on the above mentioned correlation, a "squeeze-excitation module" has been proposed in the literature [44]. At first, it clearly expresses the interdependence between channels and then adaptively calibrates the channel-related characteristic responses by the dependence to improve the quality of features generated by the network. Applying a channel attention mechanism, the information aggregation can be achieved by squeezing for the global information while the related dependencies between channels can be captured by excitation. In order to obtain the channel correlation of aggregated information, a gating mechanism is set up to learn the nonlinear interaction between channels and then realize the activation of multichannel features. The gate here can determine whether the data should be retained or discarded in a sequence, and it can pass relevant information to a longer sequence chain for prediction. The sigmoid activation function, which is adopted by the simple gating mechanism, is similar to the tanh activation function, but it controls the value between 0 and 1 instead of -1 to 1. It helps to update or discard data, because any number multiplied by 0 is 0, which will cause the value to disappear or be omitted. Any number multiplied by 1 is itself; so, the value remains unchanged or saved. The network can identify which data is unimportant, which can be omitted and which needs to be saved.

In the pansharpening process, PAN images are single-channel while MS images are multichannel. In the fusion process, the PAN image and the MS image are firstly stitched on the channel. If each channel is treated equally,

the space of the PAN image and the spectral information of the MS image cannot be well preserved. By applying the channel attention mechanism for the remote sensing image fusion, the spatial characteristics of PAN images and the spectral characteristics of MS images can be extracted in a targeted manner.

3. Two-Stream Fusion Network Based on Residual Channel Attention Mechanism

3.1. Network Framework. The network framework of this paper is shown in Figure 1. There are three main steps to generate high-resolution multispectral images. Firstly, the two-stream CNN architecture extracts features from PAN and LMS images, and then these features are merged on the channel to form a compact feature map. Such feature maps simultaneously represent the spatial information and spectral information of the PAN and LMS images. Before feature extraction, the PAN image will be downsampled to fit the size of the LMS image. Secondly, the fused feature map passes through the residual channel attention group. The residual attention group uses residual connection to allow the residual learning of shallow features. It contains multiple residual channel attention modules that use a combination of the channel attention mechanism and the residual connection. The attention mechanism can adaptively adjust the characteristics of each channel by modeling the interdependence between feature channels. This attention mechanism enables the proposed network to focus on more useful channels and improve the learning ability of the channel feature. The residual connection stacks several transformation layers, which allows a large amount of shallow information to skip connections, simplifying the flow of information. Thirdly, the output of the residual attention group is finally enlarged by the deconvolution layer, and the enlarged features are reconstructed by a convolution layer so as to obtain the final high-resolution multispectral image.

3.2. Two-Stream Fusion Network. PAN images and LMS images contain different information. PAN has abundant spatial information while LMS is the carrier of spectral information. Convolution neural network (CNN) has great feature to represent the ability and reconstructs feature images [45] very well. As a result, fusing PAN and LMS in the feature domain is considered as shown in Figure 2.

First of all, the input of the two images of PAN and LMS is represented by Xp and Xm, and the features extracted by CNN are represented by S_p^1 and S_m^1. The superscript 1 represents the extraction of features from layer 1. In order to fully obtain spectral and spatial information, multiple convolution layers are used to extract layered features, and then the two are fused together and concatenated on the channel. To preserve features and prevent information loss without using pooling layers, batch normalization, and ReLU, they are simply connected together to implement a fusion strategy.

The specific formula is as follows:

$$\begin{aligned} S_p^1 &= W_{3\times3}^1 \times X_p + b_{3\times3}^1, \\ S_p^2 &= W_{3\times3}^2 \times S_p^1 + b_{3\times3}^2, \\ S_m^1 &= W_{3\times3}^1 \times X_m + b_{3\times3}^1, \\ S_m^2 &= W_{3\times3}^2 \times S_m^1 + b_{3\times3}^2, \\ C &= \left[S_p^2, S_m^2\right]. \end{aligned} \tag{3}$$

$W^1$and W^2 represent the weights of the first and the second convolution layers. b^1 and b^2 represent the bias of the first and the second convolution layers. 3×3 indicates the size of the convolution kernel. S_p^2 and S_m^2 represent the convolution output of layer 2. $C = [S_p^2, S_m^2]$ means concatenating the two tensors together. Then, the characteristics of the connection are inputted into the network to form a more compact representation, and follow-up operations are carried out. The Concat is the spliced operation. The proposed network accepts two inputs and has a two-stream architecture; so, it is called a two-stream fusion network.

3.3. Residual Attention Group. The input PAN image and MS image can be processed by the dual-stream fusion network to form a compact feature representation, but these feature data more or less contain some redundant information. The previous neural network method treats the information between channels equally; so, the expressive ability of the network is hindered. At the same time, previous experiments [46] prove that fairly deep networks are the most direct way to improve the performance of the deep learning of networks. However, it is very difficult to train deeper network layers which are prone to result in accuracy saturation and network degradation. In order to solve these problems, the residual attention group is introduced to construct deep network as shown in Figure 3.

The residual attention group consists of the residual attention module, the convolution layer, and the residual connection. The residual connection in the group allows the residual learning of shallow features. The residual attention group contains a plurality of residual attention blocks as shown in Figure 4.

Each residual attention block contains some simple transformation layers, channel attention mechanisms, long residual connections, and short residual connections. Long residual connection and short residual connection allow shallow information to propagate backward directly through identity mapping, which is beneficial to the flow of information. Channel attention mechanism adaptively assigns different weights to each channel by modeling the interdependence between characteristic channels, which allows the network to focus on more useful channels and enhance the learning ability.

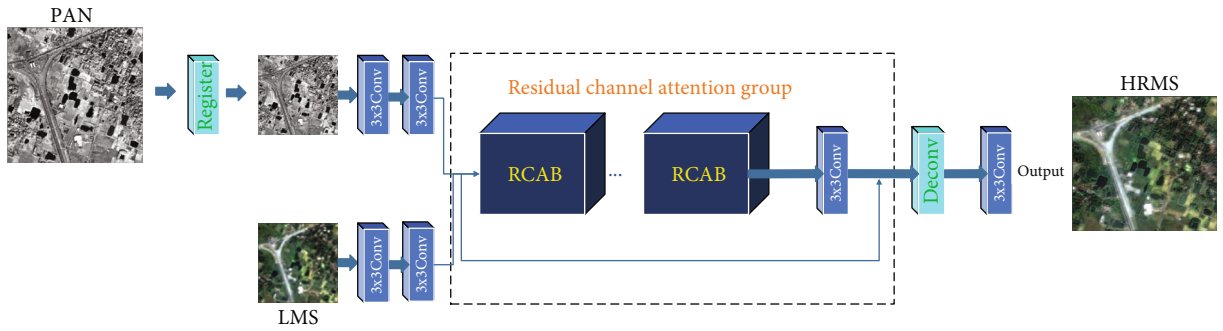


Figure 1: Network framework.

