Hindawi
International Journal of Antennas and Propagation
Volume 2022, Article ID 3063965, 20 pages
https://doi.org/10.1155/2022/3063965

Research Article

Gaussian Function Fusing Fully Convolutional Network and Region Proposal-Based Network for Ship Target Detection in SAR Images

Peipei Zhang[1], **Guokun Xie**[1] and **Jinsong Zhang**[2]

[1]*ZTE Communication Institute, Xi'an Traffic Engineering Institute, Xi'an 710300, China*
[2]*National Laboratory of Radar Signal Processing, Xidian University, Xi'an 710071, China*

Correspondence should be addressed to Peipei Zhang; zhangpeipei698@163.com

Received 21 January 2022; Accepted 6 May 2022; Published 27 May 2022

Academic Editor: Stefano Selleri

Recently, ship target detection in Synthetic aperture radar (SAR) images has become one of the current research hotspots and plays an important role in the real-time detection of sea regions. The traditional SAR ship detection methods usually consist of two modules, one module named land-sea segmentation for removing the complicated land regions, and one module named ship target detection for realizing fine ship detection. An algorithm combining the Unet-based land-sea segmentation method and improved Faster RCNN-based ship detection method is proposed in this paper. The residual convolution module is introduced into the Unet structure to deepen the network level and improve the feature representation ability. The K-means method is introduced in the Faster RCNN method to cluster the size and aspect ratio of ship targets, to improve the anchor frame design, and make it more suitable for our ship detection task. Meanwhile, a fine detection algorithm uses the Gaussian function to fuse the confidence value of sea-land segmentation results and the coarse detection results. The segmentation and detection results on the established segmentation dataset and detection dataset, respectively, demonstrate the effectiveness of our proposed segmentation methods and detection methods.

1. Introduction

Synthetic aperture radar (SAR) is an active microwave remote sensing sensor [1, 2]. Compared with optical sensors, it has the capabilities of all-day, all-weather, multi-angle, and long-distance monitoring [3–5]. SAR has been widely used in civil fields such as marine rescue, law enforcement, and other fields in marine real-time monitoring and detection [6–9]. With the continuous development of high-resolution SAR imaging technology, a large number of SAR images can be used for marine ship detection [10, 11]. Meanwhile, ship target detection in SAR images has become one of the current research hotspots and plays an important role in the real-time detection of sea areas. Therefore, it is of great significance to study ship target detection algorithms in SAR images [11–14].

At present, many scholars have studied ship target detection in SAR images. An effective method is to divide the detection process into two steps [14–16]: the first step is land-sea segmentation, and the second step is ship target detection, as shown in Figure 1. Since the ship target to be detected in the SAR image is not necessarily in the pure sea background, and the gray value of the land part in the SAR images is very close to the ship target, it is necessary to remove the land part to avoid a large number of false targets on the land in the subsequent detection process [4, 17].

In the research of SAR image sea-land segmentation, there are lots of methods based on threshold [18, 19] and its improvement, such as the Otsu algorithm [20]. According to the difference of gray characteristics between target and background, a threshold is calculated to maximize the variance between these regions, and the best threshold is

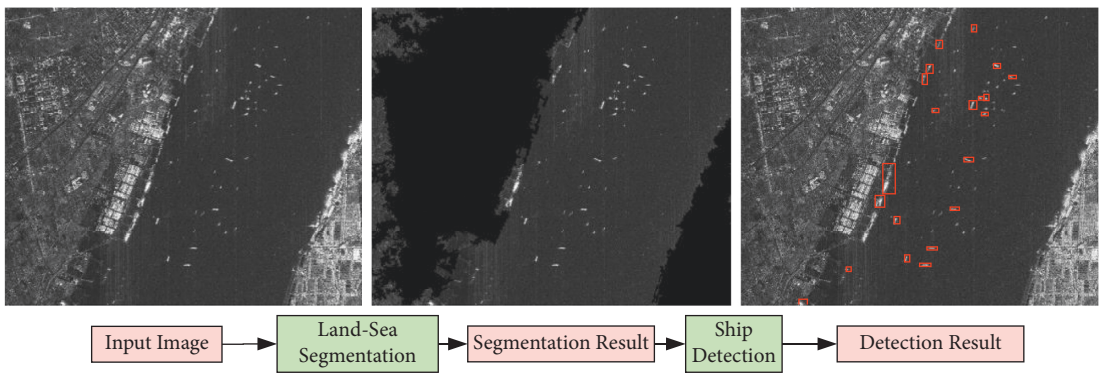
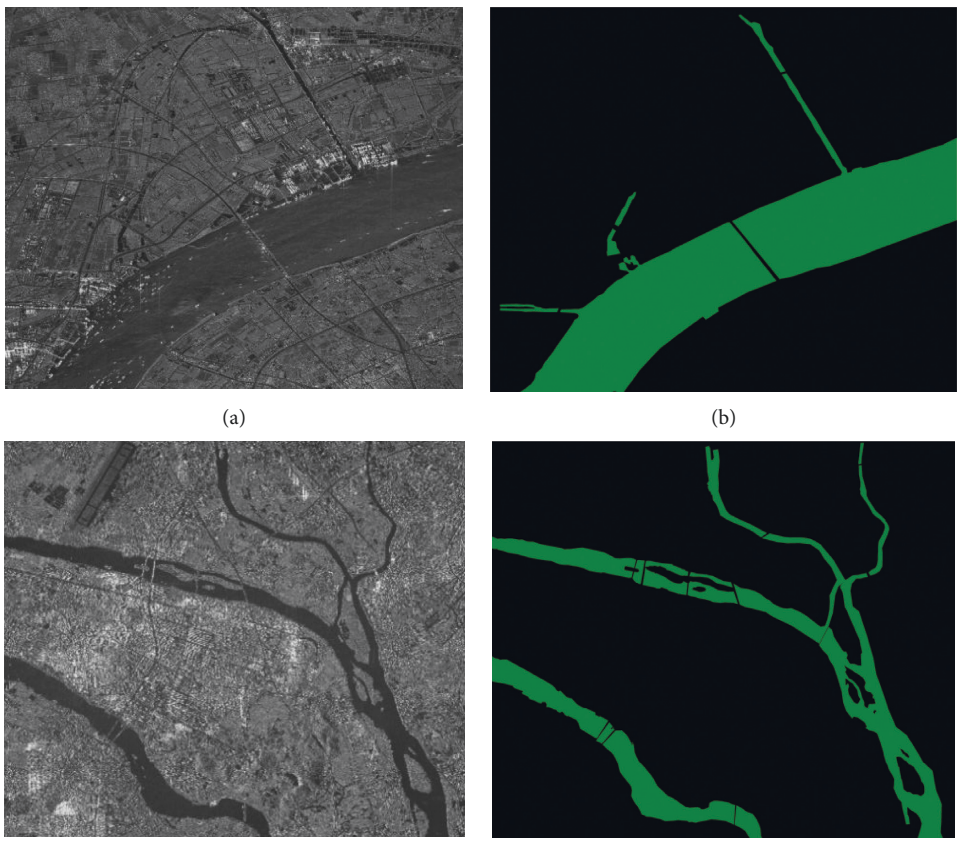


Figure 1: The flowchart of the usual ship detection method.

used for image binarization to realize sea-land segmentation [20, 21]. Although this method is simple and easy to operate, the segmentation accuracy is relatively poor under strong sea clutter and complicated land background. There are also some methods based on edge detection and its improvement [18, 22, 23]. Isikdogan et al. [24]proposed the average ratio method to connect the edge points detected in the SAR image into a closed curve to form the region to be segmented. There is also Markov random field (MRF) segmentation methods based on Bayesian theory [19]. These contour feature–based segmentation methods make use of the local structures of SAR images. Although the segmentation performance has been improved by these methods, their complex computation leads to slow speed. Sea-land segmentation in SAR image is essentially a special form of semantic segmentation. Inspired by the great success of deep learning in target classification, there have been many types of research using deep learning for semantic segmentation [7, 8, 25, 26]. With the continuous development of deep learning, there are many segmentation models with better effects, which are based on Fully Convolutional Network (FCN) [27]. FCNs replace the fully connected layer in traditional CNNs with convolution layers and pooling layers. This strategy can significantly decrease the computation complexity and improve the segmentation accuracy. Inspired by FCN, Unet [22, 28] proposes the typical encoding-decoding segmentation structure and introduces skip connection between encoding and decoding modules and further combines high-dimensional and low-dimensional features which can improve the segmentation performance, especially in medical segmentation. In this paper, considering the inadequate connection between adjacent layers, the residual connection is introduced into the encoding-decoding module of Unet to improve the accuracy of sea-land segmentation.

For the module of ship target detection, there are many detection methods; the main point is to directly detect ships in the SAR image. The most classic ship target detection method is the constant false alarm rate (CFAR) algorithm [29–31]. CFAR was first proposed by Finn in 1996, and then scholars from all over the world put forward a series of improved algorithms, such as the two-parameter CFAR algorithm [15, 30]. CFAR algorithm is now the most widely used algorithm in the field of SAR image target detection. For different sea clutter situations, researchers have proposed CFAR algorithms based on different clutter statistical models, including Gaussian distribution model, Rayleigh distribution model, lognormal distribution model, Weibull distribution model, K distribution model, and G0 distribution model. In recent years, due to the development of deep learning in the field of target detection, some researchers have proposed a ship target detection method based on deep learning in SAR images [10, 12, 32–35]. These detection algorithms are mainly divided into two categories. The first is a two-stage target detection algorithm represented by Fast RCNN (region CNN) [17, 33, 36]. In these type of detection methods, the selective search method or the region proposal network firstly generates the candidate target bounding boxes, then the classification network classifies the extracted target boxes as target or background. Based on target localization and target classification, the detection accuracy of two-stage detection methods is high. However, too much convolution layers also lead low computation efficiency. The second methods are the single-stage target detection algorithm, mainly represented by SSD (single shot multibox detector) [37–40] and YOLO (you only look once). This kind of algorithm does not need to generate candidate regions, but directly uses the regression method for target detection, and takes into account both detection efficiency and accuracy. Especially, Li et al. [41] proposed ship target detection in SAR images based on generated countermeasure network. These deep-learning-based methods have achieved good performance in ship detection. However, if the datasets cannot meet the requirements, the detection effect of these algorithms will become worse when they encounter SAR images that are quite different from the scene of the dataset [38, 42]. In this paper, considering the special sample distribution in our ship detection dataset, the K-means is introduced into the two-stage detection method to realize accurate localization of ship targets.

This paper introduces the deep-learning-based detection method into large-scale SAR ship detection. Through analyzing the detection results of SAR images within land and sea regions, we find that the false alarm rate of detection is too high due to the existence of land background. To overcome this issue, an algorithm combining the Unet-based [19] land-sea segmentation method and Faster RCNN-based [9] ship detection method is proposed in this



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

Figure 2: Images and corresponding ground truths in the segmentation dataset.

paper. More specifically, we first introduce the residual convolution module into the encoding module and decoding module of the Unet structure. With this residual module, the high-dimensional feature can be extracted from input SAR images and the clearer water-land boundary information can be retained in the corresponding segmentation results. Then, according to the characteristics of ship targets in the dataset, K-means is introduced to cluster their size and length-width ratio. Based on the clustering results, an anchor frame suitable for SAR ship targets in SAR images is designed for a Faster RCNN detection framework. The feature extraction module, region generation network, recognition module, and loss function of our detection network is introduced in detail. Finally, according to the results of land-sea segmentation, the sea confidence value modeled by Gaussian function weight is established and introduced into our detection frame, and the detection result is further improved. The results on the sea-land segmentation dataset, ship detection dataset, and large-scene SAR images demonstrate the effectiveness of our ship detection method.

