Research Article

Combining 3D-CNN and Squeeze-and-Excitation Networks for Remote Sensing Sea Ice Image Classification

Yanling Han ⓘ,[1] Cong Wei ⓘ,[1] Ruyan Zhou ⓘ,[1,2] Zhonghua Hong ⓘ,[1] Yun Zhang,[1,2] and Shuhu Yang ⓘ[1]

[1]*College of Information, Shanghai Ocean University, Shanghai 201306, China*
[2]*Key Laboratory of Fisheries Information, Ministry of Agriculture, Shanghai, China*

Correspondence should be addressed to Ruyan Zhou; ryzhou@shou.edu.cn

Sea ice is one of the most prominent marine disasters in high latitudes. Remote sensing technology provides an effective means for sea ice detection. Remote sensing sea ice images contain rich spectral and spatial information. However, most traditional methods only focus on spectral information or spatial information, and do not excavate the feature of spectral and spatial simultaneously in remote sensing sea ice images classification. At the same time, the complex correlation characteristics among spectra and small sample problem in sea ice classification also limit the improvement of sea ice classification accuracy. For this issue, this paper proposes a new remote sensing sea ice image classification method based on squeeze-and-excitation (SE) network, convolutional neural network (CNN), and support vector machines (SVMs). The proposed method designs 3D-CNN deep network so as to fully exploit the spatial-spectrum features of remote sensing sea ice images and integrates SE-Block into 3D-CNN in-depth network in order to distinguish the contributions of different spectra to sea ice classification. According to the different contributions of spectral features, the weight of each spectral feature is optimized by fusing SE-Block in order to further enhance the sample quality. Finally, information-rich and representative samples are chosen by combining the idea of active learning and input into SVM classifier, and this achieves superior classification accuracy of remote sensing sea ice images with small samples. In order to verify the effectiveness of the proposed method, we conducted experiments on three different data from Baffin Bay, Bohai Bay, and Liaodong Bay. The experimental results show that compared with other classical classification methods, the proposed method comprehensively considers the correlation among spectral features and the small samples problems and deeply excavates the spatial-spectrum characteristics of sea ice and achieves better classification performance, which can be effectively applied to remote sensing sea ice image classification.

1. Introduction

Sea ice is one of the most prominent marine disasters in the polar and mid- and high-latitude regions [1]. Freezing, melting, and drifting of sea ice have major impact on production operations in coastal areas and on the sea [2]. Therefore, in order to quickly and accurately assess sea ice conditions, timely forecast sea ice disasters, and ensure the safety of personal and property, research on sea ice detection has important significance [3], and sea ice classification is an important part of sea ice detection [4].

Remote sensing technology can acquire large-scale data rapidly and efficiently [5], and it provides a kind of effective mean for sea ice detection. At present, remote sensing technology has been widely used in sea ice detection. During recent years, the data sources commonly used include synthetic aperture radar [6], multispectral satellite images with medium or high-spatial resolution (e.g., MODIS and Landsat), and hyperspectral images [7–9]. Especially for multispectral and hyperspectral remote sensing data, they have the advantages of wide coverage, high resolution, rich spectral information and spatial information, multiple data sources, and low data cost, which provides abundant data support for sea ice detection [10]. However, remote sensing images contain tens to hundreds of bands, and there is a strong correlation between the spectral bands. In order to

achieve accurate classification result of sea ice, it is necessary to distinguish the differences between different spectral bands and measure their contributions to sea ice classification. At the same time, because of the particularity of environmental condition, it is difficult to obtain the labeled samples of sea ice, which also limit the improvement of sea ice classification accuracy. Therefore, these problems bring enormous challenge for remote sensing sea ice images classification.

Traditional remote sensing image classification methods include the maximum likelihood method, minimum distance method, K-means clustering method, etc., but these methods based on spectral statistical features have relatively lower accuracy. Therefore, researchers apply machine learning algorithms such as neural network (NN) and SVM to classify remote sensing images. Studies have shown that the remote sensing image classification method based on machine learning algorithms can achieve better classification results than traditional statistical methods [11, 12]. In particular, the SVM method has a good performance in solving small sample, high dimensional, and nonlinear classification problems, so it has been widely used. However, both SVM and NN belong to the shallow learning algorithm. Due to the non-homomorphic spatial structure and spectral information with high correlations contained in remote sensing sea ice image [13, 14], it is difficult to extract the deep features of remote sensing images effectively and achieve higher classification accuracy by using the shallow learning method with limited computational units [15].

Compared with shallow learning methods, deep learning methods have better expressive ability and can automatically extract deep hidden features, thus avoiding complex manual feature extraction processes [16–19]. CNN is a specially designed deep learning structure, and it is widely used in image recognition and image classification considering interpixel spatial correlation [20–23]. Therefore, remote sensing image classification based on CNN has attracted special research interest [24]. Liu et al. used the Siamese convolution network to classify remote sensing images and achieved better classification results [25]. Chen et al. proposed a 3D-CNN model which utilized local hyperspectral data cubes as input to excavate spatial and spectral information. Zhao and Du developed a local patch-based CNN spatial feature extraction architecture [17, 23, 25–27]. However, most of these methods improve network performance through spatial information, without considering the contribution difference of different spectra for classification result. Momenta proposed a squeeze-and-excitation network (SENet) structure in 2017; the core idea is that the network learns the feature weights through loss function; the weights of the effective feature maps are significant, and those of the ineffective feature maps are small in order to train the model to achieve better results. Experiments show that the SE-Block structure can be embedded in other network structures and can achieve superior results [28].

Based on the above research, this paper proposes a new remote sensing sea ice image classification method which integrates squeeze-and-excitation (SE) network, convolutional neural network (CNN), and SVM classifier (SE-CNN-SVM).

The proposed method uses 3D-CNN network to fully exploit the spectral and spatial characteristics of sea ice and distinguishes the contributions of different sea ice spectral features by combining SE-Block, that is, increasing the weight of the effective features, suppressing or reducing the weight of the invalid or ineffective features to further enhance the quality of samples. Finally, the rich and representative samples are chosen by combining the idea of active learning and input into SVM classifier to achieve high classification performance of sea ice remote sensing image in the case of small number of samples. The contributions of this paper are as follows: (1) this paper proposes a new method SE-CNN-SVM for remote sensing sea ice image classification. This method designs and constructs the 3D-CNN model, which can simultaneously extract the spatial information and spectral information of sea ice images and fully exploit the spatial-spectrum characteristics of sea ice hidden in remote sensing data. (2) Due to the high correlation among multiple spectral channels in remote sensing sea ice data, and the different channels have different degrees of discrimination for sea ice classification, the proposed method combines 3D-CNN network with the SE-Block to distinguish the different contributions of different spectral features and weight the spectral channels in order to improve the sample quality further in sea ice classification. (3) Because SVM has obvious advantages in solving small samples and high dimensional nonlinear problems, the proposed method extracts the spatial spectral feature and weight based on the 3D-CNN fusing SE-Block, combining the active learning method to choose the rich and representative samples and input into SVM classifier for classification, which achieves superior sea ice classification performance in the case of small number of samples.

The remainder of this paper is organized as follows. In Section 2, we introduce the overall framework and relative technology. Then, the proposed method is described in Section 3 in detail. Experiments and results are demonstrated in Section 4. Finally, Section 5 concludes the paper.

2. The Framework for Sea Ice Image Classification

In this section, we illustrate the design of 3D-CNN and SE-CNN model for sea ice image classification. Using these methods, we extract deep spectral-spatial features from multi or hyperspectral data and feed them into SVM classifier for classification. Figure 1 shows the entire framework for sea ice images classification. Four major issues were investigated: (1) data preprocessing, (2) building the SE-CNN network, (3) sea ice image classification based on SVM, and (4) classification accuracy evaluation. The methods are discussed in detail in the following sections.

