An Anchor-Free Detection Method for Ship Targets in High-Resolution SAR Images

Zhongzhen Sun, Muchen Dai, Xiangguang Leng, *Member, IEEE*, Yu Lei, Boli Xiong, Kefeng Ji, *Member, IEEE*, and Gangyao Kuang, *Senior Member, IEEE*

***Abstract*—With the rapid development of earth observation technology, high-resolution synthetic aperture radar (HR SAR) imaging satellites could provide more observational information for maritime surveillance. However, there are still some problems to detect ship targets in HR SAR images due to the complex surroundings, targets defocusing, and diversity of the scales. In this article, an anchor-free method is proposed for ship target detection in HR SAR images. First, fully convolutional one-stage object detection (FCOS) as the base network is applied to detect ship targets, achieving better detection performance through pixel-by-pixel prediction of the image. Second, the category-position (CP) module is proposed to optimize the position regression branch features in the FCOS network. This module can improve target positioning performance in complex scenes by generating guidance vector from the classification branch features. At the same time, target classification and boundary box regression methods are redesigned to shield the adverse effects of fuzzy areas in the network training. Finally, to evaluate the effectiveness of CP-FCOS, extensive experiments are conducted on High-Resolution SAR Images Dataset, SAR Ship Detection Dataset, IEEE 2020 Gaofen Challenge SAR dataset, and two complex large-scene HR SAR images. The experimental results show that our method can obtain encouraging detection performance compared with Faster-RCNN, RetinaNet, and FCOS. Remarkably, the proposed method was applied to SAR ship detection in the 2020 Gaofen Challenge. Our team ranked first among 292 teams in the preliminary contest and won seventh place in the final match.**

***Index Terms*—Category-position (CP), fully convolutional one-stage object detection (FCOS), high-resolution (HR), ship detection, synthetic aperture radar (SAR).**

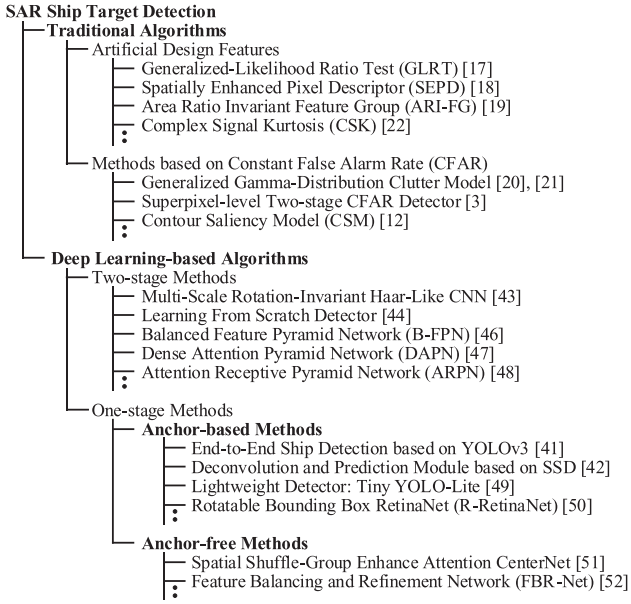

Fig. 1. Taxonomy of SAR ship target detection methods.

I. Introduction

SYNTHETIC aperture radar (SAR) can provide massive high-resolution (HR) images in day-and-night and all-weather conditions, which has unique advantages over other sensors such as optical sensor, infrared sensor, hyperspectral sensor, and so on [1], [2]. For this reason, it plays a significant role in maritime management and monitoring. Recently, with the rapid development of spaceborne SAR-imaging technologies, the quantity and quality of SAR data have been improved continuously. Therefore, more and more scholars conduct research on ship target detection in HR SAR images [3]–[12]. However, it is still challenging to detect ship targets in HR SAR images because of the complex surroundings and other tough problems, e.g., sidelobes, targets defocusing [13]–[16].

In the field of SAR ship detection, many algorithms have been proposed, which can be mainly divided into traditional algorithms [3], [12], [17]–[22] and deep learning based algorithms [41]–[52], as shown in Fig. 1. Generally speaking, traditional algorithms require artificial design features to distinguish between background and ship targets in SAR images. For example, Iervolino and Guida [17] considered both the sea clutter and the signal backscattered from ship targets in SAR images, and proposed a generalized-likelihood ratio test (GLRT) detector. Lang *et al.* [18] proposed a spatially enhanced pixel descriptor (SEPD) to enable the spatial structure information of ship targets, which improved the separability between ship targets and sea clutter. Leng *et al.* [19] defined the area ratio invariant feature group (ARI-FG) to modify the traditional detectors. Due

Manuscript received April 14, 2021; revised June 14, 2021 and July 6, 2021; accepted July 20, 2021. Date of publication July 26, 2021; date of current version August 18, 2021. This work was supported by the National Natural Science Foundation of China under Grant 62001480. *(Corresponding author: Boli Xiong.)*

Zhongzhen Sun, Xiangguang Leng, Yu Lei, Boli Xiong, Kefeng Ji, and Gangyao Kuang are with the College of Electronic Science and Technology, National University of Defense Technology, Changsha 410073, China (e-mail: sunzhongzhen14@163.com; luckight@163.com; 994208066@qq.com; blxiong@nudt.edu.cn; jikefeng@nudt.edu.cn; kuangyeats@hotmail.com).

Muchen Dai is with the Troop No. 65052 of PLA, Baicheng 137100, China (e-mail: 906182992@qq.com).

Digital Object Identifier 10.1109/JSTARS.2021.3099483

to the difference in scattering distribution between ocean and ship targets, detection methods based on constant false alarm rate (CFAR) [3], [12], [20]–[22] have been widely studied. Ao *et al.* [20] designed an efficient multiscale CFAR detector with a generalized Gamma distribution clutter model to detect candidate targets in the sea. Li *et al.* [3] proposed a superpixel-level two-stage CFAR method for ship detection in HR SAR images. However, these traditional algorithms generally produce better detection performance for some specific scenes, and their adaptability is insufficient. Besides, these methods based on CFAR severely rely on the appropriateness of the statistical model of clutter [21]. And the distribution parameters may be incorrectly estimated due to the contaminated background, which will cause severe missing detection [22].

On the one hand, the characteristics obtained by manual analysis are not stable enough, and the designed parameters are complex and changeable. Thus, it is often difficult to achieve better performance when the algorithms are transferred to other scene images. On the other hand, with the development of HR SAR, conventional detectors may suffer from performance deterioration due to the changes in target imaging quality and the influence of complex surroundings. Specifically, conventional resolution theory may not be strictly applied to the increased resolution SAR imagery. Besides, the ship targets appear to be more structured and shaped in HR SAR images and nevertheless contain many weak scatters around target regions [12], which brings great challenges for ship detection.

In recent years, the target detection methods based on convolutional neural networks (CNNs) have become a research hotspot in the field of remote sensing image target detection. Compared with the traditional methods, CNN can automatically learn the essential features from large amounts of image data and has higher accuracy and better robustness in the target detection [23], [24]. Generally speaking, according to the strategy of the target detection algorithm, CNN-based detection methods can be divided into two categories: two-stage methods and one-stage methods. The two-stage methods divide the target detection task into two parts, generating candidate regions and classifying, regressing the regions. Among them, RCNN [25] is the first deep learning based method applied to target detection. RCNN uses the selective search (SS) method to obtain possible candidate regions of the targets. After that, the obtained candidate regions are sent to the CNN network to extract features, and then the support vector machine [26] is adopted to complete the target classification. Although RCNN has better detection accuracy than traditional methods, it has poor real-time performance due to the multistage implementation of training and testing. Then, SPP Net [27] proposed an idea of sharing feature convolution graphs. It only needs to convolve the input image once, avoiding the problem of repeated calculations. In addition, this method adds spatial pyramid pooling, which solves the output problem of multiscale images. Compared with RCNN, it has better detection accuracy and speed. Based on the previous research, Girshick [28] proposed Fast R-CNN. It presents multitask learning, which simultaneously performs target classification and position regression. At the same time, Fast R-CNN introduces ROI pooling, which is equivalent to a single-layer SPP layer. This method improves the feature utilization rate, thereby improving the detection efficiency.

Subsequently, Ren *et al.* [29] proposed Faster R-CNN, which replaced SS by introducing the region proposal network (RPN). RPN can share convolutional features with the classifier, thus further improving the detection speed and accuracy. Mask RCNN [30] adds a semantic segmentation network on the basis of Faster R-CNN, and replaces ROI Pooling with bilinear interpolation (ROI Align). This method builds a multitask integrated network and enhances multilevel feature fusion. Cai and Vasconcelos [31] proposed Cascade R-CNN, which designs a multilevel cascade structure for obtaining a more accurate target position. Different from two-stage methods, one-stage methods do not generate candidate regions and directly obtain the detection result from the input image. The typical one-stage methods mainly include YOLO [32], SSD [33], R-FCN [34], and RetinaNet [35]. They are mainly composed of a backbone network and a detection head. The backbone network is mainly used for feature extraction, which can obtain features of different scales and different levels of abstraction. Based on the extracted features and supervised information, the detection head learns target category and position regression. During training, through multitask loss joint training, the two branches of category prediction and position regression can be carried out at the same time. YOLO directly divides the picture into 7×7 grids as target candidate regions. Then the network directly predicts the target location and category. SSD combines the regression thinking of YOLO and the anchor mechanism of Faster R-CNN. Compared with two-stage methods, one-stage methods perform category prediction and position regression simultaneously, which have a faster detection speed. However, the network structure design makes it difficult for the one-stage methods to have higher detection accuracy over the two-stage methods.

Furthermore, according to whether the anchors are used, CNN-based target detection methods can be divided into anchor-based and anchor-free methods. Anchor-based methods are represented by Faster -RCNN and RetinaNet. In these algorithms, a large number of anchors are set in advance, and then some anchors are removed according to the actual targets. However, dense candidate regions consume a lot of computing resources, and there are only a small number of positive samples in the candidate regions, which brings about the problem of extremely imbalanced positive and negative samples. In addition, the use of anchors as candidate regions requires setting various parameters, and the detection performance is greatly affected by these parameters. Therefore, anchor-based methods have poor generalization ability. Differently, anchor-free methods such as CenterNet [36], FSAF [37], fully convolutional one-stage object detection (FCOS) [38], FoveaBox [39], and CornerNet [40] convert the target detection from a classification problem to a regression problem. These methods do not use anchors to predict the target but directly use the features extracted by the network to predict the target category and location. Therefore, anchor-free methods have simple network architectures and occupy a small number of computing resources, which are more suitable for practical scenarios. However, although these methods can alleviate the imbalance between positive and negative samples, they also have

the problem of semantic ambiguity caused by overlapping target borders.

