



Article

Multi-Scale Ship Detection Algorithm Based on a Lightweight Neural Network for Spaceborne SAR Images

Shanwei Liu ^{1,*}, Weimin Kong ^{1,2}, Xingfeng Chen ², Mingming Xu ¹, Muhammad Yasir ¹, Limin Zhao ² and Jianguo Li ²

¹ College of Oceanography and Space Informatics, China University of Petroleum (East China), Qingdao 266580, China; z20160097@s.upc.edu.cn (W.K.); xumingming@upc.edu.cn (M.X.); ls1801004@s.upc.edu.cn (M.Y.)

² Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100101, China; chenxf@aircas.ac.cn (X.C.); zhaolm@aircas.ac.cn (L.Z.); lijg@aircas.ac.cn (J.L.)

* Correspondence: 20110052@upc.edu.cn

Abstract: The current limited spaceborne hardware resources and the diversity of ship target scales in SAR images have led to the requirement of on-orbit real-time detection of ship targets in spaceborne synthetic aperture radar (SAR) images. In this paper, we propose a lightweight ship detection network based on the YOLOv4-LITE model. In order to facilitate the network migration to the satellite, the method uses MobileNetv2 as the backbone feature extraction network of the model. To solve the problem of ship target scale diversity in SAR images, an improved receptive field block (RFB) structure is introduced, enhancing the feature extraction ability of the network, and improving the accuracy of multi-scale ship target detection. A sliding window block method is designed to detect the whole SAR image, which can solve the problem of image input. Experiments on the SAR ship dataset SSDD show that the detection speed of the improved lightweight network could reach up to 47.16 FPS, with the mean average precision (*mAP*) of 95.03%, and the model size is only 49.34 M, which demonstrates that the proposed network can accurately and quickly detect ship targets. The proposed network model can provide a reference for constructing a spaceborne real-time lightweight ship detection network, which can balance the detection accuracy and speed of the network.

Keywords: real-time ship detection; SAR images; YOLO; lightweight network



Citation: Liu, S.; Kong, W.; Chen, X.; Xu, M.; Yasir, M.; Zhao, L.; Li, J. Multi-Scale Ship Detection Algorithm Based on a Lightweight Neural Network for Spaceborne SAR Images. *Remote Sens.* **2022**, *14*, 1149. <https://doi.org/10.3390/rs14051149>

Academic Editors: Fernando Bação and Eric Vaz

Received: 31 December 2021

Accepted: 23 February 2022

Published: 25 February 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 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), an essential aerospace remote sensor, is characterized by its ability to achieve all-weather observation of the ground [1]. Based on this advantage, SAR technology has developed rapidly [2–4], and has a wide range of applications in target detection [5], disaster detection, military operations [6], and resource exploration [7].

SAR is an essential tool for maritime surveillance [8]. As the maritime trade and transportation carrier is an important military object, it is of great significance to realize the real-time detection of ship targets in spaceborne SAR images [9]. Due to the satellite's limited computing and storage resources, the accuracy, detection speed, and model size of the target detection algorithm are required simultaneously [10]. At present, the methods of ship target detection [11] in SAR images are mainly divided into two types: traditional detection methods, and target detection methods based on deep learning [12].

Traditional ship detection algorithms in SAR images generally detect ship targets by manually selectively extracting features such as gray level, contrast ratio, texture, geometric size, scattering characteristics, histogram of oriented gradient (HOG) [13], and scale-invariant feature transform (SIFT) [14]. Generally, they can achieve better detection performance in simple scenes with less interference. The constant false alarm rate (CFAR) [15–17] detection algorithm is widely used as a contrast-based target detection algorithm for SAR ship detection. However, each pixel point in the CFAR detection algorithm

is involved in the calculation of distribution parameter estimation multiple times, and the calculation of background clutter distribution is extensive. The detection speed cannot meet the demand of real-time.

In recent years, deep learning technology has developed rapidly natural image recognition. R-CNN [18] introduced convolutional neural networks (CNN) into the field of target detection, which has brought new research ideas to target detection, and its application in SAR images has an ample exploration space. Currently, the algorithms based on convolutional neural networks mainly used in ship detection in SAR images include two-stage detection methods represented by R-CNN, Fast R-CNN [19], and Faster R-CNN [20]. This kind of algorithm takes a series of candidate regions as the candidate boxes of samples, and then produces secondary corrections based on the candidate regions to obtain the detection results, so that they have high detection accuracy. However, their network structures are complex, there are many parameters, and the recognition speed is slow, which cannot meet the real-time requirements of ship detection tasks. Furthermore, SSD [21] and YOLO [22–24] series algorithms based on single-stage regard the target detection problem as a regression analysis problem of target location and category information. They directly output the detection results through a neural network model, with high speed and accuracy, and are more suitable for ship detection tasks with near real-time detection requirements [25].

Although the above algorithms have good detection performance, applying them directly to ship detection in SAR images is difficult. In addition, there are still some challenges in the deep learning-based ship detection method in SAR images [26,27]: (1) Due to the unique imaging technology of SAR, there are more scattering noise and sea clutter in SAR images, and the phenomenon of side flap is also severe, which will cause the contrast between the ship and the sea to decrease. This leads to a decline in detection accuracy. Furthermore, the interference from the land, islands, and other natural factors increases the false alarm rate. (2) Ships have arbitrary directionality and multi-scale in SAR images. Various ships are of different sizes and scales, reflected in SAR images as different numbers of pixels, especially for small-scale ships. Fewer pixels are easily confused with SAR image speckle noise, and there is little information for position refinement and classification compared to large ships. Meanwhile, the orientation of ship targets in satellite images taken from the air vary greatly, and can change between 0° and 360° , which improves the detection difficulty and leads to poor detection and recognition accuracy. (3) SAR images cannot be directly input to the network for detection if the scene is enormous. It is assumed that the SAR image of the large scene is now input into the network. In this case, the ship target will be resampled to a few or even just one pixel, seriously affecting the detection accuracy.

To solve the above problems, Kang et al. [28] added context features to the corresponding region of interest to make the background information help to eliminate false alarms. In order to better obtain the salient features in the image and suppress clutter, some studies use the attention mechanism. Du et al. [29] introduced important information into the network, so that the detector could focus more on the target area. Zhao et al. [30] proposed an extended attention block to enhance the feature extraction ability of the detector. Unlike the horizontal bounding box method, An et al. [31] and Chen et al. [32] adopt a directional bounding box, which is better for densely arranged objects. Cui et al. [33] proposed a multiscale ship detection method based on a dense attention pyramid network (DAPN) in SAR images. Wang et al. [34] improved the performance of detecting multiscale ships by enhancing the ability of feature extraction and the nonlinear relationship between different features. Wu et al. [35] proposed a new ship detection network, called case segmentation assisted ship detection network (ISASDet), which uses case segmentation to promote ship detection. Shi et al. [36] proposed an adaptive sliding window algorithm to extract the connected water region, and proposed the ship's suspicious target region.

The above improvements are based on specific hardware and storage resources. Although these methods improve the detection performance of the model to a certain extent, it is still difficult to meet the development needs of real-time mission planning on satel-

lites [37]. However, the hardware resources on the satellite are limited, thus reducing the network parameters and compressing the network model on the premise of ensuring the network performance has also become a challenge in the target detection task on the satellite [38]. Chollet et al. [39] constructed deep separable convolution by combining the branching structure of inception, and reduced the computation of convolution operation on the premise of ensuring the model's accuracy through the BottleNeck method. Zhang et al. [40] proposed a high-speed SAR ship detection method based on depth separable convolutional neural network (DS-CNN) using a combination of multi-scale detection mechanism, cascade system, and anchor box mechanism.

According to the above analysis, considering the sea observation, the situation of ship detection, and the resource space of the satellite system, we propose a lightweight detection method for multi-scale ships in satellite SAR images. A real-time ship detection model on satellites robust enough to multi-scale objects, complex backgrounds, and large scene SAR images under the platform of limited computing resources is implemented. Aiming at the diversity of ship target scales in ship detection in spaceborne SAR images, the input of large scene images, and the limited resources on board, the single-stage target detection method is improved, respectively. The main contributions of our work are as follows:

- In order to make the model meet the development needs of real-time mission planning on satellites, we use the improved MobileNetv2 as the backbone feature extraction network of the YOLOv4-LITE model. In the Path Aggregation Network (PANet) structure, the depthwise separable convolution is used to replace the standard convolution, which ensures the lightness and detection speed of the network, and enables the constructed network model to meet the requirements of limited computing resources on the satellite;
- In order to solve the problem of ship target scale diversity, the RFB structure is introduced and improved to enhance the learning characteristics of the lightweight model, obtain more effective information by increasing the receptive field of the network model, and improve the accuracy of multi-scale ship detection, especially the detection accuracy of small target ships. This makes full use of fewer parameters to extract features effectively, and builds a real-time ship detection model on satellite to meet the development needs of real-time mission planning on satellite;

The K-means algorithm is used to cluster the data set, and a sliding window blocking method is designed to solve the problem of image input. At the same time, based on the sliding window blocking method, a quadratic non-maximum suppression (NMS) operation is added to the output of the network, and the *CIoU* (Complete Intersection over Union)-NMS combination method is used to suppress the repeated frame selection of a ship caused by the sliding window blocking method. The organizational structure of this paper is as follows. In Section 2, we will introduce two detection methods, respectively, and describe the detailed process of our proposed method. Section 3 describes the sliding window blocking method and the quadratic non maximum suppression operation. The detailed experimental process is shown in Section 4. See Section 5 for conclusions.

