remote sensing

MDPI

Communication

A Novel Deep Learning Network with Deformable Convolution and Attention Mechanisms for Complex Scenes Ship Detection in SAR Images

Peng Chen [1], Hui Zhou [2,*], Ying Li [3], Peng Liu [1] and Bingxin Liu [1]

[1] Navigation College, Dalian Maritime University, Dalian 116026, China; chenpeng@dlmu.edu.cn (P.C.); liupeng@dlmu.edu.cn (P.L.); gisbingxin@dlmu.edu.cn (B.L.)
[2] School of Computer and Software, Dalian Neusoft Information University, Dalian 116023, China
[3] Environmental Information Institute, Dalian Maritime University, Dalian 116026, China; yldmu@dlmu.edu.cn
* Correspondence: zhouhui@neusoft.edu.cn; Tel.: +86-411-8483-2287

Abstract: Synthetic aperture radar (SAR) can detect objects in various climate and weather conditions. Therefore, SAR images are widely used for maritime object detection in applications such as maritime transportation safety and fishery law enforcement. However, nearshore ship targets in SAR images are often affected by background clutter, resulting in a low detection rate, high false alarm rate, and high missed detection rate, especially for small-scale ship targets. To address this problem, in this paper, we propose a novel deep learning network with deformable convolution and attention mechanisms to improve the Feature Pyramid Network (FPN) model for nearshore ship target detection in SAR images with complex backgrounds. The proposed model uses a deformable convolutional neural network in the feature extraction network to adapt the convolution position to the target sampling point, enhancing the feature extraction ability of the target, and improving the detection rate of the ship target against the complex background. Moreover, this model uses a channel attention mechanism to capture the feature dependencies between different channel graphs in the feature extraction network and reduce the false detection rate. The designed experiments on a public SAR image ship dataset show that our model achieves 87.9% detection accuracy for complex scenes and 95.1% detection accuracy for small-scale ship targets. A quantitative comparison of the proposed model with several classical and recently developed deep learning models on the same SAR images dataset demonstrated the superior performance of the proposed method over other models.

Keywords: SAR image; ship target detection; deformable convolutional networks; channel attention

Citation: Chen, P.; Zhou, H.; Li, Y.; Liu, P.; Liu, B. A Novel Deep Learning Network with Deformable Convolution and Attention Mechanisms for Complex Scenes Ship Detection in SAR Images. *Remote Sens.* **2023**, *15*, 2589. https://doi.org/10.3390/rs15102589

Academic Editors: Angelica Lo Duca, Emanuele Salerno and Claudio Di Paola

Received: 3 April 2023
Revised: 5 May 2023
Accepted: 12 May 2023
Published: 16 May 2023

1. Introduction

The continuous monitoring of ship targets in harbors and marine areas is an important practical task that is widely used in various maritime fields such as combating illegal fishing, oil spill monitoring, and traffic management [1,2]. Synthetic aperture radar (SAR) has become an important method for ship detection at sea because it is unaffected by weather, has a large imaging area, and has a constant resolution far from the observed object [3–5]. However, accurate ship detection remains challenging. Currently, several detection methods can easily miss ships that are in close proximity to each other.

Numerous manually extracted features have been used for ship detection [6,7]. Recently, deep learning models have been widely used for ship detection in SAR images [8–10]. Ship detection is a combination of target localization and classification. Target classification determines whether the input image contains the desired object, and target location identifies the location of the target [11–13]. For example, Wang et al. used SDD models to detect ships in complex backgrounds in SAR images and used alternate learning methods to improve the accuracy [14]. Kang used a fast region with a convolutional neural net-

work (CNN) to obtain initial ship detection results and adjusted the final results using an adaptive threshold alarm rate [15].

In the SAR image ship target-detection task, target detection is better in a simple sea clutter background, because the grayscale characteristics of the target are significantly higher than those of the sea clutter [16–18]. However, in actual SAR image imaging processes, backgrounds such as ports, islands, and buildings appear in SAR images, which can cause confusion because of their high grayscale characteristics, resulting in low detection and high false alarm rates for deep learning target-detection algorithms [19]. The complex backgrounds mentioned in this study refer to SAR images with backgrounds, such as ports and islands, for ship targets in complex scenes. For the target-detection problem of complex scene interference, some scholars solve this problem from the perspective of improving the backbone network [20,21]. Wu et al. proposed a SAR image ship small target-detection algorithm to improve the network structure of the feature pyramid by redesigning the underlying residual units to solve the contradiction between the perceptual field and localization and introduced a balance factor to optimize the small target weights in the loss function [22]. Zhang et al. proposed a feature-fusion-based ship target-detection algorithm based on a multiscale single-shot detection framework with enhanced network feature extraction by adding deconvolution and pooling feature fusion modules [23]. He et al. proposed the Deformable Feature Fusion You Only Look Once (DFF-Yolov5) algorithm based on YOLOv5, which improved the YoloV5 model in two aspects: feature refinement and multifeature fusion in the feature extraction network [24]. Other scholars consider introducing attention mechanisms to solve the above problems [25,26]. Liu et al. improved the target-detection method by integrating the detection frame length and width as parameters, performing curve optimization of the loss function, and combining it with the coordinate attention mechanism to detect ship targets [27]. Li et al. used multiple receptive field integration and channel domain attention to enhance the resistance of features to scale and environmental changes [28]. SAR ship detection can also occur in scenes with numerous ships, high density, and a small target size (less than 15 pixels $\times$ 15 pixels), which further increases the difficulty of target extraction and recognition. When using a deep learning model of ship targets for detection, the CNN can only extract relatively regular features in the target area. For SAR ship targets with complex backgrounds, standard CNNs are susceptible to interference from background coastal information when extracting features, which affects the expressiveness of the feature extraction network [14,29]. However, because SAR image ship targets have different scale information, the extraction of the semantic information of small targets is comparatively lower with an increase in the number of feature-mapping layers when feature extraction is performed by convolution. Therefore, the above method still has limitations for nearshore ship targets in complex backgrounds.

To counter the challenges of low detection rates of ship targets and small target detection in SAR images in complex scenes, classical Feature Pyramid Networks (FPNs) have been introduced by multiple authors to implement multiscale SAR ship detection [30]. For example, Lin et al. added a compressed incentive-based module to the top of an FPN to focus on more important channel features, thus enhancing the description of the top-level semantic information [31]. Li et al. embedded a convolutional block attention module into the feature fusion branch of an FPN, which enables a joint feature channel and spatial attention to facilitate the representation of multilevel ship features [32]. Zhao et al. proposed a field-of-view attentional FPN to enhance background discrimination by extending the field-of-view range of the contextual background [33]. Liu et al. designed a scale-migratable FPN to further enhance the fusion benefits of multiscale features and ultimately improve detection accuracy [34]. Recently, Li et al. proposed a feature self-attention-guided FPN to refine more representative multiscale features, thereby facilitating multiscale information flow [35]. However, the multiscale SAR ship detection performance of these existing detection models remains limited. On the one hand, most of them use classical regular convolution kernels to extract ship features, which cannot perform multiscale

modeling of ship deformations owing to different incidence angles, resolutions, and other factors, resulting in limited multiscale feature expression capability. On the other hand, the multiscale feature fusion approach they adopt cannot provide global content, and the multiscale feature fusion methods they use cannot effectively perceive the global content, which leads to limited fusion benefits and is not conducive to a more comprehensive and adequate multiscale feature representation. This study made structural improvements to the FPN. First, a deformable CNN was used to replace the original CNN to improve the expression ability of the network features. Next, a target-detection network was used to extract features, and a channel attention mechanism was introduced to extract the weights of the feature channels to further enhance target feature extraction in complex backgrounds. Finally, through multiple sets of comparison experiments and an analysis of the detection results of high-density small target ship detection and SAR image ship target detection in complex environments, the practicality and effectiveness of the proposed algorithm for the detection of ship targets in complex backgrounds of SAR images were verified.

