



## Article

# A CFAR-Enhanced Ship Detector for SAR Images Based on YOLOv5s

Xue Wen, Shaoming Zhang, Jianmei Wang \*, Tangjun Yao and Yan Tang

College of Surveying and Geo-Informatics, Tongji University, Shanghai 200092, China; 2131967@tongji.edu.cn (X.W.); zhangshaoming@tongji.edu.cn (S.Z.); 2111331@tongji.edu.cn (T.Y.); 2211258@tongji.edu.cn (Y.T.)

\* Correspondence: jianmeiw@tongji.edu.cn

**Abstract:** Ship detection and recognition in Synthetic Aperture Radar (SAR) images are crucial for maritime surveillance and traffic management. Limited availability of high-quality datasets hinders in-depth exploration of ship features in complex SAR images. While most existing SAR ship research is primarily based on Convolutional Neural Networks (CNNs), and although deep learning advances SAR image interpretation, it often prioritizes recognition over computational efficiency and underutilizes SAR image prior information. Therefore, this paper proposes YOLOv5s-based ship detection in SAR images. Firstly, for comprehensive detection enhancement, we employ the lightweight YOLOv5s model as the baseline. Secondly, we introduce a sub-net into YOLOv5s, learning traditional features to augment ship feature representation of Constant False Alarm Rate (CFAR). Additionally, we attempt to incorporate frequency-domain information into the channel attention mechanism to further improve detection. Extensive experiments on the Ship Recognition and Detection Dataset (SRSDdV1.0) in complex SAR scenarios confirm our method's 68.04% detection accuracy and 60.25% recall, with a compact 18.51 M model size. Our network surpasses peers in mAP, F1 score, model size, and inference speed, displaying robustness across diverse complex scenes.

**Keywords:** SAR images; ship detection; CFAR; sub-net; frequency-domain information



**Citation:** Wen, X.; Zhang, S.; Wang, J.; Yao, T.; Tang, Y. A CFAR-Enhanced Ship Detector for SAR Images Based on YOLOv5s. *Remote Sens.* **2024**, *16*, 733. <https://doi.org/10.3390/rs16050733>

Academic Editor: Mingsheng Liao

Received: 26 December 2023

Revised: 26 January 2024

Accepted: 7 February 2024

Published: 20 February 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Synthetic Aperture Radar (SAR) leverages the synthetic aperture principle to achieve high-resolution microwave imaging, offering characteristics such as all-weather capability, high resolution, and extensive coverage. Unlike optical remote sensing imagery, SAR enables Earth surface observations without being constrained by weather conditions. Ships, significant subjects in remote sensing images, play a crucial role in various applications, including military surveillance, combating illegal resource exploitation and waterway management [1–3]. The use of SAR imagery for ship detection and identification has become a prominent research focus, posing a central question in this field.

Due to the top-down acquisition of remote sensing images, large image dimensions, and highly complex scenes, strong clutter signals from rough sea surfaces interfere with SAR image-based ship detection. This interference significantly diminishes the performance of ship detection [4]. Various methods have been proposed for SAR ship target detection, including traditional approaches such as CFAR, template matching, and trailing edge detection. These methods often rely on manually designed features, exhibiting limited generalization capabilities. Among them, CFAR algorithms, known for their adaptability and false alarm reduction capabilities, stand as the most widely applied algorithms in ship detection. This algorithm was initially introduced by An et al. [5] and rooted in ship detection from the Ottawa Defense Research Center in 2001 [6]. Subsequently, scientists have made significant enhancements to CFAR detection algorithms from various perspectives, resulting in numerous CFAR-based detection algorithms. However, given the rising volume of satellite remote sensing data and the escalating complexity of SAR scenes, traditional

algorithms relying on manually designed features, such as CFAR operators, are no longer able to meet the demands for detection speed and accuracy.

To address the limitations of traditional methods in complex scenarios, deep learning-based approaches have emerged and achieved remarkable success. Notably, methods based on Convolutional Neural Network (CNN) have demonstrated significant potential in computer vision tasks, giving rise to precise and robust SAR ship target detection techniques. In comparison to traditional model-driven approaches, deep learning-based methods offer advantages such as high automation, fast processing speed, and strong model transferability [7]. According to network architecture, deep learning-based ship detection methods can be categorized into single-stage and two-stage detection methods. Single-stage ship detection methods transform the problem of target localization in a frame into a regression task, eliminating the need for target proposal generation. Two-stage ship detection methods first generate a series of target proposals and subsequently determine the presence of ship targets. In general, single-stage algorithms tend to exhibit lower detection accuracy than two-stage algorithms, but the former often feature lighter networks and faster detection speeds.

The SAR ship detection model based on CNN has achieved significant performance improvement [8]. Unfortunately, it has widely abandoned traditional manual features. However, we believe that blindly abandoning these features is unwise, as they not only possess elegant functionality but also exhibit advantages in multiple sensor and diverse scene applications through fine-tuning. The interpretability of these manual features provides decision transparency for SAR target recognition technology, aiding in mitigating potential decision risks in high-risk applications, such as military reconnaissance and precision strikes, and gaining user trust. Also, despite the success of CNN in computer vision, in SAR ship detection, SAR images exhibit rich noise and complex backgrounds, while ship targets are often densely distributed, varied in size, and oriented arbitrarily. The “black box” nature of neural networks limits reliability and credibility, particularly in high-risk scenarios like military reconnaissance, where user understanding and trust in decision making are crucial. Additionally, most deep learning methods solely utilize spatial domain information from SAR images, neglecting frequency domain information, which results in decreased detection performance in sea clutter scenarios [9]. Consequently, despite the encouraging outcomes achieved by current deep learning-based SAR ship detection methods, there remains substantial room for improvement in terms of detection accuracy and efficiency.

To enhance ship target detection performance in complex scenes within SAR images, this paper proposes a rotation-based ship detection method based on the YOLOv5s [10] framework. This method prioritizes addressing challenges such as ship target rotation, noise, and signal interference in SAR images while maintaining detection speed. Our approach innovatively combines deep learning with traditional handcrafted features by introducing sub-net into YOLOv5s to learn CFAR features, thereby enhancing the representation capability of ship targets. Additionally, we incorporate frequency domain information into the channel attention mechanism using FcaNet [11] to further improve ship detection performance. Furthermore, we enable ship target detection in arbitrary orientations, enhancing detection precision. We conducted extensive experiments on publicly available SAR ship detection datasets, and the results demonstrate significant improvements in accuracy and recall compared to existing methods. Our method exhibits robustness in various complex scenarios. These experimental findings suggest the potential practical value of our approach in the field of SAR ship detection.

In summary, the main contributions of this paper can be summarized as follows:

- (1) Aiming at the complex backgrounds encountered in SAR images, we propose an end-to-end network structure based on a single-stage object detection algorithm. This network achieves high ship detection accuracy while maintaining a fast speed. We incorporate handcrafted feature extraction and attention mechanisms into the network, ensuring the effectiveness of ship detection in SAR images.

- (2) We design a sub-net that supervises feature extraction in the main network, helping our model learn more handcrafted features and highlighting the differences between ships and backgrounds, thereby overcoming the challenges of ship detection in complex backgrounds.

The remaining sections of this paper are organized as follows. Section 2 introduces related work. The method is described in Section 3. Experimental results and ablation studies are presented in Section 4. Finally, Section 5 summarizes the paper.

## 2. Related Work

### 2.1. Handcraft Feature-Based Methods

Handcraft feature-based methods can be classified into three categories: using polarization features, geometric features, and backscatter features.

Methods utilizing polarization characteristics [12–14] distinguish targets from the background by exploiting the scattering difference between ships and sea clutter, achieving ship detection. However, this approach requires accurate scattering models, which poses significant challenges. Additionally, this method is highly sensitive to the detection environment and prone to interference from sea clutter, leading to severe performance degradation.

Geometric feature-based detection methods [15] employ artificial designs to capture the shape, size, texture, and other characteristics of ship targets. This involves constructing a template library for template matching and ship detection. However, this approach involves pixel-wise matching between the entire SAR image and the designed templates, resulting in high computational costs and slow detection speed. Furthermore, the performance of this method heavily relies on the quality of the designed templates, making it expensive and highly dependent on expert experience, ultimately resulting in poor robustness. Moreover, the presence of sea clutter greatly affects the accurate matching between templates and ship targets.