At present, deep learning based methods for ship detection in SAR images are closely related to the development of deep learning and the enrichment of SAR image datasets. With the release of many public datasets of ship targets in SAR images and the development of target detection algorithms based on deep learning, more and more researchers focus on the end-to-end SAR ship detection methods based on deep learning [41]–[52]. Zhu *et al.* [41] proposed an end-to-end ship detection method based on YOLOv3. Chen *et al.* [42] added a deconvolution module and a prediction module on the basis of SSD. Ai *et al.* [43] applied the multiscale rotation-invariant Haar-like (MSRI-HL) feature to CNN. Deng *et al.* [44] designed a condensed backbone network, and adopted a feature reuse strategy. Han *et al.* [45] studied how the detection performance varies for images with different complexity, backgrounds, surroundings, and quality. Zhang *et al.* [46] proposed a novel balanced feature pyramid network (B-FPN) to improve detection accuracy. In order to reduce the interference of complex background and enhance the salient features of ship targets in SAR images, some studies adopt the attention mechanism in their networks. Cui *et al.* [47] applied a convolutional block attention module to the pyramid structure, which improves the performance for multiscale ship detection. Zhao *et al.* [48] proposed an attention receptive pyramid network (ARPN), which design a dilated attention block to enhance the relationships among nonlocal features and refine information at different feature maps. Chen *et al.* [49] designed a novel attention mechanism to help the detector focus more on the salient regions containing ships and combined it with knowledge distillation to improve the performance. Yang *et al.* [50] proposed R-RetinaNet to solve the problems such as mismatch of feature scale, contradictions between different learning tasks, and the unbalanced distribution of positive samples. In the past two years, due to the concise framework and the ability to make full use of information, the anchor-free methods have been widely applied to SAR ship detection [51], [52]. Cui *et al.* [51] added the spatial shuffle-group enhance attention module into CenterNet, which can suppress noise and reduce false positives (FP) caused by coastal and inland interferences. Fu *et al.* [52] proposed a feature-balanced and optimized anchor-free network to solve multiscale ship target detection problem. This network leverages the proposed attention module to balance the different levels of features extracted by the pyramid so that the network can obtain more small ship target information. At the same time, a feature refinement module is proposed to enhance semantic features. The CNN-based detection algorithms can automatically capture the multidimensional features of target, which are more suitable for SAR ship detection in complex scenes. However, the detection performance of these existing algorithms still needs to be improved. First of all, affected by the factors in HR SAR imaging, it is difficult for small ship targets to be fully detected. Meanwhile, there is substantial scattering interference in the land scene, and the public networks have a poor ability to detect near-shore ship targets.

Furthermore, the anchor-based methods have a complex network structure and take up more computing resources, which are not suitable for practical scenarios. Compared with anchor-based methods, the anchor-free methods can alleviate the imbalance of positive and negative samples to a certain extent. But there is also the problem of imbalance between positive and negative samples and sematic ambiguity caused by overlapping target borders. Some samples of ships in HR SAR image are shown in Fig. 2.

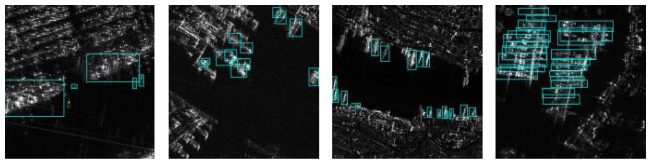

Fig. 2. Samples of ships in the HR SAR image. Rectangles with blue color represent ground truth targets. (a) Multiscale ships. (b) and (c) Inshore ships with interference. (d) Overlapping ships.

Aiming at the above problems, an improved anchor-free method is proposed for ship targets detection in HR SAR images called the category-position refinement convolutional one-stage object detection (CP-FCOS). The main contributions of our work can be summarized as follows:

1) To improve the detection performance for ship targets in complex scenes HR SAR images, the FCOS as the base network is applied to our model. This network can improve the detection efficiency through pixel-by-pixel prediction of the SAR images and avoid intensive anchors.
2) The category-position (CP) module is proposed to optimize the position regression branch features in the FCOS network. This module can improve target positioning performance in complex scenes by generating guidance vectors from classification branch features.
3) Considering the imbalance between positive and negative samples during network training and the semantic ambiguity caused by overlapping target borders, we redesign the target classification and boundary box regression methods to shield the adverse effects of fuzzy areas.
4) We conduct extensive experiments on SAR Ship Detection Dataset (SSDD), High-Resolution SAR Images Dataset (HRSID), IEEE 2020 Gaofen Challenge SAR dataset, and two complex large-scene HR SAR images to verify the proposed improvements. More importantly, our method was successfully applied to SAR ship detection in the 2020 Gaofen Challenge.

The rest of this article is organized as follows. Section II illustrates the proposed method in detail. Next, the experimental results on several dataset and the comparison with CNN-based methods are explained in Section III. Finally, Section IV concludes this article.

II. Proposed Method

In this section, the details of the proposed network structure, the key technical points of the proposed method, and each branch of the proposed method will be introduced. At first, the key points of the FCOS network are introduced, which are adopted as the

baseline in our work. Next, the overall scheme of the proposed network structure is described in detail. Then, the proposed CP module is illustrated. Finally, the target classification and regression methods are redesigned by analyzing the existing problems in anchor-free methods.

A. Fully Convolutional One-Stage Object Detection

The fully convolution one-stage object detection is a one-stage anchor-free target detection algorithm. The target Center-ness was proposed for the first time in this network, which is mainly used to remove low-quality detection areas far from the center of the target. And this network is a state-of-the-art model in anchor-free detection algorithms. At the same time, compared with the classic Faster R-CNN and other two-stage detection algorithms, it has higher detection accuracy and faster detection speed. Therefore, the FCOS network is applied to SAR ship detection in this article. The advantages and problems of FCOS are fully analyzed and studied in this article, and the model is further improved by analyzing the existing problems.

The framework of the FCOS algorithm is mainly composed of three parts: a feature extraction network, an FPN, and a three-branch detection head network. The various pretraining models such as ResNet-34, ResNet-50, ResNet-101, and ResNet-152 [53] can be used as the backbone of the feature extraction network to extract multilevel features of the input image. Then, the acquired high-level semantic elements and low-level spatial features are fused through FPN [54], generating multiscale fusion feature maps. The low-level spatial features contain more detailed information, which is suitable for accurately positioning the target. At the same time, high-level semantic features have more semantic information, which is eligible for classification. Finally, the target classification and position regression are performed on the fused feature maps through the detection head network.

In the following, we present the specific implementation steps of the FCOS detector.

1) Regression Method: Generally speaking, the anchor-based methods need to set various sizes and aspect ratios anchors on the feature map to predict the target position. However, a pixel-based detection strategy is applied in the FCOS network. In other words, each pixel on the feature map is used for regression. First, each pixel on the feature maps is mapped back to the original input image. Then, according to the position of the mapped pixel, it is determined where the original pixel is in and what the corresponding category label is. If both are satisfied, the pixel on the feature maps is a positive sample. Otherwise, it is a negative sample. At the same time, the target regression parameters, as shown in Fig. 3, are calculated in the network. In addition, if the mapped pixel is in multiple bounding boxes, it is regarded as a fuzzy sample. And the bounding box with the smallest area is regarded as the regression target of this pixel on the feature map. Notably, the regression is performed on the FPN multiscale feature maps, so the number of fuzzy samples mentioned above is small [38].

Compared with the anchor-based methods, which determine the positive and negative samples by calculating the intersection over union (IOU) of the preset anchors and the actual bounding boxes, the FCOS regression method mentioned above can obtain more positive samples. And the distribution of positive and negative samples in the network is more reasonable, which is conducive to the improving detection performance. However, all pixels in the bounding box will be treated as positive samples in the network so that some pixels in the background may be incorrectly labeled as positive samples. Therefore, this regression method directly applied to head network will interfere with the learning of positive sample features, which will affect the detection performance of ship targets.

Algorithm 1: FCOS for the SAR Image.

Input: The testing SAR image.
Step-1: Initialize the feature extraction network parameters with the pre-training model. Then, the input images are sent to the backbone network to extract the multi-scale features and obtain *C3*, *C4*, and *C5* feature maps.
Step-2: The multi-scale feature maps are fused to build a feature pyramid.
Step-3: The detection head network is applied to regresses the target category and position on the multi-scale features of the feature pyramid.
Step-4: Train the network to obtain the final model parameters.
Step-5: The trained model is used for the test data, and the target detection results are obtained on the multi-scale feature map of the feature pyramid.
Step-6: The Non-Maximum Suppression (NMS) [55] algorithm is executed to select the target and obtain the final detection result.
Output: Corresponding detection results.

Fig. 3. Regression method of FCOS. *l*, *t*, *r*, and *b* are the distances from pixel to the left, top, right, and bottom of the bounding box.

2) Center-Ness Strategy: The FCOS proposes the Center-ness strategy to suppress low-quality prediction bounding boxes in the detection results. And this strategy does not introduce any additional hyper-parameters. Specifically, the center-ness prediction branch is introduced into the detection head network, which can share parameters with the classification branch. The calculation method of Center-ness is shown in the following equation:

$$\text{Centerness} = \sqrt{\frac{\min(l,r)}{\max(l,r)} \times \frac{\min(t,b)}{\max(t,b)}}. \tag{1}$$

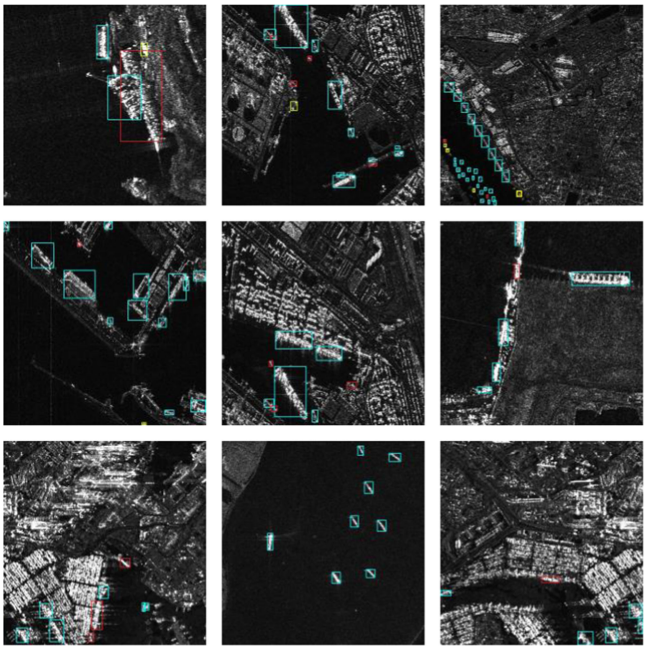

Fig. 4. Detection results of the FCOS network. Rectangles with blue color are true ships. Rectangles with yellow color are false alarms. Rectangles with red color are the missing ships.

The centrality ranges from 0 to 1, and the cross-entropy loss is used for training. When testing, the final score of the target detection result is multiplied by the predicted Center-ness and the prediction score of the classification branch. Therefore, the Center-ness can reduce the weight of the bounding boxes farther from the center of the target, which might be removed by the NMS algorithm.

3) Detection Performance Analysis: To show the advantages of the FCOS network clearly, the detection results for SAR ship targets based on this network are shown in Fig. 4. Specially, we first train the detector from the HRSID [57] dataset and then leverage the backbone, ResNet-50 [53], to extract features. The IOU threshold in experiments is set as 0.5. It can be observed that FCOS can accurately detect ship targets in the sea area. Nevertheless, there is a severe problem of the missing ships in the sea-land area, especially for small-scale ships. Furthermore, ships docked in parallel are more likely to be missed or incorrectly located. And the strong scattering objects in the land area are mistakenly detected as ship targets. Based on the above analysis, we can draw the following conclusions: The FCOS can be applied as the base network to enhance ship detection performance in SAR images. However, the feature extraction capability of the FCOS network needs to be further improved to avoid false alarms and missing detections. Therefore, it is necessary to enhance the FCOS network feature extraction capability and improve the network detection performance.

B. Overall Scheme of the Proposed Network Structure

Aiming at the problems of the FCOS network, an improved anchor-free method based on FCOS is proposed for HR SAR ship targets detection. The overall scheme of the proposed network structure is illustrated in Fig. 5, which can be divided into three parts: the feature extraction network, the FPN, and the detection head network.

1) Feature Extraction Network: Generally speaking, with the increasement of the network layers, richer semantic information can be obtained to improve the network's performance. However, simply increasing the network depth can easily cause gradient dispersion or gradient explosion problems. Therefore, the ResNet [53] is proposed to cleverly avoid the shortcomings of simply increasing the depth of the network. Its internal convolution residual units adopt cross-layer connections, which can effectively alleviate the problem of gradient disappearance caused by network deepening, and extract deeper feature information at the same time. To choose the best backbone, we leveraged several networks to test the proposed method. Through our experimental verification (see Section III-A for more detailed discussion), ResNet-50 is the most suitable backbone network to extract HR SAR image features. Compared with ResNet-101 and ResNet-152, ResNet-50 has a shallower network layer, which can avoid the overfitting problem. In addition, we found that SAR images are different from optical images in that their backgrounds are more straightforward, and their textures are single. Therefore, there is no need for a deeper network to extract too high-level semantic information. Then, as for ResNet-18 and ResNet-34, their feature extraction capabilities are insufficient. And the comparison between these networks shows that using ResNet-50 as the backbone can provide a stable performance improvement.