2. Materials and Methods

2.1. Revisiting FPN

Because of the difference in imaging mechanisms between SAR and optical images, distinguishing ship targets from maritime false targets and background noise in SAR images is difficult and often confusing [36]. The use of an FPN can effectively eliminate the background environment and scattering noise, thereby improving the recognition accuracy of ship targets in SAR images. First, feature mapping was obtained by convolution using the backbone network, where the convolution layers formed feature-mapping layers from bottom to top in the following order: {C1, C2, C3, C4, C5}, and then by undergoing upsampling through the top-down pathway to obtain feature maps from higher pyramid layers. {C1, C2, C3, C4, C5} are laterally connected with the upsampling results through a 1×1 convolution kernel (256 channels) to form new feature maps {M2, M3, M4, M5}.

$$\mathrm{M5} = \mathrm{C5}.\mathit{Conv}\ (256, (1, 1)),$$

$$\mathrm{M4} = \mathit{UpSampling}\ (\mathrm{M5}) + \mathrm{C4}.\mathit{Conv}\ (256, (1, 1)),$$

$$\mathrm{M3} = \mathit{UpSampling}\ (\mathrm{M4}) + \mathrm{C3}.\mathit{Conv}\ (256, (1, 1)),$$

$$\mathrm{M2} = \mathit{UpSampling}\ (\mathrm{M3}) + \mathrm{C2}.\mathit{Conv}\ (256, (1, 1)).$$

Finally, to remove the confusion effect, feature maps from m2 to m5 were obtained using a 3×3 convolution, and a sequence of FPN feature maps {P2, P3, P4, P5} was obtained. The FPN structure is illustrated in Figure 1 [37].

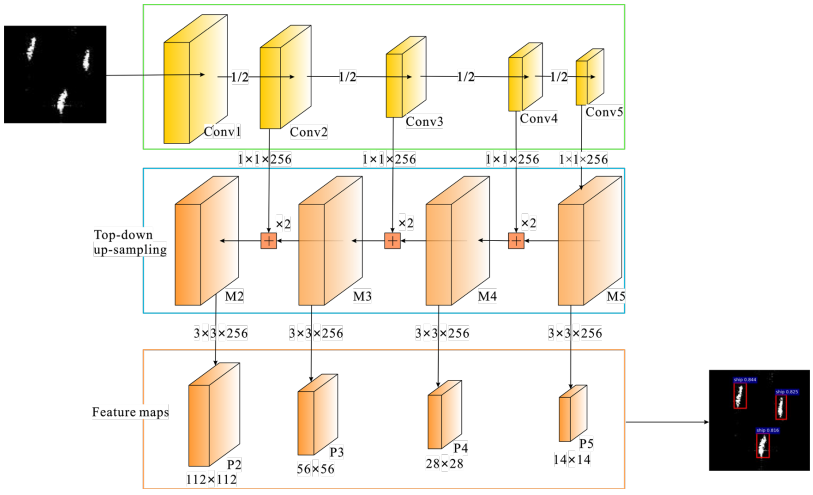


Figure 1. Structure of the Feature Pyramid Network (FPN) network.

2.2. Improved Structure of FPN with Deformable CNNs

The backbone network of the original FPN target-detection algorithm used a CNN for feature extraction. Standard convolution is not sufficiently flexible for the shape-receptive field of the target, and the efficiency of convolution naturally decreases, whereas deformable convolution uses irregular shapes, which addresses this problem. Compared with the standard convolution kernel, the pixels used for convolution by the deformable convolution kernel are not shifted by a fixed step size in the x- and y-directions relative to the central pixel; however, a new convolution kernel is used to record x and the offset in the y-direction. Figure 2 shows the learning process for the deformable convolution. After the traditional convolution layer, a biased convolution layer was added. The convolution kernel of this layer was identical to that of an ordinary convolution kernel. The output deviation was the same as that of the input feature map. The generated channel dimension is 2N, where N is the number of channels of the input feature map, which implies that the x-direction was recorded twice separately, as well as the offset features in the y-direction.

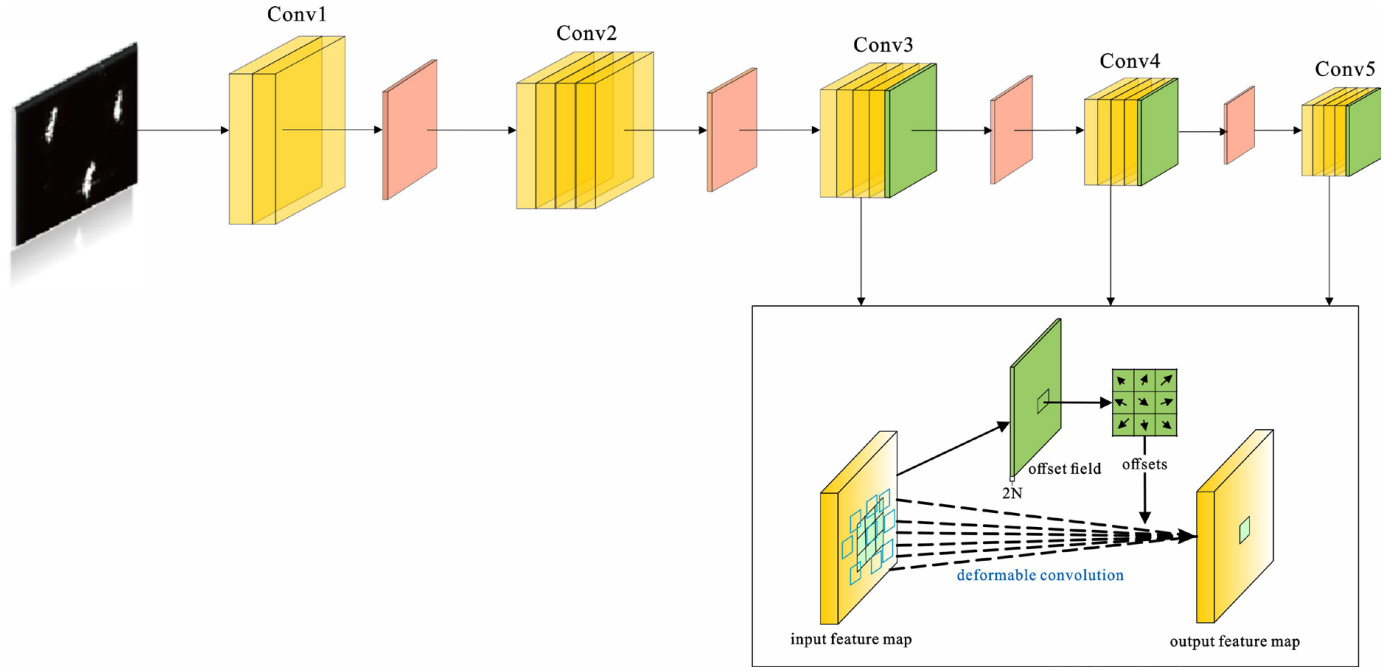


Figure 2. Deformable convolutional neural network (CNN).

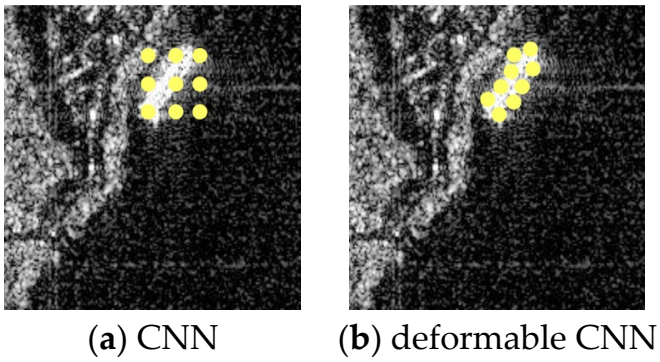
To improve the FPN target-detection model, {C1, C2, C3, C4, C5} feature maps were first generated based on the input images using conventional convolution kernels. To avoid redundant computations, deformable convolution kernels were used after the convolution layers conv3, conv4, and conv5 of the FPN; that is, the {C3, C4, C5} feature maps were used as the input, and another convolution layer was added to each feature layer to learn the deformation offsets in the x- and y-directions of the deformable convolution. During training, the convolutional kernels for generating output features and offsets were simultaneously learned using an interpolation algorithm and backpropagation [38]. Specifically, for the initial position p_0 of the input feature map, the output was obtained as y after an ordinary convolution operation.

$$y(p_0) = \sum_{p_n \in \Re} w(p_n) \cdot x(p_0 + p_n) \quad (1)$$

where w is the network weight parameter, $\Re$ is the specified convolution region, and p_n traverses this convolution region by adding a learning position offset Δp_n to the ordinary convolution y. The deformed convolution is formulated as follows:

$$y(p_0) = \sum_{p_n \in \Re} w(p_n) \cdot x(p_0 + p_n + \Delta p_n) \quad (2)$$

A deformable convolutional neural network can extract the features of an input image more effectively by offsetting the convolutional kernel and adjusting its shape according to the actual situation. As shown in Figure 3, the use of deformable CNNs can improve the feature representation capability of SAR image target-detection networks.



