A Prediction and Prior Information Guided SAR Ship Detection Method

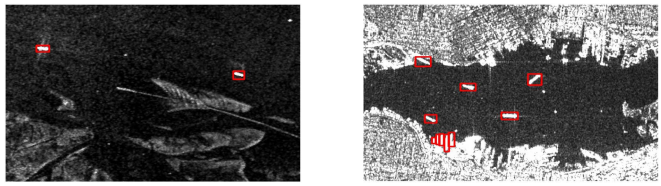
Yao Wang, Ganggang Dong, Shuai Shao, and Hongwei Liu

***Abstract*—Data-driven ship detection methods via deep learning algorithms are the recent research hotspot. In this family of models, it is essential to divide the positive and the negative samples. The commonly used strategy is called label assignment. In the previous detectors, label assignment resulted from the handcrafted heuristic techniques. Therefore, it needs to tune the hyperparameters, and improper settings will deteriorate model performance. Moreover, a significant inconsistency between the training and the testing objective is available. To address these issues, a prediction and prior information guided label assignment technique is proposed. A network specific to ship detection is then presented to improve multiscale detection performance. First, the IoU prediction is specified as the estimation of localization precision. The quality of the candidate anchor is evaluated by a combination of classification and positioning. It reduces the inconsistency between training and testing. Besides, the learning status of the current model and the anchors' prior information are exploited simultaneously. Optimum positive samples are selected in an adaptive manner. Finally, a multiscale ship detection network is designed, concentrating on small ships' rich contextual information. After feature fusion and feature enhancement on different scales, shallow texture and deep semantic information are combined to detect multiscale ships. Multiple experiments are conducted on SSDD and HRSID datasets, and the results demonstrate the advantage of the proposed method compared with advanced detectors.**

***Index Terms*—HRSID, label assignment, SSDD, ship detection.**

I. Introduction

AS AN active microwave sensor, synthetic aperture radar (SAR) can produce massive high-resolution images throughout the day and under all weather conditions. SAR ship detection is illustrated in Fig. 1, and its task is to locate the precise positions of all ships in the image. Ship detection is a fundamental component of the SAR ocean image interpretation. Detecting ship targets provides a broad range of application perspectives [33]. Ship detection serves to monitor and manage specific maritime zones, bays, and ports. Illegal fishing and smuggling can be found in real time [23]. Researches on SAR target detection algorithms date back to the 1980 s and can be divided into traditional and deep learning-based methods. As a traditional representative and widely used method, the constant false-alarm rate (CFAR) algorithm distinguishes between target and clutter while maintaining the constant value of false alarm rate [24]. CFAR calculates the detection threshold according to the statistical characteristics of the clutter around the target [25]. The result of CFAR-based methods severely relies on the statistical model of clutter. Model mismatch leads to performance decline and weak generalization ability when dealing with data including various sea clutter. In conclusion, traditional methods are difficult to achieve satisfactory detection results in complex scenarios [1].



(a) (b)

Fig. 1. Examples of SAR ship detection. (a) Offshore scene. (b) Inshore scene.

Benefitting from the development of artificial intelligence, object detection algorithms based on deep learning have achieved a great success. The convolutional neural network (CNN) can automatically extract image features and generate more robust results. However, label assignment, which is the ship detectors' crucial process, has yet to be investigated sufficiently. Label assignment refers to distinguishing positive and negative samples during the training phase. Its result determines the learning objectives of each position of the feature map, which directly impacts model performance. Anchor-free detectors like FCOS assign points in the feature map area corresponding to ground truth as positive samples [17]. Two-stage detectors like Faster R-CNN assign one or more anchors as positive samples according to their intersection over union (IoU) with the ground-truth (GT) box being maximum or its IoU exceeding a certain threshold [3]. One-stage detectors like the Yolo series divide an image into multiple grid cells [2], [4]. When the center of a target falls into a grid, this grid is in charge of predicting that target. Anchors will be assigned as positive samples if their bounding box overlaps a GT box more than any other. The traditional label assignment strategies mainly follow a fixed rule based on manual habits. They are always suboptimal for a variety of ships in complex SAR scenes. Moreover, traditional label assignment strategies require researchers to spend plenty of time tuning hyperparameters (the number of positive samples, the threshold for dividing positive samples, and the artificially

Manuscript received 28 January 2023; revised 15 May 2023 and 25 June 2023; accepted 9 July 2023. Date of publication 24 July 2023; date of current version 4 August 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 61971324, Grant 61525105, Grant 6210010665, and in part by the National Key Laboratory of Radar Signal Processing Fund. *(Corresponding author: Ganggang Dong.)*

The authors are with the National Key Laboratory of Radar Signal Processing, Xidian University, Xi'an 710071, China (e-mail: dongganggang@nudt.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2023.3298483

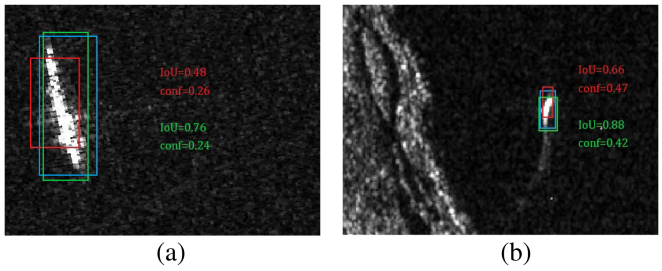


Fig. 2. Visualization of two drawbacks brought by the significant misalignment between the training objective and the testing one. The blue bounding boxes denote the ground truth and the red bounding boxes mark the detection results retained via NMS. The green bounding boxes mark the detected boxes removed via NMS. They are yielded by Yolo v4. (a) Offshore scene. (b) Inshore scene.

set ratio of positive and negative samples), which adds tedious and unnecessary work. Furthermore, inappropriate settings of hyperparameters can significantly affect the performance of models. It restricts the ability to generalize the model to various datasets. Especially, there is a discrepancy between the training (minimizing both classification and localization loss of anchors) and the testing stage (expurgating duplication of bounding boxes only according to the classification score via nonmaximum suppression) [5], [39]. The cases of inconsistency between classification confidence and localization accuracy are shown in Fig. 2. The process of NMS uses confidence as the only indicator and accurate localized bounding boxes are not always the ones with the highest confidence score. The red boxes are the final detection results retained by the NMS, but the removed green boxes are more accurate. The detector may miss the ship in Fig. 2(a) when the IoU of the retained detected box is under 0.5. Besides, a less precise detected box may suppress an accurately localized one in Fig. 2(b).

Recently, extensive experiment results have revealed the great potential of making label assignment strategies adaptive and prediction-aware [6], [7], [8]. However, compared to selecting positive samples and then calculating the loss, cost-based methods must calculate the cost of all the anchors. Most calculations are superfluous and occupy vast amounts of memory resources. To solve these problems, this article proposes a new solution called prediction and prior information guided label assignment (PPIG) to sample high-quality positives. The main contributions of this article are as follows:

1) A novel label assignment strategy, PPIG, is proposed for SAR ship detection. The proposed method eliminates the need for manual hyperparameter tuning. In addition, PPIG assigns positive samples in an adaptive manner by integrating the current learning status of the model and anchors' prior information.
2) IoU prediction is specified as the estimation of location accuracy. The detection quality of bounding boxes is measured by classification and localization to reduce the discrepancy between the training and testing objectives.
3) A multiscale ship detection network is designed, concentrating on small ships' rich contextual information. The proposed network extracts more semantic information about various ships through feature enhancement and feature fusion on different scales.

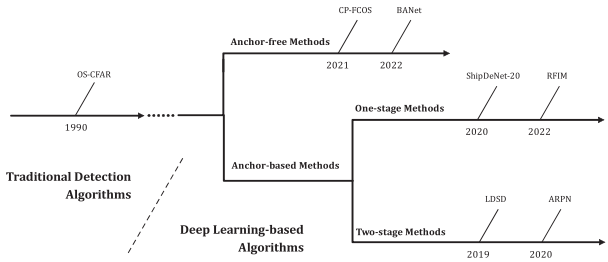


Fig. 3. Some representative SAR ship detection methods.

The remainder of this article is structured as follows. Section II provides an overview of related work. Section III elaborates on the proposed methods. Section IV presents experimental results on two datasets and related analyses. Finally, certain conclusions are reached in Section V.

II. Related Work

Because of the difference between optical and SAR images, obtaining satisfactory results in ship detection is still tricky. Many networks have been designed specifically focusing on the characteristics of SAR data [34]. Ship detectors can also be divided into anchor-free, one-stage, and two-stage methods. Some representative detection methods are presented in Fig. 3. Anchor-free models find the key points of ship targets and then predicts the corresponding position [26]. Sun et al. [27] proposed the category-position module in the FCOS network (CP-FCOS), which utilizes the information from classification branch features to improve the ship position regression performance. Hu et al. [28] designed a balance attention network (BANet) with a local attention module and a nonlocal attention module to improve the generalization capability for multiscale ship detection. Besides, one-stage models only predict once to obtain all the ship targets' position, so the detection speed is faster. Zhang and Zhang [29] designed a lightweight ship detection network called "ShipDeNet-20" with only 20 convolution layers, and the accuracy was still competitive. Yang et al. [30] devised a receptive field increased module and introduced a coordinate attention module to suppress false negatives. In comparison, two-stage models achieve accurate detection performance through coarse localization and position refinement of ship targets. Deng et al. [31] designed a condensed backbone network with dense blocks learning deep ship detector from scratch and proposed a new loss function to improve the recall rate. Zhao et al. [32] proposed the attention receptive pyramid network to refine feature representation, which combines the receptive fields block and the convolutional block attention module as the lateral connection.

