A Ship Detection Method via Redesigned FCOS in Large-Scale SAR Images

Mingming Zhu 1, Guoping Hu 2,*, Hao Zhou 2, Shiqiang Wang 2, Ziang Feng 2 and Shijie Yue 2

1 Graduate College, Air Force Engineering University, Xi'an 710051, China; ming_paper@163.com
2 Air and Missile Defense College, Air Force Engineering University, Xi'an 710051, China; haozhou_kgd@163.com (H.Z.); wsq_kgd@163.com (S.W.); feng_ziang@163.com (Z.F.); yueshi_jie@163.com (S.Y.)
* Correspondence: hgp_kgd123@163.com

Abstract: Ship detection in large-scale synthetic aperture radar (SAR) images has achieved breakthroughs as a result of the improvement of SAR imaging technology. However, there still exist some issues due to the scattering interference, sparsity of ships, and dim and small ships. To address these issues, an anchor-free method is proposed for dim and small ship detection in large-scale SAR images. First, fully convolutional one-stage object detection (FCOS) as the baseline is applied to detecting ships pixel by pixel, which can eliminate the effect of anchors and avoid the missing detection of small ships. Then, considering the particularity of SAR ships, the sample definition is redesigned based on the statistical characteristics of ships. Next, the feature extraction is redesigned to improve the feature representation for dim and small ships. Finally, the classification and regression are redesigned by introducing an improved focal loss and regression refinement with complete intersection over union (CIoU) loss. Experimental simulation results show that the proposed R-FCOS method can detect dim and small ships in large-scale SAR images with higher accuracy compared with other methods.

Keywords: synthetic aperture radar (SAR); ship detection; anchor-free

1. Introduction

Due to the continuous improvements in the quantity and quality of synthetic aperture radar (SAR) images [1–3], ship detection has been widely applied for maritime management and surveillance. Ship detection has become a subject task with important theoretical research and practical applications, so that it has attracted more and more scholars’ attention [4–8]. For example, Li et al. [9] introduced the superpixels into constant false alarm rate (CFAR) ship detection method. Salembier et al. [10] applied graph signal processing based on Maxtree representation for ship detection. Lin et al. [11] proposed a ship detection method via superpixels and fisher vectors. Wang et al. [12] proposed the local contrast of fisher vectors (LCFVs) for detecting ships. However, these traditional methods still have limitations. On the one hand, it is difficult to design appropriate manual features for ships, resulting in weak generalization ability. On the other hand, the time cost is high due to the complex detection process.

Benefiting from the excellent performance of convolutional neural networks (CNN) in object classification, research on CNN-based object detection is booming. Object detection methods proposed in the past ten years can be roughly divided into two categories. The first are the two-stage detectors, such as faster region-based CNN (Faster R-CNN) [13]. The second are the one-stage detectors, such as you only look once (YOLO) [14], single shot multi-box detector (SSD) [15], and RetinaNet [16]. The two-stage detectors can achieve relatively good detection performance but have high time costs. However, the one-stage detectors have lower computational costs but have a weaker detection performance.

As a result of the emergence of large-scale SAR datasets [17–21], CNN-based detectors have been introduced for ship detection [22–26]. For example, Deng et al. [27]...
proposed a method to detect small and densely clustered ships, and it is trained from scratch. Chen et al. [28] proposed a ship detection method via an adaptive recalibration mechanism. Jin et al. [29] proposed P2P-CNN via using the ship and its surroundings to deal with the small ship problem. Gao et al. [30] proposed SAR-Net for achieving a balance between speed and accuracy. These methods are anchor-based detectors, whose detection accuracies are related to the quality of the pre-defined anchors. Although the anchor-based methods work well in detecting ships, they still have certain shortcomings. Firstly, the anchor box will introduce additional hyperparameters, which are set based on prior knowledge. When the detection object changes, the hyperparameters need to be reset. Setting inappropriate anchor boxes will result in poor detection performance, thereby reducing the generalization ability of the network. Secondly, considering the sparsity of ships, most anchor boxes tend to contain empty backgrounds, and only a few anchor boxes contain ships. This will result in far more negative samples than positive samples, resulting in an unbalanced design. Finally, a large number of anchor frames are redundant because of the sparsity of ships, thereby bringing additional computational costs.

Anchor-free methods directly predict class and location instead of using the anchor box. Anchor-free methods consist of key-point-based methods [31–33] and center-based methods [34–38]. The key-point-based methods, such as CenterNet [33] and representative points (RepPoints) [31], first detect the key points and then combine the key points for object detection. Center-based methods, such as fully convolutional one-stage object detection (FCOS) [35], FoveaBox [36], and feature selective anchor-free (FSAF) [37], directly detect objects by the center point and bounding box. Although the anchor-free methods [39–41] for ship detection are still under development, they have shown a better performance potential and a trade-off between speed and accuracy. However, the performance of the anchor-free method still needs to be improved due to the following issues. First, there is a lot of scattering interference from islands or the sea in SAR images. Secondly, the sparsity of ships cannot be ignored. Finally, some ships are small and dim, that is, the proportion of ships is small, and the object scattering is weak.