(**a**) CNN (**b**) deformable CNN

Figure 3. Compared with the CNN (**a**), the deformable CNN (**b**) extracts the features of the input image by adjusting its shape according to the actual situation by shifting the convolutional kernel.

2.3. Channel Attention Mechanism Introduced

In the complex backgrounds of SAR images, distinguishing nearshore ship targets is difficult because of increased interference. To improve the expression ability of features in images, in the target-detection model FPN, for feature-mapping layers of different scales, the channel attention mechanism is used to capture different channel maps, compute the feature dependencies between them, and calculate the weighted values of all the channel maps. The feature weight vector was learned to explicitly model the correlation between the feature channels. To compute the channel attention efficiently, we squeeze the spatial dimension of the input feature map. For aggregating spatial information, average pooling and max pooling have been commonly adopted to compute spatial statistics. as shown in Figure 4. The dot product of the original feature layer (F) of any H × W × C and the feature weight vector **w** were determined to obtain the feature layers with different levels of importance for different channels. The layers with channel weights were then merged at each layer in the FPN manner; that is, a new feature map layer is obtained, as shown in Figure 5.

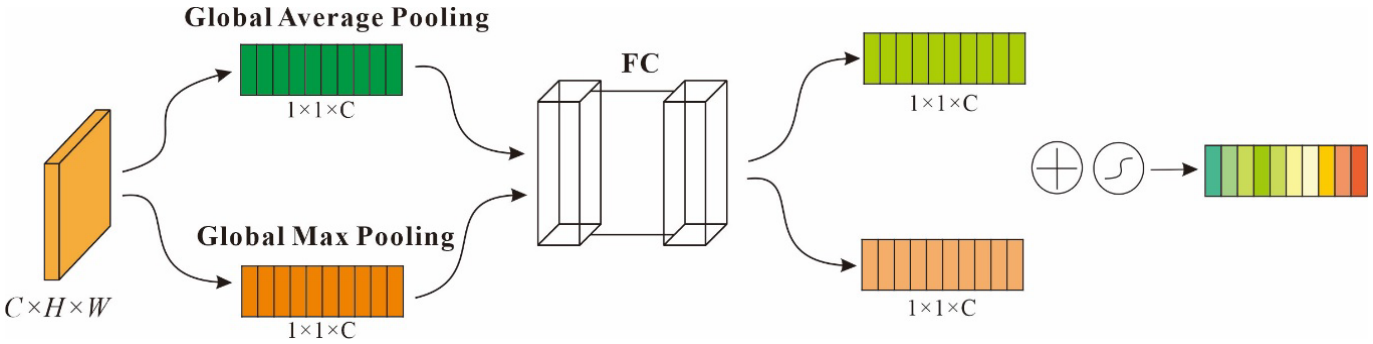


Figure 4. Channel attention mechanism.

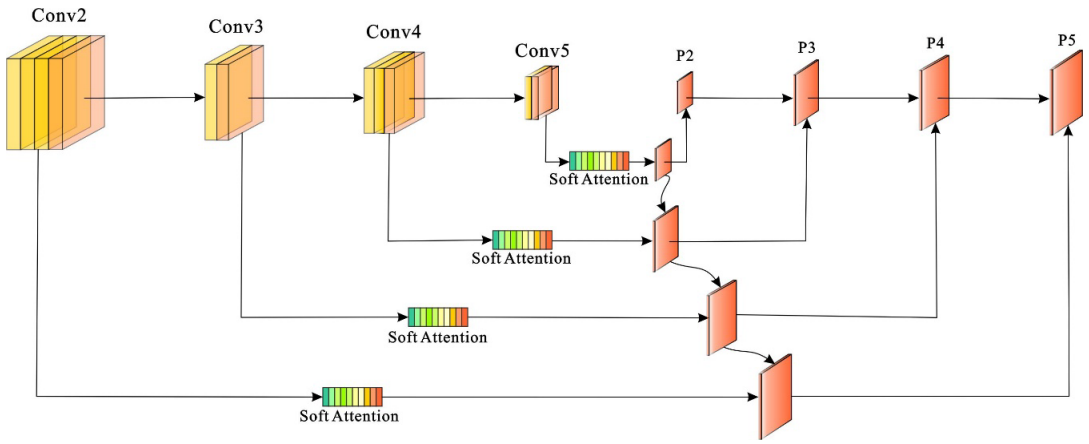


Figure 5. Introducing the soft attention mechanism to FPN. The feature weight vector was used in each convolution layer, which obtained different levels of importance for different channels.

First, using the feature layer (F) of any H × W × C as the input, the global average pooling *AvgPool* and the maximum pooling *MaxPool* of the space were conducted, where the pool size was H×W, and the channel description row vectors $\mathbf{F_{avg}}$ and $\mathbf{F_{max}}$ of the two 1 × 1 × C were obtained. Two fully connected layers were shared, and the ReLU activation function was used to fit the complex correlations between the channels. Then, we added the description row vectors of the two channels and obtained the feature weight vector **w** of 1 × 1 × C using the sigmoid activation function. The original feature layer (H × W × C) and feature weight vector **w** were multiplied to obtain feature layers with different channel importance values, as shown in Figures 4 and 5, respectively. The region of interest was determined using a sliding window operation on the reconstructed feature map.

$$\begin{aligned} \mathbf{w} &= sigmoid(TFC(\mathbf{F_{avg}}) + TFC(\mathbf{F_{max}})), \\ \mathbf{F_{avg}} &= \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} AvgPool(\mathbf{F}), \\ \mathbf{F_{max}} &= \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} MaxPool(\mathbf{F}). \end{aligned} \tag{3}$$

2.4. Loss Function

The overall loss of the model includes classification loss L_{class} and bounding box loss L_{box}, the classification loss adopts the cross-entropy loss function, and the bounding box loss includes the location loss L_1 the of ground truth b_i and predicted box $\hat{b}_{\sigma(i)}$ and Intersection over Union (IoU) loss L_{GIoU}.

$$L_{class} = -\frac{1}{N} \sum_{i=1}^{N} (p_i \log \hat{p}_{\hat{\sigma}i} + (1 - p_i) \log(1 - \hat{p}_{\hat{\sigma}i})), \tag{4}$$

$$L_{box}\left(y_i, \hat{y}_{\sigma(i)}\right) = \sum_{1}^{N} \left[\lambda_1 \|b_i - \hat{b}_{\sigma(i)}\| + \lambda_{GIoU} L_{GIoU}\left(b_i, \hat{b}_{\sigma(i)}\right)\right], \tag{5}$$

where N is the number of prediction frames, p_i is the true category probability, and $\hat{p}_{\hat{\sigma}i}$ represents the probability of predicting the $\hat{\sigma}_i$ th ship target. λ_1 and λ_{GIoU} are the corresponding loss function penalty factors.

