Research Article

Remote Sensing Sea Ice Image Classification Based on Multilevel Feature Fusion and Residual Network

Yanling Han,[1,2] **Pengxia Cui** (ID),[1] **Yun Zhang** (ID),[1] **Ruyan Zhou,**[1,2] **Shuhu Yang,**[1,2] **and Jing Wang**[1,2]

[1]*Shanghai Marine Intelligent Information and Navigation Remote Sensing Engineering Technology Research Center, Shanghai Ocean University, Shanghai, China*
[2]*Key Laboratory of Fisheries Information, Ministry of Agriculture, Shanghai Ocean University, Shanghai, China*

Correspondence should be addressed to Yun Zhang; y-zhang@shou.edu.cn

Sea ice disasters are already one of the most serious marine disasters in the Bohai Sea region of our country, which have seriously affected the coastal economic development and residents' lives. Sea ice classification is an important part of sea ice detection. Hyperspectral imagery and multispectral imagery contain rich spectral information and spatial information and provide important data support for sea ice classification. At present, most sea ice classification methods mainly focus on shallow learning based on spectral features, and the good performance of the deep learning method in remote sensing image classification provides a new idea for sea ice classification. However, the level of deep learning is limited due to the influence of input size in sea ice image classification, and the deep features in the image cannot be fully mined, which affects the further improvement of sea ice classification accuracy. Therefore, this paper proposes an image classification method based on multilevel feature fusion using residual network. First, the PCA method is used to extract the first principal component of the original image, and the residual network is used to deepen the number of network layers. The FPN, PAN, and SPP modules increase the mining between layer and layer features and merge the features between different layers to further improve the accuracy of sea ice classification. In order to verify the effectiveness of the method in this paper, sea ice classification experiments were performed on the hyperspectral image of Bohai Bay in 2008 and the multispectral image of Bohai Bay in 2020. The experimental results show that compared with the algorithm with fewer layers of deep learning network, the method proposed in this paper utilizes the idea of residual network to deepen the number of network layers and carries out multilevel feature fusion through FPN, PAN, and SPP modules, which effectively solves the problem of insufficient deep feature extraction and obtains better classification performance.

1. Introduction

Sea ice disasters are one of the marine disasters that should not be underestimated. They mostly occur in polar regions and mid-to-high dimensional regions. China's Bohai Bay is located in a mid-to-high dimensional area. The region has a developed economy and heavy maritime traffic. However, the Bohai Bay area has different degrees of icing every winter [1], which affects marine activities such as marine fisheries, marine oil and gas resource development, and other social practices. According to statistics, the economic loss caused by sea ice disasters has reached hundreds of millions, and the affected population has reached tens of thousands. Therefore, in order to avoid more economic losses and casualties, it is very necessary to detect sea ice in the Bohai Bay area [2].

Sea ice image classification is an important part of sea ice detection. The remote sensing images currently used for sea ice classification mainly include SAR images and optical images (multispectral images (such as Landsat, Sentinel, MODIS, and so on) and hyperspectral images). Among them, the SAR image contains rich feature information. Its imaging feature is that it can penetrate clouds and is less affected by the environment. However, due to sensor limitations, its data are relatively singular, which is not

conducive to the detailed classification of multiple types of sea ice. The optical image has a high spectral resolution, contains rich spectral and spatial information, can extract detailed features of different types of sea ice, and provides effective data support for accurate sea ice classification. The current sea ice classification methods mostly use traditional supervised classification methods, including SVM, decision tree, maximum likelihood, minimum distance, etc. For example, literature [3] is based on the SVM algorithm, which combines the backscatter coefficient, GLCM, and sea ice density to classify the sea ice image. Literature [4] improved the classification accuracy of sea ice images in Liaodong Bay by building decision trees and gray-level co-occurrence matrix (GLCM). Literature [5] used SVM and maximum likelihood method to conduct comparative experiments on Landsat images, in which SVM classification is closer to the actual situation. Literature [6] used the discriminant function to calculate the distance to the center, and the result shows that the minimum distance method can obtain higher classification accuracy when analyzing remote sensing images. The above supervised classification algorithms are all shallow models and cannot extract deep features in hyperspectral and multispectral images, which limits the further improvement of classification accuracy.

In recent years, as deep learning technology has continuously made important progress in the field of image classification, it has gradually gained attention in the field of remote sensing applications. Literature [7] enters the extracted texture features into the 3D-CNN model, makes full use of the spatial spectral characteristics of remote sensing sea ice images, and achieves high classification accuracy. Literature [8] used CNN and DBN for sea ice classification to evaluate the performance of different types of deep learning in SAR image sea ice classification and its influencing factors. Research shows that the use of deep learning can improve the classification accuracy to a certain extent, but in pixel-level classification, the input size of hyperspectral and multispectral images is small, the number of layers of deep learning models is limited, and the possibility of model optimization is less. In 2015, He et al. [9] proposed a deep residual network, which won the image classification and object recognition in the ImageNet large-scale visual recognition competition, alleviating the problem of gradient disappearance and solving the problem of insufficient network layers. Literature [10] proposed a model based on spectral-spatial residual network. The spectral residual module and the spatial residual module are designed to extract spectral features and spatial features, respectively, and achieve high classification accuracy on the standard data set, effectively alleviating the problem of network degradation and deepening the number of network layers. Literature [11] is based on embedding two modules of skip connection and covariance pooling in the residual network model, fusing different levels of feature information, and achieving high classification accuracy in the standard dataset. Literature [12] added top-down feature fusion on the basis of feature pyramid networks (FPNs) and improved the loss function. The improved mask R-CNN algorithm improved the classification accuracy of image recognition. Literature [13] proposed a RP-SSD (residual and pyramid SSD) algorithm based on residual network and improved feature pyramid, which significantly improves the detection performance of small targets. Literature [14] fused SPP net and PA net to improve the quality of algorithm feature fusion. The above literature shows that the combination of residual network, feature pyramid, and SPP net is an idea to improve the accuracy of image recognition in recent years. Therefore, this paper applies the combination of feature pyramid and residual network in the field of sea ice classification to improve classification accuracy.