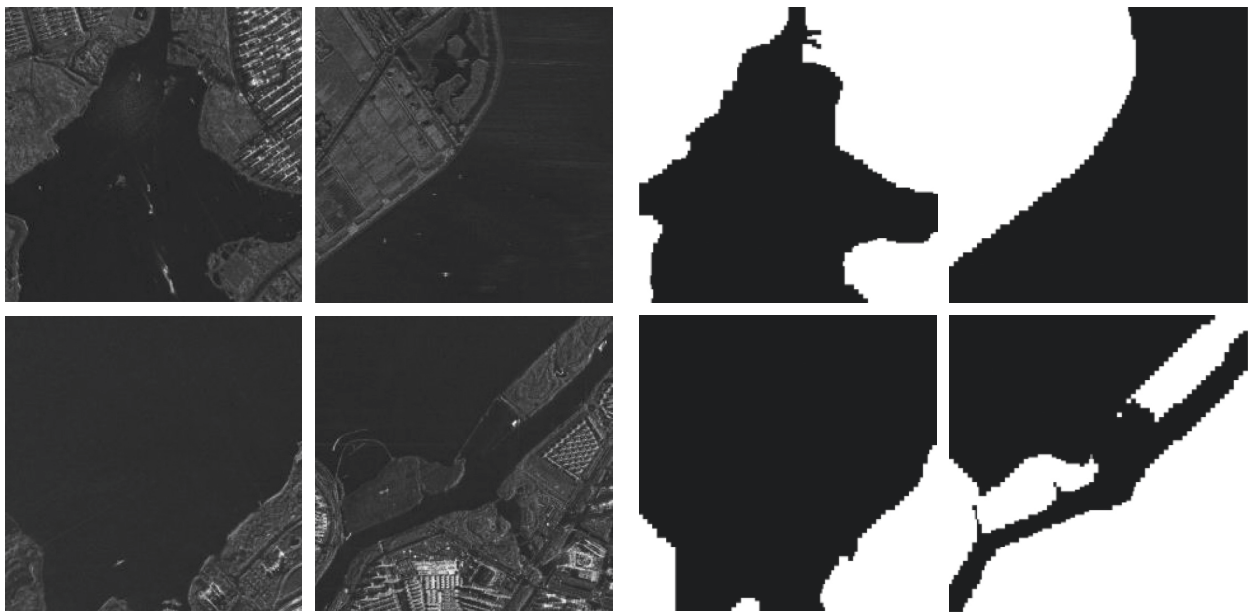
This paper begins in Section 2 which shows the results of our sea-land segmentation dataset and ship detection dataset. In Section 3, the sea-land segmentation method and detection method are explained. The experimental results and corresponding analyses are described in Section 4. Section 5 concludes this paper.

2. Dataset

The SAR sea-land segmentation dataset and ship detection dataset are shown as follows

2.1. Sea-Land Segmentation Dataset. In this paper, we select 5 high-resolution and large-scene Gaofen-3 SAR images to form the land-sea segmentation dataset. Gaofen-3 SAR system is a C-band multi-mode SAR satellite of China with multi-polarization [4]. These selected images acquired with spotlight mode and HH/HV polarization belong to the second-level data of the Gaofen-3 system which is processed through radiometric calibration and geometric rectification. The two dimension resolution of these images is $1\,\text{m} \times 1\,\text{m}$. Two Gaofen-3 images used in our dataset are shown in Figures 2(a) and 2(c). These images include complex land backgrounds and large-scale water regions. Meanwhile, some ships are sparsely distributed in the water regions. Using these high-resolution images to train the segmentation model, we believe the well-trained model can effectively discriminate between water regions and land regions.

We use the Labelme annotation tool to mask the SAR images to get the masked ground truth. It is noticeable that we only label the water regions with a large area as the ships are usually driving or moored at open large-scale water regions. The annotated labels corresponding to Figures 2(a)



(a) (b)

Figure 3: Training data and ground truths in sea-land segmentation dataset. (a) Training images. (b) Ground Truths.

and 2(c) are shown in Figures 2(b) and 2(d). One can see that the annotated labels can well represent the true water regions and background. Since the large size of these images is not conducive to directly use in subsequent network training, it is necessary to preprocess the original large image. First, downsampling the original large image to reduce the resolution to 1/4 of the original, and then clipping it according to the standard image size without overlapping, and filling the insufficient size with 0 pixels directly. After clipping, the dataset is augmented by rotation, left-right, and up-down flipping, and the labeled images are synchronously operated. Finally, 14216 small images and their labels with size 512×512 pixels are obtained, which completes the production of the dataset. The dataset is used for training (12796) and testing (1420) at a ratio of 9 : 1. Four training images and their corresponding annotated labels are shown in Figure 3. In the labeled image, 255 pixels represent land while 0 pixel represents the sea.

2.2. Ship Detection Dataset. In this paper, 14 large-scene spaceborne SAR images are taken as the source of ship detection dataset while 5 images come from the above land-sea segmentation dataset. All these images with high-resolution of 1–3 m contain different ship targets. Two images of them are shown in Figure 4. One can see that there are lots of ships of various shapes and sizes in the open water region. Since the reflection energy of ship targets is larger than that of water regions and other backgrounds, the intensity of SAR ships is usually higher than other backgrounds. However, the strong sea clutter in the sea region and the complex building in the land background with high intensity can also lead to a high false alarm in the detection results. Thus, using these images to construct the ship detection dataset, the well-trained detection model will have a good detection performance and robustness to the real complicated ship detection task.

Similar to the land-sea segmentation dataset, the size of the original image in the ship detection dataset is too large to be directly sent into the detection model. Thus, we still crop the region containing the ship targets into small SAR patches with size of 600×600 pixels. In the process of clipping, the multi-scale clipping method is used to get different sizes of small images, to obtain more feature information in the training process. The small images are flipped up and down, left and right to expand the data, and a total of 1242 images are obtained. The training set, verification set, and test set are divided by 7 : 2 : 1. To train the ship target detection network, the dataset of ship target is made with the VOC2007 format standard. A few cropped images are shown in Figure 5.

3. Proposed Method

As shown in Figure 6, the proposed method consists of three modules, the residual structure-based Unet for sea-land segmentation, the K-means-based Faster RCNN for coarse ship detection, and the confidence weighting with Gaussian function for final ship detection. A detailed description of these modules is shown as follows.

3.1. Residual Structure-Based Unet for Sea-Land Segmentation. Although the original Unet structure [28] transmits the high-resolution features of the encoding module, the high-resolution edge information before pooling operation does not pass through any convolution layer with the skip connection [4, 43]. Thus, the learned high-resolution edge information by above skip connection does not contain enough high-resolution edge information of the input image. Based on the Unet network structure, the

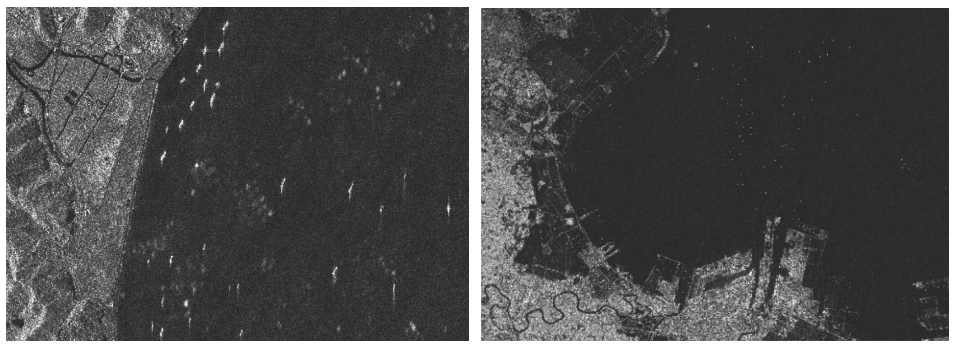


Figure 4: Two images in our ship detection dataset.

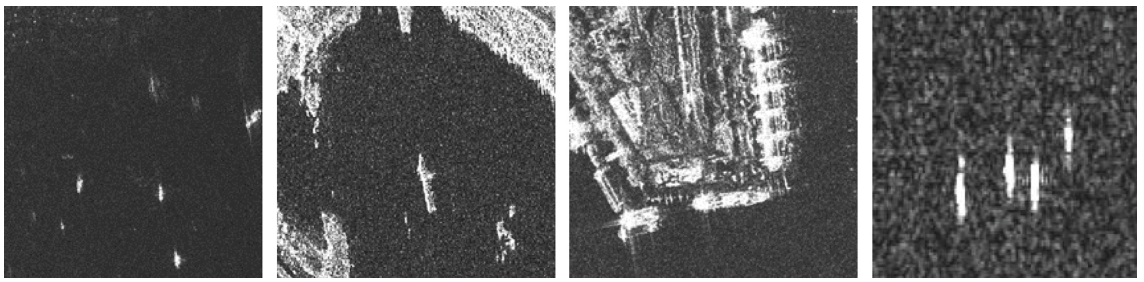


Figure 5: A few images in our ship detection dataset.

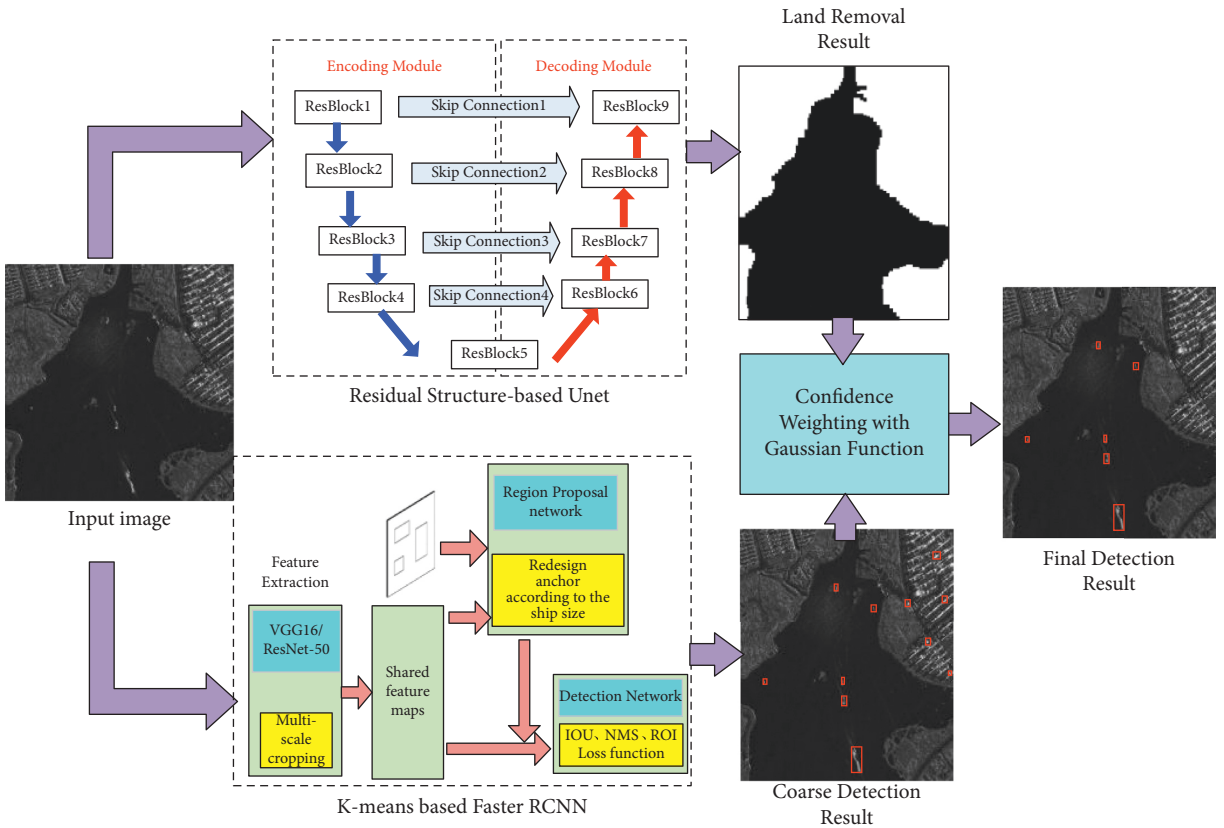


Figure 6: Flow chart of the proposed method.