2.1. 3D-CNN. In recent years, CNN has achieved great success in the fields of image recognition and target detection. It can automatically extract features that are effective for classification results from images, thus avoiding the process of manually designing and extracting features. Due
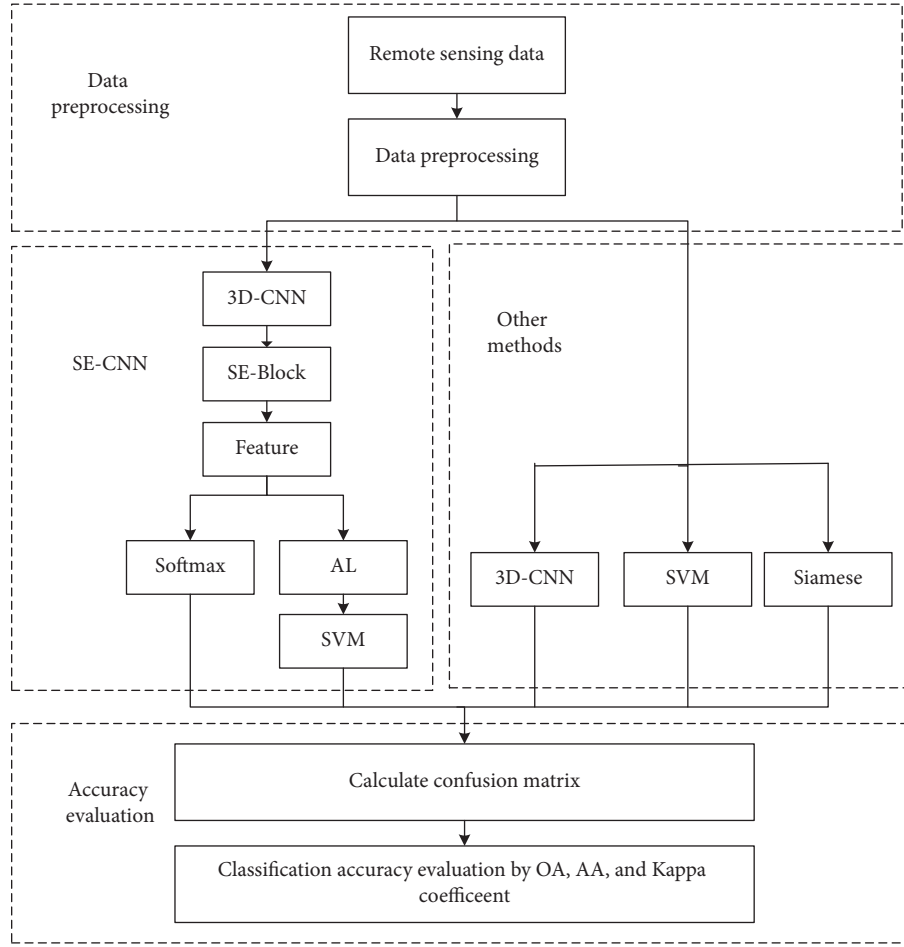
FIGURE 1: General framework for remote sensing sea ice image.

to the special three-dimensional structure of the multi/hyperspectral remote sensing image, part of the information in the image will be lost if classification is performed using spectral feature-based 1D-CNN and spatial feature-based 2D-CNN methods. Therefore, this paper uses the 3D-CNN structure and performs convolution operations through 3D convolution kernels and extracts simultaneously spatial and spectral features.

2.1.1. 3D-CNN Structure. The CNN is generally composed of a convolutional layer, a pooled layer, a fully connected layer, and a softmax classification layer as shown in Figure 2. The convolutional layer performs nonlinear feature extraction on images by using an activation function; the fully connected layer integrates the extracted features and then obtains the probability value of each type of tag through the softmax function, thereby predicting the label of the image. Before the final prediction value is obtained, the network hidden layer minimizes the error between the predicted value and the true value through the loss function and finally determines the classification performance of the model. This paper uses the Adam optimizer to update and calculate the network parameters that affect model training and model output

and to approximate or reach the optimal value in order to minimize the loss function.

In 3D-CNN, the value of the neuron at (x, y, z) is

$$v_{ij}^{xyz} = g\left(\sum_m \sum_{p=0}^{P_i-1} \sum_{q=0}^{Q_i-1} \sum_{r=0}^{R_i-1} w_{ijm}^{pqr} v_{(i-1)m}^{(x+p)(y+q)(z+r)} + b_{ij} \right), \quad (1)$$

where i denotes ith layer of neurons and j denotes a jth feature map; P_i and Q_i represent the height and width of the convolution kernel; R_i is the dimension of the convolution kernel along the spectral dimension; m represents the number of features connected to the previous layer, related to the feature dimension of each layer; w_{xyz}^{pqr} is the weight of the (p, q, r)th neuron connected to the mth feature; b_{ij} is the deviation of the jth feature map on the ith layer neuron; and g is the activation function; the activation function used in this paper is the ReLU function. ReLU function is represented by the following formula:

$$g(x) = \mathrm{ReLU}(x) = \begin{cases} x, & \text{if } x > 0, \\ 0, & \text{if } x \le 0. \end{cases} \quad (2)$$

2.1.2. Optimizer. The optimizer is used to update and calculate network parameters that affect model training and
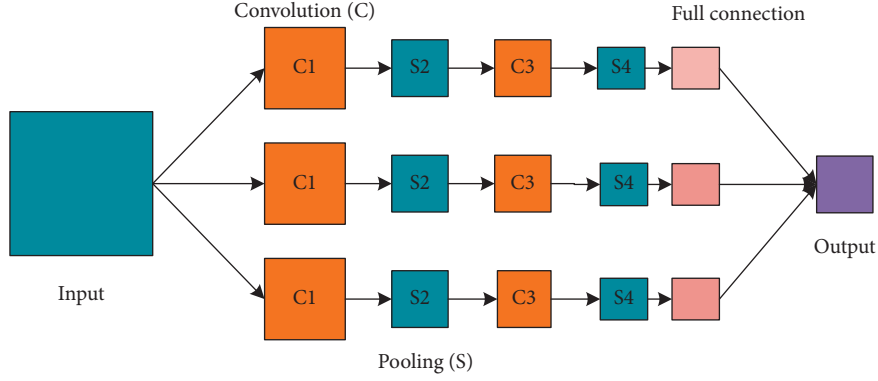
FIGURE 2: CNN model consisting of convolution layers, pooling layers, and full connection layer.

model output and to approximate or reach optimal values, thereby minimizing the loss function. Among them, the Adam optimizer combines the advantages of AdaGrad and RMSProp optimizer and has the characteristics of high efficiency, convenient implementation, and parameter updation without gradient transformation. Therefore, this paper uses the Adam optimizer, and Adam's update rules are as follows.

First, calculate the exponential moving average of the gradient; m_0 is initialized to 0.

$$g_t = \nabla_\theta J(\theta_{t-1}), \tag{3}$$

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t. \tag{4}$$

Second, calculate the exponential moving average of the square of the gradient; v_0 is initialized to 0.

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2. \tag{5}$$

Third, the deviation correction is performed on the gradient mean m_t and the gradient variance v_t.

$$\widehat{m}_t = \frac{m_t}{(1 - \beta_1^t)}, \tag{6}$$

$$\widehat{v}_t = \frac{v_t}{(1 - \beta_2^t)}. \tag{7}$$

Fourth, update the parameters; the initial learning rate α is multiplied by the ratio of the gradient mean to the square root of the gradient variance.

$$\Theta_t = \theta_{t-1} - \alpha^* \frac{\widehat{m}_t}{\sqrt{\widehat{v}_t} + \varepsilon}. \tag{8}$$

In equations (4)–(8), β_1 represents the exponential decay rate, which controls the weight assignment, usually taking a value close to 1, with a default of 0.9; β_2 represents the exponential decay rate, which weights the mean of the gradient squares, with a default of 0.99; $\varepsilon = 10^\wedge - 8$, which prevents the denominator from being 0.

2.2. Squeeze-and-Excitation Net.
The remote sensing sea ice image contains rich spatial information and spectral information. Different spectral features are suitable for

distinguishing different types of sea ice, and there are strong correlations among spectral dimensions. SENet can automatically gain the importance of each feature channel by learning. And according to this importance, the features effective for the classification are improved and those less effective for the classification are suppressed. The weighted feature can effectively improve the classification performance of remote sensing sea ice image. The implementation of SE-Block is shown in the following formula:

$$U = F_{tr}(X),$$
$$X \in R^{H' \times W' \times C'}, \tag{9}$$
$$U \in R^{H \times W \times C},$$

$$u_c = v_c * X = \sum_{s=1}^{C'} v_c^s * X^s, \tag{10}$$

where X indicates the input sample. For simplicity, in the flowing notation, we take F_{tr} to be a convolutional operator. Let $V = [v_1, v_2, \ldots, v_c]$ denote the learned set of filter kernels, where v_c refers to the parameters of the c-th filter. The output is $U = [u_1, u_2, \ldots, u_c]$; X^s represents the sth input. U represents the characteristics obtained after convolution. In the SE-Block structure, squeeze and excitation are two very important operations. A diagram illustrating the structure of an SE block is shown in Figure 3.

