IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 20, 2019, accepted December 10, 2019, date of publication December 13, 2019,
date of current version December 23, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2959662

Coastline Extraction Method Based on Convolutional Neural Networks–A Case Study of Jiaozhou Bay in Qingdao, China

XIAO-YING LIU[1], RUI-SHENG JIA[1,2], QING-MING LIU[1], CHAO-YUE ZHAO[1], AND HONG-MEI SUN[1,2]

[1]College of Computer Science and Engineering, Shandong University of Science and Technology, Qingdao 266590, China
[2]Shandong Province Key Laboratory of Wisdom Mine Information Technology, Shandong University of Science and Technology, Qingdao 266590, China

Corresponding authors: Rui-Sheng Jia (jrs716@163.com) and Hong-Mei Sun (shm0221@163.com)

This work was supported in part by the Natural Science Foundation of Shandong Province, China, under Grant ZR2018MEE008, and in part by the Key Research and Development Program of Shandong Province, China, under Grant 2017GSF20115.

ABSTRACT The traditional edge detection-based shoreline extraction method is severely disturbed by noise, and it is difficult to obtain a continuous coastline. In response to the above problems, we propose a coastline extraction method based on convolutional neural networks. Firstly, we replace the standard convolution with the Mini-Inception structure in the backbone network to extract multi-scale features of the object, and all the multi-scale features are concatenated. Then, we use the leaky-ReLU activation function instead of the ReLU activation function to avoid the problem that ''dead'' neurons cannot learn the effective features of remote sensing images. Finally, the network fully exploits multi-level information of objects to perform the image-to-image prediction. We carried out experiments on the remote sensing images of Jiaozhou Bay in Qingdao. The experimental results showed that our method could effectively extract the coastline automatically, and the producer's accuracy and the user's accuracy were higher than the comparison methods.

INDEX TERMS Coastline extraction, remote sensing images, convolutional neural networks, backbone networks, activation function.

I. INTRODUCTION

The coastline is the boundary between land and sea. Under the influence of natural environment and human development, the coastline has been in a state of change. Accurately grasp the location of the coastline, the process of change, and the trend of future coastlines to guide coastal aquaculture, coastal zone development, navigation and transportation are of great significance.

The methods of coastline extraction using remote sensing images can be divided into two categories: manual visual interpretation [1] and automatic computer interpretation [2]. The visual interpretation has the advantages of high interpretation accuracy and continuous coastline extraction, but it is susceptible to subjective factors, the efficiency is very low, and the manual time-consuming is long. Automatic computer interpretation has become the main research direction of scholars at home and abroad because of its high efficiency and reusability. At present, the methods of coastline automatic extraction mainly include threshold segmentation-based method [3], index analysis-based method [4], [5], active contour model-based method [6], [7], region growing-based method [8]–[10], etc. These methods usually rely on the feature knowledge of remote sensing images. According to the actual situation of the research area, we need to gradually adjust various parameters to obtain the optimal result of the visual judgment. The process is time-consuming and laborious, and it is difficult to determine the parameters. However, the edge detection-based method [11]–[14] can detect the positions of the step change of gray value by the spatial relationship of coastline with Roberts, Canny, Sobel, Prewitt, Log, and other operators. The method is simple and effective and does not need to rely on the feature knowledge of remote sensing images to determine the parameters, so the method based on edge detection becomes the main method of coastline automatic extraction.

The associate editor coordinating the review of this manuscript and approving it for publication was Fan-Hsun Tseng.

However, it is sensitive to noise and easy to lose edge information when processing remote sensing images with complex edges and obvious noise. In recent years, the edge detection methods based on convolutional neural networks have become a research hotspot. In 2015, Xie and Tu [15] proposed a holistically-nested edge detection network (HED). For the first time, it performed image-to-image edge detection by using fully convolutional neural networks, and the detection effect obtained was far better than that of Canny. However, the network did not make full use of the features of the middle convolution layer, so that many details were lost. In 2017, Yu *et al.* [16] proposed a deep end-to-end network for category-aware semantic edge detection (CASENet). The network not only performed single edge detection, but also judged the edge category, but its detection accuracy was not high, and it could not meet the remote sensing images coastline extraction task with high precision requirements. In the same year, Liu *et al.* [17] proposed an accurate edge detector using richer convolutional features (RCF). It had a more detailed feature fusion, and the performance had been greatly improved. However, because the edge information and background noise of remote sensing images are more complex, the effect is not ideal when it is used in the coastline extraction task.

In order to overcome these limitations and improve the accuracy of coastline extraction, based on RCF, we propose a coastline extraction method based on convolutional neural networks. This paper mainly refers to the following research contents: (1) Aiming at the complex background noise and large edge density of remote sensing images, and the network structure is designed from the perspective of enhancing feature extraction ability. This paper introduces the Mini-Inception structure [18] and uses this structure to redesign the backbone network so that the multi-scale features of the objects can be extracted, and the multi-scale features can be effectively aggregated to make full use of all convolution features. In addition, the leaky-ReLU [19] activation function is used to replace the ReLU [20], [27] activation function, which avoids the problem that dead neurons cannot learn useful features of remote sensing images, thus improving the accuracy of coastline extraction. (2) The accuracy and effectiveness of the proposed method were verified by experiments on remote sensing images of Jiaozhou Bay in Qingdao and its adjacent waters.

II. RELATED WORK

RCF is a convolutional neural network for edge detection proposed by Liu *et al.* [17]. It encapsulates all convolutional features into more discriminative representation, making full use of rich feature hierarchies. Its network structure is shown in Figure 1.

RCF uses the backbone network to extract feature information and generate feature maps. The backbone network used includes convolution layers of VGG16 [21] and ResNet50/101 [22]. In Figure 1, the backbone network takes VGG16 as an example. The network is divided into

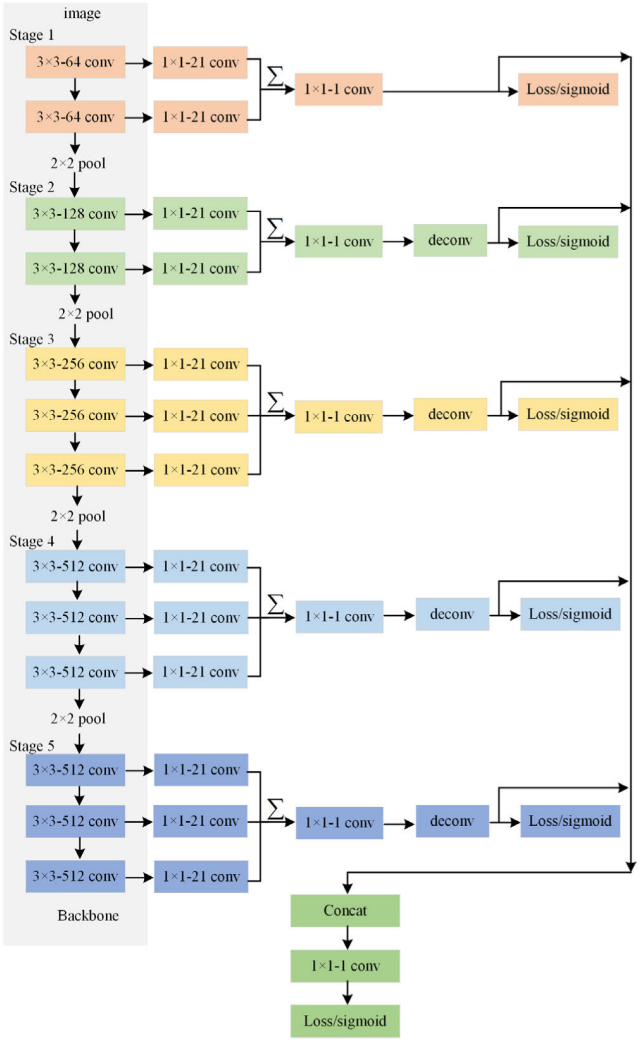


FIGURE 1. RCF network structure. The backbone network takes VGG16 as an example.