A group of methods based on backscatter features widely employs CFAR detection. These algorithms demonstrate reliability across diverse environments and applications, requiring minimal human intervention. The CFAR algorithm was initially introduced by An et al. [5] and rooted in ship detection at the Ottawa Defense Research Center in 2001 [6]. Subsequently, scientists have made significant improvements to CFAR detection algorithms from various perspectives, resulting in numerous CFAR-based detection algorithms. Leng et al. [16] introduced a bilateral CFAR algorithm that combines SAR image intensity and spatial distribution, reducing the blurred effects caused by the SAR platform and sea clutter. In simple scenarios, CFAR methods yield favorable results. However, for small ships and complex maritime scenes, modeling challenges result in higher false alarm rates and inferior detection performance. With the increasing volume of satellite remote sensing data and the growing complexity of SAR scenes, traditional ship detection methods based on handcrafted features can only leverage simple low-level features of SAR images. They lack generalization, rendering them unsuitable for complex SAR image detection tasks.

### 2.2. Deep Learning-Based Methods

To address the challenge of insufficient generalization, deep learning-based methods have emerged and achieved significant success. Particularly, methods based on Convolutional Neural Networks (CNNs) have shown substantial potential in computer vision tasks, thereby advancing the precise and robust detection of SAR ship targets. Compared to traditional model-driven approaches, deep learning-based methods offer advantages such as high automation, rapid processing speed, and strong model transferability. Based on network architecture, deep learning-based ship detection methods can be categorized into two-stage and one-stage detection methods.

Two-stage ship detection methods generate target proposals before confirming ship presence, involving classification, regression, and segmentation tasks with a connected segmentation sub-network, enhancing accuracy. RCNN [17] paved the way, followed

by Fast R-CNN [18] in 2015 and Faster R-CNN [19] in the same year. In 2016, Dai et al., introduced the region-based fully convolutional network (R-fcn) [20] with location-sensitive ROI pooling.

Despite the potential advantages mentioned earlier, two-stage methods pose challenges with proposal generation and multiple tasks, leading to complexity and computational overhead. In response, single-stage methods like SSD [21] emerged, directly detecting targets from densely sampled anchor points, utilizing uniform dense sampling and strategies like aspect ratios and scales. The YOLO series [10,22,23] exemplifies classic single-stage algorithms, consistently improving performance.

Deep learning-based SAR ship detection, whether two-stage or single-stage, often features large models and deep architectures [24]. Zhang et al. [25] proposed a lightweight SAR ship detector, “ShipDeNet-20”, which is several tens to even hundreds of times lighter than other detectors. This contributes to real-time SAR applications and future hardware implementations. Despite advancements, methods predominantly use spatial domain information, neglecting the potential of frequency domain information. In addition, Liu et al. [26] highlighted challenges in feature extraction from the horizontal region of interest (HROI) in remote sensing images. While these methods show promise, there is room for improvement in terms of accuracy and efficiency.

### 2.3. Fusion-Based Methods

In recent years, the integration of handcrafted traditional features with deep learning has become a crucial research direction [27,28]. In the domain of SAR image ship detection, based on the differences in the fusion approaches, three primary methods are prominent: the two-stage fusion method, direct embedding method, and multi-branch fusion method.

The two-stage fusion method employs a two-stage processing pipeline that combines traditional operators with neural networks, demonstrating effectiveness [29]. Through initial preprocessing, leveraging traditional operators such as CFAR, classical operations are performed on SAR images for feature extraction. However, this method may lose information when dealing with complex electromagnetic scattering scenarios. Additionally, the two-stage fusion method introduces computational overhead, impacting the feasibility of real-time applications.

The direct embedding method directly embeds traditional features into neural networks, either at the input layer or intermediate layers. This method offers advantages but also presents challenges. For instance, introducing traditional features at the input layer, as demonstrated by MSRIHL-CNN [30], enhances the combination of low-level texture and deep features but may face difficulties in complex pattern recognition of high-level features. Embedding traditional features into intermediate layers, as shown by HOG-ShipCLSNet [31], requires a delicate balance between traditional and deep features for optimal performance.

The multi-branch fusion method designs a multi-branch structure to incorporate traditional features into the network. The innovation of such methods aims to organically fuse traditional and deep learning features, with each branch focusing on different types of features. The decision to use the multi-branch approach is based on its capability to comprehensively capture various target features. However, it is essential to recognize the potential for further improvement. Practices like collaborative tasks in MTL-Det [32] may introduce complexity during training and require a substantial amount of labeled data. Similarly, the effective yet optimizable feature fusion method in a single-stage ship detection network [33] could further be optimized to reduce computational demands and enhance real-time applicability.

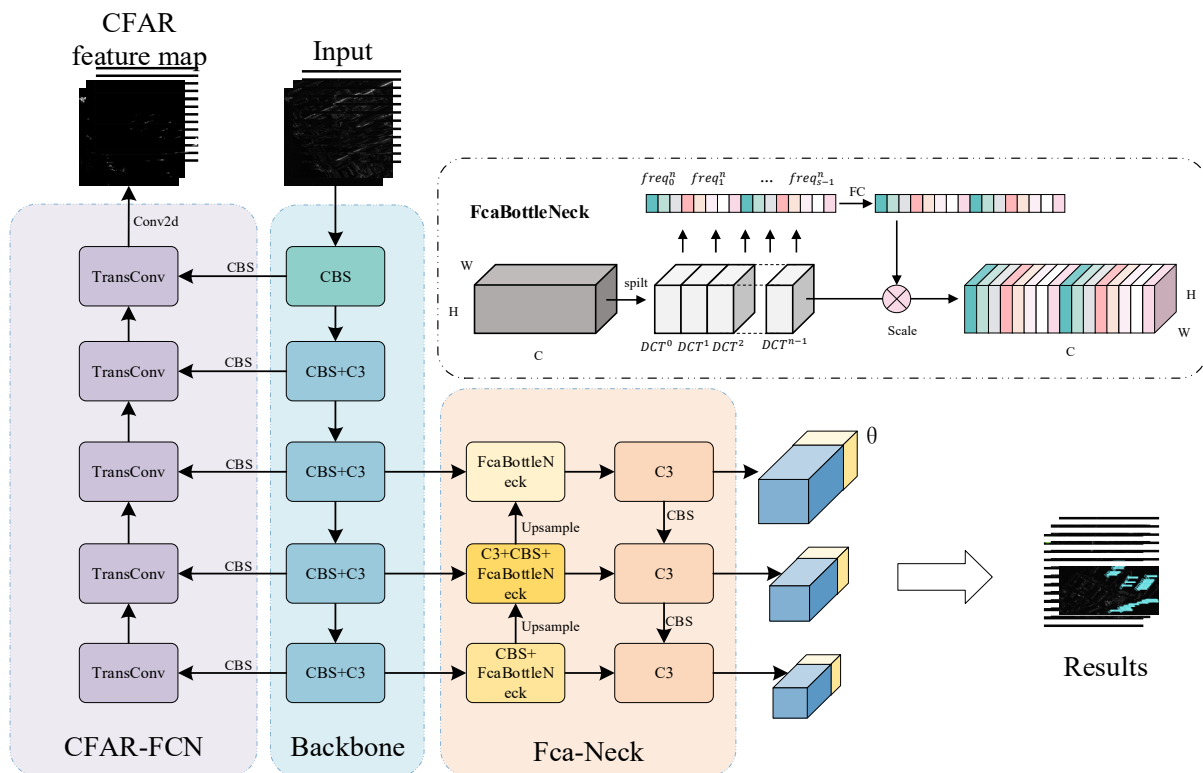
To achieve a balance between the speed and accuracy of ship detection in SAR images, this study is anchored in the YOLOv5s single-stage network. Simultaneously, by harnessing the comprehensive feature-capturing capability of the multi-branch network approach, sub-net is incorporated beyond the backbone network to seamlessly integrate traditional CFAR features into deep learning methods.

### 3. Method

In this section, we provide a detailed exposition of the architecture of the proposed network designed for achieving fast and accurate ship detection in SAR images.

#### 3.1. The Overall Framework

Figure 1 illustrates the overall structure of the method proposed in this paper. We have chosen the YOLOv5s model as the base network. Unlike other detection networks in the YOLO series, YOLOv5 has not been formally introduced in relevant literature. Nevertheless, the detection speed of YOLOv5 is notably faster than any previous versions. Furthermore, this network exhibits robust performance in detecting multi-scale and small objects. The model comprises four versions: s, m, l, and x, among which the s version is most commonly used. This architecture is relatively concise and allows for faster execution and inference while maintaining accuracy. Our network primarily consists of three components: the backbone network, CFAR feature constraint sub-net (CFAR-FCN), and FcaNet Channel Attention Module (Fca-Neck).



