


## Article

# A Fast and Lightweight Detection Network for Multi-Scale SAR Ship Detection under Complex Backgrounds

Jimin Yu <sup>1</sup>, Guangyu Zhou <sup>1</sup>, Shangbo Zhou <sup>2,\*</sup>  and Maowei Qin <sup>1</sup>

<sup>1</sup> College of Automation, Chongqing University of Posts and Telecommunications, Chongqing 400065, China; yujm@cqupt.edu.cn (J.Y.); s200303026@stu.cqupt.edu.cn (G.Z.); s190331093@stu.cqupt.edu.cn (M.Q.)

<sup>2</sup> College of Computer Science, Chongqing University, Chongqing 400044, China

\* Correspondence: shbzhou@cqu.edu.cn

**Abstract:** It is very difficult to detect multi-scale synthetic aperture radar (SAR) ships, especially under complex backgrounds. Traditional constant false alarm rate methods are cumbersome in manual design and weak in migration capabilities. Based on deep learning, researchers have introduced methods that have shown good performance in order to get better detection results. However, the majority of these methods have a huge network structure and many parameters which greatly restrict the application and promotion. In this paper, a fast and lightweight detection network, namely FASC-Net, is proposed for multi-scale SAR ship detection under complex backgrounds. The proposed FASC-Net is mainly composed of ASIR-Block, Focus-Block, SPP-Block, and CAPE-Block. Specifically, without losing information, Focus-Block is placed at the forefront of FASC-Net for the first down-sampling of input SAR images at first. Then, ASIR-Block continues to down-sample the feature maps and use a small number of parameters for feature extraction. After that, the receptive field of the feature maps is increased by SPP-Block, and then CAPE-Block is used to perform feature fusion and predict targets of different scales on different feature maps. Based on this, a novel loss function is designed in the present paper in order to train the FASC-Net. The detection performance and generalization ability of FASC-Net have been demonstrated by a series of comparative experiments on the SSDD dataset, SAR-Ship-Dataset, and HRSID dataset, from which it is obvious that FASC-Net has outstanding detection performance on the three datasets and is superior to the existing excellent ship detection methods.

**Keywords:** deep learning; ASIR-Block; CAPE-Block; ship detection; SAR



**Citation:** Yu, J.; Zhou, G.; Zhou, S.; Qin, M. A Fast and Lightweight Detection Network for Multi-Scale SAR Ship Detection under Complex Backgrounds. *Remote Sens.* **2022**, *14*, 31. <https://doi.org/10.3390/rs14010031>

Academic Editor: Stefano Perna

Received: 9 November 2021

Accepted: 16 December 2021

Published: 22 December 2021

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



**Copyright:** © 2021 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) possesses advantages of reconnaissance and all-weather imaging [1]. With the development of airborne and spaceborne SAR, it has been widely used in military and civil fields, such as Gaofen-3, Sentinel-1, TerraSAR-X, and Radarsat-2. As a basic maritime task, SAR ship detection has important value in maritime traffic control, fishery management, and maritime emergency rescue [2]. Up to now, the SAR ship detection field can be roughly divided into two development stages: traditional methods and deep learning-based methods.

Traditional methods include the following three categories: (1) polarization information [3,4]; (2) wavelet transform [5,6]; (3) statistical characteristics [7,8]. Constant false alarm rate (CFAR) [9,10] is the most widely used method among these traditional ones. The CFAR detector calculates the detection threshold adaptively by estimating the statistics of the background clutter and maintains a constant false alarm probability. However, these traditional methods are cumbersome in manual design, complicated in the calculation process, and weak in migration capabilities, which restrict the applications of migration. In addition, these traditional methods require very high professional knowledge for researchers and can easily cause over-fitting problems.