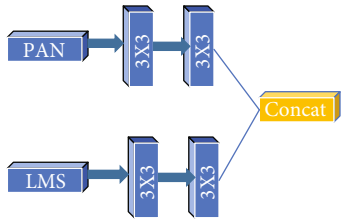


Figure 2: Two-stream fusion network.

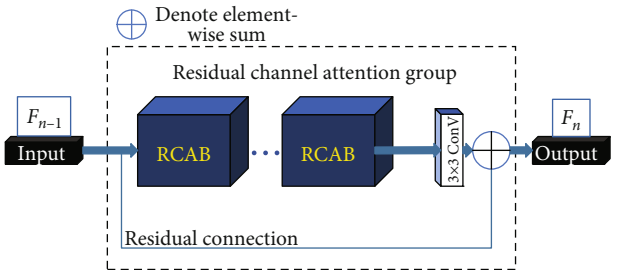


Figure 3: Residual attention group.

The specific formula is as follows:

$$\begin{aligned} X_{b-1} &= \delta\left(F_{b-1} \times W^1_{3\times3} + b^1_{3\times3}\right), \\ X_b &= W^2_{3\times3} \times X_{b-1} + b^2_{3\times3}, \\ F_b &= CA(X_b) + F_{b-1}. \end{aligned} \tag{4}$$

F_{b-1} and F_b are the input and output of the residual attention block, respectively. X_{b-1} is the output of the input F_{b-1} after the convolution of ReLU. X_b is obtained after the next convolution, and it can be used as the input of the channel attention mechanism. W^1 and W^2 represent the weight of the first and the second convolution layers, respectively, while b^1 and b^2 represent their bias. 3×3 indicates the size of the convolution kernel. δ represents the ReLU function while CA represents the channel attention mechanism function.

In the pansharpening process, channel features are dealt with equally by the previous CNN-based method of remote sensing image fusion. In the network, there is only one local receptive field for each convolution kernels in the convolution layer [47]. Therefore, the convoluted output cannot make use of context information outside the local area. In order to allow more information features to be paid attention to by the network, the interdependence between characteristic channels is made use of to generate channel attention mechanism (CAM) as shown in Figure 5.

The channel attention mechanism uses global average pooling to transform global spatial information into channel descriptors in order to obtain channel statistics z. Suppose $X = [X_1, X_2, \cdots, X_c]$ is the input of a channel attention mechanism. For example, there is a feature map with C channels, and its size is (H, W). The descriptor Z_c of the C channel is the average of the values on this channel (global average pooling).

The specific formula is as follows.

$$\begin{aligned} z_c &= f_{GP}(x_c) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} x_c(i, j), \\ w &= S(W_U \delta(W_D z)), \\ \tilde{x_c} &= w_c \times x_c. \end{aligned} \tag{5}$$

$f_{GP}()$ is the global average pooling function, and $x_c(i, j)$ is the value of x_c at (i, j) that is the characteristic in the layer c. In addition to the global pooling, the aggregation technology is introduced. The channel dependency is completely captured from the aggregated information by the global average pooling. After that, the sigmoid activation function is introduced to learn the nonlinear interaction between channels. $S()$ denotes the sigmoid activation function, and $\delta()$ denotes the ReLU activation function. W_U is the weight set of the descending-dimensional convolution layer while W_D is that of the ascending-dimensional convolution layer. The dimension of the action channel of the descending-dimensional convolution layer is reduced, and the dimension reduction ratio is set to r. After the descending-dimensional signal is activated by the ReLU activation function, the number of channels is increased by r times through the ascending-dimensional convolution layer so that the weight coefficient of each channel can be obtained. Then, the final channel statistic w is obtained, which is used to re-adjust the weight and feature map of the input. x_c, w_c,

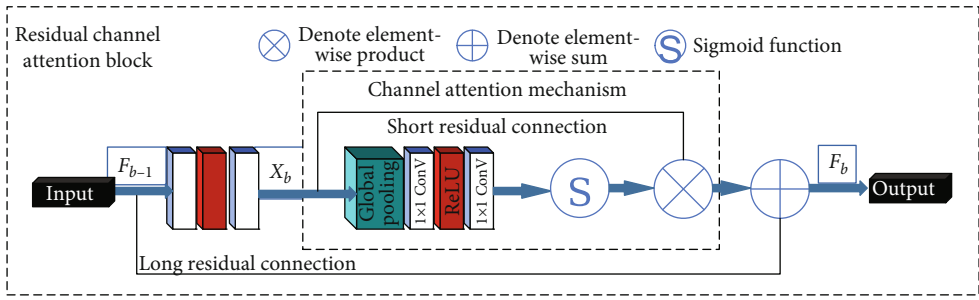


Figure 4: Residual attention block.

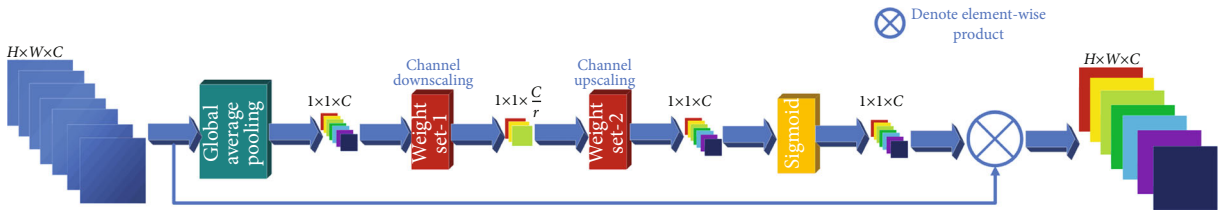


Figure 5: Channel attention mechanism.

and x_c are the weights and feature maps of layer c channels, respectively. The input is a feature of $H \times W \times C$. Firstly, the global average pooling of a space is carried out to get a $1 \times 1 \times C$ channel description. Then, the weight coefficient of each channel is obtained through a descending-dimensional convolution layer and an ascending-dimensional convolution layer. The new feature can be obtained by multiplying the weight coefficient and the original feature. In fact, the whole process is actually a weighted redistribution of the feature in different channels.