five stages. Firstly, the feature fusion is carried out by using the method of feature superposition and summation in each stage. Then, the fusion features are up-sampled and enlarged to the same size as the original image, which is convenient for the later stage supervision learning and the final feature fusion. Finally, the up-sampling features on five stages are concatenated, and the final detection results are output with 1×1-1 convolution fusion features.

The background noise and edge information of remote sensing images are more complicated than that of common images. When using convolution layers of VGG16 as the backbone network of RCF for the coastline extraction task, it is found that there are discontinuous areas in the coastline extracted. While RCF using ResNet50/101 as the backbone network, although the accuracy of coastline extraction has increased, the model contains more parameters, increases the calculation cost, and increases the requirements of hardware.

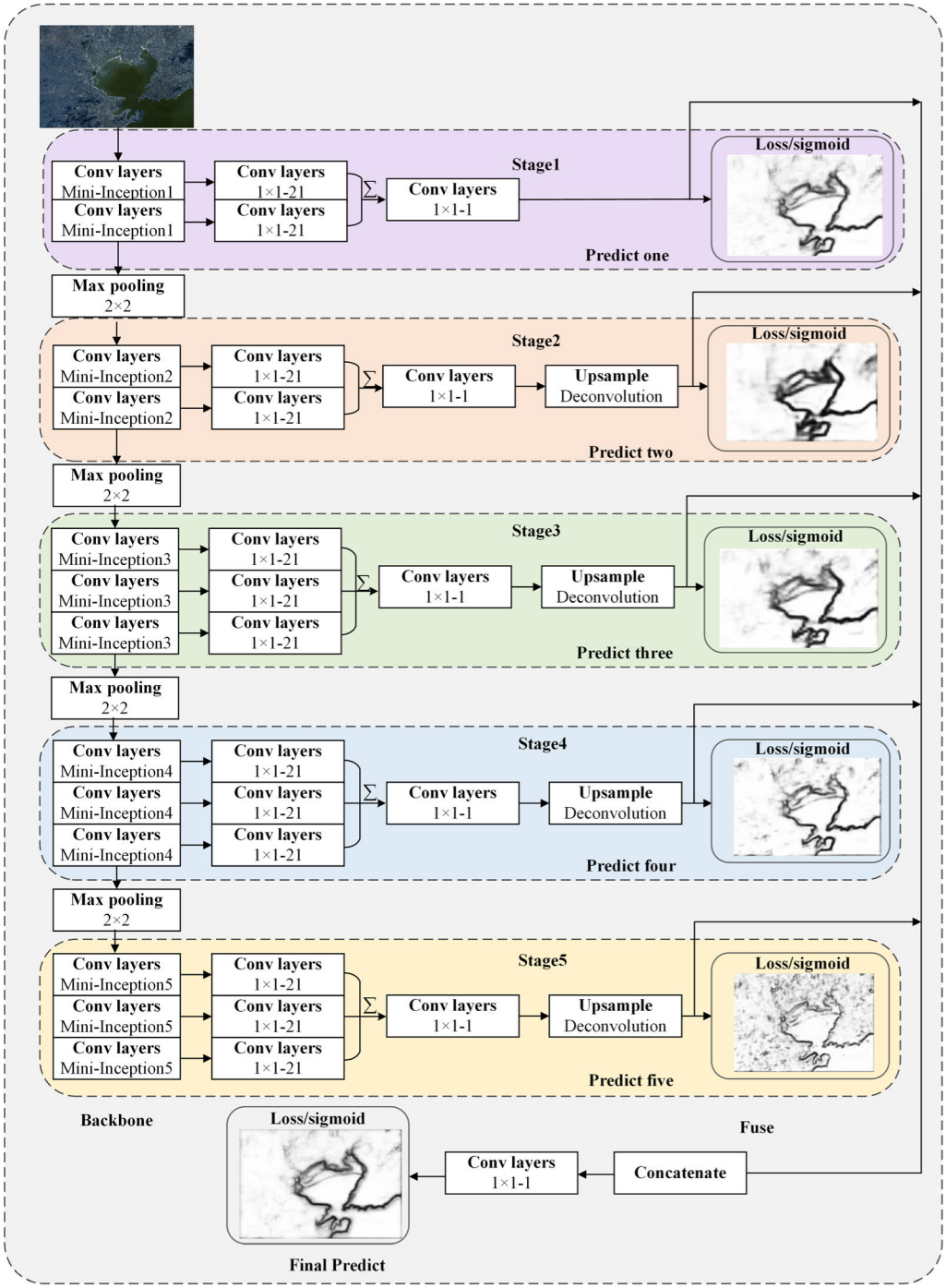


FIGURE 2. Our network architecture. The input is an image with arbitrary sizes, and our network outputs an edge possibility map in the same size.

III. PROPOSED ALGORITHM

A. NETWORK ARCHITECTURE

To overcome the shortcomings of RCF in using remote sensing images to extract coastline, we use Mini-Inception structure to replace the standard convolution in the backbone network, strengthen the feature extraction ability of backbone network for remote sensing images, and encapsulate all convolution features into more discriminative representation, making full use of multi-scale information and multi-level information of the objects. The ReLU activation function is used in RCF. As the training progresses, ''dead'' neurons will be generated. These neurons cannot continue to learn the useful features of remote sensing images. In order to avoid the appearance of ''dead'' neurons, we use the leaky-ReLU activation function to enable all neurons to produce effective predictions, which in turn improves the accuracy

of shoreline extraction. The network structure is shown in Figure 2.

Our network is divided into five stages, and Mini-Inception structures in each stage are used to extract convolution features. Each Mini-Inception structure is connected to a convolution layer with kernel size 1×1 and channel depth 21. Moreover, after convolution, all feature maps in each stage are accumulated to obtain hybrid features. After that, the convolution layer of 1×1-1 is followed. Then the deconvolution is used for up-sampling to get the output of the same size as the original image. Finally, in order to utilize all the feature hierarchies, the up-sampled feature maps of the five stages are concatenated, and the 1×1 convolution kernel is used to fuse the features, and the final result is output. The Mini-Inception structures and the pooling layers constitute the backbone network for extracting features. The backbone network will be described in detail below.

