

Article



Improved Ship Detection Algorithm Based on YOLOX for SAR Outline Enhancement Image

Sen Li, Xiongjun Fu * and Jian Dong

The School of Integrated Circuits and Electronics, Beijing Institute of Technology, Beijing 100081, China * Correspondence: fuxiongjun@bit.edu.cn

Abstract: Synthetic aperture radar (SAR) ship detection based on deep learning has the advantages of high accuracy and end-to-end processing, which has received more and more attention. However, SAR ship detection faces many problems, such as fuzzy ship contour, complex background, large scale difference and dense distribution of small targets. To solve these problems, this paper proposes a SAR ship detection method with ultra lightweight and high detection accuracy based on YOLOX. Aiming at the problem of speckle noise and blurred ship contour caused by the special imaging mechanism of SAR, a SAR ship feature enhancement method based on high frequency sub-band channel fusion which makes full use of contour information is proposed. Aiming at the requirement of light-weight detection algorithms for micro-SAR platforms such as small unmanned aerial vehicle and the defect of spatial pooling pyramid structure damaging ship contour features, an ultra-lightweight and high performance detection backbone based on Ghost Cross Stage Partial (GhostCSP) and lightweight spatial dilation convolution pyramid (LSDP) is designed. Aiming at the characteristics of ship scale diversity and unbalanced distribution of channel feature information after contour enhancement in SAR images, four feature layers are used to fuse contextual semantic information and channel attention mechanism is used for feature enhancement, and finally the improved ship target detection method based on YOLOX (ImYOLOX) is formed. Experimental tests on the SAR Ship Detection Dataset (SSDD) show that the proposed method achieves an average precision of 97.45% with a parameter size of 3.31 MB and a model size of 4.35 MB, and its detection performance is ahead of most current SAR ship detection algorithms.

Keywords: SAR ship detection; deep learning; YOLOX; high average precision; lightweight network

1. Introduction

Synthetic aperture radar (SAR) is a microwave sensor for imaging based on the scattering characteristics of electromagnetic waves, which can be observed all day and all weather and has certain cloud and ground penetrating ability. It has unique advantages in marine monitoring, mapping, and the military. With the continuous exploitation of marine resources, people have begun to pay attention to the monitoring of marine vessels. Therefore, SAR ship detection algorithm is of great significance in the fields of territorial sea security, maritime law enforcement, and marine ecological protection.

There are some differences between ship target detection in SAR image and optical image. Contour information is one of the important features for target detection. However, since the SAR uses electromagnetic scattered echoes for vector synthesis, the coherently superimposed scattered echoes inevitably experience random fading in amplitude and phase, presenting as light and dark interlaced speckle noise. The special imaging mechanism of SAR causes the outline of the ship to be unclear, which is more serious in the complex background of the ship close to the coast and berthing in the port. In addition, the SAR ship detection algorithm should also consider the limited computing resources of hardware devices in practical application scenarios of micro-SAR platforms such as small unmanned aerial vehicles, the complexity of the algorithm determines the practicability.



Citation: Li, S.; Fu, X.; Dong, J. Improved Ship Detection Algorithm Based on YOLOX for SAR Outline Enhancement Image. *Remote Sens*. 2022, 14, 4070. https://doi.org/ 10.3390/rs14164070

Academic Editor: Deodato Tapete

Received: 14 July 2022 Accepted: 17 August 2022 Published: 20 August 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/). The development of SAR ship detection algorithm is divided into two stages: the traditional detection algorithm represented by constant false alarm rate (CFAR) and the deep learning detection algorithm represented by convolutional neural network.

The most widely used traditional SAR ship detection method is the CFAR detection algorithm, such as literature [1–3] are all algorithms for SAR ship detection based on CFAR. The CFAR algorithm adjusts the detection threshold according to the statistical characteristics of the background local clutter, and judges whether it is a target according to the threshold. The algorithm relies on the modeling of clutter statistical characteristics, so it is sensitive to complex coastline, sea clutter and coherent speckle noise, with low detection accuracy and poor generalization. At present, SAR ship detection based on deep learning is gradually replacing traditional methods, which is becoming the mainstream research direction.

Since the sensational impact of AlexNet in the ImageNet image classification challenge in 2012, deep learning has started a rapid revival in the field of computer vision. Thanks to the high accuracy and reliability of the deep learning method and the simple end-to-end processing, these have led to the rapid development of convolutional neural networkbased target detection algorithms, resulting in the gradual replacement of traditional target detection algorithms. The current mainstream target detection methods based on deep learning are mainly divided into two-stage and one-stage target detection algorithms. Typical representatives of two-stage target detection algorithms are R-CNN [4], Fast R-CNN [5], Faster R-CNN [6] and so on. These two-stage algorithms first generate target candidate regions in the first stage, and then classify and identify and position the candidate regions in the second stage. Typical representatives of one-stage target detection algorithms are SSD [7], RetinaNet [8], YOLO series, etc. Among them, YOLO series target detection algorithms have strong vitality due to their superior performance and have been updated and iterated several versions. These one-stage algorithms do not need to generate candidate regions, and directly predict the category and position coordinate information of the target in one step, and the detection speed is faster.