IoU reflects the degree of coincidence between the prediction box and the ground truth, and the larger the coincidence, the greater the value; therefore, it works as an optimization function. GIoU introduces the minimum enclosure box and can avoid the problem that, when the prediction box and ground truth do not overlap, the gradient is 0, and the model cannot be optimized. Given the coordinates of the ground truth are **gt**, and the calculated predicted box coordinates are **pb**, IoU can be obtained through calculations.

$$\begin{array}{l} \mathbf{pb} = (x^p_{min}, x^p_{max}, y^p_{min}, y^p_{max}), \\ \mathbf{gt} = (x^g_{min}, x^g_{max}, y^g_{min}, y^g_{max}), \end{array} \tag{6}$$

$$\begin{array}{l} A_p = (x^p_{max} - x^p_{min}) \times (y^p_{max} - y^p_{min}), \\ A_g = (x^g_{max} - x^g_{min}) \times (y^g_{max} - y^g_{min}), \end{array} \tag{7}$$

$$\begin{array}{l} I_{pg} = \begin{cases} (x^I_2 - x^I_1) \times (y^I_2 - y^I_1), \text{ if } x^I_2 > x^I_1, y^I_2 > y^I_1 \\ 0, \text{ otherwise} \end{cases}, \\ U_{pg} = A_p + A_g - I_{pg}, \end{array} \tag{8}$$

$$\begin{array}{l} x^c_{min} = min\left(x^p_{min}, x^g_{min}\right), x^c_{max} = max\left(x^p_{max}, x^g_{max}\right), \\ y^c_{min} = min\left(y^p_{min}, y^g_{min}\right), y^c_{max} = max\left(y^p_{max}, y^g_{max}\right), \end{array} \tag{9}$$

$$A_c = (x^c_{\max} - x^c_{\min}) \times (y^c_{\max} - y^c_{\min}), \tag{10}$$

$$IoU = \frac{I_{pg}}{U_{pg}}, GIoU = IoU - \frac{|A_c - U_{pg}|}{|A_c|}, \tag{11}$$

$$L_{GIoU} = 1 - GIoU, \tag{12}$$

where $x^I_1 = max\left(x^p_{min'} x^g_{\min}\right)$, $x^I_2 = min(x^p_{min}, x^g_{min})$, $y^I_1 = max\left(y^p_{min}, y^g_{min}\right)$, $y^I_2 = min\left(y^p_{min}, y^g_{min}\right)$, I_{pg} is the intersection of the predicted box and ground truth, and U_{pg} is the union of the predicted box and ground truth.

3. Results and Discussion

3.1. Implement Details

3.1.1. Dataset

The SAR marine dataset is currently the largest SAR dataset available for multiscale ship detection and was constructed by Wang et al. [39] and labeled by SAR experts. It entails 102 Chinese Gaofen-3 images and 108 Sentinel-1 images. The dataset comprises 43,819 ship chips with a resolution of 256 pixels in both range and azimuth. For Gaofen-3, the image modes included Ultrafine Strip-Map (UFS), Fine Strip-Map 1 (FSI), Full Polarization 1 (QPSI), Full Polarization 2 (QPSII), and Fine Strip-Map 2 (FSII). Furthermore, the resolution of the SAR images in the dataset ranged from 3 m to 10 m. For Sentinel-1, the imaging modes were S3 Strip-Map (SM), S6 SM, and IW-mode. The details of these images, including their resolutions, incidence angles, and polarizations, are summarized in Table 1. Ship objects have distinct scales and backgrounds. Furthermore, some ships were present in complex scenes, which were divided into three categories: offshore, island, and harbor, as shown in Figure 6. SAR ship detection can also occur in scenes with numerous ships, high density, and a small target size (less than 15 pixels × 15 pixels; Figure 6). An overview of the dataset is presented in Table 2. The training, verification, and testing sets constitute 70%, 20%, and 10% of the dataset, respectively.

Table 1. Detailed dataset information for original SAR imagery.

Sensor	Polarization	Imaging Mode	Resolution Rg. × Az. (m)
GF-3	Single	UFS	3 × 3
	Dual	FSI	5 × 5
	Full	QPSI	8 × 8
	Dual	FSII	10 × 10
	Full	QPSII	25 × 25
Sentinel-1 SLC	Dual	SM	1.7 × 4.3~ 3.6 × 4.9

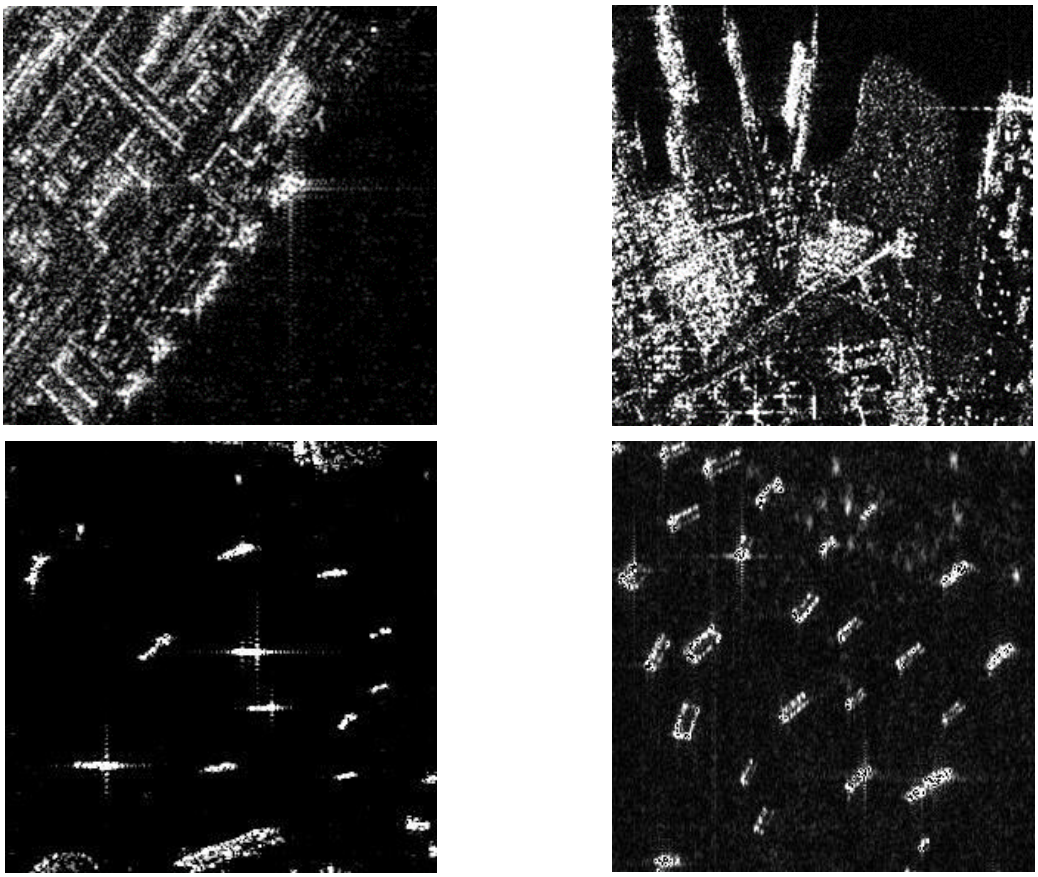


Figure 6. Complex scenes and high-density small target ships in SAR images.

Table 2. Overview of the dataset.

DataType	Image	Complex Scenes	High-Density Small Target Scenes
Training	21,420	12,840	8580
Verification	6120	3660	2460
Testing	3060	1836	1224

3.1.2. Evaluation Metrics

For detection evaluation, the evaluation indicators corresponding to each algorithm are calculated and categorized into the precision, recall rate, and F1 Score defined as follows:

$$\begin{aligned} &Precision = N_{TD}/N, \\ &Recall = N_{TD}/N_{GT}, \\ &F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall}, \end{aligned} \tag{13}$$

where N_{TD} is the number of ship targets detected correctly (true detection, TD), N_{GT} is the actual number of ship targets (ground truth, GT), and N is the total number of ship targets detected. The average precision (AP) with the adaptive IoU threshold was used as a metric based on the precision-recall (PR) curves, and AP is defined as

$$AP = \int_0^1 P(R)dR, \tag{14}$$

Different IoU thresholds can be used to calculate different numbers of interest and detect different N_{TD}. Each IoU threshold corresponds to an AP value, and mAP denotes its mean, which assesses the detection effect of the model, where n denotes the number of

scene categories. Overall, the *mAP* and precision-recall curves were employed to evaluate the proposed method.

$$mAP = \frac{\sum_{i=1}^{n} APi}{n}. \quad (15)$$

3.1.3. Implementation Details

The experiment platform was Ubuntu16.0, the GPU was NVIDIA Tesla V100, and the development platform was Paddle X. During the experiment, the empirical learning rate of the semantic model was 0.0001, the batch size was 24, and the dataset was randomly arranged in each iteration.

