

SAR Ship Detection Based on End-to-End Morphological Feature Pyramid Network

Congxia Zhao, Xiongjun Fu, Jian Dong, Rui Qin, Jiayun Chang, Ping Lang

Abstract—Intelligent ship detection based on high-precision synthetic aperture radar (SAR) images plays a vital role in ocean monitoring and maritime management. Denoising is an effective preprocessing step for target detection. Morphological network based denoising can effectively remove speckle noise, while the smoothing effect of which blurs the edges of the image and reduce the detection accuracy. The fusion of edge extraction and morphological network can improve detection accuracy by compensating for the lack of edge information caused by smoothing. This paper proposes an end-to-end lightweight network called morphological feature-pyramid (Mor-FP) Yolo v4-tiny for SAR ship detection. First, a morphological network is introduced to preprocess the SAR images for speckle noise suppression and edge enhancement, providing spatial high-frequency information for target detection. Then, the original and preprocessed images are combined into the multi-channel as an input for the convolution layer of the network. The feature pyramid fusion structure is used to extract the high-level semantic features and shallow detailed features from the image, improving the performance of multi-scale target detection. Experiments on the public SAR ship detection data set (SSDD) and AIR SARShip-1.0 show that the proposed method performs better than the other CNN-based methods.

Index Terms—CNN, morphological network, feature pyramid fusion, SAR target detection.

I. INTRODUCTION

SHIP detection plays an important role in ocean inspection and maritime management [1]. Due to its feature of all-weather, full-time observation, and high resolution for targets, SAR ship detection has been a research hotspot in recent years.

Traditional SAR ship detection mainly focuses on ship target and waves detection. Waves detection is not the mainstream method because waves do not exist all the time.

Ship target detection method is mainly based on statistical features [2-4], scattering characteristics [5], and transform domain [6], among which the constant false alarm rate (CFAR)[7-12] is the most widely used classical one in traditional ship detection. Although the previous ship detection methods have excellent performance in some specific scenes, they have many shortcomings, such as complex feature extraction process, sensitivity to speckle noise, and higher false positives in complex backgrounds, which limit the accuracy of these methods. Besides, due to the complexity and huge calculation of these methods, it is difficult for them to meet real-time requirements.

Recently, with the development of deep learning, the methods based on convolution neural networks (CNNs) have become the mainstream for target detection. Deep learning can learn image features automatically and owns stronger feature extraction capabilities. Deep learning based target detection algorithms are mainly divided into one-stage [13-18] and two-stage detection algorithms [19-23] according to the method of anchor regression. The two-stage methods generally possess higher accuracy. L. Zhao *et al.* [24] innovatively combined squeeze and excitation mechanism with Faster R-CNN to make better detection performance. J. Jiao *et al.* [25] introduced dense connection and a new training strategy into Faster R-CNN to reduce the weight of easy examples, bringing an excellent performance in detecting small-scale ships and the interference of inshore complex background. Z. Cui *et al.* [26] proposed a dense attention pyramid network (DAPN) to combine the highlight features from densely connected CBAM with the global unblurred features, and fed them to the detection network to detect ships. The one-stage detectors use a variety of techniques to reduce the amount of calculation, making the one-stage methods become the mainstream target detection algorithms in practical applications. X. Qi *et al.* [27] proposed a single-stage detector based on the attention mechanism, which improved the detection accuracy for small objects under the complicated background. R. Yang *et al.* [28] proposed an improved one-stage object detection framework based on RetinaNet and rotatable bounding box, which performed well in rotate target detection. Some lightweight algorithms with more practical value have been proposed for higher detection efficiency, including Yolo v4-tiny, YOLOX-tiny, while few of them have been used for ship detection.

It is a common problem that SAR images are regarded as optical images to process, while the characteristics of which are ignored in current methods. SAR images are produced in

This work was supported by 111 Project of China under Grant B14010. (Corresponding author: Xiongjun Fu.)

Congxia Zhao and Xiongjun Fu are with the School of Integrated Circuits and Electronics, Beijing Institute of Technology, Beijing 100081, China (e-mail: zcx09890989@163.com; fuxiongjun@bit.edu.cn).

Jian Dong is with the Shijiazhuang campus, Army Engineering University (e-mail: radarvincent@sina.com).

Rui Qin is with the School of Computer Science and Technology, Chongqing University of Posts and Telecommunications (e-mail: qinrui@cqupt.edu.cn).

Jiayun Chang and Ping Lang are with the School of Integrated Circuits and Electronics, Beijing Institute of Technology, Beijing 100081, China (e-mail: 824400828@qq.com; langping911220@bit.edu.cn).

the microwave/millimeter wave band, and they are different from the optical images consisting of image data acquired by visible and partial infrared band sensors. On the one hand, SAR images only contain the information of a single band, while the optical images usually consist of three-color channels, RGB, or HSV. The difference makes the CNNs-based algorithms aiming at optical images not entirely applicable for SAR image detection. On the other hand, there is a large amount of speckle noise in SAR images caused by the basic principle of coherent imaging, which causes shadows on the image and brings a low signal-to-noise ratio, affecting the SAR target detection results. Driven by these problems, the detection algorithms that combine CNNs with feature extraction and fusion algorithms have recently become a crucial research direction. R. Qin *et al.* [29] introduced the wavelet speckle reduction network into the CNN framework for target recognition and achieved high test accuracy. Jiang J *et al.* [30] adopted non-subsampling Laplacian pyramid decomposition (NSLP) as a step for preprocessing to extract features and fed them into CNNs, bringing excellent performance. J. Ai *et al.* [31] synthesized haar wavelet transform with CNN to classify ships in SAR images and had superior discrimination.

Although the above algorithms have achieved good results, the feature extraction methods are relatively fixed. There is an incomplete fit for feature extraction algorithms and CNNs because of the use of fixed kernel functions and parameters for feature extraction. Setting the kernel functions and parameters as variables that can be adjusted adaptively through backpropagation can obtain more suitable features. The features are conducive to target detection and provide more accurate information for the network. Morphological image processing has achieved great success in many aspects, including image segmentation [32-34], object shape detection [35-37], and filtering [38-41]. The morphological network proposed by Ranjan [42] combines traditional morphological algorithms with the idea of convolution kernels so that morphological operators can be trained through backpropagation. However, the multi-layer stacked structure of the deep morphological network is not suitable for SAR

image detection algorithms. Therefore, this paper proposes a structure that combines the morphological network with the edge extraction algorithm and integrates it into the SAR ship detection network. Considering the real-time requirements for SAR image detection, this paper takes the Yolo v4-tiny as the basic network, which owns a superior balance on detection effect and speed.

Moreover, as the scales of different ships vary in SAR images, the detection methods face great difficulties in detecting multi-scale ships [26], especially small-scale ships, which have similar characteristics with clustered targets and may be lost in the deep feature maps [43-44]. To enhance the detection effect of multi-scale ships, the feature pyramid fusion structure is absorbed to improve the network.

The main contributions of this article can be summarized as follows:

- (1) A novel imaging preprocessing structure is proposed in this paper, which combines the deep morphological network with the edge extraction so that the features with edge information and less noise can be provided for the detection network. It offers more effective prior information for the detection network, making the network more effective.
- (2) The feature pyramid fusion structure is introduced into Yolo v4-tiny to get detailed information for detection. Through extracting multilevel features, the network gets feature maps of different sizes so that the detection capability of multi-scale targets is enhanced, especially on small targets.

The remainders of this article are organized as follows. Section II presents an introduction of the proposed method, Section III describes experiments and results analysis, and Section IV concludes the paper.

II. METHODOLOGY

In this section, we present an introduction of Mor-FP Yolo v4-tiny in detail. First, an overview of the proposed method is presented. Then, the structure of the morphological preprocessing module is explored. Finally, the feature pyramid fusion structure is concerned.