When the target detection algorithm based on deep learning was just emerging, it could not be applied in the field of SAR ship detection because SAR images were not easy to obtain and there were no public data sets. Fortunately, Li et al. [9] published the first SAR ship detection data set (SSDD) in 2017, which promoted the rapid development of this field. In 2021, Zhang et al. [10] corrected the mislabeling of the initial version of the SSDD dataset and standardized the usage standards. At present, the SSDD dataset has become one of the most widely used datasets for SAR ship detection. In recent years, Cui et al. [11] proposed a dense connection pyramid with enhanced attention mechanism, which makes full use of context features and improves the detection ability of small ships, but this method uses the super large network Resnet101 as the backbone, and the number of training iterations is as high as 50,000. Jiang et al. [12] designed a lightweight SAR ship detection network based on YOLOV4-Tiny, and used non-subsampling Laplace transform to construct a multi-channel SAR image, which enhances the ship contour features, but this method has low adaptability to SAR images with unbalanced channel features. Yu et al. [13] proposed a convolutional network based on a bidirectional convolutional structure, which greatly improves the effectiveness of SAR image feature extraction through the exchange of information between the upper and lower channels, but the convolution operation in two directions by this method increases the number of model parameters and reduces the model inference speed. Zhang et al. [14] proposed a quad feature pyramid network composed of Deformable Convolutional FPN, Content-Aware Feature Reassembly FPN, Path Aggregation Space Attention FPN and Balance Scale Global Attention FPN, which improves the ability of network feature extraction and fusion, but this method uses a large network ResNet50 as the backbone and stacks four feature pyramids with different functions in a serial manner, resulting in high network complexity and slow detection speed. Ma et al. [15] designed an anchor-free framework with skip connections and aggregated nodes based on key point estimation and attention mechanism to fuse multi-resolution features, which improves the

detection ability of multi-scale ship objects. Zhu et al. [16] proposed an anchor-free ship detection algorithm based on fully convolutional one-stage object detection and adaptive training sample selection, which eliminated the influence of anchor points. However, neither of the above [15,16] methods mentioned the amount of network parameters or the size of the model. It can be seen that the SAR ship detection algorithm based on deep learning still has further development space in terms of lightweight and high average precision.

Aiming at the above problems, this paper proposes a SAR ship feature enhancement method based on high-frequency sub-band channel fusion that makes full use of contour information. This method enhances the important profile features of the ship while reducing the influence of speckle noise and provides more effective feature information for the detection network. In addition, this paper selects the advanced YOLOX of the YOLO series as the basic network. It is the first anchor-free target detection algorithm with superior performance in the YOLO series. According to the characteristics of SAR ship detection, this paper makes targeted improvements to YOLOX. Through experimental verification using SSDD dataset on mobile computing platform, the proposed method achieves higher average accuracy with fewer model parameters.

The main contributions of our work are as follows:

- A new SAR ship outline enhancement preprocessing method is proposed, which combines the original single channel SAR image with the outline extracted image, enhances the ship outline feature while weakening the influence of speckle noise, and improves the network's ability to extract key features;
- (2) According to the actual situation that the hardware equipment of micro-SAR platform such as small unmanned aerial vehicle has limited computing resources and the disadvantage that the spatial pooling pyramid structure damages the contour features, the paper takes YOLOX as the basic network, uses Ghost module to carry out lightweight design of Cross Stage Partial (CSP) structure and proposes a lightweight spatial dilation convolution pyramid (LSDP) structure improved by spatial pooled pyramid, and finally designs an ultra lightweight high-performance backbone for SAR ship detection, which is named Ghost Cross Stage Partial Darknet (GhostCDNet);
- (3) A multi-scale feature pyramid structure based on channel attention enhancement (EMFP) is proposed to highlight channel features that make important contributions to object detection. Contextual multi-scale feature fusion using four-layer feature maps enhances the detection capability of multi-scale ship targets.

The remainder of this paper is organized as follows: the second section introduces the proposed method; the third section describes the experiments and analysis of the results; and the fourth section summarizes this paper.

2. Methods

2.1. Overall Architecture

The workflow and overall structure of the method proposed in this paper are shown in Figure 1. The method consists of the following two parts:

One part is the SAR ship outline enhancement preprocessing method shown in the red dashed box in Figure 1 corresponding to contribution 1 described in Chapter 1. The outline enhancement preprocessing method based on side window mean filtering (SWOE) to smooth the speckle noise and extract the SAR ship contour for SAR ship feature enhancement preprocessing. The side window mean filtering is a side-window filtering that uses a mean filtering kernel. Different from the traditional filtering that places the pixels to be processed at the center of the window, the side-window filtering places the pixels to be processed at the edge or diagonal point of the window, which improves the ability to preserve the outline [17], so that the mean filtering kernel can well preserve the outline feature of the ship when smoothing the speckle noise.