**Figure 1.** The Overall Framework. CFAR-FCN stands for CFAR feature constraint sub-net; Backbone represents the feature extractor; Fca-Neck is the FcaNet Channel Attention Module.

Firstly, the input image is resized to an appropriate size (e.g., 512). We employ the backbone of YOLOv5 as the main structure of the overall network and incorporate the FcaNet bottleneck into the neck to enhance low-frequency domain features. Furthermore, we establish a sub-net mechanism to constrain and incentivize feature learning in the backbone network. In particular, we treat the backbone part of the main network as the “contracting path”, where feature map sizes decrease layer by layer while channel counts increase. Subsequently, we connect this with an “expanding path” consisting of multiple layers of deconvolutional layers to output feature maps, which are supervised using CFAR feature maps. CFAR supervision allows us to obtain more meaningful semantic information from upsampled feature maps and finer-grained information from early traditional feature maps.

Our network predicts three different scales of bounding boxes. After feature extraction by the backbone network, we further extract features based on the fused feature map using Fca-Neck with the channel attention mechanism. Finally, multiple convolutional layers are connected after Fca-Neck in each branch, and three different scales of outputs are used to predict rotated boxes, including the center point, height, width, and rotation angle. In this method, we treat the angle prediction as an output with 180 categories. Therefore, for four bounding box offsets, one object confidence, six categories, and 180 angles, the predicted tensor size is  $N \times M \times [3 \times (4 + 1 + 6 + 180)]$ , where  $N$  and  $M$  represent the height and width of the tensor, respectively. The final detection results are obtained after applying Non-Maximum Suppression (NMS) filtering to the predicted boxes.

### 3.2. CFAR-FCN

The objective of ship target detection in SAR ship images primarily emphasizes high detection rates and low false alarm rates. The CFAR algorithm achieves target detection by comparing pixel grayscale values with a threshold within a specific region, effectively representing the electromagnetic scattering characteristics of targets in the form of images and providing rich target information. In this paper, we construct a multi-task network structure with two network branches. One serves as the main branch for ship target detection and regression, while the other branch acts as a sub-net to constrain and supervise the backbone network, enabling it to learn the spatial features of ship targets present in CFAR feature maps. The CFAR-FCN is not involved in inference.

As a typical image segmentation network, U-Net [34] has been widely applied in the field of SAR image water body segmentation in recent years. Considering the cost and constraints of SAR image acquisition, obtaining a large-scale dataset is relatively challenging. U-Net, through jump connections, propagates shallow feature information to deeper layers, merging relevant feature information, making the network model more suitable for small-sample SAR dataset segmentation tasks. Therefore, in this paper, we refer to the overall network architecture of U-Net and construct a sub-net for enhancing ship target features, as shown in Figure 1.

This sub-net consists of a contracting path and an expanding path. In this module, the contracting path corresponds to the backbone network and follows the typical architecture of a convolutional network. Each block in the expanding path includes a  $2 \times 2$  convolution (“deconvolution”) that reduces the number of feature channels by half. It is connected with the modules in the contracting path through convolution, followed by an activation function. In the final layer, a  $1 \times 1$  2D convolution reduces the number of output channels to match the number of channels in the feature map. Ultimately, this branch outputs a feature map of size  $512 \times 512 \times 3$  (taking an input image size of 512 as an example). The CFAR operators are used in the filtering results of each SAR image, yielding CFAR feature maps. During the training phase, the loss is quantified by computing the Euclidean distance between the CFAR feature maps and the output of this branch, followed by backpropagation to update the model parameters. Through the gradient backpropagation of this branch, the backbone network learns CFAR features, endowing this branch with the capability to simulate CFAR feature maps.

### 3.3. Fca-Neck

Differing from conventional optical images, the imaging of objects in SAR images is solely related to their radar signal reflectivity, resulting in relatively monotonous information content, low resolution, and a strong correlation with the image signal-to-noise ratio. On the other hand, SAR images exhibit strong speckle noise, which interferes with target identification and detection. Furthermore, in the application of attention mechanisms in typical neural networks, there is a tendency to overlook the extraction and collaboration of frequency domain information. We attempt to exploit the frequency domain features of SAR images.

We use the two-dimensional Discrete Cosine Transform (2D-DCT) as weights to aggregate frequency information based on an attention mechanism. This yields the frequency band with the highest energy aggregation, which helps refine dense multi-target feature maps, reducing false alarms and enhancing the accuracy of dense multi-target detection.

The typical basis functions of two-dimensional (2D) DCT [35] are as

$$B_{h,w}^{i,j} = \cos\left(\frac{\pi h}{H}\left(i + \frac{1}{2}\right)\right) \cos\left(\frac{\pi w}{W}\left(j + \frac{1}{2}\right)\right). \quad (1)$$

And 2D-DCT can be represented as

$$\begin{aligned} f_{h,w}^{2d} &= \sum_{i=0}^{H-1} \sum_{j=0}^{W-1} x_{i,j}^{2d} B_{h,w}^{i,j} \\ \text{s.t. } h &\in 0, 1, \dots, H-1, w \in 0, 1, \dots, W-1 \end{aligned} \quad (2)$$

where  $f^{2d} \in \mathbb{R}^{H \times W}$  is the 2D DCT spectrum,  $x^{2d} \in \mathbb{R}^{H \times W}$  is the input,  $H$  is the height of  $x^{2d}$ , and  $W$  is the width of  $x^{2d}$ .

Channel attention mechanisms are widely used in CNNs. They employ scalars to represent and assess the importance of each channel. Let  $X \in \mathbb{R}^{H \times W \times C}$  be the image feature tensor in the network,  $C$  be the number of channels,  $H$  be the height of the feature, and  $W$  be the width of the feature. The scalar in channel attention is often seen as a compression problem because it must represent the entire channel, and therefore, only a single scalar can be used. Thus, the attention mechanism can be written as

$$att = \text{sigmoid}\left(\text{fc}\left(\text{compress}\left(x^{2d}\right)\right)\right), \quad (3)$$

where  $att \in \mathbb{R}^C$  is the attention vector,  $\text{sigmoid}$  is the function of  $\text{sigmoid}$ ,  $\text{fc}$  is a mapping function like a fully connected layer or 1D convolution, and  $\text{compress} : \mathbb{R}^{C \times H \times W} \mapsto \mathbb{R}^C$  is a compression method. After obtaining attention vectors for all  $C$  channels, each channel of the input  $x^{2d}$  is scaled by the corresponding attention value,

$$\tilde{x}_{:,i,:}^{2d} = att_i x_{:,i,:}^{2d}, \text{ s.t. } i \in 0, 1, \dots, C-1, \quad (4)$$

where  $\tilde{X}$  is the output of the attention mechanism,  $att_i$  is the  $i$ th element of the attention vector, and  $X_{:,i,:}$  is the  $i$ th input channel. In this paper, we employ FcaNet [11] for weight allocation, as detailed in Figure 1. In contrast to the global average pooling (GAP) used in typical networks, FcaNet extends GAP to 2D discrete cosine transform (2DDCT). In FcaNet, GAP is considered as a special case of 2DDCT, with its result being directly proportional to the lowest-frequency component of the 2DDCT. Therefore, utilizing FcaNet enables more efficient aggregation of frequency domain information.

### 3.4. Loss Function

This network employs a multi-task loss to optimize the supervisory branch and the final prediction network. The loss function consists of three parts: the CFAR-FCN branch loss, the original YOLO loss function, and the angle classification prediction loss. The formulas are as

$$L_{all} = \lambda_1 L_{box} + \lambda_2 L_{obj} + \lambda_3 L_{cls} + \lambda_4 L_{CFAR} + \lambda_5 L_{angle}, \quad (5)$$

where  $L_{box}$ ,  $L_{obj}$ ,  $L_{cls}$ ,  $L_{CFAR}$ , and  $L_{angle}$  represent the detection box regression loss, the confidence loss, the classification loss, the CFAR branch loss, and the angle classification loss, respectively.  $\lambda_n$  are the corresponding weight coefficients that control the importance of these losses, with a default setting of 1.