3.4. Loss Function. The loss function is an important factor that affects the quality of the fused image. In this paper, L1 loss function is adopted to optimize the network and give the panchromatic image of training sample and the low-resolution multispectral image as well as the high-resolution multispectral image. The L1 loss function is defined as

$$L_{1\_}\text{loss}(\Theta) = \frac{1}{N}\sum_{i=1}^{N}\left|\Phi\left(X^i_{\text{PAN}}, X^i_{\text{MS}}; \Theta\right) - X^i_{\text{HMS}}\right|, \quad (6)$$

where $|\cdot|$ denotes the L1-norm, which computes the mean absolute error between the generated and the reference data. Θ represents the network parameters, and N represents the number of training samples. X^i_{PAN}, X^i_{MS}, and X^i_{HMS} represent PAN images, LMS images, and reference multispectral images, respectively.

4. Experiment and Analysis

4.1. Dataset and Experiment Settings. The applicability of the network is verified by two different satellite datasets, GaoFen-2 [26] and WorldView-2 [28]. The detailed data sets are shown in Table 1. According to the Wald [48], the original MS image is used as the reference image, and then the PAN image is downsampled to be registered with the LMS image for training and testing.

The spatial resolutions of the MS image and PAN image of the GF-2 satellite are 3.2 m and 0.8 m, respectively, while the corresponding images of the WV-2 satellite are 4 m and 1 m, respectively. During training, the original ms is used as a reference, ms and pan are downsampled and trained on the degenerate model, and then the pan after ms upsampling and downsampling remains the same size and input into the model for training.

That is, WV-2 image uses the downsampled and then upsampled 4 m resolution ms and the 4 times downsampled 4 m resolution pan as the network input image and the original 4 m resolution ms image as the ideal image. The resolution of the ms image of GF-2 is 3.2 m, the resolution of the panchromatic image is 0.8 m, and the downsampled and then upsampled ms of 3.2 m resolution and the 4 times downsampled pan of 3.2 m resolution are used as the network. The input image is the original 3.2 m resolution MS image as the ideal image.

There are 3 residual attention blocks in the setting of the experiment. Except for the convolution kernel of 1×1 in the channel attention mechanism, that of 3×3 is adopted in the other. Downsampling 4 times to register the LMS image, the number of extracted feature layers is 64, and the ratio of channel dimensionality reduction in the channel attention mechanism is 16. The Adam optimizer is adopted to optimize the network parameters. The initial learning rate is 0.001 and every 200 epoch learning rate multiplied by 0.5. The experiment is conducted in Pytorch and Python3.7 environments, including Intel Xeon CPU, a 11 GB NVIDIA Tesla V100 PCIe GPU, and 16 GB RAM.

4.2. Evaluation Method. In the image fusion experiment part, select peak signal-to-noise ratio (PSNR), structural similarity (SSIM), spectral angle mapping (SAM), relative

Table 1: Datasets.

Type of sensor	GaoFen-2	WorldView-2
PAN resolution	0.8 m	1 m
MS resolution	3.2 m	4 m
Bands number	4	5
Width	45 km	45 km
Return days	5 days	5 days
Train patches	2099	2372
Test patches	128	440
PAN image size	256×256	256×256
LMS image size	64×64	64×64

global error (ERGAS), spatial correlation coefficient (SCC), and general image quality index (Q). These 6 indicators are used to evaluate the fusion results.

(1) Peak signal-to-noise ratio (PSNR) [49], which is mainly used to evaluate the sensitivity error of fused images' quality, is an important indicator to measure the difference between two images. The higher the value of PSNR is, the more similar the two images are

$$\mathrm{PSN}R = 10 \lg \left(\frac{(2^n - 1)^2}{\mathrm{MSE}} \right). \tag{7}$$

MSE represents the mean square error of the fused image and the reference image.

(2) Structural similarity (SSIM) [50] is an indicator that measures the similarity of two images, in which the picture is highly structured. There is a strong correlation between adjacent pixels. The value of SSIM is between 0 and 1. $SSIM = 1$ means that the two images are exactly alike:

$$\mathrm{SSIM}(x, y) = [l(x, y)]^{\alpha} [c(x, y)]^{\beta} [s(x, y)]^{\gamma}. \tag{8}$$

$l(x, y)$ compares luminance; $c(x, y)$ compares contrast; $s(x, y)$ compares structure. X represents the fused image, and Y represents the reference image. L, C, and S are used to make a comparison of the brightness information, the contrast information, and the structure information between the reference image and the fused image, respectively.

(3) Spectral angle mapping (SAM) [51] reflects the degree of spectral distortion and calculates the angle between the corresponding pixels of the fused image and the reference image on the triad. $SAM = 0$ means that there is no spectral distortion:

$$\mathrm{SAM}\left(v, v'\right) = \cos^{-1} \left(\frac{\sum_{i=1}^{N} v_i v_i'}{\sqrt{\sum_{i=1}^{N} v_i^2} \sqrt{\sum_{i=1}^{N} v_i'^2}} \right). \tag{9}$$

There are N bands in every image. v and v', respectively, represent the vector set of pixels in different bands of the fused image and the reference image with N components.

(4) Erreur Relative Global Adimensionnelle de Synthese (ERGAS) [52] demonstrates the spectral information contained in the whole band of the image by the resolution ratio of the PAN image to the MS image. The lower the value of ERGAS is, the better the fusion effect is

$$\mathrm{ERGAS} = 100 \frac{R_p}{R_M} \sqrt{\frac{1}{N} \sum_{i=1}^{\mathrm{N}} \frac{\mathrm{RMSE}^2(F_i)}{\bar{F}_i^2}}. \tag{10}$$

R_P and R_M represent the resolution of the PAN image and the MS image, respectively. F_i is each band component of the fused image, and $\bar{F}_i$ is the average value of the band component of the fused image.