With the gradual maturity of deep learning theory, its application fields are increasingly becoming wider. Deep convolutional neural networks (DCNN) have been shown high

reliability and accuracy in target detection. DCNN can learn stable and efficient features for target detection automatically. Therefore, traditional feature extraction methods are gradually being replaced by convolutional neural networks. The current mainstream target detection methods based on DCNN consist of two categories: two-stage and one-stage.

For the two-stage networks, region proposals containing the approximate position information of the target are generated in the first stage. The second stage mainly consists of fine-tuning the target category and specific location in the region proposals. Representatives of two-stage networks are region-based fully convolutional networks (R-FCN) [11] and faster region-based CNN (Faster R-CNN) [12]. Different from the two-stage networks, the region proposal stage is not necessary in one-stage networks and can directly generate target classification probabilities and locate coordinates. Typical one-stage networks include Single Shot Multibox Detector (SSD) [13], You Only Look Once (YOLO) [14], and Retina-Net [15]. Two-stage networks have advantages in accuracy but low speed. On the contrary, one-stage networks have advantages in detection speed which is conducive to applications on mobile devices with high real-time requirements. With the development of one-stage networks, their detection performance has gradually surpassed two-stage networks and have become the mainstream method in the field of target detection.

Ship detection methods based on deep learning in SAR images have shown outstanding detection performance. A SAR ship region extraction method based on binarized normalized gradient and Fast R-CNN was proposed in [16,17]. After that, based on the context area for SAR ship detection, Kang et al. [18] proposed a multi-layer fusion convolutional neural network which consists of a region proposal network (RPN) with high network resolution and a target detection network with contextual features. YOLO-V2-reduced for SAR ship detection, which was proposed by Chang et al. [19], reduced the parameters and layers based on YOLO-v2, thereby the detection speed improved. In [20], a squeeze and excitation rank mechanism was designed to enhance the SAR ship detection ability of Faster R-CNN. Automatic ship detection on GF-3 multi-resolution image was realized by Wang et al. [21] based on Retina-Net and focal loss.

It was Zhang et al. who designed a lightweight depthwise separable convolution neural network (DS-CNN) based on anchor box mechanism, concatenation mechanism, and multi-scale detection mechanism [22]. To enhance the features of the low-level layers and high-level layers, a new two-way feature fusion module includes a semantic aggregation block and a feature reuse block was designed by Zhang et al. [23].

For multi-scale SAR ship detection, Cui et al. [24] designed a dense attention pyramid network (DAPN) that connects the convolutional block attention module (CBAM) close to each cascaded feature map from the top to the bottom of the pyramid network. For ship detection in high-resolution SAR images, Wei et al. [25] designed a high-resolution ship detection network (HR-SDNet) which makes full use of high-resolution and low-resolution convolutional feature maps through a new high-resolution feature pyramid network (HRFPN). Zhang et al. [26] proposed a quad feature pyramid network (Quad-FPN) for multi-scale SAR ship detection under complex backgrounds. Due to the difficulty of SAR image labeling, using less SAR images to achieve better detection results has become a hot research direction. Rostami et al. [27] proposed a new framework to train a deep neural network for classification without a large number of SAR images. Zhang et al. [28] proposed a multitask learning-based object detector (MTL-Det) that can learn more discriminative target features without increasing the cost of manual labeling.

Although the above methods have shown excellent performances, there are still the following three problems. (1) Most methods focus too much on improving detection accuracy, while the detection speed, which is particularly important in emergency military decision-making and maritime rescue, has been ignored to a certain extent. (2) Most methods have a huge network scale and numerous parameters, which leads to greater challenges in hardware migration (if the number of parameters of a network is less than 4 million, the network can be transplanted to a field-programmable gate array (FPGA) or digital signal processing (DSP) [29]). (3) The aforementioned methods need to be strengthened for multi-scale SAR ship detection under complex backgrounds. Therefore,

a fast and lightweight detection network is proposed for multi-scale SAR ship detection under complex backgrounds in this paper. While increasing the detection speed and reducing the parameters, the proposed FASC-Net maintains a quite satisfactory detection performance. The main contributions of our work are summarized as follows.