III. Proposed Methods

The significant inconsistency between the training and testing objectives affects the detection performance and the anchors'

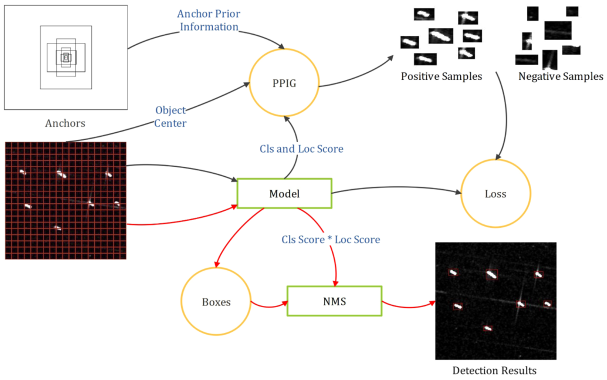


Fig. 4. Structure of the proposed methodology. The black lines indicate the training stage and the red lines indicate the test phase.

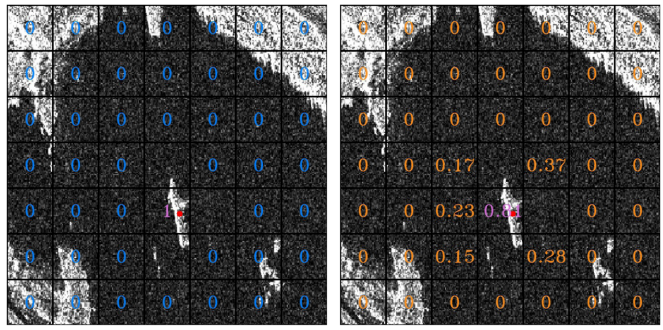


Fig. 5. Difference between confidence and IoU prediction in learning objectives. (a) Confidence. (b) IoU prediction.

prior information has been ignored in some methods. Furthermore, unnecessary calculations consumed large quantities of memory resources. To solve these problems, this article proposes a novel label assignment for SAR ship detection. The overall framework with PPIG is illustrated in Fig. 4. First, a single convolution layer is added to predict the IoU of each anchor without sacrificing efficiency. The predicted IoU reflects the positioning accuracy of the anchors and is conducive to selecting appropriate positive samples in the training stage [9]. Second, PPIG integrates the anchor classification and localization quality scores with the prior information. It leverages semantic feature to guide the label assignment by using the model's learning status. Additionally, introducing IoU prediction makes it prone to assign anchors containing more target information as positive samples. Third, the quality of the candidate boxes is measured by classification and localization simultaneously. It is utilized in the NMS process to reduce the discrepancy between training and testing objectives. Finally, since the candidate anchors of poor quality are excluded from the positive sample, the false alarm rate will decrease in the test stage.

Besides, the gap between optical deep learning methods and SAR ship detectors is a great challenge [10]. The ship size in the SAR image is significantly smaller than the targets in the optical image [11] and the small ships are difficult to detect. It leads to a performance decline to directly migrate optical models to SAR images [12]. To solve this problem, MSDNet is designed to extract more semantic and texture information about multiscale ships.

A. *Phase 1. IoU Prediction*

At the training stage, detectors optimize both the classification loss and the localization loss. However, at the testing stage, detectors only expurgate the duplication of bounding boxes based on the classification score via NMS [15]. The discrepancy leads to a mismatch problem between the classification scores and localization quality. It makes the training objective suboptimal and diminishes detector performance. Anchors' IoU with ships can be calculated directly from GT information during training. However, the detected boxes' IoU with its corresponding ship is unavailable in the test. To solve this problem, only one convolution layer is added to predict the IoU of each anchor with its corresponding ship. The sigmoid activation function is used to obtain valid IoU values. Furthermore, the IoU prediction is trained jointly with the classification and localization. IoU prediction loss is defined as Focal loss [16] between the predicted IoU and the true IoU. Specifically, an anchor quality indicator is devised based on classification score and localization estimated accuracy simultaneously. The intuitive way is illustrated in (1) to multiply the confidence score by the predicted IoU. The evaluating quality function is obtained as follows:

$$\delta(\omega, i) = \text{conf}(\omega, i) \times \text{pred\_IoU}(\omega, i) \tag{1}$$

where ω and i are candidate anchors and the indexes of ships. $\text{conf}(\omega, i)$ and $\text{pred\_IoU}(\omega, i)$ refer to the predicted confidence scores and the predicted IoU of corresponding candidate anchor, respectively. Ship detectors should only distinguish between foreground and background. Therefore, the confidence score is instead of classification. $\delta(\omega, i)$ can act as a scoring function reflecting the prediction quality of the candidate anchors. The score ranks all the detected boxes in the subsequent NMS computation procedure. The inconsistency between the training objective and the testing one is then reduced. This article uses a single convolution layer to enhance the detector's accuracy without sacrificing efficiency.

To intuitively explore the differences between the predicted IoU and confidence, the learning objectives of the two are shown in Fig. 5. For the sake of clarity, Fig. 5 results in a grid of $7 * 7$ instead of the experimental minimum of $19 * 19$. The predicted IoU and confidence are one-to-one corresponding to the initial anchor. Moreover, each feature map matches three initial anchors of different sizes. Therefore, the learning objectives are corresponding to the initial anchor [138, 70]. The positive sample's grid is marked by the red dot in the ship's center, and the learning objective is in the lavender font. The learning objectives of the negative samples of confidence and the predicted IoU are marked in blue and orange, respectively. The areas without fonts are ignored samples. The result of Fig. 5 intuitively demonstrates that the predicted IoU can learn more

knowledge about the target than confidence. The fundamental is similar to the label smoothing technique. Confidence only contains information about whether the grid has an object but has no idea about the IoU between the object and the predicted bounding box. In the proposed method, the confidence score is used to distinguish whether the center of the ship target falls within the grid.

B. Phase 2. A Prediction and Prior Information Guided Label Assignment Method

Traditional label assignment follows a hand-designed fixed rule. However, it cannot be optimal for various ships in complex SAR scenes. If the low-quality prediction is forced to assign as a foreground sample, the convergence direction of the network could be suboptimal, leading to degradation of detection performance. To this end, PPIG is devised to adaptively assign positive samples according to the quality of prediction and prior information. Let N and ξ_i denote the number of ships and clustering of anchors produced by the pixel closest to the ith ship's center, respectively. The quality of the ith ship's candidate anchor ω is represented by $\Phi\ (\omega, i)$. PPIG considers selecting positive samples as finding the optimal anchor $\hat{\omega}$. As presented in (2) and (3), the positive sample $\hat{\omega}$ is the anchor with the highest quality $\Phi(\omega, i)$

$$\hat{\omega} = \arg\max_{i}^{N} \Phi\ (\omega, i) \tag{2}$$

$$\Phi\ (\omega, i) = \underbrace{\mathbb{1}\ [\omega \in \xi_i]}_{\text{center rule}} \times [\underbrace{\mathrm{IoU}_{\text{anchor\_}GT}\ (\omega, b_i)}_{\text{anchor prior knowledge}} + \underbrace{\text{conf}\ (\omega, i) \times \text{pred\_IoU}\ (\omega, i)}_{\text{prediction-aware anchor quality score}}]. \tag{3}$$

Here $\mathrm{IoU}_{\text{anchor\_}GT}(\omega, b_i)$ is the IoU between the candidate anchors and the ships. PPIG simultaneously takes into account the center rule [17], anchors' prior knowledge, and prediction-aware quality score. First, to suppress the low-quality anchors generated by the locations far from the center of ships, PPIG follows the center rule. Positive samples are selected only from anchors predicted by the pixel closest to the center of a ship instance in the predefined feature map layer. Besides, considering the ship bounding box b_i prior IoU with ship instance is conducive to allowing positive samples to contain more target information.

Additionally, the IoU of the anchor and its corresponding ship is used as the prior knowledge. After finding the grid where the GT center of the ship is located, the IoU could be calculated shown in Fig. 7. The prior knowledge reflects the scale and aspect ratio similarity between the anchor and the corresponding ship. In the early stage of training, the model has yet to converge. Therefore, the current learning state of the model gives inaccurate results of IoU prediction and confidence. In this case, the positive sample selected according to the prediction quality may not be the best choice. Introducing prior information can avoid the suboptimal positive sample selection caused by the inaccurate prediction.

Algorithm 1: PPIG.