2.2.1. Squeeze Operation.
Squeeze operations are implemented through global average pooling, which is used to obtain dependencies between channels. We perform feature compression along the spatial dimension, turning each two-dimensional feature channel into a real number. This real number has a global receptive field to some extent, and the output dimension matches the number of input feature channels. It characterizes the global distribution of responses on feature channels and allows layers close to the input to obtain global receptive fields, which is very useful in many tasks. We opt for the simplest aggregation technique, global average pooling, which can be realized by formula (11). Formula (11) converts the input of H^*W^*C into the output of 1^*1^*C, that is, the F_{sq} operation in Figure 3. The result is
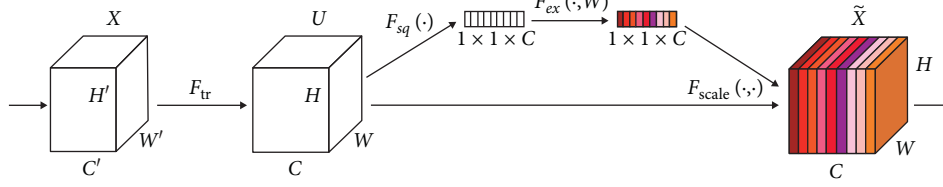
FIGURE 3: A squeeze-and-excitation block.

equivalent to indicating the numerical distribution of the C feature maps of the layer.

$$z_c = F_{sq}(u_c) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} u_c(i, j). \tag{11}$$

2.2.2. Excitation Operation. Excitation operation is used to generate weights for each feature channel by parameters. And the parameters are learned to explicitly model the correlation between feature channels. To make use of the information aggregated in the squeeze operation, we follow it with a second operation which aims to fully capture channelwise dependencies. To fulfil this objective, the function must meet two criteria: first, it must be flexible (in particular, it must be capable of learning a nonlinear interaction among channels) and second, it must learn a non-mutually-exclusive relationship since we would like to ensure that multiple channels are allowed to be emphasized (rather than enforcing a one-hot activation). To meet these criteria, we opt to employ a simple gating mechanism with a sigmoid activation:

$$z_c = F_{ex}(z, W) = \sigma(g(z, W)) = \sigma(W_2 \delta(W_1 z)), \tag{12}$$

where σ refers to the sigmoid function, δ refers to the ReLU function, $W_1 \in R^{(C/r) \times C}$, and $W_2 \in R^{C \times (C/r)}$. The final output of the block is obtained by rescaling the transformation output U with the activations:

$$\tilde{X}_c = F_{scale}(u_c, s_c) = s_c \cdot u_c, \tag{13}$$

where $\tilde{X} = [\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_c]$ and $F_{scale}(u_c, s_c)$ refers to channelwise multiplication between the scalar s_c and the feature map $u_c \in R^{H \times W}$.

2.3. SVM. Support vector machine (SVM) is a machine learning algorithm based on statistical learning theory, which adopts structural risk minimization criteria, and it minimizes the sample error while reducing the upper bound of the model generalization error, thereby improving the generalization ability of the model. The main idea of SVM is to use the kernel transform to transform the linear indivisible problem of low-dimensional space into high-order space for accurate classification. It is widely used in all aspects of remote sensing data processing. The basic mathematical form of SVM is shown in the following formula [29]:

$$\min_{w,b} \Phi(w, b) = \frac{1}{2}(w \cdot w). \tag{14}$$

Restrictions:

$$y_i[(w \cdot x_i) + b] \geq 1. \tag{15}$$

Introduce the Lagrange multiplier α_i to solve the equation as

$$\min_{w,b,a} : L(w, b, \alpha) = \frac{1}{2}(w.w) - \sum_{i=1}^{k} \alpha_i \{ y_i[(w.x_i) + b] - 1. \tag{16}$$

Seek partial deviations for w and b; then, get

$$\sum_{i=1}^{k} \alpha_i y_i = 0,$$

$$w = \sum_{i=1}^{k} \alpha_i y_i x_i. \tag{17}$$

Substitute formula (17) into (16) and get

$$W(\alpha) = \sum_{i=1}^{k} \alpha_i - \frac{1}{2} \sum_{i=1}^{k} \sum_{j=1}^{k} \alpha_i \alpha_j y_i y_j (x_i \cdot x_j). \tag{18}$$

Through the duality theorem, we can find the optimal solution of $w = \sum_{i=1}^{k} \alpha_i^* y_i^* x_i$, and taking any $\alpha_i \neq 0$, you can find b.

For high-dimensional space, if the inner product $K(x, x')$ is used instead of the dot product in the optimal classification plane, it is equivalent to mapping the original feature space to a high feature space. The optimization function at this time is

$$\max w(\alpha) = \sum_{i=1}^{k} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j K(x_i \cdot x_j). \tag{19}$$

Restrictions: $\sum_{i=1}^{n} y_i \alpha_i = 0, \alpha_i \geq 0$, where α_i is the Lagrange multiplier corresponding to formula (16). The optimal classification function obtained after solving the above problem is

$$f(x) = \text{sgn}[(w.x) + b] = \text{sgn}\left[\sum_{i=1}^{n} \alpha_i y_i K(x_i, x) + b\right], \tag{20}$$

where the kernel function $K(x_i \cdot x)$ in this paper chooses the RBF kernel function, that is,

$$K(x_i, x) = \exp\left[-\frac{\|x - x_i\|^2}{\sigma^2}\right]. \tag{21}$$

3. Proposed Sea Ice Image Classification Algorithm Combining 3D-CNN and Squeeze-and-Excitation Networks

3.1. Algorithm Framework. The implementation framework of the SE-CNN-SVM method in this paper is shown in Figure 4. The method comprises three modules: 3D-CNN module, SE-Block, and SVM classification module.

3.1.1. 3D-CNN Module. First, the original remote sensing sea ice data are preprocessed and the sample library is obtained. Then, the sample library is divided into training samples and test samples according to different strategies. Consequently, the 3D-CNN network model is established, the related network parameters are determined, and the training samples are input into the established CNN network for model training.

3.1.2. SE-Block. The SE-Block module includes two operations, squeeze and excitation. The squeeze operation is a global average pooling of the features obtained from the convolution of the last layer in the CNN; the feature maps after the squeeze are reduced by a fully connected layer and are nonlinearized by the ReLU activation function, hereafter upgraded through the fully connected layer, and then weight is activated by sigmoid, that is, the excitation operation. The weight of the excitation output represents the importance of each feature channel after feature selection, and then multiply the previous features by the resulting weights, and the recalibration of the original features in the channel dimension is completed. Finally, the rescaled features are converted into one-dimensional vector and input to the fully connected layer, and the weights of the parameters in the network are updated by the loss function, and the network optimization is finished.

3.1.3. Classification Module. Firstly, the sample features obtained by the previous module are normalized, and the sea ice classification is performed by the SVM classifier. Aiming at the small samples problem caused by the difficulty in obtaining sea ice samples, the module incorporates the idea of active learning ideas and uses a combination of uncertainty and differential strategies for sampling. And this method is a combination of best versus second-best (BVSB) and enhanced clustering-based diversity (ECBD) [30]. By this method, more representative samples which are suitable for sea ice classification are chosen and input into classifier, and superior classification results are obtained with fewer labeled samples.

3.2. Algorithm Implementation. The description of Algorithm 1 is shown below.

4. Experiments and Results

The experimental results are generated on a personal computer equipped with Intel Core i5-4590 with 3.30 GHz
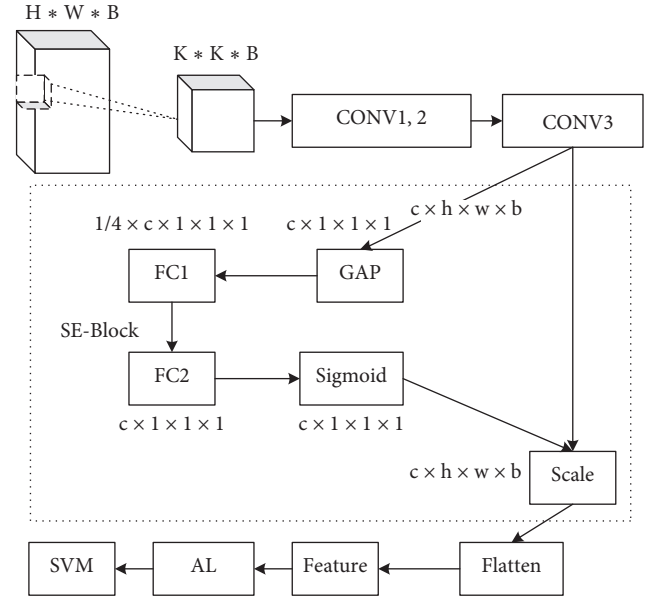

FIGURE 4: Flowchart of the proposed combination of SE-CNN-SVM.

and Nvidia GeForce GT 705. The personal computer's memory is 16G. These proposed methods are implemented by TensorFlow library. The SVM classifier is implemented by Lib-SVM library.