1. A Channel-Attention Path Enhancement block (CAPE-Block) is designed by adding a bottom-up enhancement path with a channel-attention mechanism based on feature pyramid networks (FPN), which is used to shorten the path of information transmission and enhance the precise positioning information stored in the low-level feature maps.
2. A fast and lightweight detection network is designed based on CAPE-Block, ASIR-Block, Focus-Block, and SPP-Block for multi-scale SAR ship detection under complex backgrounds.
3. A novel loss function is designed for the training of FASC-Net. Binary cross-entropy loss is used to calculate the object loss and classification loss, and *GloU* loss is used to calculate the loss of the prediction box. Three hyperparameters  $\lambda_1$ ,  $\lambda_2$ , and  $\lambda_3$  are introduced to balance the weights of the three sub-losses.
4. Comparing with other excellent methods (e.g., Faster R-CNN [17], SSD [13], YOLO-V4 [30], DAPN [24], HR-SDNet [25], and Quad-FPN [26]), a series of comparative experiments and ablation studies on the SSDD dataset [31], SAR-Ship-Dataset [32], and HRSID dataset [33] illustrate that our FASC-Net achieves higher mean average precision (mAP) and faster detection speed with smaller number of parameters.

The rest of our paper is organized as follows. Section 2 introduces the datasets and structure of FASC-Net. Section 3 introduces the evaluation criteria, data augmentation methods, and a series of comparative experiments. Ablation studies are used to discuss the effects of key technologies in Section 4. Section 5 concludes this paper.

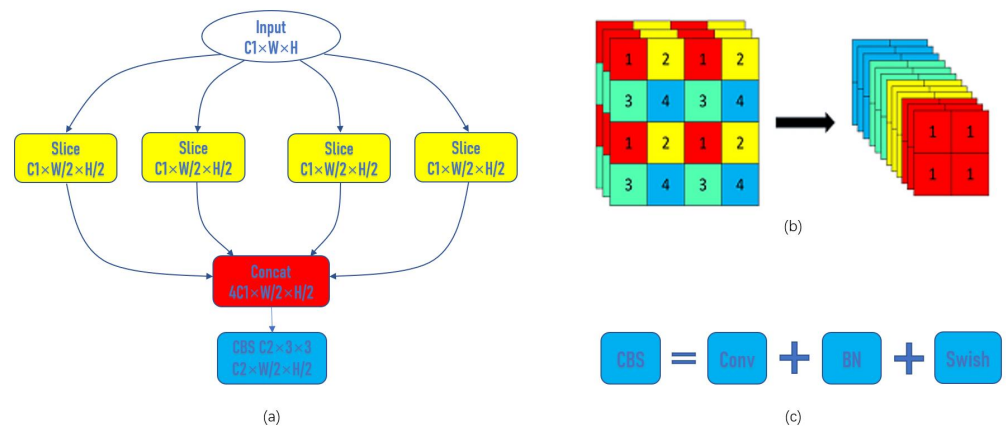
## 2. Materials and Methods

### 2.1. Backbone

#### 2.1.1. Focus-Block

As shown in Figure 1, Focus-Block is the first module of the backbone which slices the input image. Input  $C1 \times W \times H$  represents that the width and height of channels of the input are  $W$  and  $H$  respectively, and the number of channels of the input is  $C1$ , and  $C2$  represents the number of the output channels. The specific operation is to get a value from every other pixel in each channel of the input image, which is similar to the neighbor down-sampling algorithm. In this way, four complementary images with the same shape are obtained. The width and height have become half of the input image and the number of channels has become four times that of the input image. The most important thing is that there is no information loss, as shown in Figure 1b. Just re-arrange all the pixels by getting a value for every other pixel without changing the value of those pixels, so there will be no loss of information. Finally, perform a CBS operation on the newly obtained images. The CBS operation includes convolution, batch normalization, and non-linear activation, as shown in Figure 1c. The activation function of CBS is the Swish activation function.