Input: I is a set of ship ground-truth boxes on the SAR image
b_i is the ship bounding box
ℓ is the number of feature pyramid levels
ω is a set of candidate anchors
Output: P is a set of positive samples

1: **for** each ground-truth $i \in I$ **do**
2: build an empty set for candidate anchors of the ground-truth i: $\omega \leftarrow \emptyset$;
3: **for** each level $j \in [1, \ell]$ **do**
4: $\xi_i \leftarrow$ pick anchors predicted by the pixel closest to the center of a ship instance in the feature map;
5: $\omega = \omega \cup \xi_i$;
6: **end for**
7: compute IoU between ω and i: $\mathrm{IoU}_{\text{anchor\_}GT}\ (\omega, b_i)$;
8: decode the confidence from network predictions: conf (ω, i);
9: decode the predicted IoU from network predictions: pred_IoU (ω, i);
10: compute the quality of the ship i with the candidate anchors ω: $\Phi\ (\omega, i)$;
11: find the optimal positive sample $\hat{\omega}$ according to $\hat{\omega} = \arg\max \sum_i^N \Phi\ (\omega, i)$;
12: $P = P \cup \hat{\omega}$;
13: **end for**
14: **return** P;

Finally, the quality of anchors should be evaluated reflecting the learning status of the current model and the prediction quality should be based on classification score and localization precision simultaneously. PPIG dispenses with hyperparameter tuning and improves the detector's performance by dynamically assigning the optimal predictions as positive samples. Algorithm 1 describes how the PPIG works in the training phase. For each ground truth i on the image, its candidate anchors are found according to the center rule. As described in Lines 3 to 6, on each pyramid level, candidate anchors ω are only picked from anchors predicted by the pixel closest to the center of a ship instance. Then, the true IoU $\mathrm{IoU}_{\text{anchor\_}GT}\ (\omega, b_i)$ between these candidates and the ground-truth box b_i is computed in Line 7. The confidence conf (ω, i) and the predicted IoU pred_IoU (ω, i) are decoded from the network prediction in Line 8 and Line 9. Because the result of PPIG should reflect the learning status of the current model and the quality of candidate anchors should be evaluated based on classification and localization simultaneously. The quality of the candidate anchors ω: $\Phi\ (\omega, i)$ is calculated in accordance with (3) in Line 10. Finally, the candidate anchor with the highest quality is selected as the final positive sample $\hat{\omega}$ in Lines 11 and 12. For each ship, only one positive sample is selected for the loss calculation. Compared to assigning multiple positive samples to a ship, the proposed method reduces the false alarm rate by excluding low-quality samples from the training objectives.

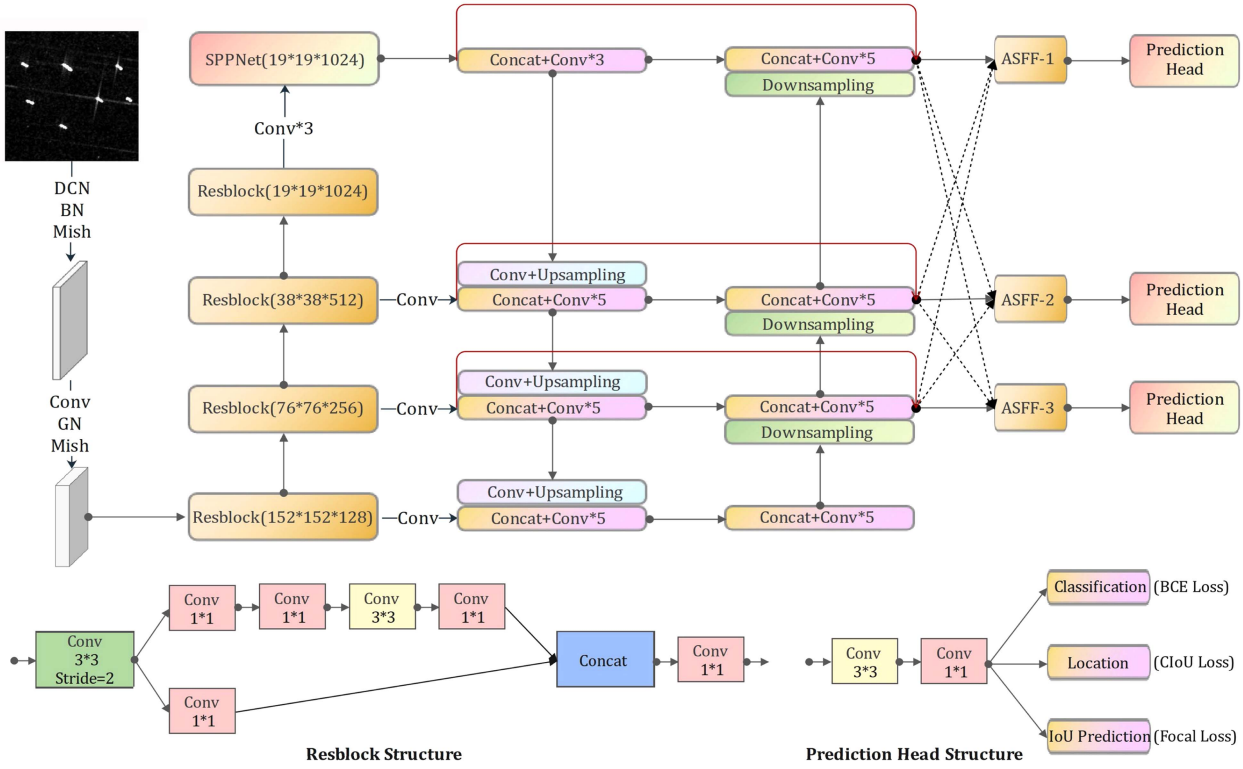


Fig. 6. Configuration of MSDNet.

C. Phase 3. Multiscale Ship Detection Network

In view of the difficulty of small ship detection, MSDNet is designed based on small ship detection via deformable convolutional networks (SSD-DCN) [13]. MSDNet extends ideas of feature fusion and feature enhancement to boost performance further. In the ship detection network, the high-level feature maps usually have larger receptive fields with more semantic information but lack specific information on the ship target. In contrast, low-level feature maps can provide relatively more ship target information, especially for small targets. The multilevel feature fusion (MLFF) is proposed to extract multiscale features. PANet and MLFF are shown in Fig. 8 [14]. P3 is added in the bottom-to-up and up-to-bottom feature fusion. It takes full advantage of ship target information in low-level feature maps and makes the detectors focus more on the small ships. The obtained multiscale fusion features contain rich contextual information. Besides, many times feature fusion will lead to forgetting some knowledge in the original feature map. To solve this problem, the feature reuse branches are added to P4, P5, and P6. The feature reuse branches avoid missing contextual information while extracting more small-ship detailed information. MLFF fuses and refines shallow texture features and deep semantic features, aiming to integrate the location and classification information into the detection head.

Moreover, an intense network will not extract more semantic information about ships, but will produce feature redundancy. The structure of MSDNet is illustrated in Fig. 6. MSDNet expurgates the times (2,8,8,4, respectively) of repetition of Resblock in the backbone to improve the model's efficiency. Additionally,

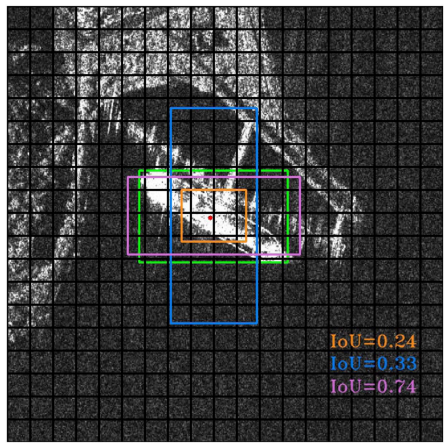


Fig. 7. Schematic diagram of prior knowledge calculation. The rectangle with green color marks the ground truth and rectangles with other colors mark the candidate anchors.

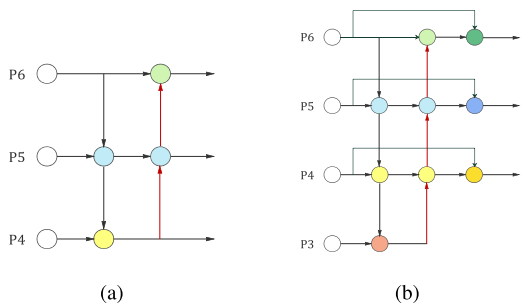


Fig. 8. Feature pyramid network design. (a) Original PANet. (b) Proposed method called MLFF.