Figure 1. The process of the SAR ship detection method proposed in this paper and the network architecture of ImYOLOX.

The other part is the ImYOLOX ship detection network shown in the blue dashed box in Figure 1 corresponding to contribution 2 and contribution 3 described in Chapter 1. According to the characteristics of SAR ship detection, the network has improved the backbone and neck parts of YOLOX, which designs the backbone GhostCDNet based on GhostCSP and LSDP, and the multi-scale feature pyramid based on channel attention enhancement. The YOLOX target detection network uses anchor-free detection heads, which no longer needs to obtain a priori box by clustering, and integrates advanced technologies such as the leading label assignment strategy SimOTA, and classification and regression decoupling [18]. The YOLOX network represents the advanced target detection level of YOLO series, which has achieved superior detection performance on optical image data sets. However, the imaging mechanism of SAR images is different from that of optical images. In addition, micro-SAR platforms such as small unmanned aerial vehicles do not have sufficient computing resources. Therefore, under the premise of ensuring the detection performance, it is necessary to design a lightweight target detection network according to the characteristics of SAR ship detection. The method proposed in this paper will be described in detail from the following aspects.

2.2. SAR Ship Feature Enhancement Method Based on High Frequency Subband Outline Information Fusion

The outline information of ships contains unique features that are difficult to have by interference such as coastline, wharf, coherent speckle noise, etc. Enhancing the outline features of ships can effectively improve the accuracy and sensitivity of ship target detection

and is especially suitable for ship detection in complex offshore background conditions. However, due to the special imaging mechanism of synthetic aperture radar, speckle noise will inevitably appear in SAR images, which seriously affects the imaging quality and reduces the accuracy of target detection. Traditional outline detection filters can not effectively suppress the speckle noise while extracting the ship outline information. Besides, the SAR image denoising and outline extraction method based on deep learning needs to train the deep neural network for image preprocessing before target detection, which greatly increases the complexity, parameters and training time of the model and is not conducive to deployment on mobile devices. Therefore, this paper proposes a simple and efficient SAR ship outline enhancement method based on high frequency sub-band channel fusion, that is, SAR ship feature enhancement through outline extraction based on side window mean filtering (SWOE).

The SWOE which draws on the idea of side window filtering [17] can effectively suppress speckle noise, extract the outline information of ships, and generate new high-frequency channel images for edge feature enhancement. According to the angle distribution characteristics and shape characteristics of the ship's approximate spindle shape, we define 12 side window mean filters with different shapes and directions, denoted as SW_i^n , $n \in [0, 11]$, as shown in Figure 2, where 2r = 6 is the length of the window and (x, y) is the position of the target pixel *i*.



Figure 2. Schematic diagram of the designed 12 side-window mean filtering with different shapes and orientations.

By applying a mean filter to the target pixel *i* in each side window, we can obtain 12 outputs, denoted as I_n , $n \in [0, 11]$. During calculation, overlapping sub windows can be used to reduce repeated calculation.

$$I_n = \frac{1}{M} \sum_{j \in SW_i^n} \omega_{ij} q_j, M_n = \sum_{j \in SW_i^n} \omega_{ij}, n \in [0, 11]$$

$$\tag{1}$$

$$I_{SW} = \arg\min \|q_i - I_n\|_2^2, n \in [0, 11]$$
(2)

where q_i is the pixel value of the target pixel, and ω_{ij} is the weight of the neighbor pixel *j* of the target pixel *i* within the side window range.

In order to retain the outline information, the output of the side window mean filtering should be as close as possible to the input at the outline. Therefore, in the output I_n of the side window mean filtering, I_{SW} with the minimum Euclidean distance from the input pixel value is selected as the input of the Sobel operators. The Sobel operators used is shown in Figure 3.

-1	-2	-1	-1	0	1	-2	-1	0	0	1	2
0	0	0	-2	0	2	-1	0	1	-1	0	1
1	2	1	-1	0	1	0	1	2	-2	-1	0

Figure 3. Sobel operators.

Figure 4 shows the process of constructing a outline enhanced three-channel SAR images based on SWOE. First, the side window mean filtering is used to smooth the speckle noise while retaining the outline information to the greatest extent. Then, the outline feature extraction is performed by the Sobel operators to obtain the high frequency outline information of the ship. The original image as the first and second channels and the high frequency outline information as the third channel are combined into outline enhanced three-channel SAR image. This method not only retains all the information in the original SAR image, but also smooths the speckle noise and enhances the outline features in one of the sub-channels. Compared with the original image, the contour feature information is enriched.