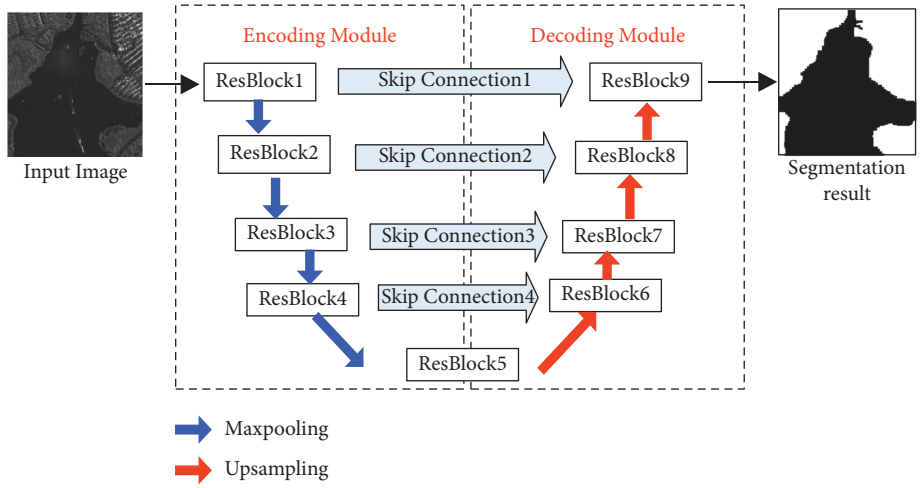


Figure 7: The network architecture of land-sea segmentation.

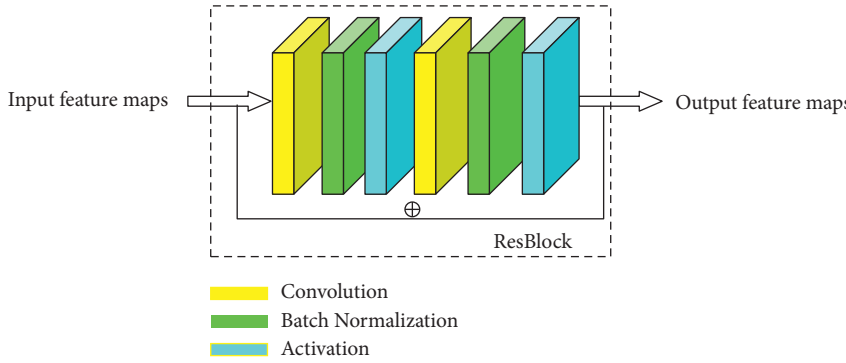


Figure 8: Residual blocks in segmentation network.

main body of our land-sea segmentation network is divided into two modules: encoding module and decoding module. The residual network is introduced to the original Unet network to form the residual convolution module to deepen the network level. Meanwhile, the residual network has the advantage of avoiding the deepening network gradient vanishing [44]. With the above improvements, the architecture of the sea-land segmentation network based on our improved Unet structure is shown in Figure 7.

In Figure 7, the encoding module is composed of multiple residual convolution modules [45] and max-pooling layers. With the pooling layers in the encoding module, the size of the feature map decreases. The decoding module is composed of residual convolution modules and bilinear interpolation upsampling. The upsampling layers recover the size of the feature map to the same size as the input image through layer-by-layer upsampling. Through the skip connection, the size of each upsampling result and the size of corresponding encoding feature maps are spliced. Thus, the high-resolution low-level features are connected to the decoding module, and the high-level information and low-level information are fused to improve the segmentation accuracy. The final output segmentation layer is implemented by the convolution layer. Since the image only contains two categories of sea and land, the sigmoid function can be used to complete the pixel classification. The detailed structure of the segmentation network is shown as follows.

3.1.1. The Architecture of the Residual Block. The structure of the residual block in our land-sea segmentation network is shown in Figure 8. Each residual block contains two convolution layers with kernel size 3×3. After each convolution layer, the batch normalization (BN) [45] operation is carried out and the activation function is used to activate the residual convolution block. What is more, the shortcut connection is realized by using one convolution layer and adding it to the output feature maps to get more abundant combination information.

3.1.2. Encoding Module. The encoding module of our segmentation network is shown in Table 1, which contains five residual blocks and four downsampling layers. Given the input SAR image with size $512 \times 512 \times 1$ pixels, these residual blocks and pooling layers can automatically extract high-dimensional features from the input image. The kernel size of all convolution layers in the main network is 3×3 while that in the shortcut network is 1×1. The downsampling is completed by the max-pooling with the kernel size and step size of 2. The size of output feature maps of different

Table 1: The structure of the encoding module in our segmentation network.

	Input $512 \times 512 \times 1$		Output feature maps
Resblock1	Main	Conv $(3 \times 3 \times 32)$/BN/ReLU Conv $(3 \times 3 \times 32)$/BN/ReLU	$512 \times 512 \times 32$
	Shortcut	Conv $(1 \times 1 \times 32)$	
	Max-pooling (2×2)		$256 \times 256 \times 32$
Resblock2	Main	Conv $(3 \times 3 \times 64)$/BN/ReLU Conv $(3 \times 3 \times 64)$/BN/ReLU	$256 \times 256 \times 64$
	Shortcut	Conv $(1 \times 1 \times 64)$	
	Max-pooling (2×2)		$128 \times 128 \times 64$
Resblock3	Main	Conv $(3 \times 3 \times 128)$/BN/ReLU Conv $(3 \times 3 \times 128)$/BN/ReLU	$128 \times 128 \times 128$
	Shortcut	Conv $(1 \times 1 \times 128)$	
	Max-pooling (2×2)		$64 \times 64 \times 128$
Resblock4	Main	Conv $(3 \times 3 \times 256)$/BN/ReLU Conv $(3 \times 3 \times 256)$/BN/ReLU	$64 \times 64 \times 256$
	Shortcut	Conv $(1 \times 1 \times 256)$	
	Max-pooling (2×2)		$32 \times 32 \times 256$
Resblock5	Main	Conv $(3 \times 3 \times 512)$/BN/ReLU Conv $(3 \times 3 \times 512)$/BN/ReLU	$32 \times 32 \times 512$
	Shortcut	Conv $(1 \times 1 \times 512)$	

Table 2: The structure of the decoding module in our segmentation network.

	UpSampling2D(2, 2), Resblock5 Skip connection4		Concatenate
Resblock6	Main	Conv2D $(3 \times 3 \times 256)$/BN/ReLU Conv2D $(3 \times 3 \times 256)$/BN/ReLU	Add $64 \times 64 \times 256$
	Shortcut	Conv2D $(1 \times 1 \times 256)$	
	UpSampling2D (2, 2) Skip connection3		Concatenate
Resblock7	Main	Conv2D $(3 \times 3 \times 128)$/BN/ReLU Conv2D $(3 \times 3 \times 128)$/BN/ReLU	Add $128 \times 128 \times 128$
	Shortcut	Conv2D $(1 \times 1 \times 128)$	
	UpSampling2D (2, 2) Skip connection2		Concatenate
Resblock8	Main	Conv2D $(3 \times 3 \times 64)$/BN/ReLU Conv2D $(3 \times 3 \times 64)$/BN/ReLU	Add $256 \times 256 \times 64$
	Shortcut	Conv2D $(1 \times 1 \times 64)$	
	UpSampling2D (2, 2) Skip connection1		Concatenate
Resblock9	Main	Conv2D $(3 \times 3 \times 32)$/BN/ReLU Conv2D $(3 \times 3 \times 32)$/BN/ReLU	Add $512 \times 512 \times 32$
	Shortcut	Conv2D $(1 \times 1 \times 32)$	
	Conv2D $(1 \times 1 \times 1)$/sigmoid $512 \times 512 \times 1$		

operations is shown in the right column of Table 1. With the feature extraction in the encoding module, 512 feature maps are obtained, and the size of the final feature map is reduced to 32×32, which is 1/16 of the size of the original input image.

3.1.3. Decoding Module. The structure of the decoding module of our land-sea segmentation network is shown in Table 2, which contains four residual blocks and four upsampling layers. With layer-wise upsampling, the size of the final feature map is restored to the same size as the input. Then, the dimension concatenation operation is performed between each upsampling result and the corresponding size of the encoding feature maps by skip connection. The concatenated feature maps are used as the input of the next residual block in the decoding module.

The bilinear interpolation [22, 46]is used to upsample feature maps, which is a commonly used upsampling method in semantic segmentation. Assuming the value of $p_{11}(x_1, y_1)$, $p_{12}(x_1, y_2)$, $p_{21}(x_2, y_1)$ and $p_{22}(x_2, y_2)$ in feature maps, F is known, the value of any points in this feature map can be computed as

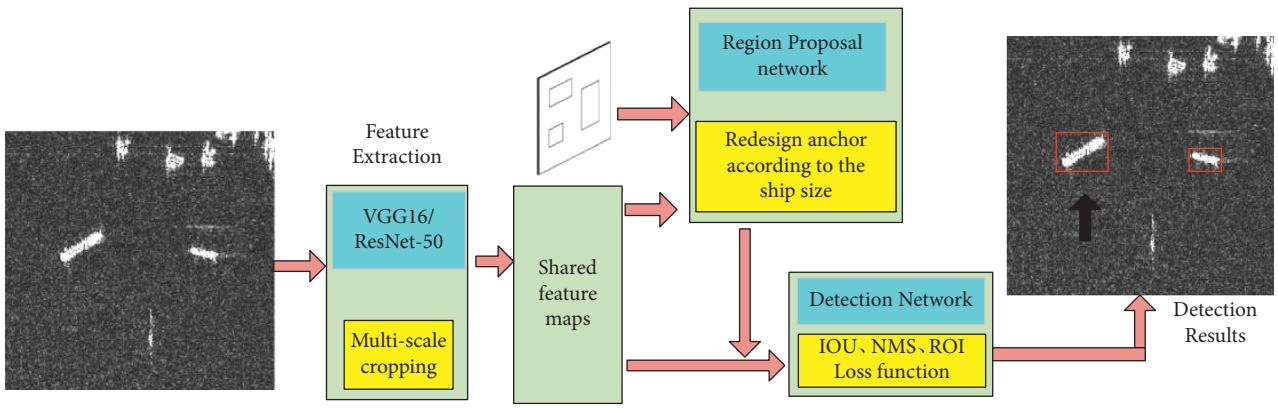


Figure 9: Ship detection network architecture.

$$F(q) \approx \frac{(x_2 - x)(y_2 - y)}{(x_2 - x_1)(y_2 - y_1)} F(p_{11}) + \frac{(x_2 - x)(y - y_1)}{(x_2 - x_1)(y_2 - y_1)} F(p_{12}) + \frac{(x - x_1)(y_2 - y)}{(x_2 - x_1)(y_2 - y_1)} F(p_{21}) + \frac{(x - x_1)(y - y_1)}{(x_2 - x_1)(y_2 - y_1)} F(p_{22}). \quad (1)$$

3.1.4. Network Partition Layer and Loss Function. The last layer of the segmentation network completes the pixel-by-pixel classification of the input image and outputs the probability of its mapping to each category, where the highest probability means its final category. Through a series of operations, such as encoding module, decoding module, and skip connection, the proposed network extracts sufficient feature information of the input image. To ensure that the output and input size of the last convolution layer in the network are the same, the convolution kernel with the size of 1 and the depth of 1 is used, and then the activation operation is carried out to complete the pixel by pixel classification. Because the sea-land segmentation network is a kind of binary classification, the sigmoid activation function is selected in the final output layer to map the feature map value to 0–1 [4, 28]. It is noticeable that the pixel with confidence higher than 0.5 is decided as land region and the pixels with confidence lower than 0.5 is decided as water region.