For the loss function related to the regression of detection box parameters, the complete-IoU ( $CIoU_{Loss}$ ) [36] is used for calculation. The specific formula for calculation is as

$$L_{box} = \sum_i^{N_p} CIoU_{Loss}(P_{box}, T_{box}), \quad (6)$$

where  $N_p$  represents the number of prediction layers (with a default value of 3),  $P_{box} \in R^{N_t \times (x_c, y_c, w, h)}$  represents the bounding boxes predicted by the model,  $T_{box} \in R^{N_t \times (x_c, y_c, w, h)}$  represents the corresponding ground truth bounding boxes, and  $N_t$  represents the number of ship targets.

The losses  $L_{obj}$ ,  $L_{cls}$ , and  $L_{angle}$  are computed using binary cross-entropy (BCE) logits loss, *BCEWithLogits* named  $L_{BCEL}$ , defined as:

$$L_{BCEL} = -\sum_{n=1}^N [x_i^* \log(\delta(x)) + (1 - x_i^*) \log(\delta(1 - x))]. \quad (7)$$

The specific formulas for  $L_{obj}$ ,  $L_{cls}$ , and  $L_{angle}$  are as follows:

$$L_{obj} = \sum_i^{N_p} BCEL(P_{obj}, T_{obj}), \quad (8)$$

$$L_{cls} = \sum_i^{N_p} BCEL(P_{cls}, T_{cls}), \quad (9)$$

$$L_{angle} = \sum_i^{N_p} BCEL(P_{\theta}, T_{\theta}), \quad (10)$$

where  $P_{obj} \in R^{N_p \times W_i \times H_i}$  represents the predicted offset vector,  $T_{obj} \in R^{N_p \times W_i \times H_i}$  is the ground truth vector,  $W_i$  ( $i = 1, 2, 3$ ) is the width of the predicted layer feature map, and  $H_i$  ( $i = 1, 2, 3$ ) is the height of the predicted layer feature map.  $P_{cls} \in R^{N_t \times N_c}$  represents the probability distribution of various class predictions,  $T_{cls} \in R^{N_t \times N_c}$  is the ground truth probability distribution, and  $N_c$  is the number of ship categories (default value is 1).

In typical angle classification, angles are usually treated as categories to calculate the loss function, neglecting the relationship between angles. In this paper,  $P_{\theta} \in R^{N_t \times L_{angle}}$  and  $T_{\theta} \in R^{N_t \times L_{angle}}$  represent the CSL-adjusted angle labels and predictions, respectively.

Furthermore, to encourage the sub-net to learn features from the CFAR maps effectively and thus provide supervision to the detection backbone network, the loss function for the CFAR-FCN branch is calculated using Euclidean distance.

#### 4. Experiments and Analysis

In this section, we evaluate the detection performance of the proposed method through experiments. First, we introduce the dataset and relevant experimental settings and evaluation criteria. Then, we compare this method with other CNN-based rotation ship detection methods to demonstrate its superiority.

##### 4.1. Dataset and Experimental Settings

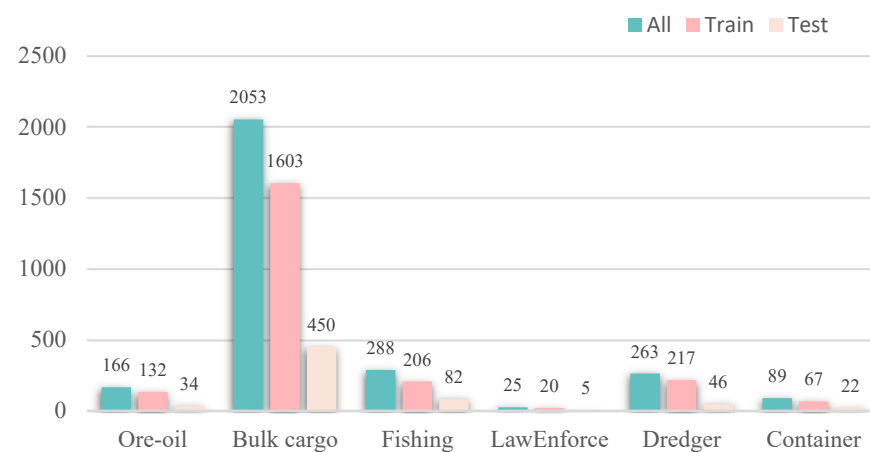
###### 4.1.1. Dataset

Lei et al. [37] released the SRSDD-v1.0 dataset in 2021, which was constructed from more than 30 large-scene images captured from five locations using China's GF-3 satellite, with an image resolution of 1 m. These large-scene images were cropped into small patches of  $1024 \times 1024$  pixels each. The dataset contains a total of 666 images, all of which are cut from the original s, and container ships. The number of images including land cover is 420, which contain 2275 ships. The number of images with only the sea in the background is 246, which contain 609 ships. It is worth noting that this dataset is the first publicly available SAR ship dataset with varying resolutions and classifications. Therefore, we conduct ship detection and recognition experiments based on this dataset. Table 1 provides additional details about SRSDD.

**Table 1.** The basic parameters of SRSDD.

Parameter	Value
Number of Images	666
Waveband	C
Image Size	1024 × 1024
Image Mode	Spotlight Mode
Polarization	HH, VV
Resolution (m)	1
Ship Classes	6
Position	Nanjing, Hongkong, Zhoushan, Macao, Yokohama

The ships in SRSDD are annotated with rotated bounding boxes and categorized into six classes: oil tankers, bulk carriers, fishing boats, law enforcement vessels, dredgers, and container ships. Experts inspected the corresponding SAR images and optical images, providing rotated bounding boxes and class labels for each ship target. This ensures their authenticity and accuracy. Figure 2 displays the quantity for each class. From the graph, it can be observed that bulk carriers constitute the majority of the dataset, while law enforcement vessels make up almost one-tenth of the bulk carriers. The overall distribution of the training and testing datasets is consistent. Additionally, considering the issue of having more offshore scenes than nearshore scenes in existing SAR datasets, SRSDD places emphasis on selecting nearshore scenes during sampling. In SRSDD, nearshore scenes account for 63.1%, while offshore scenes make up 36.9%. To ensure the fairness of experimental results, our experiments follow the same protocol as in the literature [37], which uses 532 images for training and 134 images for testing. In both the training and testing sets, the categories, widths, heights, and angle distributions of the ships are similar to ensure the effectiveness of the test results. The specific distribution of ship quantities can be seen in Figure 2.

**Figure 2.** The quantity distributions of six categories of ships in SRSDD-v1.0 dataset.

#### 4.1.2. Implementation

All experiments in this paper were conducted on a PC with an Intel® Core™ i7-10875H CPU @2.30 GHz × 16 and a GeForce RTX 2060 Mobile GPU, using PyTorch for implementation. The operating system was Ubuntu, and the CUDA version was 11.7. We employed the stochastic gradient descent (SGD) algorithm as the optimizer, with each minibatch containing 20 images. Weight updates for the network used an initial learning rate of  $1 \times 10^{-2}$ , weight decay of  $5 \times 10^{-4}$ , and momentum of 0.937. Lastly, we utilized rotation non-maximum suppression (R-NMS) with an IoU threshold of 0.5 to remove duplicate detections and set the confidence threshold to 0.3. The computer and deep learning environment configurations for our experiments are presented in Table 2.

**Table 2.** Experimental setup and environment.

Project	Model/Parameter
CPU	Intel® Core™ 7-10875H
RAM	16 GB
GPU	GeForce RTX 2060 Mobile
SYSTEM	Ubuntu22.4
CODE	Python3.8
FRAMEWORK	CUDA11.7/torch 1.13

#### 4.1.3. Metrics

To evaluate the detection performance of our method, we employed four evaluation metrics: precision, recall, F1 score, and average precision (AP). Precision and recall are defined as follows:

$$\text{Precision} = \frac{TP}{TP + FP}, \quad (11)$$

$$\text{Recall} = \frac{TP}{TP + FN}, \quad (12)$$

where  $TP$  represents the number of true-positive detections,  $FP$  represents the number of false positives, and  $FN$  represents the number of ships that were not detected. Precision is defined as the proportion of true ship detections among all positive predictions made by the algorithm. Precision and recall are typically negatively correlated, meaning that as one goes up, the other tends to go down. A higher precision indicates fewer false positives, while a higher recall means fewer false negatives. The F1 score is used to provide a balanced measure of both precision and recall and is defined as

$$F1 = 2 \times \frac{P \times R}{P + R}. \quad (13)$$

Furthermore, we can obtain the precision–recall curve and calculate the mean Average Precision ( $mAP$ ):

$$AP = \int_0^1 P(R) dR, \quad (14)$$

$$mAP = \frac{1}{k} \sum_{i=1}^k AP_i \quad (15)$$

where  $k$  is the total number of target categories, and  $P(R)$  is the precision–recall curve, which is obtained from the four components determined in information retrieval: true positives ( $TP$ ), true negatives ( $TN$ ), false positives ( $FP$ ), and false negatives ( $FN$ ). Additionally, it is set that a detected target in a bounding box is considered a correct ship when the IoU between the bounding box and a single ground truth is greater than 0.5.