2) Feature Pyramid Network: As we all know, the convolution operation of the feature extraction network can obtain rich high-level semantic information, but the spatial position information will inevitably be reduced. In order to achieve accurate classification and regression of target, the feature map pyramid is built to construct multiscale feature maps and predict multiscale ship targets. FPN is specifically composed of three parts: bottom-up, top-down, and horizontal connections. Bottom-up connection is the process for the feature extraction network to extract image features. Top-down connection is an up-sampling process. The horizontal connection is added to combine the up-sampling results with the feature maps generated by the CNN. Feature maps $C2$, $C3$, $C4$, and $C5$ are first extracted by the ResNet-50. The $P5$ feature map is acquired after the convolution with a kernel size of 3×3 and stride of 1 is conducted on $C5$. Similarly, the convolution with a kernel size of 1×1 and the stride of 1 are conducted on $C4$. After that, the added $P5$ up-sampling results are sent to the convolution with a kernel size of 3×3 and the stride of 1 to produce $P4$ feature map. According to the above operation, the $P3$ and $P2$ feature maps are obtained. Specially, the $P6$ feature map is acquired after the convolution with a kernel size of 3×3 and stride of 1 is conducted on $P5$. Each feature map ($P2$ to $P5$) channel C in FPN is 256, independently sent to the subsequent detection head to complete classification and regression.

3) Detection Head Network: The structure of the detection head network is shown in Fig. 6, which contains three prediction branches and a CP attention feature optimization module. First, the two prediction branches are designed for target classification

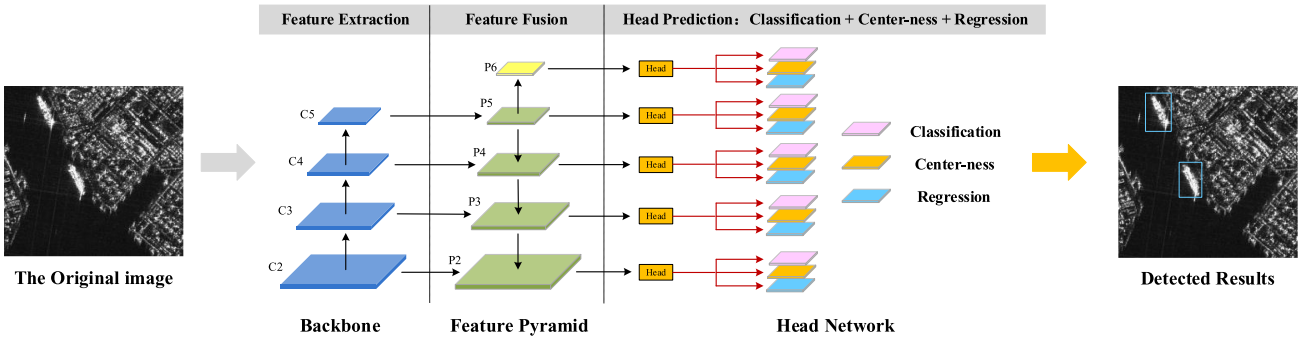

Fig. 5. Architecture of the proposed model CP-FCOS.

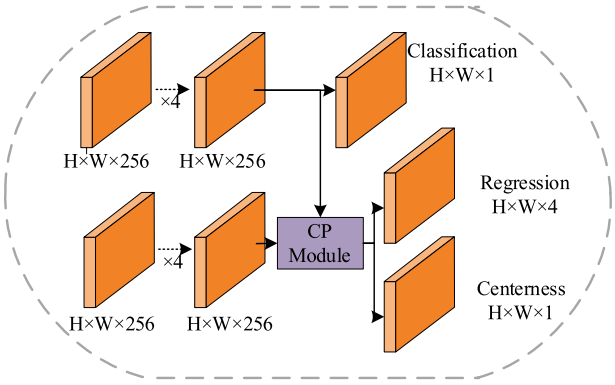

Fig. 6. Structure of detection head network.

and location regression, respectively. Then, the CP module is proposed to combine two prediction branches features, generating a guidance vector to optimize the regression feature. Meanwhile, the target classification network directly connected to the feature pyramid is a fully connected network (FCN). Each fused feature map P_i is subjected to a series of convolution processes; these processes can be summarized as

$$L_{i1} = \mathrm{ReLU}(\mathrm{Conv}_{3\times3}(P_i))\ i = 2, 3, 4, 5, 6 \tag{2}$$

$$L_{ij+1} = \mathrm{ReLU}(\mathrm{Conv}_{3\times3}(L_{ij}))\ j = 1, 2, 3 \tag{3}$$

where P_i represents the FPN output features, L_{ij} represents convolution processing features, $\mathrm{Conv}_{3\times3}$ denotes 3×3 convolutional layer, and the channels of these processes are 256. The final classification feature is produced as shown in the following equation:

$$F_c = \mathrm{Sigmoid}(\mathrm{Conv}_{3\times3}(L_{i4})) \tag{4}$$

where F_c represents the final output of the classification branch, which is the $H \times W$ heat-map of target confidence. Another small FCN paralleled to the classification network is added for bounding box regression. Differently, the size of the output feature of the regression branch is $H \times W \times 4$. The third branch is the Center-ness branch, which represents the distance offset between the pixel (i, j) with the center pixel. When testing, the final score is multiplied by the Center-ness score and the classification score.

Algorithm 2: Proposed CP-FCOS Detector.

Input: The testing SAR image.
Step-1: Initialize the feature extraction network parameters with the pre-training model. Then, the input images are sent to the ResNet-50 to obtain *C2*, *C3*, *C4*, and *C5* feature maps.
Step-2: A feature pyramid is adopted in FPN. The fused features {*P2, P3, P4, P5, P6*} are produced for the detection head.
Step-3: Detection head network is applied to target classification and regression on the multi-scale features of the feature pyramid. Especially, the category-position (CP) module is added to the regression branch to optimize features.
Step-4: Train the network with the Center-ness loss function to obtain the improved model parameters.
Step-5: Testing image with the trained model, and the bounding box score of the target detection result is multiplied by the Center-ness score and the prediction score of classification branch.
Step-6: The Non-Maximum Suppression (NMS) algorithm is executed to select the target and obtain the final detection result.
Output: Corresponding detection results.

The specific implementation steps of the proposed detector are summarized in Algorithm 2.

C. CP Module

As analyzed in Section II-A, the FCOS network can achieve better detection performance through pixel-by-pixel prediction of the SAR images. However, there is also an imbalance between positive and negative samples and semantic ambiguity caused by overlapping target borders. Therefore, the CP module (see Section II-C for more detailed discussion) is proposed to optimize the position regression branch features in the FCOS network. This module can improve the detector performance in complex scenes by generating guidance vector from the classification

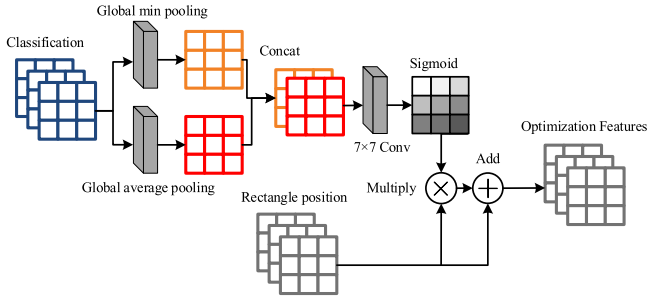

Fig. 7. CP module.

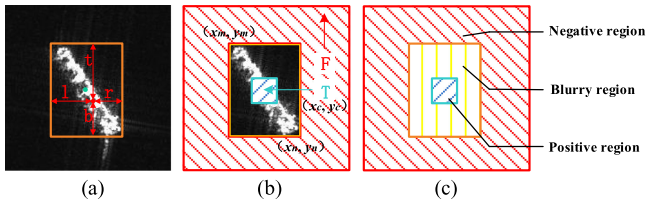

Fig. 8. Regression target and image region division. (a) Rectangle position. (b) Classification method. (c) Image region.

branch features. Fig. 7 shows its architecture in detail. The CP module first generates the features $F_{\min}$ and F_{avg} by using the global min pooling and the global average pooling along the channel axis, respectively. And the features are aggregated by a concatenation operation. Then, one 7×7 convolutional layer and the sigmoid function are applied to the aggregated feature maps. After multiplied with regression branch feature, the pixel addition operation is performed to obtain the final refinement feature F_{cp}. This process is computed as follows:

$$F_{\min} = \text{MinPool}(F_{cla}) \tag{5}$$

$$F_{\text{avg}} = \text{AvgPool}(F_{cla}) \tag{6}$$

$$F_{\text{con}} = \text{Sig}(\text{Conv}_{3\times3}(\text{Concat}(F_{\min}, F_{\text{avg}}))) \tag{7}$$

where MinPool and AvgPool represent the min pooling and average pooling, Concat is the concatenation operation, and Sig is the sigmoid function. Finally, the output feature maps F_{con} are combined with the position feature maps F_{pos} as follows:

$$F_{\text{cp}} = (F_{\text{pos}}(\oplus (F_{\text{con}} \otimes F_{\text{pos}})) \tag{8}$$

where $\oplus$ represents the elementwise addition, and $\otimes$ represents the elementwise multiplication. The output features are used for the target regression and the calculation of Center-ness. Through the optimization with CP module, it can be found that the regression features of the network will contain more position information. In addition, the feature optimization process is beneficial to improve the accuracy of Center-ness, which can improve the target detection performance.

D. Classification and Regression Redesign

In the complex sea-land scenarios, ship targets often dock side by side, which brings huge challenges to deep-learning target detection. At the same time, there is the problem of semantic ambiguity caused by overlapping target borders in the network. Therefore, the target classification and boundary box regression methods are redesigned in our model to shield the adverse effects of fuzzy areas in the FCOS training. Specifically, the regression range of the bounding boxes in multiscale FPN feature maps is limited in our model. First, the regression value (l, t, r, b) of ship targets on each feature map is calculated. Then, if the pixel position meets the conditions as follows:

$$\max(l, t, r, b) > M_i \text{ or } \min(l, t, r, b) < M_{i-1} \tag{9}$$

then, it will be set as a negative example, and the category and position regression will not be performed on the feature map. Here, M_i is the maximum regression distance of the ith feature map. Specially, $[M_1, M_2, M_3, M_4, M_5, M_6]$ are preset to [0, 32, 64, 128, 256, $+Inf$] in the proposed model.

Generally speaking, the large aspect ratio is one of the most significant characteristics of a ship target, and the ship's bow direction is randomly distributed at 360 angles. When using rectangular to mark a target, the ship target pixel occupies a small area in the image, but the pixel in the background occupies most of the area. Some background pixels in the bounding box may be incorrectly labeled by the FCOS network regression method. Therefore, all pixels in the bounding box area cannot simply be regarded as targets. According to the characteristic of the SAR ship target, the target classification and bounding box regression method are redesigned in this article. And the image region is divided into three parts: positive region, negative region, and blurry region. The regression target and image region division are shown in Fig. 8.