Therefore, a novel method called redesigned FCOS (R-FCOS) is proposed for dim and small ship detection. Specifically, we redesign the anchor-free detector FCOS to address the issues described above, and the architecture of R-FCOS is shown in Figure 1. The contributions are summarized as follows:

Figure 1. The overall architecture of R-FCOS. W, H denote the width and height of the feature map.
1. It is shown that R-FCOS can eliminate the effect of anchors and avoid missing detection of small ships.
2. Considering the particularity of SAR ships, the sample definition was redesigned based on the statistical characteristics of these ships.
3. The feature extraction was redesigned to improve the feature representation for dim and small ships.
4. The classification and regression stages were redesigned by introducing an improved focal loss and bounding box refinement with complete intersection over union (CIoU) loss.

The rest of the paper is organized as follows. Section 2 introduces the proposed method based on FCOS. In Section 3, the details of experiments and the analysis of the results are exhibited. Section 4 describes the discussion. Finally, the conclusion is presented in Section 5.

2. Materials and Methods

In this section, the FOCS network, the baseline of the proposed method, is introduced. Then, the R-FCOS method, including redesigned sample definition, feature extraction, classification, and regression are described in detail. Finally, the loss function is given.

2.1. Baseline

The core idea of FCOS is to detect objects pixel by pixel, that is, to directly predict the distance \((l, t, r, b)\) from the center point to the four sides of the bounding box, as shown in Figure 2. Figure 3 shows the architecture of the FCOS network, which includes a backbone network, a feature pyramid network (FPN), a classification branch, a regression branch, and a center-ness branch. The backbone network is used to extract feature maps of the input images. In FPN, the different feature pyramid levels are used to detect multi-scale objects. The classification branch and regression branch are to achieve object classification and location, respectively. The center-ness branch is to evaluate the “center-ness” of a location. The “center-ness” represents the degree of coincidence of the position with the center of the ground truth bounding box, which is defined as:

\[
\text{center-ness} = \frac{\min(l, r)}{\max(l, r)} \times \frac{\min(t, b)}{\max(t, b)}
\]  

Figure 2. The predicted 4D vector of FCOS.
The total training loss as follows:

\[
L\left(\{p_{x,y}\}, \{t_{x,y}\}\right) = \frac{1}{N_{\text{pos}}} \sum_{x,y} L_{\text{cls}}(p_{x,y}, c^*_{x,y}) + \frac{1}{N_{\text{pos}}} \sum_{x,y} \left[c^*_{x,y} \geq 1\right] L_{\text{reg}}(t_{x,y}, t^*_{x,y}) + \frac{1}{N_{\text{pos}}} \sum_{x,y} L_{\text{CE}}(\text{center-ness}_{x,y}, \text{center-ness}^*_{x,y})
\]

(2)

where \(p\) is the classification score, \(t = (l, t, r, b)\) is the regression prediction, and \(N_{\text{pos}}\) is the number of positive samples. \(L_{\text{cls}}, L_{\text{reg}},\) and \(L_{\text{CE}}\) denote loss function [42], generalized intersection over union (GloU) loss [42], and binary cross entropy loss, respectively, and \(\left[c^*_{x,y} \geq 1\right]\) denotes the indicator function as follows:

\[
\left[c^*_{x,y} \geq 1\right] = \begin{cases} 1 & c^*_{x,y} \geq 1 \\ 0 & \text{otherwise} \end{cases}
\]

(3)

### 2.2. Sample Definition Redesign

SAR images and natural images have obvious differences in imaging conditions and object categories. Therefore, the previous sample definition methods based on the intersection over union (IoU) threshold, or the scale range will not be suitable for SAR ship detection. Therefore, we introduced a new sample definition method, which defines the sample threshold according to the statistical characteristics of SAR ships. The specific process is as follows:

**Step 1:** For a ground-truth box \(g \in G\) on each pyramid level, the L2 distances between the predicted boxes \(D_i \subseteq D\) on the \(i\)th pyramid level and the ground-truth \(g\) are computed.

**Step 2:** The L2 distances are sorted from small to large, and the first \(k\) corresponding predicted boxes \(A_i\) are selected.