Based on the above research, this paper proposes a multilevel feature fusion sea ice classification method based on residual network. We use the residual network to deepen the number of network layers and solve the problem of difficulty in extracting sea ice depth features caused by the small image input size in sea ice detection and the limited number of network layers. The idea of fusing FPN, PAN, and SPP is to perform multilevel fusion of different levels of features extracted from multiple residual blocks, fully mining different levels of depth features, fusing multiscale feature information, fully mining different scale sea ice features in the image, and further improving the accuracy of sea ice classification.

The rest of this article is organized as follows. Second 2 introduces the overall framework, theoretical methods, and algorithm ideas of this article in detail. Section 3 introduces the dataset and related experimental settings and discusses the experimental results and the influence of the experimental parameters on the results. The work of this paper is summarized in Section 4.

2. Theoretical Method

The overall framework of this article is shown in Figure 1. It is divided into three parts in total. The first part is the data preprocessing part, which mainly uses ENVI to make sample labels and stores them in the required format through MATLAB. The second part is the comparison between the algorithm framework of this paper and the three algorithms of SVM, CNN, and traditional residual network. The third part is the accuracy evaluation part, which mainly calculates the confusion matrix to obtain the overall classification accuracy and kappa coefficient. The main idea of the algorithm in this paper is to use the residual network to deepen the number of network layers, alleviate the accuracy drop caused by the excessive number of network layers, and use the three modules of FPN, PAN, and SPP to perform multilevel and multiscale fusion of the extracted features. We make full use of the deep features of the mining, distinguish the types of sea ice more effectively, and solve the problem of limited network layers due to the small input size in sea ice detection. First, perform principal component analysis on the original image, select its first principal component as input and input it into the convolutional layer, then add three residual blocks to deepen the number of network layers, use FPN and PAN to extract features of different layers from the three residual blocks, use multiscale fusion of the features extracted from the residual blocks
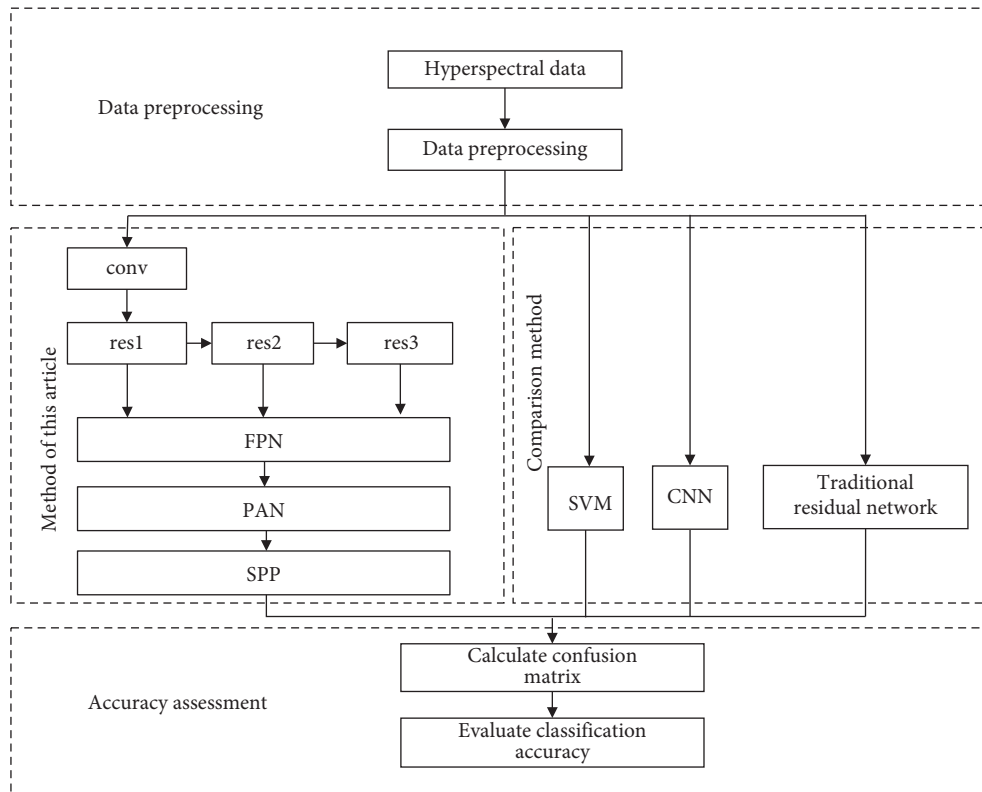
FIGURE 1: Overall framework of this article.

through upsampling and downsampling, and use the features between different layers in the residual block to improve the utilization of each layer to extract features. Finally, the SPP module is used to pool according to the size of three convolution kernels of different sizes, and the three features are spliced with the features before pooling, making full use of the features extracted from each layer. Therefore, by deepening the number of network layers and multiscale fusion, this paper fully excavates the deep features in the original image and improves the classification accuracy of the image.