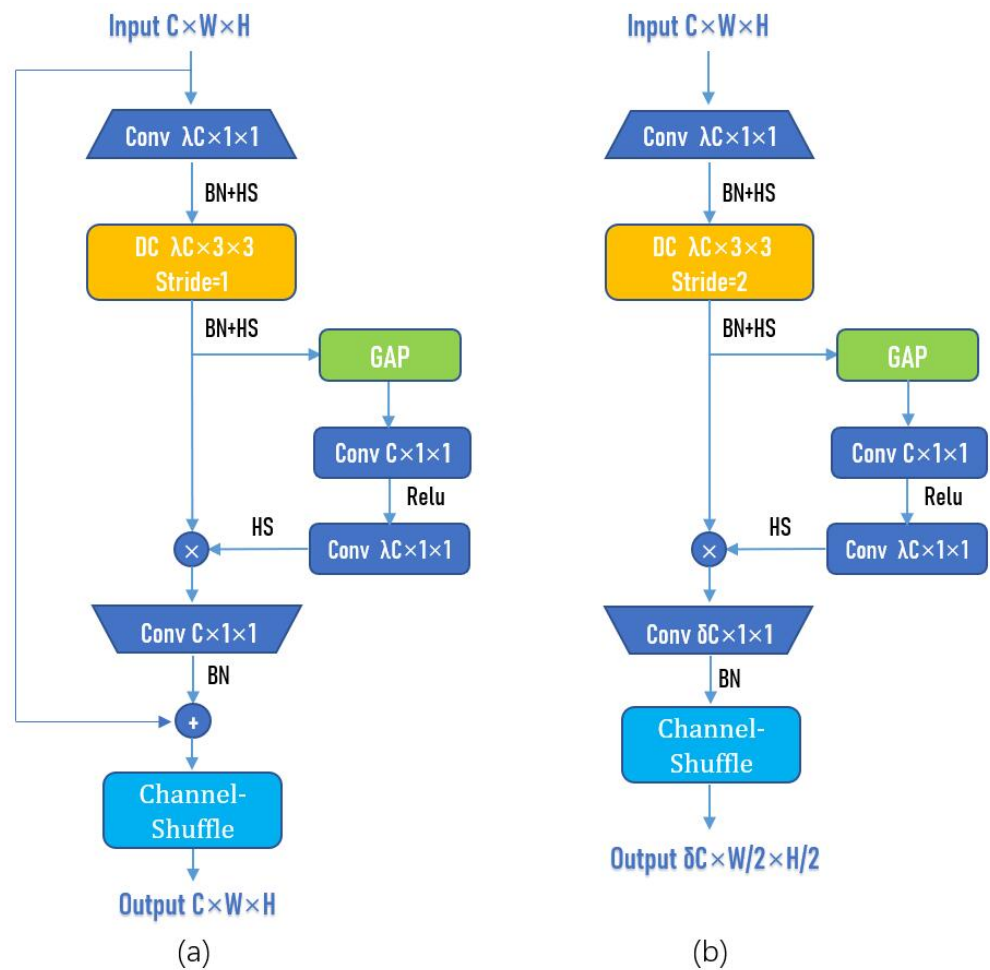
When high-resolution images are input to the network, some detailed information may be lost if a conventional convolution lay is used for the first down-sampling, and there will be a large number of parameters in the conventional convolution layer. Compared with the conventional convolution lays, Focus-Block can reduce the number of parameters, and CUDA memory as well as increase forward and backward speed while minimally impacting mAP. Focus-Block is generally used for the first down-sampling, and the higher the resolution of the input images, the more obvious its advantages.



**Figure 1.** Structure of Focus-Block. (a) Specific operation process of Focus-Block. (b) Schematic diagram of Focus-Block. (c) Structure of CBS.