2. MobileNet-YOLOv4-LITE Algorithm Principles and Improvements

2.1. MobileNetv2 Network

MobileNet is a new lightweight network proposed by the Google team in 2017 [41], which has fewer network parameters and lower operation costs. The MobileNetv2 [42] network inherits the advantages of the MobileNet v1 network, and introduces the inverted residuals structure and bottleneck structure. Unlike the traditional residual structure, the inverted residual first increases and then decreases the dimension of the network. The number of channels in the network increases by introducing 1×1 convolution, and then the standard convolution is replaced by depth separable convolution. Compared with the standard convolution of the backbone feature extraction network of YOLOv4, the amount of calculation of deep separable convolution is reduced by approximately nine times, and

the amount of parameters is reduced accordingly, which further improves the network performance.

The MobileNetv2 network includes standard convolution, depth-wise separable convolution of reverse residual structure, and avgpool. The network structure is shown in Figure 1.

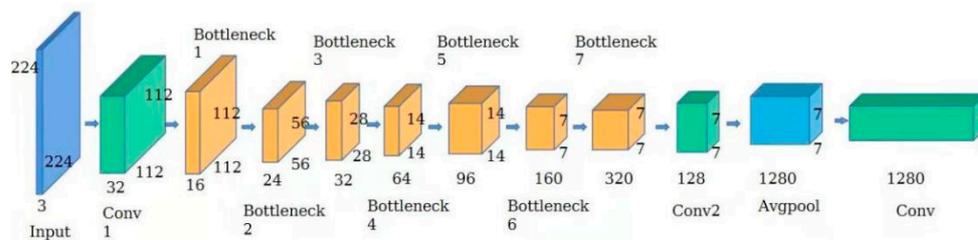


Figure 1. The MobileNetv2 network structure.

In order to use the MobileNetv2 network as the backbone feature extraction network of the YOLOv4-LITE detection framework, the network input and output image sizes need to be changed. Firstly, the input image size is adjusted to meet the input requirements of the MobileNetv2 network; secondly, the number of network parameters and operation costs are fully considered to ensure the accuracy of the training network, and the subsequent layer of the output channel number increase from 320 to 1280 in Figure 1 is removed; finally, the modification to meet the dimensional articulation requirements is carried out. The modified network structure and its parameter settings are shown in Table 1, where input is the input feature map size; t denotes the expansion coefficient of the intermediate convolution channel; c denotes the number of output channels; n denotes how many times the layer is repeated; and s denotes the stride of convolution.

Table 1. Modified MobileNetv2 network structure by layer.

Layer	Input	Operator	t	c	n	s
1	416 × 416 × 3	Conv2d	-	32	1	2
2	208 × 208 × 32	Bottleneck	1	16	1	1
3	208 × 208 × 16	Bottleneck	6	24	2	2
4	104 × 104 × 24	Bottleneck	6	32	3	2
5	52 × 52 × 32	Bottleneck	6	64	4	2
6	26 × 26 × 64	Bottleneck	6	96	3	1
7	26 × 26 × 96	Bottleneck	6	160	3	2
8	13 × 13 × 160	Bottleneck	6	320	1	1

2.2. MobileNet-YOLOv4-LITE Algorithm and Improvements

YOLOv4 is an algorithm introduced by Bochkovskiy et al. through a series of improvements to YOLOv3, which pushes the accuracy and speed of the YOLO family of algorithms to a new level [43]. Compared with YOLOv3, the YOLOv4 network structure has been improved in many ways. The improvements are mainly reflected in the following three aspects: first, the CSPnet architecture is incorporated into DarkNet53 to generate the CSPDarkNet53 backbone feature extraction network, which significantly reduces the number of network parameters while enhancing the extraction capability of the backbone feature extraction network; second, the neck feature enhancement network adopts the combination of path aggregation network (PANnet) and spatial pyramid pooling (SPP) module to enhance the extraction ability of different feature layers of the network and enable more efficient feature fusion; third, the LeakyReLU activation function in the backbone feature extraction network is modified into the Mish activation function, which can effectively

improve the generalization ability of the model and the optimization ability of the results and improve the quality of the results. Equation (1) is as follows:

$$f(x) = x \times \tanh(\log(1 + e^x)), \tag{1}$$

MobileNet-YOLOv4-LITE is a detection model based on the YOLOv4 algorithm framework, using the MobileNet series network as the backbone feature extraction network of the YOLOv4 detection framework, and the structure of the model is shown in Figure 2, where MobileNetv2 is the Backbone part, SPP and PAnet are the Neck part, and Head is the output part.

The bottleneck structure is the core part of the MobileNetv2 network. Each bottleneck comprises two standard convolutions and one deep separation convolution (Dwise). Among them, Relu 6 is used as the activation function. Relu 6 is obtained by limiting the maximum output value of the Relu activation function to six, ensuring better numerical resolution in floating-point operation when the equipment calculation conditions are limited. If the activation range of Relu is not limited, the output range is 0 to positive infinity. Suppose the activation value is tremendous and distributed in an extensive range. In that case, it is impossible to accurately describe such an extensive range of values in floating-point operation and how to bring accuracy loss. The calculation method is shown in Equation (2):

$$Relu6 = \min(6, \max(0, x)), \tag{2}$$

In addition, compared with the Mish activation function, Relu6 can make the model learn sparse features earlier, prevent numerical explosion, and avoid the loss of accuracy caused by numerical distribution in an extensive range. The amount of calculation is smaller, which is more suitable for the needs of target detection on satellites.

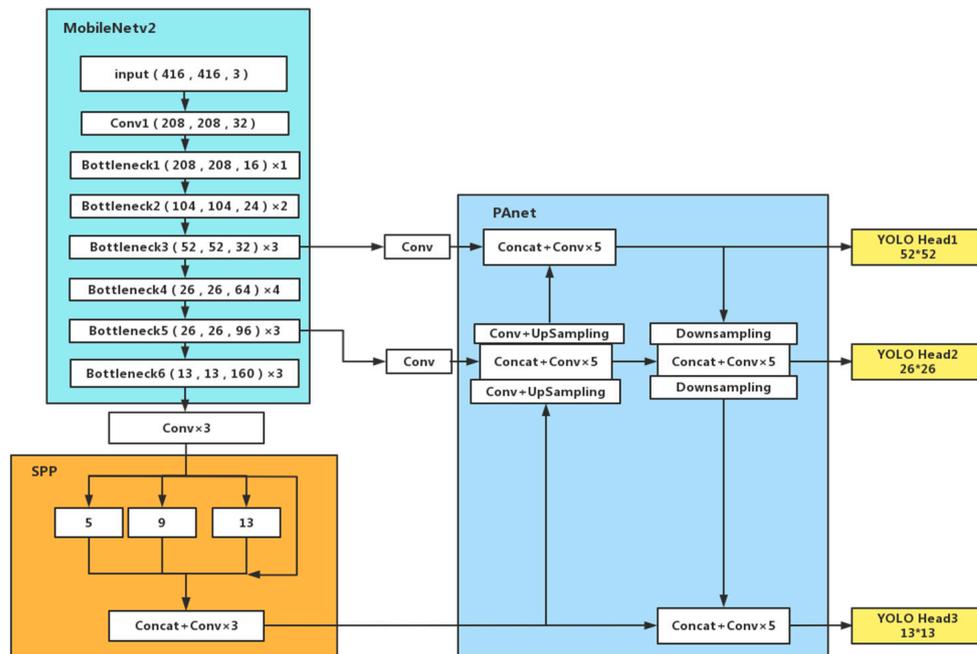


Figure 2. MobileNet-YOLOv4-LITE network structure.

Various ships are of different sizes and scales, which are reflected in the SAR images as objects with different pixels. Ship detection becomes a challenging task for SAR images due to its multi-scale feature. Especially for small-scale ships, fewer pixels are easily confused with SAR image speckle noise. Compared with large ships, existing ship detection methods are not sensitive to small ships, resulting in poor accuracy. Thus, the critical issue to improve the detection accuracy of multi-scale ships is how to make full use of the features of different scales to obtain more information.

The RFB structure starts from simulating human vision's receptive field and eccentricity. It draws on the Inception network and the idea of dilated convolution, which can effectively increase the receptive field, extract multi-scale features, and enhance the ability of lightweight convolutional neural networks to learn in-depth features [44]. By adding a small amount of calculation, the detection accuracy of the model is effectively improved. At the same time, in feature extraction, the area containing the ship target in the image plays the leading role as the effective area. However, due to the significant difference in the size of ship targets in SAR images, small target ships are easily confused with speckle noise, resulting in the inability to extract ship features during feature extraction effectively. Unlike the max pooling operation of the SPP structure, the dilated convolution and residual connection in the RFB structure can enhance the receptive field of the effective region without changing the scale features, and maximize the preservation of small ship objects in the feature map.

RFB structure extracts features using standard convolution and dilated convolution on branches of different scales. The standard convolution simulates the receptive fields of different scales in the receptive field group. The dilated convolution increases the receptive fields, while keeping the size of the characteristic map unchanged. The original RFB structure is a three-branch structure composed of 1×1 , 3×3 , and 5×5 convolution cores, and the dilated convolution of dilation rate = 1, dilation rate = 3, and dilation rate = 5 is introduced into each branch to increase the receptive field. Finally, the output and input ports of the three branches are connected by residuals to achieve the purpose of fusing different characteristics.