2.1. Principle of Residual Network.
Network depth is a hot topic in deep learning. In theory, the more the network layers are, the better the effect can be achieved. However, the input size of the sea ice image is small because the convolution calculation is a process of multiplying matrix inner product and then adding all the new matrix values to obtain a value. Therefore, the input image will be reduced during the convolution process, which leads to the limitation of the number of network layers. So, this paper improves the traditional residual network to deepen the number of network layers. The main idea of the residual network is shown in Figure 2. The identity mapping prevents the error from increasing. This is the key to the residual network to solve the problem of the disappearance of the gradient caused by the increase in the number of network layers in deep learning. We pass the input x to the output as the initial result, and the output result is $Y = F(x) + x$, when $F(x) = 0$, $Y = x$, which is
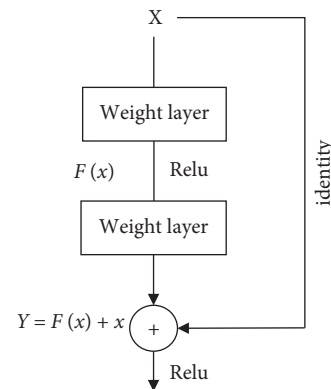

FIGURE 2: Residual network design ideas.

the identity mapping. Therefore, the target value of the residual network is the difference between Y and x, that is, $F(x) = Y - x$.

2.2. Improved Residual Network.
This paper adds FPN, PAN, and SPP modules based on the residual network. FPN is top-down. The high-level features are merged with the low-level features through upsampling to enhance the utilization of features. PAN is bottom-up. Using the fusion features of FPN, the lower-level features are passed up. The combination of FPN and PAN can fully exploit the features between different layers. The SPP module is proposed to solve the problem that the input of the fully connected layer

requires a fixed data size. Generally, the feature is mapped into several equal parts through maximum pooling. Finally, when the fully connected layer is input, the features are expanded into a one-dimensional matrix, which is spliced in the channel dimension, and the local features and the overall features are merged to enrich the expressive ability of the feature map. The combination of residual network with FPN, PAN, and SPP modules not only solves the problem that the number of network layers is limited by the input size but also makes full use of the features extracted from the residual block, thereby improving the classification accuracy.

Figure 3 shows the network structure diagram of the algorithm in this paper. Conv means the convolutional layer, resx means the residual block, and there are three residual blocks. Each residual block contains three convolutional layers and a pooling layer. Ups means upsampling, dws means downsampling, concat means splicing between features, and FC means fully connected layer. We first input the data into the convolutional layer to extract low-level features, input the features into the residual block, then upsample the features extracted from the third residual block, and stitch them with the features extracted from the second residual block. Then, the spliced feature is continuously upsampled and then spliced with the feature extracted from the first residual block to obtain a new feature, and the new feature is downsampled and spliced with the feature of the corresponding size. Finally, it is spliced with the third residual block and the new feature and then input into the SPP module; among them, the SPP module has three maximum pooling, and the most obtained three new features are stitched with the input SPP module to obtain a one-dimensional matrix and input to the fully connected layer, and finally the image is classified.

2.3. Algorithm Description. Based on the above algorithm analysis, Algorithm 1 described in this article is as follows.

3. Experimental Results and Analysis

3.1. Experimental Data Description. The Bohai Bay has always been an important area for studying sea ice conditions in our country. The first experimental data in this paper consist of a hyperspectral sea ice image, which was taken in the Bohai Sea on January 23, 2008. The image size selected in the experiment is 442×212, and 176 bands remain after excluding some of the disturbed bands, with a resolution of 30 m, as shown in Figure 4. The second experimental data consist of a multispectral image, downloaded from the official website of the European Space Agency (ESA) and taken in the Bohai Sea on February 11, 2018. The size of the image selected in the experiment is 400×400, the number of bands is 13, and the different resolutions in each band are processed to the same resolution of 10 m. The processed image is shown in Figure 5.

The two sets of experimental data in this paper are based on the combination of spectral curves and Google Maps, using pixel points as sample labels. The hyperspectral images are divided into three categories: white ice, sea water, and gray ice.

The sample labels are 2363, and the ratio of training samples to test samples is 1 : 9; the specific numbers are shown in Table 1. The multispectral image is divided into four categories: white ice, sea water, gray ice, and land. There are a total of 8025 sample labels, and the ratio of training samples to test samples is 1 : 9. The specific numbers are shown in Table 2. Figures 6 and 7 show the hyperspectral average spectral curve and the multispectral average spectral curve, respectively. The color of the curve corresponds to the label color of Figures 4 and 5. The vertical axis values of the two graphs are quite different, and the categories are easier to distinguish.

3.2. Preprocessing and Experimental Settings. The data preprocessing part uses the PCA algorithm to reduce the dimensionality of the image. The main features of different bands are concentrated in the first band and used as input. The training sample is random input, and the test sample is the total sample minus the training sample. The random input of the sample makes the accuracy of the network model not completely consistent each time, so all the experiments in this article are trained five times to ensure the stability of the experimental results, making the experimental results more convincing.

Deep learning contains a large number of training parameters, and different parameters have a certain impact on the experimental results. Therefore, some parameters are fixed in the experiment process of this article. The learning rate is 0.0005, the discard ratio is 0.5, the batch size is 20, and the number of iterations is 10,000. The other parameters and network settings of the algorithm in this paper are shown in Table 3. There are three comparison algorithms in this paper, namely, SVM [15], CNN [16], and traditional residual network [17]. The radial basis function used by SVM is used as the sum function. The parameter g and penalty factor c are all obtained after 5 times cross validation. Other parameters of the CNN are shown in Table 4, and other parameters of the traditional residual network are shown in Table 5.