## 4.2. Experimental Results

### 4.2.1. Quantitative Analysis of Results

To evaluate the comprehensive performance of our method, we compared the detection results of our proposed model with those of eight other rotation detectors, as shown in Table 3. Labels C1–C6 correspond to bulk carriers, fishing vessels, law enforcement vessels, dredgers, general cargo ships, and container ships, respectively. The detection results of other methods are sourced from [37]. From the results, it can be observed that among the two-stage detection networks, O-RCNN [38] exhibits the best overall detection performance on the SRSDD dataset, with  $mAP$  and F1 scores of 56.23% and 60.64, respectively. In the single-stage detection models, BBAVectors [39] and R-FCOS [37,40] show the best overall performance, but their  $mAP$  values are both less than 50, and F1 scores are below 45. In contrast, our proposed model achieves a  $mAP$  of 61.07 and an F1 score of 63.91, significantly outperforming BBAVectors and R-FCOS. Moreover, our method processes the SRSDD

dataset almost six times faster than other methods, with a model size of only 18.51M, which is notably better than the original lightweight model R-FCOS with 244M [40].

**Table 3.** Detection results of different CNN-based methods on SRSDD-v1.0.

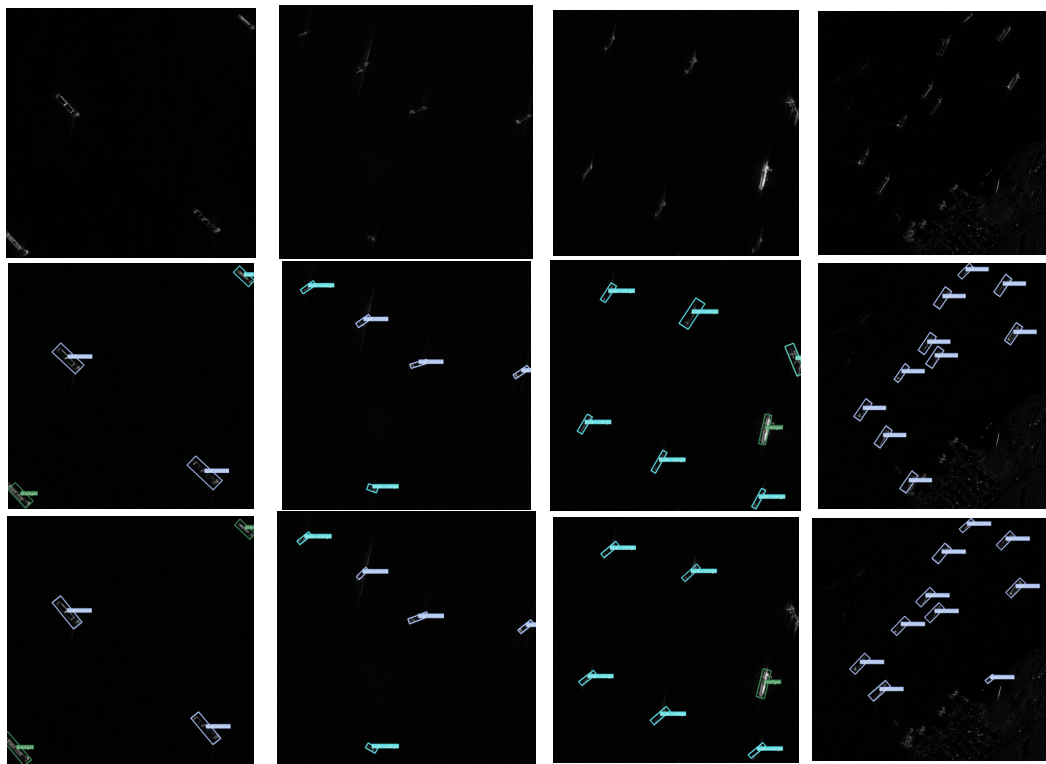
Model	Category	Precision (%)	Recall (%)	mAP	F1	FPS	Model (M)
FR-O [37]	Two-stage	57.12	49.66	53.93	53.13	8.09	315
ROI [37,41]	Two-stage	59.31	51.22	54.38	54.97	7.75	421
Gliding Vertex [37,42]	Two-stage	57.75	53.95	51.50	55.79	7.58	315
O-RCNN [37,38]	Two-stage	64.01	57.61	56.23	60.64	8.38	315
R-RetinaNet [37]	One-stage	53.52	12.55	32.73	20.33	10.53	277
R3Det [37,43]	One-stage	58.06	15.41	39.12	24.36	7.69	468
BBAVectors [37,39]	One-stage	50.08	34.56	45.33	40.90	3.26	829
R-FCOS [37,40]	One-stage	60.56	18.42	49.49	28.25	10.15	244
our method	One-stage	<b>68.04</b>	<b>60.25</b>	<b>61.07</b>	<b>63.91</b>	<b>56.18</b>	<b>18.51 *</b>

\* With the best results highlighted in bold.

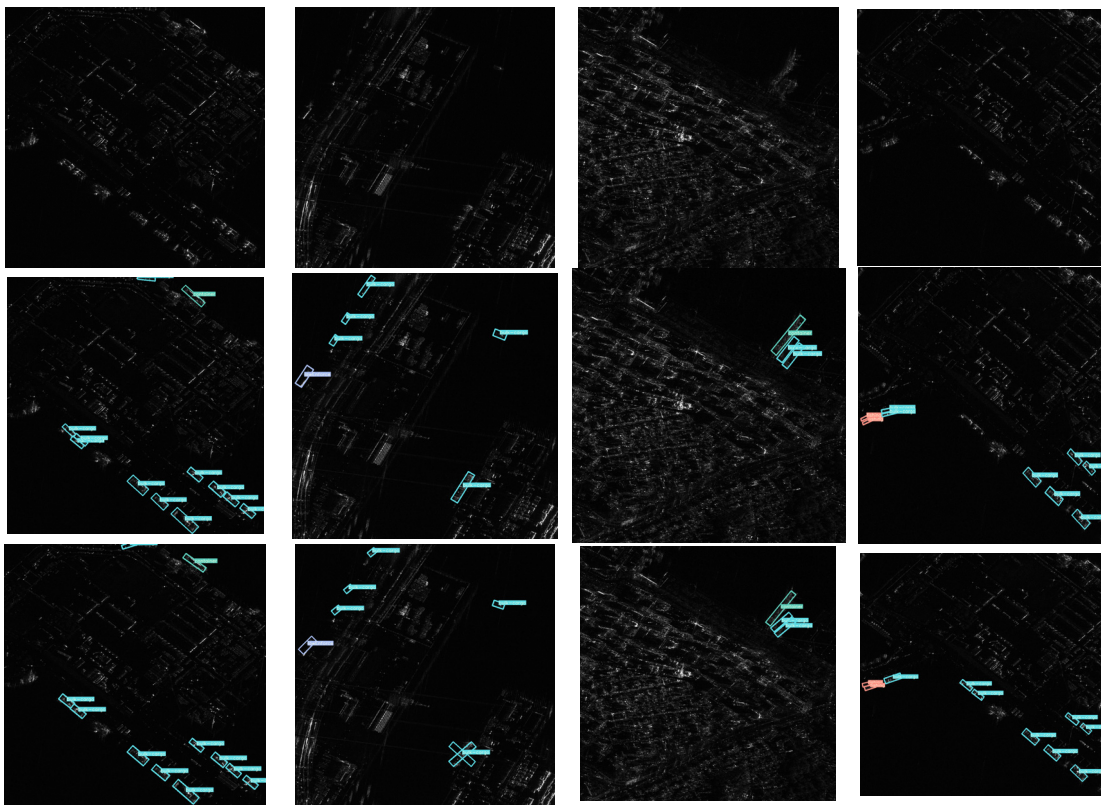
The results indicate that our model can effectively detect ship targets in complex SAR images. In terms of operational efficiency across different models, two-stage detection models achieve a maximum FPS of only 8.38 when processing images on the SRSDD dataset. Among the other single-stage detection models, R-RetinaNet [37] is the fastest, but its FPS is only 10.53. In contrast, as shown in Table 3, our proposed model achieves an FPS value of 56.18, demonstrating a significant advantage in processing efficiency over other single-stage or two-stage detection models.