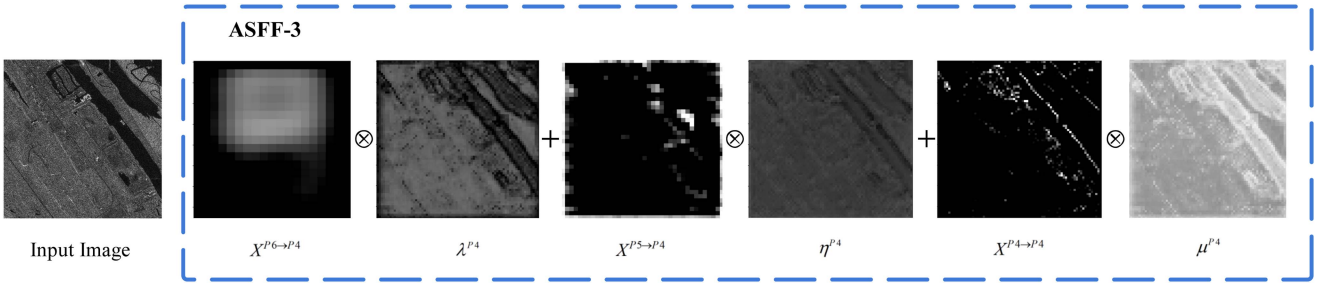


Fig. 9. Illustration of the adaptively spatial feature fusion (ASFF) module.

to suppress the inconsistency between the different scales of the feature pyramid, MSDNet deploys ASFF to improve the scale-invariance of features [18]. Moreover, MSDNet halves the channel dimension of ASFF processing's output by half using a 1×1 convolution. It reduces the loss of semantic information by preventing a rapid decrease in the number of channels of the feature map. As shown in Fig. 9, different levels of feature maps are gathered in the different levels of the ASFF module and fused according to the spatial weights learned by the network. Since feature maps at different levels have different resolutions and channel numbers, convolution and pooling operations are used to resize them to the same scale. $X^{Pa\to Pb}$ refers to the feature maps resized from level Pa to level Pb. (i, j) represents the specific position. $\lambda^{P4}(i,j)$, $\eta^{P4}(i,j)$, $\mu^{P4}(i,j)$ denote the spatial adaptive weights for the different levels of feature maps. According to [18], we force $\lambda^{P4}(i,j) + \eta^{P4}(i,j) + \mu^{P4}(i,j) = 1$ and $\lambda^{P4}(i,j), \eta^{P4}(i,j), \mu^{P4}(i,j) \in [0,1]$. Thus, the equation of feature fusion at level P4 is shown as follows:

$$Y^{P4} = \lambda^{P4}(i,j) \times X^{P6\to P4} + \eta^{P4}(i,j) \times X^{P5\to P4} + \mu^{P4}(i,j) \times X^{P4\to P4}. \tag{4}$$

Y^{P4} implies the output feature maps of the ASFF module. By introducing ASFF, contradictory information may be filtered out, and some discriminative clues may dominate at the corresponding position. Finally, MSDNet adds an IoU prediction layer to replace the classification layer by considering ship detection as a binary classification problem. MSDNet's prediction head is shown in the right corner of Fig. 6. The IoU prediction head is trained jointly with the confidence head and location head. Loss of IoU prediction is also calculated for positive and negative samples instead of all anchors. Furthermore, the overall loss function can be divided into three parts contributing to the confidence, location, and IoU prediction [19].

IV. Experimental Results

A. Implementations

1) Datasets: Experiments are conducted to evaluate the advantage of the proposed methods on SSDD and HRSID dataset [20]. SSDD is a public SAR ship detection dataset containing images with different resolutions, sensors, polarizations, sea states, and ship types [36]. There are 1160 SAR images and 2358 ships with different sizes in the SSDD. HRSID consists of 5604 high-resolution SAR images with 800×800 pixels [37]. These images are cropped from 136 panoramic SAR imageries containing 16 951 ships. HRSID includes SAR images of different polarizations, sea states, and sensors. The resolution is 0.5 m, 1 m, and 3 m, respectively. It is noted that detectors will achieve high performance in the condition of a relatively simple test set. This article uses a relatively complex test set (including more inshore images and fewer offshore images) and the partitioning of datasets follows the random partition mode. The training subset, validating subset, and test subset are divided in the proportion: $7:2:1$.

2) Training Settings: All evaluations were implemented with the PyTorch framework. Faster R-CNN, Yolo v3 [21], Yolo v4, Yolo v5, EfficientDet, and CenterNet [22] were trained with pretrained models for contrastive experiments and only the proposed methods were trained from scratch. The Adam method is used to train the network. The beginning learning rate is set to 0.001, and the weight decay mode is cosine. The freeze and unfreeze epoch batch size is set to 8 and 2, respectively.

3) Evaluation Metric: The precision, recall, average precision (AP), and F1 score are adopted to evaluate the detection performance. These widely used evaluation metrics are calculated by three components, true positive (TP), false positive (FP), and false negative (FN). The precision represents the correctness of all predicted samples and the recall indicates the coverage of all targets. In general, precision and recall are defined as

$$\text{Precision} = \frac{TP}{TP + FP} \tag{5}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{6}$$

where TP denotes the number of correctly detected targets, FP indicates the number of false alarms, and FN represents the number of missing targets. Besides, a detected bounding box is recognized as a correct prediction when its IoU with the corresponding ground truth is higher than 0.5. Since precision and recall affect each other, AP and F1 score are used to evaluate the overall performance of detection algorithms. AP and F1 scores are calculated as follows:

$$AP = \int_0^1 P(R)dR \tag{7}$$

$$F_1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{8}$$

where $P(R)$ is the precision-recall curve and R represents recall.

TABLE I
EXPERIMENTS RESULTS OF THE IOU PREDICTION IN PPIG AND USING A SINGLE IOU BRANCH AS A CLASSIFICATION RESULT

Method	AP	Precision	Recall
Conf∗Pred_IoU (PPIG)	94.05%	83.44%	95.05%
Conf	92.88%	82.35%	94.59%
Pred_IoU	93.73%	83.33%	94.59%

TABLE II
INFLUENCE OF THE PROPOSED METHODS ON SSDD

MSDNet	IoU Prediction	PPIG	AP
			86.05%
✓			89.41%
✓	✓		89.59%
✓	✓	✓	**94.05%**

The bold value used to highlight the best value.

TABLE III
EXPERIMENTAL RESULTS OF PANET AND MLFF ON SSDD

Method	AP	Precision	FPS
Yolo v4 (PANet)	86.05%	72.87%	19.32
Yolo v4+MLFF	88.04%(+2.01%)	82.50%(+9.63%)	18.92

B. Ablation Experiments

First, experiments are reported in Table I to investigate the difference between the IoU prediction in PPIG and using a single IoU branch as a classification result.

According to Table I, PPIG is better than using a single IoU branch as the classification result. The result demonstrates the superiority of evaluating the prediction quality of the detected boxes based on classification and positioning. It is also worth noting that IoU prediction has fewer false alarms when the recall rate is the same as the confidence. Confidence only contains information about whether the grid has an object, but IoU prediction can predict the IoU between the object and the bounding box. Therefore, IoU prediction can learn more information related to the object.

In addition, some ablation experiments on SSDD are performed to evaluate the advantage of the proposed methods reported in Table II. MSDNet improves the AP by 3.36% compared with Yolo v4 [23]. It reveals that MSDNet makes the general detector more suitable for SAR ship detection. With the introduction of IoU prediction, AP has been further improved. It demonstrates the potential to measure the detected boxes based on both classification and localization. Besides, PPIG boosts the AP from 89.59% to 94.05%. The result demonstrates the success in assigning the positive samples adaptively according to integrating the model's current learning status and anchors' prior information.

Besides, experiments on SSDD are conducted to prove the advantages of MLFF over PANet reported in Table III. MLFF improves the AP by 2.01% and the precision by 9.63% compared with PANet. Besides, the inference speed of MLFF decreases by 0.4 FPS compared to PANet. The significant increase in detection accuracy is worth the slight increase in inference time.

TABLE IV
STUDY RESULTS OF PPIG ON SSDD

Method	AP	Precision	Recall
Yolo v3	83.03%	77.59%	85.88%
Yolo v3+PPIG	87.58%(+4.55%)	80.90%	88.93%
Yolo v5m	91.66%	72.92%	93.51%
Yolo v5m+PPIG	93.74%(+2.08%)	79.05%	95.04%
Yolo v5l	92.08%	81.79%	94.27%
Yolo v5l +PPIG	94.53%(+2.45%)	84.00%	96.18%
RetinaNet	82.87%	68.12%	85.29%
RetinaNet +PPIG	85.22%(+2.35%)	72.82%	87.81%

TABLE V
COMPARISONS ON TWO EXTRA LABEL ASSIGNMENT METHODS

Method	AP	Precision	Recall	F1-score
MSDNet	89.41%	76.43%	90.54%	82.89%
$\alpha(\omega, i)$	91.04%	75.28%	91.89%	82.76%
$\beta(\omega, i)$	91.42%	71.78%	92.79%	80.94%
PPIG	94.05%	83.44%	95.05%	88.87%

The comparison between the detection results of PANet and MLFF in Yolo v4 on the HRSID dataset is shown in Fig. 10. The false alarm of MLFF is fewer than that of PANet, and the detection ability of small ships is also improved.

In addition, in order to verify the transferability of PPIG, some ablation experiments on different detection frameworks have been conducted, as shown in Table IV. Compared with the Yolo v3, Yolo v5m, Yolo v5l, and RetinaNet, PPIG increases the AP by 4.55%, 2.08%, 2.45%, and 2.35%, respectively [38]. Precision and recall rates are also improved in the above four methods. The results show that the application of PPIG brings a noticeable improvement. It demonstrates that PPIG is a general and robust algorithm. Therefore, the proposed methods can be ported to other SAR detection frameworks to improve the performance.