3.3. Analysis of Results. Table 6 shows the classification results of Bohai Bay hyperspectral images based on different algorithms in 2008. The final classification results are obtained by training label samples. The algorithm in this paper has achieved the best classification results with an overall classification accuracy of 93.01% and a kappa coefficient of 88.87%. Table 7 shows the classification results of Bohai Bay multispectral images based on different algorithms in 2018. The algorithm in this paper has achieved the best classification results, with an overall classification accuracy of 90.41% and a kappa coefficient of 87.52%. In the experiment, SVM is traditional machine learning, and the extracted features are shallow features. The deep features in hyperspectral images are not fully utilized, so the classification accuracy is low. The CNN is limited in the number of network layers of the CNN model due to the small input sample size of the hyperspectral image and cannot fully extract the deep features of the hyperspectral image. Because the traditional residual network deepens the network level, it can further
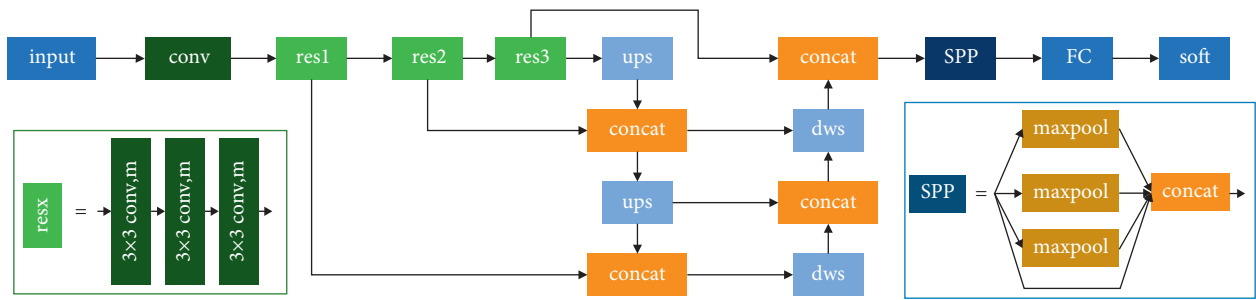
FIGURE 3: Improved residual network.

Start:
    Input: raw hyperspectral image
 (i) Sample label
 (1) Tag the original hyperspectral images to form a sample library;
 (2) Divide the sample library into training samples and test samples according to a certain proportion;
 (ii) Extract spatial features
 (3) Obtain the first principal component of the original image through the PCA algorithm as input;
 (iii) Improved residual network
    Training phase:
 (4) Randomly input the training samples into the first convolutional layer according to the iterative batch, and the size of the convolution kernel is $k \times k$;
 (5) Use the output in step (4) as the input in the first residual block. The residual block contains three residual units and a pooling layer. The size of the convolution kernel in each layer is $k \times k$. The number of convolution kernels is m;
 (6) Take the output of each layer as the input of the next residual block in turn, and there are three residual blocks in total;
 (7) After upsampling the output of the third residual block in step (6) to expand the feature size (the output result is called feature S), it is spliced with the output of the second residual block to obtain a new feature ST;
 (8) After upsampling feature ST to expand the feature size (the output result is called feature STR), it is spliced with the output of the first residual block to obtain a new feature SF;
 (9) After downsampling feature SF to reduce the feature size, splice with feature STR to obtain a new feature SFI;
 (10) After downsampling feature SFI to reduce the size, splice with feature SS to obtain a new feature SSI;
 (11) Input feature SSI into the SPP module, where the SPP module contains pooling of different convolution kernel sizes, and the step size is set to f;
 (12) Splicing the output of different pooling in SPP with feature SSI to obtain a new feature SSE;
 (13) Input feature SSSSSSS to the fully connected layer;
 (14) Iterate the model until convergence;
 (15) Model training completed;
    Testing phase:
 (16) Input the test sample into the trained model, calculate the confusion matrix, and get the classification accuracy;
 (17) The test is completed.
    Output: overall classification accuracy, kappa coefficient, and confusion matrix
    End

ALGORITHM 1:Improve residual network algorithm.

extract the deep features in the hyperspectral image and obtain higher accuracy, but because it only inputs the features extracted from each layer to the residual unit of the next layer, the features between different layers are not deeply fused and mined, which limits the further improvement of the classification accuracy of hyperspectral images. The method in this paper uses the residual network to increase the depth of the network and uses its shortcut connection to solve the problem of gradient disappearance caused by the deepening of the network; at the same time, the FPN and PAN modules are used to reuse the features between layers. Finally, the SPP module is used to generate the output of the same size,

which not only solves the problem of the limited number of network layers but also fully excavates the depth feature information in the hyperspectral image to further improve the classification accuracy, thus obtaining the highest classification accuracy.