1) BACKBONE NETWORK

With the improvement of the spatial resolution of satellite sensors, the texture of remote sensing images is clearer, and the background is more complex, which makes the task of coastline extraction more challenging. When using a shallow network to process remote sensing images, the extraction capability of the network is limited due to the limited range of perception. Using the deeper network, because of the increase of the receptive field, we can get better detection accuracy, but the deeper network structure represents a vast amount of parameters, which increases the consumption of computing resources. Therefore, we introduce the Mini-Inception structure, which aims to increase the feature extraction capability of the backbone network without significantly increasing the amount of calculation. Its structure is shown in Figure 3. Furthermore, based on the 13 convolutional layers of VGG16, the backbone network is designed using the Mini-Inception structure. The details of the backbone network are shown in Figure 4.

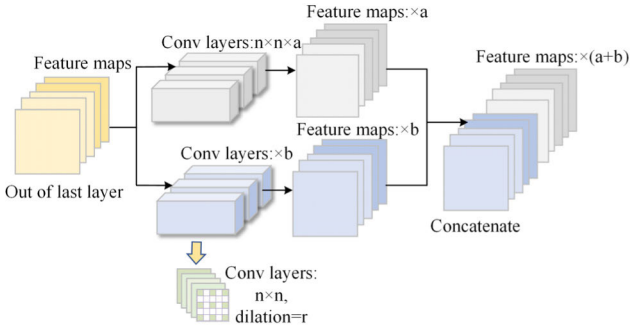


FIGURE 3. Mini-Inception structure. $n \times n$ denotes the size of the convolution kernel, a and b denote the number of convolution kernels, r denotes the dilated rate.

As shown in Figure 3, the essence of the Mini-Inception structure is that the input data enters two branches separately for parallel computing. The one uses a standard $n \times n$ convolution. The other uses a dilated convolution [24] with kernel size $n \times n$ and dilated $= r$, which is equivalent to the convolution with the convolution kernel size of $[n+(n-1)\times(r-1)] \times [n+(n-1)\times(r-1)]$, but the actual parameter amount is the same as the standard $n \times n$ convolution, that is, it does not increase any calculation cost. Finally, the feature maps of the two branches are concatenated as an output. The Mini-Inception structure uses the dilated convolution to increase the receptive field of different branch convolution kernels. For the same input object, it can extract features of different scales at the same time, so the structure is used instead of the standard convolution to enhance the feature extraction ability of the backbone network for remote sensing images.

As shown in Figure 4, in the backbone network, 13 Mini-Inception structures are used to replace thirteen 3×3 standard convolutional layers. The total number of convolution kernels in each convolution layer is based on the setting of 13 convolution layers in VGG16, that is, the total number of convolution kernels used in Mini-Inception1-5 are 64, 128, 256, 512, 512 respectively, and then they are equally allocated to two branches. The input object enters into two branches for parallel calculation. The one uses the standard 3×3 convolution to extract features. The other uses the dilated convolution with kernel size 3×3 and dilation $= 2$ to extract features, which is equivalent to 5×5 convolution kernel, and then the feature maps of the two branches are concatenated and used as input to the next convolutional layer.

2) ACTIVATION FUNCTION

The activation function is the key to the artificial neural network to achieve nonlinear expression, which directly affects the final training effect of the network. The ReLU activation function is used in RCF, and the output after each convolution is:

$$F_n(Y) = Max(0, W_n * F_{n-1}(Y) + B_n) \tag{1}$$

where $Max(0, \cdot)$ corresponds to the ReLU activation function, W_n is the convolution kernel size, B_n is the bias, $F_{n-1}(Y)$ is the feature map of the last layer output.

It can be seen from formula (1) that when its input is negative, it will be forced to convert to 0, resulting in that the corresponding weights of neurons cannot be updated during the training process. These neurons are called ''dead'' neurons. And these ''dead'' neurons cannot be activated by any incoming data, which means they are unable to continue to learn any features. However, remote sensing images have large edge densities and complex features. If we continue to use the ReLU activation function, we will lose much useful information. The leaky-ReLU activation function is especially proposed to solve the problem of neuron death by ReLU. When the input data is negative, it can also produce an effective prediction. Therefore, we choose the leaky-ReLU activation function, and the output after convolution is:

$$F_n(Y) = Max(\alpha * (W_n * F_{n-1}(Y) + B_n), \quad W_n * F_{n-1}(Y) + B_n) \tag{2}$$

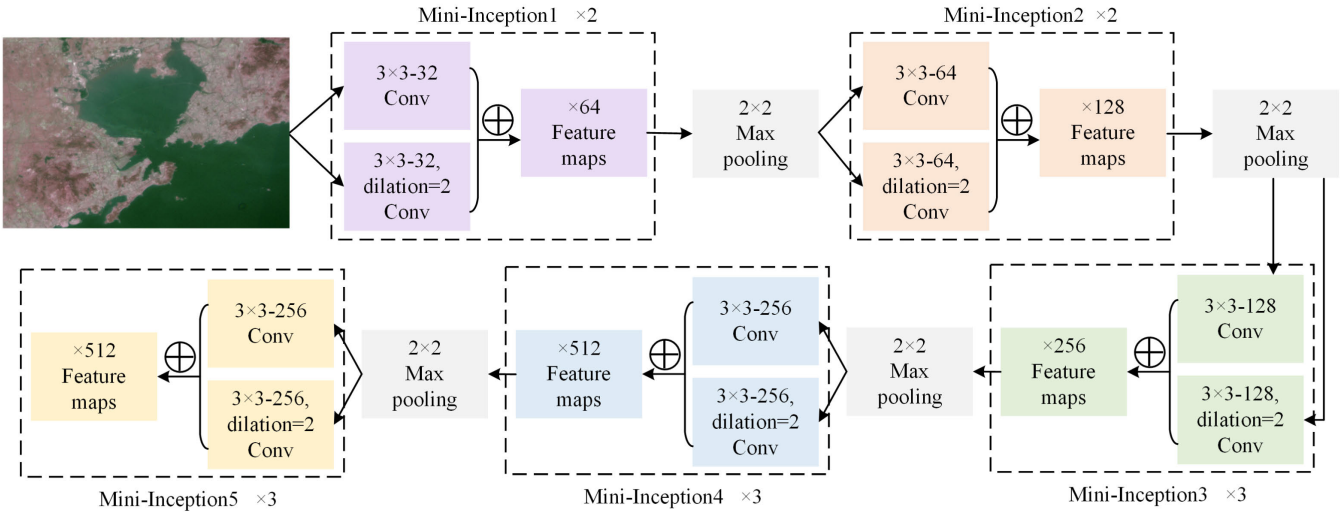


FIGURE 4. Detail of backbone network. × 2 denotes two Mini-Inception structures, × 3 denotes three Mini-Inception structures. The backbone network uses a total of 13 Mini-Inception structures.

where W_n, $F_{n-1}(Y)$, B_n is the same as formula (1), α is the parameter in the leaky-ReLU activation function, and the expression of the leaky-ReLU activation function is $Max(\alpha x, x)$.