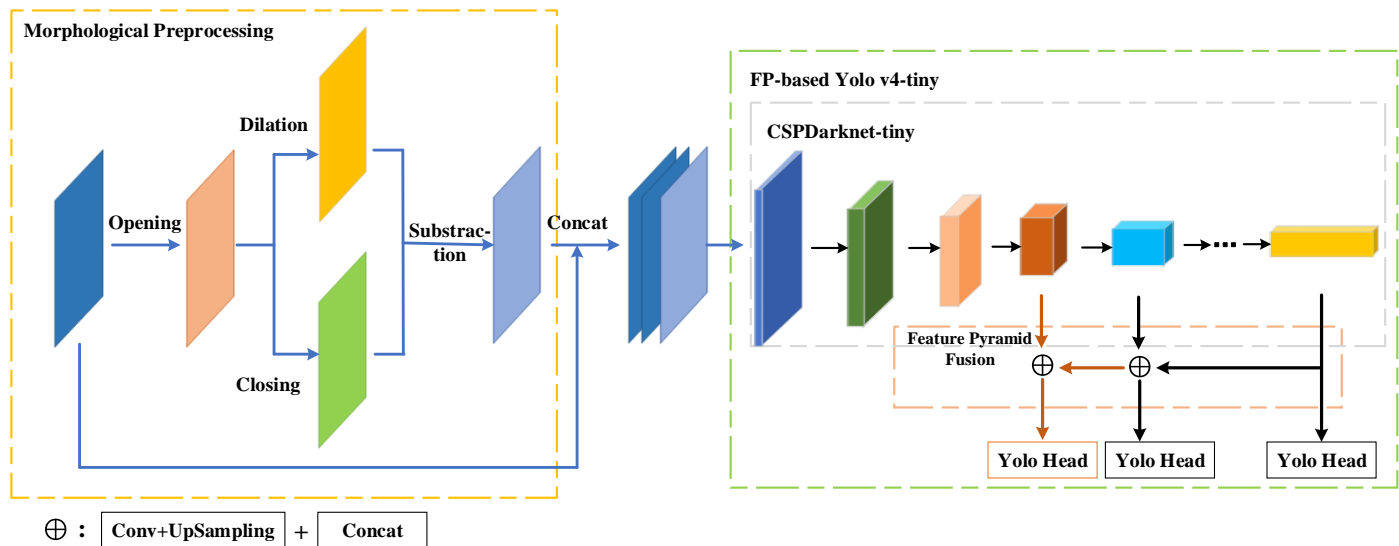


Fig 1 Processing flow of the proposed network.

A. Processing Flow of the Mor-FP Yolo v4-tiny

The processing procedure of the Mor-FP Yolo v4-tiny is shown in Fig 1.

Firstly, the original images are sent into the morphological preprocessing module which reduces the speckle noise and enhances the edge information with the trainable morphological kernels adjusted by the subsequent network. Thus, the feature maps with the same size and relative position as the original images are obtained. Second, the feature maps are combined with the original images to recover the information lost in the denoising and edge extraction procedures. Third, a Yolo v4-tiny network combined with the feature pyramid fusion structure tailored for SAR ship detection is designed. At last, the Yolo head module gives the final results.

B. Feature Enhancement Based on Morphological Preprocessing Module

The morphology preprocessing module is composed of denoising and contour extraction. The generated feature map is used to expand the image channel to construct a multi-channel SAR image.

1) Construction of morphological module

Classical morphological algorithms are defined by dilation and erosion, which have great effects in image processing, such as boundary extraction and denoising. Considering the SAR images processed in this paper are all grayscale images, the grayscale morphological operations are used during processing.

Let $f(x, y)$ be the original grayscale image and $b(x, y)$ be the structure element. The equations defining dilation and erosion are:

$$(f \oplus b)(s, t) = \max\{f(s - x, t - y) + b(x, y) \mid (s - x), (t - y) \in D_f, (x, y) \in D_b\} \quad (1)$$

$$(f \ominus b)(s, t) = \min\{f(s + x, t + y) - b(x, y) \mid (s + x), (t + y) \in D_f, (x, y) \in D_b\} \quad (2)$$

where \oplus is dilation and \ominus is erosion. D_f, D_b present the domain of definition of f and b respectively.

The image processing effect is always greatly influenced by the shape and values of the kernel. The kernels of morphology are usually chosen as fixed shapes, such as ellipse, rectangle, cross, which have fixed values. The choice of kernels is based on professional experience. Therefore, converting the kernels of morphology ($b(x, y)$) into trainable parameters set initialized randomly and optimized through back-propagation makes the morphological module adaptively match the target detection tasks, in either directivity or boundary thickness. The proposed morphological module is composed of trainable morphological layers.

The morphological module is composed of a denoising part and an edge extraction part, as shown in the orange dashed box in Fig 2. Considering the speckle noise in the SAR image is light noise, an opening operator is used to reduce the speckle noise. The opening operator is composed of an erosion operator and a dilation operator. The erosion operation sets the

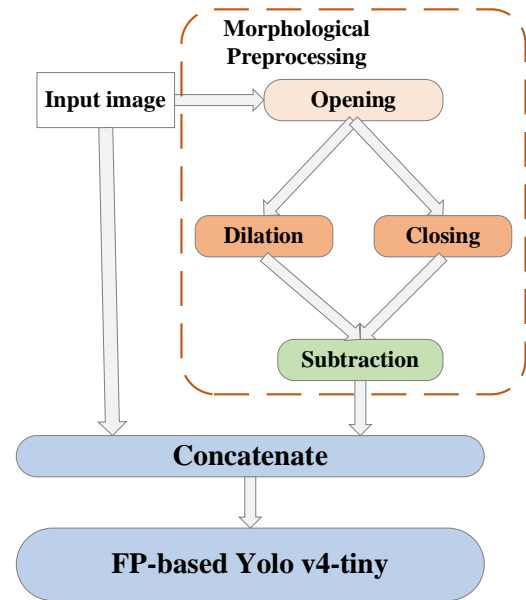


Fig 2 Design of morphological module.

central pixel as the minimum value of the difference between the kernel function and the adjacent pixels, while the dilation operator sets the central pixel as the maximum value of the sum of the kernel function and the adjacent pixels. Therefore, the opening operator can make the image blurred. Moreover, when the kernels of the operator become larger, the image becomes fuzzier. In this model, the sizes of the kernels in this algorithm are set as small as possible, 3×3 .

Fig 3 shows the results of morphological operation. Fig 3 (a) shows the original images, and (b) are the images after denoising operation. It is shown that the images in Fig 3 (b) are more blurred. Then the edge extraction module combined with morphological network is adopted to process the image after denoising to enhance the edge information of the image.

The edge extraction part is composed of dilation, closing, and subtraction:

$$g = (f \oplus b_1) - (f \bullet b_2) \quad (3)$$

where \bullet designates grayscale closing operator. b_1, b_2 present different morphological elements trained by the network.

The dilation operator provides the coarsening of the image, and the closing operator has a smoothing effect, whose coarsening effect is weaker than that of dilation. As the homogeneous regions are unaffected, the subtraction operation tends to eliminate homogeneous areas. Therefore, the results are the edge of the area, producing a difference-like effect, as shown in Fig 3(c). The preprocessing process reduces the impact of speckle noise, enhances the boundary information, and can provide clearer prior information of the target boundary for the subsequent detection network so that the entire network performs better.

Finally, we incorporate morphological module into the network as first few layers, which together with the subsequent Yolo v4-tiny form the entire target detection network.