In the experiments in this article, it can be seen from Figures 6 and 7 that the value of the spectral curve of seawater is the lowest among all categories, and the reflectance of sea ice is clearly distinguished from other categories of sea ice, and a higher classification accuracy is obtained. However, white ice and gray ice are divided according to the thickness of sea ice. As two different types of sea ice, it is classified between the same type of sea
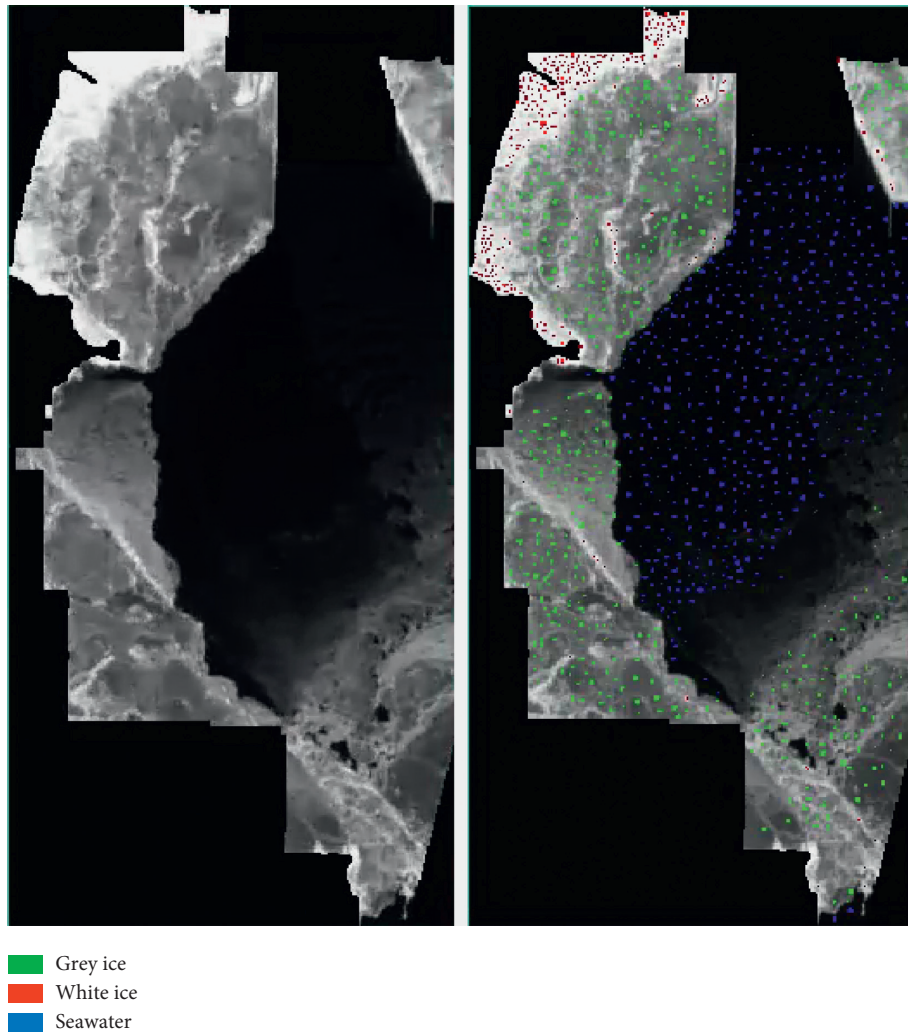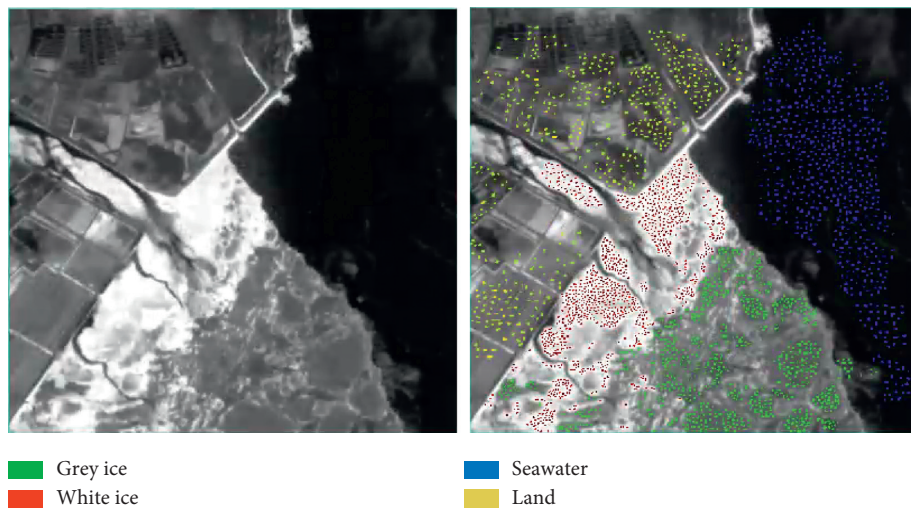
Grey ice
White ice
Seawater

Figure 4: Hyperspectral image.

Grey ice   Seawater
White ice   Land

Figure 5: Multispectral image.

TABLE 1: Number of hyperspectral image samples.

Category	Number of samples	10% training	90% test
White ice	307	30	277
Seawater	962	96	866
Gray ice	1094	109	985
Total	2363	235	2128

TABLE 2: Number of samples of multispectral image.

Category	Number of samples	10% training	90% test
White ice	2060	206	1854
Seawater	2201	220	1981
Gray ice	1840	184	1656
Land	1924	192	1732
Total	8025	802	7223
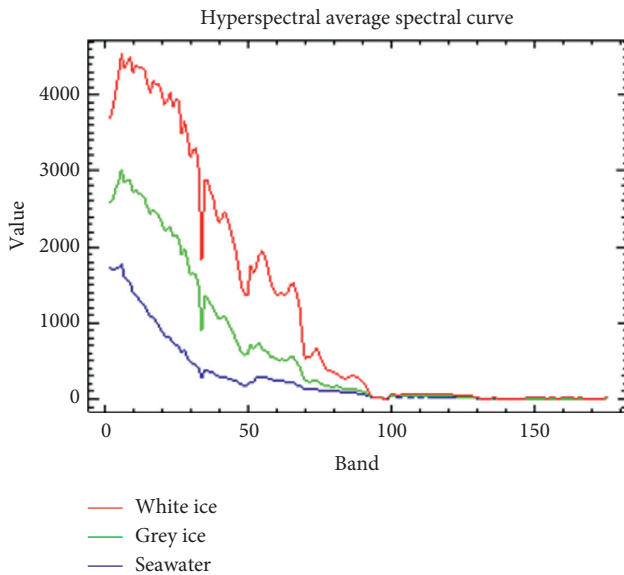
FIGURE 6: Average spectral curve of hyperspectral image.