In order to better improve the detection accuracy of the network model for multi-scale ship targets in SAR images, we improved the original RFB structure. We add a dilated convolution with a dilation rate = 7, based on the original three branches. By connecting the output and input of the four branches with residuals, features which are more recognizable to multi-scale ship targets in SAR images are generated. Meanwhile, to consider the model parameters, we replace the 7×7 convolutions with 1×7 and 7×1 convolutions. At the same time, the convolution size of 3×3 in the second branch is replaced by the convolution of 1×3 and 3×1 . The convolution of 1×5 and 5×1 replaces the convolution of 5×5 in the third branch, generating a lightweight receptive field enhancement structure that is more suitable for multi-scale ship detection in SAR images. Experiments verify the performance of the improved RFB structure. The improved RFB structure is shown in Figure 3.

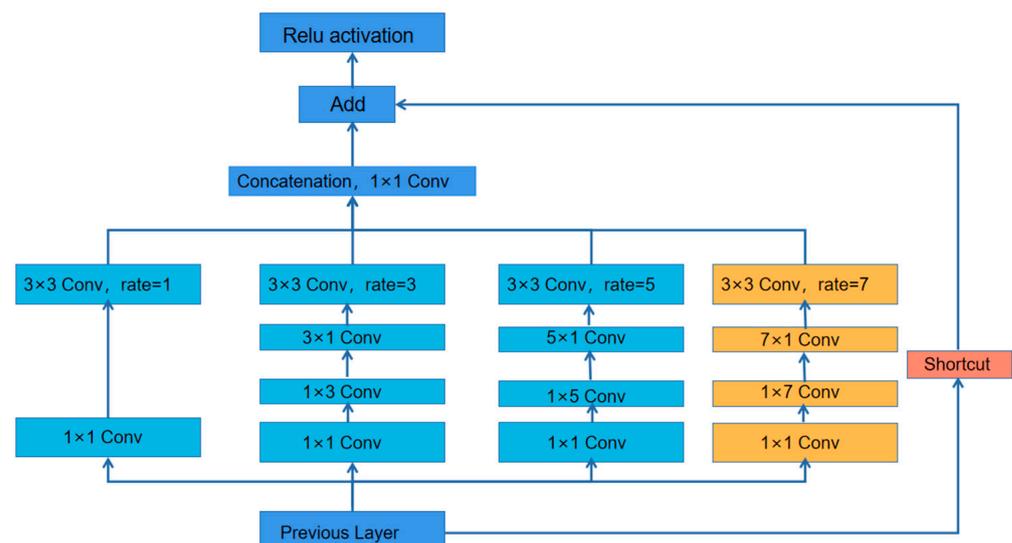


Figure 3. Improved RFB structure. Orange is the increased 7×7 dilation convolution branch, which can better extract multi-scale features.

The parameters calculated by the model are mainly concentrated in the PANet structure and 3×3 convolution. In order to further reduce the number of parameters, the depth separable convolution is used to replace the 3×3 convolution of PANet in MobileNet-YOLOv4-LITE.

At the same time, the effective receptive field [45] plays a vital role in the feature map. The ship in the SAR image accounts for a small proportion of the original image, thus the effective receptive field area is easily lost in the feature extraction process. Compared with the maximum pooling operation of the SPP module, the hollow convolution and residual connection of the RFB module can enhance the receptive field of the effective area without changing the scale feature, and retain the small-sized ship target in the feature map to the greatest extent. Therefore, the SPP part of the network model is replaced with an improved RFB module, and the improved RFB module is added to the output ends of the feature enhancement network 32 and 96 scales. The improved YOLOv4-LITE-MR network structure is shown in Figure 4.

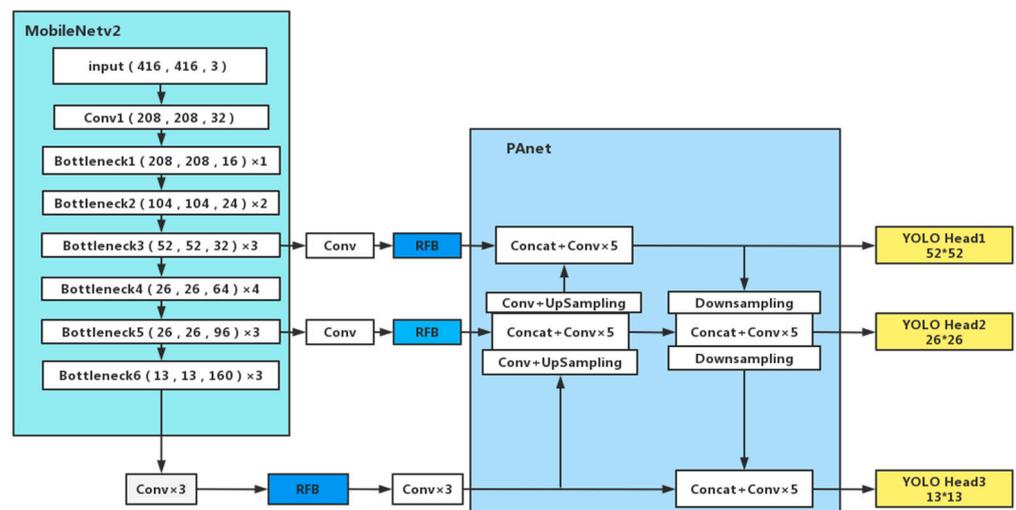


Figure 4. YOLOv4-LITE-MR network structure. The blue part is the added position of the improved RFB structure.

3. Sliding Window Partition Method

All target detection methods based on deep learning have strict restrictions on image input, and all input images will be adjusted to fixed pixel size. The imaging scene of the SAR image is enormous, therefore the image needs to be divided into blocks to avoid the size of the ship target from being resampled to a few or even only one pixel, which affects the detection performance.

It was inspired by the idea of the CFAR algorithm to detect images through sliding window technology. A sliding window detection method is designed and used, as shown in Figure 5. The square area with side length L is used to traverse the whole image in both horizontal and vertical directions, set an overlap of the sub-image with a length of S to avoid damage to the ship by the block, and complete the sliding window block process of the entire image.

The subgraph size is $L \times L$, and the length of the overlapping part is S (Figure 5). Among them, L is set at 1650 pixels. The size of S should be larger than the number of pixels along ship's length in the image. It is set to 320 pixels, according to the ship size and image resolution. The above parameters can be changed adaptively based on the image's resolution.

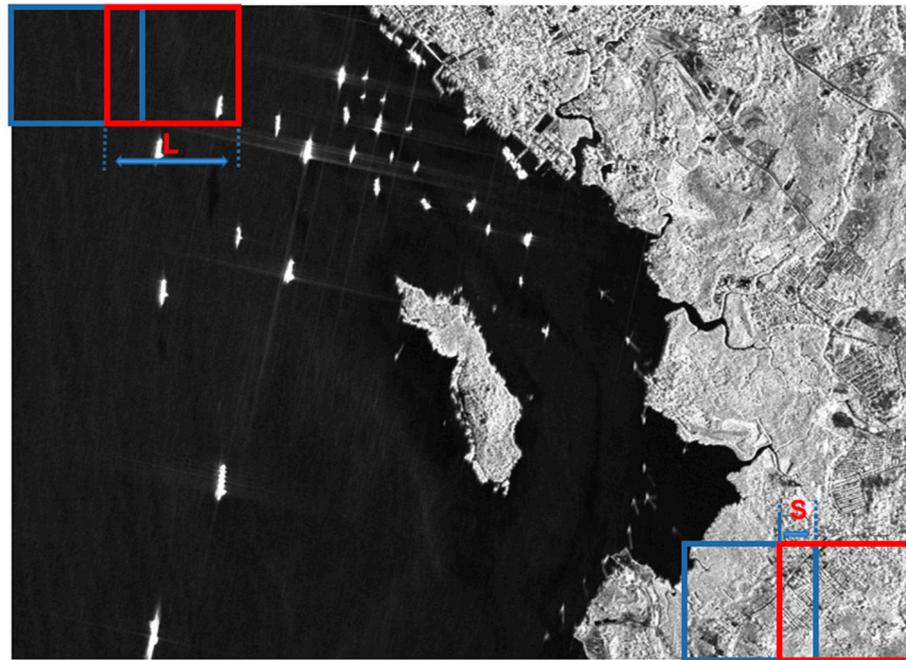


Figure 5. Schematic diagram of sliding window partition method. Blue and red are adjacent sub-blocks, L is the length and width of sub-blocks, and S is the degree of overlap between sub-blocks.

However, in practical applications, it is found that the sliding window detection method will bring a large number of false alarms, mainly due to repeated frame selection of a ship. In order to solve this problem, we added a non-maximum suppression (NMS) operation in the output part of the network to exclude the detection box where the intersection over union (IoU) [46] is less than the threshold, and to suppress the ships selected repeatedly.

At the same time, it is found that different IoU strategies used in the sliding window blocking method have different inhibitory effects on the repeated frame selection of ship targets. In order to determine the best combination of the IoU method and NMS, the experiments of IoU , $DIoU$ [47], and $CIoU$ [48] methods are carried out, respectively. Finally, it is determined to add the combination of $CIoU$ loss and NMS to the output port of the prediction network. $CIoU$ is set to 0.3, eliminating the detection frames more minor than the threshold, suppressing repeated false positives, and selecting targets in multiple frames.