**Step 3:** For the different pyramid level \(i\), the total candidate positive samples are computed as follows:

\[
A_i = \bigcup_i A_i
\]

(4)

**Step 4:** The IoU between \(A_i\) and \(g\) is computed as follows:

\[
O_{i} = \text{IoU}(A_i, g)
\]

(5)

**Step 5:** The mean and standard deviation are computed as follows:

\[
\mu_g = \text{Mean}(O_g)
\]

(6)

\[
\sigma_g = \text{Std}(O_g)
\]

(7)
Step 6: The IoU threshold for \( g \) is as follows:

\[
T_g = \mu_g + v_g
\]  
(8)

Step 7: For each candidate \( a \in A_g \), if \( \text{IoU}(a, g) \geq T_g \) and the center of \( a \) in \( g \), then

\[
\text{Positive}_g = \text{Positive}_g \cup a
\]  
(9)

Step 8: Finally, all positive samples are as follows:

\[
\text{Positive} = \bigcup_{g \in G} \text{Positive}_g
\]  
(10)

Step 9: The rest are negative samples as follows:

\[
\text{Negative} = D - \text{Positive}
\]  
(11)

2.3. Feature Extraction Redesign

As the feature level increases, the resolution of the feature map decreases while the semantic information contained increases. For dim and small ship detection, high-resolution feature representation is essential. This is because the low-resolution feature maps contain very little object information due to the small ships with less pixels and weak scattering. Previous methods usually restore the obtained low-resolution features to high-resolution, such as Hourglass [43], SegNet [44], and the deconvolution network (DeconvNet) [45]. Different from the previous methods, we no longer add additional operations of feature resolution recovery. In other words, the high-resolution feature representation persists throughout the entire computing stage. In addition, the multi-resolution features are repeatedly fused to obtain rich semantic information and high-resolution feature representation. In summary, we redesign feature extraction via the same-resolution feature convolution (SFC) module, multi-resolution feature fusion (MFF) module, and feature pyramid (FP) module, as shown in Figure 1.

At first, the input image is down-sampled via two \( 3 \times 3 \) stride-2 convolutions.

\[
F_0 = \text{Conv}_{3 \times 3, s=2}(\text{Conv}_{3 \times 3, s=2}(\text{Input}))
\]  
(12)

where \( F_0 \) denotes the feature map, and \( \text{Conv}_{3 \times 3, s=2} \) denotes the operation of \( 3 \times 3 \) stride-2 convolution.

Two types of SFC modules are used to extract feature maps, of which two residual blocks, i.e., Basicblock and Bottleneck, are the main components, as shown in Figure 4. The computation processes of Basicblock and Bottleneck are summarized as follows:

\[
f(F) = \text{Conv}_{1 \times 1}(F) + \text{Conv}_{1 \times 1}(\text{Conv}_{3 \times 3}(\text{Conv}_{1 \times 1}(F)))
\]  
(13)

\[
h(F) = F + \text{Conv}_{3 \times 3}(\text{Conv}_{3 \times 3}(F))
\]  
(14)

where \( \text{Conv}_{1 \times 1} \) and \( \text{Conv}_{3 \times 3} \) denote the operations of \( 1 \times 1 \) stride-1 convolution and \( 3 \times 3 \) stride-1 convolution, respectively; \( f(\cdot) \) and \( h(\cdot) \) denote the operations of Bottleneck and Basicblock, respectively.
The MFF module achieves information fusion between features through convolution and up-sample. Figure 5 shows the fusion process of three-resolution features. Given three-resolution features \( \{F_{31}, F_{32}, F_{33}\} \), the output features \( \{F_{41}, F_{42}, F_{43}, F_{44}\} \) are as follows:

\[
F_{41} = F_{31} + \text{Conv}_{1 \times 1}(\text{Upsample}_{s=2}(F'_{32})) + \text{Conv}_{1 \times 1}(\text{Upsample}_{s=4}(F'_{33})) \quad (15)
\]

\[
F_{42} = \text{Conv}_{1 \times 1}(\text{Conv}_{3 \times 3,s=2}(F'_{31})) + F'_{32} + \text{Conv}_{1 \times 1}(\text{Upsample}_{s=2}(F'_{33})) \quad (16)
\]

\[
F_{43} = \text{Conv}_{3 \times 3,s=2}(\text{Conv}_{3 \times 3,s=2}(F'_{31})) + \text{Conv}_{3 \times 3,s=2}(F'_{32}) + F'_{33} \quad (17)
\]

\[
F_{44} = \text{Conv}_{3 \times 3,s=2}(\text{Conv}_{3 \times 3,s=2}(\text{Conv}_{3 \times 3,s=2}(F'_{31}))) + \text{Conv}_{3 \times 3,s=2}(\text{Conv}_{3 \times 3,s=2}(F'_{32})) + \text{Conv}_{3 \times 3,s=2}(F'_{33}) \quad (18)
\]

where \( \text{Upsample}_{s=2} \) and \( \text{Upsample}_{s=4} \) denote the operations of bilinear up-sample with stride-2 and stride-4, respectively.