Furthermore, the label assignment method is designed in (9) and (10) to evaluate the necessity of prior information further

$$\hat{\omega} = \arg\max_{i}^{N} \alpha(\omega, i) \tag{9}$$

$$\alpha(\omega, i) = 1[\omega \in \xi_i] \times [\text{conf}(\omega, i) \times \text{pred\_IoU}(\omega, i)] \tag{10}$$

Another label assignment method is designed in (11) and (12) to evaluate the optimal combination of prediction and prior information

$$\hat{\omega} = \arg\max_{i}^{N} \beta(\omega, i) \tag{11}$$

$$\beta(\omega, i) = 1[\omega \in \xi_i] \times [IoU_{\text{anchor\_GT}}(\omega, b_i) \times \text{conf}(\omega, i) \times \text{pred\_IoU}(\omega, i)]. \tag{12}$$

The performance comparisons on label assignment separately in the SSDD dataset are shown in Table V. According to the performance of (9) and PPIG, it concludes that anchors' prior information is prone to the anchor containing more target pixels as positive samples and contributes to improving the model's accuracy. Therefore, it is necessary to introduce prior information to label assignments. Besides, comparing the performance

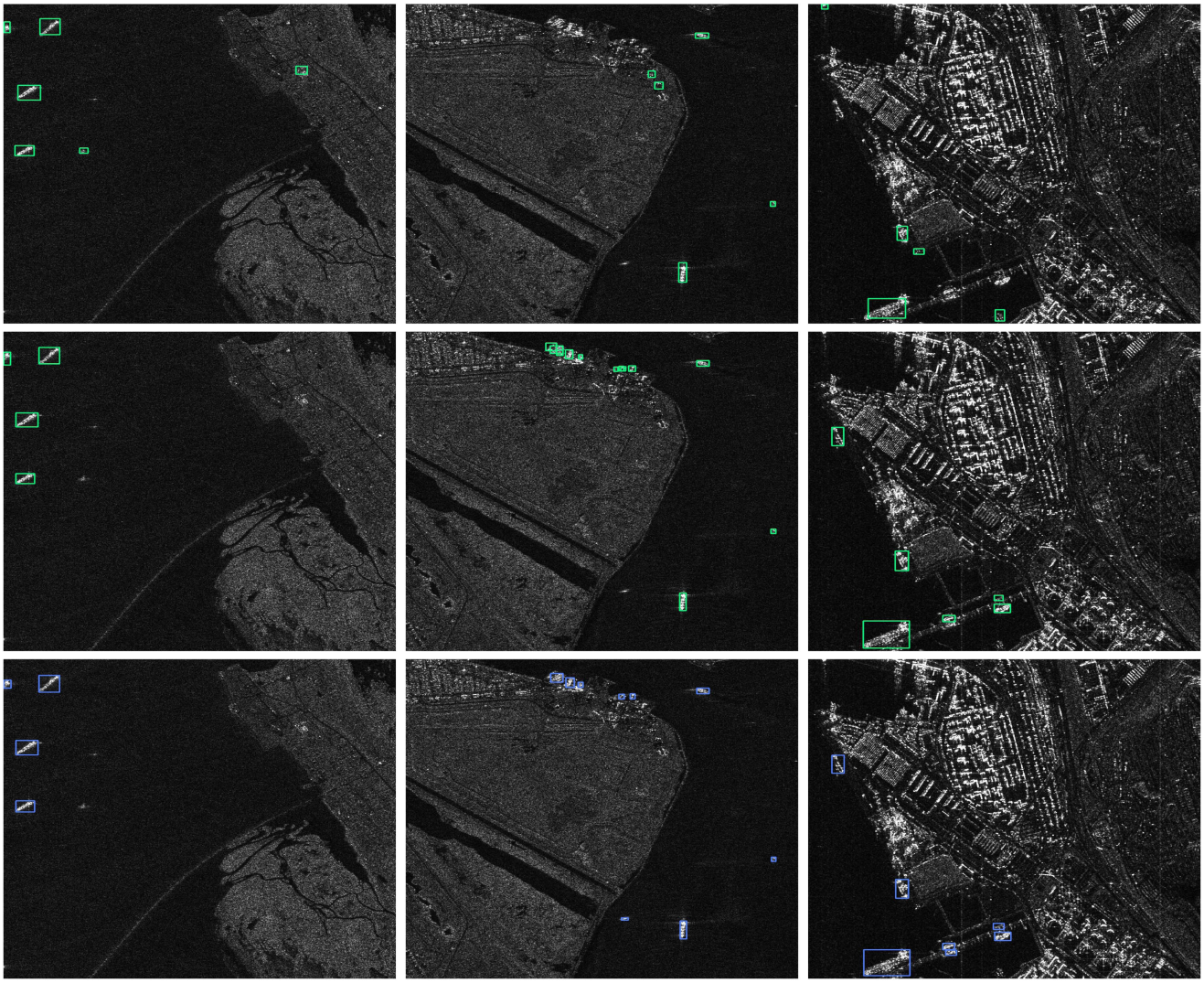


Fig. 10. Visual cases on HRSID dataset. The first row is the detection results of Yolo v4 with PANet. The second is the detection results of Yolo v4 with MLFF. The third row is the ground truth. Rectangles with green color mark the predicted ships. Rectangles with blue color mark the ground truth ships.

TABLE VI
INFLUENCE OF THE PROPOSED METHODS ON HRSID

MSDNet	IoU Prediction	PPIG	AP
			74.91%
✓			82.25%
✓	✓		83.42%
✓	✓	✓	**86.11%**

The bold value is used to highlight the best value.

of (11) with PPIG, prior information is just as crucial as the model's current learning status in assigning positive samples. The results demonstrate that PPIG is the optimal combination of the prediction-aware and anchor prior information.

Finally, ablation experiments of the proposed methods on HRSID are performed to evaluate the robustness and broad applicability. The quantitative comparisons are displayed in Table VI. Compared with the SSDD dataset, MSDNet has a more outstanding performance improvement on the HRSID dataset. MSDNet improves AP by 7.34%. The number of images, the number of pixels per image, and the total number of ships on HRSID are 4–8 times higher than SSDD. As the data increases, designing a specialized detection network can lead to even more significant performance improvement. The PPIG provides 2.69% AP gains. Result comparisons revealed that an appropriate label assignment strategy could improve the detection performance. The deployment of IoU prediction, PPIG, and MSDNet achieves the highest AP. It demonstrates that all the proposed methods are indispensable for improving the detection performance.

C. Comparative Experiments

In order to evaluate the advantage of PPIG comprehensively, state-of-the-art (SOTA) label assignment methods (ATSS [7], o2o [8], POTA [9]) are evaluated in Table VII. These label assignment methods are deployed on MSDNet to control variables and parameters that are not related to label assignment are set the same. All of them follow standard practices, but have limitations. The AP and recall of PPIG are superior to the SOTA