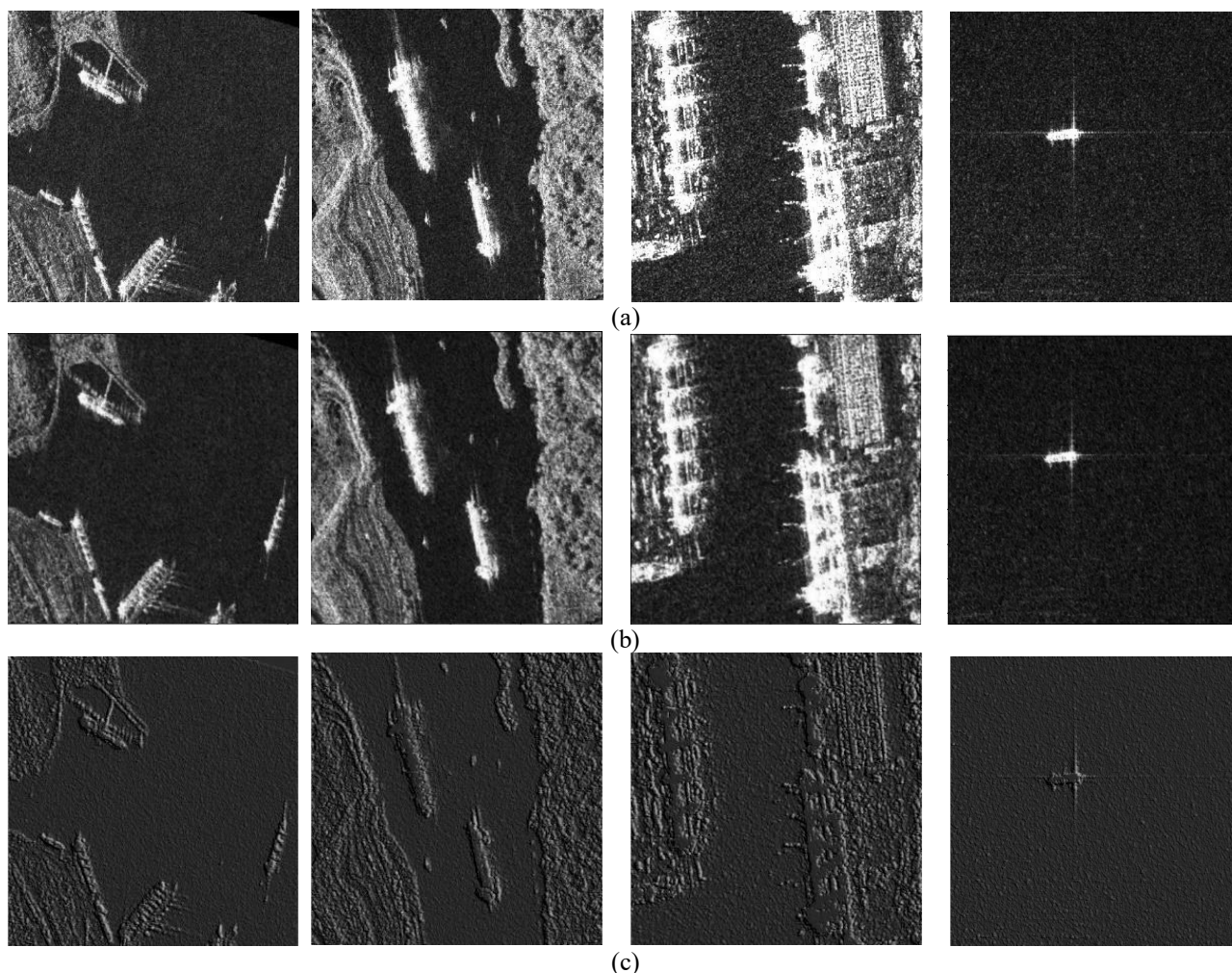


Fig 3 Results of morphological operation.
(a) Original images (b) denoising images (c) images after edge enhancement

2) Multi-channels image construction

The enhanced images obtained after edge extraction operation are shown in Fig 3(c). Then, the original images are combined with the images processed by the morphological module. More narrowly, the original images are taken as the first and second channel, and the morphology enhanced images are taken as the third channel, as shown in Fig 2. This method retains all the information of the original images and enhances the edge information. Compared with the case that the three channels are all original images, this method reduces the influence of speckle noise, enhances high-frequency contour information, and greatly enriches the information contained in the training data. In summary, the data from preprocessing part can guide the network to make feature mining and selection of the original image and edge information. The result of mining and selection can be mapped into the CNN to improve model training efficiency, reduce noise interference and enhance the sensitivity and accuracy of detection.

C. FP-based Yolo v4-tiny

SAR image ship detection is mainly used in the military

domain and marine safety, which have high demand on timeliness, so it is necessary to choose a lightweight and efficient detection algorithm as the backbone of the network. Therefore, we take the Yolo v4-tiny network as the basic detection algorithm.

Based on YOLOv3, Yolo v4 takes CSPDarknet53 as the backbone of the network and introduces SPP and PANet into the network to enhance the effect of feature extraction.

Yolo v4-tiny is a lightweight network of Yolo v4. In order to get a faster detection speed, Yolo v4-tiny takes the CSPdarknet-tiny network as the backbone. CSPdarknet-tiny network is successively composed of two convolution layers, three CSPBlock modules, and two convolution layers. CSPBlock module takes the infrastructure of CSPNet, adopting convolution layers and the structure of ResNet in local transition layers. In other words, the CSPBlock module is composed of convolution layers, skip connections and feature concatenate, as shown in Fig 4. Compared with the 60 million parameters in Yolo v4, Yolo v4-tiny has only 6 million parameters, which gives it better training speed and detection speed performance.

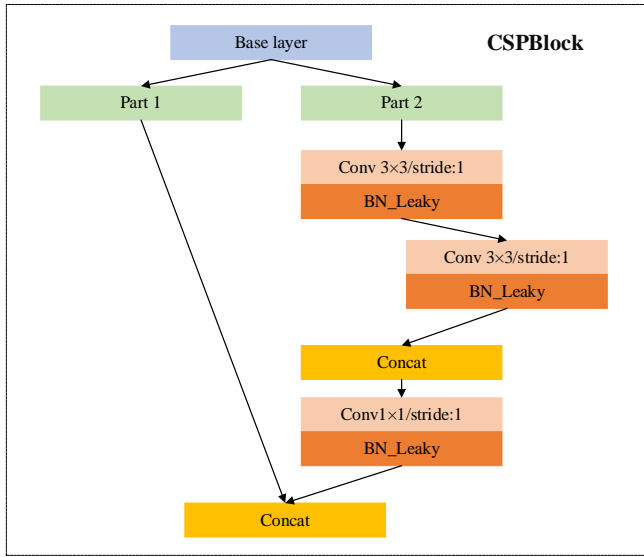


Fig 4 Structure of CSPBlock

Yolo v4 uses three feature maps as the input of the detection part, while Yolo v4-tiny uses two smaller feature maps as the final feature maps to be detected, which causes the decrease of small targets detection rate. However, in the SAR ship images, the targets occupy a small space and the target sizes vary a lot. It is known that high-level features reflect abundant semantic information, while low-level features have better target resolution and more target details. To improve the detection rate of small targets, the network needs to integrate more low-level detailed information. In this way, it is conducive to extracting not only semantic features of large-scale targets, but also detailed features of small-scale targets, ensuring the detection capability and detection accuracy of multi-scale targets.

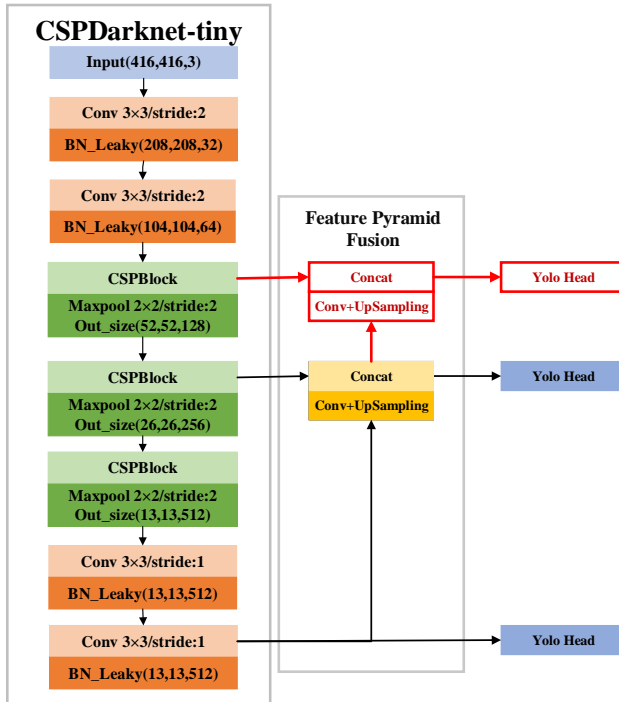


Fig 5 Structure of FP-based Yolo v4-tiny

Therefore, we add a detection module for shallow feature maps with larger sizes into the network to detect small targets, which is called a feature pyramid fusion structure. We take the feature maps of 52×52 to fuse with other feature maps of 26×26 and 13×13 based on the actual size of the target. The constructed network is shown in Fig 5, where the red part is the added part of the network.

III. EXPERIMENT RESULTS

In this section, the experiments are implemented to evaluate the detection performance of the Mor-FP Yolo v4-tiny. At first, the data sets and the detailed settings of related experiments are illustrated. Then the best combination of the morphological preprocessing module is explored, and ablation experiments are implemented to evaluate the performance of each component of the network. At last, the algorithm proposed in this paper is compared with other detection methods.

A. Data Sets and Settings

In our experiments, the performance of the proposed method is evaluated and analyzed on two different SAR ship data sets, SSDD and AIR SARShip-1.0.

1) *SSDD*: The SSDD data set is constructed by Li et al [45], containing multi-scale ships in different environments, including different polarization modes, resolutions, and scenes. The data is mainly obtained from RadarSat-2, TerraSAR-X, and Sentinel-1 sensors with four polarization modes: HH, HV, VV, and VH. Moreover, the resolution of SAR images in the data set ranges from 1 to 15m. Ship targets in the images are located in large areas of sea and nearshore, which are various and abundant.

There are 1160 images in the data set and 2456 ships in the images. The images are cut into sizes of about 500×500 pixels and labeled manually. To make the data set easier to process, we transform the image size to 416×416 , as shown in Fig 6, and convert the annotation information into a standard XML format. The label of each target is represented as (x, y, w, h) . (x, y) denotes the top-left coordinate of the rectangle label. w represents the width of the box and h represents the height.

2) *AIR SARShip-1.0*: AIR SARShip-1.0 data set is a multi-scenario, multimode SAR ship data set published by Aerospace Information Research Institute [46]. The data set contains 31 large views with 1-m and 3-m spatial resolution under single-polarization from Gaofen-3, which is a C-band multi-polarization high-resolution synthetic aperture radar satellite.

The images of the data set have different sea conditions, scenes, and the number of ships, and most of the image sizes are 3000×3000 pixels. To make the data set easier to process, we cut the raw images into 416×416 . Therefore, we can get 930 images with ship targets. The annotation information is the same as SSDD.