It is noticeable that the sea-land segmentation task only predicts "sea" and "land" categories. Assuming the size of the input image is $W \times H$, $t_{(w,h)}$ means that the label of the pixel (w, h) in the input image is "sea" while $f(z_{(w,h)})$ is the probability that the pixel (w, h) in the image is predicted to be "sea", the number of samples in each training is N. The cross-entropy function is used as the loss function

$$C = -\sum_{w=1,h=1}^{W,H} \sum_{n=1}^{N} t_{(w,h)} \log\left(f\left(z_{(w,h)}\right)\right) + \left(1 - t_{(w,h)}\right)\left(1 - \log\left(f\left(z_{(w,h)}\right)\right)\right). \quad (2)$$

Using the land-sea segmentation dataset and our segmentation network, we can train a good segmentation model. The detailed parameter setting and training process are shown in experimental results.

3.2. K-Means Based Faster RCNN for Coarse Ship Detection. Through the ship target detection in SAR images, we can obtain the distribution of ships in the sea, port area, and even the complex sea-land junction area, which provides strong technical support for marine monitoring tasks. Nowadays, the CFAR algorithm is widely used in ship target detection based on SAR images. It needs to fit the sea clutter model according to different sea conditions, and it is greatly affected by the detection scene, noise, and other factors [15, 30, 33]. Some parameters of the CFAR detection algorithm are adjusted according to the specific scene, which makes the algorithm unadaptive.

SAR images are different from optical images in that the ship target in SAR image exists in the form of a point target, occupying only a small number of pixels. At this time, if the Faster RCNN algorithm is directly used for ship detection in SAR images, the detection effect is not very ideal. Therefore, it is necessary to improve the network structure by combining the characteristics of SAR images and build a ship detection network based on region extraction following SAR images. Figure 9 shows the overall architecture of our ship detection network proposed in this paper. K-means is used to aggregate the size and length-width ratio of ships, and the anchor frame suitable for the ship targets in SAR images is designed. Meanwhile, the IOU threshold in the non-maximum suppression (NMS) algorithm is also modified according to the location accuracy of our SAR ship detection dataset. The detailed description of the feature extraction module, region-proposal extraction network, recognition network, and loss function is shown as follows.

3.2.1. Feature Extraction, Region Proposal Network, and Classification Network. There are lots of widely used feature extraction networks, e.g., VGGNet [47], ResNet [48], and Xception [49]. Compared with the most typical VGGNet, ResNet is deeper and its computational complexity is lower than VGGNet, which speeds up the training speed of this network. Meanwhile, ResNet also has good adaptability to different tasks, like target recognition and object detection. According to the above analysis, the ResNet is taken as the feature extraction network in our detection model. The ResNet with different configurations is shown in Table 3.

Table 3: The structures of different ResNet models.

name	Output size	18-layer	34-layer	50-layer	101-layer	152-layer
Conv 1	112 × 112	7×7, 64, stride = 2				
		3×3, max pooling, stride = 2				
Block 1	56 × 56	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
Block 2	28 × 28	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 8$
Block 3	14 × 14	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 23$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 36$
Block 4	7 × 7	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$
	1 × 1	Average pool, 1000-d fc, softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

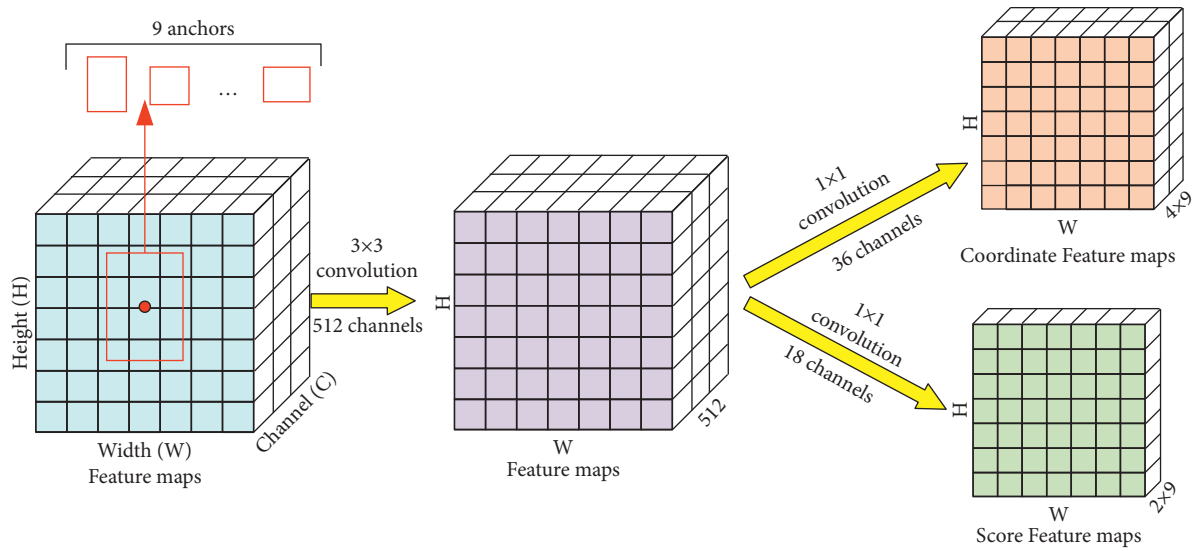


Figure 10: The implementation details of RPN in our detection network.

Since the ResNet-50 has fewer parameters and lower Flops than ResNet-10 and ResNet-152, and it also has a powerful feature representing ability, we select ResNet-50 as our feature extraction network of ship target detection in SAR images.

The region proposal network (RPN) [17, 33] is used to extract candidate regions based on the shared convolution feature map, and the whole detection process is unified into an overall framework so that the end-to-end training can be truly realized. A small network is sliding on the output convolution feature maps of the last shared convolution layer, as shown in Figure 10. This small network takes the spatial window on the input convolution feature maps as the input and obtains 512 feature maps with size $W \times H$. The post-processing is connected to two fully connected layers of the same level: the regression layer and the frame classification layer, as shown in Figure 10, which are realized by two convolution layers of the same level.

At each sliding window position, multiple regional proposals are predicted simultaneously, in which the anchor is located at the center of the relevant sliding window. If the maximum number of possible proposals for each sliding position is K, K anchors are generated at each sliding position. For convolution feature maps of size $w \times h$, there are $w \times h \times k$ anchors. After then, each anchor in the regression layer gets 4 outputs from forwarding features, corresponding to the offset of each proposal coordinate from the original image coordinate, while each anchor in the classification layer gets 2 outputs from forwarding feature maps, corresponding to the probability that each proposal is a foreground (target) or background. By default, each sliding position generates 9 anchors.

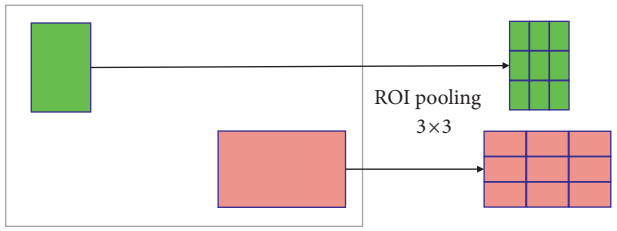


Figure 11: The structure of the ROI pooling layer.

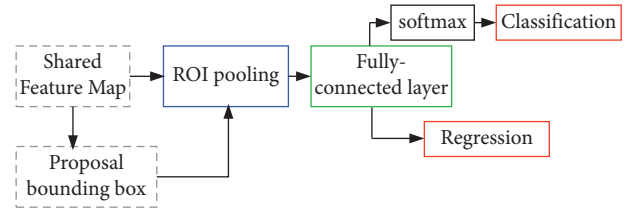


Figure 12: The implantation details of the classification module.

The detection network is used to classify the region of interest (ROI) with location and background information. The sizes of obtained ROIs are different while the sizes corresponding to the feature maps are also different. However, the number of neurons in the fully connected layers is fixed. To overcome this question, the ROI pooling layer is used to extract the feature maps corresponding to the candidate regions, and all the feature maps with different sizes are divided into the same scale, and then the output results are pooled to the same size, as shown in Figure 11.

The last two fully connected layers are the regression layer and classification layer of the prediction box, which are used to make more accurate regression and classification of candidate regions to achieve the final target location and specific category. The structure of the recognition module is shown in Figure 12. Using the classification module to classify the extracted proposal bounding box by RPN, the category can be determined and the coordinate is also modified to get a more accurate predicted box.

3.2.2. K-Means for Determining Anchors. There are three fixed length-width ratios and three scales of the anchors in the original RPN [17]. The main reason lies in that the target distribution in VOC is fixed and it does not need to revise the anchors. Different from VOC data, the ships in SAR images are various with different scales and different length-width ratios. Thus, we should revise the anchors in our detection network according to the characteristics of the ship dataset. To overcome the above issue, a K-means method is introduced to our detection to provide more accurate anchors.

K-means clustering algorithm is based on partition [6]. The cluster generated by clustering is a collection of data objects, which can improve the similarity of objects in the cluster as much as possible, and clusters differ greatly. Given a dataset with N elements, K-means clustering is carried out as follows. Firstly, based on the initial cluster center (centroid), the algorithm calculates the distance from each sample to the cluster center and then divides the sample into the nearest cluster center. Then, the average value of each new clustering data object is calculated to obtain a new clustering center. Based on this, the samples are reclassified and iterated continuously. Each iteration can make the same cluster closer until the best cluster is obtained.

K points are selected from the ship detection dataset as the initial centroid, but the initial centroid selection has a certain impact on the clustering results. To overcome this problem, K-means ++ is introduced to cluster the anchors of the ship detection dataset, as shown in Figure 13. This process is as follows:

(1) Randomly select a point from the ship detection data as the clustering center;

(2) The shortest distance between each point in the data and the existing cluster center is calculated, that is, the distance between each point and the nearest cluster center $L(x_i) = \arg\min\|x_i - o_j\|^2$, where $j = 1, 2 \ldots k$;

(3) The data with the shortest distance is selected as the new clustering center;

(4) Repeat steps (2) and (3) until K cluster centers are obtained; The K-means algorithm is initialized by the K clustering centers obtained by the above method. With the initialized clustering centers and the clustering method in Figure 13, the anchors appropriate for our ship detection network can be attained.

3.2.3. Loss Function. The loss function of RPN can be expressed as:

$$L(\{p_i\},\{t_i\}) = \frac{1}{N_{\text{cls}}}\sum_i L_{\text{cls}}(p_i, p_i^*) + \lambda\frac{1}{N_{\text{reg}}}\sum_i p_i^* L_{\text{reg}}(t_i, t_i^*), \quad (3)$$

where i is the index of the anchor, p_i is the probability of predicting the anchor i as the ship target. p_i^* is the label of anchor i. t_i represents the four parameterized coordinates of the prediction box while t_i^* is the parameterized coordinates of the groundtruth. N_{cls} is the number of candidate frames involved in classification, N_{reg} is the number of candidate frames involved in regression.