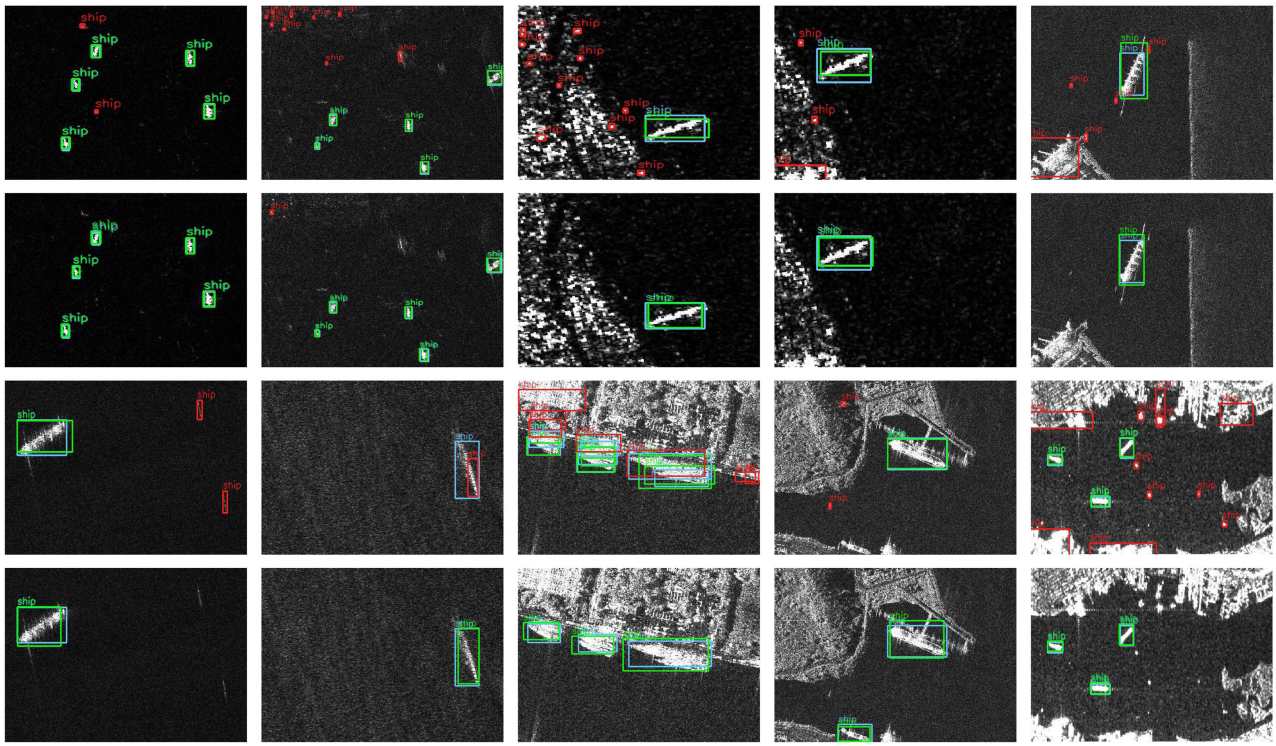


Fig. 11. Visual detection results of Yolo v4 and MSDNet with PPIG on SSDD. The first and the third rows are visual detection results obtained by Yolo v4. The second and fourth rows are visual detection results obtained by MSDNet (PPIG). Rectangles with green color mark the predicted ships. Rectangles with blue color mark the ground truth ships. Rectangles with red color mark the false alarm.

TABLE VII
STUDY RESULTS OF LABEL ASSIGNMENT

Method	AP	Precision	Recall	F1-score
MSDNet	89.41%	76.43%	90.54%	82.89%
ATSS	90.69%	80.16%	90.99%	85.23%
o2o	91.84%	80.86%	93.24%	86.61%
POTA	92.05%	**86.16%**	92.79%	**89.35%**
PPIG	**94.05%**	83.44%	**95.05%**	88.87%

The bold values are used to highlight the best values.

methods ATSS, o2o, and POTA. It demonstrates that the neglect of the prior information of anchors causes a performance decline. Deploying PPIG improves the model's AP by 4.64%, precision by 7.01%, recall by 4.51%, and F1-score by 5.98%. It reveals the potential of assigning the positive samples according to the model's learning status and the prior information of anchors simultaneously.

The performances of the proposed method, faster RCNN, CenterNet, EfficientDet, Yolo v3, Yolo v4, Yolo v5, and SSD-DCN [13] are shown in Table VIII. MSDNet with PPIG is superior in all performance indicators. The improvement is based on a proper label assignment and a network specially designed for SAR ship detection.

Furthermore, to demonstrate the advantage of the proposed methods on different datasets, the comparisons with other state-of-the-art detectors are shown in Table IX [35]. AP of MSDNet with PPIG has achieved 86.11%, which is 10.43% higher than Yolo v5l. The analysis of precision and recall rate concludes

TABLE VIII
DETECTION RESULTS OF CNN-BASED METHODS ON SSDD

Method	AP	Precision	Recall	F1-score
Faster R-CNN (ResNet50)	75.95%	59.68%	83.33%	69.55%
CenterNet (ResNet50)	80.36%	76.13%	83.33%	79.57%
EfficientDet-d0 (EfficientNet)	84.70%	80.00%	86.49%	83.12%
Yolo v3 (DarkNet53)	83.03%	77.59%	85.88%	81.52%
Yolo v4 (CSPDarkNet53)	86.05%	72.87%	91.22%	81.02%
Yolo v5m	91.66%	72.92%	93.51%	81.94%
Yolo v5l	92.08%	81.79%	94.27%	87.59%
SSD-DCN	87.28%	82.77%	88.74%	85.05%
MSDNet	89.41%	76.43%	90.54%	82.89%
MSDNet + PPIG	**94.05%**	**83.44%**	**95.05%**	**88.87%**

The bold values are used to highlight the best values.

TABLE IX
DETECTION RESULTS OF CNN-BASED METHODS ON HRSID

Method	AP	Precision	Recall	F1-score
Faster R-CNN (ResNet50)	43.14%	49.89%	51.47%	50.67%
CenterNet (ResNet50)	75.92%	71.69%	78.50%	74.94%
EfficientDet-d0 (EfficientNet)	72.89%	69.21%	75.43%	72.19%
Yolo v3 (DarkNet53)	76.24%	68.54%	79.81%	73.75%
Yolo v4 (CSPDarkNet53)	76.92%	68.4%	80.02%	73.76%
Yolo v5m	74.87%	80.91%	76.08%	78.42%
Yolo v5l	75.68%	81.81%	76.93%	79.29%
SSD-DCN	81.53%	80.53%	83.13%	81.81%
MSDNet	83.42%	81.54%	84.68%	83.08%
MSDNet(PPIG)	**86.11%**	**86.31%**	**87.22%**	**86.76%**

The bold values are used to highlight the best values.

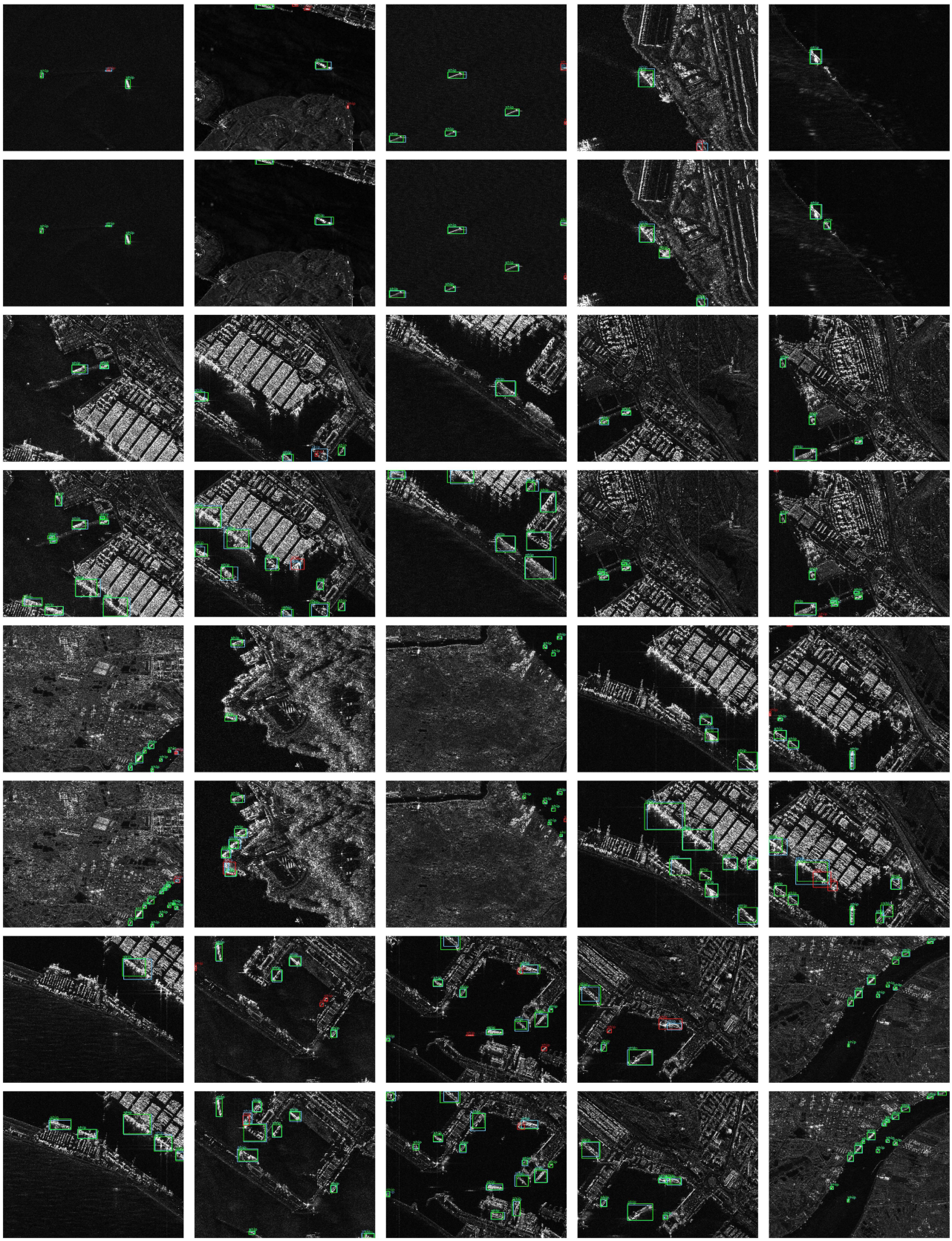


Fig. 12. Visual detection results on HRSID of Yolo v4 and MSDNet with PPIG. The first, third, fifth, and seventh rows are visual detection results obtained by Yolo v4. The second, fourth, sixth, and eighth rows are visual detection results obtained by MSDNet with PPIG. Rectangles with green color mark the predicted ships. Rectangles with blue color mark the ground truth ships. Rectangles with red color mark the false alarm.