Figure 4. Processing flow of SAR image outline enhancement based on SWOE method.

We show the results of constructing outline enhanced three-channel SAR images based on SWOE by taking the near shore and far sea situations in the SSDD dataset as examples. As shown in Figure 5, Figure 5a shows the original SAR images, Figure 5b shows the images after the side window mean filtering, and Figure 5c shows the high frequency outline information obtained after SWOE processing. Figure 5d shows the outline enhanced three-channel SAR images obtained by channel fusion of the original image and high-frequency outline information. It can be seen that the images in Figure 5b smooth the speckle noise while preserving the outline information to the maximum extent, the intensity of the speckle noise is significantly reduced, and the outline of the ship is clear. The images in Figure 5c contain the complete outline information of the ship and are only disturbed by a small part of severe speckle noise. Figure 5d shows the result of splicing the original image as the first and second channels and the outline information as the third channel, the edge of the ship is highlighted, which enhances the outline feature information contained in the original SAR image.







(d)

Figure 5. Examples of processing results for both near-shore and far-ocean situations in the SAR ship detection dataset SSDD: (a) original SAR image; (b) SAR image after side window mean filtering; (c) SAR image after outline extraction; and (d) SAR image after outline enhancement.

2.3. A Lightweight and Powerful Backbone Based on GhostCSP and LSDP

The backbone CSPDarkNet of YOLOX-Tiny is a high-performance target detection backbone designed based on CSPNet [19] and DarkNet53 [20], with large depth and width. The characteristics of SAR image sample data are few and ships have multi-scale features, which will cause redundancy in large networks and may even cause overfitting during training. In addition, the computing resources of mobile platform hardware devices are limited, which is not enough to support the operation of large deep learning networks.

Therefore, according to these characteristics of SAR ship detection, we compress the channel number of CSPDarkNet and make improvements in lightweight and multi-scale feature fusion, and design a light and strong backbone GhostCDNet, as shown in the dotted box on the left of Figure 1, These improvements are mainly reflected in the following two aspects. First, the standard convolution before CSP is replaced by depthwise separable convolution, and the lightweight design of CSP is carried out using ghost module. Second, the spatial pooling pyramid module is improved using depthwise separable convolution and dilated convolution.

Ghost module [21] is a lightweight convolution module designed for embedded devices, and its structure is shown in Figure 6. In the feature map of a fully trained convolutional neural network, there are many similar feature maps. This redundant feature information promotes the feature extraction of the neural network but causes some unnecessary parameters. The Ghost module uses fewer parameters to generate these similar feature maps. The specific method is to first use the convolution to compress the feature channel to obtain the intrinsic feature map, then use the grouped convolution to expand the feature information, and finally use the identity mapping to splice the intrinsic feature map and the extended feature map to increase the number of feature channels. In order to design a lightweight backbone network, we greatly compress the number of channels of feature maps, use depth-wise separable convolution to replace the standard convolution before each layer of CSP module, and use Ghost module to redesign the CSP module, called GhostCSP, as shown in Figure 6.



Figure 6. Ghost Module and GhostCSP Module.

Due to different ship sizes and different heights of equipment during imaging, the scales of ship targets in SAR images are very different and have the characteristics of multi-scale distribution. The spatial pooling pyramid structure extracts multi-scale feature information by using parallel max-pooling operations of different sizes, but the max-pooling operation to obtain a larger receptive field will greatly reduce the resolution of the feature map and lose important details about the ship outline. Dilated convolution can effectively solve the problem of detail loss caused by acquiring a large receptive field. Modules such as ASPP [22] and RBF [23] are all structures that use dilated convolution to build receptive fields of different sizes to obtain multi-scale target information. Inspired by this, a new structure based on spatial pooling pyramid improvement is designed using depthwise separable convolution and dilated convolution, called Lightweight Spatial Dilated Convolution Pyramid (LSDP). LSDP has a variety of receptive field sizes and has the ability to capture feature information at different scales, while reducing the loss of ship outline information. The structure of LSDP is shown in Figure 7.



Figure 7. The structure of Lightweight Spatial Dilated Convolution Pyramid (LSDP).