2.1.2. ASIR-Block

ASIR-Net is a lightweight fully convolutional neural network for SAR automatic target recognition. Compared with existed excellent target recognition networks, it uses fewer parameters to achieve higher recognition accuracy. ASIR-Block is the primary block in ASIR-Net which consists of Channel-Shuffle mechanism, Inverted-Residual block, and Channel-Attention mechanism, as shown in Figure 2.



**Figure 2.** Structure of ASIR-Block. (a) ASIR-Block-1. (b) ASIR-Block-2.

ASIR-Block can be divided into ASIR-Block-1 and ASIR-Block-2. HS, BN, DC, and GAP represent Hard-Swish activation function, batch normalization, depthwise convolution, and global average pooling, respectively.  $\lambda$  and  $\delta$  are hyperparameters-scaling factors and are set by  $\lambda = 4$  and  $\delta = 2$  in this paper. The main function of  $\lambda$  is to increase the number of channels of the input feature map before passing the attention mechanism. The larger the value of  $\lambda$ , the larger the number of channels of the input feature map, and the more parameters in the attention mechanism, so more features can be extracted. However, with the increase of  $\lambda$ , the detection speed of the network will decrease correspondingly. A series of experiments prove that  $\lambda = 4$  is the best balance between detection accuracy and speed. The main function of  $\delta$  is to control the number of channels of the output feature map. The value of  $\delta$  draws on many classic deep learning networks, such as YOLO-V4 and EfficientNet-v2. The scaling factors of feature extraction modules of these algorithms are 2, so the value of  $\delta$  is also set to 2 in this paper.

In ASIR-Block-1, set the stride of depthwise convolution as 1, and the padding is set as the same. The number of convolutional kernels of the last  $1 \times 1$  convolutional layer and the number of channels of the input are the same, so as to ensure that the shape of the output is the same as that of the input, and a shortcut connection can be used.

Shortcut connections can prevent network degradation caused by gradient divergence. The main function of ASIR-Block-1 in FASC-Net is to extract more detailed features with fewer parameters. In ASIR-Block-2, the stride of depthwise convolution is set as 2, the padding is set as the same, and the number of convolutional kernels of the last  $1 \times 1$  convolutional layer is set as  $\delta C$ . The main function of ASIR-Block-2 in FASC-Net is down-sampling. The width and height of the output have become half of the input, and the number of channels of the output has become  $\delta$  times that of the input.

### 2.1.3. SPP-Block

The Spatial-Pyramid-Pooling (SPP) block was proposed in SPP-Net [34] which greatly improved the training speed, test speed, and mAP of R-CNN. The structure of the SPP-Block is very simple and easy to understand, as shown in Figure 3. It uses Maxpool operations of different kernel sizes for feature fusion. The stride of CBS and Maxpool is set as 1 and the padding is set as the same. The function of SPP-Block is to increase the receptive field. It has little effect on the running speed of the entire network, but the performance is significantly improved owing to the fact that it can separate out the most significant context features.

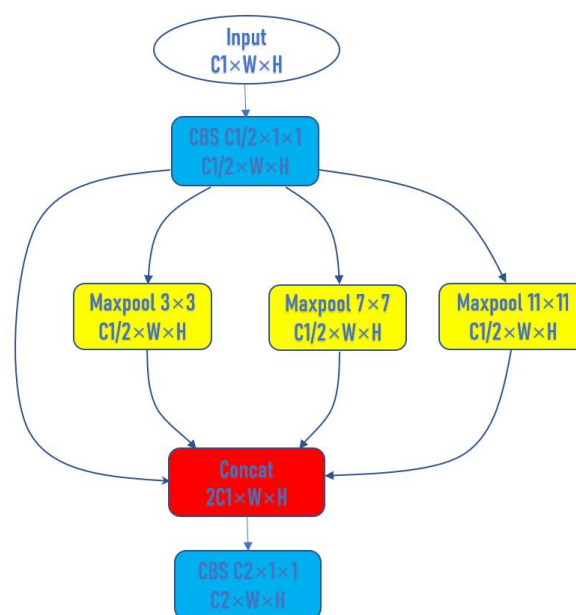


Figure 3. Structure of the SPP-Block.