The cross-entropy loss function is used as the classification loss function [36]

$$L_{\text{clc}}(p_i, p_i^*) = -\log(p_i p_i^* + (1 - p_i)(1 - p_i^*)). \quad (4)$$

The regression loss is $L_{\text{reg}}(t_i, t_i^*) = R(t_i^* - t_i)$ where

$$R(x) = \text{smooth}(x) = \begin{cases} 0 \cdot 5x^2, & |x| < 0, \\ |x| - 0 \cdot 5, & \text{other}. \end{cases} \quad (5)$$

For the boundary box regression, the following four coordinates are parameterized [36]

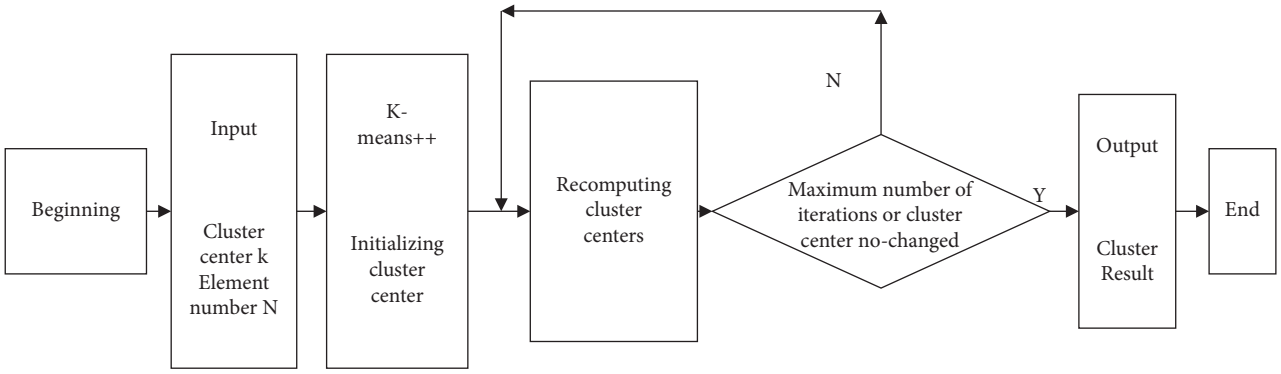


Figure 13: K-means clustering process of our detection network.

$$\begin{aligned} t_x &= \frac{(x - x_a)}{w_a}, \quad t_y = \frac{(y - y_a)}{h_a}, \\ t_w &= \log\left(\frac{w}{w_a}\right), \quad t_h = \log\left(\frac{h}{h_a}\right), \\ t_x^* &= \frac{(x^* - x_a)}{w_a}, \quad t_y^* = \frac{(y^* - y_a)}{h_a}, \\ t_w^* &= \log\left(\frac{w^*}{w_a}\right), \quad t_h^* = \log\left(\frac{h^*}{h_a}\right), \end{aligned} \tag{6}$$

where x, y, w, and h represent the center coordinates, width, and height of the prediction box. The parameters x, x_a , and x^* represent the coordinate values of the predicted box, anchor box, and groundtruth, respectively (similarly for y, w , and h). t_x and t_y are the position translation of the anchor box relative to the prediction box while t_w and t_h are the scale factors. t_x^* and t_y^* are the offset of prediction box relative to ground truth, while t_w^* and t_h^* are the scale factors.

For the recognition part, the loss function still includes two parts: the classification loss of the target and the regression loss of the prediction box, using softmax classifier, the classification part uses the cross-entropy loss function, and the regression loss function is the same as above.

3.3. Fine Detection with Gaussian Function. As shown above, since the radar cross section of the ship is higher than the sea clutter, the ship usually shows a higher brightness than the sea in the attained images [12, 42]. In addition, in the SAR image, the land region also shows a high brightness which makes them likely to be mistakenly detected as a ship target. Although the proposed K-means-based Faster RCNN can detect most ships, some small and isolated land areas and targets on land will be mistakenly detected as ships. In this section, we try to remove the false alarm according to the sea surface confidence value. More specifically, we first attain the probability of judging the background in the detection frame as sea surface using a two-dimensional Gaussian function on the sea-land segmentation result. Then, the confidence threshold is used to correct the detection results, and the ship targets with low sea surface confidence value in the detection results are eliminated, to further improve the detection accuracy.

The probability density function of our two-dimensional Gaussian distribution is as follows:

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{\left(\left((x-\mu_1)^2/\sigma_1^2\right) - 2\rho\left((x-\mu_1)(y-\mu_2)\right)/\sigma_1\sigma_2 + \left((y-\mu_2)^2/\sigma_2^2\right)\right)}{2\left(1-\rho^2\right)}\right\}, \tag{7}$$

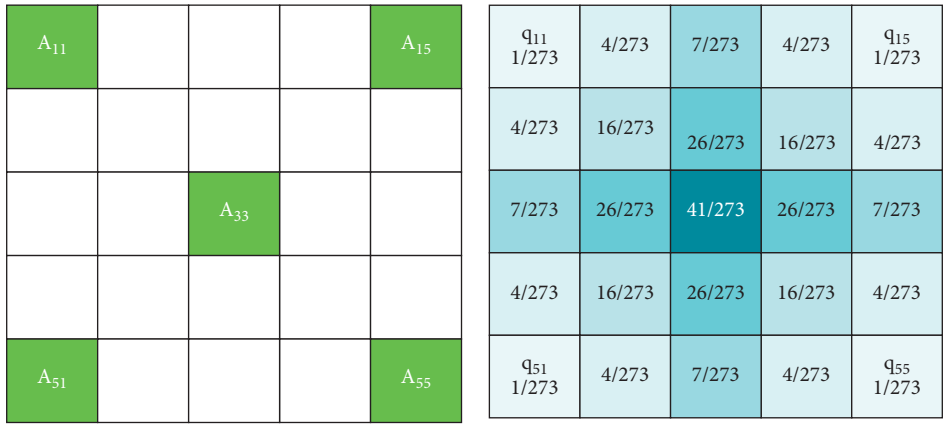
where μ_1, μ_2, σ_1, σ_2 , and ρ are constants, and $\sigma_1, \sigma_2 > 0$, $|\rho| < 1$.

To ensure that the weight of the pixels closer to the center in the detection bounding box is greater, the two-dimensional Gaussian function in (7) is used to multiply the land-sea segmentation results in the detection results to generate the probability that the detected ship target is located on the sea surface, that is, the sea surface confidence value. The detailed operation process is as follows:

(1) The horizontal and vertical directions of the detection frame are divided into 5 identify blocks, as shown in Figure 14(a). Each area is represented as $A_{ij}\,(1 \le i, j \le 5)$ the number of pixels in each area is $K = M \times N/25$, and the pixel value is $p_{ij}^k (k = 1 \ldots K)$;

(2) The element number of the two-dimensional Gaussian function is 5×5, as shown in Figure 14(b). The probability distribution of them is $q_{ij}\,(1 \le i, j \le 5)$;

(3) Each region is multiplied by the corresponding Gaussian probability point to get the probability that the detection bounding box is judged as the sea



(a) (b)

Figure 14: The calculation process of confidence value.

surface, that is, the confidence value of the sea surface is $C = \sum_{i,j}\sum_k p_{ij}^k/K$.

(4) If the confidence value of one predicted bounding box is less than 0.5, this bounding box is classified as background and is eliminated from the detection results.

With the above steps, the false alarm in ships caused by land background can be removed, and the detection accuracy can also be improved.

4. Experimental Results

Since the proposed method consists of three parts: the residual structure-based Unet for sea-land segmentation, K-means-based Faster RCNN for coarse ship detection, and the confidence weighting with Gaussian function, the effectiveness of these parts are validated in this section.

4.1. Experimental Results of Sea-Land Segmentation

4.1.1. Evaluation Criteria and Training Details. In this paper, three evaluation indexes, pixel accuracy (PA), mean pixel accuracy (MPA), and mean intersection over Union (MIoU) [4], are selected to analyze the sea-land segmentation effect. The calculation of these evaluation indexes is defined based on the confusion matrix, that is, the matrix of statistical model classification results. Since there are only two categories in the sea-land segmentation network, the confusion matrix is defined in Table 4.

The pixel accuracy (PA) means the ratio of the number for the corrected classified pixels to the number for total pixels in the detection result. PA also means the ratio of the sum of diagonal elements to the sum of total elements in the confusion matrix. In the sea-land segmentation model, PA can be expressed as:

$$\mathrm{PA} = \frac{T_{\mathrm{sea}} + T_{\mathrm{land}}}{T_{\mathrm{sea}} + F_{\mathrm{sea}} + T_{\mathrm{land}} + F_{\mathrm{land}}}. \quad (8)$$

Table 4: The confusion matrix of sea-land segmentation.

Confusion Matrix		Predicted Value	
		Sea	Land
Ground truth	Sea	T_{sea}	F_{land}
	Land	F_{sea}	T_{land}

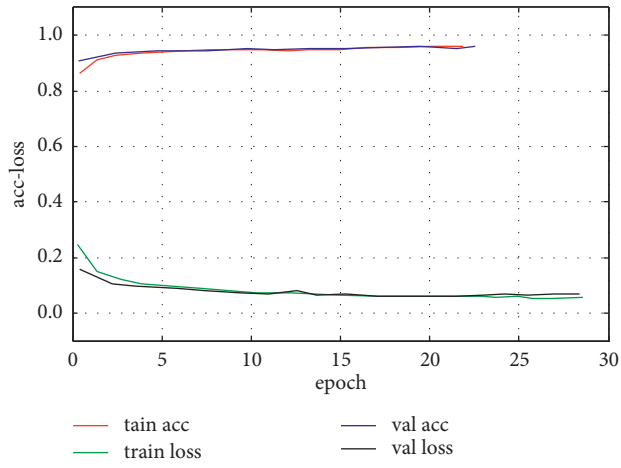


Figure 15: Training and testing curve for loss and accuracy.

For mean pixel accuracy (MPA), the proportion of the predicted value of each class that belongs to this class is calculated, respectively, and then the average is calculated by accumulation. In the confusion matrix, the accuracy of each class CPA is equal to the ratio of the value on the diagonal to the sum of the elements in the corresponding column. In the sea-land segmentation model, it can be expressed as:

$$\mathrm{MPA} = \frac{\mathrm{CPA}_{\mathrm{sea}} + \mathrm{CPA}_{\mathrm{land}}}{2}, \quad (9)$$

where $\mathrm{CPA}_{\mathrm{sea}} = T_{\mathrm{sea}}/(T_{\mathrm{sea}} + F_{\mathrm{sea}})$ is the accuracy of sea pixels; $\mathrm{CPA}_{\mathrm{land}} = T_{\mathrm{land}}/(T_{\mathrm{land}} + F_{\mathrm{land}})$ represents the pixel accuracy.

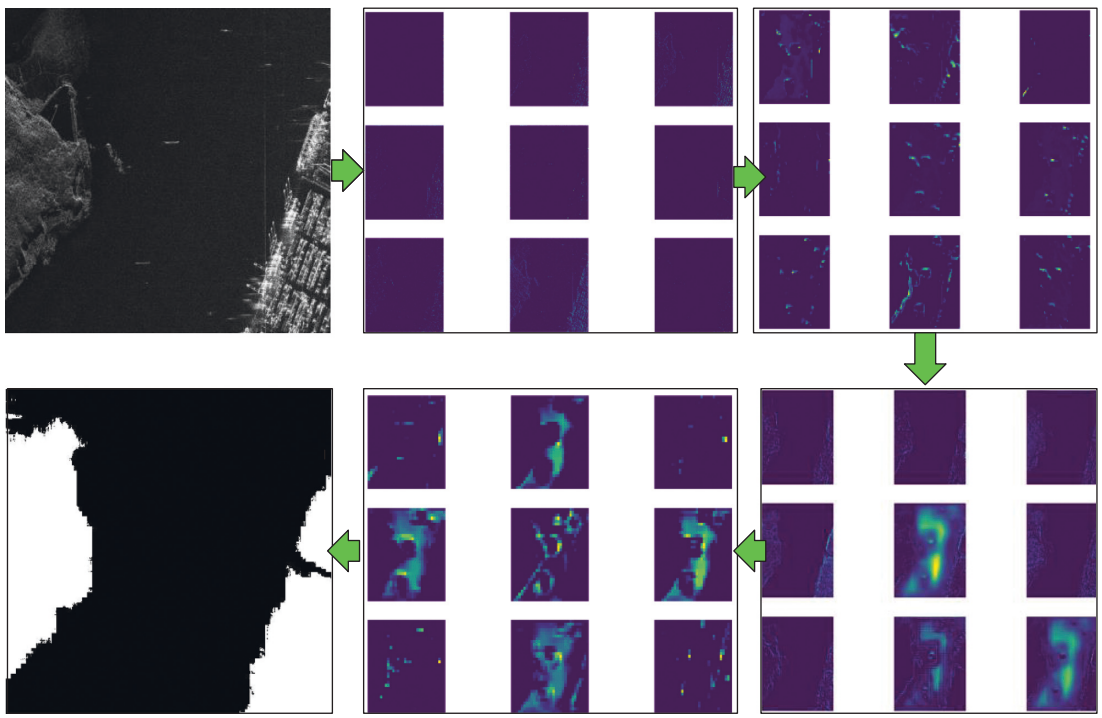


Figure 16: The extracted feature maps for a SAR image.