The final output features are up-sampled and concatenated for the subsequent FP module.

\[
F_{\text{out}} = \text{Concat}(F''_{41} + \text{Upsample}_{s=2}(F''_{42}) + \text{Upsample}_{s=4}(F''_{43}) + \text{Upsample}_{s=8}(F''_{44})) \quad (19)
\]

where \( \text{Upsample}_{s=8} \) denotes the operation of bilinear up-sample with stride-8, and Concat denotes the concatenation operation.
The dimension of $P_{in}$ is set to 256 by a $1 \times 1$ convolution.

$$P_{in} = \text{Conv}_{1 \times 1}(F_{out})$$ (20)

The FP module is constructed through a set of average pooling operations.

$$P_{\varphi}^{out} = \text{AvgPool}_{s=2^\varphi}(P_{in}), \varphi = 1, 2, 3, 4$$ (21)

where $\text{AvgPool}_{s=2^\varphi}$ denotes the operation of average pooling with stride-$2^\varphi$; $\varphi$ is the pyramid level.

Finally, a $3 \times 3$ convolution is appended on each level to obtain the final feature map called $\{P_0, P_1, P_2, P_3, P_4\}$.

$$P_{\varphi} = \text{Conv}_{3 \times 3}(P_{\varphi}^{out}), \varphi = 0, 2, 3, 4$$ (22)

### 2.4. Classification and Regression Redesign

To utilize the geometric features and contextual information of dim and small ships, we implemented a new feature representation for the bounding box via the features of nine fixed points, as shown in Figure 6. The deformable convolution (Dconv) is introduced to achieve the above feature representation. Specifically, these offsets between the other eight points and $(x, y)$ are used as the offsets of the Dconv.

![Figure 6](image)

**Figure 6.** A new feature representation for the bounding box.

The purpose of object detection is to identify the categories of all objects and give their localizations. Object detectors generally use bounding boxes with category labels and classification scores to represent the detection results. Some bounding boxes with accurate localizations may be eliminated due to low classification scores. This indicates that the classification score is not suitable for estimating detection accuracy. Therefore, we use an IoU score (IS) to simultaneously express classification and localization accuracy.

Due to the sparsity of ships, most of the regions in SAR images are negative samples, and the sample imbalance is unavoidable. To solve this issue, an improved focal loss function is proposed to predict IS. Specifically, an asymmetric training sample weighting method is applied to focal loss. Focal loss is defined as:

$$L_{FL}(p, z) = \begin{cases} 
-\alpha(1 - p)^\gamma \log(p) & \text{if } z = 1 \\
-(1 - \alpha)p^\gamma \log(1 - p) & \text{otherwise}
\end{cases}$$ (23)

where $z \in \{-1, +1\}$ is class label, $p \in \{0, 1\}$ is predicted probability, and $\alpha, \gamma$ are the weighting factor and tunable focusing parameter, respectively.

Next, we used an asymmetric sample weighting to improve focal loss function.

$$L_{IFL}(IS, q) = \begin{cases} 
-q(q\log(IS) + (1 - q) \log(1 - IS)) & q > 0 \\
-\alpha IS^\gamma \log(1 - IS) & q = 0
\end{cases}$$ (24)
where $IS$ and $q$ denote the predicted and target $IS$, respectively.

There is a lot of scattering interference from islands and the sea in SAR images. Therefore, we introduced a regression refinement branch to improve localization accuracy by a refinement factor $(\Delta l, \Delta t, \Delta r, \Delta b)$. The final output of the regression branch is $(\Delta l \cdot l', \Delta t \cdot t', \Delta r \cdot r', \Delta b \cdot b')$. To improve the speed and accuracy of loss convergence, a CIoU loss is introduced into the regression branch:

$$L_{\text{CIoU}}(d, g) = 1 - \text{IoU} + \frac{\rho^2(d, g)}{\Delta^2} + \beta$$

(25)

where $d$ and $g$ denote the center point of predicted and ground truth box, respectively, $\rho$ is the Euclidean distance, $\Delta$ denotes the diagonal distance of the smallest rectangle containing the predicted and ground-truth box, and $\beta$ is an impact factor.

2.5. Loss Function

The entire network is trained with a multi-task loss function as follows:

$$L = \frac{1}{N_{\text{pos}}} \sum_j L_{\text{IFL}}(IS_j, q_j) + \frac{1.5}{N_{\text{pos}}} \sum_j \frac{q_j L_{\text{CIoU}}(d_j, g_j)}{q_j} + \frac{2}{N_{\text{pos}}} \sum_j \frac{q_j L_{\text{CIoU}}(d'_j, g'_j)}{q_j}$$

(26)

where $d_j$ denotes the initial box at location $j$ on feature map, and $d'$ and $g$ are the refined and ground-truth box, respectively.