(5) The spatial correlation coefficient (SCC) [53] demonstrates the spatial correlation between the fused image and the multispectral image. The value of the SCC reflects the correlation degree between the reference multispectral image and the fused image. If the value of SCC is closer to 1, it indicates that the closer the relationship between the reference multispectral image and the fused image is, the better the fusion effect is

$$\mathrm{SCC}(P, F) = \frac{\sum_{i=1}^{N} \sum_{i=1}^{N} (P_i - u_P)^2 (F_i - u_F)^2}{\sqrt{\sum_{i=1}^{N} (P_i - u_P)^2 \sum_{i=1}^{N} (F_i - u_F)^2}}. \tag{11}$$

u_P denotes the mean value of PAN(P) while u_F denotes the mean value of the fusion image(F).

(6) Q [54] is the average general image quality indicator. The value range of it is (-1, 1). The closer the value is to 1, the more similar the two images are, and the better the fused image quality is

$$Q = \frac{4 \sigma_{\mathrm{MF}} \mathrm{MF}}{\left(\sigma_M^2 + \sigma_F^2\right)\left(M^2 + F^2\right)}. \tag{12}$$

σ_M and σ_F denote the gray-scale covariance of the fusion image and the MS image, respectively.

Since it is not possible to directly obtain high-resolution multispectral images, objective indicators without reference images are used in the original image fusion experiment to evaluate the fusion results.

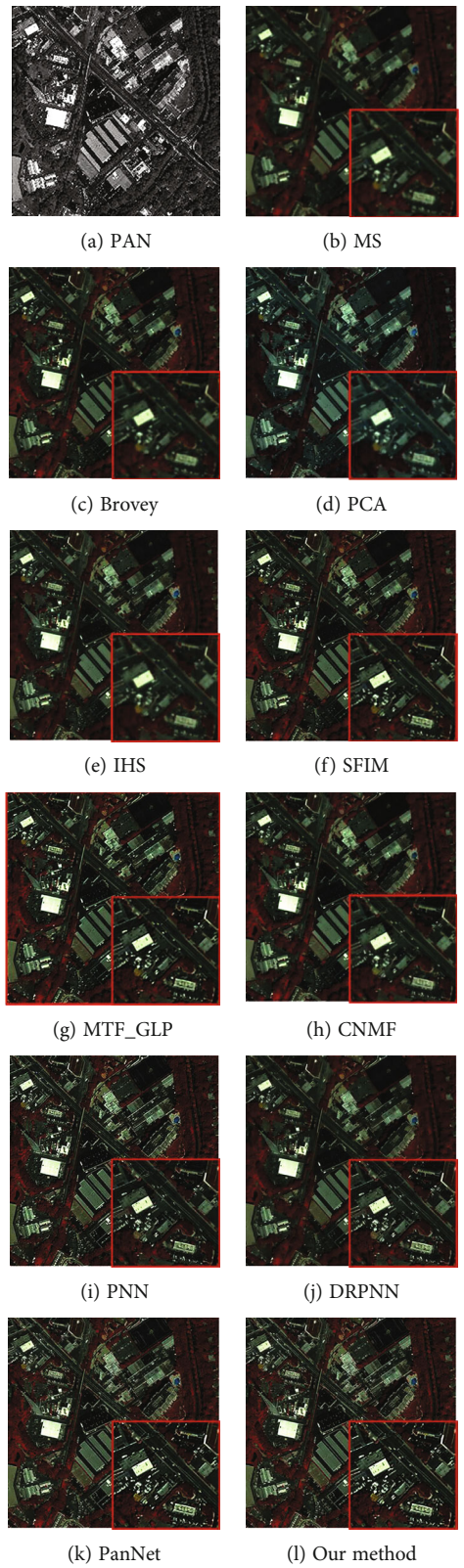


Figure 6: The results of different remote sensing image fusion algorithms under the WorldView-2 satellite datasets.

The nonreference quality indicator (QNR) [55] includes the index $\mathbf{D}_\lambda$ of evaluating the loss of spectral details and $\mathbf{D}_S$ of evaluating the loss of spatial details. The spectral detail loss $\mathbf{D}_\lambda$ aims to explore the disparity of the relationship between the bands of the fused image and that of the multispectral image. The smaller the $\mathbf{D}_\lambda$ is, the more spectral

Table 2: Evaluation indicators of different remote sensing image fusion algorithms under WorldView-2 satellite datasets.

Reference comparison						
Methods	SCC	SSIM	Q	PSNR	ERGAS	SAM
Brovey	0.8481	0.7095	0.5476	43.1899	5.8498	0.1004
PCA	0.8903	0.7031	0.4028	44.2140	4.7886	0.1069
IHS	0.8732	0.6831	0.4642	44.2607	5.0181	0.0951
SFIM	0.9139	0.7472	0.6146	44.6207	3.8994	0.0937
MTF_GTP	0.8858	0.6751	0.4254	44.8155	4.8367	0.0867
CNMF	0.9271	0.7706	0.6139	45.7792	3.8437	0.0888
PNN	0.9281	0.7791	0.6252	45.9311	3.7992	0.0885
DPRNN	0.9312	0.7825	0.6371	46.6754	3.5910	0.0851
PanNet	0.9350	0.7838	0.6215	46.2696	3.5771	0.0816
Our algorithm	0.9422	0.7891	0.6432	47.5762	3.4545	0.0656
Nonreference comparison						
Methods	$\mathbf{D}_\lambda$		$\mathbf{D}_S$		QNR	
Brovey	0.1002		0.1063		0.8041	
PCA	0.0911		0.0401		0.8724	
IHS	0.0897		0.1143		0.8063	
SFIM	0.0649		0.1257		0.8175	
MTF_GLP	0.0624		0.0764		0.8643	
CNMF	0.0595		0.1687		0.7818	
PNN	0.0523		0.1022		0.8508	
DPRNN	0.0448		0.0540		0.9036	
PanNet	0.0323		0.0274		0.9412	
Our algorithm	0.0204		0.0128		0.9671	