IoU loss function is needed in the detection frame regression in the target detection task. The prediction box is defined as B , and the actual box is B_{GT} , the IoU loss function Equation (3) as follows:

$$IoU = \frac{|B \cap B_{GT}|}{|B \cup B_{GT}|} \quad (3)$$

$$L_{IoU} = 1 - \frac{|B \cap B_{GT}|}{|B \cup B_{GT}|} \quad (4)$$

However, as shown in Figure 6, the IoU loss function has two problems: first, when the prediction box and the target box do not intersect, IoU is 0. Then, the loss function is not differentiable, and the gradient direction is lost. In the IoU loss function, it is impossible to optimize the disjoint case of two boxes. Second, when the size of the two prediction boxes are the same, the two IOUs are also equal, and the IoU loss cannot distinguish the difference in the intersection of the two.

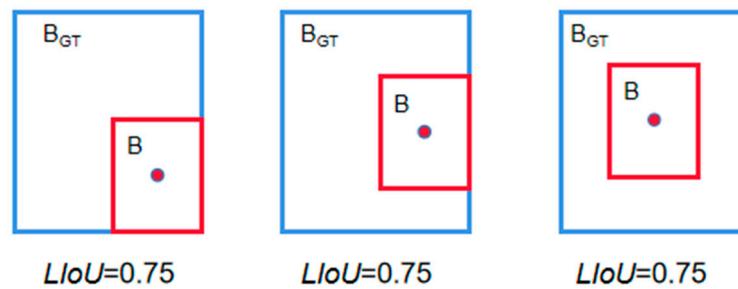


Figure 6. Limitations of *IoU* losses. Red rectangles represent prediction boxes, and blue rectangles represent reality boxes.

Compared to the *IoU* loss function, the *DIoU* loss function is more consistent with the mechanism of target box regression and converges faster without the risk of divergence. The *DIoU* loss is defined as:

$$L_{DIoU} = 1 - IoU + \frac{\rho^2(b, b_{GT})}{c^2}, \tag{5}$$

Among them, b and b_{GT} are the center points of B and B_{GT} , respectively. ρ represents the Euclidean distance between the two centers; and c is the diagonal length of the minimum enclosing box that can enclose B and B_{GT} , as shown in Figure 7.

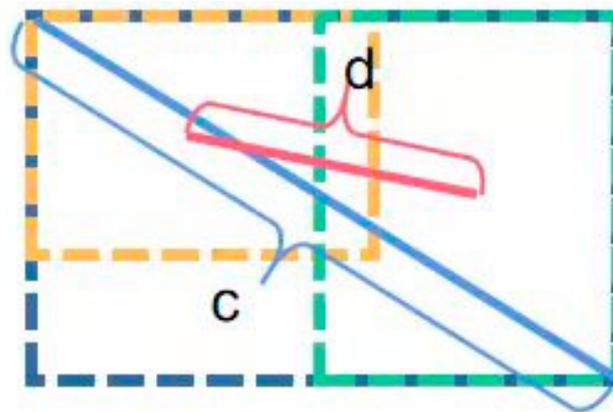


Figure 7. Schematic diagram of *DIoU*. c is the diagonal length of the smallest enclosing box covering two boxes, and $d = \rho^2(b, b_{GT})$ is the distance of central points of two boxes.

DIoU has two distinct advantages: (1) When B and B_{GT} do not intersect, it can guide the direction movement of B and directly minimize the distance B between and B_{GT} , convergence speed is fast; (2) it can be applied to non-maximum suppression (NMS) to make NMS more reasonable.

However, when the center points of the two frames coincide, that is, the values of c and d remain unchanged, *DIoU* cannot accurately determine the position of the anchor frame, which brings a loss of accuracy. Therefore, it is necessary to introduce the aspect ratio of the anchor frame at this time. *CIoU* can effectively solve this problem and make the prediction frame more consistent with the actual frame. The calculation formula is shown in Equation (6):

$$CIoU = \frac{\rho^2(b, b_{GT})}{c^2} + \alpha v, \tag{6}$$

where α is the weight function, and v is used to measure the consistency of the aspect ratio.

$$v = \frac{4}{\pi^2} \left(\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h} \right)^2, \tag{7}$$

$$\alpha = \frac{v}{(1 - IoU) + v'} \quad (8)$$

The $CIoU$ loss is defined as:

$$L_{CIoU} = 1 - IoU + \frac{\rho^2(b, b_{GT})}{c^2} + \alpha v, \quad (9)$$

4. Experiment

4.1. Experimental Environment and Data Introduction

The experimental software and hardware configuration are shown in Table 2. In the experiment, all comparison methods were run on the same platform.

Table 2. Experimental software and hardware configuration.

Configuration	Parameter
CPU	Intel(R) Core (TM) i9-10980XE
RAM	128G
GPU	NVIDIA GeForce RTX2080Ti
Operating system	Windows10
Development tools	PyTorch, python3.6, CUDA10

The experiment uses the SSDD [49] dataset, which is commonly used in SAR ship target detection. The SSDD dataset has 1160 images and 2456 ship targets, with an average of 2.12 ships per image, and contains ships of different sizes, from small target ships of 7×7 pixels to large target ships of 211×298 pixels. The images in this dataset have multiple polarization patterns, different resolutions, and far and near sea scenes, which can better verify the algorithm's effectiveness.

GF-3 image, as shown in Figure 8. In order to verify the detection effect of the proposed model in the whole image, three scenes of high-resolution SAR data are used to test the ship detection performance of the proposed method. The data are from the GF3 satellite, the shooting area is Malacca Strait, and some parameters are shown in Table 3.

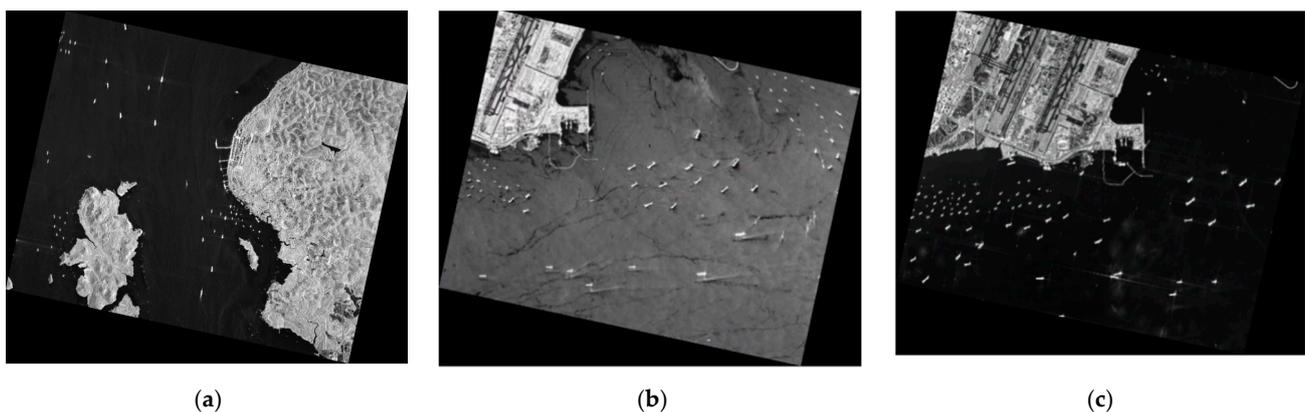


Figure 8. Whole SAR image data, (a) data1, (b) data2, and (c) data3.

Table 3. Test data parameters.

Date	Imaging Time	Pixel Size	Product Type	Resolving Power/m	Polarization Mode
data1	11 January 2020	$15,770 \times 13,926$	SLC	0.56×0.33	HH
data2	21 October 2018	$15,344 \times 13,586$	SLC	0.56×0.34	HH
data3	28 October 2018	$16,208 \times 13,798$	SLC	0.56×0.31	HH

The three images have apparent speckle noise, sea clutter, and sidelobe phenomenon from the imaging situation. Figure 9 shows the ships under the background of different interference factors. The number of ships in a single scene image is between 80 and 110. The scale span is large, the largest ship is 300 m, the smallest is about 20 m, and it is mainly distributed in the open sea. It can be seen that the test data used in this time include multi-scale, high-density, and high sea state scenes, which is very challenging and meaningful in the ship detection task and can effectively verify the effectiveness of the proposed method.

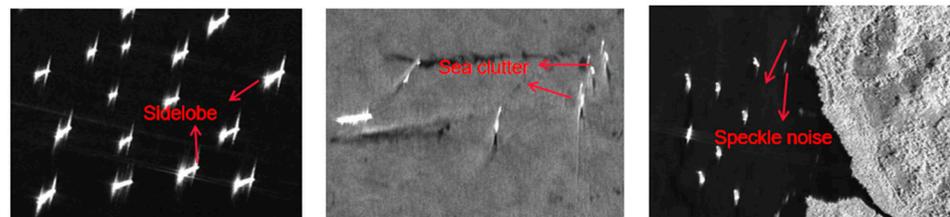


Figure 9. Graphical representation of different interference factors in images.

To ensure the reliability of the experimental data, the SSDD dataset was randomly divided into a training set, a validation set, and a test set according to the ratio of 7:2:1 during training. In the training process, the mosaic data enhancement method was used, the training epoch was set to 100, the training batch size was taken as 8, the initial learning rate was 0.001, and then the learning rate was adjusted to 0.0001 after 50 rounds of training. Cosine annealing learning rate was used for training weights, the momentum factor was 0.9, and the image input size was 416×416 .