#### 4.2.2. Qualitative Analysis of Results

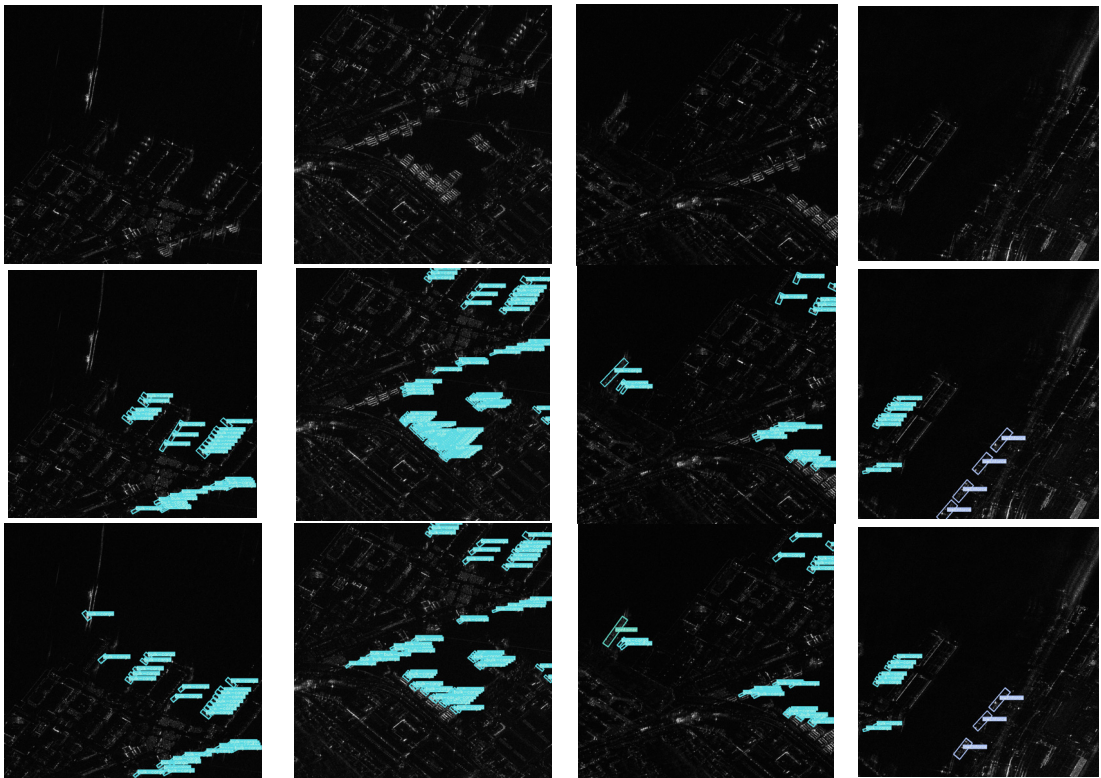
Furthermore, we visualize the detection and recognition results of our proposed model in nearshore and offshore scenarios in the SRSDD dataset, as shown in Figures 3–5.



**Figure 3.** Detection results in offshore scenes. The first row is the original SAR images, the second row is the ground truth, and the third row is the OBB prediction results.



**Figure 4.** Detection results in inshore scenes. The first row is the original SAR images, the second row is the ground truth, and the third row is the OBB prediction results.



**Figure 5.** Detection results in dense array scenes. The first row is the original SAR images, the second row is the ground truth, and the third row is the OBB prediction results.

Figure 3 presents the detection and classification results for ships at a distance from the shore. From the SAR images, it is evident that our network successfully suppresses false alarms in SAR images and can effectively suppress interference noise that closely resembles ship characteristics. The detection results highlight our network's strong adaptability to different scenes. Due to the utilization of the CFAR-FCN in our model, it enhances ship features while suppressing noise, making the network more robust and reducing the impact of noise on detection and classification results to some extent.

Figure 4 displays the detection and classification results for ships in nearshore scenarios. From the image, it is evident that our network successfully detects and classifies ship targets of different categories, even when complex backgrounds dominate most of the picture. This is attributed to our network's ability to suppress land features, thereby enhancing the accuracy of nearshore ship detection.

Figure 5 presents the detection and classification results for densely arranged ship scenarios. Typically, due to complex backgrounds, false alarms are a common occurrence in the detection results for onshore ships. Furthermore, the dense arrangement of nearshore ships poses a challenge to detection and classification. From the detection results, it is evident that our network accurately detects and classifies densely arranged ships in nearshore scenes while suppressing false alarms in these scenarios. This is attributed to our adoption of a rotation detection mechanism, which yields improved regression results for ship detection boxes in nearshore settings.

#### 4.2.3. Ablation Experiments

In this section, we conducted a series of experiments to validate the effectiveness of key improvements in the network. The comparative results of each module are shown in Table 4. These improvements include the CFAR-FCN branch and the frequency domain Fca-Neck module. Furthermore, we qualitatively explained the improvements brought about by these introduced modules based on the test results. The results show that the addition of these improvements gradually enhances the network's detection accuracy.

**Table 4.** Comparative analysis of the effects of each module in ablation experiments.

Model	Precision (%)	Recall (%)	mAP	F1	FPS	Param(M)	Model (M)
YOLOv5s + CFAR + low	68.04	<b>60.25</b>	<b>61.07</b> *	<b>63.91</b>	56.18	9.52	18.51
YOLOv5s + CFAR	<b>69.84</b>	56.65	57.56	62.56	60.98	8.98	17.48
YOLOv5s + low	67.21	58.06	56.49	62.3	78.74	7.84	15.7
YOLOv5s	68.57	56.34	52.04	61.86	<b>84.75</b>	<b>7.51</b>	<b>14.67</b>

\* With the best results highlighted in bold. YOLOv5s means the normal YOLOv5s with rotated detector; CFAR means CFAR-FCN with CFAR feature map; low means low-frequency information.

- Effectiveness of CFAR-FCN

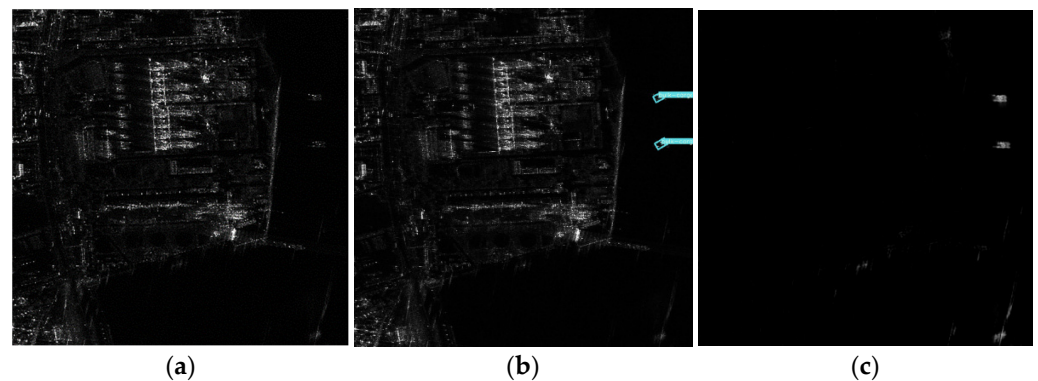
Table 5 presents the results of ablation studies on the CFAR-FCN network branch. We compared a series of detection results. Except for the CFAR-FCN network part, the three networks used in the experiments were identical. One network employed YOLOv5s+CFAR feature maps, using feature maps filtered by the CFAR operator as the output labels for the CFAR-FCN branch. Another network used YOLOv5s+shipseg segmentation maps, where segmentation maps that segmented ship targets and assigned 0 to other areas were used as the output labels for the network branch. The last network was our base network, which is the normal YOLOv5s with rotated detector. The results indicate that using the CFAR-FCN network can improve the accuracy of detection and classification to a certain extent, effectively enhancing the overall detection performance of the network. Leveraging the CFAR-FCN network branch allows the fusion of CFAR ship features from traditional handcrafted operators, highlighting ship information, thus improving detection and classification accuracy. Additionally, from the experimental results, it can be seen that CFAR feature maps outperform regular ship segmentation maps in terms of detection performance. This is because handcrafted operators enhance effective ship features.