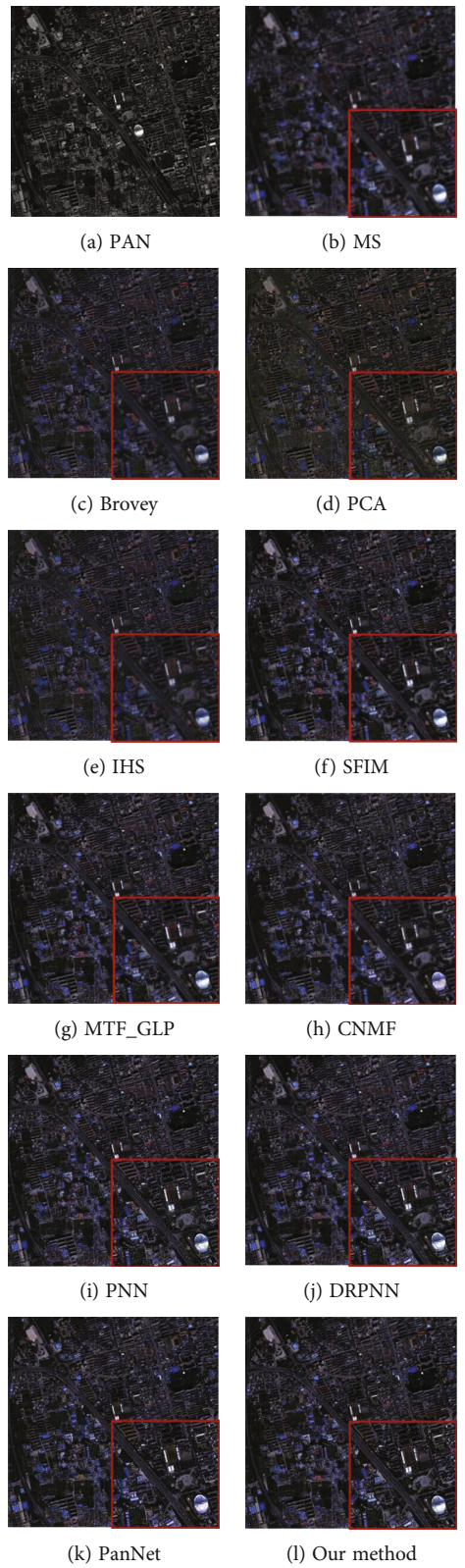
features are retained among the fused image bands. The spatial detail loss $\mathbf{D}_S$ aims to explore the disparity between the bands of the fused image and the panchromatic image. The smaller the $\mathbf{D}_S$ is, the closer the spatial structure feature of each band of the fused image to the panchromatic image is. Moreover, the smaller the spectral detail loss and spatial detail loss of the fused image are, the higher the corresponding QNR is. The ideal value of $\mathbf{D}_\lambda$ and $\mathbf{D}_S$ is 0, and that of QNR indicator is 1.

$$\mathrm{QNR} = (1 - D_\lambda)^\alpha (1 - D_S)^\beta. \tag{13}$$

4.3. Experimental Result Analysis. In this part, the superiority of the algorithm in this paper is verified by a large amount of quantitative and visual evaluation. Nine remote sensing image fusion algorithms are compared in the experiments, including methods based on component substitution, multiresolution analysis, and deep learning. There are Brovey [11], PCA [12], IHS [13], SFIM [15], MTF_GLP [16], CNMF [17], PNN [26], DPRNN [27], and PanNet [28].In the simulation experiment, the existing MS image is used as a reference, the PAN image is sampled to the same size as the MS image, and the same multiple of the MS image is taken as the LMS image. The fusion results are consistent with the existing multispectral image size and measure the relevant indicators.

4.4. WorldView-2 Datasets. The experimental results of different remote sensing image fusion algorithms by means of WorldView-2 satellite datasets are demonstrated in Figure 6 below [26]. Figure 6(a) is PAN image, and Figure 6(b) is MS image. Figures 6(c)–6(l) are images of Brovey, PCA, IHS, SFIM, MTF_GLP, CNMF, PNN, DPRNN, PanNet, and the algorithm in this paper, respectively. For a more accurate observation, an enlarged view in the middle area is provided in the lower right corner of the picture.

According to the results in Figure 6, it is clear that the results of the experiment obtained by Brovey, PCA, and IHS algorithms have a certain degree of spectral distortion; especially, the images obtained by PCA algorithm are obviously blue. The result details of SFIM, CNMF, and MTF_GLP algorithm are vague, the lack of detail information is serious, and there are some artifacts on the road. Compared with the traditional algorithms based on the deep learning such as Brovey, PCA, and IHS, PNN, DPRNN, PanNet, and the algorithm in this paper have an obvious improvement in spatial features, and the details are better preserved; the overall visual effect is better. However, the algorithm in this paper performs better in terms of spectrum. Consequently, it gets good results from the subjective evaluation. As shown in Table 2, further analysis indicates the numerical comparison of various remote sensing image fusion algorithms in evaluation indicators.



(a) PAN (b) MS

(c) Brovey (d) PCA

(e) IHS (f) SFIM

(g) MTF_GLP (h) CNMF

(i) PNN (j) DRPNN

(k) PanNet (l) Our method

Figure 7: The results of different remote sensing image fusion algorithms under the GaoFen-2 satellite datasets.

As for the referenced evaluation indicators, it is better when the values of SCC, SSIM, Q, and PSNR are larger while those of ERGAS and SAM are smaller. In the nonreferenced evaluation indicators, it is better when the values of $\mathbf{D}_\lambda$ and $\mathbf{D}_S$ are smaller while those of QNR are larger. From the table, it is obvious that the maximum value is obtained in

Table 3: Evaluation indicators of different remote sensing image fusion algorithms under GaoFen-2 satellite datasets.