3) LOSS FUNCTION

The definition of the loss function is essential for the performance of convolutional neural networks. The training process of the generating model is the optimization process of the loss function. Our network implements a two-class task that classifies each pixel as an edge pixel and a non-edge pixel. In a remote sensing image, the percentage of edge pixels belonging to the coastline is very small, and the number of positive and negative samples is unbalanced. In order to avoid the problem of low classification detection performance caused by the imbalance of positive and negative samples, we choose the cost-sensitive loss function [17] as the loss function. It is defined as follows:

The loss of each pixel is calculated as follows:

$$l(X_i; W) = \begin{cases} \alpha \cdot \log(1 - P(X_i; W)), & if\ y_i = 0 \\ 0, & if\ 0 < y_i \le \eta \\ \beta \cdot \log P(X_i; W), & otherwise \end{cases} \tag{3}$$

in which, the parameters α and β are calculated as follows:

$$\alpha = \lambda \cdot \frac{|Y^+|}{|Y^+| + |Y^-|} \tag{4}$$

$$\beta = \lambda \cdot \frac{|Y^-|}{|Y^+| + |Y^-|} \tag{5}$$

the sum of all pixels loss in the image is taken as the final loss, which is calculated as follows:

$$L(W) = \sum_{i=1}^{|I|} \left(\sum_{k=1}^{K} l\left(X_i^{(k)}; W\right) + l\left(X_i^{fuse}; W\right) \right) \tag{6}$$

in formula (4), $l(X_i; W)$ denotes the loss of pixel i, y_i is the edge probability of the pixel i in the ground truth, $P(\cdot)$ is the standard sigmoid function, X_i is the activation value(CNN feature vector), W denotes all the parameters that will be learned in our architecture, the value of parameter η is preset. In equations (4) and (5), Y^+ denotes the set of pixels with edge probability greater than η, as a positive sample set, Y^- denotes the set of pixels with edge probability less than η, as a negative sample set, and the hyper-parameter λ is used to balance positive and negative samples. In formula (6), $|I|$ denotes the number of pixels in image I, $L(W)$ denotes the loss of image I, $l\left(X_i^{(k)}; W\right)$ denotes loss of pixel i in stage k, $l\left(X_i^{fuse}; W\right)$ is the loss of pixel i in the fuse layer.

B. COASTLINE EXTRACTION PROCESS

In order to achieve the coastline extraction of remote sensing images, it is necessary to pre-train the convolutional neural network model, and then input the remote sensing image into the trained model for coastline extraction. The coastline extraction process is shown in Figure 5, which is divided into two parts: training and testing.

The training process is as follows:

Step1: We mix augmentation data of the remote sensing images with augmentation data of BSDS500 [25] as the training data set $I = \{\{I_{1x}, I_{1y}\}., \{I_{2x}, I_{2y}\}, \ldots, \{I_{nx}, I_{ny}\}\}$, in which, I_{nx} denotes original image, and I_{ny} denotes ground truth, n denotes the number of images in the training data set I.

Step2: The training data set I is input to the convolutional neural network for training, the pre-trained ImageNet [30] model is used to initialize the backbone network, and the network parameters are trained via backpropagation to obtain the model 'Training completed CNN_1'.

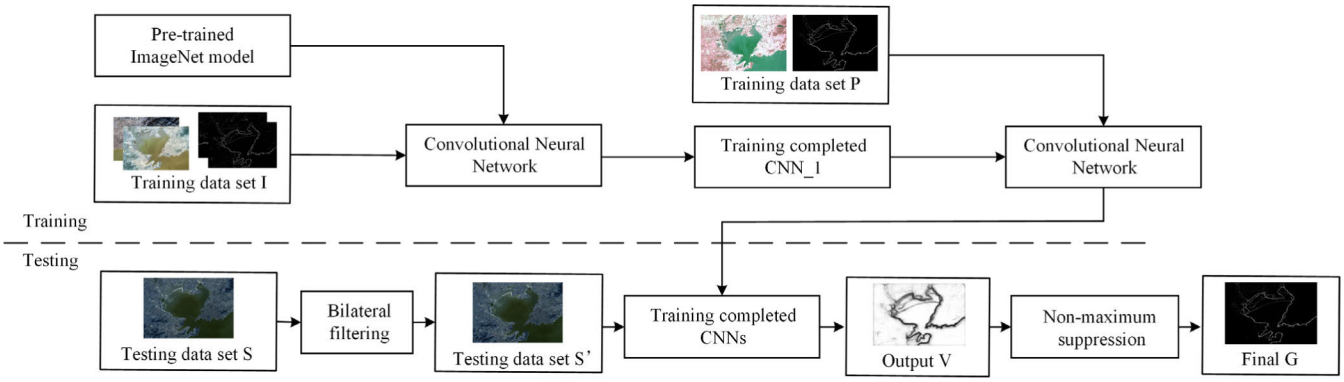


FIGURE 5. Coastline extraction process.

Step3: We use augmentation data of the remote sensing images as training data set $P = \{\{P_{1x}, P_{1y}\}, \{P_{2x}, I_{2y}\}, \ldots, \{P_{nx}, P_{ny}\}\}$, in which, P_{nx} denotes original image, and P_{ny} denotes ground truth.

Step4: The training data set P is input to the convolutional neural network, and the whole network is initialized by 'Training completed CNN_1' obtained by Step2, and fine-tuned by backpropagation to obtain the final model 'Training completed CNNs'.

The testing process is as follows:

Step1: We use multiple remote sensing images as testing data set $S = \{S_1, S_2, \ldots, S_m\}$, and then the image in S is subjected to bilateral filtering [23] to obtain the testing data set $S' = \{S'_1, S'_2, \ldots, S'_m\}$.

Step2: The testing data set S' is input into the trained model 'Training completed CNNs' for testing to obtain an initial coastline extraction set $V = \{V_1, V_2, \ldots, V_m\}$.

Step3: Finally, the initial coastline extraction set $V = \{V_1, V_2, \ldots, V_m\}$ is refined by the non-maximum suppression method to obtain the final coastline extraction result set $G = \{G_1, G_2, \ldots, G_m\}$.

IV. EXPERIMENTS

A. DATASET

Through the data distribution system, 40 different high-resolution remote sensing images of Jiaozhou Bay and its adjacent sea areas were obtained, with the spatial resolution of 16m and 50m. The remote sensing images were cut according to different proportions, and 200 images of different sizes were obtained. According to the data format of BSDS500 [25], the coastline was annotated with ArcGIS interpretation, and the JZBay_data set was made. 80 images were randomly selected as the testing data set, and 48 images with the spatial resolution of 16m were used as Test_1, and 32 images with the spatial resolution of 50m were used as Test_2. After that, the remaining 120 images are expanded to 8000 using data enhancement methods such as inversion, scaling and rotation, and regard them as the training data set, which is recorded as Train_1. Considering that the quality and quantity of the training set will affect the quality of the model,

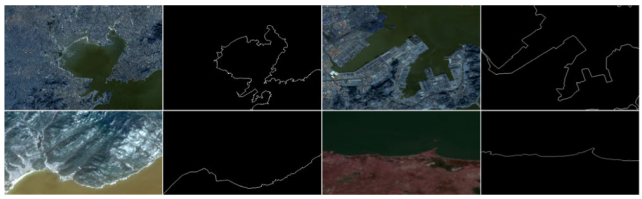


FIGURE 6. Some images and their ground truth edge maps in the JZBay_data set.

we take the augmentation training data set in BSDS500 and the data in the Train_1 as the training data set, a total of 36800, which is recorded as Train2. Some images in data set are shown in Figure 6.