3. Results

All experiments in this paper are performed on a computer equipped with RTX 2080Ti GPU and Intel i9-9820X CPU. The operating system used is Ubuntu 16.04, and the basic framework is pytorch. The average precision (AP) and frames per second (FPS) are used as the evaluation metrics.

3.1. Dataset

The Large-Scale SAR Ship Detection Dataset-v1.0 (LS-SSDD-v1.0) [20] was constructed from Sentinel-1 satellites and included 15 images with $24,000 \times 16,000$ pixels, a sample image is shown in Figure 7. To facilitate network training and testing, the original image is split into 9000 sub-images with $800 \times 800$ pixels for training (6000) and testing (3000). According to the scale division for bounding box in Microsoft Common Objects in COntext (MS COCO), the area below 1024 pixels is a small ship, the area from 1024 pixels to 9216 pixels is a medium ship, and the area above 9216 pixels is a large ship. The distributions of ship shape and number are shown in Figure 8 and Figure 9, respectively.

Figure 7. A sample image of LS-SSDD-v1.0.
3.2. Ablation Study

Since the proposed method is similar in structure to FCOS, FCOS is used as the baseline. Figure 10 shows the detection results of FCOS, FCOS with sample definition redesign (FCOS+SDR), FCOS with sample definition redesign and feature extraction redesign (FCOS+SDR+FER), and FCOS with sample definition redesign, feature extraction redesign, and classification and regression redesign (FCOS+SDR+FER+CRR).

Figure 8. The distribution of the ship’s shape.

Figure 9. The distribution of the ship’s number.
3.2.1. Analysis of Sample Definition Redesign

As shown in Figure 10, the AP of FCOS+SDR is 1.3% higher than that of FCOS. This suggests that sample definition redesign is effective for improving detection accuracy by redefining the sample threshold according to the statistical characteristics of SAR ships.

3.2.2. Analysis of Feature Extraction Redesign

As shown in Figure 10, the APs of FCOS+SDR+FER are 2.3% and 1.0% higher than those of FCOS and FCOS+SDR, respectively, suggesting that feature extraction redesign is effective for improving detection accuracy. The main reason is that feature extraction redesign can obtain rich semantic information and high-resolution feature representation for dim and small ships.

3.2.3. Analysis of Classification and Regression Redesign

As shown in Figure 10, the APs of R-FCOS are 3.2% and 0.9% higher than those of FCOS and FCOS+SDR+FER, respectively. The main reason is that the classification and regression redesign is used to deal with these issues, i.e., complex surroundings, sparsity of ships, and dim and small ships. Specifically, the nine-point feature representation is used to extract geometric features and contextual information for dim and small ships. The improved focal loss function is used to predict IS. Finally, the regression refinement branch with CIoU loss is used to improve localization accuracy.

3.3. Comparison with Other Methods

Based on the same experimental environment, the ship detection performance of R-FCOS was compared with those of other methods such as Faster RCNN [13], SSD [15], RetinaNet [16], YOLOv3 [46], RepPoints [31], FSAF [37], and FoveaBox [36]. As illustrated in Figure 11, we can draw the following conclusions:

1. The AP of R-FCOS is better than those of other methods. Specifically, the AP of R-FCOS is 75.5%, which is 9.2%, 17.7%, 4.7%, 8.9%, 3.2%, 4.9%, and 6.1% higher than Faster RCNN, SSD, RetinaNet, YOLOv3, RepPoints, FSAF, and FoveaBox, respectively.
2. The AP of SSD is the worst and is 17.7% lower than our method. Although SSD uses high-resolution features to detect small objects, it contains less semantic information,
resulting in unsatisfactory detection results. In addition, SSD reduces the input image size to 300×300, which destroys the object information in the image.

3. The APs of anchor-free methods such as RepPoints, FSAF, FoveaBox, and R-FCOS are generally better than those of anchor-based methods except for RetinaNet. This shows that the anchor-free method is more suitable for ship detection.

4. The FPS of SSD is the highest, and that of Faster RCNN is the lowest. Although the FPS of our method is only 52.0, it already meets the real-time requirements.

Figure 11. The bar graph of different methods on LS-SSDD-v1.0.