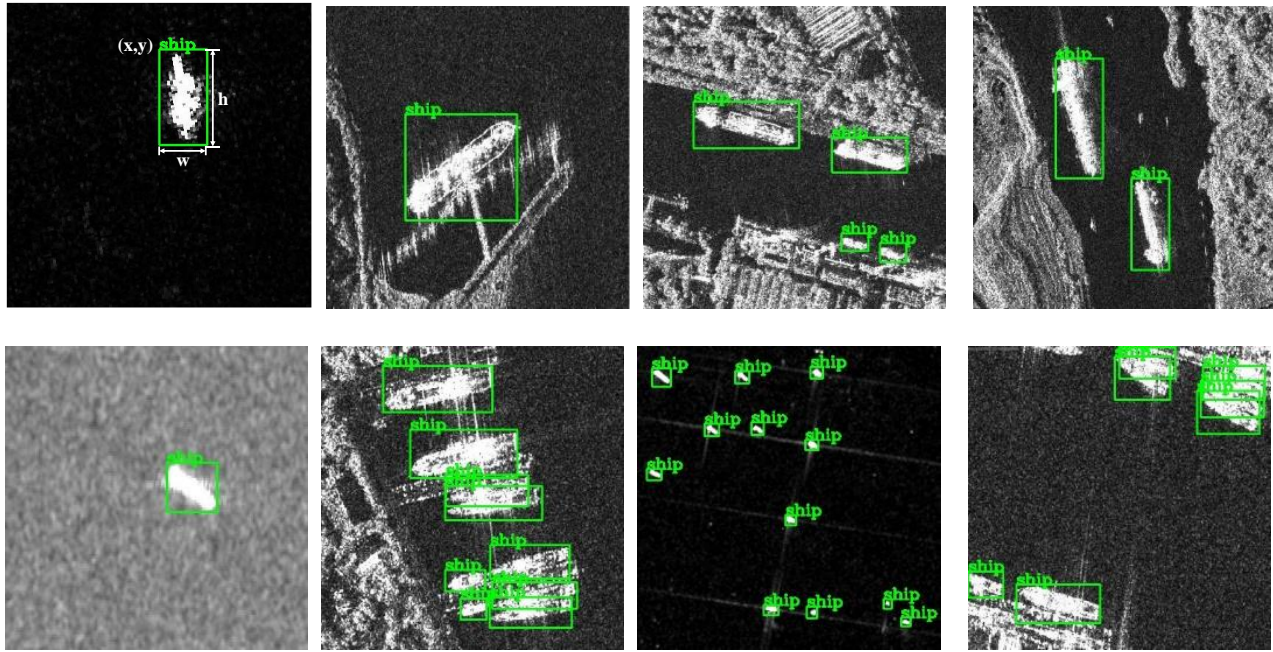


Fig 6 Sample images and labels on SSDD.

3) *Settings*: In the experiment of this chapter, the data sets are randomly divided into three parts according to the ratio of 8:1:1, which are respectively the training set, validation set, and test set. Before training, we obtain the different sizes of anchor boxes through the K-means algorithm. Since the Yolo head of each scale in this chapter sets three anchors, we obtain nine anchors, whose sizes are shown in Table 1 and Table 2 for SSDD and AIR SARShip-1.0, respectively.

Because the number of remote sensing images is limited, CSPdarknet-tiny is pretrained on the PASCAL VOC data set and then domain-specifically fine-tuned to adapt to remote sensing images. The commonly used ADAM algorithm [47] is taken as the gradient optimization algorithm. Moreover, the training process is divided into two parts. First, we freeze the parameters in the backbone while training the parameters in the morphological operation part, feature pyramid fusion part, and detection part. In this stage, the initial learning rate is set

Table 1 The sizes of 9 anchors on SSDD.

Number of Anchors	1	2	3	4	5	6	7	8	9
Width	10	10	12	19	20	34	34	53	92
Height	13	26	40	20	62	77	28	141	47

Table 2 The sizes of 9 anchors on AIR SARShip-1.0.

Number of Anchors	1	2	3	4	5	6	7	8	9
Width	10	21	29	40	53	64	81	87	93
Height	26	37	62	26	70	48	25	124	43

to 0.001, the batch size is set to 32, and the training epoch is set to 50. And we take CosineAnnealing as the learning rate adjustment method. In the latter 100 epochs of unfrozen training, we set a lower initial learning rate to 0.0001. The other settings are the same as the first stage.

All experiments are implemented in the Keras framework and carried out on a computer with an Nvidia GeForce RTX 3090 card. The operating system is Linux with CUDNN v8.

B. Evaluation Criteria

To evaluate the performance of the detection algorithm quantitatively, we adopt the following indexes to contrast different algorithms, including precision, recall, average precision (AP), and F1-score.

These evaluation criteria are calculated based on four components: true positive (TP), true negative (TN), false positive (FP), false negative (FN). In this paper, TP and TN indicate the number of correct detected ships and correct backgrounds, respectively. FP represents the number of false alarms, and FN is the number of undetected ships. In order to judge whether the detected frame is correct, Intersection over Union (IoU) is introduced. IoU is calculated as the ratio of the overlap between the bounding box and the single true box:

$$IoU = \frac{S_{\cap}}{S_{\cup}} \quad (4)$$

where S_{\cap} denotes the area of intersection of predict frame and true frame, while S_{\cup} is the area of the concurrent set of the two.

A detected box can be judged right if IoU is greater than the standard, set to 0.5 there.

The formulas of precision rate, recall rate, AP, F1-score are as follows:

$$precision = \frac{TP}{TP + FP} \quad (5)$$

$$recall = \frac{TP}{TP + FN} \quad (6)$$

$$AP = \int_0^1 P(R) dR \quad (7)$$

$$F1 = \frac{2 \times precision \times recall}{precision + recall} \quad (8)$$

Precision indicates the correct proportion of all predicted targets. The recall represents the proportion of correctly located and the proportion of identified targets in the total number of targets. We use AP and F1-score to evaluate the balance between precision and recall. AP computes the average value of precision over the interval from recall=0 to recall=1. F1-score is the harmonic average of the two. The higher the two values, the better the detection performance.

C. Evaluation of Morphological Structures

In this section, some trainable morphological operators are selected for comparative experiments that are carried out on SSDD.

First, four denoising operators are selected for experiments, and are shown as follows:

$$c_1 = f \otimes a_1 \quad (9)$$

$$c_2 = f \ominus a_2 \quad (10)$$

$$c_3 = f \bullet a_3 \quad (11)$$

$$c_4 = f \circ a_4 \quad (12)$$

where \oplus is the dilation, \ominus is the erosion, \circ, \bullet designate grayscale opening and closing operator, respectively. a_1, a_2, a_3, a_4 present different morphological elements trained by the network.

We take Yolo v4-tiny combined with the feature pyramid fusion module as the basic network. Except for the different

morphological denoising operators, the structure of networks and parameter settings are all the same. Table 3 provides the results of the denoising operations. It is noticed that the methods carried out with different morphological denoising operators perform differently. As the boundary blurs, some information loses and the detection performance becomes worse. In these operators, the methods of opening and closing with more information saved are better than those of dilation and erosion. Meanwhile, the method of c_4 performs the best with a similar testing speed.

Next, we take the same settings with the denoising operators for the edge extraction operations and select four edge extraction operators, and are shown as follows:

$$g_1 = (f \bullet b_1) - f \quad (13)$$

$$g_2 = f - (f \circ b_2) \quad (14)$$

$$g_3 = (f \oplus b_3) - (f \bullet b_4) \quad (15)$$

$$g_4 = (f \circ b_5) - (f \ominus b_6) \quad (16)$$

where $b_1 \sim b_6$ present different morphological elements trained by the network.

The detection performance of these operators is shown in Table 4. It can be observed that the methods combined with morphological edge extraction perform better than the basic network. Among them, the edge extraction method of g_3 has the best overall detection performance that AP achieves 95.44% and F1-score achieves 0.91. Furthermore, the testing time of the method increases by 0.02s approximately.

Then, we explore the methods that combine the above two parts. The morphological operators are shown as follows:

$$h_1 = (f \circ d_1) \oplus d_2 - (f \circ d_1) \bullet d_3 \quad (17)$$

$$h_2 = (f \bullet d_4) \oplus d_5 - (f \bullet d_4) \circ d_6 \quad (18)$$

$$h_3 = (f \circ d_7 \bullet d_8) \oplus d_9 - (f \circ d_7 \bullet d_8) \bullet d_{10} \quad (19)$$

$$h_4 = (f \bullet d_{11} \circ d_{12}) \oplus d_{13} - (f \bullet d_{11} \circ d_{12}) \bullet d_{14} \quad (20)$$

where $d_1 \sim d_{14}$ present different morphological elements trained by the network.