4.1. Data Description. To verify the performance method SE-CNN-SVM for sea ice remote sensing image classification, we utilized three remote sensing datasets in our experiments: Baffin Bay image captured by Earth Observing-1 (EO-1) satellite, Bohai Bay image capture by EO-1 satellite, and Liaodong Bay image captured by Landsat-8 satellite.

The first dataset utilized in our experiments is the Baffin Bay images, which are hyperspectral images acquired on a marine area of Baffin Bay in northwest Greenland on April 12, 2014. The original image dataset has a spatial resolution of 30 m. It comprises 2395×1769 pixels (which includes background pixels) and has 242 bands, among which 176 bands are used for the analysis after removing the bands with low signal-to-noise and water absorption. In the experiment, according to the spectral curve and the reference data Landsat-8, the Hyperion images are mainly divided into three categories: seawater, white ice, and gray ice, and a total of 3190 labeled samples are selected as samples of the network. The scene, shown in Figure 5(a), is a hyperspectral dataset that is a false color image composed of three bands: R: 115, G: 102, and B: 91. Figure 5(a) shows the RGB image of Baffin Bay image. Figure 5(b) shows magnified view of part of the area in Figure 5(a). The corresponding class legends are shown in Figure 5(c). The number of training data for each category in Baffin Bay data is shown in Table 1.

The second dataset utilized in our experiments is the images of Bohai Bay, which is from the EO-1 hyperspectral sea ice dataset on January 23, 2008, and the image size selected in the experiment is $442 * 212$. According to the spectral curve, the image dataset is roughly divided into four

Input:

(1) The labeled samples set: data blocks of size $K \times K \times B$, K represents the image size of the input network and B represents the number of bands.

(2) The training samples S and test samples set T according to the training strategy.

(3) The batch size for training network: N

(4) The learning rate of the network: α

Begin:

SE-CNN Model:

(1) N training samples are randomly selected from S and input into a preestablished CNN.

(2) Squeeze operation: global average pooling of the features obtained by the CONV3 layer.

(3) Excitation operation: reduce the dimension of the feature maps from the squeeze operation, nonlinearize by the ReLU activation function, then upgrade through the fully connected layer, and then perform weight activation through sigmoid.

(4) Recalibration of the original features: according to the output of excitation operation, reweight the original spectral features by multiplication to the previous features.

(5) Convert the characteristics of the network training into a one-dimensional vector and input into the fully connected layer of 3D-CNN network, update the parameter weights in the network with the loss function, and optimize the network through the Adam optimizer.

(6) Repeat steps 1–5 until the network converges.

(7) Input T into the trained network to obtain prediction labels, and the classification accuracy is calculated by the confusion matrix.

(8) SE-CNN end;

(9) Save the features after training network.

SVM Model:

(10) Uniform normalization of saved features.

(11) According to the BVSB-ECBD algorithm, M feature samples with rich information content are selected.

(12) Optimize the parameters c and g of RBF kernel function by using the grid optimization method.

(13) Obtain the classification accuracy based on SVM classifier.

(14) SVM Model end

Output: Overall accuracy, Kappa coefficient;

End

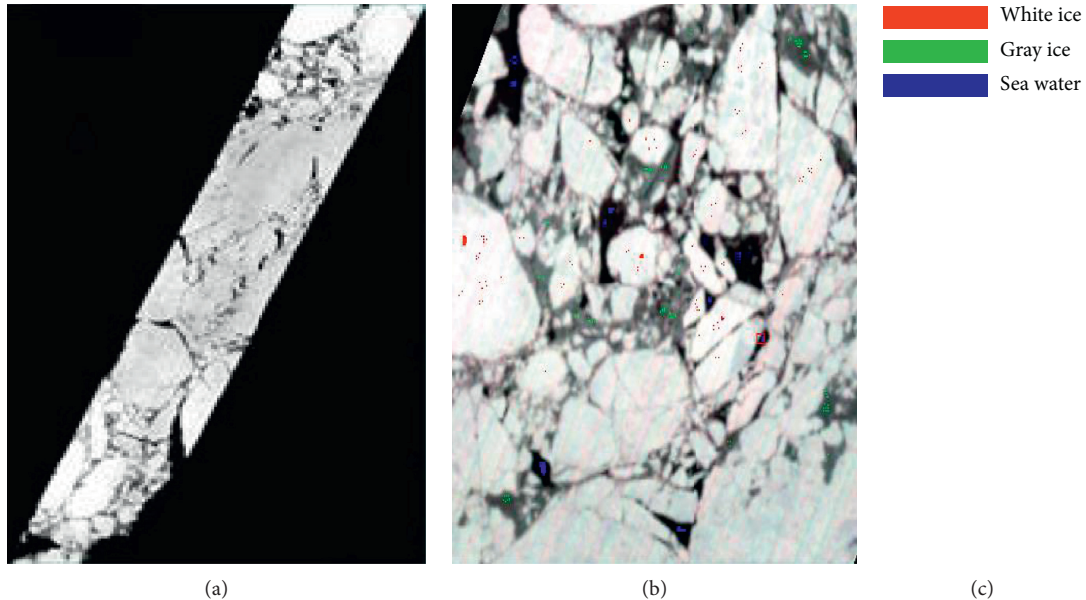ALGORITHM 1: SE-CNN-SVM algorithm.

FIGURE 5: (a) False color image composed of R: 115, G: 102, and B: 91; (b) magnified view of part of the area in (a); (c) class legend of Baffin Bay image.

categories: white ice, gray white ice, gray ice, and sea water, and a total of 1247 labeled samples are selected as samples of the network. The scene, shown in Figure 6(a), is a hyperspectral dataset that is a false color image composed of three bands: R: 93, G: 105, and B: 84. Figure 6(a) shows the RGB image of Bohai Bay image. Figure 6(b) shows marked

TABLE 1: Number of training data for each category in Baffin Bay data.

Class name	Number	5 per class	10 per class	20 per class
White ice	2098	5	10	20
Gray ice	783	5	10	20
Seawater	309	5	10	20
Total	3190	15	30	60

(a)   (b)   (c)
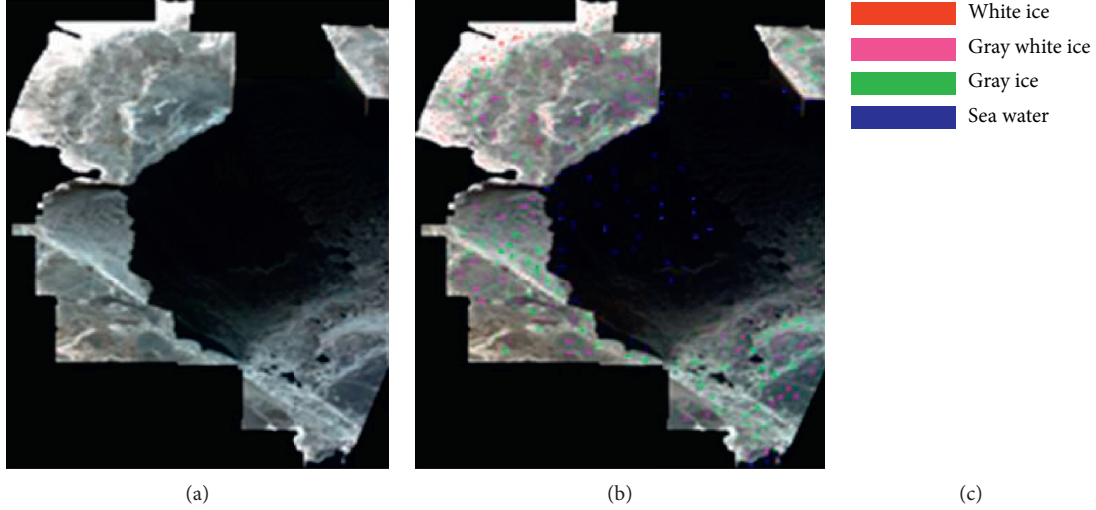
White ice
Gray white ice
Gray ice
Sea water

FIGURE 6: (a) False color image composed of R: 95, G: 105, and B: 84; (b) marked sample distribution map in the experiment; (c) class legend of Bohai Bay image.

TABLE 2: Number of training samples per category in the Bohai Bay data.

Class name	Number	10% training	20% training	30% training
White ice	142	14	28	42
Gray white ice	398	39	78	117
Gray ice	511	51	102	153
Seawater	196	19	38	57
Total	1247	123	246	369

sample distribution map in the experiment. The corresponding class legends are shown in Figure 6(c). The number of training samples per category in the Bohai Bay data is shown in Table 2.