For the mean intersection over Union (MIoU), the IOU of predicted value and real value of each class is calculated, respectively, and then the average is obtained by accumulation. In the confusion matrix, the IOU of each category is the ratio of the value of the element on the diagonal to the sum of the values of all the elements in the corresponding column. In the sea-land segmentation model, it can be expressed as:

$$\mathrm{MIoU} = \frac{\mathrm{IoU}_{\mathrm{sea}} + \mathrm{IoU}_{\mathrm{land}}}{2}, \tag{10}$$

where $\mathrm{IoU}_{\mathrm{sea}} = T_{\mathrm{sea}}/(T_{\mathrm{sea}} + F_{\mathrm{sea}} + F_{\mathrm{land}})$ refers to the IoU of sea region; $\mathrm{IoU}_{\mathrm{land}} = T_{\mathrm{land}}/(T_{\mathrm{land}} + F_{\mathrm{land}} + F_{\mathrm{sea}})$ means the IoU of land.

The experiment was carried out using the hardware of Windows10 Xeon E5-2643 V3 3.50Ghz, the memory size of 64 GB, and configured NVIDIA Titan XP GPU with 12 G memory. The experiment uses python programming language, builds networks based on Tensorflow's keras deep learning framework, and uses opencv, numpy, Matplotlib, and other libraries.

The land-sea segmentation dataset in Section 2.1 is used to train the proposed segmentation network. In the process of training, the training data are randomly divided into training sets and verification set at a ratio of 7 : 2, with 9952 training sets and 2844 verification sets. Most of the training parameters are initialized by using a truncated normal distribution with the mean value of 0, the standard deviation of 0.05, and the constant deviation of 0.1. Then, the adaptive moment (Adam) algorithm is used to update parameters with adaptive learning rates to improve learning efficiency. The moving average decay of batch standardization is set to 0.9, and the normalized dropout probability is 0.5. The batch size is 4, and the number of training epochs is 30, which is selected considering memory limitation and learning time.

The training curve of the segmentation model is shown in Figure 15. The red curve is the accuracy of the training process while the blue curve is the accuracy of the test set. The green curve is the change curve of the loss function on the training set, and the black curve is the change curve of the loss function on the test set. One can see that the proposed segmentation model has a good training process on the constructed sea-land segmentation dataset. The training process is stable and the testing accuracy is also significantly improved with the training process.

To show the feature extraction process of our residual structure-based Unet model, we take a sub-image cut out from an SAR image to be segmented as an example, and output its feature maps in the encoding stage and decoding stage. Since the dimension of the last feature maps is relatively large (1024), that is, there are 1024 sub-feature maps in this layer, this paper only selects 9 sub-feature maps, as shown in Figure 16, which shows the intermediate result maps in the feature extraction process. In Figure 16, the images from left to right and from top to bottom are, respectively, SAR sub-image, feature maps in the encoding pooling layer, feature maps in the encoding stage convolution layer, feature maps in the decoding stage upsampling layer, feature maps in the decoding stage convolution layer, and the final segmentation result map. It can be seen that with the deepening of the network, the features become more and more obvious, and the marine and land features

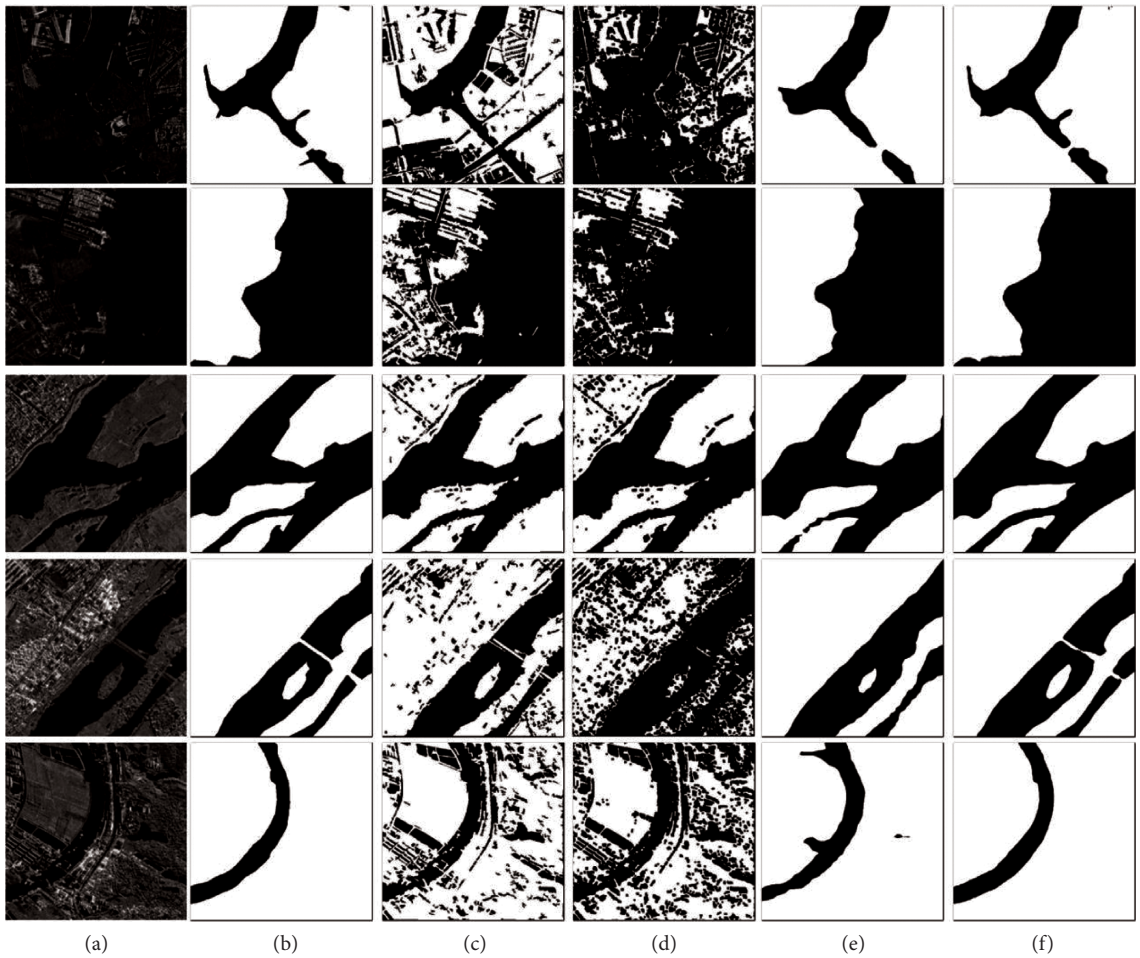


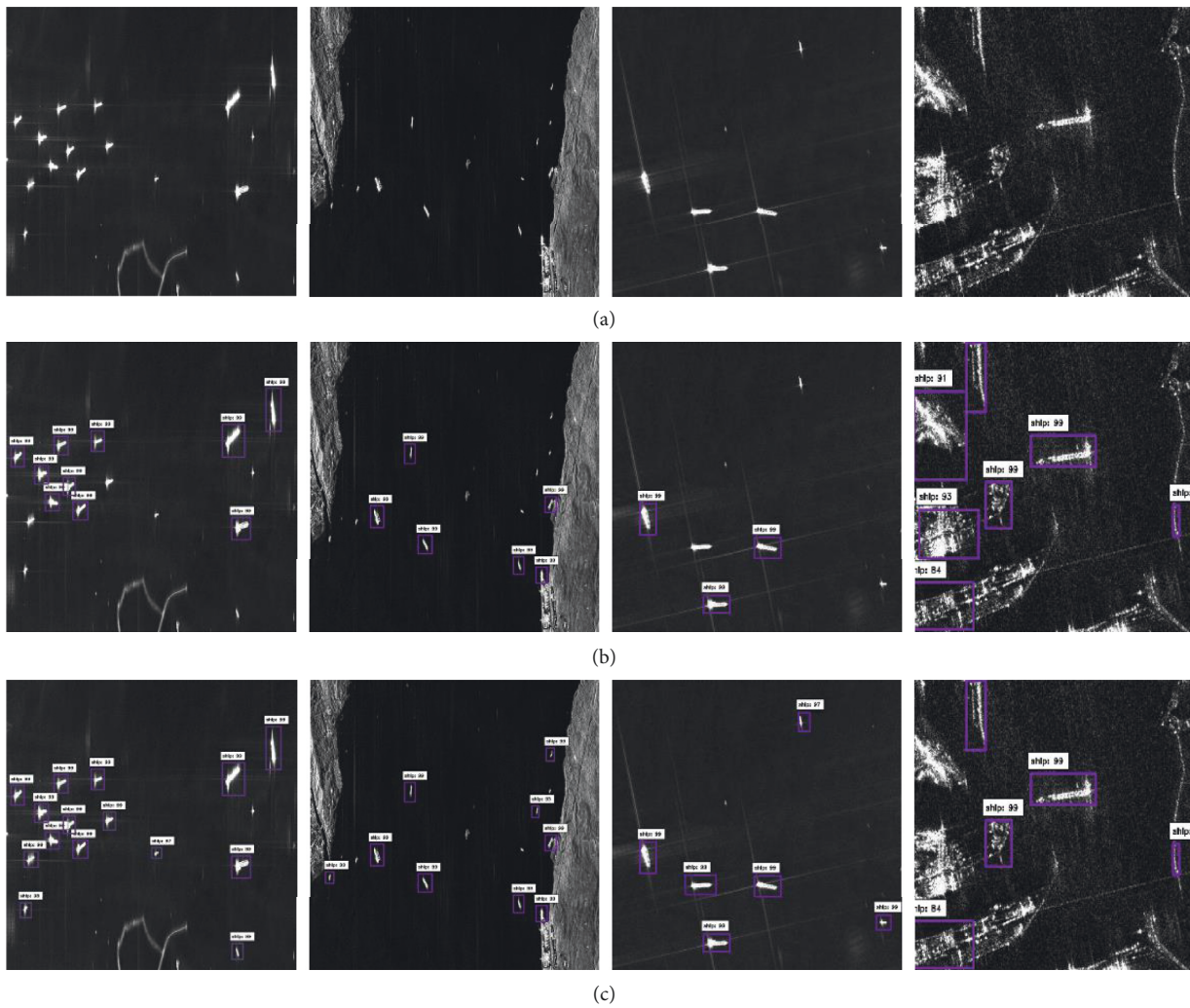
Figure 17: Experimental results of different methods. (a) Input image. (b) Ground truth. (c) Levelset. (d) Otsu. (e) Unet. (f) Proposed method.

are effectively distinguished. Thus, both the high-dimensional features and accurate features are extracted in our segmentation network.