3.2. Experimental Process

Experiments verified the effects of the improvements and optimizations on various parts of the FPN structure, as shown in Tables 3 and 4. Since the channel attention mechanism enhances the model feature traction capability, the improved model with this mechanism introduced alone improves the mAP by 2.1% in complex scenes and 6.1% in high-density small target scenes over the original model. Similarly, after introducing the deformable convolutional network alone, the mAP improves by 7.7% in complex scenes and 9.5% in high-density small target scenes compared to the original model. This is because the complex background information blurs the ship target location information, and also the localization information of small targets is not obvious after multi-layer convolution, so it is important to enhance the location and feature information using variable convolution. In the final step, the two improvements were combined, and the regression loss function was replaced with a new GIoU loss function. As a result, the mAP increases to 87.9% in complex scenarios and 95.1% in high-density small target scenarios. The effectiveness of the proposed method even in complex scenarios is demonstrated.

Table 3. Comparison of detection results in complex scenes.

FPN	Channel Attention	Deformable CNN	Improved Loss Function	SAR Ships in Complex Scenes			
				Precision (%)	Recall (%)	F1 Score	mAP (%)
√				83.4	71.3	0768.	79.4
√	√			85.5	73.2	0.789	81.5
√		√		89.1	78.1	0.833	87.1
√	√	√		91.2	78.1	0.841	87.9
√	√	√	√	91.7	78.1	0.844	87.9

The check mark "√" indicates that the technique was used in training.

Table 4. Comparison of detection results in high-density small target scenes.

FPN	Channel Attention	Deformable CNN	Improved Loss Function	High-Density Small Target Scenes			
				Precision (%)	Recall (%)	F1 Score	mAP (%)
√				87.7	75.3	0.810	83.8
√	√			89.9	75.9	0.823	89.9
√		√		95.3	85.4	0.900	93.3
√	√	√		96.2	92.8	0.946	95.1
√	√	√	√	96.5	93.0	0.947	95.1

The check mark "√" indicates that the technique was used in training.

The proposed model is compared with FPNs incorporating different mechanisms, following their training protocols, using the reported default parameters and iterations on each benchmark. Figure 7 shows the precision-recall curves of different models on the SAR dataset in different scenarios.

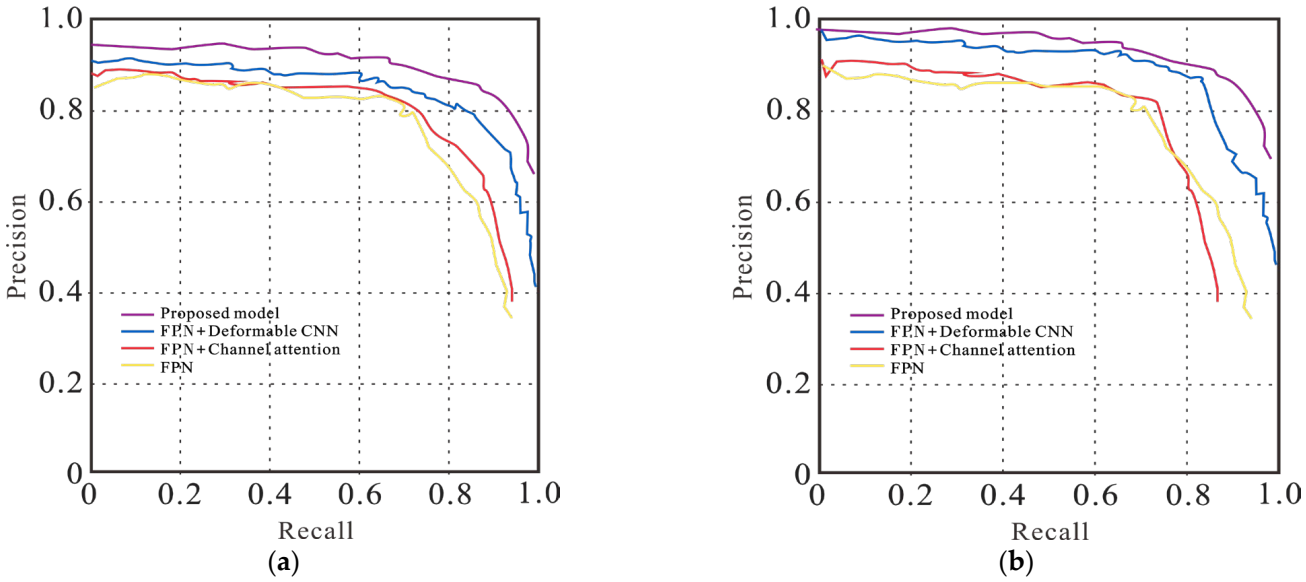


Figure 7. Precision-recall (PR) curves for SAR ships in (**a**) complex and (**b**) high-density small target scenes.

The results in Figures 8 and 9 show that the proposed model can detect multiscale ship objects in various scenes. As shown in Figure 8, the proposed model can detect the ship when it is located in a complex scene. Moreover, as shown in Figure 9, the model can produce accurate detection results even when multiple dense ship targets are present.

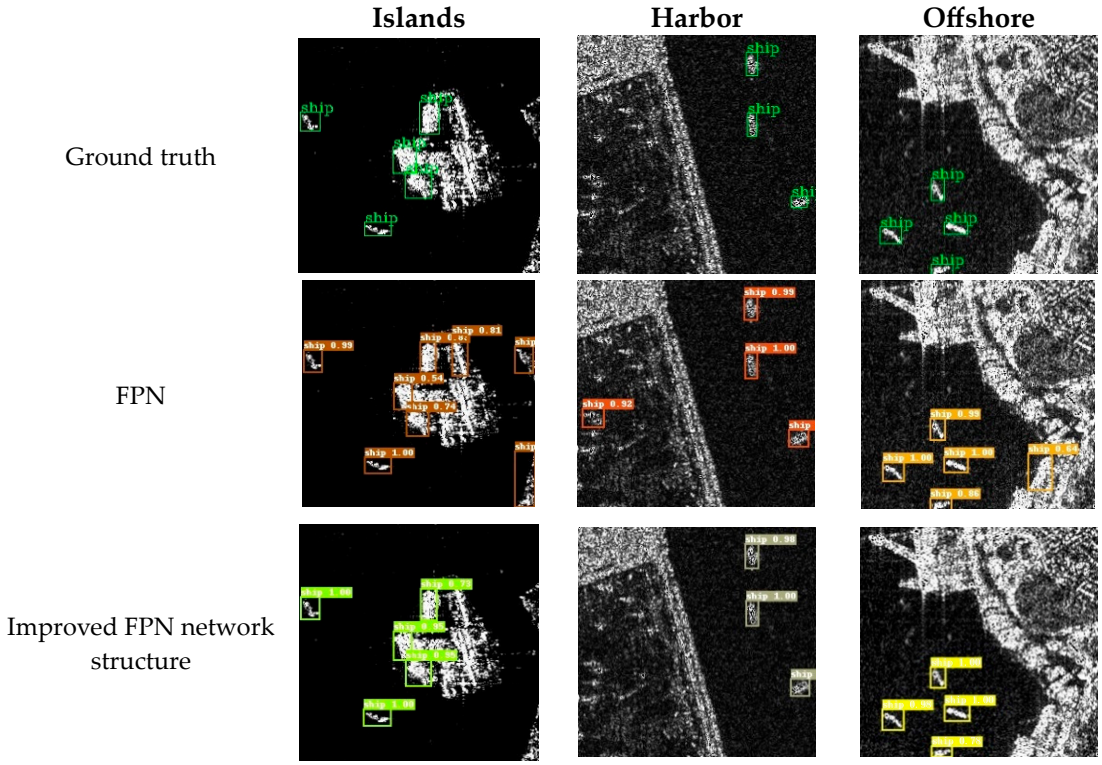


Figure 8. Qualitative results of FPN and improved FPN on SAR Ships dataset in complex scenes. The first, second, and third rows represent the ground truth, FPN detection results, and improved FPN network structure detection results, respectively. Most of the prediction boxes are matched, and those error results from FPN were detected by our improved method.