The third dataset utilized in our experiments is an image of Liaodong Bay, which is a Landsat-8 dataset acquired on a section of coastal waters in the northeast of Bohai Sea on January 24, 2016. It has a 15 m spatial resolution and comprises 596×373 pixels. We have identified three classes of ice: white ice, gray ice, and white-gray ice. The scene, shown in Figure 7(a), is a hyperspectral dataset that is a false color image composed of three bands: R: 6, G: 5, and B: 4. Figure 7(a) shows the RGB image of Liaodong Bay image. Figure 7(b) shows magnified view of part of the area in Figure 7(a). The corresponding class legends are shown in Figure 7(c). The number of training data for each category Liaodong Bay data is shown in Table 3.

4.2. Network Structure Design.
In the three experiments, we design the network structure which contains seven different functional layers; they are the input layer, three convolutional layers, SE-Block, a full connection layer, and output layer, respectively. The learning rate of the model is set to 0.0005 and batch number is set to 25. The convolution layer of each layer uses the ReLU activation function, the sliding step size of the convolution kernel is [1, 1, 1], and the number of convolution kernels per layer is 2, 4, and 8, respectively. In SE-Block, the global average pooling size is [1, 1, 1]; the first fully connected layer (FC1) neuron number is 2, using the ReLU activation function; the second fully connected layer (FC2) neuron number is 8, using sigmoid activation function. The final fully connected layer uses the ReLU activation function with a dropout value of 0.5. The network structure of the three datasets is shown in Table 4.

4.3. Network Parameter Tuning

4.3.1. The Effect of Input Image Size on Classification Performance.
In the experiment, we input a three-dimensional data block with a neighborhood size of $K \times K \times B$ into

(a)   (b)   (c)
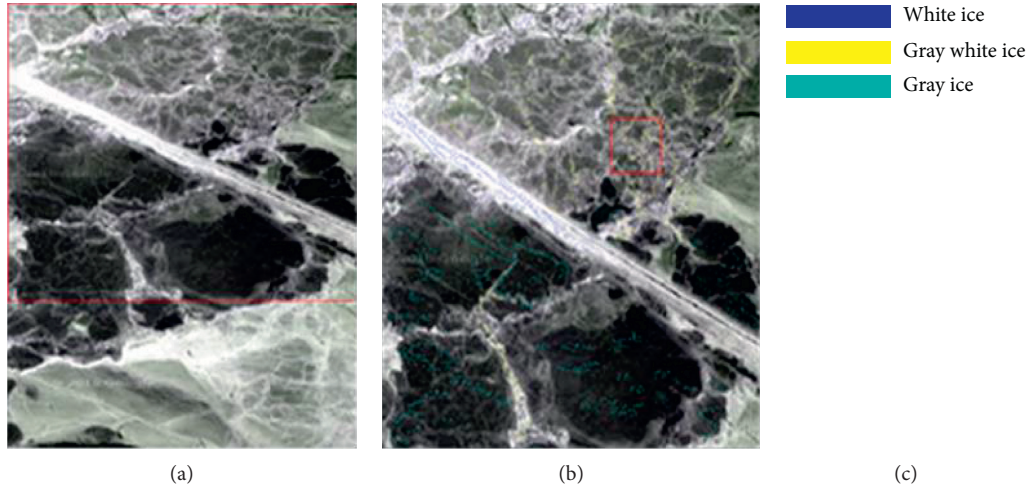
White ice
Gray white ice
Gray ice

FIGURE 7: (a) False color image composed of R: 6, G: 5, and B: 4; (b) magnified view of part of the area in (a); (c) class legend of Liaodong Bay image.

TABLE 3: Number of training data for each category Liaodong Bay data.

Class name	Number	5 per class	10 per class	20 per class
White ice	247	5	10	20
Gray ice	535	5	10	20
Seawater	676	5	10	20
Total	1458	15	30	60

TABLE 4: The network structure of the three datasets.

Network	Baffin Bay image	Bohai Bay image	Liaodong Bay image
Image size	$2395 \times 1769 \times 176$	$442 \times 212 \times 176$	$596 \times 373 \times 7$
Input layer	$5 \times 5 \times 176$	$5 \times 5 \times 176$	$5 \times 5 \times 7$
Conv1	$3 \times 3 \times 7$	$3 \times 3 \times 7$	$3 \times 3 \times 7$
Conv2	$3 \times 3 \times 5$	$3 \times 3 \times 3$	$3 \times 3 \times 3$
Conv3	$1 \times 1 \times 1$	$1 \times 1 \times 1$	$1 \times 1 \times 1$
SE-Block	(1) Global average Pooling (Conv3) (2) FC1 (units = 2, ReLU) (3) FC2 (units = 8, sigmoid) (4) Scale		
FC	128	128	128
Class num	3	4	3

TABLE 5: The effect of different sizes of network images on accuracy and time.

SE-CNN		Bohai Bay image		
		10% training	20% training	30% training
$5 \times 5 \times B$	Acc (%)	70.70	73.37	78.71
	Time (s)	338	396	694
$7 \times 7 \times B$	Acc (%)	70.80	73.85	78.55
	Time (s)	1090	1302	1514

TABLE 6: The effect of different sizes of network images on accuracy and time.

SE-CNN		Liaodong Bay image		
		5 per class	10 per class	20 per class
$5 \times 5 \times B$	Acc (%)	90.09	93.60	96.45
	Time (s)	20	39	44
$7 \times 7 \times B$	Acc (%)	93.78	91.07	96.77
	Time (s)	85	97	151

the deep network and use the label category of the central pixel as the category of the sample. However, different sizes of input image will affect the classification accuracy and model training time. We analyzed the effects of different sizes of input image on classification performance by using experimental data from Bohai Bay and Liaodong Bay, respectively. The experimental results are shown in Tables 5 and 6. Among them, the accuracy and time data are the average of 5 experimental results.

In the Bohai Bay data, when the input image size is changed from $5 \times 5 \times B$ to $7 \times 7 \times B$ in different proportions of training data, taking 10% and 20% samples size for each type randomly as training samples, the classification accuracy increased by 0.30% and 0.47% respectively, but the network training time has increased by 2.2 times and 2.3 times; when taking 30% sample size of each type was randomly selected as the training samples, the classification accuracy decreased by 0.16%, and the training time increased by 1.2 times.

In the Liaodong Bay data, when the input image size is changed from $5 \times 5 \times B$ to $7 \times 7 \times B$ in each type of training data, taking 5 per class and 10 per class samples for each type randomly as training samples. The classification accuracy decreased by 0.31% and 2.52%, and the network training time increased by 3.3 times and 1.5 times, respectively; when taking 20 per class samples for each type randomly as training samples, the classification accuracy increased by 0.32%, but the training time increased by 2.4 times.

According to the experimental results, comprehensively considering the classification accuracy and training time,
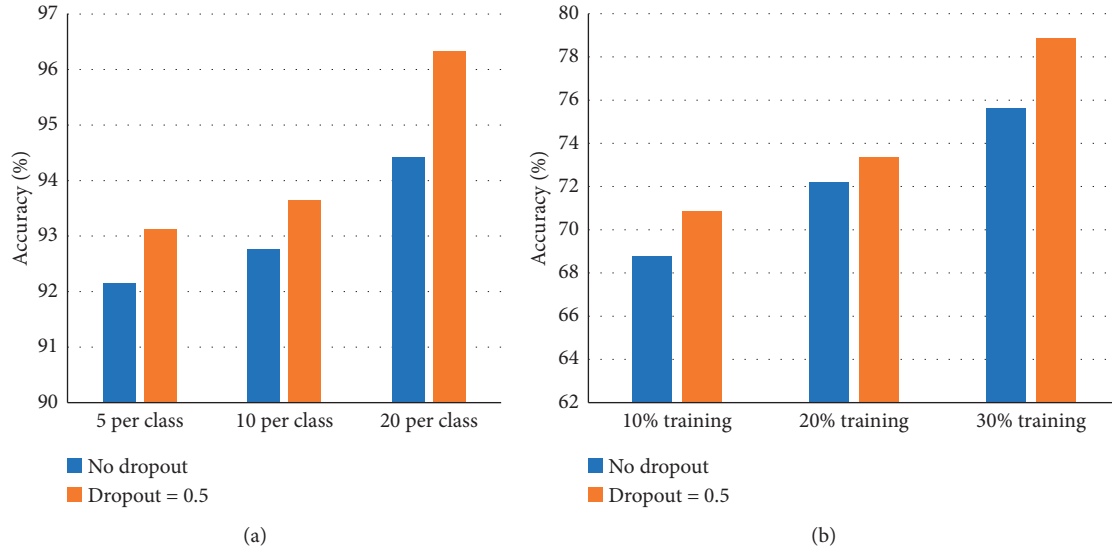
FIGURE 8: (a) Baffin Bay data SE-CNN structure with or without dropout layer experiment results when the number of trainings per class is 5, 10, and 20; (b) Bohai Bay data SE-CNN structure with or without dropout layer experiment results when the number of trainings per class is 5, 10, and 20.

this paper sets the image size of the input network to $5 \times 5 \times B$, and B indicates the number of bands.