LSDP first uses 1×1 depth-wise separable convolution to compress the channel, and then uses depth-wise separable convolutions of different sizes and dilated convolutions with different dilation rates constructed from them to form multiple parallel branches, with dilation rates of 3, 5, 7 and 9 respectively. In addition, there is another branch that is not processed. Finally, the outputs of these parallel branches are spliced together to form a feature map *y* with multi-scale feature information. The process is summarized as follows:

$$x_i = Conv_{1 \times 1}(x), i \in [0, 4]$$
(3)

$$x'_{i} = Conv_{n \times n}^{rate=n} \{ Conv_{n \times n}(x_{i}) \}, i \in [1, 4], n \in \{3, 5, 7, 9\}$$

$$\tag{4}$$

$$y = Conv_{1\times 1} \{ concat \{ x_0, x_1', x_2', x_3', x_4' \} \}$$
(5)

2.4. Multiscale Feature Pyramid Based on Channel Attention Enhancement

Among the features extracted from the backbone network, the deeper the feature map, the larger the receptive field, which can capture more abundant global information, but its spatial resolution is low, so it is not easy to capture the feature information of small targets. On the contrary, the shallower the feature map, the smaller the receptive field, and has more accurate local information, but the global information representation ability is weak. Feature maps at different levels can capture different feature information, so feature pyramid fusion is a good choice, especially for detecting ship targets with large changes in size.

YOLOX uses three feature maps of the backbone network as input to the neck section. Considering that the proportion of space occupied by ships in the SAR image is very small and there are many small ship targets, we add a shallow feature in the neck part to fuse the features of the context semantic information of the four feature maps of 104×104 , 52×52 , 26×26 and 13×13 , so as to enhance the network's ability to detect small targets. In addition, since the features of different channels in each feature layer contribute to the network to extract important information differently, in order to enhance the feature channels that make important contributions, we add an ultra-light channel attention module ECA [24] to the input part of neck, which is an efficient channel attention. The multi-scale feature pyramid based on channel attention enhancement (EMFP) constructed by us is shown in Figure 8.



Figure 8. Structure of multi-scale feature pyramid based on channel attention enhancement (EMFP).

3. Experiments

In this section, the performance of the proposed method for SAR ship detection will be verified by experiments. Firstly, the software environment and hardware equipment of the experiment are described, and then the SAR image data set used is introduced. Finally, the effectiveness of each module is evaluated through ablation experiment and compared with other latest detection algorithms.

3.1. Experimental Environment

These experiments were carried out on a portable mobile laptop. The details of the software environment and hardware configuration of these experiments are shown in Table 1. All experiments were carried out on this mobile platform.

Table 1. Experimental environment.

Configuration	Title 2				
CPU	AMD Ryzen 7 5800 H				
GPU	NVIDIA GeForce RTX 3070 8 GB				
Operating system	Windows10				
Development Environment	Pytorch1.10.0				

3.2. Dataset and Experimental Settings

We use the publicly available SAR ship detection dataset SSDD for experimental validation. The SSDD dataset contains various ships in different environmental backgrounds near shore and far sea, and the weather conditions during imaging are not exactly the same. The dataset is collected by TerraSAR-X, RadarSat-2 and Sentinel-1 in Yantai, China and Visakhapatnam, India. The imaging resolution is 1–15 m, with a total of 1160 images, including 2456 ships. Due to the high acquisition cost of SAR image data, the dataset contains fewer sample images. If the dataset is randomly divided, there will be great uncertainty in the distribution of data samples, which will destroy the consistency of scene distribution between the training set and the test set and is not conducive to the performance comparison between different detection methods. Therefore, the publisher of the SSDD dataset stipulates a strict division standard [10], that is, the images whose data file numbers end with 1 and 9 are required to be determined as the test set. According to this regulation, we divide the dataset into training set, validation set, and test set according to the ratio of 7:1:2, as shown in Table 2. The original SSDD dataset and the new dataset SSDD-EE after SAR image edge enhancement preprocessing according to the method proposed in Section 2.2 are divided in this way, and the division of both training sets and validation sets is guaranteed to be exactly the same.

Table 2. SSDD dataset division.

	Training Set	Validation Set	Test Set
File name	others	others	* 1.jpg, * 9.jpg
Number	812	116	232
Ratio	7	1	2
Size	416 imes 416	416 imes 416	416 imes 416

* n.jpg refers to the last digit of the document number in SSDD dataset is n, e.g., * 1.jpg means 000001.jpg, 000011.jpg, 000021.jpg, etc.

The training and testing of the networks are carried out in the experimental environment described in Section 3.1. Since the number of images in the SSDD dataset is limited, we use the pre-training method to load the pre-trained model trained with the VOC2007 dataset to initialize the network parameters. The training process is divided into two parts: freezing training and unfreezing training. First, load the pre-training model, and perform frozen training in the first 50 epochs, the batch size is set to 32 and the initial learning rate is 0.001, the cosine annealing method is used to dynamically adjust the learning rate, only train and adjust the part behind the backbone network. Unfreeze training is performed in the last 450 epochs, at this time, the batch size is set to 16, and the initial learning rate is 0.0001, the cosine annealing method is also used to dynamically adjust the learning rate.

3.3. Evaluation Indicators

As we all know, the performance evaluation indicators of target detection algorithms include Precision, Recall and Average Precision (AP). Below we briefly introduce the physical meaning and calculation method of these indicators.