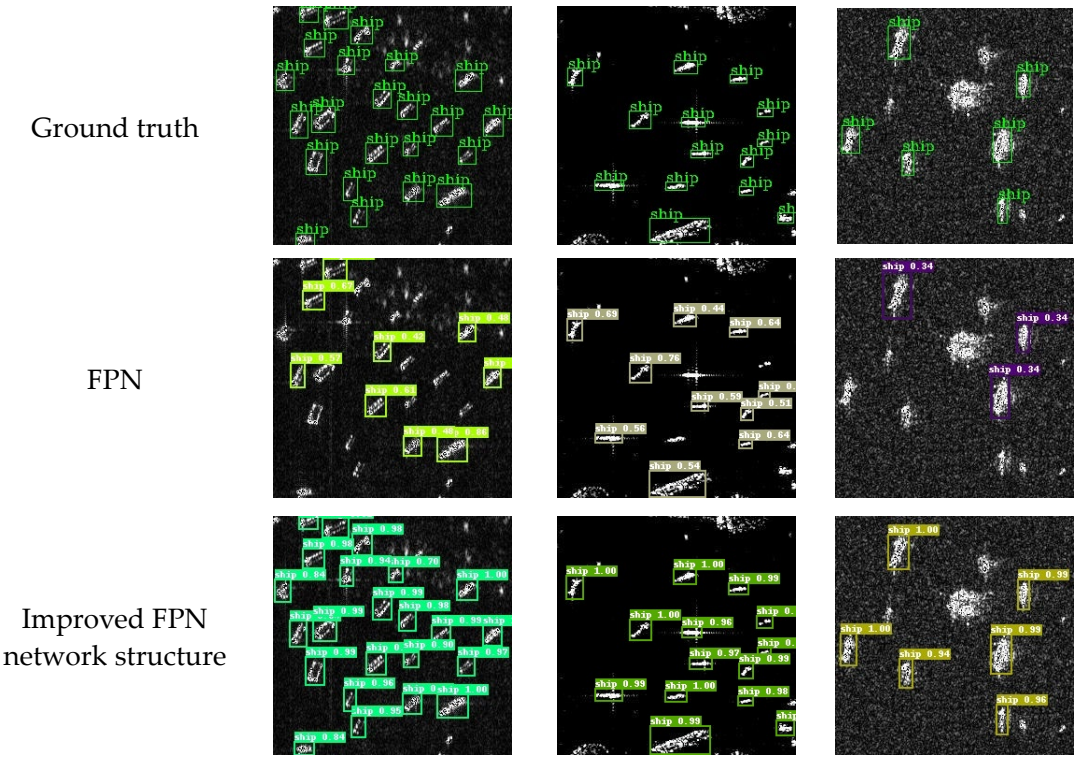


Figure 9. Qualitative results of FPN and improved FPN on SAR ship datasets in the high-density small target dataset. The first, second, and third rows represent the ground truth, FPN detection results, and improved FPN network structure detection results, respectively. The improved FPN network structure has higher accuracy.

3.3. Experiments on HRSID

Other validation experiments were performed on the public dataset HRSID. The original SAR image used to construct HRSID [40] includes 99 Sentinel-1B images, 36 TerraSAR-X images, and 1 TanDEM-X image. It is designed for ship detection based on CNN. It is also divided 70:30, respectively, into a training set and a testing set. According to statistics, the total number of ships marked in HRSID is 16,951. The number of small ships accounted for 54.8% of all ships.

FPNs were used to train these models to obtain baseline results. Then, channel attention and deformable CNNs were added to the models, and their original losses were replaced with GIoU in the final bounding box refinement stage. Similar to the above experiments, we gradually improved the models. The final results using the dataset HRSID are presented in Table 5.

Table 5. Comparison among detection results in HRSID.

FPN	Channel Attention	Deformable CNN	Improved Loss Function	Precision (%)	Recall (%)	F1 score	mAP (%)
√				88.2	92.1	0.901	88.2
√	√			88.7	92.6	0.906	88.7
√		√		89.2	93.0	0.911	89.3
√	√	√		89.3	93.2	0.912	89.3
√	√	√	√	89.6	93.3	0.914	89.6

The check mark "√" indicates that the technique was used in training.

The F1 score and mAP of the proposed method herein are increased by 0.013 and 1.4% as compared with the original object detection model, respectively. The small improvement observed in the experiments demonstrates that the channel attention mechanism and deformable convolution are useful in target-detection tasks. On the other hand, it also shows

that although deformable CNNs can adjust their shape by transforming the convolution kernel according to the actual situation, and channel attention can adjust weights to better extract features from the input image, they perform better in complex scenes or for small target detection.

3.4. Comparison with Other Models

Currently, the model proposed in this paper is compared with several classical and recently developed deep learning models, such as Yolov5 [27], CBAM Faster R-CNN [15], SSD [9], Mask R-CNN [36], and MS-FPN [30]. These models can achieve good detection results in general ship detection tasks. Figure 10 shows the ship recognition results of these models and results with the proposed model under different scenarios. The aforementioned models were executed on a Tesla V100, and the same training strategy was used to train the SAR ocean dataset with 105 iterations. The batch size was set to 32, and the initial learning rate was set to 0.00001. Tables 6 and 7 show that the proposed model has the highest detection accuracy. The mAP reaches 87.9% in complex scenes, which is 2.1% higher than the next-best model, Mask R-CNN, and 4.8% higher than the CBAM Faster R-CNN, which also incorporates the attention mechanism This indirectly reflects that the proposed model was more effective in identifying ships under extremely complex scenes. Meanwhile, the mAP was 95.1% in high-density small target scenes, which was 3.7% higher than that of the suboptimal model Mask R-CNN and 5.2% higher than that of the attention mechanism model CBAM Faster R-CNN. This indirectly reflects the improvement of the proposed model in terms of multiscale ship detection performance.

Table 6. Comparison of mAP from multiple models in complex scenes.

Model	SAR Ships in Complex Scenes			
	Precision (%)	Recall (%)	F1 Score	mAP (%)
Yolo v5	78.2	76.1	0.771	76.3
CBAM Faster R-CNN	63.3	85.5	0.727	83.1
SSD	83.5	78.6	0.809	79.5
Mask R-CNN	91.7	77.9	0.842	85.8
MS-FPN	89.3	77.7	0.831	87.9
Proposed method	91.7	78.1	0.844	87.9

Table 7. Comparison of mAP from multiple models in high-density small target scenes.

Model	High-Density Small Target Scenes			
	Precision (%)	Recall (%)	F1 Score	mAP (%)
Yolo v5	92.8	87.1	0.898	84.4
CBAM Faster R-CNN	91.4	77.7	0.839	89.9
SSD	94.1	71.3	0.822	70.0
Mask R-CNN	95.9	93.8	0.948	91.4
MS-FPN	92.9	93.2	0.930	92.9
Proposed method	96.5	93.0	0.947	95.1