4.2. Cluster Analysis of Datasets

A rectangular border is usually used to represent the detected target in target detection. An anchor box is a predefined set of boxes whose width and height are selected to match the width and height of the target in the dataset. Instead of using a predefined combination of aspect ratio and scale, the anchor box in the YOLO algorithm adopts a K-means clustering approach to learn from the training set. The different anchors are obtained from the training set.

Compared with the PASCAL VOC dataset classification in the original network model, ship detection is a two-class classification problem, which only targets one type of ship data. In addition, the multi-scale nature of ships in the ship dataset is also different from other datasets. If the anchor calculated from the VOC dataset continues to be used, the training speed and accuracy of the model will be affected. Meanwhile, Euclidean distance is used in the traditional K-means clustering [50] method to calculate the distance between each sample and the selected cluster center point. For multi-scale ship data, the size of the label box is too large, which leads to errors.

The IoU method can avoid this problem, thus the $1 - IoU$ method is used instead of the Euclidean distance to calculate the distance between the sample and the cluster centroid. The smaller Dis , the higher the degree of similarity, and the distance Equation (10) follows:

$$Dis = 1 - IoU, \quad (10)$$

Since the algorithm performs detection, each detection layer allocates three anchor boxes, thus nine cluster centers are selected. After multiple calculations, the optimal anchor box size is shown in Table 4.

Table 4. Optimized anchor box.

Feature Layer	Anchor Box
(52, 52, 256)	(9, 11), (18, 11), (10, 21)
(26, 26, 512)	(13, 25), (30, 16), (18, 42)
(13, 13, 1024)	(29, 48), (93, 33), (45, 93)

4.3. Evaluation Indicators

The *IoU* threshold between the prediction frame and the actual frame of the target is set to 0.5, and the target's location is correctly predicted when the *IoU* value is more significant than 0.5. The accuracy, precision, recall, and mean accuracy (*mAP*), and frames per second (*FPS*), which are commonly used for target detection, are selected as the evaluation metrics for the algorithm's detection accuracy and detection speed, respectively. In addition, the parameter quantity of the added position of the RFB structure and the figure of merit (*FOM*) of the ship detection effect in the whole SAR image are compared. The parameter quantity directly affects the size of the network model, which also an essential factor in the target detection task. The calculation Equations (11)–(16) is as follows:

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

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

$$Fom = \frac{TP}{TP + FN + FP} \quad (13)$$

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

$$mAP = \frac{1}{N} \sum_{i=1}^N AP, \quad (15)$$

$$FPS = \frac{FrameNum}{ElapsedTime} \quad (16)$$

In the equations mentioned above, *TP* is the number of positive samples predicted to be positive, *FP* is the number of negative samples predicted to be positive, and *FN* is the number of positive samples predicted to be negative. The larger the final calculated value, the higher the accuracy. Framenum represents the number of images processed, and ElapsedTime represents the time taken to process the image. *FPS* is the number of pictures that the model in one second can detect. The larger the *FPS* value, the faster the detection speed of the model.

4.4. Experimental Results and Analysis

In order to verify the effect of the improved network in ship target detection, three experiments are designed. (1) The MobileNet-YOLOv4 model is used to carry out the improved RFB module, and the feature fusion experiments are carried out on the modules with different positions. The effect of the improved RFB module and the influence of the added positions on the model accuracy are discussed. (2) Under the same software and hardware environment and dataset, the applicability of the improved network in the task of satellite-borne ship target detection is verified according to comparing the detection performance of YOLOv3, SSD, YOLOv4, MobileNet-YOLOv4, and the improved MobileNet-YOLOv4 network. (3) The improved network is used to detect the entire GF-3 image, and the network model and sliding window block method are used to verify the detection performance of ships with significant scale differences in complex scenes.

According to the above experimental results, the additional position of the RFB structure is determined, and the YOLOv4-LITE-MR model shown in Figure 4 is obtained.

4.4.1. Improved RFB Module and Module Adding Location Experiment

In order to verify the effect of adding RFB module and improving RFB module on the accuracy of the algorithm and the optimization of depth-separable convolution on the number of parameters, a set of comparison experiments were set up based on the MobileNet-YOLOv4-LITE network to compare and analyze the effect of replacing only the backbone feature extraction network model, the depth-separable convolution replacement model, the original RFB structure, and the improved RFB structure trained with SSDD dataset under the same experimental conditions on the algorithm accuracy, the specific experimental results of which are shown in Table 5, where Params is the total number of parameters, GFLOPs is the number of floating-point operations to measure the complexity of the model, RFB scale is the maximum perceptual field scale of the RFB structure, and *mAP* is the average accuracy mean of the model.

Table 5. Accuracy comparison of RFB structure improvement.

Method	Params/ 10^6	GFlops	RFB Scale	<i>mAP</i> /%
Replace backbone network	38.64	13.36	0	93.45
Depth separable convolution replacement	10.38	3.82	0	93.32
Add RFB structure	10.79	4.27	5	94.06
Add improved RFB structure	10.78	4.26	7	94.59

It can be seen from Table 5 that the number of parameters required to train the MobileNetv2 network model before convolutional replacement is 38.64×10^6 . The number of parameters of the network model after replacement with depth-separable convolution is 10.38×10^6 , which only accounts for 26.86% of the size of the replaced backbone network model. The final computational volume of the improved network is 4.26 GFLOPs, which is 68.11% less than the GFLOPs of the original network, and the improved RFB structure has a specific improvement in model size and accuracy over the original RFB model. It can be seen that the improved RFB structure has a larger perceptual field scale and most minor parameters, which shows the characteristics of lightweight and high accuracy.

In introducing the RFB structure, which enhances the network's ability to extract features at different scales, the backbone network outputs 52×52 , 26×26 , and 13×13 feature maps, respectively. The improved RFB structure for features extracts the 13×13 feature maps. Comparative experiments were carried out to verify the effect of the RFB structure added in the network on the accuracy of the model. The experimental results are shown in Table 6.

Table 6. Comparison of model accuracy for adding RFB additions at different locations.

Location	Params/ 10^6	GFlops	<i>mAP</i> /%
26×26	10.69	4.04	93.96
52×52	10.44	4.04	94.35
$26 \times 26 + 52 \times 52$	10.77	4.26	95.03

The results in Table 6 show that the richer the detailed information of shallow features in two different locations, the higher the accuracy improved by RFB structure. The improved RFB structure is added in front of the two input feature maps of PANet due to its performance.

4.4.2. Experiments on the Detection Performance of Different Detection Networks on SSDD Dataset

In order to verify the detection performance of the improved model for ship targets, Faster-CNN, SSD, YOLOv3 algorithm, YOLOv4 algorithm, MobileNet-YOLOv4-LITE algorithm, and YOLOv4-LITE-MR are compared in the same hardware and software

environment. The detection performance of the proposed network models is evaluated by comparing different evaluation metrics, including *mAP*, detection speed *FPS*, and model size. The experimental results are shown in Table 7.

Table 7. Comparison of performance of different models.

Model	Input	<i>mAP</i> /%	Precision/%	Recall/%	<i>FPS</i>	Model Size/MB
Faster-CNN	416 × 416	74.98	50.98	84.89	18.66	108.12
SSD	300 × 300	85.17	93.73	43.86	83.32	90.60
YOLOv3	416 × 416	93.58	92.99	92.67	46.08	234.98
YOLOv4	416 × 416	94.59	94.88	93.43	38.57	244.30
MobileNet-YOLOv4-LITE	416 × 416	93.32	91.23	92.67	59.88	46.45
YOLOv4-LITE-MR	416 × 416	95.03	94.04	92.56	47.16	49.34

As can be seen from Table 7, compared with the MobileNet-YOLOv4-LITE algorithm, the YOLOv4-LITE-MR algorithm improves the *mAP* and accuracy of the model by 2.29% and 2.91%, respectively, and the improvement of the model recall is smaller. Compared with the YOLOv4 model, the *mAP* of the YOLOv4-LITE-MR model increased by 0.44%. The final improved model has a significant improvement in the above indexes, and has a great improvement in the model size and *FPS*. Compared to the YOLOv4 network model, the size is reduced by 197.96 M, and the *FPS* is increased by 22.27%. The above comparison results indicate that the proposed YOLOv4-LITE-MR algorithm is more suitable for ship detection tasks on the SSDD dataset.

4.4.3. Experiments on the Detection Effect of Sliding Window Chunking Method in the Whole SAR Image

In order to verify the effectiveness of the improved algorithm and the sliding window chunking method for ship detection on the whole SAR image, the GF-3 image of the Malacca Strait region was selected for the experiment. The detection results are shown in Figure 10.

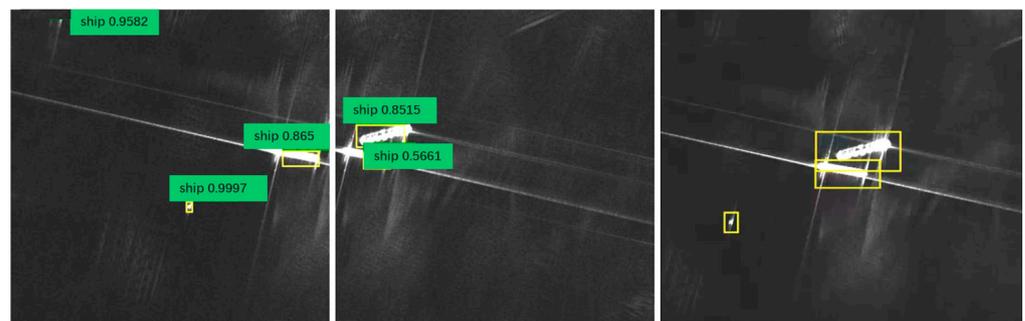


Figure 10. Sliding window chunking method effect schematic.