4.3.2. Impact of Dropout Layer on Classification Performance.
Dropout is an optimization used to solve overfitting and gradient-disappearing problems in deep learning networks. In the learning process, the partial weights or outputs of the hidden layers are randomly zeroed, thereby reducing the dependencies among nodes and improving the classification performance. This paper verified the effects of dropout on classification performance through experiments in Baffin Bay and Bohai Bay.

In the Baffin Bay data, when 5, 10, and 20 samples of each class are taken randomly as training samples, the classification accuracy increased by 0.97%, 0.89%, and 1.85% after adding the dropout layer. In the Bohai Bay data, when 10%, 20%, and 30% samples of each category are taken randomly as training samples, the classification accuracy increased by 2.10%, 1.16%, and 3.22% after adding the dropout layer. The experimental results are shown in Figure 8.

According to the experimental results, this article adds a dropout layer after the last fully connected layer. The dropout value is 0.5, which means that some parameters are randomly discarded with a probability of 50% so that the network will not be overfitting, and because the parameters are reduced, the network training speed is also faster.

4.4. Result Analysis

4.4.1. Experimental Results on Baffin Bay Dataset.
The result maps of Baffin Bay dataset using 5 methods are shown in Figure 9. Table 7 shows a comparison of experimental results in the Baffin Bay data experiment using the proposed method and several other classical methods in selecting different sample sizes. In Table 7, the proposed method

achieves a better classification effect. This shows that the deep learning method can deeply explore the intrinsic relationship among the spatial-spectrum characteristics of multi/hyperspectral remote sensing sea ice image, better extract the typical characteristics of different types of sea ice, and achieve higher classification performance under small sample conditions. When 5, 10, and 20 samples are randomly selected as training samples for each category, the classification accuracy can reach 93.98%, 94.54%, and 97.02%, respectively.

From Table 7, the SVM classification accuracy is generally low, indicating that the deep learning algorithm generally obtains better classification results than the shallow learning algorithm. The Siamese method has the lowest precision, and the Siamese method is more suitable for more classified image classifications due to its double-convolution network structure. Due to the advantages of SVM classifiers in dealing with small samples and non-linear high-dimensional feature classification problems, compared with CNN methods, CNN-SVM can obtain better classification results than CNN's own softmax classifier. The method proposed in this paper considers the small sample problem and the complex correlation among spectra, 3D-CNN is used to extract different types of sea ice features, and SE-Block is integrated to optimize the weight of each spectral feature, further distinguishing the contribution of different spectral features to sea ice classification. Finally, the SVM classification model is used for sea ice classification, which improves the separability between sea ice categories, thus achieving better classification performance. For example, when 20 samples are randomly selected for each category as the training samples, the classification accuracy is 97.02%, which is higher than the Siamese method, SVM method, CNN method, and CNN-SVM method, 14.00%, 4.97%, 2.99%, and 2.10% respectively.
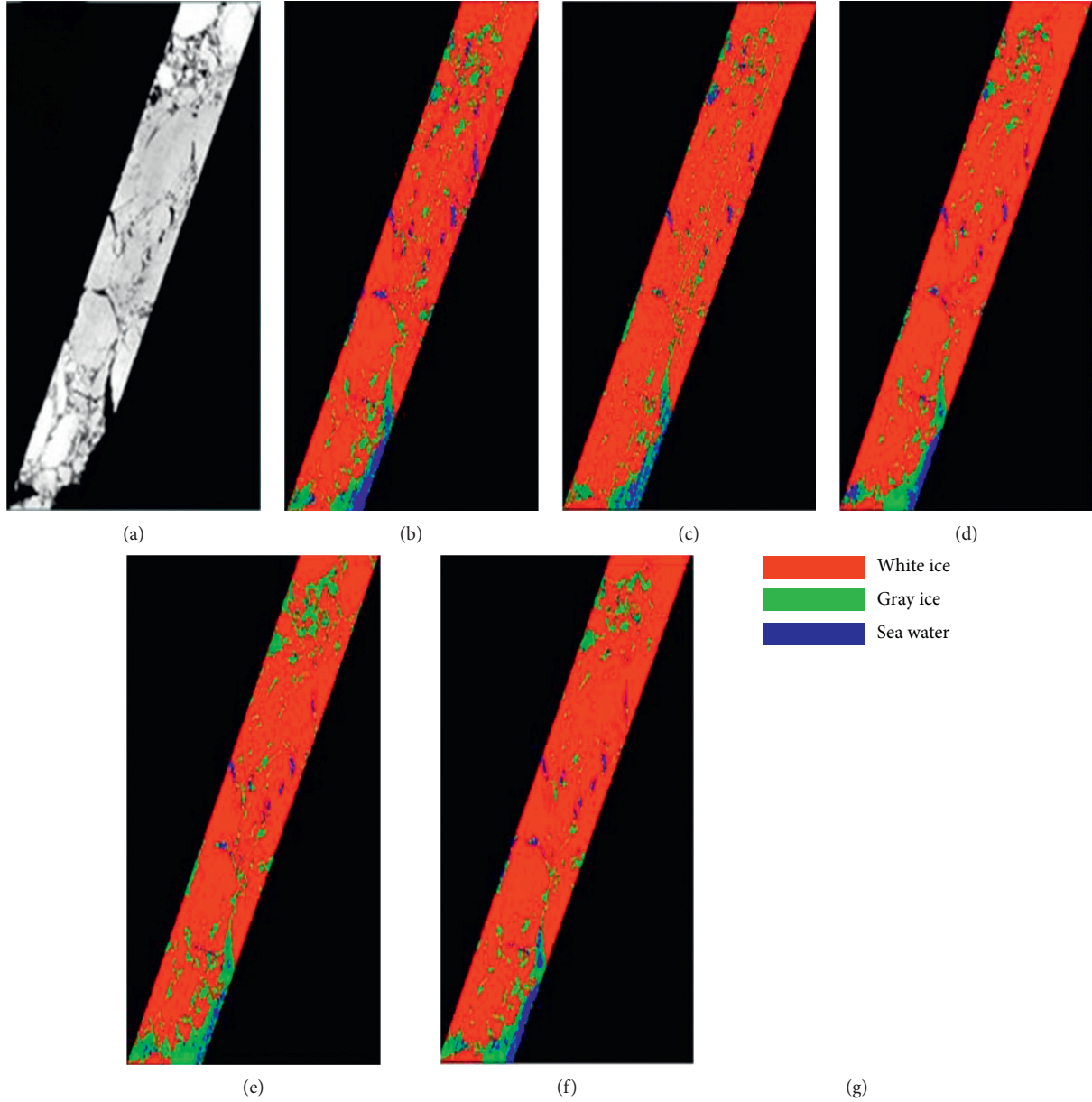
FIGURE 9: Classification maps of Baffin Bay dataset: (a) false-color composite; (b) SVM; (c) Siamese; (d) CNN; (e) CNN-SVM; (f) SE-CNN-SVM; (g) class legend of Baffin Bay image.

TABLE 7: Classification results (%) of SE-CNN-SVM and other methods.

Method	5 per class	10 per class	20 per class
SVM	90.58	91.08	92.05
Siamese	76.44	79.87	83.02
CNN	93.07	93.68	94.03
CNN-SVM	93.95	94.85	94.92
SE-CNN-SVM	93.98	94.54	97.02

4.4.2. Experimental Results on Bohai Bay Dataset. The result maps of Bohai Bay dataset using 5 methods are shown in Figure 10. Table 8 shows a comparison of experimental results in the Bohai Bay data experiment using the proposed method and several other classical methods in selecting different sample proportion. In Table 8, the proposed method achieves higher classification results under small sample conditions. When each class is randomly selected as 10%, 20%, and 30% as training samples, the classification accuracy can reach 72.58%, 74.93%, and 80.64%.