Table 3 Comparison of different denoising operators.

Denoising Operator	Recall	Precision	AP	F1	Params	Time(s)
None	83.89%	89.85%	93.39%	0.87	6,353,174	0.044
c_1	81.52%	87.76%	92.52%	0.85	6,353,201	0.064
c_2	86.73%	93.37%	93.17%	0.9	6,353,201	0.064
c_3	84.36%	92.71%	93.24%	0.88	6,353,210	0.067
c_4	88.15%	90.73%	93.3%	0.9	6,353,210	0.068

Table 4 Comparison of different edge extraction operators.

Edge Operator	Recall	Precision	AP	F1	Params	Time(s)
None	83.89%	89.85%	93.39%	0.87	6,353,174	0.044
g_1	88.63%	93.5%	95.13%	0.91	6,353,210	0.062
g_2	89.1%	90.82%	93.95%	0.9	6,353,210	0.064
g_3	89.1%	93.53%	95.44%	0.91	6,353,237	0.064
g_4	86.73%	90.59%	94.29%	0.89	6,353,237	0.065

Table 5 Comparison of different preprocessing methods.

Operator	Recall	Precision	AP	F1	Params	Time(s)
h_1 (proposed)	88.63%	93.5%	96.36%	0.91	6,353,237	0.066
h_2	87.2%	94.36%	95.55%	0.91	6,353,237	0.073
h_3	87.2%	92.46%	94.98%	0.9	6,353,255	0.083
h_4	88.63%	94.92%	95.67%	0.92	6,353,255	0.081

The results of edge extraction methods combined with denoising are illustrated in Table 5. It is obvious that the combination of the two operations has a better effect. Among the methods in Table 5, because more valid information in the image is removed with the increase of the denoising structure, the first two structures have better performance than the last two structures. Meanwhile, the testing time increases. As mentioned above, the method with closing expands the boundaries of the target and brings difficulties to detection. The method with h_1 has the best performance in the methods with higher accuracy and efficiency. To further substantiate our conclusion, it is displayed in Fig 7. that h_1 has a better performance than the other contrast methods. In summary, the use of trainable morphological denoising and edge extraction modules in the detection network can obtain more accurate boundary information and improve detection performance.

D. Ablation Study

In this section, three ablation experiments are applied on SSDD to evaluate the effects of the morphological preprocessing model and feature pyramid fusion structure in detail.

The first experiment adopts the original Yolo v4-tiny. The second experiment adds the feature pyramid fusion model based on Yolo v4-tiny. And the third experiment adopts the

morphological module as the preprocessing process of the second experiment. Except for the feature pyramid fusion structure and morphological model, the structure of the three networks and parameter settings are all the same.

Some test results can be seen in Fig 8(a) shows the ground truth, while Fig 8(b)-(d) shows the detection results of the above methods respectively. In Fig 8(a), the green rectangles represent correct ship targets, while in Fig 8(b)-(d), the red rectangles are the detection results of these three methods. In the ground truth and the results of the detection, a single rectangle represents a single target. Different scenes are shown in Fig 8, including inshore or offshore locations, from the prospect or close view, clear or degraded due to noise.

The pyramid fusion of multi-scale features contains both the image's high-level semantic and detailed information, and brings more detailed information compared with Yolo v4-tiny. As shown in the first three rows of Fig 8(c), the detection effect of the network on small targets is improved. However, the detection effect is unsatisfactory for targets inshore and targets close to or overlapping, as shown in the fourth to sixth rows of Fig 8(c). Then, the morphological model is absorbed into the FP-based Yolo v4-tiny as the preprocessing of the detection network, which reduces speckle noise, brings more edge information, and enhances the detection effect. As shown in the fourth to sixth rows of Fig 8(d), the algorithm detects more targets, and target locations are more accurate than the

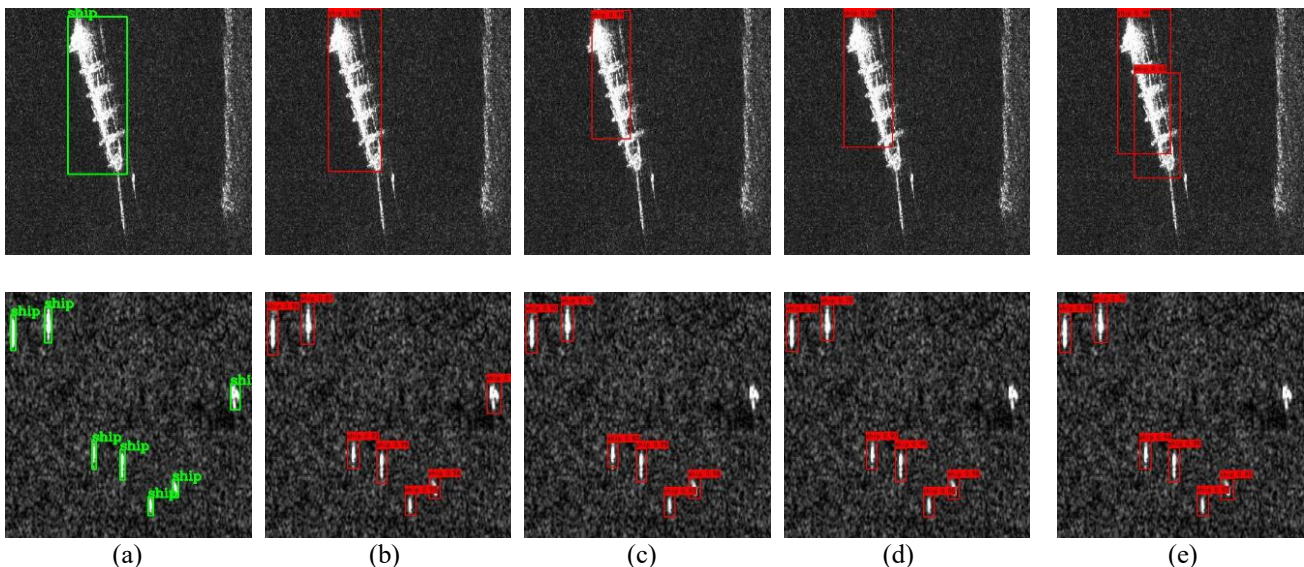


Fig 7 Effect of edge extraction methods combined with denoising. (a) Ground truth. (b)Results of h_1 . (c) Results of h_2 . (d) Results of h_3 . (e) Results of h_4 . The green rectangles in (a) are correct ship targets. The red rectangles in (b)-(e) indicate detected targets.