Different NMS strategies correspond to different inhibitory effects on ship target frame selection. Through comparative experiments, it is found that the loss of *CIoU* and the inhibition effect of NMS are the best, which mainly benefits from the fact that the loss function of *CIoU* fully considers the distance between the prediction frame of the central point and the entire frame, and introduces the aspect ratio of the anchor frame so that the prediction frame will be more in line with the actual frame. The experimental results are shown in Figure 11.

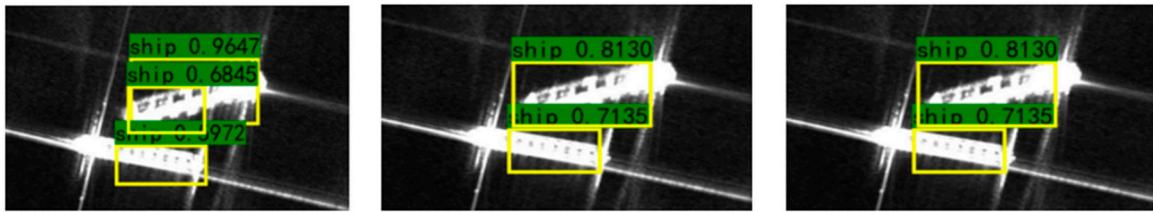


Figure 11. Comparison of *IoU-NMS*, *Diou-NMS*, and *Ciou-NMS*.

Figures 12–14 show the detection results of data1, data2, and data3 SAR images using MobileNet-YOLOv4-LITE and YOLOv4-LITE-MR, respectively. As can be seen from the enlarged part in Figures 12–14, the improved network reduces the generation of false alarm targets. At the same time, it shows that the multi-scale ship target detection has a good detection effect, the detection effect of near-shore ships has also been improved, and it has a better ability to distinguish ships and side lobes.

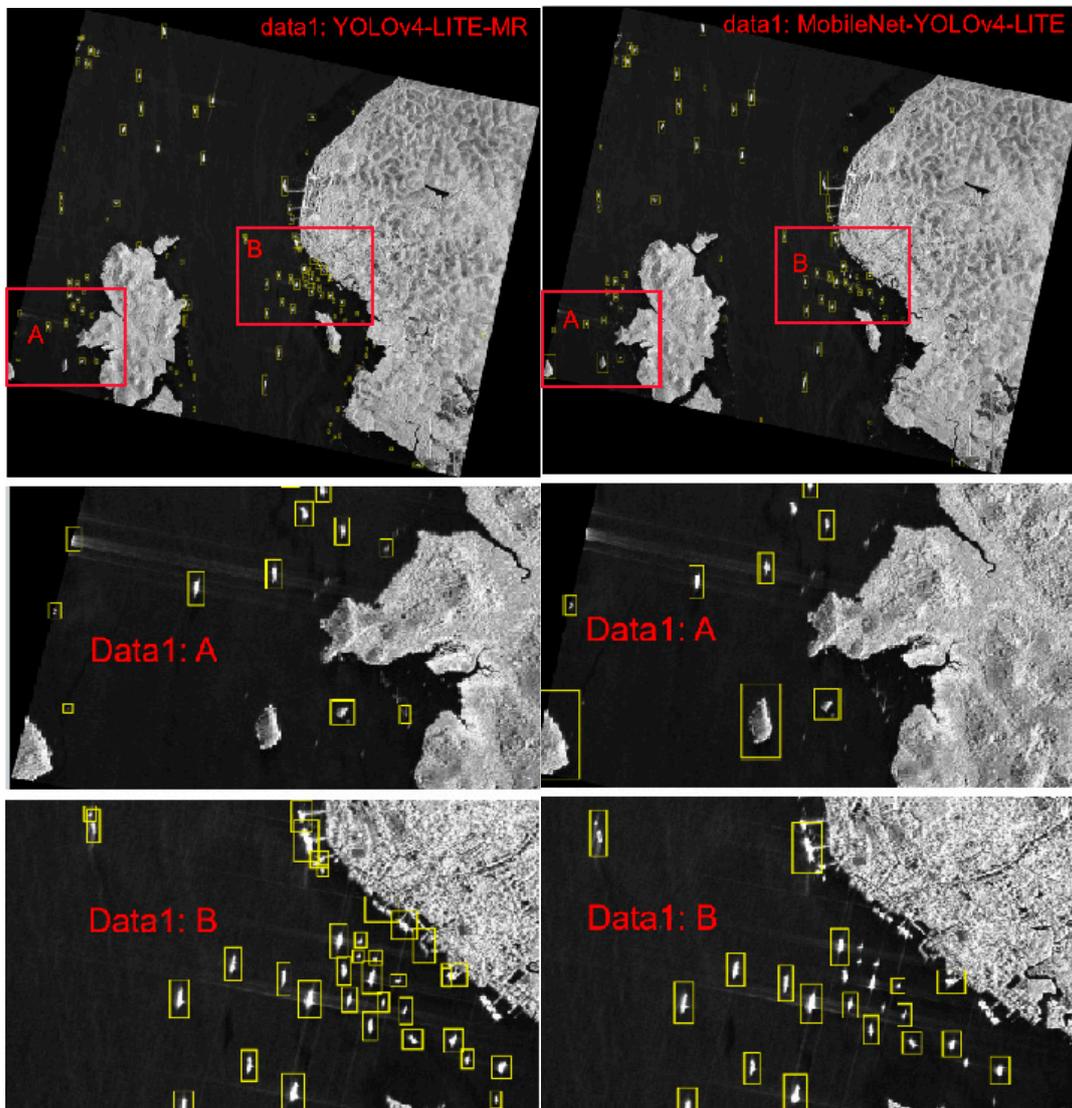


Figure 12. Comparison of the detection results of YOLOv4-LITE-MR (left column) and MobileNet-YOLOv4-LITE (right column) in data1. A and B are sub-images in data1.

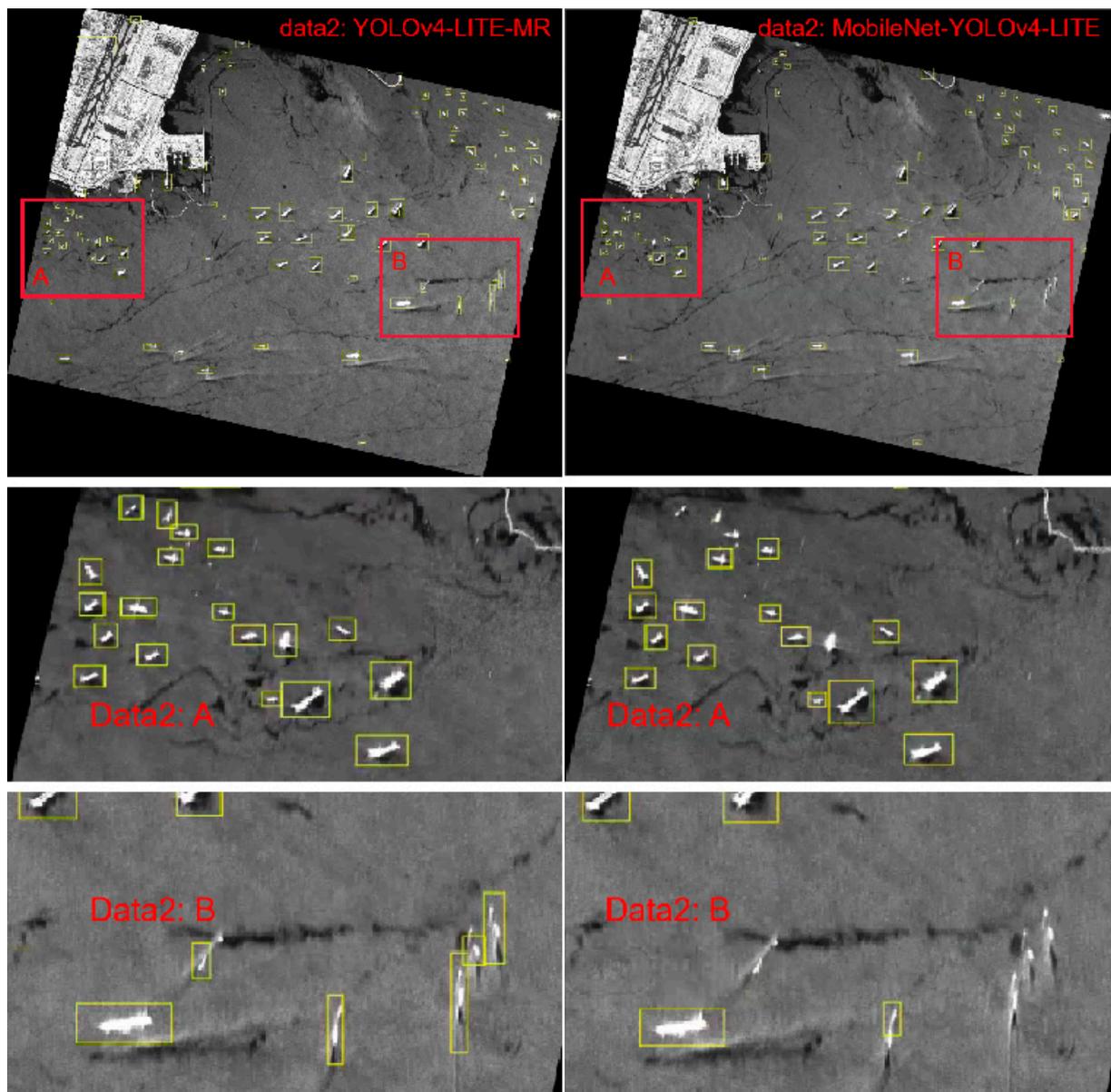


Figure 13. Comparison of the detection results of YOLOv4-LITE-MR (left column) and MobileNet-YOLOv4-LITE (right column) in data2. A and B are sub-images in data2.