From Table 8, the SVM classification accuracy is generally low, indicating that the deep learning algorithm generally obtains better classification results than the shallow learning algorithm; the Siamese method has the lowest precision; compared with CNN method and CNN-SVM, the method proposed in this paper comprehensively considers the small sample problem and the complex correlation among spectra and distinguishes the contribution of

(a) (b) (c) (d)

(e) (f) (g)

White ice
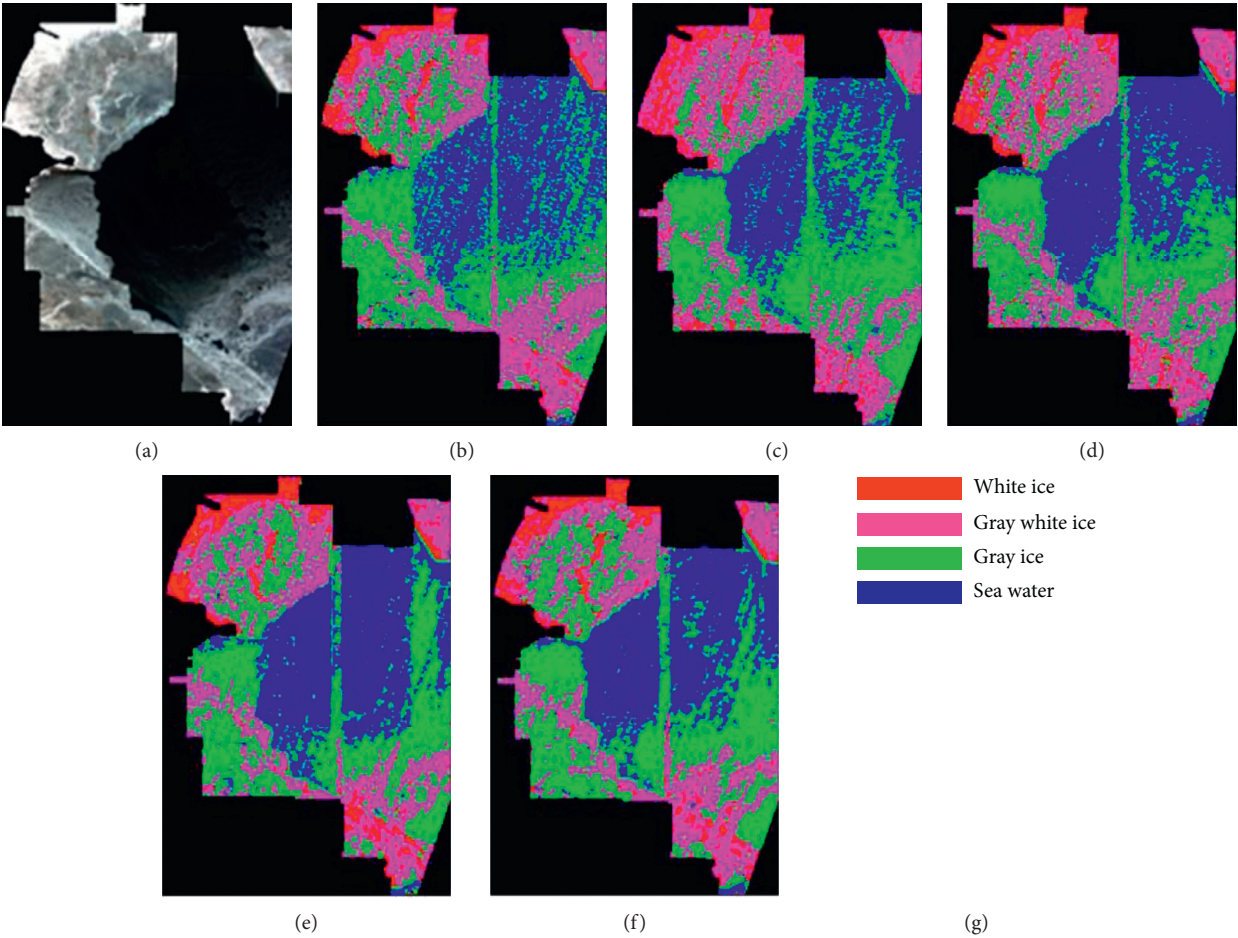Gray white ice
Gray ice
Sea water

FIGURE 10: Classification maps of Bohai Bay dataset: (a) false-color composite; (b) SVM; (c) Siamese; (d) CNN; (e) CNN-SVM; (f) SE-CNN-SVM; (g) class legend of Bohai Bay image.

TABLE 8: Classification results (%) of SE-CNN-SVM and other methods.

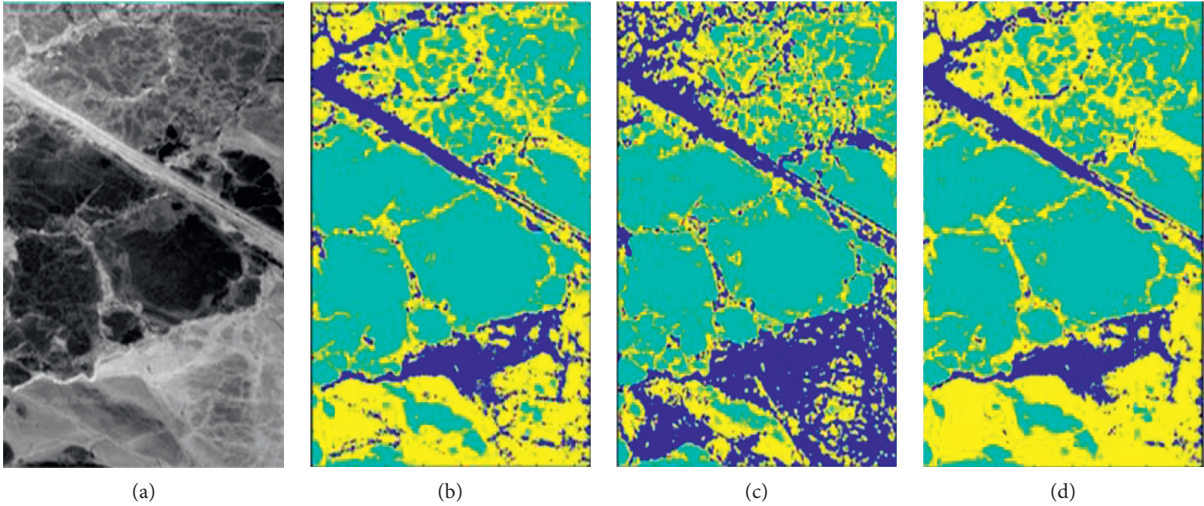| Method	10% training	20% training	30% training
SVM	64.41	67.53	73.45
Siamese	64.09	67.34	70.61
CNN	68.51	70.59	76.64
CNN-SVM	70.78	73.37	78.71
SE-CNN-SVM	72.58	74.93	80.64

(a) (b) (c) (d)
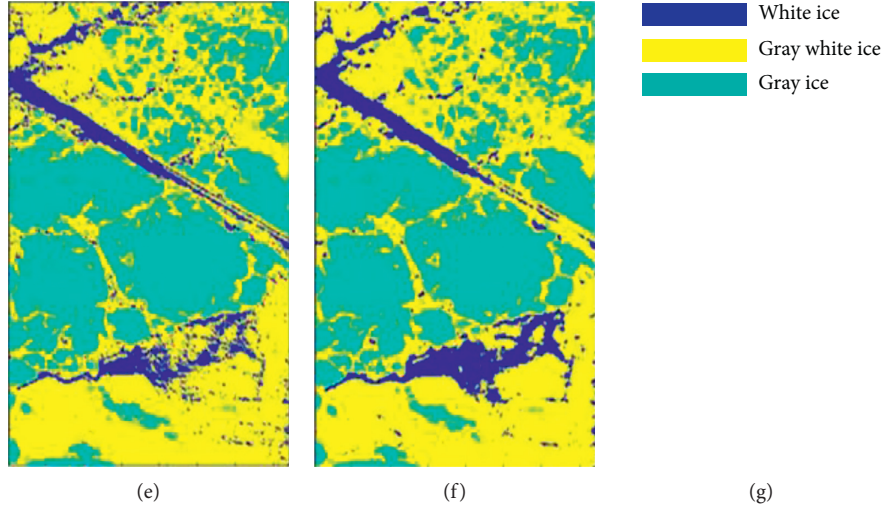
FIGURE 11: Continued.

Figure 11: Classification maps of Liaodong Bay dataset: (a) false-color composite; (b) SVM; (c) Siamese; (d) CNN; (e) CNN-SVM; (f) SE-CNN-SVM; (g) class legend of Liaodong Bay image.

Table 9: Classification results (%) of SE-CNN-SVM and other methods.

Method	5 per class	10 per class	20 per class
SVM	88.16	89.40	91.14
Siamese	81.18	83.45	85.10
CNN	90.90	91.77	94.96
CNN-SVM	93.46	93.35	95.48
SE-CNN-SVM	94.58	95.11	97.42

different spectral features to the classification of sea ice by SE-Block, and finally, the SVM classification model is used to classify and obtain better classification performance. When each class randomly selects 10%, 20%, and 30% as training samples, the classification accuracy of the proposed method is higher than the other four methods. When each class randomly selects 30% as training samples, the accuracy difference from other four methods reaches the maximum, which is 10.03%, 7.19%, 4.00%, and 1.93%, respectively.