Reference comparison						
Methods	SCC	SSIM	Q	PSNR	ERGAS	SAM
Brovey	0.8914	0.7991	0.7452	42.0577	7.9374	0.2451
PCA	0.8364	0.7284	0.6446	42.2696	9.2124	0.2716
IHS	0.8831	0.7768	0.7236	43.5003	7.8865	0.1796
SFIM	0.8713	0.7132	0.7451	43.3909	9.0278	0.1865
MTF_GTP	0.9271	0.8021	0.7467	44.5458	6.2051	0.1932
CNMF	0.9281	0.8275	0.7717	44.9722	6.8851	0.1993
PNN	0.9471	0.8021	0.7631	46.4144	5.2051	0.1432
DPRNN	0.9489	0.8275	0.7747	46.9355	4.9311	0.1333
PanNet	0.9576	0.8619	0.8395	47.7915	4.8269	0.1364
Our method	0.9681	0.8656	0.8499	48.0889	4.6449	0.1216
Nonreference comparison						
Methods	$\mathbf{D}_\lambda$		$\mathbf{D}_S$		QNR	
Brovey	0.0744		0.1968		0.7435	
PCA	0.1047		0.1564		0.7553	
IHS	0.1195		0.1165		0.7779	
SFIM	0.0606		0.1485		0.7999	
MTF_GLP	0.1069		0.1018		0.8022	
CNMF	0.1061		0.0925		0.8112	
PNN	0.1511		0.0319		0.8218	
DPRNN	0.0900		0.0926		0.8275	
PanNet	0.0734		0.0889		0.8442	
Our method	0.0824		0.0532		0.8688	

SCC, SSIM, Q, PSNR, and QNR indicators while minimum value is obtained in ERGAS, SAM, $\mathbf{D}_\lambda$, and $\mathbf{D}_S$, which show the optimal results in various indicators. The reduction of ERGAS value in the algorithm is the most visible, which is 40.9% lower than the traditional one and 3.4% lower than the pannet algorithm. Combined with the previous visual effects, it can be concluded that the algorithm in this paper has achieved good results both subjectively and objectively on WorldView-2 satellite images.

4.5. GaoFen-2 Datasets. The experimental results of different remote sensing image fusion algorithms by means of GaFen-2 satellite datasets are demonstrated in Figure 7 below [25]. Figure 7(a) is PAN image, and Figure 7(b) is MS image. Figures 7(c)–7(l) are images of Brovey, PCA, IHS, SFIM, MTF_GLP, CNMF, PNN, DPRNN, PanNet, and the algorithm in this paper, respectively. For a more accurate observation, an enlarged view in the middle area is provided in the lower right corner of the picture.

According to the results in Figure 7, it is clear that the result images of Brovey, PCA, and IHS algorithms are very blurry, and the edges of the local images are not clear. The result images of SFIM, MTF_GLP, and CNMF algorithms have a serious degree of spectral and spatial distortion. On the whole, the result images of the deep learning, PNN, DRPNN, PanNet, and the algorithm in this paper are superior. From the local enlarged image, the outline and texture of the magnified image are clearer in the algorithm in this paper. As shown in Table 3, further analysis indicates the numerical comparison of various remote sensing image fusion algorithms in evaluation indicators.

As for the referenced evaluation indicators, it is better when the values of SCC, SSIM, Q, and PSNR are larger while those of ERGAS and SAM are smaller. In the nonreferenced evaluation indicators, it is better when the values of $\mathbf{D}_\lambda$ and $\mathbf{D}_S$ are smaller while those of QNR are larger. From the table, it is obvious that the maximum value is obtained in SCC, SSIM, Q, PSNR, and qnr indicators while minimum value is obtained in ERGAS, SAM, and $\mathbf{D}_S$. $\mathbf{D}_\lambda$ is slightly inferior to the traditional algorithm with little difference. The reduction of ERGAS value in the algorithm is the most visible, which is 49.5% lower than the traditional one and 10.8% lower than the pannet algorithm. Combined with the previous visual effects, it can be concluded that the algorithm in this paper has achieved good results both subjectively and objectively on GaoFen-2 satellite images.

4.6. The Influence of Different Channel Attention Mechanisms. In this paper, an architecture based on residual channel attention mechanism is proposed. In order to illustrate the influence of the channel attention mechanism on the results of the experiment, a comparative experiment is conducted. The experiment of adding channel attention mechanism (CAM), adding residual channel attention

Table 4: The influence of different channel attention mechanisms in WorldView-2.

Reference comparison						
Methods	SCC	SSIM	Q	PSNR	ERGAS	SAM
Without CAM	0.8912	0.7525	0.6071	45.9612	3.7911	0.0922
CAM	0.9320	0.7781	0.6115	47.1254	3.4923	0.0716
RCAM	0.9422	0.7891	0.6432	44.5762	3.4545	0.0656
Nonreference comparison						
Methods	$\mathbf{D}_\lambda$		$\mathbf{D}_S$		QNR	
Without CAM	0.0497		0.0421		0.9103	
CAM	0.0367		0.0233		0.9408	
RCAM	0.0204		0.0128		0.9671	

Table 5: The influence of different channel attention mechanisms in GaoFen-2 datasets.

Reference comparison						
Methods	SCC	SSIM	Q	PSNR	ERGAS	SAM
Without CAM	0.9275	0.8266	0.8023	43.8921	4.9563	0.1639
CAM	0.9612	0.8534	0.8223	47.1221	4.7677	0.1498
RCAM	0.9681	0.8656	0.8499	48.0889	4.6449	0.1216
Nonreference comparison						
Methods	$\mathbf{D}_\lambda$		$\mathbf{D}_S$		QNR	
Without CAM	0.1033		0.0987		0.8082	
CAM	0.0894		0.0639		0.8524	
RCAM	0.0824		0.0532		0.8688	

mechanism (RCAM), and nonadding channel attention mechanism (without CAM) is compared in the evaluation indicator. The experimental results of the WorldView-2 datasets are shown in Table 4, and those of the GaoFen-2 datasets are shown in Table 5.

Several conclusions can be drawn from the table above. The data obtained from the network with RCAM module is the best in all aspects, and that obtained without any channel attention module is the worst, which illustrates that the channel attention mechanism can indeed improve the effect of the network and the RCAM module is able to help the network show better performance.

5. Conclusion

In this paper, a two-stream remote sensing image fusion network based on the residual channel attention mechanism (RCAMTFNet) is proposed. In the RCAMTFNet, the spatial features of PAN and the spectral features of MS are extracted, respectively, by a two-channel feature extraction layer. Multiresidual connections allow the network to adapt to a deeper network structure without degradation. The residual channel attention mechanism is introduced to learn the interdependence between channels, and then the correlation features among channels are adapted on the basis of the dependency. In this way, image spatial information and spectral information are extracted exclusively. What is more, pansharpening images are reconstructed across the board. The results of the experiment show that the algorithm in this paper has a better performance on various objective evaluation indicators of fused images obtained from different datasets, and it can integrate the spatial features of PAN images and the spectral information of LMS images effectively.

In the convolution neural network, the parameters of convolution are relatively large, which requires a large amount of memory at runtime. The running speed is slow, and a larger dataset is required to train the network. How to reduce the parameters of the network and improve the running speed is the research interest for the further study.