4.1.2. Segmentation Results of Different Methods. In order to prove the effectiveness of the proposed segmentation method, five images including land and sea region are selected from the test set. The LevelSet method [50, 51], threshold-based OtSU method [20], and the original Unet method [28] are compared with our segmentation method. The segmentation results of these methods are shown in Figure 17, where Figure 17(a) shows the input images, Figure 17(b) shows the ground truth, Figures 17(c)–17(e) show the segmentation results of LevelSet, Otsu and Original Unet method. Especially, the results of our residual structure-based segmentation model are shown in Figure 17(f). It can be seen from the segmentation results that the segmentation results in Figures 17(c) and 17(d) cannot represent the true sea-land distribution in Figure 17(b). After using the traditional Levelset and Otsu method to segment the dark area on the land, it is easy to be misjudged as the sea surface. Thus, the traditional methods cannot form the connected land area, and the segmentation effect is greatly affected by the image itself. The results in Figures 17(e) and 17(f) are better than those in Figures 17(c) and 17(d). Thus, the deep learning method based on the sea-land training dataset can extract abstract and powerful features from input images and make an accurate prediction. Especially, the results in Figure 17(e) are closer to the ground truth in Figure 17(b) than Figure 17(d). These results demonstrate that compared with the original Unet method, the proposed method can retain the edge contour information of sea and land, and make the segmentation result closer to the label image.

Using the three evaluation indexes, the quantitative results of the LevelSet method, Otsu method, original Unet method, and the proposed method are analyzed and shown in Table 5.

From this table, we can see that the PA, MPA, and MIoU of deep learning-based methods are much higher than that of the traditional LevelSet method and threshold-based Otsu method. Meanwhile, the proposed method also performs better than the original Unet method. For an image with 512×512 pixels in testing data, the Otsu method has the fastest running speed. Although the



(a)

(b)

(c)

Figure 18: Ship detection results.

Table 5: Index comparison of different segmentation algorithms.

Evaluation criteria	LevelSet	Otsu	Unet	Proposed method
PA (%)	52.39	41.33	95.48	97.89
MPA (%)	58.74	54.81	94.27	96.42
MIoU (%)	49.35	37.86	94.73	97.04
Speed (s)	3.2190	0.032	0.058	0.087

Table 6: Confusion matrix of ship target detection.

Confusion matrix		Predicted value	
		Ship	Background
Ground truth	Ship	T_{ship}	F_{bg}
	Background	F_{ship}	T_{bg}

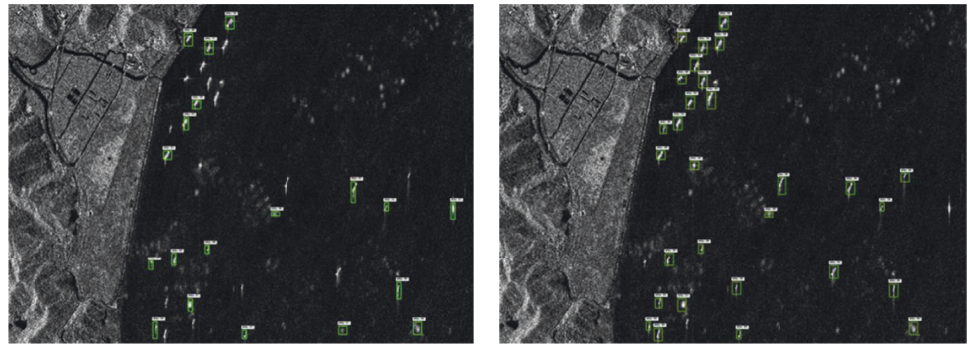
proposed method has a slower speed than Unet, it improves the PA from 95.48% to 97.89% and improves the MIoU from 94.73% to 97.04%. Thus, both the visualization results and quantitative results demonstrate the effectiveness of the proposed method.

4.2. Experimental Results of Ship Detection

4.2.1. Evaluation Criteria and Parameter Setting. In this paper, we choose the commonly used indicators to evaluate the effect of our detection model: precision

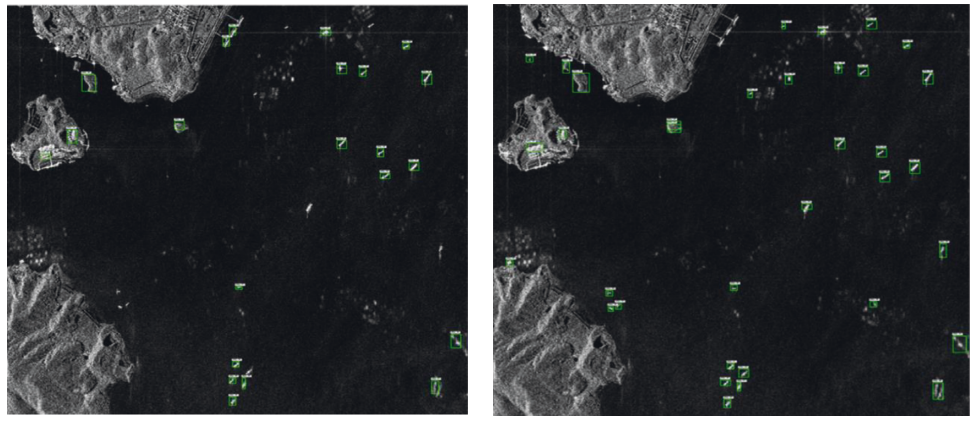
Table 7: Comparison of ship target detection results.

Method	Faster-RCNN	Proposed method
$T_{ship} + F_{bg}$	291	291
$T_{ship} + F_{ship}$	290	305
T_{ship}	239	264
Precision	82.41%	86.56%
Recall	82.13%	90.72%



(a) (b)

Figure 19: The detection results of different methods on the large-scene image I. (a) The original Faster RCNN method. (b) The proposed detection method.



(a) (b)

Figure 20: The detection results of different methods on the large-scene image II. (a) The original Faster RCNN method. (b) The proposed detection method.

(P) and recall (R) to analyze the detection effect. Like the evaluation index used in Table 6, its calculation is also defined based on the confusion matrix. Since there are only ships in the ship detection task, the detection results are background or ships. The confusion matrix is shown in Table 6.

Among them, T_{ship} is the number of detected real ship targets while F_{ship} is the number of false alarms determined by the background. F_{bg} is the number of undetected real ship targets determined by the detection model while T_{bg} is the number of real ship targets in the image. The detection accuracy P reflects the proportion of the real ship target among the targets detected by the algorithm

$$P = \frac{T_{ship}}{T_{ship} + F_{ship}}. \tag{11}$$

Recall rate R reflects the probability that the real ship target in the image is judged as a ship by the detection algorithm.

$$R = \frac{T_{ship}}{T_{ship} + F_{bg}}. \tag{12}$$

During training, the network parameters are initialized by the pretrained model on Imagenet, and then the ship detection dataset in Section 2.2 is used for training. Back-propagation and stochastic gradient descent (SGD) are used

Figure 21: Land-sea segmentation result and confidence value of test box.

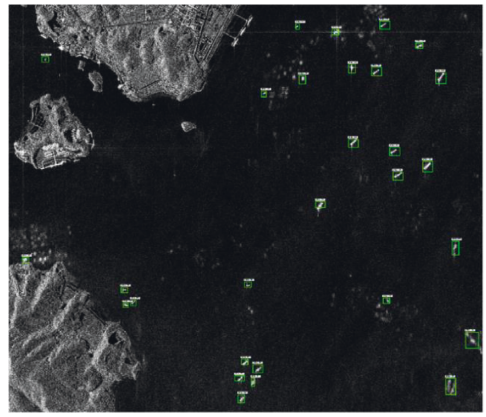


Figure 22: Fine detection result with sea-land segmentation.

for the end-to-end training of our detection model. In each gradient descent process, the proposal box generated by RPN is directly transferred to the detection module for training. In the process of backpropagation, the derivation of each stage in the network is obtained, respectively. The four loss functions, including the classification loss function and the regression function, are optimized in the whole process. In training, the initial learning rate is 0.0001, the weight decay rate is 0.0001, and the momentum is 0.9.

4.2.2. Detection Results and Analysis. After getting the well-trained detection model, the SAR images in the testing data are used to validate the effectiveness of the proposed method. When detecting ships, multiple detection bounding boxes may be marked for the same ship target. At this time, the NMS algorithm is used to abandon the bounding boxes whose coincidence ratio is higher than 0.5. The original Fast RCNN algorithm and the proposed method are used to detect the images in the test set, as shown in Figure 18. Figure 18(a) shows four images from the testing dataset. Figure 18(b) shows the detection results of the original Faster RCNN method while Figure 18(c) shows the detection results of the proposed method. It can be seen that the improved method can detect more ship targets than the original Faster RCNN method, especially in the pure sea area. When it comes to the complex background region, like the harbor with various buildings, the original Faster-RCNN generates too many false ship targets. Compared with the original Faster RCNN, the proposed method can reduce the number of missed detection. The reason mainly lies in that the K-means in our method constructs adaptive anchors for the SAR ship detection task and captures more accurate bounding boxes for each ship.

To quantitatively analyze the detection results, we use the evaluation indexes in Section 4.2.1: detection accuracy P and recall R to analyze and compare the original Faster RCNN algorithm and the proposed algorithm in this paper. The detection results are shown in Table 7. From the statistical results of the indicators in this table, it can be seen that compared with the original fast RCNN algorithm, the recall rate and accuracy rate of the improved algorithm are improved by the proposed method, in which the recall rate is increased from 82.13% to 90.72%, and the accuracy rate is increased from 82.41% to 86.56%. Thus, both the visualization results in Figure 18 and the quantitation results in Table 7 demonstrate the effectiveness of the proposed ship detection method.

4.3. Experimental Results of Gaussian Fusion. Although the accuracy rate has been improved to a certain extent, for ship target detection including land area, it can be seen from Figures 19 and 20 that there are still high false alarm targets in the land area, and further operation is needed to reduce the high false alarm rate caused by land area.

The method in Section 3.3 is used for the re-detection of Figure 20. The sea-land segmentation result of the proposed residual structure-based Unet is shown in Figure 21, and the sea level confidence values calculated by the detection box and Gaussian function are also shown in Figure 21.

After deleting the detection box whose confidence value of sea surface is less than 0.5, the accurate detection result is attained and shown in Figure 22. From the detection result in Figure 22, it can be seen that by combining the sea-land segmentation results and judging the sea confidence value of the detection box, some false targets located on the land are eliminated, and the detection performance is improved. Table 8 shows the quantitative results corresponding to Figures 19, 20, and 22. One can see that for the large-scene SAR image I and II, the improved Faster RCNN method has a higher precision rate of 87.10% and a lower recall rate of 86.20% than the original Faster RCNN method. It is noticeable that the detection accuracy of Figure 20 is not well with the precision of 75.75%, which mainly is caused by the false alarm targets in the land region. Using the sea-land segmentation results and confidence value with Gaussian function, the detection precision of Figure 20 is improved from 75.75% to 89.29%, as shown in Figure 22. Thus, the proposed fine detection strategy with Gaussian function can indeed remove the false alarm targets from the detection results.

Table 8: Comparison of ship target detection results.