FIGURE 7: Average spectral curve of multispectral image.

TABLE 3: Improved residual network parameters.

	Kernel_size_num	Stride	Activation
Conv1	8	1×1	ReLU
Conv2i	$\begin{bmatrix} 3 \times 3 & 8 \\ 3 \times 3 & 8 \\ 3 \times 3 & 8 \end{bmatrix}$	1×1	ReLU
Pool2	2×2	2×2	
Conv3i	$\begin{bmatrix} 3 \times 3 & 16 \\ 3 \times 3 & 16 \\ 3 \times 3 & 16 \end{bmatrix}$	1×1	ReLU
Pool3	2×2	2×2	—
Conv3	$\begin{bmatrix} 3 \times 3 & 32 \\ 3 \times 3 & 32 \\ 3 \times 3 & 32 \end{bmatrix}$	1×1	ReLU
Pool3	2×2	2×2	
FPN			
PAN			
SPP			

TABLE 4: CNN parameters.

	Kernel_size_num	Stride	Activation
Conv1	$4 \times 4,8$	1×1	ReLU
Pool1	2×2	2×2	—
Conv2	$5 \times 5,16$	1×1	ReLU
Pool2	2×2	2×2	—
Conv3	$4 \times 4,32$	1×1	ReLU

ice, more affected factors, and the misstorming is more serious. It can be seen from Tables 6 and 7 that the SVM method is shallow learning, and the classification accuracy of white ice and gray ice is low. The CNN can extract sea ice feature information in depth and obtain a certain improvement effect. The residual network further improves the classification accuracy of white ice and gray ice by deepening the network level. The method proposed in this paper uses the improved residual network to approach the gradient disappearance problem on the basis of deepening the network hierarchy and makes full use of the multilevel and multiscale features of remote sensing sea ice through FPN, PAN, and SPP modules to obtain the best classification effect. Compared with SVM, CNN, and traditional residual network methods, in hyperspectral images, the classification accuracy of white ice has increased by 1.47%, 3.47%, and 9.92%, respectively, and the classification 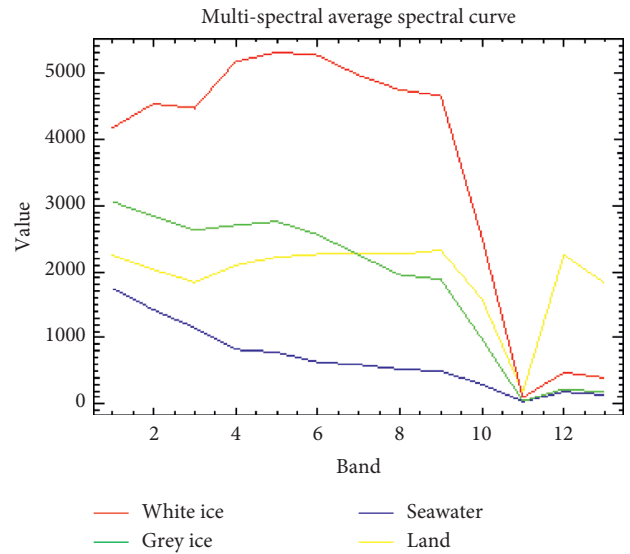accuracy of gray ice has increased by 11.06%, 11.47%, and 12.63%. In the multispectral image, the classification accuracy of white ice has increased by 8.18%, 8.36%, and 10.21%, and the classification accuracy of gray ice has increased by 6.07%, 7.86%, and 11.49%.

TABLE 5: Parameters of traditional residual network.

	Kernel_size_num	Stride	Activation
Conv1i	$\begin{bmatrix} 3 \times 3 & 8 \\ 3 \times 3 & 8 \\ 3 \times 3 & 8 \end{bmatrix}$	1×1	ReLU
Pool1	2×2	2×2	—
Conv2i	$\begin{bmatrix} 3 \times 3 & 16 \\ 3 \times 3 & 16 \\ 3 \times 3 & 16 \end{bmatrix}$	—	ReLU
Pool2	2×2	2×2	—
Conv3	$\begin{bmatrix} 3 \times 3 & 32 \\ 3 \times 3 & 32 \\ 3 \times 3 & 32 \end{bmatrix}$	—	ReLU
Pool3	2×2	2×2	

3.4. Influence of Parameters on Experimental Results

3.4.1. The Effect of Training Sample Size on Experimental Results. Taking into account the local characteristics of the distribution of sea ice categories, for each pixel, within a certain range, adjacent pixels in its spatial neighborhood belong to the same category with a high probability. Therefore, in this experiment, we take the pixel as the center and take the $M \times M$ neighborhood, and all the pixels in the neighborhood will form a data block of size $M \times M \times B$ (where $M \times M$ is the size and B is the number of bands). As the training sample of the pixel, the category of the pixel is the category of the training sample. The training sample size will affect the classification accuracy of sea ice; therefore, the selection of the training sample size takes into account the spatial information contained in the sample and the errors caused by it. The larger the size of the training sample, the relatively more the spatial information will be contained, which can increase the depth of the convolutional network and mine more feature information. However, because the surrounding samples may not belong to this category, it will also bring some errors. Taking the above factors into consideration, choosing an appropriate training sample size will obtain better classification results. Therefore, this section mainly discusses the comparative analysis experiments of sea ice classification with three sample sizes of 29×29, 27×27, and 25×25. It can be seen from Tables 8 and 9 that when the training sample size is 27×27, the overall classification accuracy and kappa coefficient are the highest. Therefore, the training sample size of 27×27 is selected for the two different images in this article.