To visually demonstrate the detection performance of R-FCOS, Figure 12 shows the comparative results for different methods on LS-SSDD-v1.0. In the first column of Figure 11, there exist some missing ships that occur in all methods. FSAF and R-FCOS have the least number of missed ships, but our method has higher classification scores. In the second column of Figure 11, most of the ships were missed by SSD, YOLOv3, and FoveaBox, and several false alarms were obtained by Faster RCNN. In addition, false alarms and missing ships existed in RetinaNet, RepPoints, and FSAF. However, our method had only one missing ship. In the third column of Figure 11, false alarms or missing ships exist in all methods, except for YOLOv3 and R-FCOS. However, the classification scores of R-FCOS were higher than those of YOLOv3. Overall, R-FCOS is better than the other methods.
Figure 12. Cont.
Figure 12. Cont.
4. Discussion

Simulation experiments were carried out on the SAR ship detection dataset (SSDD) [21] to verify the model migration ability. Figure 13 shows the detection results on SSDD. The AP of R-FCOS is 97.8%, which is 3.9%, 5.8%, 1.5%, 2.8%, 1.3%, 1.4%, and 2.2% higher than Faster RCNN, SSD, RetinaNet, YOLOv3, RepPoints, FSAF, and FoveaBox, respectively. Although the FPS of R-FCOS is only 17.5, it is acceptable. Figure 14 shows the visual results of R-FCOS on SSDD.

![Comparison results of different methods on LS-SSDD-v1.0.](image)

**Figure 12.** Comparison results of different methods on LS-SSDD-v1.0.

![The bar graph of different methods on SSDD.](image)

**Figure 13.** The bar graph of different methods on SSDD.
5. Conclusions

In this paper, an anchor-free detector R-FCOS was proposed for dim and small ship detection. We have redesigned FCOS with the aim to address the issues of complex surroundings, sparsity of ships, and dim and small ships. The sample definition redesign deals with the particularity of SAR ships. The feature extraction redesign improved the feature representation for dim and small ships. The classification and regression redesign introduced an improved focal loss and regression refinement with CIoU loss. In the experimental part, we verified the effectiveness of the sample definition redesign, feature extraction redesign, and classification and regression redesign. Experimental results on LS-SSDD-v1.0 showed that, the proposed method achieves competitive detection performance in comparison with Faster RCNN, SSD, RetinaNet, YOLOv3, RepPoints, FSAF, and FoveaBox. In addition, we verified the model migration ability of the proposed method on SSDD. However, it should be noted that, although the proposed method has better detection performance, it cannot completely eliminate all false alarms and missed detection results, which requires further analysis and research.
**Author Contributions:** Conceptualization, M.Z. and G.H.; methodology, M.Z.; software, M.Z.; validation, M.Z., H.Z., S.W., Z.F. and S.Y.; formal analysis, M.Z.; investigation, M.Z.; resources, M.Z.; data curation, M.Z.; writing—original draft preparation, M.Z.; writing—review and editing, M.Z.; visualization, M.Z.; supervision, M.Z.; project administration, G.H.; funding acquisition, S.W. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Natural Science Foundation Research Project of Shaanxi Province, China, under Grant 2020JM-345.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

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

**Abbreviations**

The following abbreviations are used in this manuscript:

- SAR: Synthetic Aperture Radar
- FCOS: Fully Convolutional One-Stage Object Detection
- CIoU: Complete Intersection over Union
- CFAR: Constant False Alarm Rate
- LCFVs: Local Contrast of Fisher Vectors
- CNN: Convolutional Neural Network
- Faster: R-CNN Faster Region-based CNN
- YOLO: You Only Look Once
- SSD: Single Shot Multi-Box Detector
- P2P-CNN: Patch-to-Pixel CNN
- RepPoints: Representative Points
- FSAF: Feature Selective Anchor-Free
- R-FCOS: Redesigned FCOS
- FPN: Feature Pyramid Network
- GloU: Generalized Intersection over Union
- IoU: Intersection over Union
- DeconvNet: Deconvolution Network
- SFC: Same-Resolution Feature Convolution
- MFF: Multi-Resolution Feature Fusion
- FP: Feature Pyramid
- Dconv: Deformable Convolution
- IS: IoU Score
- AP: Average Precision
- FPS: Frames Per Second
- LS-SSDD-v1.0: Large-Scale SAR Ship Detection Dataset-v1.0
- MS COCO: Microsoft Common Objects in COntext
- FCOS+SDR: FCOS with Sample Definition Redesign
- FCOS+SDR+FER: FCOS with SDR and Feature Extraction Redesign
- FCOS+SDR+FER+CRR: FCOS with SDR, FER, and Classification and Regression Redesign
- SSDD: SAR Ship Detection Dataset