The calculation of these indicators uses 4 components, namely True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). TP represents the number of correctly detected ships, TN represents the number of correctly detected backgrounds, FP represents the number of falsely detected ships, and FN represents the number of ships missed. The criterion for the detector to correctly detect the ship is that the Intersection over Union (IoU) between the predicted box and the ground-truth box is greater than the standard value, which is set as 0.5. The calculation announcements of Precision, Recall and AP are as follows:

$$Precision = \frac{TP}{TP + FP}$$
(6)

$$Recall = \frac{TP}{TP + FN}$$
(7)

$$AP = \int_0^1 P(R)dR \tag{8}$$

Precision is the ratio of the number of real ships to the total number of ships determined by the detector, 1-Precision is called the false detection rate. Recall is the ratio of the number of real ships detected by the detector to all real ships, 1-Recall is called the missed detection rate. However, Precision and Recall can only have a one-sided response detection performance, so AP is used to balance the two, which is the area enclosed by the Recall-Precision curve in the *Recall* $\in [0, 1]$ interval. AP can comprehensively evaluate the performance of the target detection network. The larger its value, the better the detection performance of the network.

3.4. Ablation Experimental Results and Analysis

In this section, in order to verify the effect of SWOE-based SAR image outline enhancement preprocessing method, GhostCDNet and EMFP, we design four ablation experiments using SSDD dataset.

The first experiment uses the benchmark network YOLOX-Tiny as a comparison benchmark for subsequent experiments. The second experiment uses our designed GhostCDNet to replace the backbone of YOLOX-Tiny and compress the number of convolutional channels of the Neck part and the detection head, which is mainly used to verify the detection performance of the proposed lightweight backbone. The third experiment uses EMFP to replace the original Neck part on the basis of the second experiment, which is mainly used to verify the effectiveness of the proposed feature pyramid. The fourth experiment uses the SAR image outline enhancement method based on SWOE as the SAR image preprocessing process of the third experiment, which is mainly used to verify the effect of the proposed SAR outline enhancement method on the detection performance. Four experiments are carried out step by step, which finally verifies the effectiveness and superiority of our proposed method. The training set, validation set, test set and hyperparameters are kept the same for all experiments.

The target detection performance evaluation indicators of these four comparative experiments are shown in Table 3. It can be seen that after the lightweight design and improvement of YOLOX-Tiny based on the characteristics of SAR images in experiment 2, the total params of the network decreased from 5,032,866 to 851,474, which achieved the requirements of lightweight and maintained good detection performance. Compared with YOLOX-Tiny, the AP decreased by only 0.87%. Experiment 3 uses EMFP to add a small amount of params to the model, which improves the model's ability to detect multi-scale targets. The AP increases by 0.65 percentage points, and the detection performance is close to the large-scale network YOLOX-Tiny. In experiment 4, the SAR outline enhancement preprocessing is performed by the SWOE method, which further improves the AP. Finally, compared with YOLOX-Tiny, total params decreased by about 82.74%, model size decreased by 77.58% and AP increased by 0.22%. Figure 9 shows the variation curves of Precision-Recall of the four experiments, the abscissa is Recall, the ordinate is Precision, and the area of the shaded part is the AP value.

Figure 9 shows the variation curves of Precision-Recall of the four experiments, the abscissa is Recall, the ordinate is Precision, and the area of the shaded part is the Average Precision value.

Some visualization results of the ablation experiment are shown in Figure 10, in which Figure 10a–d respectively show the experimental results of the ablation experiments No.1–No.4, including different background states of offshore, far-sea and berthed in the port. The green rectangle in the figure is the real ship label, and the yellow rectangle is the detected ship target. The first two rows of Figure 10 show the detection situation under the complex background state of berthing in the port, the third row to the seventh row show the detection situation in the offshore navigation state, in which the seventh row shows the detection results of the dense distribution of small ships in the offshore area. The last row shows the detection in the far sea background state. It can be seen from the visualization results that the large-scale network YOLOX-Tiny has superior detection performance, but in the face of SAR ship detection with the characteristics of small samples, multi-scale and outline blurring, there will still be obvious missed and false detections. The ultra-lightweight detection algorithm proposed by us not only greatly reduces the number of parameters, but also improves the missed detection and false detection in some scenarios. At the same time, it also has excellent detection performance for dense small targets, but there are still some missed detections due to the blurred outline of ships, and false detections due to ship-like reefs with blurred contours and other severe noise disturbances. These cases are significantly improved after preprocessing with our proposed SAR outline enhancement method, as shown in the results in the first and second rows in Figure 10. After using the multi-scale feature pyramid based on channel attention enhancement, the

detection ability of densely distributed small ships is significantly improved, as shown in the seventh row of results in Figure 10. From the quantitative analysis of performance evaluation indicators and the qualitative analysis of visualization results, the superiority of our proposed SAR ship detection method is confirmed.