In the FCOS network, for each location (x, y) on the feature map F_i, it will be mapped back onto the input image as ($xs + s/2$, $ys + s/2$), where s represents the total stride. If the location (x, y) falls into any ground-truth box, it will be regarded as a positive sample and the corresponding class label $c^* = 1$ (c^* represents the class label of the ground-truth box). Otherwise, it will be considered as a negative sample and $c^* = 0$ (background class). However, we redesign the positive sample region of the feature map. We expand it a certain distance from the center point of ship target, and the size of the area does not exceed the outer contour of target. We define the center pixel point (x_c, y_c), (x_1, y_1) is the corresponding point on the input image of the location (x, y). In our model, the width of the positive sample region is 0.5 times the target width in the feature map, as shown in Fig. 8. Therefore, if location (x_1, y_1) is inside a ground-truth box $T = [(x_m + x_c)/2, (y_m + y_c)/2, (x_c + x_n)/2, (y_c + y_n)/2]$, it will be considered as a positive sample, and the class label $c^* = 1$, where (x_m, y_m) and (x_n, y_n) are the top-left and bottom-right coordinates of the ground-truth box. If the location is outside ground-truth box $B = (x_m, y_m, x_n, y_n)$, it will be considered as a negative sample, and the class label $c^* = 0$. Otherwise, it is a fuzzy sample but $c^* = 1$.

Formally, during the network training stage, the label for classification in positive sample area and negative sample area are regressed by c^*, but the c^* loss of the fuzzy sample area is not calculated. The 4-D real vector $P^* = (l^*, t^*, r^*, b^*)$ are the regression value for target location in the positive region

and blurry region, but the position loss of negative region is not calculated. And the training regression vector $P^* = (l^*, t^*, r^*, b^*)$ can be formulated as

$$\begin{aligned} l^* &= x_1 - x_m \\ t^* &= y_1 - y_m \\ r^* &= x_n - x_1 \\ b^* &= y_n - y_1. \end{aligned} \tag{10}$$

It can be found that the proposed method can leverage as many target samples as possible to train classification regressor, which can reduce the effect of background pixels. Besides, we choose enough foreground samples (include background pixels in ground-truth box) to train the position regressor. Therefore, the accuracy of the location and category regression value (P^* and c^*) can be improved in the proposed network. Meanwhile, after such region division and loss calculation, the influence of overlapping bounding boxes can be reduced. Detailed analysis of improvement is explained in Section III.

E. Loss Function

The training loss function of the proposed method mainly consists of three branch loss functions

$$L_{\mathrm{CP-FCOS}} = \frac{1}{N_{\mathrm{pos}}} L_{\mathrm{cls}} + \frac{m}{N_{\mathrm{pos}}} L_{\mathrm{reg}} + \frac{n}{N_{\mathrm{pos}}} L_{\mathrm{cen}} \tag{11}$$

where L_{cls}, L_{reg}, and L_{cen} denote the classification loss, regression loss, and center-ness loss. N_{pos} is the number of positive sample pixels in the standard result. m, n, and 1 are the loss balance weight, which can adjust three branch loss proportions.

The Focal Loss [56] is adopted for L_{cls}, and it can be formulated as

$$\begin{aligned} L_{cls} = &\sum_{i,j \in R^p} -\alpha_t (1 - \overset{i,j}{\underset{cls}{\Pr}})^{\gamma} \log(\overset{i,j}{\underset{cls}{\Pr}}) \\ &+ \sum_{i,j \in R^n} -(1-\alpha_t) \Pr{}_{cls}^{i,j\,\gamma} \log(1 - \overset{i,j}{\underset{cls}{\Pr}}) \end{aligned} \tag{12}$$

where R^p and R^n denote the positive sample region and negative sample region. $\alpha_t$and γrefer to modulation factor. $\Pr_{cls}^{i,j}$ represents the predicted category of (i, j). We use the IOU LOSS to calculate the regression loss, which is defined as

$$L_{\mathrm{reg}} = \sum_{i,j \in (R^p \cup R^i)} (1 - \mathrm{IOU}(\Pr{}_{bbox}^{i,j}, Gt_{bbox}^{i,j})) \tag{13}$$

where $Gt_{bbox}^{i,j}$ represents the ground truth localization of the pixel (i, j), $\Pr_{bbox}^{i,j}$ represents the predicted result of (i, j), and R^i denotes the invalid sample area. The Center-ness loss is calculated with binary cross entropy (BCE) [38] loss

$$L_{\mathrm{cen}} = \sum_{i,j \in (R^p \cup R^i)} \mathrm{BCE}(\overset{i,j}{\underset{\mathrm{cen}}{\Pr}}, Gt_{\mathrm{cen}}^{i,j}) \tag{14}$$

where $\Pr_{\mathrm{cen}}^{i,j}$ represents predicted center-ness of the pixel (i, j), $Gt_{\mathrm{cen}}^{i,j}$ represents the truth center-ness of the pixel (i, j), and Ri denotes the invalid sample region.

III. Experiments

In this section, to evaluate the effectiveness of the proposed method, extensive experiments are conducted on HRSID [57], SSDD [58], IEEE 2020 Gaofen Challenge SAR dataset [59], and two complex large-scene HR SAR images. First, experimental implementations such as datasets, settings, and evaluation metrics will be introduced. Then, the ablation and comparison experiments will be performed to prove the effectiveness of the proposed method. Finally, the results in complex and large-scene images imply the generalization of the proposed method.

A. Implementations

1) Datasets: The HRSID proposed by Wei *et al.* [57] is established by using images from 99 Sentinel-1B imageries, 36 TerraSAR-X, and 1 TanDEM-X imagery. These large-scene imageries are cropped to 800 × 800 pixels SAR images. Therefore, it contains 5604 cropped SAR images and 16951 ships. And resolutions of these images range from 1 to 15 m. The ablation experiments are implemented on HRSID to evaluate the effectiveness of the proposed improvements. SSDD is a multiresolution, multisize, and multisensor SAR ship dataset published by Li *et al.* [58], a benchmark dataset for scholars to evaluate their algorithms. The SSDD dataset contains 1160 images and has 2456 ships with sizes ranging from the most miniature scale of 7 × 7 to the largest scale of 211 × 298. The polarization modes of these images include HH, HV, VV, and VH, resolutions also range from 1 to 15 m. IEEE 2020 Gaofen Challenge SAR dataset is provided to participate in the 2020 Gaofen Challenge on Automated High-Resolution Earth Observation Image Interpretation. It contains 1000 SAR images with spatial resolution ranging from 1 to 5 m, collected from the Chinese Gaofen-3 satellite. All the images are 1000 × 1000 pixels value, which contain multiorientations, and multiscales ships. For ship target detection in complex and large-scene HR SAR images, some typical slices from Gaofen-3 satellite are applied to our experiments. Gaofen-3 satellite has different imaging modes, such as Strip-Map (UFS), Fine Strip-Map 1 (FSI), Full Polarization 1 (QPSI), Full Polarization 2(QPSII), and so on. The original pixel values of the two complex large-scale SAR images are shown in this article are 16746 × 24919 and 12678 × 11328. They are cropped into 800 × 800 image blocks under the overlapped ratio of 25%, and these cropped images are divided into, i.e., a training set and a testing set, with the proportion of 7:3. The detailed information of these datasets is presented in Table I. Some instances of different datasets are shown in Fig. 9.

2) Settings: All the experiments are implemented in PyTorch 1.3.1, CUDA 9.2, and CUDNN 7 with Intel Core i7-8700K CPU and a NVIDIA Geforce GTX 1080Ti GPU. The PC operation system is Windows 7. The computer and deep-learning environment configuration for our experiments is presented in Table II.

Furthermore, all the models are trained with the stochastic gradient descent algorithm for 60 000 iterations with a total of two images per minibatch. The initial learning rate is set as 0.0004, and the weight decay is 0.00002. The detection threshold IOU in all experiments is set 0.5.

TABLE I
DETAILED INFORMATION OF DIFFERENT DATASET

Dataset	Size (pixel)	Image (num)	Ship (num)	Mode	Satellite	Resolution(m)
HRSID [59]	800 × 800	5604	16951	S3-SM /ST/HS	Sentinel-1B/ TerraSAR-X/ TanDEM-X	1-15
SSDD [60]	390 × 205 - 600 × 500	1160	2456	SLC /ST/SM	RadarSat-2/TerraSAR-X/Sentinel-1	1-15
IEEE 2020 Gaofen Challenge [61]	1000 × 1000	300	1550	SL/UFS	GF-3	1-3
Large-scene Images	16746 × 24919 12678 × 11328	2	979	UFS/FSII	GF-3	3-10

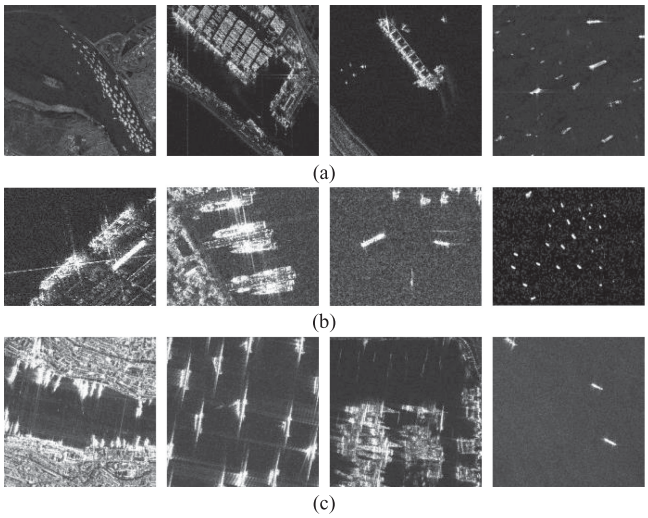

Fig. 9. Instances of different dataset. (a) HRSID. (b) SSDD. (c) IEEE 2020 Gaofen Challenge. All of them show the specialty of multiresolution, multisize ships, and substantial interference in HR SAR images.

TABLE II
ENVIRONMENT CONFIGURATION

Project	Model/Parameter
CPU	Intel i7-8700K
RAM	32GB
GPU	NVIDIA GTX Geforce 1080 Ti
System	windows 7
Code	python3.7
Framework	CUDA9.2/cudnn7/torch 1.3.1

Generally speaking, the backbone used in the proposed network will directly affect the detection performance. To choose the best backbone for our model, we conduct some experiments with different backbones based on the FCOS network; the experimental results are presented in Table III and Fig. 10. From the experimental results, we found that the detection performance of ResNet-50 is better than the other four backbones. First, the training and testing efficiency of ResNet-50 is significantly faster than ResNet-101 and ResNet-152. Second, the mAP value of network with ResNet-50 is completely better than that with ResNet-18 and ResNet-34. Furthermore, as presented in Table III, the mAP value of the network with ResNet-50 is 93.97%, which is 11.38, 6.99, 1.89, and 2.16% higher than that with ResNet-18, ResNet-34, ResNet-101, and ResNet-152.

TABLE III
DETECTION PERFORMANCE OF DIFFERENT BACKBONES

Backbone	Training Time(ms)/Iter	Testing Time(ms)/Image	mAP(%)
ResNet-18+FPN	52.59	47.94	82.59
ResNet-34+FPN	58.24	50.58	86.98
ResNet-50+FPN	62.57	55.38	93.97
ResNet-101+FPN	73.59	68.89	92.08
ResNet-152+FPN	97.75	87.12	91.81

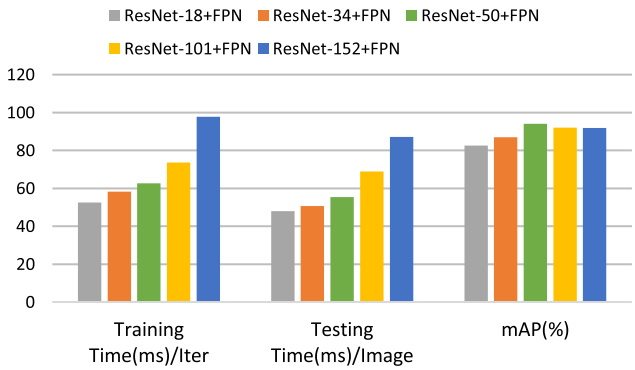

Fig. 10. Detection performance of different backbones.