Ground truth

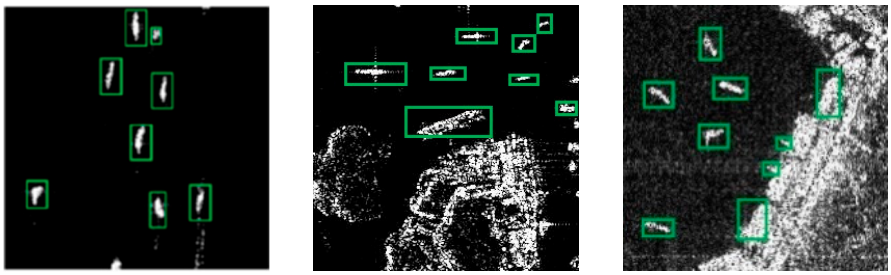


Figure 10. *Cont.*

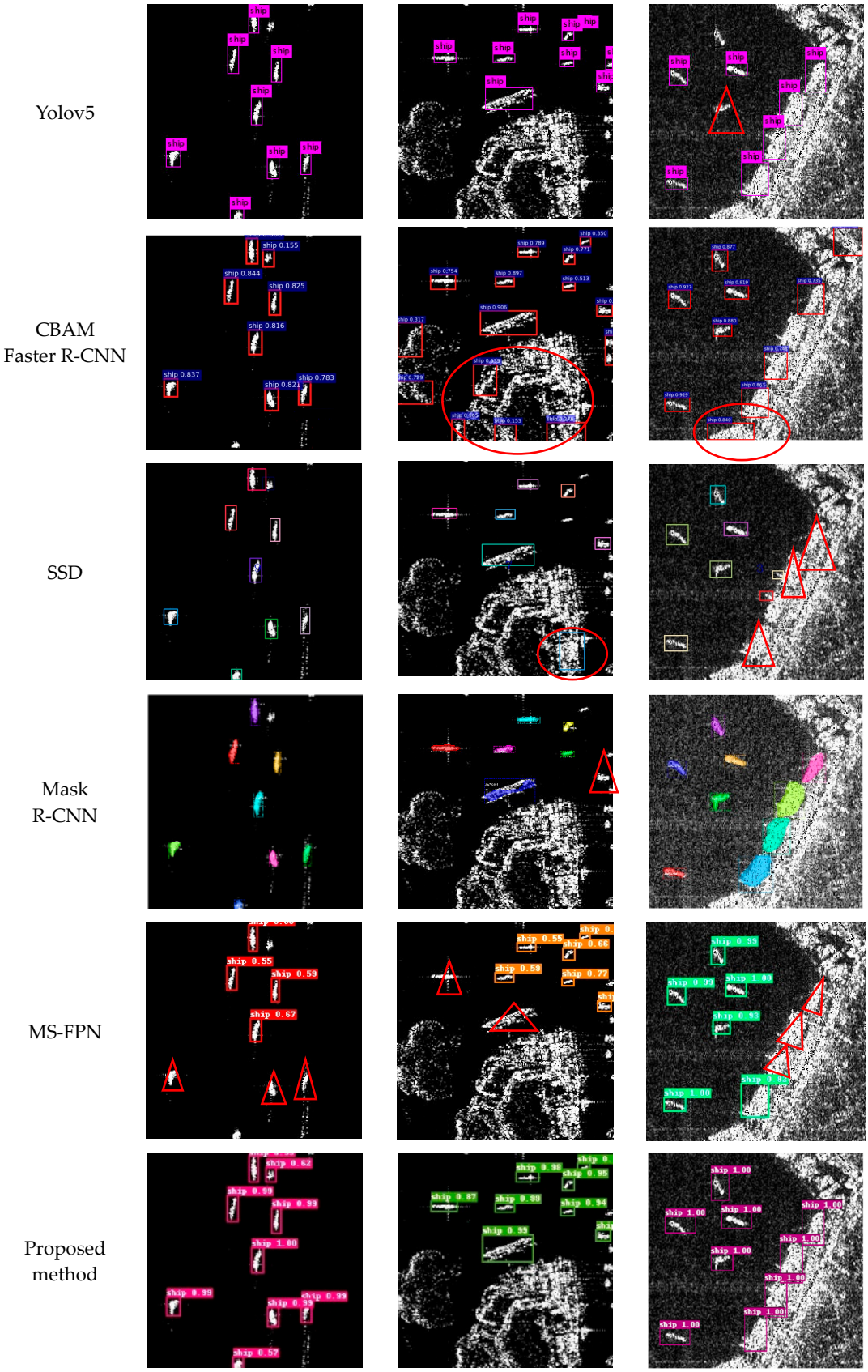


Figure 10. Comparison of detection results for different models. The triangles indicate imprecise regions, while the ovals denote missed areas.

4. Conclusions

For small target detection in complex scenes, which is prone to errors and missed detection problems, this study proposes a method based on an improved FPN target-detection model that can significantly improve ship SAR detection performance in different complex scenes and at different scales, especially the small ones. The improved FPN model in this paper consists of the deformable CNN module and the channel attention mechanism module. The deformable CNN was used to replace the original CNN, and the channel attention mechanism was introduced to extract the weights of the feature channels to optimize the expression ability of the network features. The performance of the proposed model was verified on a SAR ship dataset, showing that the mAP of ship detection via the proposed method reaches 94.7%, which can prove that the proposed method achieves a better performance than other state-of-the-art ship detection methods, such as CBAM faster R-CNN, SSD, YOLOV5, MS-FPN, and their variations.

To determine which module is most effective in improving the model, we compared the deformable CNN module, the channel attention mechanism module, their combination, and the replacement of the loss function. The results indicate that the deformable CNN module has the greatest impact, while combining the two modules further improves detection accuracy. However, we also observed that our proposed model fails to detect some very small ships against an interference background. The low recall rate in our ablation experiment suggests that there are missed detections. Additionally, introducing the GIoU loss function did not significantly improve detection accuracy.

Overall, the proposed model effectively detects ships in complex scenes and small ships in dense scenes. We have demonstrated that using deformable convolution instead of traditional convolution and introducing the channel attention mechanism is a more effective approach. However, it also needs to be noted that, although the proposed method has better detection performance and can effectively reduce false alarms. it cannot completely eliminate all false alarms for complex backgrounds. Further analysis and research are required.

Author Contributions: P.C. conceived and designed the algorithm and contributed to the manuscript and experiments; H.Z. was responsible for the construction of the ship detection dataset, constructed the outline of the manuscript, and made the first draft of the manuscript; Y.L. and B.L. supervised the experiments and were also responsible for the dataset; P.L. performed ship detection using machine learning methods. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Dalian Neusoft Institute of Information Joint Fund Project LH-JSRZ-202203, the Fundamental Scientific Research Project for Liaoning Education Department LJKMZ20222006, and the National Natural Science Foundation of CHINA 52271359.

Data Availability Statement: Owing to the nature of this research, the participants in this study did not agree that their data can be publicly shared; therefore, supporting data are not available.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Xiao, X.; Zhou, Z.; Wang, B.; Li, L.; Miao, L. Ship Detection under Complex Backgrounds Based on Accurate Rotated Anchor Boxes from Paired Semantic Segmentation. *Remote Sens.* **2019**, *11*, 2506. [CrossRef]
2. Yang, Z.; Yu, X.; Dedman, S.; Rosso, M.; Zhu, J.; Yang, J.; Xia, Y.; Tian, Y.; Zhang, G.; Wang, J. UAV Remote Sensing Applications in Marine Monitoring: Knowledge Visualization and Review. *Sci. Total Environ.* **2022**, *838*, 155939. [CrossRef] [PubMed]
3. Li, J.; Qu, C.; Shao, J. Ship Detection in SAR Images Based on an Improved Faster R-CNN. In Proceedings of the 2017 SAR in Big Data Era: Models, Methods and Applications (BIGSARDATA), Beijing, China, 13–14 November 2017; pp. 1–6.
4. Fu, Q.; Luo, K.; Song, Y.; Zhang, M.; Zhang, S.; Zhan, J.; Duan, J.; Li, Y. Study of Sea Fog Environment Polarization Transmission Characteristics. *Appl. Sci.* **2022**, *12*, 8892. [CrossRef]
5. Zhao, C.; Cheung, C.F.; Xu, P. High-Efficiency Sub-Microscale Uncertainty Measurement Method Using Pattern Recognition. *ISA Trans.* **2020**, *101*, 503–514. [CrossRef]
6. An, Q.; Pan, Z.; You, H. Ship Detection in Gaofen-3 SAR Images Based on Sea Clutter Distribution Analysis and Deep Convolutional Neural Network. *Sensors* **2018**, *18*, 334. [CrossRef] [PubMed]