Data Availability

The data used to support the results of this study can be downloaded and found at http://www.cresda.com/CN/ and http://www.digitalglobe.com/product-samples.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Key Research and Development Program of China (Grant: 2018YFB1404400), Hainan Provincial Natural Science Foundation of China (Grant: 2019CXTD400), the Scientific Research Fund Project of Hainan University (Grant: KYQD(ZR)-21007, KYQD(ZR)-21008), and the Scientific Research Fund Project for Youth Teachers of Hainan University (Grant: HDQN202103).

References

[1] X. Yao, J. Han, G. Cheng, X. Qian, and L. Guo, "Semantic annotation of high-resolution satellite images via weakly supervised learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 6, pp. 3660–3671, 2016.

[2] Q. C. Zhang, L. Hu, and J. Gow, "Output feedback stabilization for mimo semi-linear stochastic systems with transient optimisation," *International Journal of Automation and Computing*, vol. 17, no. 1, pp. 83–95, 2019.

[3] Q. Zhang and H. Wang, "A novel data-based stochastic distribution control for non-gaussian stochastic systems," *IEEE transactions on automatic control*, vol. 99, p. 1, 2021.

[4] Y. Liu, L. Dang, K. Cai, S. Li, and X. Zuo, "Research progress on models, algorithms and systems for remote sensing spatial-temporal big data processing," *IEEE journal of selected topics in applied earth observations and remote sensing*, vol. 14, pp. 5918–5931, 2021.

[5] X. Qian, S. Lin, G. Cheng, X. Yao, and W. Wang, "Object detection in remote sensing images based on improved bounding box regression and multi-level features fusion," *Remote Sensing*, vol. 12, no. 1, p. 143, 2020.

[6] X. Meng, H. Shen, H. Li, L. Zhang, and R. Fu, "Review of the pansharpening methods for remote sensing images based on the idea of meta-analysis: practical discussion and challenges," *Information Fusion*, vol. 46, pp. 102–113, 2018.

[7] Y. Liu, Y. Xie, J. Yang, X. Zuo, and B. Zhou, "Target classification and recognition for high-resolution remote sensing images: using the parallel cross-modal neural cognitive computing algorithm," *IEEE Geoscience and Remote Sensing Magazine*, vol. 8, no. 3, pp. 50–62, 2020.

[8] L. Jing, L. Zhang, and W. Shuang, "Region of interest extraction based on saliency detection and contrast analysis for remote sensing images," in *Spie Remote Sensing*, September 2016.

[9] V. R. Pandit and R. J. Bhiwani, "Component substitution based fusion of worldview imagery," in *2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, Kanpur, India, 2019.

[10] F. Palsson, J. R. Sveinsson, and M. O. Ulfarsson, "Optimal component substitution and multi-resolution analysis pansharpening methods using a convolutional neural network," in *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*, Yokohama, Japan, 2019.

[11] T. M. Tu, P. S. Huang, C. L. Hung, and C. P. Chang, "A fast intensitychuecsaturation fusion technique with spectral adjustment for ikonos imagery," *IEEE Geoscience and Remote Sensing Letters*, vol. 1, no. 4, pp. 309–312, 2004.

[12] C. D. Souza, "A Tutorial on Principal Component Analysis with the accord.net Framework," *Computer Science*, 2012, http://arxiv.org/abs/1210.7463.

[13] N. Y. Zhang and W. U. Quan-Yuan, "Information inuence on quickbird images by Brovey fusion and wavelet fusion," *Remote Sensing Technology & Application*, vol. 21, no. 1, pp. 67–70, 2006.

[14] N. Tsukamoto, Y. Sugaya, and S. Omachi, "Pansharpening by complementing compressed sensing with spectral correction," *Applied Sciences*, vol. 10, no. 17, p. 5789, 2020.

[15] X. U. Hanqiu, *Classification of Fused Imagery Base on the s_m Algorithm*, Editorial Board of Geomatics & Information Science of Wuhan University, 2004.

[16] Y. Zhang, Y. Wang, L. Yang, C. Zhang, and S. Mei, "Hyperspectral and multispectral image fusion using cnmf with minimum endmember simplex volume and abundance sparsity constraints," in *Geoscience & Remote Sensing Symposium*, Milan, Italy, 2015.

[17] K. Ren, W. Sun, X. Meng, G. Yang, and Q. Du, "Fusing china gf-5 hyperspectral data with gf-1, gf-2 and sentinel-2a multispectral data: Which methods should be used?," *Remote Sensing*, vol. 12, no. 5, 2020.

[18] K. A. Althelaya, S. A. Mohammed, and E. El-Alfy, "Combining deep learning and multiresolution analysis for stock market forecasting," *IEEE Access*, vol. 99, p. 1, 2021.

[19] Y. Zhang and U. Nauman, "Deep learning trends driven by temes: a philosophical perspective," *IEEE Access*, vol. 8, pp. 196587–196599, 2020.

[20] Y. Liu, X. Chen, Z. Wang, Z. J. Wang, R. K. Ward, and X. Wang, "Deep learning for pixel-level image fusion: recent advances and future prospects," *Information Fusion*, vol. 42, pp. 158–173, 2018.

[21] L. Kheli and M. Mignotte, "Deep learning for change detection in remote sensing images: Comprehensive review and meta-analysis," *IEEE Access*, vol. 8, pp. 126385–126400, 2020.

[22] X. Cui, C. Zou, and Z. Wang, "Remote sensing image recognition based on dual-channel deep learning network," *Multimedia Tools and Applications*, vol. 7, pp. 1–17, 2021.

[23] Y. Chen, Y. Bai, W. Zhang, and T. Mei, "Destruction and construction learning for fine-grained image recognition," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, 2019.

[24] W. Wang, Y. Cui, G. Li, C. Jiang, and S. Deng, "A self-attention-based destruction and construction learning fine-grained image classification method for retail product recognition," *Neural Computing and Applications*, vol. 32, no. 12, pp. 14613–14622, 2020.

[25] M. Rout, S. Nahak, S. Priyadarshinee, P. Mohapatra, K. D. Sa, and D. Dash, "A deep learning approach for sari fusion," in *2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT)*, Kannur, India, 2020.