Image	Method	$T_{ship} + F_{bg}$	$T_{ship} + F_{ship}$	T_{ship}	Precision (%)	Recall (%)
1	Original Faster-RCNN	31	19	16	84.21	51.61
	Improved Faster-RCNN		29	27	93.10	87.10
2	Original Faster-RCNN	29	22	14	63.63	48.28
	Improved Faster-RCNN		33	25	75.75	86.20
	Fine detection		28	25	89.29	86.20

5. Conclusion

This paper mainly studies the sea-land segmentation and ship target detection method based on deep learning technology for GF-3 SAR images, from the production of dataset to the design of sea-land segmentation algorithm and ship detection algorithm. Traditional SAR image segmentation methods are usually affected by specific scenes and have poor robustness. In this paper, an improved Unet based on fully convolution layers and residual-connection structure is proposed for SAR image sea-land segmentation. To solve the problem of a large number of data training, this paper introduces a residual convolution module into the original network to deepen the network level and redesigns the jump connection mode. Then, the improved Faster RCNN algorithm based on region extraction is studied, which is applied to ship detection in SAR images. K-means is used to cluster the size and aspect ratio of ship targets, to improve the anchor frame design and make it more suitable for our ship detection task. Since the false alarm rate of detection is too high due to the existence of a complicated land region, a fine detection algorithm combined with sea-land segmentation results is proposed. In this method, the Gaussian function fuses the confidence value of sea-land segmentation results and the coarse detection results of improved Faster RCNN. The segmentation results and detection results on the established segmentation dataset and detection dataset, respectively, demonstrate the effectiveness of our proposed segmentation method and detection method. It is noticeable that although the proposed methods achieve good detection results in ship detection dataset, the segmentation module and detection module of our method leads to higher complexity than traditional methods. Thus, if these modules can be fused into a module which can realize sea-land segmentation and ship detection simultaneously, the detection performance will also be improved. In the future, we will continue this research to attain better ship detection performance in high-resolution SAR images.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was funded by the Scientific Research Program Funded by Shaanxi Provincial Education Department, grant no. 21JK0747.

References

[1] J. Chen, J. Zhang, Y. Jin, H. Yu, B. Liang, and D. G. Yang, "Real-time processing of spaceborne SAR data with nonlinear trajectory based on variable PRF," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–12, 2022.

[2] A. Moreira, P. Prats-Iraola, M. Younis, G. Krieger, I. Hajnsek, and K. P. Papathanassiou, "A tutorial on synthetic aperture radar," *IEEE Geosci. Remote Sens. Mag.*vol. 1, pp. 6–43, 2013.

[3] J. Chen, M. Xing, X. G. Xia, J. Zhang, B. Liang, and D. G. Yang, "SVD-based ambiguity function analysis for nonlinear trajectory SAR," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 4, pp. 3072–3087, 2021.

[4] J. Zhang, M. Xing, G. C. Sun et al., "Water body detection in high-resolution SAR images with cascaded fully-convolutional network and variable focal loss," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 1, pp. 316–332, 2021.

[5] G. C. Sun, Y. Liu, M. Xing, S. Wang, L. Guo, and J. Yang, "A real-time imaging algorithm based on sub-aperture CS-dechirp for GF3-SAR data," *Sensors*, vol. 18, no. 8, p. 2562, 2018.

[6] J. Zhang, M. Xing, and Y. Xie, "FEC: a feature fusion framework for SAR target recognition based on electromagnetic scattering features and deep CNN features," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 3, pp. 2174–2187, 2021.

[7] F. Bachofer, G. Queneherve, and M. Marker, "The delineation of paleo-shorelines in the lake manyara basin using TerraSAR-X data," *Remote Sensing*, vol. 6, no. 3, pp. 2195–2212, 2014.

[8] F. Hu, G. S. Xia, J. W. Hu, and L. P. Zhang, "Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery," *Remote Sensing*, vol. 7, no. 11, Article ID 14680, 2015.

[9] M. Ma, J. Chen, W. Liu, and W. Yang, "Ship classification and detection based on CNN using GF-3 SAR images," *Remote Sensing*, vol. 10, no. 12, p. 2043, 2018.

[10] Y. Wang, C. Wang, H. Zhang, Y. Dong, and S. Wei, "Automatic ship detection based on RetinaNet using multi-resolution gaofen-3 imagery," *Remote Sensing*, vol. 11, no. 5, p. 531, 2019.

[11] Y. Wang, C. Wang, H. Zhang, Y. Dong, and S. Wei, "A SAR dataset of ship detection for deep learning under complex backgrounds," *Remote Sensing*, vol. 11, p. 765, 2019.

[12] X. Sun, Z. Wang, Y. Sun, W. Diao, Y. Zhang, and K. Fu, "AIR-SARShip-1.0: high-resolution SAR ship detection dataset," *J. Radars*, vol. 8, pp. 852–862, 2019.

[13] P. Iervolino, R. Guida, P. Lumsdon et al., "Ship detection in SAR imagery: a comparison study," in *Proceedings of the 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 2050–2053, IEEE, Fort Worth, TX, USA, July 2017.

[14] Y. Wang and H. Liu, "A hierarchical ship detection scheme for high-resolution SAR images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 10, pp. 4173–4184, 2012.

[15] R. Touzi, A. Lopes, and P. Bousquet, "A statistical and geometrical edge detector for SAR images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 26, no. 6, pp. 764–773, 1988.

[16] W. Ao, F. Xu, Y. Li, and H. Wang, "Detection and discrimination of ship targets in complex background from spaceborne ALOS-2 SAR images," *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 2, pp. 536–550, 2018.

[17] J. Li, C. Qu, and J. Shao, "Ship detection in SAR images based on an improved faster R-CNN," in *Proceedings of the 2017 SAR in Big Data Era: Models Methods and Applications (BIGSARDATA)*, pp. 1–6, Beijing, China, November 2017.

[18] X. Ding and X. Li, "Coastline detection in SAR images using multiscale normalized cut segmentation," in *Proceedings of the 2014 IEEE Geoscience and Remote Sensing Symposium*, pp. 4447–4449, July 2014.

[19] M. Li, Y. Wu, and Q. Zhang, "SAR image segmentation based on mixture context and wavelet hidden-class-label Markov random field," *Computers & Mathematics with Applications*, vol. 57, no. 6, pp. 961–969, 2009.

[20] J. Jennifer Ranjani and S. J. Thiruvengadam, "Fast threshold selection algorithm for segmentation of synthetic aperture radar images," *IET Radar, Sonar & Navigation*, vol. 6, no. 8, pp. 788–795, 2012.

[21] U. Kandaswamy, D. A. Adjeroh, and M. Lee, "Efficient texture analysis of SAR imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 9, pp. 2075–2083, 2005.

[22] W. Feng, H. Sui, W. Huang, C. Xu, and K. An, "Water body extraction from very high-resolution remote sensing imagery using deep U-net and a superpixel-based conditional random field model," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 4, pp. 618–622, 2019.

[23] C. Sukawattanavijit, J. Chen, and H. S. Zhang, "GA-SVM algorithm for improving land-cover classification using SAR and optical remote sensing data," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 3, pp. 284–288, 2017.

[24] F. Isikdogan, A. C. Bovik, and P. Passalacqua, "Surface water mapping by deep learning," *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 11, pp. 4909–4918, 2017.

[25] J. Geng, H. Y. Wang, J. C. Fan, and X. R. Ma, "Deep supervised and contractive neural network for SAR image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 4, pp. 2442–2459, 2017.

[26] J. Geng, X. R. Ma, J. C. Fan, and H. Y. Wang, "Semisupervised classification of polarimetric SAR image via superpixel restrained deep neural network," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 1, pp. 122–126, 2018.

[27] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, pp. 640–651, 2017.

[28] O. Ronneberger, P. Fischer, and T. Brox, "U-net: convolutional networks for biomedical image segmentation," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 234–241, Munich, Germany, October 2015.

[29] N. Levanon, "Detection loss due to interfering targets in ordered statistics CFAR," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 24, no. 6, pp. 678–681, 1988.

[30] M. Barkat and P. K. Varshney, "Adaptive cell-averaging CFAR detection in distributed sensor networks," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 27, no. 3, pp. 424–429, 1991.

[31] G. Gao, L. Liu, L. Zhao, G. Shi, and G. Kuang, "An adaptive and fast CFAR algorithm based on automatic censoring for target detection in high-resolution SAR images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 6, pp. 1685–1697, 2009.

[32] H. Guo, X. Yang, N. Wang, B. Song, and X. Gao, "A rotational libra R-CNN method for ship detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 8, pp. 5772–5781, 2020.

[33] M. Kang, X. Leng, Z. Lin, and K. Ji, "A modified faster R-CNN based on CFAR algorithm for SAR ship detection," in *Proceedings of the 2017 International Workshop on Remote Sensing with Intelligent Processing (RSIP)*, pp. 1–4, IEEE, Shanghai, China, May 2017.

[34] W. Dai, Y. Mao, R. Yuan, Y. Liu, X. Pu, and C. Li, "A novel detector based on convolution neural networks for multiscale SAR ship detection in complex background," *Sensors*, no. 9, p. 2547, 2020.

[35] L. De Laurentiis, A. Pomente, F. Del Frate, and G. Schiavon, "Capsule and convolutional neural network-based SAR ship classification in Sentinel-1 data," *Active and Passive Microwave Remote Sensing for Environmental Monitoring III*, vol. 11154, Article ID 1115405, 2019.

[36] Z. Lin, K. Ji, X. Leng, and G. Kuang, "Squeeze and excitation rank faster R-CNN for ship detection in SAR images," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 5, pp. 751–755, 2019.

[37] X. Zhou, J. Zhuo, and P. Krähenbühl, "Bottom-up object detection by grouping extreme and center points," in *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 850–859, Long Beach, CA, USA, June 2019.

[38] S. Wei, X. Zeng, Q. Qu, M. Wang, H. Su, and J. Shi, "HRSID: a high-resolution SAR images dataset for ship detection and instance segmentation," *IEEE Access*, vol. 8, Article ID 120234, 2020.

[39] C. Zhu, Y. He, and M. Savvides, "Feature selective anchor-free module for single-shot object detection," in *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 840–849, Long Beach, CA, USA, June 2019.

[40] Y. Gui, X. Li, and L. Xue, "A multilayer fusion light-head detector for SAR ship detection," *Sensors*, vol. 19, no. 5, p. 1124, 2019.

[41] J. Li, C. Qu, and S. Peng, "Ship classification for unbalanced SAR dataset based on convolutional neural network," *Journal of Applied Remote Sensing*, vol. 12, p. 1, 2018.

[42] B. Li, B. Liu, L. Huang, W. Guo, Z. Zhang, and W. Yu, "OpenSARShip 2.0: a large-volume dataset for deeper interpretation of ship targets in Sentinel-1 imagery," in *Proceedings of the 2017 SAR in Big Data Era: Models, Methods and Applications (BIGSARDATA)*, pp. 1–5, IEEE, Beijing, China, November. 2017.

[43] T. Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proceedings of the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2117–2125, Honolulu, HI, USA, July 2017.

[44] S. Hochreiter, "The vanishing gradient problem during learning recurrent neural nets and problem solutions," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 06, no. 02, pp. 107–116, 1998.

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

[46] Y. Mao, Y. Yang, Z. Ma, M. Li, H. Su, and J. Zhang, "Efficient low-cost ship detection for SAR imagery based on simplified U-net," *IEEE Access*, vol. 8, Article ID 69742, 2020.

[47] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, https://arxiv.org/abs/1409.1556.

[48] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, Inception-Resnet and the Impact of Residual Connections on Learning," 2017, https://arxiv.org/abs/1602.07261.

[49] F. X. Chollet, "Deep learning with depthwise separable convolutions," in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1800–1807, Honolulu, HI, USA, July 2017.

[50] R. C. P. Marques, F. N. Medeiros, and J. Santos Nobre, "SAR image segmentation based on level set approach and {\cal G} _A0 model," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 10, pp. 2046–2057, 2012.

[51] T. Xie, W. Zhang, L. Yang, Q. Wang, J. Huang, and N. Yuan, "Inshore ship detection based on level set method and visual saliency for SAR images," *Sensors*, vol. 18, no. 11, p. 3877, 2018.