7. Kang, M.; Ji, K.; Leng, X.; Lin, Z. Contextual Region-Based Convolutional Neural Network with Multilayer Fusion for SAR Ship Detection. *Remote Sens.* **2017**, *9*, 860. [CrossRef]
8. Liu, W.; Wang, Z.; Liu, X.; Zeng, N.; Liu, Y.; Alsaadi, F.E. A Survey of Deep Neural Network Architectures and Their Applications. *Neurocomputing* **2017**, *234*, 11–26. [CrossRef]
9. Zhang, J.; Lin, S.; Ding, L.; Bruzzone, L. Multi-Scale Context Aggregation for Semantic Segmentation of Remote Sensing Images. *Remote Sens.* **2020**, *12*, 701. [CrossRef]
10. Rostami, M.; Kolouri, S.; Eaton, E.; Kim, K. Deep Transfer Learning for Few-Shot SAR Image Classification. *Remote Sens.* **2019**, *11*, 1374. [CrossRef]
11. Liu, L.; Chen, G.; Pan, Z.; Lei, B.; An, Q. Inshore Ship Detection in SAR Images Based on Deep Neural Networks. In Proceedings of the IGARSS 2018–2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 25–28.
12. Li, R.; Wang, X.; Wang, J.; Song, Y.; Lei, L. SAR Target Recognition Based on Efficient Fully Convolutional Attention Block CNN. *IEEE Geosci. Remote Sens. Lett.* **2020**, *19*, 21510590. [CrossRef]
13. Zhou, G.; Yang, F.; Xiao, J. Study on Pixel Entanglement Theory for Imagery Classification. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 21706089. [CrossRef]
14. Wang, Y.; Wang, C.; Zhang, H. Combining a Single Shot Multibox Detector with Transfer Learning for Ship Detection Using Sentinel-1 SAR Images. *Remote Sens. Lett.* **2018**, *9*, 780–788. [CrossRef]
15. Kang, M.; Leng, X.; Lin, Z.; Ji, K. A Modified Faster R-CNN Based on CFAR Algorithm for SAR Ship Detection. In Proceedings of the 2017 International Workshop on Remote Sensing with Intelligent Processing (RSIP), Shanghai, China, 18–21 May 2017; pp. 1–4.
16. Jiao, J.; Zhang, Y.; Sun, H.; Yang, X.; Gao, X.; Hong, W.; Fu, K.; Sun, X. A Densely Connected End-to-End Neural Network for Multiscale and Multiscene SAR Ship Detection. *IEEE Access* **2018**, *6*, 20881–20892. [CrossRef]
17. Cui, Z.; Wang, X.; Liu, N.; Cao, Z.; Yang, J. Ship Detection in Large-Scale SAR Images via Spatial Shuffle-Group Enhance Attention. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 379–391. [CrossRef]
18. Villano, M.; Krieger, G.; Steinbrecher, U.; Moreira, A. Simultaneous Single-/Dual-and Quad-Pol SAR Imaging over Swaths of Different Widths. *IEEE Trans. Geosci. Remote Sens.* **2019**, *58*, 2096–2103. [CrossRef]
19. Wang, R.; Xu, F.; Pei, J.; Wang, C.; Huang, Y.; Yang, J.; Wu, J. An Improved Faster R-CNN Based on MSER Decision Criterion for SAR Image Ship Detection in Harbor. In Proceedings of the IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 1322–1325.
20. Hu, Q.; Hu, S.; Liu, S. BANet: A Balance Attention Network for Anchor-Free Ship Detection in SAR Images. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 21603357. [CrossRef]
21. Hu, Q.; Hu, S.; Liu, S.; Xu, S.; Zhang, Y.-D. FINet: A Feature Interaction Network for SAR Ship Object-Level and Pixel-Level Detection. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–15. [CrossRef]
22. Wu, T.; Li, B.; Luo, Y.; Wang, Y.; Xiao, C.; Liu, T.; Yang, J.; An, W.; Guo, Y. MTU-Net: Multilevel TransUNet for Space-Based Infrared Tiny Ship Detection. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 22538002. [CrossRef]
23. Zhang, F.; Wang, X.; Zhou, S.; Wang, Y.; Hou, Y. Arbitrary-Oriented Ship Detection through Center-Head Point Extraction. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 21590405. [CrossRef]
24. He, C.; Tu, M.; Xiong, D.; Tu, F.; Liao, M. Adaptive Component Selection-Based Discriminative Model for Object Detection in High-Resolution SAR Imagery. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 72. [CrossRef]
25. Yang, M.; Wang, H.; Hu, K.; Yin, G.; Wei, Z. IA-Net: An Inception–Attention-Module-Based Network for Classifying Underwater Images From Others. *IEEE J. Ocean. Eng.* **2022**, *47*, 704–717. [CrossRef]
26. Zhou, G.; Song, B.; Liang, P.; Xu, J.; Yue, T. Voids Filling of DEM with Multiattention Generative Adversarial Network Model. *Remote Sens.* **2022**, *14*, 1206. [CrossRef]
27. Ting, L.; Baijun, Z.; Yongsheng, Z.; Shun, Y. Ship Detection Algorithm Based on Improved YOLO V5. In Proceedings of the 2021 6th International Conference on Automation, Control and Robotics Engineering (CACRE), Dalian, China, 15–17 July 2021; pp. 483–487.
28. Li, L.; Zhou, Z.; Wang, B.; Miao, L.; An, Z.; Xiao, X. Domain Adaptive Ship Detection in Optical Remote Sensing Images. *Remote Sens.* **2021**, *13*, 3168. [CrossRef]
29. Rezatofighi, H.; Tsoi, N.; Gwak, J.; Sadeghian, A.; Reid, I.; Savarese, S. Generalized Intersection over Union: A Metric and a Loss for Bounding Box Regression. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 658–666.
30. Sun, Z.; Meng, C.; Cheng, J.; Zhang, Z.; Chang, S. A Multi-Scale Feature Pyramid Network for Detection and Instance Segmentation of Marine Ships in SAR Images. *Remote Sens.* **2022**, *14*, 6312. [CrossRef]
31. Lin, Z.; Ji, K.; Leng, X.; Kuang, G. Squeeze and Excitation Rank Faster R-CNN for Ship Detection in SAR Images. *IEEE Geosci. Remote Sens. Lett.* **2018**, *16*, 751–755. [CrossRef]
32. Li, J.; Xu, C.; Su, H.; Gao, L.; Wang, T. Deep Learning for SAR Ship Detection: Past, Present and Future. *Remote Sens.* **2022**, *14*, 2712. [CrossRef]
33. Zhao, Y.; Zhao, L.; Xiong, B.; Kuang, G. Attention Receptive Pyramid Network for Ship Detection in SAR Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 2738–2756. [CrossRef]

34. Liu, N.; Cui, Z.; Cao, Z.; Pi, Y.; Lan, H. Scale-Transferrable Pyramid Network for Multi-Scale Ship Detection in SAR Images. In Proceedings of the IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 1–4.
35. Li, X.; Li, D.; Liu, H.; Wan, J.; Chen, Z.; Liu, Q. A-BFPN: An Attention-Guided Balanced Feature Pyramid Network for SAR Ship Detection. *Remote Sens.* **2022**, *14*, 3829. [CrossRef]
36. Nie, X.; Duan, M.; Ding, H.; Hu, B.; Wong, E.K. Attention Mask R-CNN for Ship Detection and Segmentation from Remote Sensing Images. *IEEE Access* **2020**, *8*, 9325–9334. [CrossRef]
37. Lin, T.-Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature Pyramid Networks for Object Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2117–2125.
38. Dai, J.; Qi, H.; Xiong, Y.; Li, Y.; Zhang, G.; Hu, H.; Wei, Y. Deformable Convolutional Networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 764–773.
39. Wang, Y.; Wang, C.; Zhang, H.; Dong, Y.; Wei, S. A SAR Dataset of Ship Detection for Deep Learning under Complex Backgrounds. *Remote Sens.* **2019**, *11*, 765. [CrossRef]
40. Wei, S.; Zeng, X.; Qu, Q.; Wang, M.; Su, H.; Shi, J. HRSID: A High-Resolution SAR Images Dataset for Ship Detection and Instance Segmentation. *IEEE Access* **2020**, *8*, 120234–120254. [CrossRef]