**Table 5.** Effectiveness of CFAR-FCN.

Model	Precision (%)	Recall (%)	mAP	F1	FPS	Param (M)	Model (M)
YOLOv5s + CFAR	<b>69.84</b>	<b>56.65</b>	<b>57.56</b> *	<b>62.56</b>	60.98	8.98	17.48
YOLOv5s + shipseg	65.86	55.56	40.09	60.27	63.69	8.98	17.48
YOLOv5s	68.57	56.34	52.04	61.86	<b>84.75</b>	<b>7.51</b>	<b>14.67</b>

\* With the best results highlighted in bold.

Additionally, from Figure 6, the sub-net of this network exhibits a suppressive effect on land. This phenomenon arises from the fact that in this experiment, the CFAR feature map is obtained after land and water segmentation, that is, land is assigned as 0, and then the CFAR operator is used for the filtering calculation. In other words, the land portions in the CFAR map input to the network are all assigned a value of 0. This significantly reduces the interference caused by strong signal reflection points on land to the network's ship detection.



**Figure 6.** Effectiveness of CFAR-FCN. (a) Ground truth; (b) OBB prediction results; (c) Output of CFAR-FCN.

- Effectiveness of Fca-Neck

Table 6 presents the results of ablative studies on the frequency-domain module. In this experiment, we compared the detection results with and without the inclusion of the Fca module. The networks used in the experiment were identical except for the presence of the FcaNet module.

**Table 6.** Effectiveness of Fca-Neck.

Model	Precision (%)	Recall (%)	mAP	F1	FPS	Param (M)	Model (M)
YOLOv5s + low	<b>67.21</b>	58.06	<b>56.49</b> *	<b>62.3</b>	78.74	7.84	15.7
YOLOv5s + top	63.44	<b>58.37</b>	55.4	60.8	80.0	7.84	15.7
YOLOv5s + bot	67.11	55.87	53.26	60.97	78.13	7.84	15.7
YOLOv5s	68.57	56.34	52.04	61.86	<b>84.75</b>	<b>7.51</b>	<b>14.67</b>

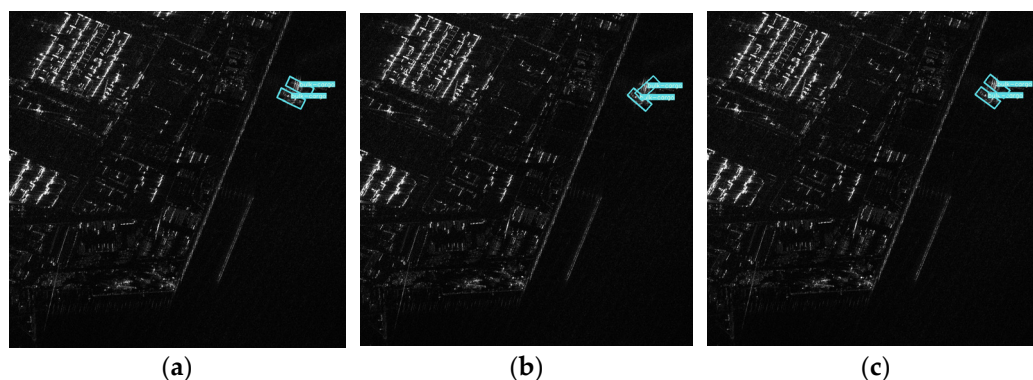
\* With the best results highlighted in bold.

The first three networks in the table incorporate the FcaNet module, and each of them utilizes a different frequency component: low-frequency (low), top-performing frequency component (top), and the optimal frequency component (bot) obtained through the neural architecture search [11].

From Table 6, it can be observed that networks with the FcaNet module exhibit higher detection accuracy, as this module enhances the features. Furthermore, the module that aggregates low-frequency information shows the best detection performance. This is attributed to the Fca-low module's ability to aggregate low-frequency information that is

often overlooked in convolutional neural networks, enabling the network to learn more useful features.

As shown in Figure 7, the integration of high-frequency features in SAR images enables rapid and accurate localization of ship targets. However, it is susceptible to interference from echo noise, leading to errors in the detection of rotational directions. Conversely, the utilization of low-frequency information effectively mitigates this issue, allowing the network to comprehensively learn the characteristic information of ship targets across different frequency bands. This enhances the network's resistance to interference.



**Figure 7.** Comparison of results of different frequency characteristics. (a) Ground truth; (b) YOLOv5s + top result; (c) YOLOv5s + low result.

## 5. Conclusions

This study introduced a network model for ship detection and recognition in complex SAR images. Firstly, the CFAR-FCN branch innovatively integrates CFAR features in the form of a sub-net, allowing the detection network to learn CFAR features while concurrently providing it with supervisory learning and feature enhancement capabilities. The Fca-Neck module aggregates useful frequency domain information, especially low-frequency information, in the image, which is often overlooked in standard convolutional neural networks. Additionally, the rotation anchors used in this study effectively reduce interference from complex backgrounds in SAR ship recognition while mitigating the problem of ground truth suppression caused by NMS in densely clustered ship scenarios.

Numerous ablation experiments and comparative studies conducted on the latest SRSDD-v1.0 dataset demonstrated the effectiveness of each module created in this paper. The experimental results indicate that the model achieves an F1 Score of 63.91, a mAP of 61.07, and an FPS of 56.18 on the SRSDD dataset, with a model size of only 18.51 M. The performance of our method on the SRSDD dataset outperforms several other approaches. Our method is notably suitable for practical devices, ensuring accuracy and meeting the real-time requirements for future SAR ship detection and recognition.

Deep learning-based SAR ship detection methods offer advantages. With extensive training data, deep learning methods can explore features that traditional algorithms cannot, resulting in improved SAR ship detection. When training data is limited, incorporating certain traditional handcrafted operators, such as feature extraction techniques, can enhance the network detection performance. However, through extensive comparative experiments, we have observed that due to the influence of image quality and resolution, our method still has noticeable limitations in ship target recognition performance. The detection accuracy is not adequately high, especially in few-shot scenarios. Therefore, we intend to conduct further research in the following areas: (1) Continuing research on improving the accuracy and quality of SAR ship detection in few-shot scenarios; (2) Continuing research on rapid detection and segmentation of ship targets in SAR images.

**Author Contributions:** Conceptualization, Y.T. and X.W.; methodology, X.W.; validation, X.W. and S.Z.; formal analysis, X.W. and Y.T.; investigation, X.W. and S.Z.; resources, S.Z.; data curation, S.Z. and T.Y.; writing—original draft preparation, X.W. and T.Y.; writing—review and editing, X.W. and J.W.; visualization, S.Z.; supervision, S.Z.; project administration, J.W.; funding acquisition, J.W. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Natural Science Foundation of China, grant number 42271367.

**Data Availability Statement:** Data are available within the article.

**Acknowledgments:** The authors would like to thank the editors and anonymous reviewers for their valuable comments and suggestions.