4.4.3. Experimental Results on Liaodong Bay Dataset. The result maps of Liaodong Bay dataset using 5 methods are shown in Figure 11. Table 9 shows a comparison of experimental results in the Liaodong Bay data experiment using the proposed method and several other classical methods in selecting different sample size. From Table 9, the proposed method achieves higher classification performance in small sample cases. When 5, 10, and 20 samples are randomly selected for each type as training samples, the classification accuracy can reach 94.58%, 95.11%, and 97.42%. When each class randomly selects 20 as training samples, the SE-CNN-SVM method is 12.32% higher than the Siamese method, 6.28% higher than the SVM method, 2.46% higher than the CNN method, and 1.94% higher than the CNN-SVM method.

5. Conclusion

Because of the high labeling cost in remote sensing sea ice image classification, the labeled samples are difficult to acquire, which causes small sample problems. At the same time, there are high correlations among multiple spectral channels in remote sensing sea ice data, and different channels have different degrees of discrimination for sea ice classification, which results in low classification accuracy of sea ice image. Aiming at above problems, this paper proposes a new convolutional neural network model for remote sensing sea ice image classification and compares the proposed method with several other classical remote sensing image classification methods. The experimental results show that compared with other methods, the proposed method SE-CNN-SVM can effectively extract feature information from remote sensing sea ice images with fewer labeled samples, weight the spectral features according to the contribution of different spectral channels in the sea ice classification, and further optimize the model structure, and it can achieve better overall classification performance overall. We can summarize the results as follows.

The convolutional neural network method can extract image features by autonomous learning and is widely used in remote sensing image classification. And 3D-CNN model can simultaneously extract the spectral and spatial features of remote sensing sea ice data, which fully exploits the sea ice feature information hidden in the remote sensing data. It meets the requirements of remote sensing sea ice image classification and achieves better classification results.

There are high correlations among multiple spectral channels in remote sensing sea ice data, and different channels have different contributions for sea ice classification. Therefore, unified processing of each spectral channel data indiscriminately will inevitably limit the improvement of classification accuracy. The proposed method integrated the SE-Block in the 3D-CNN structure and improved the network model to achieve better classification result by

increasing the weight of the effective feature and reducing the feature weight with invalid or small effect.

Compared to CNN's softmax classifier, SVM has obvious advantages in the way of solving small samples, nonlinearity, and preventing the network from entering local minimum. The proposed method combined 3D-CNN with SE-Block for spatial spectral feature extraction and weighting and integrated the active learning method to select rich and representative samples and input SVM classifier for classification, which further achieved higher classification accuracy of remote sensing sea ice image with small samples and provided a new method for remote sensing sea ice 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 there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

References

[1] Y. Han, J. Li, Y. Zhang, Z. Hong, and J. Wang, "Sea ice detection based on an improved similarity measurement method using hyperspectral data," *Sensors*, vol. 17, no. 5, p. 1124, 2017.

[2] D. Lindell and D. Long, "Multiyear arctic ice classification using ASCAT and SSMIS," *Remote Sensing*, vol. 8, no. 4, p. 294, 2016.

[3] Z. Luo and L. Sun, "Remote sensing monitoring of Bohai sea ice based on hyperspectral imagery," *Science of Surveying and Mapping*, vol. 37, no. 1, pp. 54-55, 2012.

[4] H. Xie and X. Tong, "A probability-based improved binary encoding algorithm for classification of hyperspectral images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 6, pp. 2108–2118, 2014.

[5] N. Rodriguez-Alvarez, B. Holt, S. Jaruwatanadilok, E. Podest, and K. C. Cavanaugh, "An Arctic sea ice multi-step classification based on GNSS-R data from the TDS-1 mission," *Remote Sensing of Environment*, vol. 230, Article ID 111202, 2019.

[6] C. Chen, Y. Ma, and G. Ren, "A convolutional neural network with fletcher-reeves algorithm for hyperspectral image classification," *Remote Sensing*, vol. 11, no. 11, p. 1325, 2019.

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

[8] G. Cheng, J. Han, and X. Lu, "Remote sensing image scene classification: benchmark and state of the art," *Proceedings of the IEEE*, vol. 10, no. 105, pp. 1865–1883, 2017.

[9] W. Zhao, Z. Guo, J. Yue, X. Zhang, and L. Luo, "On combining multiscale deep learning features for the classification of hyperspectral remote sensing imagery," *International Journal of Remote Sensing*, vol. 36, no. 13, pp. 3368–3379, 2015.

[10] J. Leng, T. Li, G. Bai, Q. Dong, and H. Dong, "Cube-CNN-SVM: a novel hyperspectral image classification method," in *Proceedings of the IEEE International Conference on Tools with Artificial Intelligence*, November 2017.

[11] G. Alimjan, T. Sun, Y. Liang et al., "A new technique for remote sensing image classification based on combinatorial algorithm of SVM and KNN," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 32, no. 7, Article ID 1859012, 2017.

[12] A. C. Teodoro, F. Gutierres, P. Gomes, and J. Rocha, "Remote sensing data and image classification algorithms in the identification of beach patterns," *Beach Management Tools—Concepts, Methodologies and Case Studies*, Springer, Cham, Switzerland, 2018.

[13] D. Liu, L. Han, and X. Han, "High-resolution remote sensing image classification based on deep learning," *Acta Optica Sinica*, vol. 36, no. 4, pp. 298–306, 2016.

[14] P. Liu, K. R. Choo, L. Wang, and F. Huang, "SVM or deep learning? A comparative study on remote sensing image classification," *Soft Computing*, vol. 21, no. 23, pp. 7053–7065, 2016.

[15] A. B. Hamida, A. Benoit, P. Lambert, and C. B. Amar, "3-D deep learning approach for remote sensing image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 8, pp. 4420–4434, 2018.

[16] G. Cheng, Z. Li, J. Han, X. Yao, and L. Guo, "Exploring hierarchical convolutional features for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 11, pp. 6712–6722, 2018.

[17] W. Hu, Y. Huang, L. Wei, F. Zhang, and H. Li, "Deep convolutional neural networks for hyperspectral image classification," *Journal of Sensors*, vol. 2015, Article ID 258619 12 pages, 2015.

[18] X. Yang, Y. Ye, X. Li et al., "Hyperspectral image classification with deep learning models," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 9, pp. 5408–5423, 2018.

[19] C. Zhang, X. Pan, H. Li et al., "A hybrid MLP-CNN classifier for very fine resolution remotely sensed image classification," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 140, pp. 133–144, 2018.

[20] S. Chaib, H. Yao, Y. Gu, and M. Amrani, "Deep feature extraction and combination for remote sensing image classification based on pre-trained CNN models," in *Proceedings of the Ninth International Conference on Digital Image Processing (ICDIP 2017)*, Hong Kong, China, May 2017.

[21] Q. Zhang, Z. Cui, X. Niu et al., "Image segmentation with pyramid dilated convolution based on ResNet and U-Net," in *Proceedings of the International Conference on Neural Information Processing*, Guangzhou, China, November 2017.

[22] M. Z. Alom, T. M. Taha, C. Yakopcic et al., "The history began from AlexNet: a comprehensive survey on deep learning approaches," 2018, https://arxiv.org/abs/1803.01164.

[23] Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi, "Deep feature extraction and classification of hyperspectral images based on convolutional neural networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 10, pp. 6232–6251, 2016.

[24] Y. Chen, Z. Lin, X. Zhao, G. Wang, and Y. Gu, "Deep learning-based classification of hyperspectral data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 6, pp. 2094–2107, 2014.

[25] B. Liu, X. Yu, P. Zhang, X. Tan, A. Yu, and Z. Xue, "A semi-supervised convolutional neural network for hyperspectral image classification," *Remote Sensing Letters*, vol. 8, no. 9, pp. 839–848, 2017.

[26] Y. Li, H. Zhang, and Q. Shen, "Spectral-spatial classification of hyperspectral imagery with 3D convolutional neural network," *Remote Sensing*, vol. 9, no. 1, p. 67, 2017.

[27] Z. Zhong, B. Fan, J. Duan et al., "Discriminant tensor spectral-spatial feature extraction for hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 5, pp. 1028–1032, 2015.

[28] J. Hu, L. Shen, S. Albanie et al., "Squeeze-and-excitation networks," 2017, https://arxiv.org/abs/1709.01507.

[29] K. Tan, "Hyperspectral remote sensing image classification based on support vector machine," *Journal of Infrared and Millimeter Waves*, vol. 27, no. 2, pp. 123–128, 2008.

[30] Y. Han, P. Li, Y. Zhang et al., "Combining active learning and transductive support vector machines for sea ice detection," *Journal of Applied Remote Sensing*, vol. 12, no. 2, Article ID 026016, 2018.