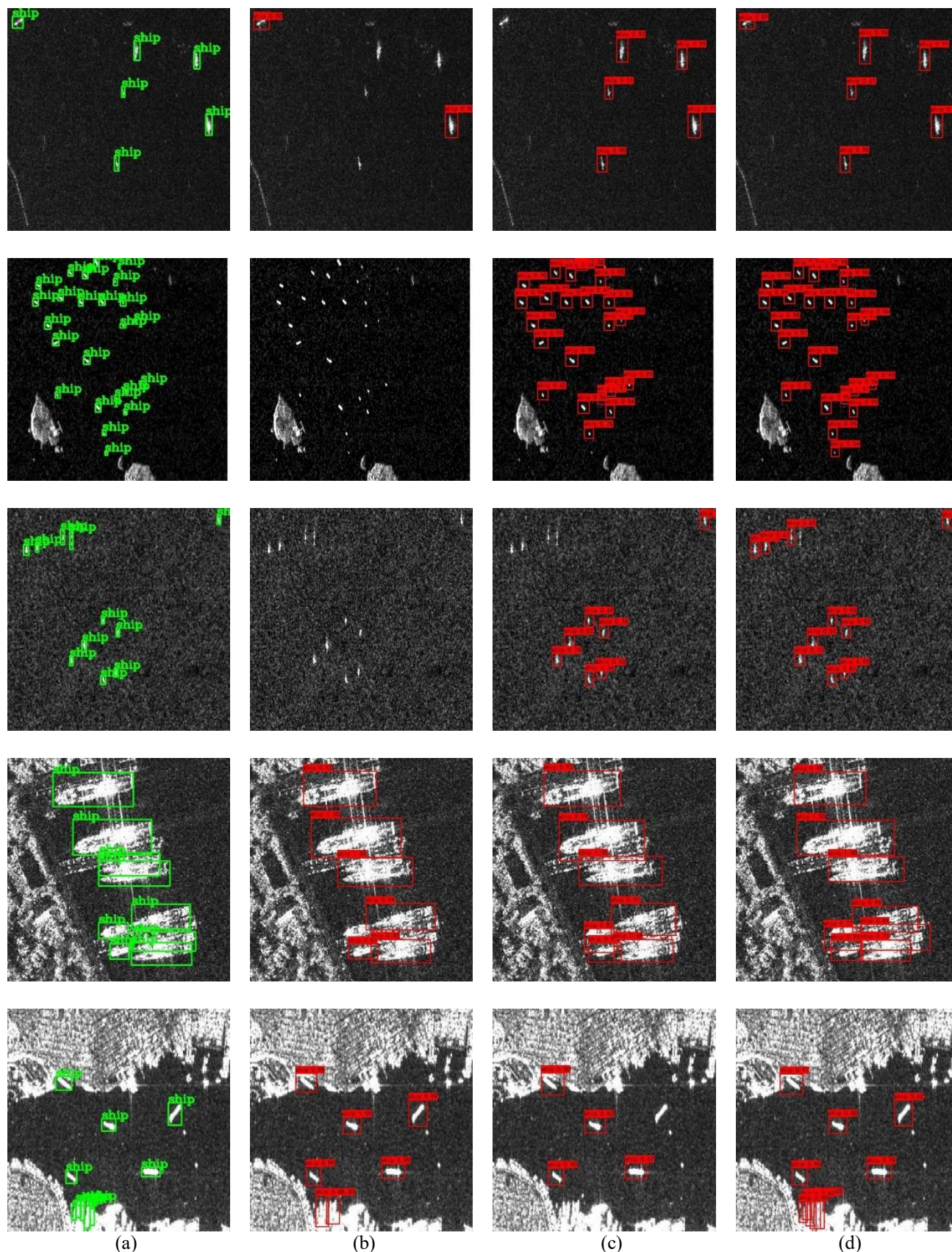


Fig 8(1) Effect of ablation studies. (a) Ground truth. (b) Results of Yolov4-tiny. (c) Results of FP-based Yolov4-tiny. (d) Results of Mor-FP Yolo v4-tiny. The green rectangles in (a) are correct ship targets. The red rectangles in (b)-(d) indicate detected targets.

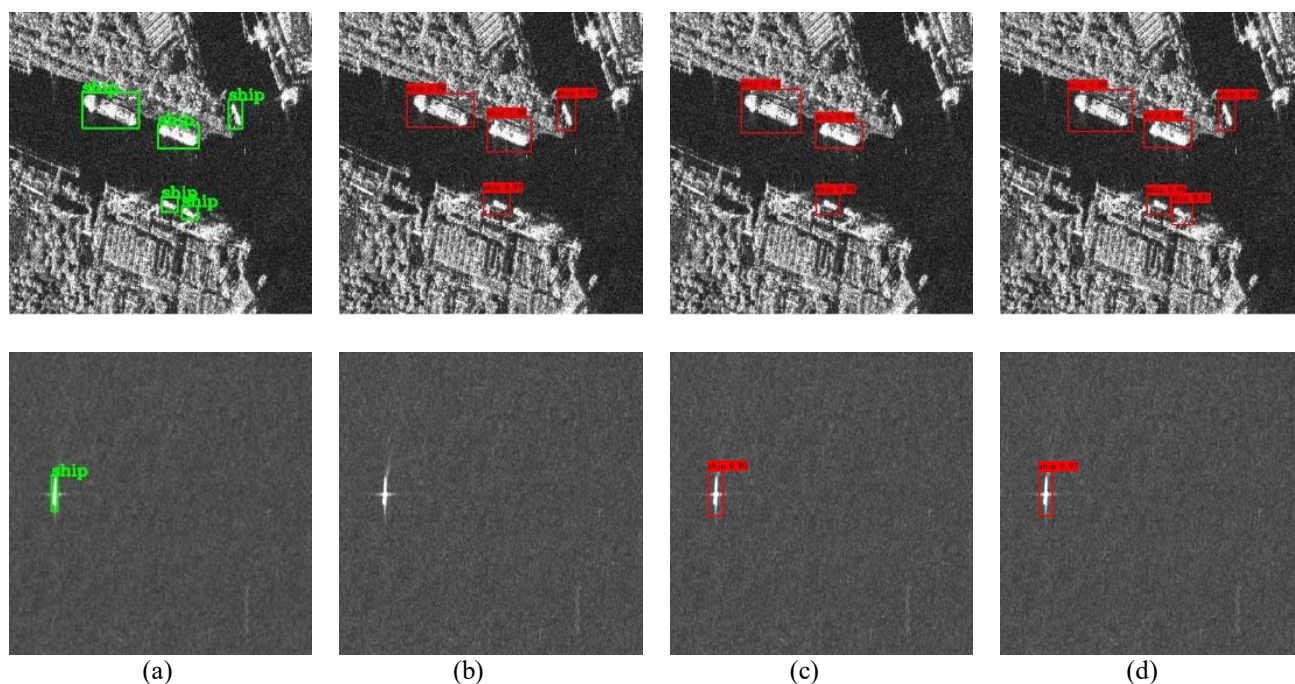


Fig 8(2) Effect of ablation studies. (a) Ground truth. (b) Results of YOLOv4-tiny. (c) Results of FP-based YOLOv4-tiny. (d) Results of Mor-FP YOLO v4-tiny. The green rectangles in (a) are correct ship targets. The red rectangles in (b)-(d) indicate detected targets.

methods aforementioned. In the fifth row of Figure 8, some ships are located in the lower part of the image, close to the coast, relatively close to each other, and challenging to distinguish. In the detection results of the YOLO v4-tiny and FP-based YOLO v4-tiny, they are only partially detected or completely undetected, while Mor-FP YOLO v4-tiny can detect more accurately with the help of the morphological model. In summary, the proposed method achieves a better detection performance than other methods.

The evaluation indicators of the detection performance of these three methods are displayed in Table 6. It can be seen that the feature pyramid fusion model brings advances in the recall, AP, and F1-score. In addition, the preprocessing module brings improvements in all evaluation criteria. Especially, the detection performance of the Mor-FP YOLO v4-tiny is 8% and 0.05 higher than YOLO v4-tiny in terms of AP and F1-score, respectively. The improvement of the F1-score signifies the balance between recall and precision. Meanwhile, the number of parameters increases, and testing time increases by 0.03s in total, as shown in Table 6. It is confirmed that the proposed method adopting feature pyramid fusion model and morphological model achieves a good detection performance for multi-scale ships in SAR images.

E. Comparison with CNNs

In this part, some CNN-based ship detection methods are contrasted with the Mor-FP YOLO v4-tiny, including SSD, CenterNet, YOLO v4, YOLO v4-tiny, YOLOx-tiny. They are tested on SSDD and AIR SARShip-1.0. As the datasets contain images of different polarization modes and resolutions, the experiments verify the robustness of the method. These methods are carried out based on the same train and test sets as the proposed method. Meanwhile, the settings of the methods adopt the default parameters. To evaluate the overall detection performance of these methods quantitatively, evaluation criteria mentioned in III-B are utilized in this part, and the results are shown in Table 7 and Table 8.

As can be seen from the results on SSDD in Table 7, the AP of the Mor-FP YOLO v4-tiny has achieved 96.36%, which is 7%, 3%, 6.2%, 8.1%, 1.39% higher than SSD, CenterNet, YOLO v4, YOLO v4-tiny, YOLOx-tiny, respectively. Although SSD, CenterNet, and YOLO v4 have outstanding performance in terms of precision, the recalls of these methods are not satisfactory, which brings the imbalance of precision and recall and a lower F1-score. Meanwhile, the proposed method has fewer parameters, about 1/3 of SSD, 1/5 of CenterNet,

Table 6 Ablation Studies of the proposed method.

Methods	Recall	Precision	AP	F1	Params	Time(s)
YOLOv4-tiny	81.04%	91.94%	88.25%	0.86	5,880,324	0.039
FP-based YOLOv4-tiny	83.89%	89.85%	93.39%	0.87	6,353,174	0.044
Mor-FP YOLO v4-tiny (proposed)	88.63%	93.5%	96.36%	0.91	6,353,237	0.066

Table 7 Comparison with Other Detection Methods on SSDD.

Networks	Recall	Precision	AP	F1	Params	Time(s)
SSD	44.74%	96.93%	89.99%	0.61	23,745,908	0.17
CenterNet	73.93%	98.11%	93.33%	0.84	32,718,597	0.21
Yolo v4	83.4%	93.91%	90.18%	0.88	64,002,306	0.32
Yolo v4-tiny	81.04%	91.94%	88.25%	0.86	5,880,324	0.047
Yolox-tiny	90.53%	91.12%	94.97%	0.91	5,050,194	0.044
Mor-FP Yolo v4-tiny (proposed)	88.63%	93.5%	96.36%	0.91	6,353,237	0.066

Table 8 Comparison with Other Detection Methods on AIR SARShip-1.0.