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

[27] Y. Wei, Q. Yuan, H. Shen, and L. Zhang, "Boosting the accuracy of multispectral image pansharpening by learning a deep residual network," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 10, pp. 1795–1799, 2017.

[28] J. Yang, X. Fu, Y. Hu, H. Yue, and J. Paisley, "Pannet: A deep network architecture for pansharpening," in *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, 2017.

[29] X. Liu, Q. Liu, and Y. Wang, "Remote sensing image fusion based on two-stream fusion network," *Information Fusion*, vol. 55, 2019.

[30] D. Christilin and D. Mary, "Residual encoder-decoder upsampling for structural preservation in noise removal," *Multimedia Tools and Applications*, vol. 80, no. 13, pp. 19441–19457, 2021.

[31] M. Riaz, H. Garg, H. M. A. Farid, and M. Aslam, "Novel q-rung orthopair fuzzy interaction aggregation operators and their application to low-carbon green supply chain management," *Journal of Intelligent and Fuzzy Systems*, vol. 1, pp. 1–18, 2021.

[32] X. Yao, X. Feng, J. Han, G. Cheng, and L. Guo, "Automatic weakly supervised object detection from high spatial resolution remote sensing images via dynamic curriculum learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 99, pp. 1–11, 2020.

[33] G. Cheng, C. Yang, X. Yao, L. Guo, and J. Han, "When deep learning meets metric learning: remote sensing image scene classi_cation via learning discriminative cnns," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 5, pp. 2811–2821, 2018.

[34] T. Zhou, S. Canu, and R. Su, "Automatic covid-19 ct segmentation using u-net integrated spatial and channel attention mechanism," *International Journal of Imaging Systems and Technology*, vol. 31, no. 3, pp. 16–27, 2020.

[35] H. Luo, C. Chen, L. Fang, X. Zhu, and L. Lu, "High-resolution aerial images semantic segmentation using deep fully convolutional network with channel attention mechanism," *IEEE journal of selected topics in applied earth observations and remote sensing*, vol. 99, pp. 1–17, 2019.

[36] B. Tolooshams, R. Giri, A. H. Song, U. Isik, and A. Krishnaswamy, "Channel-attention dense u-net for multichannel speech enhancement," 2020.

[37] Q. Liu, L. Han, R. Tan et al., "Hybrid Attention Based Residual Network for Pansharpening," *Remote Sensing*, vol. 13, no. 10, p. 1962, 2021.

[38] Y. Zheng, J. Li, Y. Li, J. Guo, X. Wu, and J. Chanussot, "Hyperspectral pansharpening using deep prior and dual attention residual network," *IEEE transactions on geoscience and remote sensing*, vol. 58, no. 11, pp. 8059–8076, 2020.

[39] W. Zhang, J. Li, and Z. Hua, "Attention based tri-unet for remote sensing image pan-sharpening," *IEEE journal of selected topics in applied earth observations and remote sensing*, vol. 14, pp. 3719–3732, 2021.

[40] Y. Wei and Q. Yuan, "Deep residual learning for remote sensed imagery pansharpening," in *International Workshop on Remote Sensing with Intelligent Processing*, Shanghai, China, 2017.

[41] Y. Wu, M. Huang, Y. Li, S. Feng, and D. Wu, "A distributed fusion framework of multispectral and panchromatic images based on residual network," *Remote Sensing*, vol. 13, no. 13, 2021.

[42] D. Lei, H. Chen, L. Zhang, and W. Li, "Nlrnet: an efficient nonlocal attention resnet for pansharpening," *IEEE transactions on geoscience and remote sensing*, vol. 99, pp. 1–3, 2021.

[43] M. Giuseppe, C. Davide, V. Luisa, and S. Giuseppe, "Pansharpening by convolutional neural networks," *Remote Sensing*, vol. 8, no. 7, p. 594, 2016.

[44] H. Jie, S. Li, S. Gang, and S. Albanie, "Squeeze-and-excitation networks," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 7132–7141, Salt Lake City, UT, 2017.

[45] X. Liu, Y. Wang, and Q. Liu, *Remote Sensing Image Fusion Based Ontwo-Stream Fusion Network*, Springer, Cham, 2018.

[46] Z. Kai, W. Zuo, Y. Chen, D. Meng, and Z. Lei, "Beyond a gaussian denoiser: residual learning of deep cnn for image denoising," *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3142–3155, 2016.

[47] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image Super-Resolution Using Very Deep Residual Channel Attention Networks," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 286–301, Munich, Germany, 2018.

[48] L. Wald and T. Ranchin, "The Arsis concept in image fusion: an answer to users needs," in *6th International Conference on Information Fusion*, pp. 181–184, Cairns, QLD, Australia, 2003.

[49] G. S. Reddy, R. Nanmaran, and G. Paramasivam, "Image restoration using Lucy Richardson algorithm for deblurring images with improved osnr, ssim, nc inc comparison with Wiener filter," *Journal of Contemporary Issues in Business and Government*, vol. 27, no. 4, p. 147, 2021.

[50] R. Bhatt, N. Naik, and V. K. Subramanian, "SSIM compliant modeling framework with denoising and deblurring applications," *IEEE transactions on image processing*, vol. 30, pp. 2611–2626, 2021.

[51] P. Li, L. Sang-Heon, H. Hung-Yao, and P. Jae-Sam, "Nonlinear fusion of multispectral citrus fruit image data with information contents," *Sensors*, vol. 17, no. 12, p. 142-, 2017.

[52] N. Konstantinos, "Dimitrios, and Oikonomidis, quality assessment of ten fusion techniques applied on worldview-2," *European Journal of Remote Sensing*, vol. 48, no. 1, pp. 141–167, 2017.

[53] K. Wang, G. Qi, Z. Zhu, and C. Yi, "A novel geometric dictionary construction approach for sparse representation based image fusion," *Entropy*, vol. 19, no. 7, p. 306, 2017.

[54] Y. Zhou, Y. Wang, Y. Kong, and M. Hu, "Multi-indicator image quality assessment of smartphone camera based on human subjective behavior and perception," in *2020 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, ELECTR Network, 2020.

[55] W. Xue, Z. Zhang, and S. Chen, "Ghost elimination via multicomponent collaboration for unmanned aerial vehicle remote sensing image stitching," *Remote Sensing*, vol. 13, no. 7, p. 1388, 2021.