Therefore, the pretrained model ResNet-50 on ImageNet is used as the backbone in the rest experiments.

3) Evaluation Metric: In our experiments, AP, AP_s, AP_m, AP_l, and $F1$ scores are utilized to evaluate the detection performance of the network, which are the same as the metric definition on the COCO dataset. The PR curve is plotted by calculating the precision–recall at different confidence levels in all experiments. Average precision (AP) is the area enclosed by the relationship curve of precision and recall. The average of multiple categories of AP is mean average precision (mAP). Especially for the ship target, mAP is AP. According to the number of pixels contained in the bounding box of the ship target, the ship is divided into small targets (0–1000 pixels), medium targets (1000–4000 pixels), and large targets (over 4000 pixels). Therefore, AP_s, AP_m, and AP_l represent AP for small ships, medium ships, and large ships. The definition of precision and recall is defined as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{15}$$

TABLE IV
EFFECT OF CP-MODULE WITH DIFFERENT POOLING METHODS

CP	Pooling	mAP	AP_s	AP_m	AP_l	F1
-	-	0.9397	0.9371	0.9887	0.9486	0.9252
√	Avg+Max	**0.9524**	**0.9501**	0.9885	**0.9818**	**0.9393**
√	Avg+Min	0.9481	0.9473	**0.9892**	0.9531	0.9294

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{16}$$

where TP (true positives), FP, and FN (false negative) refer to the number of correctly detected ships, false alarms, and missing ships. $F1$ score combines the precision and recall as follows:

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \tag{17}$$

AP is defined as

$$\text{AP} = \int_0^1 P(R)dR \tag{18}$$

where P represents precision and R represents recall. Actually, the AP metric is applied to evaluate the comprehensive detection performance of the model.

B. Model Analysis

1) Effect of CP-Module: The experiments in this section are used to investigate the impact of the proposed CP feature optimization module on the model detection performance. Significantly, the Center-sample [37] is applied to train the network in these experiments. The experimental results based on the test dataset are presented in Table IV. It turns out that the CP module can improve the detection performance obviously, especially the AP_l index. Compared with the baseline, the CP module with average-pooling and the max-pooling finally increases the AP_s and AP_l by 1.3 and 3.32%, respectively. Furthermore, the mAP and $F1$ scores are increased by 1.33 and 1.41%, implying that the proposed improved model can obtain better comprehensive detection performance. And the CP-module with average-pooling and the min-pooling can increases the value of AP_m. In other words, the AP values of large, medium, and small ships in the SAR image can be improved with the proposed module. This is because the CP-module optimizes the position regression branch features and obtains the balanced features for multiscale ships. The effectiveness of the CP module can also be verified in Fig. 11. We can observe that more false alarms exist in the baseline detector and more small-scale ships are missed. In contrast, our method with CP module shows a better performance in dealing with such complex surroundings and multiscale ships. Meanwhile, to give a more intuitive explanation, basic and fine-grained feature maps of different heads, e.g., classification, center-ness, and regression, tested on offshore and inshore ships are shown in Figs. 12 and 13, respectively.

Fig. 12 shows some visualization results of inshore ships (including detection results with and without CP module, classification, center-ness, and regression feature maps). In the detection results (the first left column of Fig. 12), rectangles with

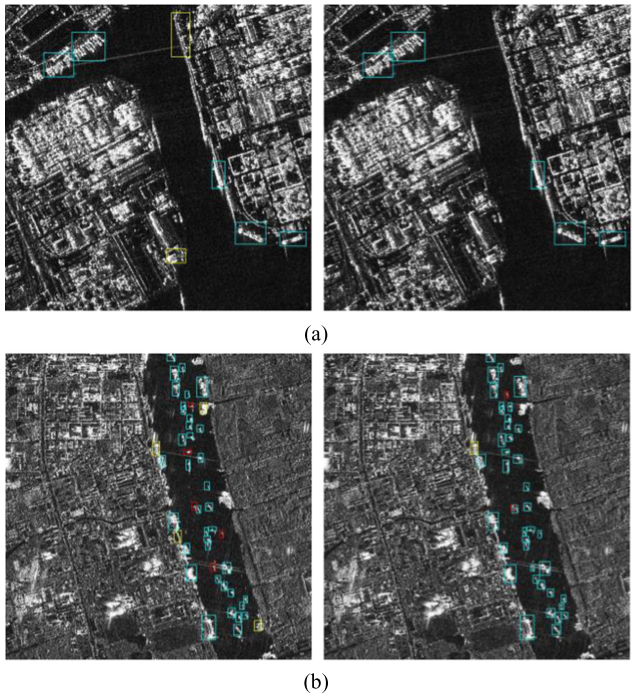

Fig. 11. Comparison results of the methods without and with CP module. Rectangles with blue color are true ships. Rectangles with yellow color are false alarms. Rectangles with red color are missing ships. (a) Results of the baseline. (b) Results of the baseline with CP module.

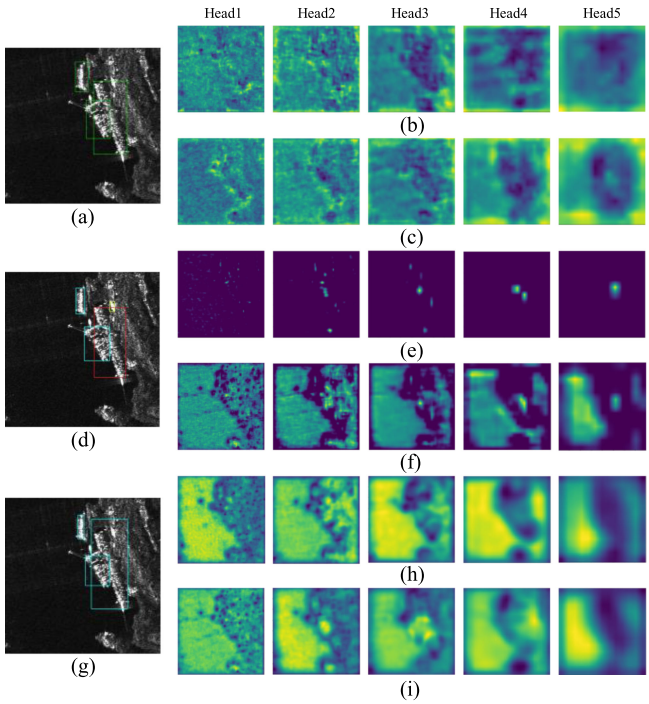

Fig. 12. Some visualization results of inshore ships. (a) Ground truths. (d) Detection results without CP module. (g) Detection results with CP module. (b) and (c) Classification feature maps. (e) and (f) Center-ness feature maps. (h) and (i) Regression feature maps. Rectangles with blue color are true ships. Rectangles with yellow color are false alarms. Rectangles with red color are missing ships.

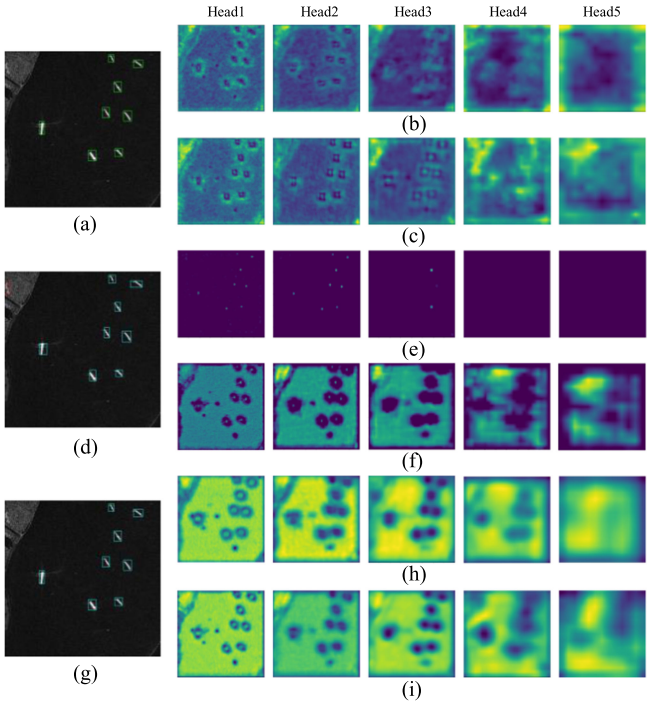

Fig. 13. Some visualization results of offshore ships. The meanings of the box color are the same as Fig. 12. (a) Ground truths. (d) Detection results without CP module. (g) Detection results with CP module. (b) and (c) Classification feature maps. (e) and (f) Center-ness feature maps. (h) and (i) Regression feature maps.

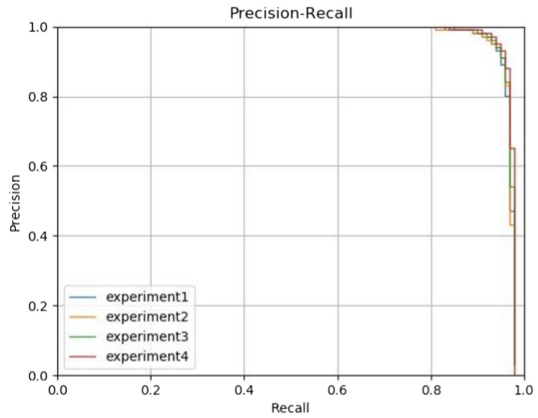

Fig. 14. PR curves of different improvement experiments.

blue color are true ships, rectangles with yellow and red color refer to false alarms and missing ships, and the green rectangles represent the ground truth (the following results are drawn like this). As shown in Fig. 12, the method with CP module performs better than the method without this module. Our method can detect parallel ships in such complex background and the locations of ships are predicted more accurately. According to the feature maps (the five columns on the right of Fig. 12), active areas in the classification, center-ness, and regression feature maps constructed by our method distribute more clearly than the baseline. Especially, the Center-ness and classification feature maps in higher head show stronger position information, which

TABLE V
EFFECT OF DIFFERENT REGRESSION METHODS

Method	mAP	AP_s	AP_m	AP_l	F1
FCOS [40]	0.9368	0.9306	0.9875	0.9451	0.9247
Center-sample[62]	0.9397	0.9371	0.9887	0.9486	0.9252
Ours	**0.9505**	**0.9470**	**0.9881**	**0.9582**	**0.9345**

is conducive to positioning these parallel ships. However, active areas in these feature maps predicted without CP module are blurry, which might lead to the missing ships or wrong detection.

Fig. 13 shows some visualization results of offshore ships (including detection results with and without CP module, and classification, center-ness, and regression feature maps). As shown in Fig. 13(e) and (f), the CP module significantly improves the Center-ness feature of Head3, Head4, and Head5 detectors. Meanwhile, the comparison between Fig. 13(h) and (i) shows that our method extracts more sufficient high-level semantic information in regression branch, proving that the proposed CP module plays an important role in enhancing the semantic information of multiscale ships.

2) Effect of Classification and Regression Redesign: In this section, the regression method proposed in this article is verified. The same parameter settings are used for training the network in these experiments. The results in Table V show that the redesigned classification and regression method yields better overall performance. According to the statistical results, it is shown that the FCOS network regression method achieves the worst detection performance. This is because the FCOS network treats all pixels in the bounding box as positive samples, which affects the model to learn target feature with incorrectly labeled background pixels. Furthermore, fixed sampling ratio is applied in the Center-sample regression to suppress the influence of semantic ambiguity. However, it only brings a slight improvement on mAP and $F1$. In comparison, the mAP value of ship target is obviously improved in our method, especially for small ships and large ships. This is because the regression method proposed in this article can effectively avoid the interference in the background pixels in bounding box on the network training, which is beneficial to the network detection performance.