**References**

1. Li, M.; Wen, G.; Huang, X.; Li, K.; Lin, S. A Lightweight Detection Model for SAR Aircraft in a Complex Environment. *Remote Sens.*, *2021*, *13*, 5020. [CrossRef]
2. Ghaderpour, E.; Pagiatakis, S.D.; Hassan, Q.K. A Survey on Change Detection and Time Series Analysis with Applications. *Appl. Sci.*, *2021*, *11*, 6141. [CrossRef]
3. Cui, Z.; Qin, Y.; Zhong, Y.; Cao, Z.; Yang, H. Target Detection in High-Resolution SAR Image via Iterating Outliers and Recursing Saliency Depth. *Remote Sens.*, *2021*, *13*, 4315. [CrossRef]
4. Pappas, O.; Achim, A.; Bull, D. Superpixel-Level CFAR Detectors for Ship Detection in SAR Imagery. *IEEE Geosci. Remote Sens. Lett.*, *2018*, *15*, 1397–1401. [CrossRef]
5. He, J.; Wang, Y.; Liu, H.; Wang, N.; Wang, J. A Novel Automatic PolSAR Ship Detection Method Based on Superpixel-Level Local Information Measurement. *IEEE Geosci. Remote Sens. Lett.* 2018, 15, 384–388. [CrossRef]

6. Yang, F.; Xu, Q.; Li, B. Ship Detection From Optical Satellite Images Based on Saliency Segmentation and Structure-LBP Feature. *IEEE Geosci. Remote Sens. Lett.* 2017, 14, 602–606. [CrossRef]

7. Wang, S.; Wang, M.; Yang, S.; Jiao, L. New Hierarchical Saliency Filtering for Fast Ship Detection in High-Resolution SAR Images. *IEEE Trans. Geosci. Remote Sens.* 2017, 55, 351–362. [CrossRef]

8. Song, S.; Xu, B.; Yang, J. SAR Target Recognition via Supervised Discriminative Dictionary Learning and Sparse Representation of the SAR-HOG Feature. *Remote Sens.* 2016, 8, 683. [CrossRef]

9. Li, T.; Liu, Z.; Xie, R.; Ran, L. An Improved Superpixel-Level CFAR Detection Method for Ship Targets in High-Resolution SAR Images. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* 2018, 11, 184–194. [CrossRef]

10. Salembier, P.; Liesegang, S.; Lopez-Martinez, C. Ship Detection in SAR Images Based on Maxtree Representation and Graph Signal Processing. *IEEE Trans. Geosci. Remote Sens.* 2019, 57, 2709–2724. [CrossRef]

11. Lin, H.; Chen, H.; Jin, K.; Zeng, L.; Yang, J. Ship Detection with Superpixel-Level Fisher Vector in High-Resolution SAR Images. *IEEE Geosci. Remote Sens. Lett.* 2020, 17, 247–251. [CrossRef]

12. Wang, X.; Li, G.; Zhang, X.-P.; He, Y. Ship Detection in SAR Images via Local Contrast of Fisher Vectors. *IEEE Trans. Geosci. Remote Sens.* 2020, 58, 6467–6479. [CrossRef]

13. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* 2015, 39, 1137–1149. [CrossRef] [PubMed]

14. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.

15. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.; Berg, A.C. SSD: Single Shot MultiBox Detector. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; pp. 21–37.

16. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollar, P. Focal Loss for Dense Object Detection. *IEEE Trans. Pattern Anal. Mach. Intell.* 2020, 42, 318–327. [CrossRef]

17. 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]

18. Song, J.; Kim, D.-J.; Kang, K.-M. Automated Procurement of Training Data for Machine Learning Algorithm on Ship Detection Using AIS Information. *Remote Sens.* 2020, 12, 1443. [CrossRef]

19. Wei, S.; Zeng, X.; Qu, Q.; Wang, M.; Su, H.; Shi, J. HRSSD: A High-Resolution SAR Images Dataset for Ship Detection and Instance Segmentation. *IEEE Access* 2020, 8, 120234–120254. [CrossRef]

20. Zhang, T.; Zhang, X.; Ke, X.; Zhan, X.; Shi, J.; Wei, S.; Pan, D.; Li, J.; Su, H.; Zhou, Y.; et al. LS-SSDD-v1.0: A Deep Learning Dataset Dedicated to Small Ship Detection from Large-Scale Sentinel-1 SAR Images. *Remote Sens.* 2020, 12, 2997. [CrossRef]

21. 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.