Figures 12–14 and Table 8 respectively show the comparison of YOLOv4-LITE-MR and MobileNet-YOLOv4-LITE and the detection accuracy combined with AIS data. The results show that the method proposed in this paper has advantages in detection accuracy and detection effect for ship detection in the whole SAR image. For data1, the accuracy and FOM reached 92.68% and 0.835, respectively, increased by 10.91% and 0.078, respectively; for data2, 88.04% and 0.757, 14.13% and 0.084 respectively; and for data3, 86.24% and 0.839, 8.26% and 0.214 respectively. To some extent, it solves the problems of large-scale SAR image input, a large number of calculations, and many false alarm targets in ship target detection. It can achieve good detection performance in complex scenes, such as high sea conditions, large ship density, and serious sidelobe phenomenon., effectively improving the detection effect of small target ships and inhibiting the repeated frame selection of ship targets.

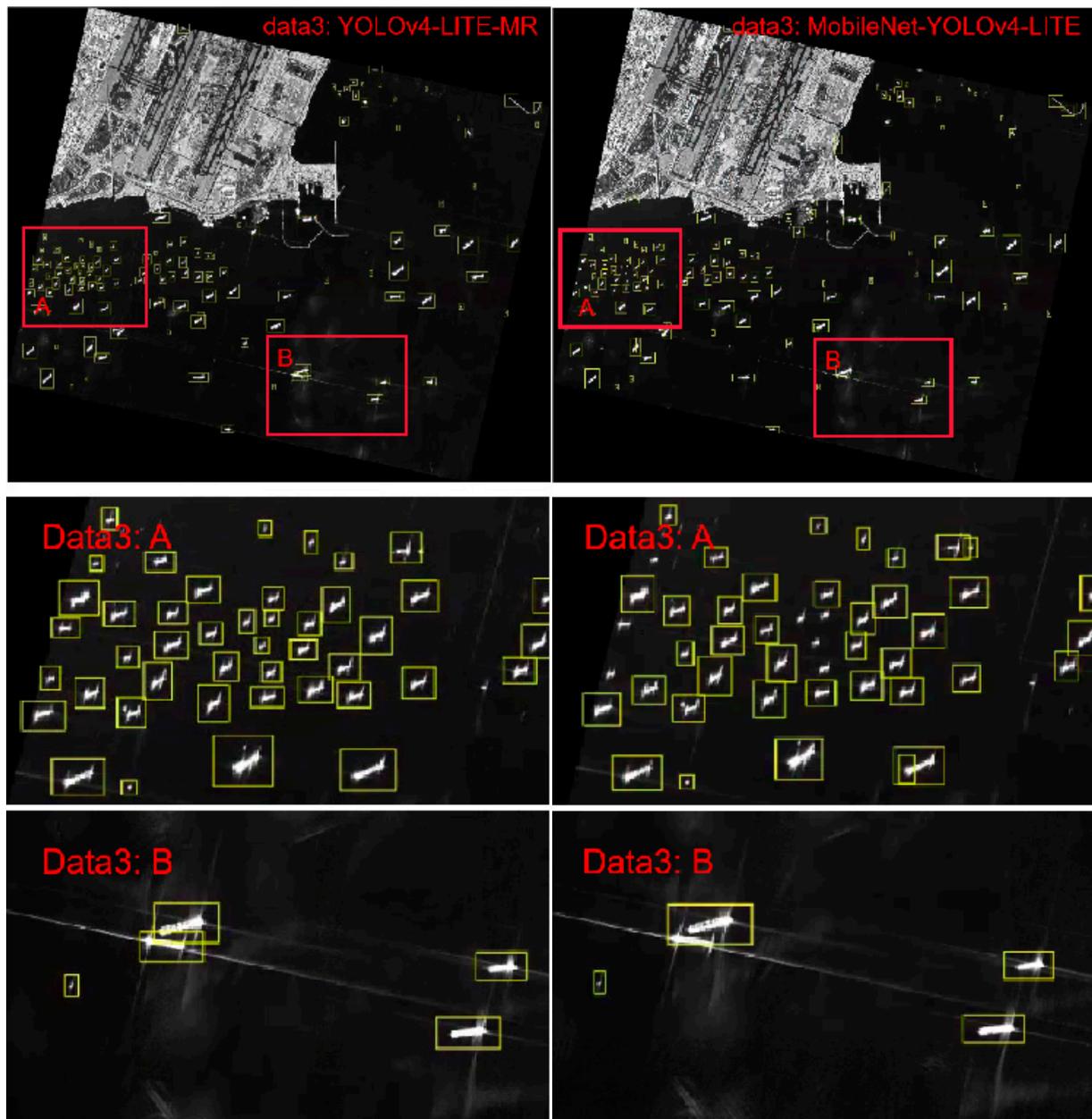


Figure 14. Comparison of the detection results of YOLOv4-LITE-MR (left column) and MobileNet-YOLOv4-LITE (right column) in data3. A and B are sub-images in data3.

Table 8. Comparison of detection accuracy of whole SAR images.

Data	Model	Total Number	Correct Detection	Missing Target	False Alarm Target	Accuracy/%	FOM
data1	YOLOv4-LITE-MR	82	76	6	9	92.68	0.835
	MobileNet-YOLOv4-LITE	82	64	18	12	81.71	0.681
data2	YOLOv4-LITE-MR	92	81	8	15	88.04	0.757
	MobileNet-YOLOv4-LITE	92	68	24	9	73.91	0.673
data3	YOLOv4-LITE-MR	109	94	15	13	86.24	0.839
	MobileNet-YOLOv4-LITE	109	85	24	27	77.98	0.625

As a lightweight network for real-time ship target detection on board, the proposed improved model is better than YOLOv4 in terms of detection speed and accuracy. Compared

to MobileNet-YOLOv4-LITE, it brings 2.29% and 2.91% improvement in the model's mean accuracy (*mAP*) and accuracy, with only a 2.89M increase in model size. It can balance the detection accuracy and detection speed of the network. However, during the experiments, it was found that there is a large area of sea area without ship targets in the sub-blocks of the SAR image of the actual scene acquired by the sliding window method. Moreover, if the target detection is also performed for these scenes, many computational resources will be wasted, and the detection speed will be reduced. It is believed that solving the above problem can significantly improve the detection speed of the actual scene SAR images and meet the demand of real-time ship target detection.

5. Conclusions

Existing SAR ship detection methods can not meet real-time satellite detection, while limited hardware resources require ship detection algorithms with lightweight and high accuracy. Based on the lightweight network, improvements are made based on the YOLOv4-LITE model. The MobileNetv2 network is used as the backbone feature extraction network. The deep separable convolution is replaced to reduce the computational overhead during network training and ensure the lightweight characteristic of the network. By adding and improving the RFB module, which can strengthen the deep features learned by the lightweight CNN model, the feature fusion effect of the model and the detection accuracy of small-scale ship targets are enhanced. The sliding window block method is designed to solve the input problem of the whole SAR image. The SAR image experiments show that the designed sliding window block method can detect small-scale ship targets and reduce the generation of false alarm targets. At the same time, it improves the detection effect of coastal ships to a certain extent. Compared with other mainstream algorithms, the improved algorithm achieves higher detection accuracy and realizes the rapid and accurate detection of ship targets. It provides a reference for building a spaceborne ship detection network that combines the actual needs, balances the detection accuracy and speed of the network, and pursues a lightweight network.

While achieving the above results, there are also shortcomings: the detection effect of ship targets in densely docked near coast areas needs to be improved, the detection of sea areas without ship targets in the sub-map will cause a waste of computing resources, and the size of the model needs to be further compressed. In future work, we will expand the research work in this area and optimize the detection method by using semantic segmentation and network pruning methods.

Author Contributions: Conceptualization, S.L. and W.K.; methodology, W.K., L.Z.; software, X.C., W.K., J.L.; validation, L.Z., M.X.; formal analysis, W.K., L.Z.; investigation, X.C.; resources, J.L.; data curation, M.Y.; writing—original draft preparation, W.K., M.Y., X.C.; writing—review and editing, S.L., X.C.; visualization, X.C., W.K.; supervision, S.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key Research and Development Program of China (Grant No. 2017YFC1405600 and Grant No. 2019YFE0126600) and the National Natural Science Foundation of China (Grant No. 42171342 and Grant No. 61761021).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: We gratefully appreciate the editor and anonymous reviewers for their efforts and constructive comments, which have greatly improved the technical quality and presentation of this study.