3) Effect of Overall Improvement: Although the CP-module and redesigned regression method make the model better adapt to ships detection, the effect of overall improvement deserves further discussion. Therefore, we conducted some experiments to verify the impact of overall improvement on the detection performance. Experiment 1 is the baseline experiment, and the two improvement measures mentioned are not adopted. The detection results of these experiments are listed in Table VI. Fig. 14 shows the PR curves of different improvements. It can be seen from Table VI that our method brings a noticeable improvement in all evaluation metrics. Specifically, compared with the baseline experiment, the model with CP-module and redesigned classification and regression method increases the mAP and $F1$ by 2.04 and 1.98%. In addition, the much higher AP_s, AP_m, and AP_l indicate that our method can achieve more accurate localization for small, medium, and large ships, especially for small ship targets. And it can be seen from Fig. 10 that the model

TABLE VI
INFLUENCE OF EACH COMPONENT IN THE PROPOSED CP-FCOS

Experiment	CP	Regression	mAP	AP_s	AP_m	AP_l	F1
1	-	-	0.9397	0.9371	0.9887	0.9486	0.9252
2	√	-	0.9524	0.9501	0.9885	0.9818	0.9393
3	-	√	0.9505	0.9470	0.9881	0.9582	0.9345
4	√	√	**0.9601**	**0.9538**	**0.9899**	**0.9896**	**0.9450**

TABLE VII
DETECTION RESULTS OF CNN-BASED METHODS ON HRSID

Method	mAP	AP_s	AP_m	AP_l	F1	Inference Time(ms)	Parameter Size(M)	Model Volume(M)
Faster-RCNN [30]	0.9441	0.9358	**0.9959**	0.9739	0.9388	130	41.4	319
RetinaNet [37]	0.8920	0.8767	0.9885	0.9466	0.8699	122	36.3	257
FCOS [40]	0.9368	0.9306	0.9875	0.9451	0.9247	**55**	**32.1**	**228**
Ours	**0.9601**	**0.9538**	0.9899	**0.9896**	**0.9450**	64	33.5	240

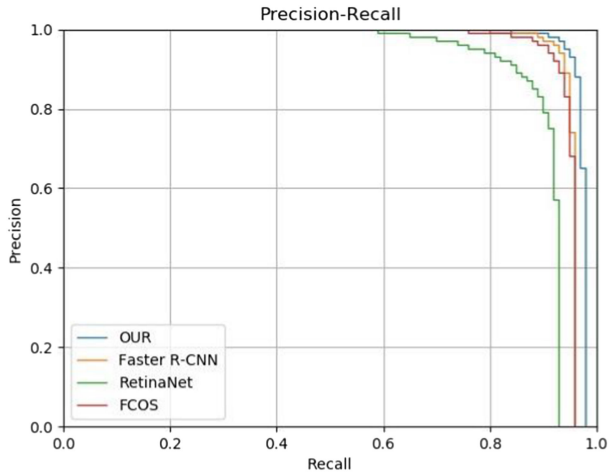

Fig. 15. PR curves of different methods on HRSID.

with overall improvement improves both the precision and recall rate considerably. This may be because the overall improvement optimizes the position regression branch features and avoids the semantic ambiguity caused by overlapping target borders.

C. Comparison With Other Methods on Different Dataset

To verify the performance of our method, we compare the proposed network with three CNN-based methods, including Faster R-CNN, RetinaNet, and FCOS, on four different datasets. Table VII presents the speed, accuracy, and computational complexity of different CNN-based methods tested on HRSID, and the corresponding PR curves are displayed in Fig. 15. As presented in Table VII, compared with the baseline FCOS network, the proposed structure can improve the detection performance in different degrees with only a slight increase in the model parameters and the inference time per image. However, more prominently, with the minimum model volume and minimum number of parameters, the detection speed of the proposed method is significantly faster than all the others. This may be because in Faster RCNN [28], anchor boxes are proposed, avoiding repeated feature computation. However, anchor boxes can result in excessively many hyper-parameters, which typically increases the adjustment time of the model. As for RetinaNet, the network leverages a confidence threshold of 0.05 to filter the first 1000 candidate boxes in each pyramid layer. After that, all candidate boxes of different layers are chosen by NMS. A large number of useless anchors cause waste of computing and storage resources, at the same time, increasing the computational complexity of the model. By contrast, the inference of the proposed method is more straightforward than Faster-RCNN and RetinaNet, it can directly obtain the classification scores and location vector on the feature maps, and then the predicted bounding boxes are chosen by threshold without any anchor calculation. Besides, it can be seen from Table VII that the mAP value and *F*1 scores of the model proposed in this article are 96.01 and 94.50%, respectively, which are better than anchor-based Faster R-CNN, RetinaNet, and anchor-free FCOS methods. In terms of AP_s, the score of our method is 1.8, 7.71, and 2.32% higher than Faster-RCNN, RetinaNet, and FCOS. AP_m of our method is 98.99%, which is close to other methods, whereas 0.6% slightly lower AP_m than Faster-RCNN. Furthermore, AP_l of our method is 94.50%, which is 0.62, 7.51, and 2.03% higher than Faster-RCNN, RetinaNet, and FCOS. This means improvements proposed in this article have a significant impact on the detection of multiscale ships. Among them, the mAP value of RetinaNet is less than 0.9, which is the worst detection result. In addition, it can be seen from Fig. 15 that the precision of our method (the blue curve in Fig. 15) under different recall rates is better than other three CNN-based methods. It is distinct that the PR curve of RetinaNet (the green curve in Fig. 15) is lower than those of the other methods. And the PR curves of Faster-RCNN (the yellow curve in Fig. 15) and FCOS (the red curve in Fig. 15) decrease sharply when the recall rate is higher than 0.8. However, the PR curve of our method is generally stable. Significantly, the precision rate of our model is always higher than those of Faster-RCNN, RetinaNet, and FCOS, especially when the recall rate is higher than 0.8. The comparison of the detection results of different methods can be seen from Fig. 16. The meanings of the boxes color are the same as Fig. 12. It can be observed that the proposed

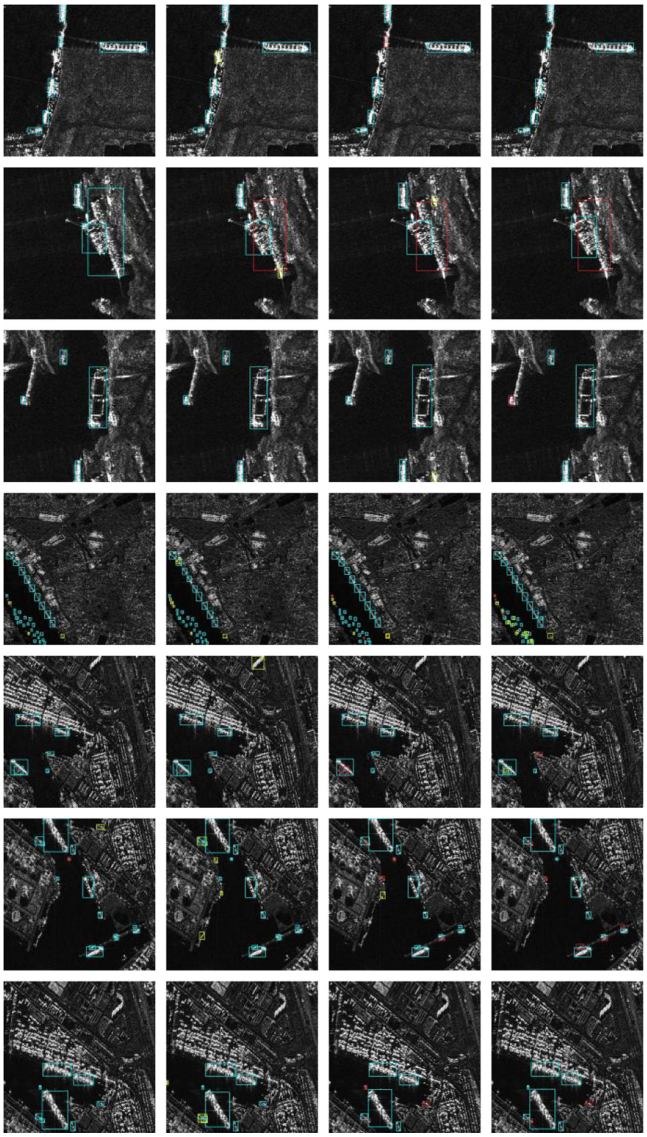

Fig. 16. Detection results of different methods on HRSID. The meanings of the boxes color are the same as Fig. 12. (a) Our methods. (b) Faster-RCNN. (c) FCOS. (d) RetinaNet.

method obtains a considerable performance on multiscale ships in complex scenes. And the proposed model is not susceptible to the interference of near-shore objects and could effectively reduce false alarms around the land. At the same time, compared with Faster-RCNN, RetinaNet, and FCOS, our model has a better detection performance for dense-docking ships. These benefit from the feature map enhanced by the CP-module, and the redesigned classification and regression method, which can effectively extract the shallow position features and high-level semantic features of multiscale ship target.

To verify the robustness and generalization ability of improvement, we perform ship detection in the SSDD dataset and apply the model to the 2020 IEEE Gaofen Challenge. Figs. 17 and 18 present the detection results in the two datasets. It can be observed that Faster-RCNN, RetinaNet, and FCOS miss some

TABLE VIII
DETECTION RESULTS OF CNN-BASED METHODS ON IMAGE 1

Method	TP	FP	FN	F1	Time(s)
Faster-RCNN [30]	54	11	7	0.8571	71.8
RetinaNet [37]	47	5	14	0.8318	75.2
Ours	55	5	6	**0.9089**	**64.6**

TABLE IX
DETECTION RESULTS OF CNN-BASED METHODS ON IMAGE 2

Method	TP	FP	FN	F1	Time(s)
Faster-RCNN [30]	12	4	6	0.7059	26.4
RetinaNet [37]	13	2	5	0.7879	27.6
Ours	15	2	5	**0.8571**	**23.3**

prominent offshore ships and produce more false alarms near the land. Specially, only one offshore ship is missed in the result of our model, which shows a better detection performance. As for the ships that are docked intensively in the port, almost all of them could be detected with the proposed network. However, a large number of false alarms exist in the detection results of Faster-RCNN, RetinaNet, and FCOS. This may be due to these models have poor feature enhancement capability, which are not well suited for ship detection in complex scenes. On the contrary, the method proposed in this article shows strong generalization performance and robustness.

Overall, extensive experiments conducted on HRSID, SSDD, and IEEE 2020 Gaofen Challenge SAR dataset show that the method proposed in this article can provide competitive performance for ship detection in HR SAR images. Most importantly, the proposed method was successfully applied to SAR ship detection in the 2020 Gaofen Challenge, and our team ranked first among 292 teams in the preliminary contest and won the seventh place in the final match.