## 2.2. CAPE-Block, Prediction, and Post-Processing

### 2.2.1. CAPE-Block

Most previous target detection networks use only top-level features for prediction, but we know that although the semantic information of low-level feature maps is less, the target location information is accurate. The semantic information of high-level feature maps is rich, but the target location information is rough. In order to fuse the rich semantic information of high-level features and the precise location information of low-level features, Lin et al. [35] proposed the feature pyramid network (FPN). The main structure of FPN includes a bottom-up path, a top-down path, and some lateral connections, as shown in Figure 4.

The bottom-up path is actually the forward process of the network, which mainly includes some feature extraction layers and down-sampling layers. The main function of the bottom-up path is to compress the size of the feature map and extract rich high-level semantic features. The top-down path mainly contains some feature fusion layers and up-sampling layers. Its main function is to combine rich semantic information with accurate target location information. Most importantly, prediction is made on each feature layer after fusion, which is different from the traditional feature fusion and prediction methods.

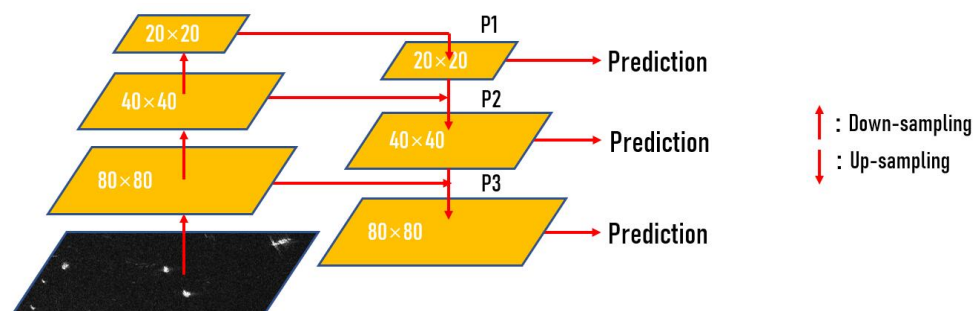


Figure 4. Structure of FPN.

In previous experiments, we found that low-level features help to identify large targets, but the path from high-level features to low-level features is too long, which increases the difficulty of positioning information flow. Therefore, we add a bottom-up enhancement path, which is used to shorten the path of information transmission and use the precise positioning information stored in the low-level features to enhance the ability of FPN.

In addition, we also found that not each channel of feature maps contains the location information of the target. Therefore, we try to design a full convolution Channel-Attention mechanism to enhance those channels that contain accurate target location information and suppress those channels in which target location information is not contained.

The principle of the channel attention mechanism is shown in Figure 5.  $C \times W \times H$  presents that the number of channels, width, and height of the input  $U$  are  $C$ ,  $W$ , and  $H$  respectively. GAP represents global average pooling and different colors represent different weights of channels. The  $1 \times 1$  convolutional layer is used to replace the fully connected layer in the traditional attention mechanism, which can effectively reduce the number of trainable parameters in the attention mechanism and speed up the training speed of the network. After passing the channel attention mechanism, the channel with more target location information will be given a larger weight coefficient, which will play a greater role in the subsequent prediction process, thereby improving the accuracy of prediction boxes. Different colors represent different weights of channels.

By introducing the Channel-Attention mechanism and bottom-up path enhancement in FPN, the channel attention path enhancement block (CAPE-Block) is obtained, as shown in Figure 6. CA represents the Channel-Attention mechanism. In CAPE-Block, ASIR-Block-2 is used to achieve down-sampling, and bilinear interpolation is used to achieve up-sampling. The function of down-sampling is to compress the size of feature maps and extract important semantic information, and the function of up-sampling is to increase the

size of feature maps to facilitate feature fusion with low-level feature maps that store the positioning information. In this way, the three feature layers p1, p2, and p3 after feature fusion have both rich semantic information and precise positioning information. p1, p2, and p3 have different receptive fields that can be used to predict targets of different sizes.