22. Cui, Z.; Li, Q.; Cao, Z.; Liu, N. Dense Attention Pyramid Networks for Multi-Scale Ship Detection in SAR Images. *IEEE Trans. Geosci. Remote Sens.* 2019, 57, 8983–8997. [CrossRef]

23. Zhao, Y.; Zhao, L.; Xiong, B.; Kuang, G. Attention Receptive Pyramid Network for Ship Detection in SAR Images. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* 2020, 13, 2738–2756. [CrossRef]

24. Li, Y.; Zhang, S.; Wang, W.-Q. A Lightweight Faster R-CNN for Ship Detection in SAR Images. *IEEE Geosci. Remote Sens. Lett.* 2020, 1–5. [CrossRef]

25. Hong, Z.; Yang, T.; Tong, X.; Zhang, Y.; Jiang, S.; Zhou, R.; Han, Y.; Wang, J.; Yang, S.; Liu, S. Multi-Scale Ship Detection From SAR and Optical Imagery Via A More Accurate YOLOv3. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* 2021, 14, 6083–6101. [CrossRef]

26. Zhang, K.; Wu, Y.; Wang, J.; Wang, Y.; Wang, Q. Semantic Context-Aware Network for Multiscale Object Detection in Remote Sensing Images. *IEEE Geosci. Remote Sens. Lett.* 2021, 19, 1–5. [CrossRef]

27. Deng, Z.; Sun, H.; Zhou, S.; Zhao, J. Learning Deep Ship Detector in SAR Images from Scratch. *IEEE Trans. Geosci. Remote Sens.* 2019, 57, 4021–4039. [CrossRef]

28. Chen, C.; He, C.; Hu, C.; Pei, H.; Jiao, L. MSARN: A Deep Neural Network Based on an Adaptive Recalibration Mechanism for Multiscale and Arbitrary-Oriented Ship SAR Detection. *IEEE Access* 2019, 7, 159262–159283. [CrossRef]

29. Jin, K.; Chen, Y.; Xu, B.; Yin, J.; Wang, X.; Yang, J. A Patch-to-Pixel Convolutional Neural Network for Small Ship Detection with PolSAR Images. *IEEE Trans. Geosci. Remote Sens.* 2020, 58, 6623–6638. [CrossRef]

30. Gao, S.; Liu, J.M.; Miao, Y.H.; He, Z.J.; Zhao, X.J. A High-Effective Implementation of Ship Detector for SAR Images. *IEEE Geosci. Remote Sens. Lett.* 2021, 19, 1–5. [CrossRef]

31. Yang, Z.; Liu, S.; Hu, H.; Wang, L.; Lin, S. RepPoints: Point Set Representation for Object Detection. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea, 27 October–2 November 2019; pp. 9656–9665.

32. Law, H.; Deng, J. Cornernet: Detecting objects as paired keypoints. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 734–750.

33. Zhou, H.; Wang, D.; Krahenbuhl, P. Objects as Points. *arXiv* 2019, arXiv:1904.07850.
34. Zhang, S.; Chi, C.; Yao, Y.; Lei, Z.; Li, S.Z. Bridging the Gap between Anchor-Based and Anchor-Free Detection via Adaptive Training Sample Selection. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), ATSS, Seattle, WA, USA, 13–19 June 2020; pp. 9756–9765.

35. Tian, Z.; Shen, C.; Chen, H.; He, T. FCOS: A Simple and Strong Anchor-free Object Detector. *IEEE Trans. Pattern Anal. Mach. Intell*. 2020. [CrossRef]

36. Kong, T.; Sun, F.; Liu, H.; Jiang, Y.; Li, L.; Shi, J. FoveaBox: Beyond Anchor-Based Object Detection. *IEEE Trans. Image Process*. 2020, 29, 7389–7398. [CrossRef]

37. Zhu, C.; He, Y.; Savvides, M. Feature Selective Anchor-Free Module for Single-Shot Object Detection. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 16–20 June 2019; pp. 840–849.

38. Zhang, H.; Wang, Y.; Dayoub, F.; Sunderhauf, N. VarifocalNet: An IoU-aware Dense Object Detector. *arXiv* 2021, arXiv:2008.13367.

39. 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*. 2021, 59, 379–391. [CrossRef]

40. Fu, J.; Sun, X.; Wang, Z.; Fu, K. An Anchor-Free Method Based on Feature Balancing and Refinement Network for Multiscale Ship Detection in SAR Images. *IEEE Trans. Geosci. Remote Sens*. 2021, 59, 1331–1344. [CrossRef]

41. Sun, Z.; Dai, M.; Leng, X.; Lei, Y.; Xiong, B.; Ji, K.; Kuang, G. An Anchor-Free Detection Method for Ship Targets in High-Resolution SAR Images. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens*. 2021, 14, 7799–7816. [CrossRef]

42. 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 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 16–20 June 2019; pp. 658–666.

43. Bulat, A.; Tzimiropoulos, G. Binarized Convolutional Landmark Localizers for Human Pose Estimation and Face Alignment with Limited Resources. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 3726–3734.

44. Badrinarayanan, V.; Kendall, A.; Cipolla, R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell*. 2017, 39, 2481–2495. [CrossRef]

45. Noh, H.; Hong, S.; Han, B. Learning Deconvolution Network for Semantic Segmentation. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; pp. 1520–1528.

46. Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. *arXiv* 2018, arXiv:1804.02767.