Table 3. Results of four ablation experiments.

Experiment Number	AP	Model Size	Params
No.1	97.23%	19.40 MB	5,032,866
No.2	96.36%	4.25 MB	851.474
No.3	97.01%	4.35 MB	868,594
No.4	97.45%	4.35 MB	868,594



Figure 9. Precision-recall curves of four ablation experiments: (**a**) experiment-1; (**b**) experiment-2; (**c**) experiment-3; and (**d**) experiment-4.



Figure 10. Cont.





3.5. Comparison with the Latest SAR Ship Detection Methods Using SSDD Datasets

To further demonstrate the advancement and superiority of our proposed method, we compare with the latest SAR ship detection methods experimentally verified using the SSDD dataset, as shown in Table 4, where pre-SSDD represents the SAR images in the original SSDD dataset were preprocessed before input into the target detection network. These detection methods are all based on horizontal rectangular box.

Table 4. Comparison with the latest SAR ship detection methods on SSDD dataset.

Methods	Dataset	AP	Model Size	Params
V4-light [12]	Pre-SSDD	90.37%	30.00 MB	6,563,542
MSSDNet [25]	SSDD	95.60%	25.80 MB	-
Mor-FP Yolov4-Tiny [26]	Pre-SSDD	96.36%	-	6,353,237
LPEDet [27]	SSDD	97.40%	-	1,488,978
ASAFE [28]	SSDD	95.19%	-	-
Unnamed method * [29]	SSDD	95.50%	10.30 MB	-
ImYOLOX (ours)	SSDD	97.01%	4.35 MB	868,594
ImYOLOX (ours)	Pre-SSDD	97.45%	4.35 MB	868,594

* The authors of this paper did not name the proposed overall method.

It should be noted that the latest SAR ship detection algorithm compared in Table 4 cannot be reproduced using the same experimental equipment, experimental environment and experimental parameters because there is no open-source code, so we directly quote the relevant performance indicators in the corresponding papers. It can be seen that our proposed method achieves higher AP detection performance with fewer parameters than the latest SAR ship detection methods.

4. Discussion

The experimental results on the SSDD dataset and the comparison results with other latest SAR ship detection algorithms demonstrate the superiority of the method proposed in this paper. However, the method proposed is based on horizontal rectangular box, as shown in Figure 11a. This horizontal rectangular box target detection can not obtain the angle distribution information of the ship and the rectangular box contains the background information of the adjacent area of the target. The rotatable rectangular box is shown in Figure 11b, it can be seen that it can more accurately represent the ship target, reduce the background information, and help to obtain the ship heading and aspect ratio information. Therefore, the follow-up research direction is how to obtain the rotation angle information of the ship, so that the bounding box can more accurately surround the ship target and further improve the detection performance.



Figure 11. Examples of horizontal rectangle bounding box and rotatable rectangle bounding box: (a) horizontal rectangle bounding box in a SAR image; and (b) rotatable rectangle bounding box in a SAR image.

5. Conclusions

In this paper, we propose an ultra-lightweight and high average precision anchor-free target detection algorithm ImYOLOX for SAR ship detection. This method is based on the advanced one-stage target detection algorithm YOLOX. Aiming at the problems faced by SAR ship detection, such as complex background, large scale differences, dense distribution of small targets and limited computing resources of hardware equipment of micro-SAR platforms such as small unmanned aerial vehicles, a light and strong backbone network GhostCDNet and a multi-scale feature pyramid based on channel attention enhancement are designed to improve YOLOX. In addition, aiming at the problems of unclear ship outline information caused by speckle noise in SAR image, a SAR ship feature enhancement method based on high frequency sub-band channel fusion is proposed, which reduces the influence of speckle noise and enhances the ship outline features. The experimental results based on the SSDD dataset show that the proposed method achieves average precision of 97.45% with 868,594 parameters and a model size of 4.35 MB. From both quantitative and qualitative perspectives, the proposed method has superior detection performance. The comparison results show that this method is superior to other SAR ship detection methods.

There are some limitations in horizontal box target detection. In the future, we will study the SAR ship rotating box target detection with angle information on this basis, so as to achieve the purpose of obtaining the angle of the ship target while detecting it. **Author Contributions:** Conceptualization, X.F.; Investigation, S.L. and J.D.; Methodology, S.L. and J.D.; Project administration, X.F.; Resources, X.F.; Validation, S.L.; Writing—original draft, S.L.; Writing—review and editing, S.L., J.D. and X.F. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by 111 Project of China under Grant B14010.

Data Availability Statement: Not applicable.

Acknowledgments: We gratefully appreciate the publishers of the SSDD dataset, the authors of YOLOX, and the editors and reviewers for their efforts and contributions.