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

## References

1. Zhang, T.; Zhang, X.; Shi, J.; Wei, S. Depthwise Separable Convolution Neural Network for High-Speed SAR Ship Detection. *Remote Sens.* **2019**, *11*, 2483. [[CrossRef](#)]
2. Zhang, T.; Zhang, X. High-Speed Ship Detection in SAR Images Based on a Grid Convolutional Neural Network. *Remote Sens.* **2019**, *11*, 1206. [[CrossRef](#)]
3. Zhang, S.; Wu, R.; Xu, K.; Wang, J.; Sun, W. R-CNN-Based Ship Detection from High Resolution Remote Sensing Imagery. *Remote Sens.* **2019**, *11*, 631. [[CrossRef](#)]
4. Wang, X.; Li, G.; Plaza, A.; He, Y. Ship Detection in SAR Images by Aggregating Densities of Fisher Vectors: Extension to a Global Perspective. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5206613. [[CrossRef](#)]
5. Wackerman, C.C.; Friedman, K.S.; Pichel, W.G.; Clemente-Colón, P.; Li, X. Automatic Detection of Ships in RADARSAT-1 SAR Imagery. *Can. J. Remote Sens.* **2001**, *27*, 568–577. [[CrossRef](#)]
6. Gao, G.; Liu, L.; Zhao, L.; Shi, G.; Kuang, G. An Adaptive and Fast CFAR Algorithm Based on Automatic Censoring for Target Detection in High-Resolution SAR Images. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 1685–1697. [[CrossRef](#)]
7. Zhang, T.; Zhang, X.; Ke, X. Quad-FPN: A Novel Quad Feature Pyramid Network for SAR Ship Detection. *Remote Sens.* **2021**, *13*, 2771. [[CrossRef](#)]
8. Shao, Z.; Zhang, X.; Zhang, T.; Xu, X.; Zeng, T. RBFA-Net: A Rotated Balanced Feature-Aligned Network for Rotated SAR Ship Detection and Classification. *Remote Sens.* **2022**, *14*, 3345. [[CrossRef](#)]
9. Zhao, J.; Zhang, Z.; Yu, W.; Truong, T.-K. A Cascade Coupled Convolutional Neural Network Guided Visual Attention Method for Ship Detection from SAR Images. *IEEE Access* **2018**, *6*, 50693–50708. [[CrossRef](#)]
10. Joseph, R.; Ali, F. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
11. Qin, Z.; Zhang, P.; Wu, F.; Li, X. FcaNet: Frequency Channel Attention Networks. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, BC, Canada, 11–17 October 2021; IEEE: Piscataway, NJ, USA, 2021.
12. Adil, M.; Buono, A.; Nunziata, F.; Ferrentino, E.; Velotto, D.; Migliaccio, M. On the Effects of the Incidence Angle on the L-Band Multi-Polarisation Scattering of a Small Ship. *Remote Sens.* **2022**, *14*, 5813. [[CrossRef](#)]
13. Marino, A. A Notch Filter for Ship Detection with Polarimetric SAR Data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2013**, *6*, 1219–1232. [[CrossRef](#)]
14. Ferrara, G.; Migliaccio, M.; Nunziata, F.; Sorrentino, A. Generalized-K (GK)-Based Observation of Metallic Objects at Sea in Full-Resolution Synthetic Aperture Radar (SAR) Data: A Multipolarization Study. *IEEE J. Ocean. Eng.* **2011**, *36*, 195–204. [[CrossRef](#)]
15. 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](#)]
16. Leng, X.; Ji, K.; Yang, K.; Zou, H. A Bilateral CFAR Algorithm for Ship Detection in SAR Images. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 1536–1540. [[CrossRef](#)]
17. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; IEEE: Piscataway, NJ, USA, 2014.
18. Girshick, R. Fast R-CNN. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; IEEE: Piscataway, NJ, USA, 2015.
19. 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.* **2017**, *39*, 1137–1149. [[CrossRef](#)] [[PubMed](#)]
20. Dai, J.; Li, Y.; He, K.; Sun, J. R-fcn: Object detection via region-based fully convolutional networks. *Neural Inf. Process. Syst.* **2016**, *29*, 1–9.

21. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; Berg, A.C. SSD: Single Shot MultiBox Detector. In *Computer Vision—ECCV 2016, Proceedings of the 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016*; Proceedings, Part I. European Conference on Computer Vision; Leibe, B., Matas, J., Sebe, N., Welling, M., Eds.; Springer International Publishing: Cham, Switzerland, 2016; pp. 21–37, ISBN 978-3-319-46448-0.
22. 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, 26 June–1 July 2016; IEEE: Piscataway, NJ, USA, 2016.
23. Redmon, J.; Farhadi, A. YOLO9000: Better, Faster, Stronger. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; IEEE: Piscataway, NJ, USA, 2017.
24. Zhang, T.; Zhang, X.; Shi, J.; Wei, S.; Wang, J.; Li, J.; Su, H.; Zhou, Y. Balance Scene Learning Mechanism for Offshore and Inshore Ship Detection in SAR Images. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 4004905. [[CrossRef](#)]
25. Zhang, T.; Zhang, X. ShipDeNet-20: An only 20 convolution layers and < 1-MB lightweight SAR ship detector. *IEEE Geosci. Remote Sens. Lett.* **2020**, *18*, 1234–1238.
26. Liu, Z.; Wang, H.; Weng, L.; Yang, Y. Ship Rotated Bounding Box Space for Ship Extraction From High-Resolution Optical Satellite Images With Complex Backgrounds. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 1074–1078. [[CrossRef](#)]
27. Zhang, T.; Zhang, X. Injection of Traditional Hand-Crafted Features into Modern CNN-Based Models for SAR Ship Classification: What, Why, Where, and How. *Remote Sens.* **2021**, *13*, 2091. [[CrossRef](#)]
28. Zhang, T.; Zhang, X. A polarization fusion network with geometric feature embedding for SAR ship classification. *Pattern Recognit.* **2022**, *123*, 108365. [[CrossRef](#)]
29. Zhang, L.; Liu, Y.; Zhao, W.; Wang, X.; Li, G.; He, Y. Frequency-Adaptive Learning for SAR Ship Detection in Clutter Scenes. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5215514. [[CrossRef](#)]
30. Ai, J.; Tian, R.; Luo, Q.; Jin, J.; Tang, B. Multi-Scale Rotation-Invariant Haar-Like Feature Integrated CNN-Based Ship Detection Algorithm of Multiple-Target Environment in SAR Imagery. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 10070–10087. [[CrossRef](#)]
31. Zhang, T.; Zhang, X.; Ke, X.; Liu, C.; Xu, X.; Zhan, X.; Wang, C.; Ahmad, I.; Zhou, Y.; Pan, D.; et al. HOG-ShipCLSNet: A Novel Deep Learning Network with HOG Feature Fusion for SAR Ship Classification. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–22. [[CrossRef](#)]
32. Zhang, X.; Huo, C.; Xu, N.; Jiang, H.; Cao, Y.; Ni, L.; Pan, C. Multitask Learning for Ship Detection From Synthetic Aperture Radar Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 8048–8062. [[CrossRef](#)]
33. Zhang, G.; Li, Z.; Li, X.; Yin, C.; Shi, Z. A Novel Salient Feature Fusion Method for Ship Detection in Synthetic Aperture Radar Images. *IEEE Access* **2020**, *8*, 215904–215914. [[CrossRef](#)]
34. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; Springer: Cham, Switzerland, 2015; pp. 234–241.
35. Ahmed, N.; Natarajan, T.; Rao, K.R. Discrete Cosine Transform. *IEEE Trans. Comput.* **1974**, *C-23*, 90–93. [[CrossRef](#)]
36. Zheng, Z.; Wang, P.; Ren, D.; Liu, W.; Ye, R.; Hu, Q.; Zuo, W. Enhancing Geometric Factors in Model Learning and Inference for Object Detection and Instance Segmentation. *IEEE Trans. Cybern.* **2022**, *52*, 8574–8586. [[CrossRef](#)]
37. Lei, S.; Lu, D.; Qiu, X.; Ding, C. SRSDD-v1.0: A High-Resolution SAR Rotation Ship Detection Dataset. *Remote Sens.* **2021**, *13*, 5104. [[CrossRef](#)]
38. Xie, X.; Cheng, G.; Wang, J.; Yao, X.; Han, J. Oriented R-CNN for Object Detection. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, BC, Canada, 11–17 October 2021; IEEE: Piscataway, NJ, USA, 2021.
39. Xu, Y.; Fu, M.; Wang, Q.; Wang, Y.; Chen, K.; Xia, G.-S.; Bai, X. Gliding Vertex on the Horizontal Bounding Box for Multi-Oriented Object Detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *43*, 1452–1459. [[CrossRef](#)]
40. Tian, Z.; Shen, C.; Chen, H.; He, T. FCOS: Fully Convolutional One-Stage Object Detection. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019*; IEEE: Piscataway, NJ, USA, 2019.
41. Law, H.; Deng, J. CornerNet: Detecting Objects as Paired Keypoints. In *Computer Vision—ECCV 2018. Part XIV, Proceedings of the 15th European Conference, Munich, Germany, 8–14 September 2018*; Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y., Eds.; Springer: Cham, Switzerland, 2018; pp. 765–781, ISBN 9783030012632.
42. Yi, J.; Wu, P.; Liu, B.; Huang, Q.; Qu, H.; Metaxas, D. Oriented Object Detection in Aerial Images with Box Boundary-Aware Vectors. In Proceedings of the 2021 IEEE Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 3–8 January 2021; IEEE: Piscataway, NJ, USA, 2021.
43. Yang, X.; Yan, J.; Feng, Z.; He, T. R3Det: Refined Single-Stage Detector with Feature Refinement for Rotating Object. *Proc. AAAI Conf. Artif. Intell.* **2021**, *35*, 3163–3171. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.