B. TRAINING SETUP

All experiments in this paper were performed on a computer with an Intel(R) Core(TM) i7-8750H CPU@2.20 GHz CPU, 8GB RAM, and accelerated with an NVIDIA GeForce GTX 1060 graphics card. We implement our network using the Pytorch framework.

The training of the model was performed in two stages. In the first stage, we trained on Train_2. The weights of Mini-Inception1-4 were initialized using the pre-trained ImageNet model, and fine-tuning through backpropagation. The weights of Mini-Inception5 structure and all 1×1 convolutional layers were randomly initialized. The training was performed by stochastic gradient descent (SGD) with mini-batch size 10 and momentum 0.9. 30K subsequent iterations were performed with a learning rate of 1e-5. In the second stage, we fine-tuned the network on Train_1 only. Fine-tuning was performed for 40K iterations with a learning rate of 1e-6.

C. EVALUATION METRICS

We use qualitative evaluation and quantitative evaluation to evaluate all experimental results. Qualitative evaluation: with the support of feature knowledge, image is evaluated by visual observation, mainly including position and shape. Quantitative evaluation: the image is evaluated by statistical parameters, and producer's accuracy (PA), user's

TABLE 1. Quantitative evaluation results (PA/UA/OE/CE) of Sobel, Canny, HED, RCF, and our method.

Methods	PA		UA		OE		CE	
	Test_1	Test_2	Test_1	Test_2	Test_1	Test_2	Test_1	Test_2
Sobel	0.665	0.456	0.647	0.433	0.335	0.544	0.353	0.567
Canny	0.824	0.783	0.850	0.752	0.176	0.217	0.150	0.248
HED	0.897	0.882	0.929	0.895	0.103	0.118	0.071	0.105
RCF	0.925	0.909	0.936	0.923	0.075	0.091	0.064	0.077
Our method	**0.948**	**0.926**	**0.954**	**0.942**	**0.052**	**0.074**	**0.046**	**0.058**

accuracy (UA), errors of omission (OE), errors of commission (CE) were calculated based on the extraction and reference results [28].

Producer's accuracy refers to the ratio of the number of pixels correctly classified as class A by the classifier to the total number of class a real references; user's accuracy refers to the ratio of the total number of pixels correctly classified into class A to the total number of pixels in which the classifier divides the pixels of the entire image into class A. The calculation formulas of the producer's accuracy and the user's accuracy are as follows:

$$\begin{cases} \eta_1 = \dfrac{o' \cap o}{o} \\ \eta_2 = \dfrac{o' \cap o}{o'} \end{cases} \tag{7}$$

where η_1 and η_2 respectively denote the producer's accuracy and user's accuracy; o and o' respectively denote the reference data of the class A obtained by manual visual interpretation and the extracted data of class A.

Errors of omission is the probability that class A pixels are misclassified to other classes; errors of commission is the probability that other classes of pixels are misclassified into class A. The equations for errors of omission and errors of commission are as follows:

$$\begin{cases} \eta_3 = 1 - \eta_1 \\ \eta_4 = 1 - \eta_2 \end{cases} \tag{8}$$

where η_3 and η_4 respectively denote the errors of omission and errors of commission, η_1 and η_2 respectively denote the producer's accuracy and user's accuracy.

D. EXPERIMENTAL RESULTS AND ANALYSIS

In order to prove the effectiveness of our method, there is a comparison with Sobel [29], Canny [26], HED [15], RCF [17], and our method on the same experimental environment. The code of HED and RCF is the source code publicly released by the author. All network models are trained and tested on the same data set and evaluated both qualitatively and quantitatively.

Two remote sensing images are randomly selected from Test_1 and Test_2, respectively. Sobel, Canny, HED, RCF, and our method are used to extract the coastline, and the coastline extraction results are qualitatively evaluated. Figure 7 shows the coastline extraction results of two randomly selected remote sensing images with the spatial resolution of 16m from Test_1, and Figure 8 shows the coastline extraction results of two randomly selected remote sensing images with the spatial resolution of 50m from Test_2.

It can be seen from Figure 7 and Figure 8 that the coastlines extracted by the Sobel and the Canny are poorly continuous, and the land texture and the weak edges inside the water body are also extracted. The coastline extracted by HED and RCF also lose some edge points, but the effect on non-shoreline area processing is much better than Sobel and Canny. In comparison, our method is minimally disturbed by noise, and almost no land texture is extracted, and the obtained coastline results are the most continuous and smooth.

In order to verify that the experimental results are not accidental, the experiments are carried out on Test_1 and Test_2 respectively, and the quantitative evaluation of five algorithms is carried out by using four evaluation indexes of producer's accuracy (PA), user's accuracy (UA), errors of omission (OE), and errors of commission (CE). The results are shown in Table 1, and the best results are shown in bold. It should be noted that since the extraction results of Sobel and Canny contain too much background noise, the quantitative evaluation does not consider the land portion, and only evaluates the coastline.

As shown in Table 1, on Test_1 and Test_2, the PA and UA of the coastline extracted by Sobel are both lower than 0.7, and CE and OE are both higher than 0.3, which means that there are more missed and misaligned pixels. The PA and UA of HED are higher than Canny, but the PA on Test_1 and Test_2 is still less than 0.90, and the UA on Test_2 is also less than 0.90, indicating that the obtained coastline information accounts for a significant increase in the proportion of reference coastline information, but still cannot meet the requirements. The PA and UA of RCF and our method are higher than 0.90, but on Test_1, the PA and UA of our method are higher than RCF by 0.023 and 0.018, respectively. On Test_2, the PA and UA of our method are 0.017 and 0.019 higher than RCF, respectively. Based on Table 1, it is proved that the coastline extracted by our method has higher accuracy and is closest to the actual coastline.