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

References

- 1. Schwegmann, C.P.; Kleynhans, W.; Salmon, B.P. Manifold adaptation for constant false alarm rate ship detection in South African oceans. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 2015, *8*, 3329–3337. [CrossRef]
- 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]
- Pappas, O.; Achim, A.; Bull, D. Superpixel-level CFAR detectors for ship detection in SAR imagery. *IEEE Geosci. Remote Sens. Lett.* 2018, 15, 1397–1401. [CrossRef]
- 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; pp. 580–587.
- Girshick, R. Fast R-CNN. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1440–1448.
- 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]
- Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.E.; 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.
- Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal loss for dense object detection. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2980–2988.
- 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 BigData Era: Models, Methods and Applications, Beijing, China, 13–14 November 2017; pp. 1–6.
- 10. Zhang, T.; Zhang, X.; Li, J.; Xu, X.; Wang, B.; Zhan, X.; Wei, S. Sar ship detection dataset (ssdd): Official release and comprehensive data analysis. *Remote Sens.* **2021**, *13*, 3690. [CrossRef]
- 11. 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]
- 12. Jiang, J.; Fu, X.; Qin, R.; Wang, X.; Ma, Z. High-speed lightweight ship detection algorithm based on YOLO-v4 for three-channels RGB SAR image. *Remote Sens.* 2021, *13*, 1909. [CrossRef]
- 13. Yu, L.; Wu, H.; Zhong, Z.; Zheng, L.; Deng, Q.; Hu, H. TWC-Net: A SAR ship detection using two-way convolution and multiscale feature mapping. *Remote Sens.* **2021**, *13*, 2558. [CrossRef]
- 14. Zhang, T.; Zhang, X.; Ke, X. Quad-FPN: A novel quad feature pyramid network for SAR ship detection. *Remote Sens.* 2021, 13, 2771. [CrossRef]
- 15. Ma, X.; Hou, S.; Wang, Y.; Wang, J.; Wang, H. Multiscale and Dense Ship Detection in SAR Images Based on Key-Point Estimation and Attention Mechanism. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–11. [CrossRef]
- 16. Zhu, M.; Hu, G.; Li, S.; Zhou, H.; Wang, S.; Feng, Z. A Novel Anchor-Free Method Based on FCOS+ ATSS for Ship Detection in SAR Images. *Remote Sens.* 2022, 14, 2034. [CrossRef]
- 17. Yin, H.; Gong, Y.; Qiu, G. Side window filtering. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 8758–8766.
- 18. Ge, Z.; Liu, S.; Wang, F.; Li, Z.; Sun, J. Yolox: Exceeding yolo series in 2021. arXiv 2021, arXiv:2107.08430.
- Wang, C.Y.; Liao, H.Y.M.; Wu, Y.H.; Chen, P.Y.; Hsieh, J.W.; Yeh, I.H. CSPNet: A new backbone that can enhance learning capability of CNN. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 390–391.
- 20. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. arXiv 2018, arXiv:1804.02767.
- Han, K.; Wang, Y.; Tian, Q.; Guo, J.; Xu, C.; Xu, C. Ghostnet: More features from cheap operations. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 1580–1589.
- Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Trans. Pattern Anal. Mach. Intell.* 2017, 40, 834–848. [CrossRef] [PubMed]
- 23. Liu, S.; Huang, D. Receptive field block net for accurate and fast object detection. In Proceedings of the European Conference on Computer Vision, Munich, Germany, 8–14 September 2018; pp. 385–400.

- Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Hu, Q. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 11531–11539.
- 25. Zhou, K.; Zhang, M.; Wang, H.; Tan, J. Ship Detection in SAR Images Based on Multi-Scale Feature Extraction and Adaptive Feature Fusion. *Remote Sens.* 2022, 14, 755. [CrossRef]
- 26. Zhao, C.; Fu, X.; Dong, J.; Qin, R.; Chang, J.; Lang, P. SAR Ship Detection Based on End-to-End Morphological Feature Pyramid Network. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 4599–4611. [CrossRef]
- Feng, Y.; Chen, J.; Huang, Z.; Wan, H.; Xia, R.; Wu, B.; Xing, M. A Lightweight Position-Enhanced Anchor-Free Algorithm for SAR Ship Detection. *Remote Sens.* 2022, 14, 1908. [CrossRef]
- Shi, H.; Fang, Z.; Wang, Y.; Chen, L. An Adaptive Sample Assignment Strategy Based on Feature Enhancement for Ship Detection in SAR Images. *Remote Sens.* 2022, 14, 2238. [CrossRef]
- 29. Yu, W.; Wang, Z.; Li, J.; Luo, Y.; Yu, Z. A Lightweight Network Based on One-Level Feature for Ship Detection in SAR Images. *Remote Sens.* **2022**, 14, 3321. [CrossRef]