Networks	Recall	Precision	AP	F1	Params	Time(s)
SSD	46.44%	96.93%	87.77%	0.62	23,745,908	0.13
CenterNet	62%	95.22%	83.66%	0.75	32,718,597	0.21
Yolo v4	79.57%	91.13%	87.71%	0.85	64,002,306	0.25
Yolo v4-tiny	70.39%	89.92%	80.3%	0.79	5,880,324	0.047
Yolox-tiny	76.36%	85.22%	88.68%	0.81	5,050,194	0.042
Mor-FP Yolo v4-tiny (proposed)	76.97%	92.13%	89.55%	0.84	6,353,237	0.084

1/10 of Yolo v4. On the other hand, lightweight networks, including Yolo v4-tiny and Yolox-tiny, have higher detection efficiency but are not good at detection accuracy. As shown in Fig 9, Mor-FP Yolo v4-tiny has taken a balance between accuracy, lightweight, and detection speed, with faster detection speed than classic CNNs and better detection accuracy than the other state-of-the-art lightweight networks. To further verify the generalization ability of the proposed method, all the algorithms are tested on the AIR SARShip-1.0. The results are shown in Table 8 and Fig 10. The proposed method also has a good performance on the AIR SARShip-1.0, demonstrating the generalization ability and robustness of the method. Therefore, it can be concluded that the proposed

method has the best performance in the mentioned algorithms. Considering the requirements of high precision, high efficiency, and lightweight for SAR ship detection, the proposed method is better than other mentioned CNN-based ship detection methods.

IV. CONCLUSION

This paper proposes a SAR ship detection method based on edge extraction and morphological network to mitigate the influence of background clutter on network training and enhance the edge information of targets. Besides, this paper

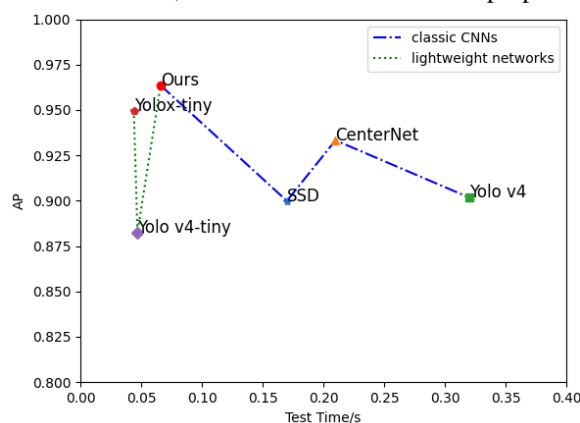


Fig 9 Effect of AP and test time of different detection algorithms on SSDD.

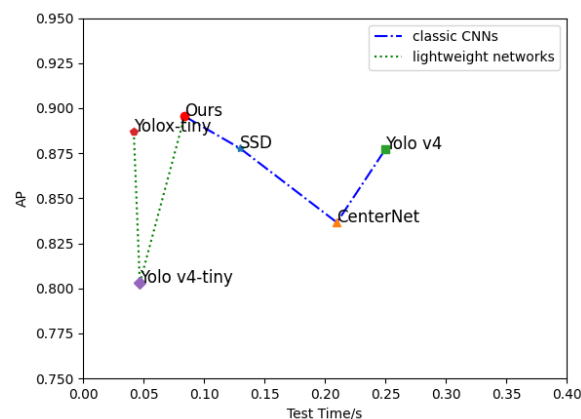


Fig 10 Effect of AP and test time of different detection algorithms on AIR SARShip-1.0.

proves the combination of the feature pyramid fusion structure with the Yolo v4-tiny can improve the detection accuracy. Experiments with different polarization modes and resolutions show that the proposed method has good robustness. The proposed network can not only be used for target detection, but also for target recognition and segmentation. Further work can be done to explore the best network structure for different application directions and the relationships between the change of morphology parameters and data resolution.

REFERENCES

- [1] S. Brusch, S. Lehner, T. Fritz, M. Soccorsi, A. Soloviev and B. van Schie, "Ship Surveillance With TerraSAR-X," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 3, pp. 1092-1103, March 2011, doi: 10.1109/TGRS.2010.2071879.
- [2] X. Wei, X. Wang, J. Chong. Local region power spectrum-based unfocused ship detection method in synthetic aperture radar images [J]. in *Journal of Applied Remote Sensing*, vol. 12, no. 1, January 2018, doi: 10.1117/1.JRS.12.016026
- [3] W. Huo, Y. Huang, J. Pei, et al. Ship detection from ocean SAR image based on local contrast variance weighted information entropy [J]. in *Sensors*, vol. 18, no. 4, April 2018, doi: 10.3390/s18041196
- [4] Yang M, Guo C. Ship Detection in SAR Images Based on Lognormal ρ -metric [J]. in *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 9, pp. 1372-1376, Sept. 2018, doi: 10.1109/LGRS.2018.2838043.
- [5] Eldhuset K. An automatic ship and ship wake detection system for spaceborne SAR images in coastal regions [J]. in *IEEE transactions on Geoscience and Remote Sensing*, vol. 34, no. 4, pp. 1010-1019, July 1996, doi: 10.1109/36.508418.
- [6] H. Shi, Q. Zhang, M. Bian, et al. A novel ship detection method based on gradient and integral feature for single-polarization synthetic aperture radar imagery [J]. in *Sensors*, vol. 18, no. 2, February 2018, doi: 10.3390/s18020563.
- [7] Fryer A C, Muller H J, Yanasse C C F, et al. A model for extremely heterogeneous clutter [J]. in *IEEE transactions on geoscience and remote sensing*, vol. 35, no. 3, pp. 648-659, May 1997, doi: 10.1109/36.581981.
- [8] Schwegmann C P, Kleynhans W, Salmon B P. Manifold adaptation for constant false alarm rate ship detection in South African oceans [J]. in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 7, pp. 3329-3337, July 2015, doi: 10.1109/JSTARS.2015.2417756.
- [9] X. Qin, S. Zhou, H. Zou and G. Gao. A CFAR Detection Algorithm for Generalized Gamma Distributed Background in High-Resolution SAR Images [J]. in *IEEE Geoscience and Remote Sensing Letters*, vol. 10, no. 4, pp. 806-810, July 2013, doi: 10.1109/LGRS.2012.2224317.
- [10] J. Liu, W. Liu, B. Chen, H. Liu and H. Li, "Detection Probability of a CFAR Matched Filter with Signal Steering Vector Errors," in *IEEE Signal Process. Lett.*, vol. 22, no. 12, pp. 2474-2478, Dec. 2015
- [11] Schwegmann C P, Kleynhans W, Salmon B P. Manifold adaptation for constant false alarm rate ship detection in South African oceans [J]. in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 7, pp. 3329-3337, July 2015, doi: 10.1109/JSTARS.2015.2417756.
- [12] O. Pappas, A. Achim and D. Bull, "Superpixel-Level CFAR Detectors for Ship Detection in SAR Imagery," in *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 9, pp. 1397-1401, Sept. 2018, doi: 10.1109/LGRS.2018.2838263.
- [13] Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
- [14] W. Liu et al., "SSD: Single shot multibox detector," in Proc. Eur. Conf. Comput. Vis. (ECCV), Amsterdam, The Netherlands, Oct. 2016, pp. 21–37
- [15] Lin T Y, Goyal P, Girshick R, et al. Focal Loss for Dense Object Detection [J]. in *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 42, no. 2, pp. 318-327, 1 Feb. 2020, doi: 10.1109/TPAMI.2018.2858826.
- [16] Fu C Y, Liu W, Ranga A, et al. DSSD: Deconvolutional Single Shot Detector [J]. in IEEE Conference on Computer Vision and Pattern Recognition, *arXivpreprint arXiv:1701.06659*, 2017.
- [17] J. Redmon and A. Farhadi, "YOLO9000: Better, Faster, Stronger," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 6517-6525, doi: 10.1109/CVPR.2017.690.
- [18] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement" in arXiv:1804.02767,2018,[online]Available:<https://arxiv.org/abs/1804.02767>.
- [19] R. Girshick, J. Donahue, T. Darrell and J. Malik, "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation," 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 580-587, doi: 10.1109/CVPR.2014.81.
- [20] R. Girshick, "Fast R-CNN," 2015 IEEE International Conference on Computer Vision (ICCV), 2015, pp. 1440-1448, doi: 10.1109/ICCV.2015.169.
- [21] S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, pp. 1137-1149, 1 June 2017, doi: 10.1109/TPAMI.2016.2577031.
- [22] K. He, G. Gkioxari, P. Dollár and R. Girshick, "Mask R-CNN," 2017 IEEE International Conference on Computer Vision (ICCV), 2017, pp. 2980-2988, doi: 10.1109/ICCV.2017.322.
- [23] Z. Cai and N. Vasconcelos, "Cascade R-CNN: Delving Into High Quality Object Detection," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 6154-6162, doi: 10.1109/CVPR.2018.00644.
- [24] Z. Lin, K. Ji, X. Leng and G. Kuang, "Squeeze and Excitation Rank Faster R-CNN for Ship Detection in SAR Images," in *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 5, pp. 751-755, May 2019, doi: 10.1109/LGRS.2018.2882551.
- [25] J. Jiao et al., "A Densely Connected End-to-End Neural Network for Multiscale and Multiscene SAR Ship Detection," in *IEEE Access*, vol. 6, pp. 20881-20892, 2018, doi: 10.1109/ACCESS.2018.2825376.
- [26] Z. Cui, Q. Li, Z. Cao and N. Liu, "Dense Attention Pyramid Networks for Multi-Scale Ship Detection in SAR Images," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 11, pp. 8983-8997, Nov. 2019, doi: 10.1109/TGRS.2019.2923988.
- [27] X. Qi, P. Lang, X. Fu et al., "A Regional Attention-based Detector for SAR Ship Detection," in *Remote Sensing Letters*, vol. 13, pp. 55-64, 2022, doi: 10.1080/2150704X.2021.1987574
- [28] R. Yang, Z. Pan, X. Jia, L. Zhang and Y. Deng, "A Novel CNN-Based Detector for Ship Detection Based on Rotatable Bounding Box in SAR Images," in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 1938-1958, 2021, doi: 10.1109/JSTARS.2021.3049851.
- [29] R. Qin, X. Fu, J. Chang and P. Lang, "Multilevel Wavelet-SRNet for SAR Target Recognition," in *IEEE Geoscience and Remote Sensing Letters*, doi: 10.1109/LGRS.2021.3050891.
- [30] Jiang J, Fu X, Qin R, Wang X, Ma Z. High-Speed Lightweight Ship Detection Algorithm Based on YOLO-V4 for Three-Channels RGB SAR Image. in *Remote Sensing*. vol. 13, no. 10, May 2021, doi: 10.3390/rs13101909.
- [31] J. Ai, R. Tian, Q. Luo, J. Jin and B. Tang, "Multi-Scale Rotation-Invariant Haar-Like Feature Integrated CNN-Based Ship Detection Algorithm of Multiple-Target Environment in SAR Imagery," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 12, pp. 10070-10087, Dec. 2019, doi: 10.1109/TGRS.2019.2931308.
- [32] J. Yin, Y. Lu, Z. Gong, Y. Jiang and J. Yao, "Edge Detection of High-Voltage Porcelain Insulators in Infrared Image Using Dual Parity Morphological Gradients," in *IEEE Access*, vol. 7, pp. 32728-32734, 2019, doi: 10.1109/ACCESS.2019.2900658.
- [33] Y. Xie et al., "Refined Extraction Of Building Outlines From High-Resolution Remote Sensing Imagery Based on a Multifeature Convolutional Neural Network and Morphological Filtering," in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 1842-1855, 2020, doi: 10.1109/JSTARS.2020.2991391.
- [34] H. Zhao et al., "Automatic Lumen Segmentation in Intravascular Optical Coherence Tomography Using Morphological Features," in *IEEE Access*, vol. 7, pp. 88859-88869, 2019, doi: 10.1109/ACCESS.2019.2925917.
- [35] Y. Li, Z. Li, B. Xu, C. Dang and J. Deng, "Low-Contrast Infrared Target