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

References

1. Dong, G.; Kuang, G.; Wang, N.; Wang, W. Classification via sparse representation of steerable wavelet frames on Grassmann manifold: Application to target recognition in SAR image. *IEEE Trans. Image Process.* **2017**, *26*, 2892–2904. [[CrossRef](#)] [[PubMed](#)]
2. Séguin, G.; Ahmed, S. RADARSAT constellation, project objectives and status. In Proceedings of the 2009 IEEE International Geoscience and Remote Sensing Symposium, Cape Town, South Africa, 12–17 July 2009; pp. II-894–II-897.
3. Wang, R.; Deng, Y.; Zhang, Z.; Shao, Y.; Hou, J.; Liu, G.; Wu, X. Double-channel bistatic SAR system with spaceborne illuminator for 2-D and 3-D SAR remote sensing. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 4496–4507. [[CrossRef](#)]
4. Wang, R.; Wang, W.; Shao, Y.; Hong, F.; Wang, P.; Deng, Y.; Zhang, Z.; Loffeld, O. First bistatic demonstration of digital beamforming in elevation with TerraSAR-X as an illuminator. *IEEE Trans. Geosci. Remote Sens.* **2015**, *54*, 842–849. [[CrossRef](#)]
5. Lan, D.; Zhaocheng, W.; Yan, W.; Di, W.; Lu, L. Survey of research progress on target detection and discrimination of single-channel SAR images for complex scenes. *J. Radars.* **2020**, *9*, 34–54.
6. Saepuloh, A.; Bakker, E.; Suminar, W. The significance of SAR remote sensing in volcano-geology for hazard and resource potential mapping. In Proceedings of the AIP Conference Proceedings, Hangzhou, China, 8 May 2017; p. 070005.
7. Schumacher, R.; Schiller, J. Non-cooperative target identification of battlefield targets-classification results based on SAR images. In Proceedings of the IEEE International Radar Conference, Arlington, VA, USA, 9–12 May 2005; pp. 167–172.
8. Olsen, R.B.; Bugden, P.; Andrade, Y.; Hoyt, P.; Lewis, M.; Edell, H.; Bjerkelund, C. Operational use of RADARSAT SAR for marine monitoring and surveillance. In Proceedings of the 1995 International Geoscience and Remote Sensing Symposium, IGARSS'95, Quantitative Remote Sensing for Science and Applications, Firenze, Italy, 10–14 July 1995; pp. 224–226.
9. Qi, B.; Shi, H.; Zhuang, Y.; Chen, H.; Chen, L. On-board, real-time preprocessing system for optical remote-sensing imagery. *Sensors* **2018**, *18*, 1328. [[CrossRef](#)]
10. Yang, Y.; Liao, Y.; Ni, S.; Lin, C. Study of Algorithm for Aerial Target Detection Based on Lightweight Neural Network. In Proceedings of the 2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE), Guangzhou, China, 15–17 January 2021; pp. 422–426.
11. Wang, Y.; Liu, H. A hierarchical ship detection scheme for high-resolution SAR images. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 4173–4184. [[CrossRef](#)]
12. Han, Z.; Chong, J. A review of ship detection algorithms in polarimetric SAR images. In Proceedings of the Proceedings 7th International Conference on Signal Processing, 2004. Proceedings. ICSP'04, Beijing, China, 31 August–4 September 2004; pp. 2155–2158.
13. Xu, F.; Liu, J.-h. Ship detection and extraction using visual saliency and histogram of oriented gradient. *Optoelectron. Lett.* **2016**, *12*, 473–477. [[CrossRef](#)]
14. Zhou, D.Z.; Zeng, L.; Zhang, K. A novel SAR target detection algorithm via multi-scale SIFT features. *J. Northwest. Polytech. Univ.* **2015**, *33*, 867–873.
15. Ai, J.; Qi, X.; Yu, W.; Deng, Y.; Liu, F.; Shi, L.; Jia, Y. A novel ship wake CFAR detection algorithm based on SCR enhancement and normalized Hough transform. *IEEE Geosci. Remote Sens. Lett.* **2011**, *8*, 681–685.
16. Kang, M.; Leng, X.; Lin, Z.; Ji, K. A modified faster R-CNN based on CFAR algorithm for SAR ship detection. In Proceedings of the 2017 International Workshop on Remote Sensing with Intelligent Processing (RSIP), Shanghai, China, 18–21 May 2017; pp. 1–4.
17. Leng, X.; Ji, K.; Yang, K.; Zou, H. A bilateral CFAR algorithm for ship detection in SAR images. *IEEE Geosci. Remote Sens.* **2015**, *12*, 1536–1540. [[CrossRef](#)]
18. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Hangzhou, China, 8 May 2017; pp. 580–587.
19. Girshick, R. Fast r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1440–1448.
20. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *Adv. Neural Inf. Process. Syst.* **2015**, *28*, 91–99. [[CrossRef](#)]
21. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; Berg, A.C. Ssd: Single shot multibox detector. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 8–16 October 2016; pp. 21–37.
22. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
23. Redmon, J.; Farhadi, A. YOLO9000: Better, faster, stronger. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 7263–7271.
24. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
25. Willburger, K.; Schwenk, K.; Brauchle, J. Amaro—An on-board ship detection and real-time information system. *Sensors* **2020**, *20*, 1324. [[CrossRef](#)]
26. Li, X.; Chen, P.; Fan, K. Overview of Deep Convolutional Neural Network Approaches for Satellite Remote Sensing Ship Monitoring Technology. In Proceedings of the IOP Conference Series: Materials Science and Engineering, Nanjing, China, 27–29 September 2019; p. 012071.
27. Zhang, K.; Luo, Y.; Liu, Z. Overview of research on marine target recognition. In Proceedings of the 2nd International Conference on Computer Vision, Image, and Deep Learning, Liuzhou, China, 25–28 June 2021; pp. 273–282.

28. Kang, M.; Ji, K.; Leng, X.; Lin, Z. Contextual region-based convolutional neural network with multilayer fusion for SAR ship detection. *Remote Sens.* **2017**, *9*, 860. [[CrossRef](#)]
29. Du, L.; Dai, H.; Wang, Y.; Xie, W.; Wang, Z. Target discrimination based on weakly supervised learning for high-resolution SAR images in complex scenes. *IEEE Trans. Geosci. Remote Sens.* **2019**, *58*, 461–472. [[CrossRef](#)]
30. Zhao, Y.; Zhao, L.; Li, C.; Kuang, G. Pyramid attention dilated network for aircraft detection in SAR images. *IEEE Geosci. Remote Sens. Lett.* **2020**, *18*, 662–666. [[CrossRef](#)]
31. An, Q.; Pan, Z.; Liu, L.; You, H. DRBox-v2: An improved detector with rotatable boxes for target detection in SAR images. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 8333–8349. [[CrossRef](#)]
32. 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 SAR ship detection. *IEEE Access* **2019**, *7*, 159262–159283. [[CrossRef](#)]
33. 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](#)]
34. Wang, Z.; Wang, B.; Xu, N. SAR ship detection in complex background based on multi-feature fusion and non-local channel attention mechanism. *Int. J. Remote Sens.* **2021**, *42*, 7519–7550. [[CrossRef](#)]
35. Wu, Z.; Hou, B.; Ren, B.; Ren, Z.; Wang, S.; Jiao, L. A deep detection network based on interaction of instance segmentation and object detection for SAR images. *Remote Sens.* **2021**, *13*, 2582. [[CrossRef](#)]
36. Shi, H.; He, G.; Feng, P.; Wang, J. An On-Orbit Ship Detection and Classification Algorithm for Sar Satellite. In Proceedings of the IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 1284–1287.
37. Deng, Y.; Yu, W.; Zhang, H.; Wang, W.; Liu, D.; Wang, R. Forthcoming spaceborne SAR development. *J. Radars* **2020**, *9*, 1–33.
38. Bouguettaya, A.; Kechida, A.; Taberkit, A.M. A survey on lightweight CNN-based object detection algorithms for platforms with limited computational resources. *Int. J. Inform. Appl. Math.* **2019**, *2*, 28–44.
39. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1251–1258.
40. 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](#)]
41. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv* **2017**, arXiv:1704.04861.
42. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.-C. Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 4510–4520.
43. Bochkovskiy, A.; Wang, C.-Y.; Liao, H.-Y.M. Yolov4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934.
44. Liu, S.; Huang, D. Receptive field block net for accurate and fast object detection. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 385–400.
45. Luo, W.; Li, Y.; Urtasun, R.; Zemel, R. Understanding the effective receptive field in deep convolutional neural networks. In Proceedings of the 30th International Conference on Neural Information Processing Systems, Barcelona, Spain, 5–10 December 2016; pp. 4905–4913.
46. Yu, J.; Jiang, Y.; Wang, Z.; Cao, Z.; Huang, T. Unitbox: An advanced object detection network. In Proceedings of the 24th ACM International Conference on Multimedia, Amsterdam, The Netherlands, 15–19 October 2016; pp. 516–520.
47. Zheng, Z.; Wang, P.; Liu, W.; Li, J.; Ye, R.; Ren, D. Distance-IoU loss: Faster and better learning for bounding box regression. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; pp. 12993–13000.
48. 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.* **2021**, 1–13. [[CrossRef](#)]
49. 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 (BIGSAR DATA), Beijing, China, 13–14 November 2017; pp. 1–6.
50. Lee, S.; Hayes, M.H. Properties of the singular value decomposition for efficient data clustering. *IEEE Signal Process. Lett.* **2004**, *11*, 862–866. [[CrossRef](#)]