D. Validation on Complex and Large-Scene SAR Images

In this section, we use two complex and large-scene SAR images collected from the GF-3 satellite to test the practicability of the proposed network. Table I presents their imaging information, and we compared the proposed method with Faster-RCNN and RetinaNet. Some detailed detection results of two large-scene images are listed in Tables VIII and IX. (TP represents the number of correctly detected ships, FP is the number of the false alarms, and FN is the number of missing ships.) From Table VIII, it can be seen that most ships in the first image can be successfully detected with our method. There are only five false alarms meanwhile six missing ships. The $F1$ score is 90.89%, which is 5.18% and 7.71% higher than Faster-RCNN and RetinaNet. As presented in Table IX, in the second large-scene image, the $F1$ score of the proposed method is 85.71%, which is obviously higher than 70.59% of Faster-RCNN and 78.79% of RetinaNet. In order to evaluate the actual efficiency of different methods applied in large scenes, we take the total processing time of the whole scene image as the evaluation standard of speed. Significantly, the overall processing time includes the detection time of all slices and the

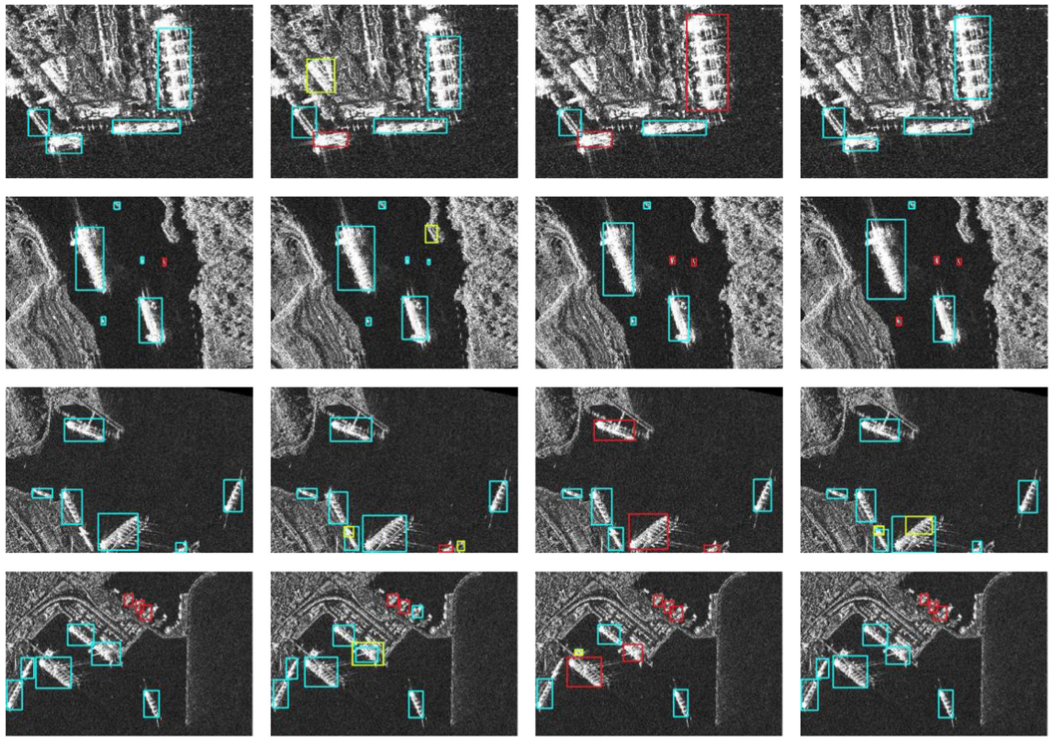

Fig. 17. Detection results of different methods on SSDD. The meanings of the boxes color are the same as Fig. 12. (a) Our methods. (b) Faster-RCNN. (c) FCOS. (d) RetinaNet.

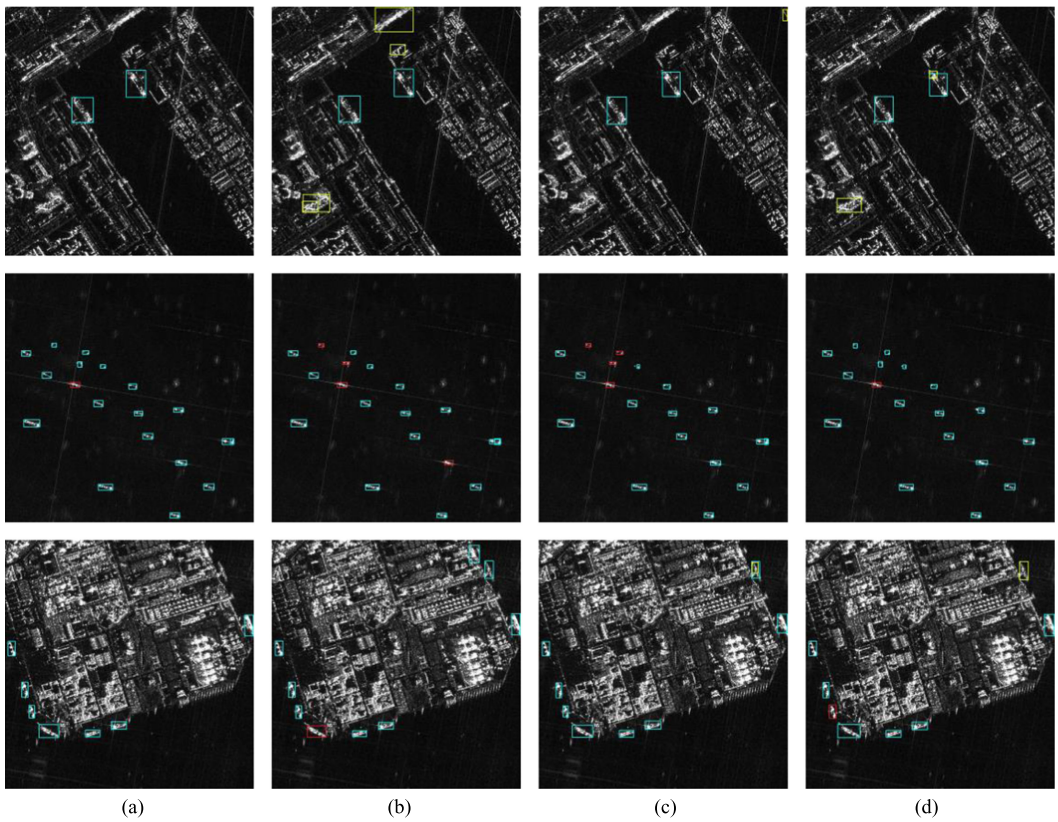

Fig. 18. Detection results of different methods on 2020 IEEE Gaofen Challenge Dataset. The meanings of the boxes color are the same as Fig. 12. (a) Our methods. (b) Faster-RCNN. (c) FCOS. (d) RetinaNet.

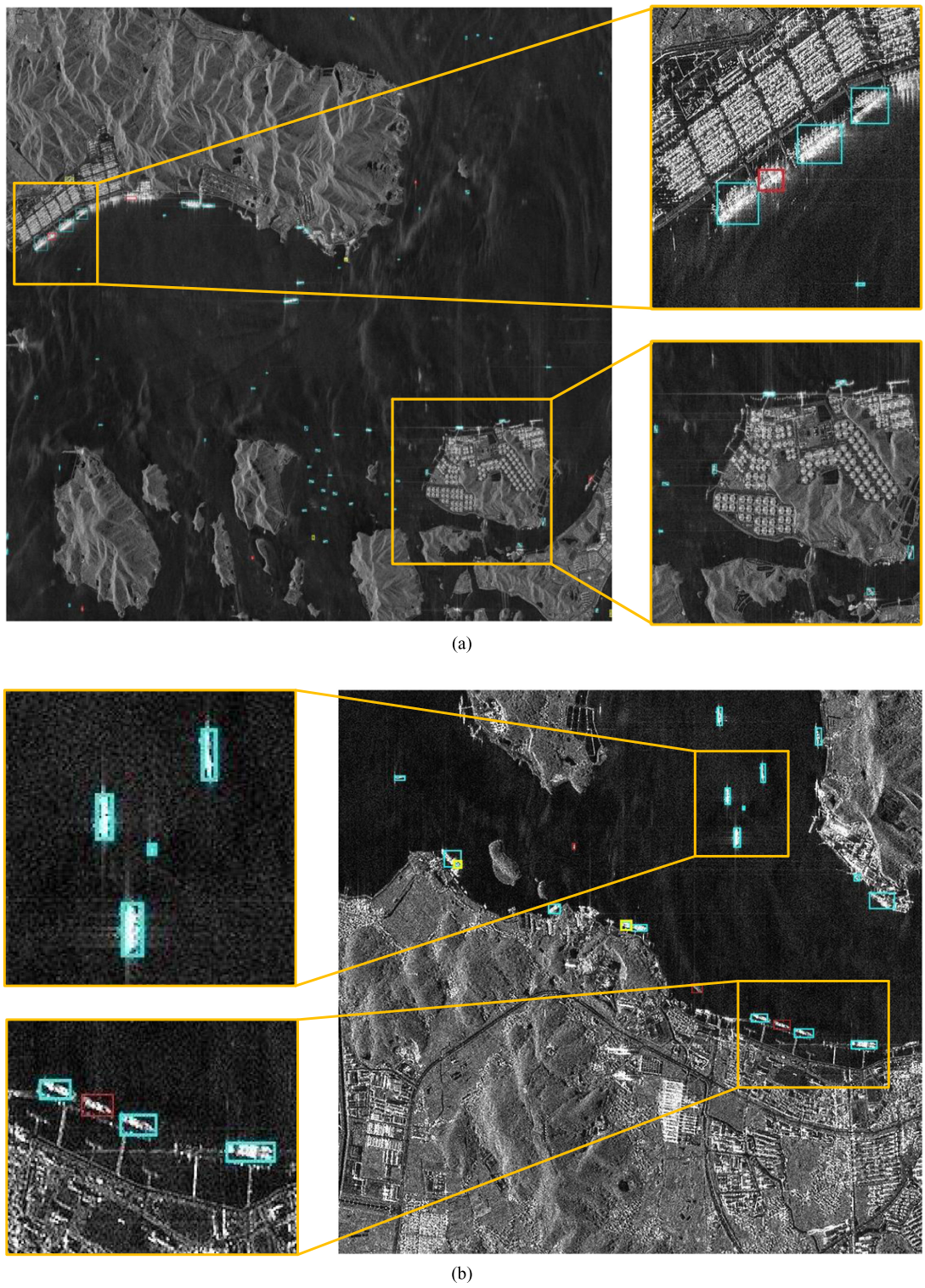

Fig. 19. Detection results in complex and large-scene SAR images. The meanings of the boxes color are the same as Fig. 12. Two specific areas marked with yellow rectangles are enlarged. (a) Image 1. (b) Image 2.

merging time of the entire large-scene image. It can be seen that the proposed method obviously has better detection speed than Faster-RCNN and RetinaNet since there is no need for our anchor-free method to design complex anchors that consume lots of computing resources.

Fig. 19 shows some detection results on the two large-scene SAR images. Two specific areas marked with yellow rectangles are enlarged and shown on the side of Fig. 19. It can be clearly observed that only one ship is missed in the complex scenes where ships are docked densely. At the same time, the proposed

method can effectively reduce false alarms at a higher detection speed. In other words, the method proposed in this article has a better generalization performance in terms of accuracy and efficiency for SAR ship detection.

IV. Conclusion

An improved anchor-free method is proposed in this article to improve the ship detection performance in HR SAR images. As the special anchor-free methods, FCOS can reduce false alarms and avoid missing ships through pixel-by-pixel prediction of the image. However, there is also a problem of the imbalance between positive and negative samples and semantic ambiguity caused by overlapping target borders in this network. Therefore, the CP module is proposed in this article to optimize the target position features. At the same time, in order to reduce the impact of the target fuzzy areas, the classification and boundary box regression methods are redesigned in our model. The extensive experiments show that the improvement proposed in this article can obtain an encouraging detection performance in terms of accuracy and speed. The proposed method was significantly applied to SAR ship detection in the 2020 Gaofen Challenge, and our team achieved a competitive result. Furthermore, in the future, we believe that the anchor-free detectors will achieve more fantastic performance in the field of SAR ship detection.

Acknowledgment

The authors would like to thank the Aerospace Information Research Institute, Chinese Academy of Sciences for providing the data used in this article, and for organizing the Gaofen Challenge on Automated High-Resolution Earth Observation Image Interpretation. Meanwhile, the authors would also like to thank the pioneer researchers in SAR ship detection and other related fields.

References

[1] X. Leng, K. Ji, S. Zhou, X. Xing, and H. Zou, "Discriminating ship from radio frequency interference based on noncircularity and non-Gaussianity in Sentinel-1 SAR imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 1, pp. 352–363, Jan. 2018.