- Detection Based on Multiscale Dual Morphological Reconstruction," in *IEEE Geoscience and Remote Sensing Letters*, doi: 10.1109/LGRS.2021.3080986.
- [36] B. Wang, X. Tan, D. Song and L. Zhang, "Rapid Identification of Post-Earthquake Collapsed Buildings via Multi-Scale Morphological Profiles With Multi-Structuring Elements," in *IEEE Access*, vol. 8, pp. 122036-122056, 2020, doi: 10.1109/ACCESS.2020.3007255.
- [37] A. Wang, X. He, P. Ghamisi and Y. Chen, "LiDAR Data Classification Using Morphological Profiles and Convolutional Neural Networks," in *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 5, pp. 774-778, May 2018, doi: 10.1109/LGRS.2018.2810276.
- [38] Q. Zhang, X. Tu, F. Li and Y. Hu, "An Effective Chatter Detection Method in Milling Process Using Morphological Empirical Wavelet Transform," in *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 8, pp. 5546-5555, Aug. 2020, doi: 10.1109/TIM.2019.2958470.
- [39] C. Wang, W. Pedrycz, M. Zhou and Z. Li, "Sparse Regularization-Based Fuzzy C-Means Clustering Incorporating Morphological Grayscale Reconstruction and Wavelet Frames," in *IEEE Transactions on Fuzzy Systems*, vol. 29, no. 7, pp. 1826-1840, July 2021, doi: 10.1109/TFUZZ.2020.2985930.
- [40] G. Boato, D. Dang-Nguyen and F. G. B. De Natale, "Morphological Filter Detector for Image Forensics Applications," in *IEEE Access*, vol. 8, pp. 13549-13560, 2020, doi: 10.1109/ACCESS.2020.2965745.
- [41] G. Treece, "Morphology-Based Noise Reduction: Structural Variation and Thresholding in the Bitonic Filter," in *IEEE Transactions on Image Processing*, vol. 29, pp. 336-350, 2020, doi: 10.1109/TIP.2019.2932572.
- [42] Ranjan Mondal, Pulak Purkait, Sanchayan Santra, Bhabatosh Chanda: Morphological Networks for Image De-raining. *DGCI 2019*: 262-275
- [43] Z. Cui, X. Wang, N. Liu, Z. Cao and J. Yang, "Ship Detection in Large-Scale SAR Images Via Spatial Shuffle-Group Enhance Attention," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 1, pp. 379-391, Jan. 2021, doi: 10.1109/TGRS.2020.2997200.
- [44] J. Fu, X. Sun, Z. Wang and K. Fu, "An Anchor-Free Method Based on Feature Balancing and Refinement Network for Multiscale Ship Detection in SAR Images," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 2, pp. 1331-1344, Feb. 2021, doi: 10.1109/TGRS.2020.3005151.
- [45] J. Li, C. Qu and J. Shao, "Ship detection in SAR images based on an improved faster R-CNN," in *2017 SAR in Big Data Era: Models, Methods and Applications (BIGSAR DATA)*, 2017, pp. 1-6, doi: 10.1109/BIGSAR DATA.2017.8124934.
- [46] SUN Xian, WANG Zhirui, SUN Yuanrui, et al. AIR-SARShip-1.0: High-resolution SAR ship detection dataset[J]. *Journal of Radars*, 2019, 8(6): 852-862. doi: 10.12000/JR19097
- [47] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization", *Proc. Int. Conf. Learn. Representations*, May 2015.



Congxia Zhao was born in 1996. She received her B.Eng. degree from Jilin University in 2018. Currently, she is a PhD student in the School of Integrated Circuits and Electronics, Beijing Institute of Technology (BIT), Beijing, China. Her research interests include deep learning, signal processing, synthetic aperture radar image detection and recognition.



Xiongjun Fu was born in 1978. He received the B.Eng. degree and the Ph.D. degree from the Beijing Institute of Technology (BIT), China, in 2000 and 2005 respectively. He is currently the vice dean of the School of Integrated Circuits and Electronics, Beijing Institute of Technology, and an professor and

PhD Supervisor with BIT. His current research interests include radar system, radar signal processing, waveform design, and automatic target recognition.



Jian Dong received the B.Eng. degree in signal and information processing and the Ph.D. degree in navigation, guidance and control from Beijing Institute of Technology (BIT), BJ, China, in 2006 and 2009, respectively. He is currently a lecturer and master candidate supervisor with BIT.

His research interests include radar signal processing, inverse synthetic aperture radar imaging, waveform design, automatic target recognition and micro-Doppler feature extraction.



Rui Qin was born in 1990. He received his Master degree in Electronic Science and Technology from the Wuhan University of Technology, Wuhan, China in 2015 and Ph.D. degree in Information and Communication Engineering from Beijing Institute of Technology, Beijing, China in 2021.

He is now working in the School of Computer Science and Technology, Chongqing University of Posts and Telecommunications. His research interests include deep learning, signal processing, synthetic aperture radar target recognition and image classification.



Chang Jiayun was born in 1989. She received her master degree from Beijing Institute of Technology (BIT), China, in 2016. Currently, she is pursuing her Ph.D. degree in the School of Integrated Circuits and Electronics, BIT. Her research interests include automatic target recognition and radar signal processing.



Ping Lang received B.E degree in Machine Design Engineering from North University of China, Taiyuan, China in 2015 and M.E degree in Machine Science and Technology from Beijing Institute of Technology, Beijing, China in 2018. He is currently a PhD student in signal and information processing at School of Integrated Circuits and Electronics, Beijing Institute of Technology. He was a visiting PhD student at School of Computing, University of Leeds, UK, from Jul. 2019 to Oct.2019.

His research interests are radar signal processing and machine learning, including explainable machine learning algorithm, radar target detection and recognition, and radar system design.