3.4.2. The Influence of the Number of Convolution Kernels on Experimental Results. In deep learning, the more the convolution kernels are, the more the parameters will be and the more the features will be extracted, which will improve the classification accuracy to a certain extent. However, if the number of convolution kernels is too large, it will cause overfitting problems and affect the improvement of sea ice classification accuracy. In addition, the more convolutionary cores, the parameters of participating operations are

increased, and the computational complexity is higher, and the cost of time consumption is more, so choosing the appropriate number of convolution kernels is the focus of this section. In the experiment, we set the candidate value of the number of convolution kernels according to the empirical value (generally, the number of convolution kernels is 16, and a value of about 16 is taken as the candidate value), and through experimental analysis to determine the optimal number of convolution kernels, 4, 8, 16, and 32 convolution kernels were set in the experiment. In this section, the number of convolution kernels in the first layer of convolution is discussed. The number of remaining layers is increased by a factor of 2. The optimal parameters are selected for the hyperspectral image and multispectral image of Bohai Bay. It can be seen from Table 10 that when the convolution kernel is 8, the overall classification accuracy and kappa coefficient are the highest. Therefore, the number of convolution kernels for the first layer of hyperspectral image convolution is set to 8. It can be seen from Table 11 that when the convolution kernel is 16, the overall classification accuracy and kappa coefficient are the highest, so the number of convolution kernels for the first layer convolution of the multispectral image is set to 16.

3.4.3. The Effect of Convolution Kernel Size on Experimental Results. The size of the convolution kernel in deep learning is also an important parameter that affects accuracy. In general, the larger the convolution kernel, the larger the receptive field and the more the information contained in the receptive field. More features can be extracted to further improve the classification accuracy. However, a larger convolution kernel will increase the amount of calculation and reduce the depth of the model, which will affect the improvement of classification accuracy. This section discusses the influence of the convolution kernel size on the experimental results. Four convolution kernel sizes of 2×2, 3×3, 5×5, and 7×7 are selected for experiments. It can be seen from Tables 12 and 13 that the classification accuracy and kappa coefficient are the highest when the convolution kernel is 3×3. Therefore, the size of the convolution kernel in the hyperspectral image and multispectral image experiments in this article is set to 3×3.

4. Summary and Outlook

In order to fully mine the deep features in remotely sensed sea ice images, this paper proposes a multilevel feature fusion remote sensing sea ice image classification method based on residual network. The residual unit in the residual network is used for feature extraction, and the number of network layers is deepened through the principle of identity mapping, and the FPN, PAN, and SPP modules are used to fuse the features extracted from different residual blocks to fully excavate the multilevel and multiscale depth feature information in the remotely sensed sea ice data to further improve the accuracy of sea ice classification. The experimental results show that compared with other learning methods, this method

TABLE 6: Classification results of different algorithms for hyperspectral images.

Model	SVM	CNN	Traditional residual network	Method of this article
OA (%)	81.92	90.64	90.87	93.01
Kappa (%)	72.83	84.84	85.55	88.87
White ice	70.73	72.20	74.20	80.65
Seawater	84.38	96.38	96.81	97.69
Gray ice	80.36	91.42	91.83	92.99

TABLE 7: Classification results of different algorithms for multispectral images.

Model	SVM	CNN	Traditional residual network	Method of this article
OA (%)	79.47	87.38	87.83	90.41
Kappa (%)	74.22	83.53	84.28	87.52
White ice	77.25	85.43	86.36	87.46
Seawater	84.73	91.44	90.84	95.69
Gray ice	78.81	84.88	86.67	90.30
Land	83.42	88.43	88.74	88.33

TABLE 8: The effect of input sample size of hyperspectral image on experimental results.

Input sample size (%)	29×29	27×27	25×25
OA	92.16	93.01	91.99
Kappa	87.52	88.87	87.40

TABLE 9: The influence of the input sample size of the multispectral image on the experimental results.

Input sample size (%)	29×29	27×27	25×25
OA	89.33	90.41	87.00
Kappa	86.16	87.52	82.99

TABLE 13: The influence of the size of the multispectral image convolution kernel on the experimental results.

Kernel size (%)	2×2	3×3	5×5	7×7
OA	84.27	90.41	81.84	81.32
Kappa	79.44	87.52	76.36	75.89

TABLE 10: The influence of the number of hyperspectral image convolution kernels on the experimental results.

Number of kernels (%)	4	8	16	32
OA	91.26	93.01	92.92	92.41
Kappa	86.17	88.87	88.64	88.04

TABLE 11: The influence of the number of multispectral image convolution kernels on the experimental results.

Number of kernels (%)	4	8	16	32
OA	82.19	86.17	90.41	89.64
Kappa	76.89	82.05	87.52	86.59

TABLE 12: The effect of hyperspectral image convolution kernel size on experimental results.

Kernel size (%)	2×2	3×3	5×5	7×7
OA	91.72	93.01	91.27	91.55
Kappa	86.82	88.87	86.09	86.47

obtains the best classification results. The specific summary is as follows:

(1) Hyperspectral images and multispectral images contain rich spectral information and spatial information. Traditional machine learning can only extract shallow features and cannot make full use of the influential deep features, which affects the improvement of classification accuracy. However, deep learning technology has obtained better classification results due to its good deep feature extraction ability.