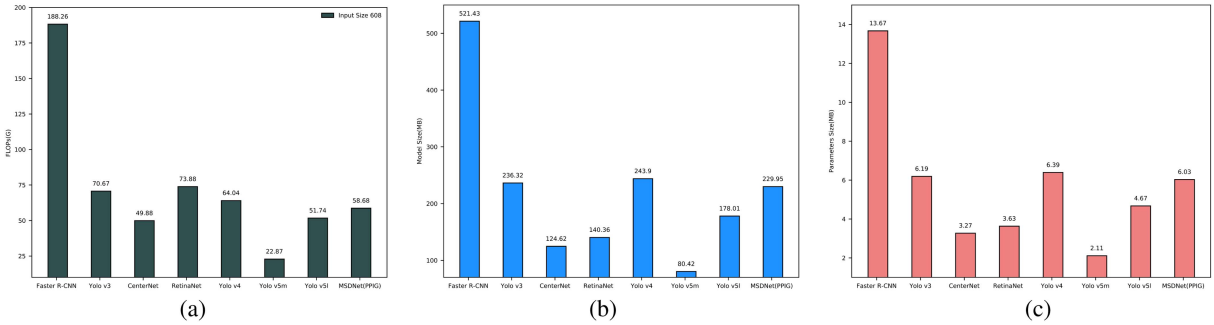


Fig. 13. FLOPs, model size, and parameter size of different detection networks. (a) FLOPs. (b) Model size. (c) Parameter size.

that the proposed method detects more ship instances with lower false alarm rates. Consequently, the much higher AP value of the proposed method confirms the necessity of designing a matched label assignment for SAR ship detection.

Fig. 11 shows some detection results predicted by Yolo v4 and the proposed method on SSDD. The first and second columns are a comparison of the off-shore results and the other columns are a comparison of the in-shore results. It can be seen that the proposed method detects more ships while generating fewer false alarms, especially in complex SAR scenarios. In addition, the proposed method gets more accurate bounding boxes than Yolo v4. The mean IoU between all the correct detections and their corresponding ships is calculated to compare the accuracy of the prediction boxes quantitatively. This article proposes this indicator to measure the quality of the prediction results. As shown in Fig. 14, the proposed method improves the mean IoU by 0.129. It indicates that the proposed method achieves a higher recall rate while maintaining a lower false alarm rate, and the prediction results are also closer to ground truth.

Some detection results on HRSID are presented in Fig. 12. The offshore scenarios as difficult samples are mainly used to compare the performance gap between the two methods. The proposed method has fewer false alarms in the detection results, particularly with islands and land. Meanwhile, the proposed method detected more multiscale ships in offshore scenes. MSDNet with PPIG has excellent detection performance in port and other complex scenarios. Additionally, as for predicting ships correctly, detected boxes predicted by MSDNet with PPIG are closer to ship instances than Yolo v4. As shown in Fig. 14, the proposed method improves the mean IoU between all the correct detections and their corresponding ships by 0.026. The comparisons of detection results demonstrate the advantage of the proposed method.

The computational complexity of the proposed MSDNet with seven CNN-based methods, including Faster R-CNN, Yolo v3, CenterNet, RetinaNet, Yolo v4, Yolo v5 m, and Yolo v5l, are compared in Fig. 13. Generally, the computational complexity of the algorithm is measured by FLOPs, model size, and parameter size. The proposed MSDNet's FLOPs decreases by 5.36 G (about 8.37%) compared to Yolo v4, which saves computing resources. Besides, the proposed method's model size decreases the 13.95 MB compared to Yolo v4, which can better meet the requirements of mobile or embedded devices. The proposed method decreases the 0.36 MB parameter size compared to Yolo v4. MSDNet (PPIG) occupies less memory usage and needs less algorithm initialization time.

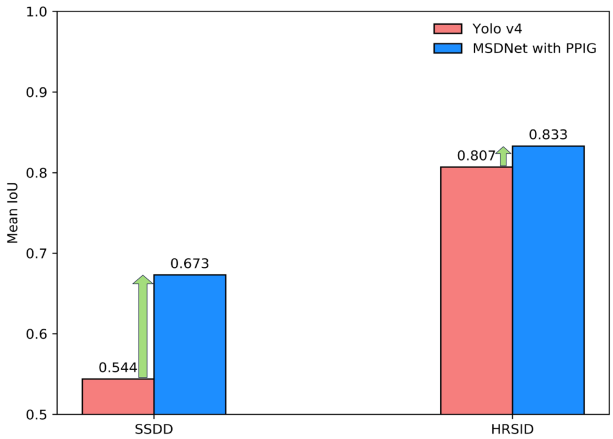


Fig. 14. Comparison of the mean IoU between all the correct detections and their corresponding ships from Yolo v4 and the proposed method.

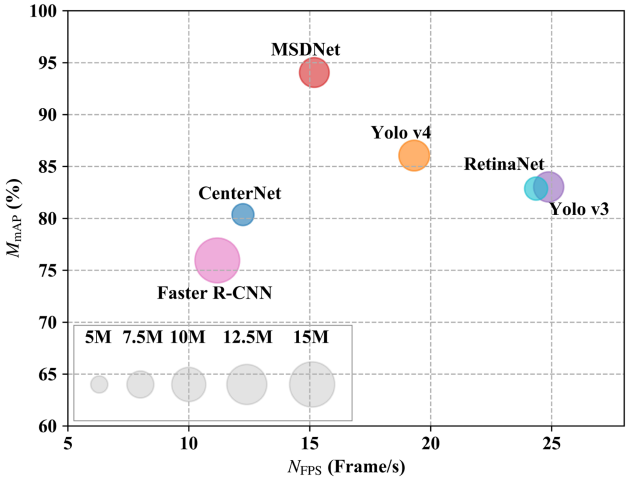


Fig. 15. Comparison between FPS, parameter size, and AP of different detection networks.

The comparison between FPS, parameter size, and AP of the proposed and SOTA methods is shown in Fig. 15. The detection speed of MSDNet (PPIG) is 15.19 FPS, and the AP is 94.05%. Compared with Yolo v4, the inference speed of MSDNet (PPIG) is 4.13 FPS slower, but its detection accuracy is much better (the AP is 8% higher). The significant increase in detection accuracy is worth the slight increase in inference time.

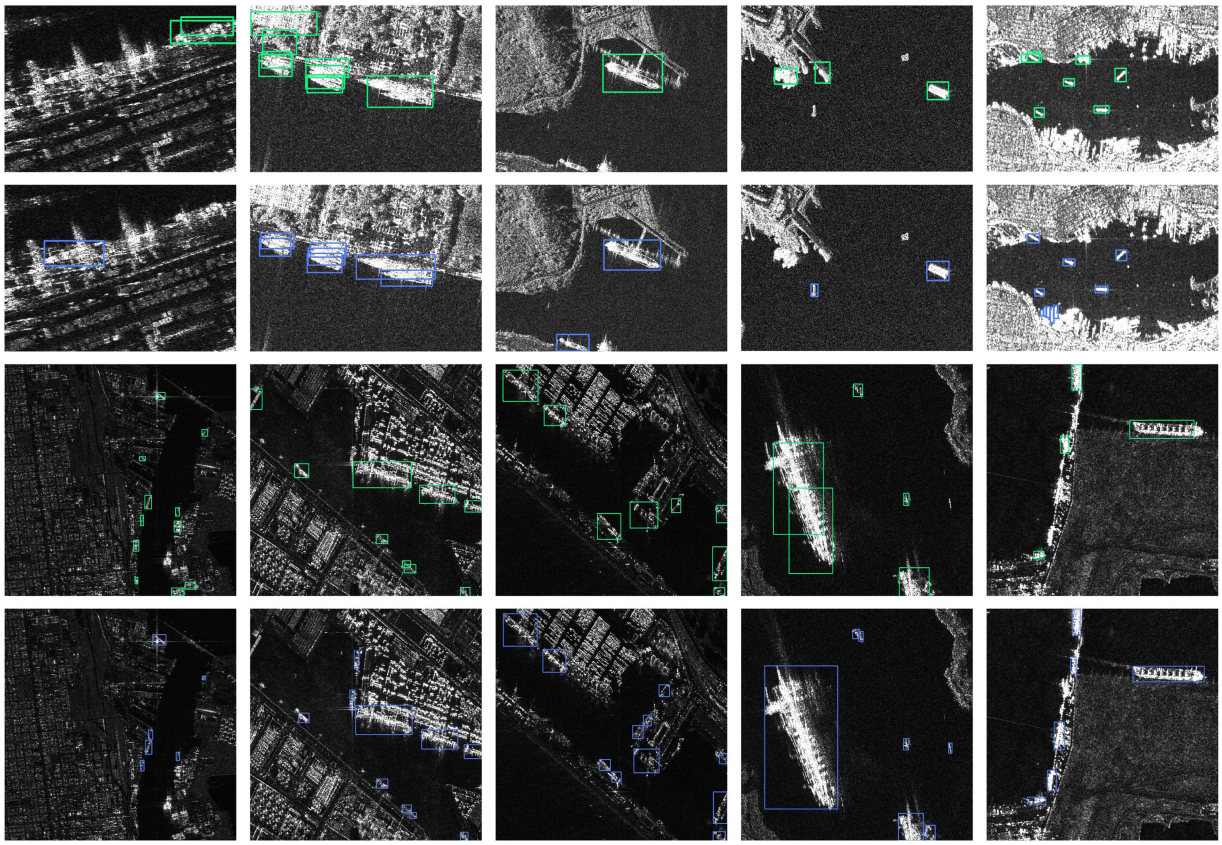


Fig. 16. Visual failure cases on SSDD and HRSID of MSDNet with PPIG. The first and third rows are visual detection results of SSDD and HRSID, respectively. The second and fourth rows are ground truth of SSDD and HRSID, respectively. Rectangles with green color mark the predicted ships. Rectangles with blue color mark the ground truth ships.