To prove that our network does not significantly increase the calculation cost, it compares with parameters, model size, and time of HED and RCF based on convolutional neural networks. All the experiments are carried out on an Intel(R) Core(TM) i7-8750H CPU@2.20GHz, 8GB RAM computer and accelerated with an NVIDIA GeForce GTX

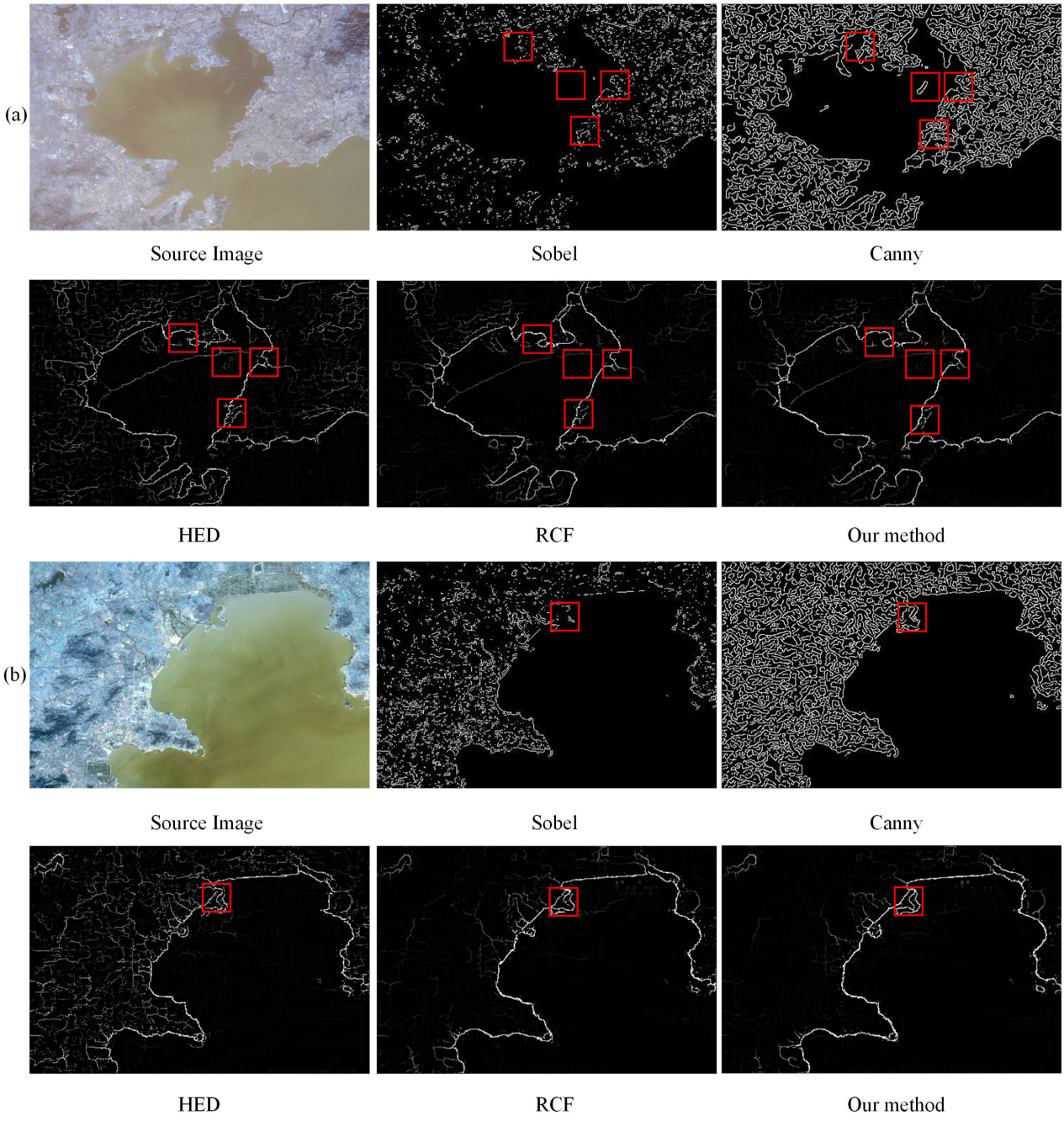


FIGURE 7. Coastline extraction results of Sobel, Canny, HED, RCF, and our method. (a) GF1_WFV1 remote sensing image with the spatial resolution of 16m on December 26, 2017; (b) GF1_ WFV2 remote sensing image with the spatial resolution of 16m on March 27, 2018.

1060 graphics card. The time is the average test time of 80 images in the testing data set. The results are shown in Table 2.

As shown in Table 2, the parameters of our network are the same as RCF and 0.1M less than HED. The model size of our network is 0.01MB and 0.68MB larger than RCF and HED, respectively. The average test time is 0.024s slower than HED and 0.0191s slower than RCF. The backbone network of HED, RCF, and our network are all designed based on VGG16, but HED does not fuse the features extracted from the intermediate convolution layer of each stage.

TABLE 2. Comparison of Parameters(M), Model Size(MB), Time(s) with HED, RCF, and our method.

Methods	Parameters/M	Model Size/MB	Time/s
HED	14.7	112.29	0.3854
RCF	14.8	112.96	0.3903
Our method	14.8	112.97	0.4094

RCF and our network combine all the features of convolutional layer extraction. So HED's parameters and model size are smaller than RCF and our network. Our network