[2] Z. Lin, K. Ji, X. Leng, and G. Kuang, "Squeeze and excitation rank faster R-CNN for ship detection in SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 5, pp. 751–755, May 2018.

[3] T. Li, Z. Liu, R. Xie, and L. Ran, "An improved superpixel-level CFAR detection method for ship targets in high-resolution SAR images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 1, pp. 184–194, Jan. 2018.

[4] L. Du, H. Dai, Y. Wang, W. Xie, and Z. Wang, "Target discrimination based on weakly supervised learning for high-resolution SAR images in complex scenes," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 1, pp. 461–472, Jan. 2020.

[5] P. Salembier, S. Liesegang, and C. Lopez-Martinez, "Ship detection in SAR images based on Maxtree representation and graph signal processing," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 5, pp. 2709–2724, May 2019.

[6] X. Wang, G. Li, and X. Zhang, "Contrast of contextual Fisher vectors for ship detection in SAR images," in *Proc. IEEE Int. Radar Conf.*, 2020, pp. 198–202.

[7] H. Lin, H. Chen, K. Jin, L. Zeng, and J. Yang, "Ship detection with superpixel-level Fisher vector in high-resolution SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 2, pp. 247–251, Feb. 2020.

[8] R. Wang *et al.*, "Context semantic perception based on superpixel segmentation for inshore ship detection in SAR image," in *Proc. IEEE Radar Conf.*, 2020, pp. 1–6.

[9] X. Wang, C. Chen, Z. Pan, and Z. Pan, "Fast and automatic ship detection for SAR imagery based on multiscale contrast measure," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 12, pp. 1834–1838, Dec. 2019.

[10] X. Zhang, T. Xie, L. Ren, and L. Yang, "Ship detection based on superpixel-level hybrid non-local MRF for SAR imagery," in *Proc. 5th Asia–Pac. Conf. Intell. Robot Syst.*, 2020, pp. 1–6.

[11] X. Wang, G. Li, X. Zhang, and Y. He, "Ship detection in SAR images via local contrast of Fisher vectors," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 9, pp. 6467–6479, Sep. 2020.

[12] S. Wang, M. Wang, S. Yang, and L. Jiao, "New hierarchical saliency filtering for fast ship detection in high-resolution SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 1, pp. 351–362, Jan. 2017.

[13] S. Yuan, Z. Yu, C. Li, and S. Wang, "A novel SAR sidelobe suppression method based on CNN," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 1, pp. 132–136, Jan. 2021.

[14] G. Xiong, F. Wang, L. Zhu, J. Li, and W. Yu, "SAR target detection in complex scene based on 2-D singularity power spectrum analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 9993–10003, Dec. 2019.

[15] J. Han, G. Li, and X. Zhang, "Refocusing of moving targets based on low-bit quantized SAR data via parametric quantized iterative hard thresholding," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 56, no. 3, pp. 2198–2211, Mar. 2020.

[16] G. Chen, G. Li, Y. Liu, X. Zhang, and L. Zhang, "SAR image despeckling based on combination of fractional-order total variation and nonlocal low rank regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 3, pp. 2056–2070, Mar. 2020.

[17] P. Iervolino and R. Guida, "A novel ship detector based on the generalized-likelihood ratio test for SAR imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 8, pp. 3616–3630, Aug. 2017.

[18] H. Lang, Y. Xi, and X. Zhang, "Ship detection in high-resolution SAR images by clustering spatially enhanced pixel descriptor," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5407–5423, Aug. 2019.

[19] X. Leng, K. Ji, X. Xing, S. Zhou, and H. Zou, "Area ratio invariant feature group for ship detection in SAR imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 7, pp. 2376–2388, Jul. 2018.

[20] 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 J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 2, pp. 536–550, Feb. 2018.

[21] G. Gao, K. Ouyang, Y. Luo, S. Liang, and S. Zhou, "Scheme of parameter estimation for generalized Gamma distribution and its application to ship detection in SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 3, pp. 1812–1832, Mar. 2017.

[22] X. Leng, K. Ji, S. Zhou, and X. Xing, "Ship detection based on complex signal kurtosis in single-channel SAR imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6447–6461, Sep. 2019.

[23] X. Sun, P. Wang, C. Wang, Y. Liu, and K. Fu, "PBNet: Part-based convolutional neural network for complex composite object detection in remote sensing imagery," *ISPRS J. Photogramm. Remote Sens.*, vol. 173, pp. 50–65, 2021.

[24] Q. He, X. Sun, Z. Yan, and K. Fu, "DABNet: Deformable contextual and boundary-weighted network for cloud detection in remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, to be published, doi: 10.1109/TGRS.2020.3045474.

[25] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 580–587.

[26] H. Wang and D. Hu, "Comparison of SVM and LS-SVM for regression," in *Proc. Int. Conf. Neural Netw. Brain*, 2005, pp. 279–283.

[27] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904–1916, Sep. 2015.

[28] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 1440–1448.

[29] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[30] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, pp. 386–397, Feb. 2020.

[31] Z. Cai and N. Vasconcelos, "Cascade R-CNN: Delving into high quality object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 6154–6162.

[32] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 779–788.
[33] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, and S. Reed, "SSD: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 21–37.
[34] J. Dai, Y. Li, K. He, and J. Sun, "R-FCN: Object detection via region-based fully convolutional networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 379–387.
[35] T. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, pp. 318–327, Feb. 2020.
[36] X. Zhou, D. Wang, and P. Krähenbühl, "Objects as points," Apr. 2019, *arXiv:1904.07850v2*, [Online]. Available: https://arxiv.org/abs/1904.07850
[37] C. Zhu, Y. He, and M. Savvides, "Feature selective anchor-free module for single-shot object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 840–849.
[38] Z. Tian, C. Shen, H. Chen, and T. He, "FCOS: Fully convolutional one-stage object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 9626–9635.
[39] T. Kong, F. Sun, H. Liu, Y. Jiang, L. Li, and J. Shi, "FoveaBox: Beyond anchor-based object detector," *IEEE Trans. Image Process.*, vol. 29, pp. 7389–7398, 2020.
[40] H. Law and J. Deng, "CornerNet: Detecting objects as paired keypoints," *Int. J. Comput. Vis.*, vol. 128, no. 3, pp. 642–656, 2020.
[41] M. Zhu *et al.*, "Rapid ship detection in SAR images based on YOLOv3," in *Proc. 5th Int. Conf. Commun., Image Signal Process.*, 2020, pp. 214–218.
[42] Y. Chen, J. Yu, and Y. Xu, "SAR ship target detection for SSDv2 under complex backgrounds," in *Proc. Int. Conf. Comput. Vis., Image Deep Learn.*, 2020, pp. 560–565.
[43] 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," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 10070–10087, Dec. 2019.
[44] Z. Deng, H. Sun, S. Zhou, and J. Zhao, "Learning deep ship detector in SAR images from scratch," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 6, pp. 4021–4039, Jun. 2019.
[45] L. Han, T. Zheng, W. Ye, and D. Ran, "Analysis of detection preference to CNN based SAR ship detectors," in *Proc. Inf. Commun. Technol. Conf.*, 2020, pp. 307–312.
[46] T. Zhang, X. Zhang, J. Shi, S. Wei, J. Wang, and J. Li, "Balanced feature pyramid network for ship detection in synthetic aperture radar images," in *Proc. IEEE Radar Conf.*, 2020, pp. 1–5.
[47] Z. Cui, Q. Li, Z. Cao, and N. Liu, "Dense attention pyramid networks for multi-scale ship detection in SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 8983–8997, Nov. 2019.
[48] Y. Zhao, L. Zhao, B. Xiong, and G. Kuang, "Attention receptive pyramid network for ship detection in SAR images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 2738–2756, May 2020, doi: 10.1109/JSTARS.2020.2997081.
[49] S. Chen, R. Zhan, W. Wang, and J. Zhang, "Learning slimming SAR ship object detector through network pruning and knowledge distillation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 1267–1282, 2021, doi: 10.1109/JSTARS.2020.3041783.
[50] 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," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 1938–1958, Jan. 2021, doi: 10.1109/JSTARS.2021.3049851.
[51] Z. Cui, X. Wang, N. Liu, Z. Cao, and J. Yang, "Ship detection in large-scale SAR images via spatial shuffle-group enhance attention," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 1, pp. 379–391, Jan. 2021.
[52] 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," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 2, pp. 1331–1344, Feb. 2021.
[53] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
[54] T. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 936–944.
[55] A. Neubeck and L. Van Gool, "Efficient non-maximum suppression," in *Proc. 18th Int. Conf. Pattern Recognit.*, 2006, pp. 850–855.
[56] T. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2999–3007.
[57] 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, pp. 120234–120254, 2020.
[58] J. Li, C. Qu, and J. Shao, "Ship detection in SAR images based on an improved faster R-CNN," in *Proc. SAR Big Data Era, Models, Methods Appl.*, 2017, pp. 1–6.
[59] 2020 Gaofen Challenge on Automated High-Resolution Earth Observation Image Interpretation, 2020. [Online]. Available: http://en.sw.chreos.org

Zhongzhen Sun received the B.S. degree in communication engineering, in 2018, from the National University of Defense Technology, Changsha, China, where he is currently working toward the M.S. degree in electronics and communication engineering with the State Key Laboratory of Complex Electromagnetic Environment Effects on Electronics and Information System, National University of Defense Technology.

His research interests include intelligent electronic countermeasures and assessment, and SAR image target recognition.

Muchen Dai received the B.S. degree in electronic information engineering from Tsinghua University, Beijing, China, in 2018, and the M.S. degree in information and communication engineering from the National University of Defense Technology, Changsha, China, in 2020.

His research interests include remote sensing information processing and SAR automatic target recognition.

Xiangguang Leng (Member, IEEE) received the B.S. degree in remote sensing science and technology from Wuhan University, Wuhan, China, in 2013, and the M.S. and Ph.D. degrees in information and communication engineering from the National University of Defense Technology, Changsha, China, in 2015 and 2019, respectively.

His research interests include remote sensing information processing, synthetic aperture radar image interpretation, and marine surveillance.

Yu Lei received the B.S. degree in electronic and information engineering from Dalian Minzu University, Dalian, China, in 2019, where he is currently working toward the M.S. degree in electronics and communication engineering with the State Key Laboratory of Complex Electromagnetic Environment Effects on Electronics and Information System, National University of Defense Technology, Changsha, China.

His research interests include intelligent electronic countermeasures and assessment, and SAR image target intelligent interpretation.

Boli Xiong received the B.S. degree in electronic engineering, M.S. degree in photogrammetry and remote sensing, and Ph.D. degree in communication engineering from the National University of Defense Technology, Changsha, China, in 2004, 2006, and 2012, respectively.

He is currently an Associate Professor with the College of Electronic Science and Technology, National University of Defense Technology. His research interests include SAR/PolSAR image registration and change detection, SAR automatic target recognition, and machine learning.

Kefeng Ji (Member, IEEE) received the M.S. and Ph.D. degrees in information and telecommunication engineering from the National University of Defense Technology, Changsha, China, in 1999 and 2003, respectively.

He is currently a Professor with the College of Electronic Science and Technology, National University of Defense Technology. His research interests include synthetic aperture radar (SAR) image interpretation, target detection and recognition, feature extraction, and SAR and automatic identification system association.

Gangyao Kuang (Senior Member, IEEE) received the B.S. and M.S. degrees in geophysics from the Central South University of Technology, Changsha, China, in 1988 and 1991, respectively, and the Ph.D. degree in communication and information from the National University of Defense Technology, Changsha, in 1995.

He is currently a Professor with the College of Electronic Science and Technology, National University of Defense Technology. His research interests include remote sensing, synthetic aperture radar (SAR) image processing, change detection, SAR ground moving target indication, and classification with polarimetric SAR images.