Fig. 16 shows some failure cases of the proposed method on the SSDD and HRSID datasets. It can be seen that the most common situation is false positive predictions, especially in coastal areas such as ports. Improving ship detection networks' ability to discriminate between sea and land is the next research point. Besides, the proposed method is challenging to deal with large or tiny targets because the initial anchor size set by the anchor-based algorithm cannot cover all targets in the dataset. Even if the anchor-free algorithm removes the constraint brought by the anchor, it is still challenging for the network to learn the exact width and height of the extreme-size targets. Almost all detection algorithms are faced with the difficulty of predicting objects of extreme size. A single picture of HRSID contains a broader range of scene changes, and the ship scale is more various. Therefore, detecting a huge ship into two objects or missing small ships is more common than SSDD. Although the MLFF alleviates the difficulty of multiscale ship detection, further research is needed to balance huge and tiny ship detection performance.

V. Conclusion

This article proposes PPIG for ship detection and designs a novel detection network refining semantic information of multiscale ships. Specifically, the work has three crucial implications:

1) It reveals that label assignment is a crucial procedure that significantly affects the current SAR ship detector's performance. Researchers should draw more attention to introducing a proper label assignment.
2) Four critical points of devising a proper label assignment are demonstrated. First, the method should be adaptively to discard hyperparameters tuning. Second, the method should reflect the model's current learning status. Third, it should simultaneously evaluate the detected boxes' quality according to classification and localization. Fourth, the anchor prior information should not be ignored, and it is conducive to assigning the optimal anchors as positive samples.
3) Ship detectors should consider multiscale object detection instead of migrating optical models directly to SAR images. Feature enhancement and fusion enable the network to make full use of features on different scales and improve the detector's recall rate.

References

[1] J. Cui, H. Jia, H. Wang, and F. Xu, "A fast threshold neural network for ship detection in large-scene SAR images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 6016–6032, 2022.
[2] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, Art. no. 6517.
[3] 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.
[4] A. Bochkovskiy, C. Y. Wang, and H. Y. M. Liao, "YOLOv4: Optimal speed and Accuracy of object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2021, Art. no. 13024.

[5] K. Kim and H. Lee, "Probabilistic anchor assignment with IoU prediction for object detection," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 355–371.
[6] J. Wang, L. Song, Z. Li, H. Sun, J. Sun, and N. Zheng, "End-to-end object detection with fully convolutional network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2021, Art. no. 15844.
[7] S. Zhang, C. Chi, Y. Yao, Z. Lei, and S. Z. Li, "Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2020, Art. no. 9756.
[8] P. Sun et al., "What makes for end-to-end object detection?," in *Proc. Int. Conf. Mach. Learn.*, 2021, Art. no. 9934.
[9] S. Wu, X. Li, and X. Wang, "IoU-aware single-stage object detector for accurate localization," *Image Vis. Comput.*, vol. 97, 2020, Art. no. 103911.
[10] 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 2019.
[11] B. Li, B. Liu, W. Guo, Z. Zhang, and W. Yu, "Ship size extraction for Sentinel-1 images based on dual-polarization fusion and nonlinear regression: Push error under one pixel," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4887–4905, Aug. 2018.
[12] T. Yue, Y. Zhang, P. Liu, Y. Xu, and C. Yu, "A generating-anchor network for small ship detection in SAR images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 7665–7676, 2022.
[13] Y. Wang, G. Dong, and H. Liu, "Small ship detection via deformable convolutional network," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2021, Art. no. 3537.
[14] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path aggregation network for instance segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, Art. no. 8759.
[15] A. Neubeck and L. Van Gool, "Efficient non-maximum suppression," in *Proc. Int. Conf. Pattern Recognit.*, 2006, pp. 850–855.
[16] 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.
[17] 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.
[18] S. Liu, D. Huang, and Y. Wang, "Learning spatial fusion for single-shot object detection," 2019, *arXiv:1911.09516*.
[19] Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, and D. Ren, "Distance-IoU loss: Faster and better learning for bounding box regression," in *Proc. IEEE Conf. AAAI Artif. Intell.*, 2020, pp. 12993–13000.
[20] 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.
[21] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018. [Online]. Available: https://arxiv.org/abs/1804.02767
[22] K. Duan, S. Bai, L. Xie, H. Qi, Q. Huang, and Q. Tian, "CenterNet: Keypoint triplets for object detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2019, pp. 6568–6577.
[23] 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.
[24] G. H. Robertson, "Computation of the noncentral F distribution (CFAR) detection," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES- 12, no. 5, pp. 568–571, Sep. 1976.
[25] S. Blake, "OS-CFAR theory for multiple targets and nonuniform clutter," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 24, no. 6, pp. 785–790, Nov. 1988.
[26] H. Law and J. Deng, "CornerNet: Detecting objects as paired keypoints," *Int. J. Comput. Vis.*, vol. 128, no. 3, pp. 642–656, Jan. 2020.
[27] Z. Sun et al., "An anchor-free detection method for ship targets in high-resolution SAR images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 7799–7816, 2021.
[28] Q. Hu, S. Hu, and S. Liu, "BANet: A balance attention network for anchor-free ship detection in SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5222212.
[29] T. Zhang and X. Zhang, "ShipDeNet-20: An only 20 convolution layers and $<$ 1-MB lightweight SAR ship detector," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 7, pp. 1234–1238, Jul. 2021.
[30] X. Yang, X. Zhang, N. Wang, and X. Gao, "A robust one-stage detector for multiscale ship detection with complex background in massive SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5217712.
[31] 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.
[32] 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, 2020.
[33] 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.
[34] 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.
[35] M. Tan, R. Pang, and Q. V. Le, "EfficientDet: Scalable and efficient object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 10781–10790.
[36] T. Zhang et al., "SAR ship detection dataset (SSDD): Official release and comprehensive data analysis," *Remote Sens.*, vol. 13, no. 18, 2021, Art. no. 3690.
[37] S. Wei, X. Zeng, Q. Qu, M. Wang, H. Su, and J. Shi, "HRSID: High-resolution SAR images dataset for ship detection and instance segmentation," *IEEE Access*, vol. 8, pp. 120234–120254, 2020.
[38] G. Jocher, "YOLOv5," 2020. [Online]. Available: https://github.com/ultralytics/yolov5
[39] B. Jiang, R. Luo, J. Mao, T. Xiao, and Y. Jiang, "Acquisition of localization confidence for accurate object detection," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 784–799.

Yao Wang received the B.S. degree in electronic and information engineering from Xidian University, Xi'an, China, in 2019. He is currently working toward the Ph.D. degree in signal processing with the National Key Laboratory of Radar Signal Processing, X'ian, China.

His research interests include object detection, computer vision, and deep learning.

Ganggang Dong received the M.S. and Ph.D. degrees in information and communication engineering from the National University of Defense Technology (NUDT), Changsha, China, in 2012 and 2016, respectively.

He is currently an Associate Professor with Xidian University. His research interests include, but not limited to deep learning, SAR imaging, radar target detection and recognition, cognitive radio, image interpretation. He has authored more than 40 scientific papers in peer-reviewed journals and conferences, including IEEE TIP, TGRS, TIM, JSTARS, GRSL, SPL, SPIE JARS, Pattern Recognition, and IGARSS. He received more than 1200 citations in Google Scholar.

Dr. Dong was awarded the 2017 Excellent Doctoral Thesis of the Chinese Institute of Electronics.

Shuai Shao received the Ph.D. degree in signal and information processing from Xidian University, Xi'an, China, in 2020.

He is an Associate Professor with the National Key Laboratory of Radar Signal Processing, Xidian University. His research interests include radar signal processing, inverse synthetic aperture radar imaging, and target recognition.

Hongwei Liu received M.S. and Ph.D. degrees in electronic engineering from Xidian University, Xi'an, China, in 1995 and 1999, respectively.

He worked with the National Key Laboratory of Radar Signal Processing, Xidian University. From 2001 to 2002, he was a Visiting Scholar with the Department of Electrical and Computer Engineering, Duke University, Durham, NC, USA. He is a Professor with the National Key Laboratory of Radar Signal Processing, Xidian University. His research interests include radar automatic target recognition, radar signal processing, and adaptive signal processing.