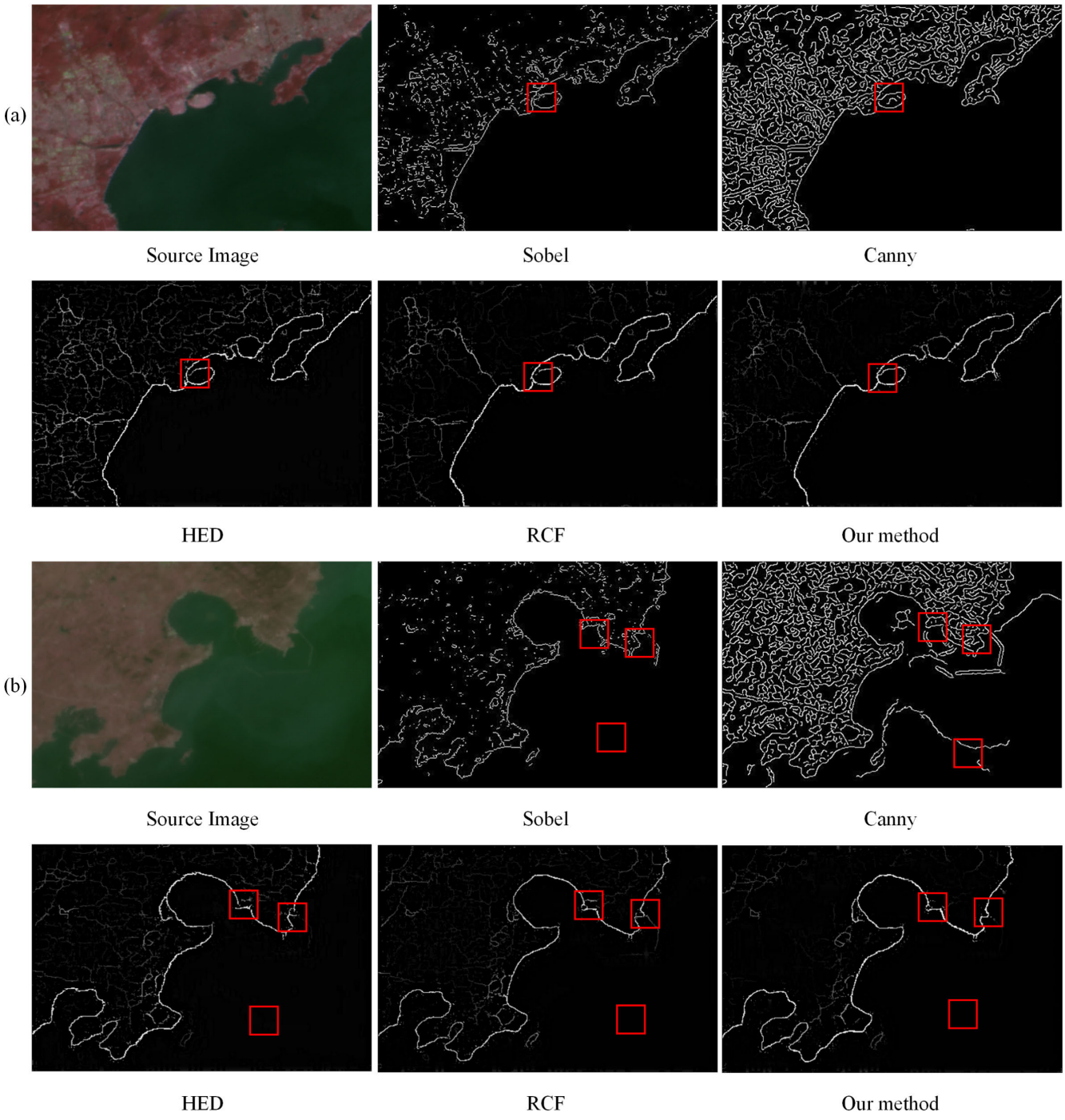


FIGURE 8. Coastline extraction results of Sobel, Canny, HED, RCF, and our method. (a) GF4_PMI remote sensing image with the spatial resolution of 50m on September 1, 2017; (b) GF4_PMI remote sensing image with the spatial resolution of 50m on February 6, 2018.

replaces all the standard convolutions in the backbone network with the Mini-Inception structure, which increases the size of the convolution kernel using dilated convolution. Therefore, compared with RCF, our network does not increase the computational complexity and time cost.

V. MODEL ANALYSIS

For the nature of remote sensing images, we have improved the network, including the backbone network structure and the activation function. Therefore, we need to verify the effectiveness of each part for the final performance.

A. ANALYSIS ON BACKBONE NETWORK STRUCTURE

The role of the Mini-Inception structure is to extract multi-scale features of the same input object simultaneously. Using this structure instead of the standard convolution in the backbone network can better extract the features of remote sensing images. In order to verify the validity of the Mini-Inception

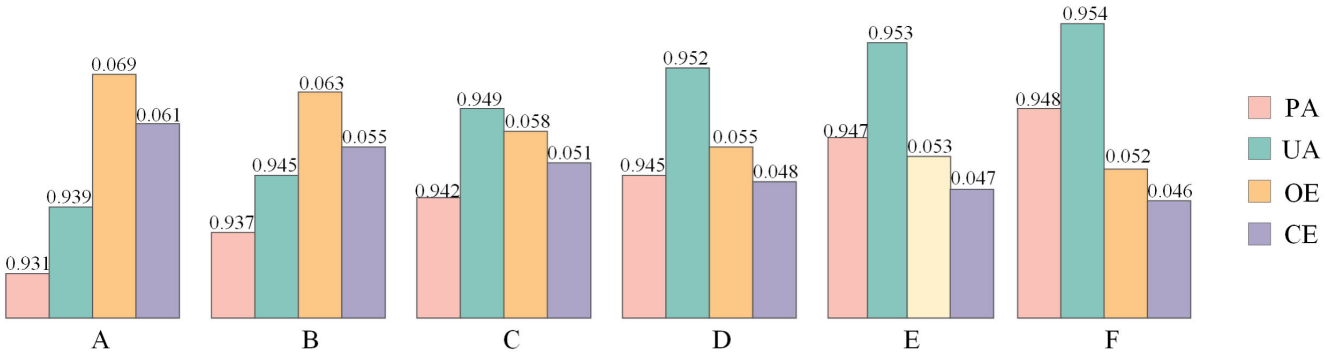


FIGURE 9. The performance (PA/UA/OE/CE) of models A, B, C, D, E, F. PA and UA take the data between 0.92 and 0.96 as the ordinate, OE and CE take the data between 0.04 and 0.07 as the ordinate.

TABLE 3. Comparison of model performance (PA/UA/OE/CE) using the ReLU activation function and the leaky-ReLU activation function.

Methods	PA		UA		OE		CE	
	Test_1	Test_2	Test_1	Test_2	Test_1	Test_2	Test_1	Test_2
Method_relu	0.943	0.919	0.947	0.937	0.057	0.081	0.053	0.063
Our method	0.948	0.926	0.954	0.942	0.052	0.074	0.046	0.058

structure and the design of the backbone network structure is optimal, a set of comparative experiments is done. On the one hand, the backbone network does not use the Mini-Inception structure, but only uses the standard 3×3 accumulation, which is recorded as model A. On the other hand, starting from stage5, the five stages add the Mini-Inception structure stage by stage, which is recorded as model B, C, D, E, and F respectively, and model F is our final network structure. The evaluation results of the six models on Test_1 are shown in Figure 9.

As shown in Figure 9, from model A to model F, PA and UA have been increasing, and the deeper layer uses the Mini-Inception structure, the more precision is improved. It is proved that the use of the Mini-Inception structure is always beneficial to the result. It is optimal to use the Mini-Inception structure in all the convolutional layers in the backbone network instead of the standard 3×3 convolution layer, which is beneficial to the improvement of coastline extraction accuracy.

B. ANALYSIS ON ACTIVATION FUNCTION

In order to verify that the leaky-ReLU activation function has a positive impact on the final results of the model, a set of comparative experiments was performed. On the one hand, based on our network but using the ReLU activation function training model, it is recorded as Method_relu. On the other hand, we use the leaky-ReLU activation function, which is our final network model. The test results on Test_1 and Test_2 are shown in Table 3.

As can be seen from Table 3, both PA and UA are higher than the ReLU activation function when using the leaky-ReLU activation function, and both OE and CE are lower than using the ReLU activation function. It can be proved that the leaky-ReLU activation function has a positive impact on the final results of the model, which can improve network performance and improve the accuracy of coastline extraction.

VI. CONCLUSION

In this paper, we propose a new convolutional neural network structure for coastline extraction of remote sensing images. Our network can better extract the features of remote sensing images, and make full use of all convolution features and multi-level information of objects to achieve image-to-image prediction. We conduct a comparative experiment on the remote sensing images of Jiaozhou Bay in Qingdao. The experimental results show that the proposed method has some advantages compared with other methods, which can improve the accuracy of coastline extraction, and does not significantly increase the computational cost and reasoning time. Besides, our network has high portability, and it can also be used for the pixel-level sea-land segmentation task of remote sensing images.

REFERENCES

[1] C. Kuenzer, M. Ottinger, G. Liu, B. Sun, R. Baumhauer, and S. Dech, "Earth observation-based coastal zone monitoring of the yellow river delta: Dynamics in China's second largest oil producing region over four decades," *Appl. Geography*, vol. 55, Dec. 2014, pp. 92–107.

[2] W. Li and P. Gong, "Continuous monitoring of coastline dynamics in western Florida with a 30-year time series of landsat imagery," *Remote Sens. Environ.*, vol. 179, pp. 196–209, Jun. 2016.