(2) The residual network can use its identity mapping characteristics to deepen the number of network layers and solve the problem of limited network layers in image classification due to the small input size of the sample. At the same time, the degradation problem caused by the deepening of the network layer is alleviated. Therefore, the residual network can further extract the features of the sea ice image and improve the classification accuracy.

(3) The method in this paper uses the advantages of residual network in deepening the number of network layers and alleviating the degradation

caused by too deep network layers. Secondly, the FPN and PAN modules are combined to connect low-level features and high-level features, which improves the entire feature level and shortens the information path between low-level and high-level features, fully excavates the features extracted by the residual network, and merges the features extracted by the residual block to different degrees to realize the deep feature complementarity between layers. Finally, the SPP module is used to fix the features into a one-dimensional vector and input to the fully connected layer to realize the fusion of local and global features, enrich the information of the final feature map, and further improve the accuracy of image classification.

Data Availability

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

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

Yanling Han and Yun Zhang conceived and designed the research framework. Pengxia Cui was responsible for data collection and processing. Yanling Han completed the algorithm design and data analysis and is the main author of the manuscript. Pengxia Cui, Yanling Han, and Shuhu Yang contributed to original draft preparation.

Acknowledgments

References

[1] W. Kuiqiao, X. Ying, and H. Yimeng, "Application of MODIS data in sea ice remote sensing," *Marine Forecast*, vol. S1, pp. 44–49, 2005.

[2] Z. Xi, Z. Jie, M. Junmin, S. Tengfei, and L. Meijie, "Research on polarization synthetic aperture radar sea ice classification method based on polarization scattering features: taking Bohai sea ice classification as an example," *Acta Oceanologica Sinica*, vol. 35, no. 5, pp. 95–101, 2013.

[3] H. Liu, H. Guo, and L. Zhang, "SVM-based sea ice classification using textural features and concentration from RADARSAT-2 dual-pol ScanSAR data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 4, pp. 1601–1613, 2015.

[4] W. Shuli, W. Zhiyong, and W. Lei, "Study on classification of sea ice in Liaodong Bay based on landsat-8 and sentinel-1A," *Beijing Surveying and Mapping*, vol. 33, no. 12, pp. 1486–1492, 2019.

[5] W. Mingwei, W. Zhiping, Z. Chunxia, M. Yue, Z. Kai, and Y. Fanlin, "Accuracy evaluation and comparison of remote sensing image classification based on maximum likelihood and support vector machine method," *Journal of Shandong University of Science and Technology*, vol. 35, no. 3, pp. 25–32, 2016.

[6] D. C. Feng, G. Chen, K. L. Xiao, W. Y. Du, and X. Y. Wu, "Remote sensing image classification based on minimum distance method," *Journal of North China Institute of Aeronautics and Astronautics*, vol. 22, no. 3, pp. 1-2, 2012.

[7] H. Yanling, G. Yi, Z. Yun, J. Wang, and S. Yang, "Hyperspectral sea ice image classification based on the spectral-spatial-joint feature with deep learning," *Remote Sensing*, vol. 11, no. 18, 2019.

[8] H. Dongmei, L. Minghui, S. Wei, and W. Jian, "Performance evaluation of convolutional neural network and deep belief network in SAR image ice water classification," *Journal of Image and Graphics*, vol. 23, no. 11, pp. 1720–1732, 2018.

[9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the 2016 IEEE Conference on Computer Vision And Pattern Recognition (CVPR)*, vol. 6, pp. 770–778, Las Vegas, NV, USA, June 2016.

[10] W. Chuntao, X. Bolin, L. Qianqian, B. Feng, and L. Zhihao, "Hyperspectral remote sensing image classification based on spectral-spatial residual network model," *Geographic Information World*, vol. 27, no. 3, pp. 42–48, 2020.

[11] W. Peng, L. Rui, X. Xuejing, and L. Peidong, "Optical remote sensing image scene classification algorithm based on residual network," *Laser and Optoelectronics Progress*, vol. 50, no. 2, pp. 1–19, 2020.

[12] Y. Lichun and L. Jinqing, "A flame image recognition algorithm based on improved mask R-CNN," *Computer Engineering and Applications*, vol. 56, no. 21, pp. 194–198, 2020.

[13] L. Hong, L. Yang, S. Mingwen, L. Chuanxiu, and Z. Zhaolei, "Oilfield job site target detection algorithm based on residual network and improved feature pyramid," *Science Technology and Engineering*, vol. 20, no. 11, pp. 4442–4450, 2020.

[14] H. Linquan, J. Liangwei, and G. Xiaofeng, "Improved YOLOv3 real-time video helmet wearing detection algorithm," *Modern Computer*, vol. 30, pp. 32–38, 2020.

[15] Z. Jun and Y. Huizhong, "Semi-supervised soft-sensing method of phenol concentration based on SVM," *Computers and Applied Chemistry*, vol. 30, no. 12, pp. 1453–1456, 2013.

[16] O. Xianfeng, X. Canqun, Z. Xiyang, P. Xin, and S. Yunsuo, "CNN-based digital character recognition algorithm for license plate," *Journal of Chengdu University of Technology*, vol. 19, no. 4, pp. 26–30, 2016.

[17] B. Haoyang, H. Ruikang, M. Kexin, and G. Ce, "Face age estimation based on residual network," *Computer Knowledge and Technology*, vol. 13, no. 14, pp. 169-170, 2017.