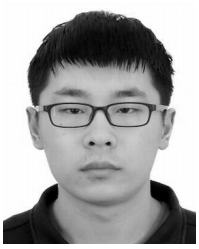
[3] A. Rasuly, R. Naghdifar, and M. Rasoli, "Monitoring of caspian sea coastline changes using object-oriented techniques," *Procedia Environ. Sci.*, vol. 2, no. 10, pp. 416–426, 2010, doi: 10.1016/j.proenv.2010.10.046.

[4] M. Li and X. C. Zheng, "A second modified normalized difference water index (SMNDWI) in the case of extracting the shoreline," *Mar. Sci. Bull.*, vol. 18, no. 3, 2016, pp. 15–27.

[5] S. K. Mcfeeter, "The use of the normalized difference water index (NDWI) in the delineation of open water features," *Int. J. Remote Sens.*, vol. 17, no. 7, pp. 1425–1432, 1996.
[6] K. Cao, J. Fan, X. Wang, X. Wang, J. Zhao, and F. Zhang, "Coastline automatic detection based on high resolution SAR images," in *Proc. 4th Int. Workshop Earth Observ. Remote Sens. Appl. (EORSA)*, Guangzhou, China, Jul. 2016, pp. 43–46, doi: 10.1109/EORSA.2016.7552763.
[7] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *Int. J. Comput. Vis.*, vol. 1, no. 4, pp. 321–331, 1988.
[8] J. Fan, K. Cao, J. Zhao, D. Jiang, and X. Tang, "A hybrid particle swarm optimization algorithm for coastline SAR image automatic detection," in *Proc. 12th World Congr. Intell. Control Automat. (WCICA)*, Guilin, China, Jun. 2016, pp. 822–825, doi: 10.1109/WCICA.2016.7578256.
[9] K. White and H. M. El Asmar, "Monitoring changing position of coastlines using thematic mapper imagery, an example from the Nile Delta," *Geomorphology*, vol. 29, nos. 1–2, pp. 93–105, Aug. 1999.
[10] M. H. Xie, Y. F. Zhang, and K. Fu, "Algorithm of detection coastline from SAR images based on seeds growing," *J. Graduate School Chin. Acad. Sci.*, vol. 24, no. 1, pp. 93–98, 2007.
[11] L. J. Wang, Z. Niu, D. G. Zhao, Y. X. Li, Z. B. Wang, and D. Kuang, "The study of coastline extraction and validation using ETM remote sensing image," *Remote Sens. Technol. Appl.*, vol. 25, no. 2, pp. 235–239, Apr. 2010.
[12] K. G. Karantzalos, D. Argialas, and A. Georgopoulos, "Towards automatic detection of coastlines from satellite imagery," in *Proc. 14th Int. Conf. Digit. Signal Process.*, Santorini, Greece, vol. 2, Jul. 2002, pp. 897–900.
[13] T. Asaka, Y. Yamamoto, S. Aoyama, K. Iwashita, and K. Kudou, "Automated method for tracing shorelines in L-band SAR images," in *Proc. Conf. Asia–Pacific Conf. Synth. Aperture Radar (APSAR)*, Tsukuba, Japan, Sep. 2013, pp. 325–328.
[14] D. Huiying, X. Peng, L. Qian, and X. Hongli, "The water coastline detection approaches based on USV vision," in *Proc. IEEE Int. Conf. Cyber Technol. Automat., Control, Intell. Syst. (CYBER)*, Shenyang, China, Jun. 2015, pp. 404–408.
[15] S. Xie and Z. Tu, "Holistically-nested edge detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 1395–1403.
[16] Z. Yu, C. Feng, M.-Y. Liu, and S. Ramalingam, "Casenet: Deep category-aware semantic edge detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 5964–5973.
[17] Y. Liu, M. M. Cheng, X. Hu, K. Wang, and X. Bai, "Richer convolutional features for edge detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 3000–3009.
[18] Q. Ha, K. Watanabe, T. Karasawa, Y. Ushiku, and T. Harada, "MFNet: Towards real-time semantic segmentation for autonomous vehicles with multi-spectral scenes," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Vancouver, BC, Canada, Sep. 2017, pp. 5108–5115, doi: 10.1109/IROS.2017.8206396.
[19] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *Proc. ICML*, Jun. 2013, vol. 30, no. 1, p. 3.
[20] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proc. 14th Int. Conf. Artif. Intell. Statist.*, Jun. 2011, pp. 315–323.
[21] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: https://arxiv.org/abs/1409.1556
[22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[23] N. H. Kaplan and I. Erer, "Bilateral filtering-based enhanced pansharpening of multispectral satellite images," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 11, pp. 1941–1945, Nov. 2014.
[24] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," 2015, *arXiv:1511.07122*. [Online]. Available: https://arxiv.org/abs/1511.07122
[25] P. Arbeláez, M. Maire, C. Fowlkes, and J. Malik, "Contour detection and hierarchical image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 5, pp. 898–916, May 2011.
[26] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986, doi: 10.1109/TPAMI.1986.4767851.
[27] Y. Tian, R. S. Jia, S. H. Xu, R. Hua, and M. D. Deng, "Super-resolution reconstruction of remote sensing images based on convolutional neural network," *J. Appl. Remote Sens.*, vol. 13, no. 4, 2019, pp. 1–12.
[28] C. Chen, J. Fu, S. Zhang, and X. Zhao, "Coastline information extraction based on the tasseled cap transformation of Landsat-8 OLI images," *Estuarine, Coastal Shelf Sci.*, vol. 217, pp. 281–291, Feb. 2019.
[29] S. Gupta and S. G. Mazumdar, "Sobel edge detection algorithm," *Int. J. Comput. Sci. Manage. Res.*, vol. 2, no. 2, pp. 1578–1583, Feb. 2013.
[30] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.

XIAO-YING LIU was born in Shandong, China, in 1995. She received the B.S. degree from the Shandong University of Science and Technology, China, in 2018, where she is currently pursuing the M.S. degree. Her research interests include image processing and deep learning.

RUI-SHENG JIA is currently a Professor with the College of Computer Science and Engineering, Shandong University of Science and Technology, China. He has more than 30 first author publications and has more than 25 coauthor publications. His research interests include artificial intelligence, big data processing, information fusion, and micro seismic monitoring and inversion.



QING-MING LIU was born in Shandong, China, in 1995. He received the B.S. degree from the Shandong University of Science and Technology, China, in 2018, where he is currently pursuing the M.S. degree. His research interests include image processing and deep learning.

CHAO-YUE ZHAO was born in Shandong, China, in 1996. She received the B. S. degree from Shandong Women's University, China, in 2018. She is currently pursuing the M.S degree with the Shandong University of Science and Technology. Her research interests include image processing and deep learning.

HONG-MEI SUN received the B.S. and M.S. degrees in computer science from the Shandong University of Science and Technology, China, in 1995 and 2005, respectively. She is currently a Lecturer with the College of Computer Science and Engineering, Shandong University of Science and Technology. She is also the Leader of the Key Research and Development Projects of Shandong Province, China. She has four first author publications and has five coauthor publications. Her research interests include micro seismic monitoring technology and software engineering.