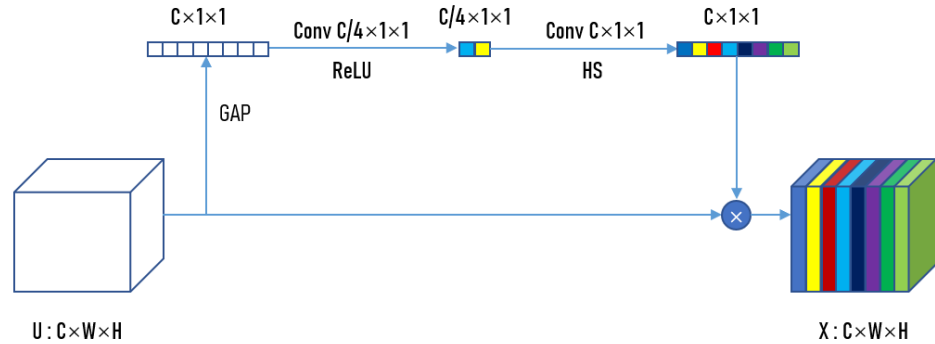


Figure 5. The principle of the channel attention mechanism.

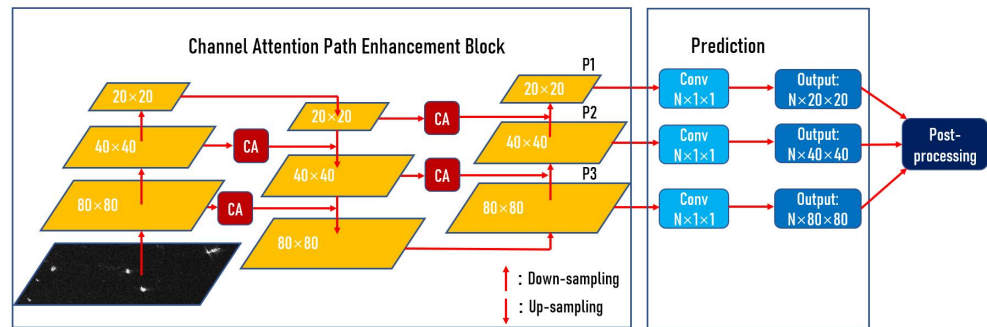


Figure 6. Structure of CAPE-Block and prediction.

### 2.2.2. Prediction

Prediction only consists of three ordinary  $1 \times 1$  convolution layers. The number of convolution kernels of the  $1 \times 1$  convolution layer is  $N$  which can be obtained by:

$$N = (N_c + 4 + 1) \times N_a, \tag{1}$$

where  $N_c$  denotes the number of target categories,  $N_a$  denotes the number of anchors on each pixel of three prediction feature layers. In this paper,  $N_c$  is set as 1, and  $N_a$  is set as 3, so  $N = 18$ . The values of  $N_c$  and  $N_a$  can be changed to apply the proposed FASC-Net to target detection in other fields.

The specific sizes of anchors on the three prediction feature layers are shown in Table 1. The anchors on the P3 layer have a small receptive field and are mainly used to predict small targets. The anchors on the P2 layer have medium receptive fields and are mainly used to predict medium-sized targets. The anchors on the P1 layer have a larger receptive field and are mainly used to predict large targets.

Table 1. The specific sizes of the anchors on the three prediction feature layers.

Layers	Anchors		
P1	(116, 90)	(156, 198)	(373, 326)
P2	(30, 61)	(62, 45)	(59, 119)
P3	(10, 13)	(16, 30)	(33, 23)

### 2.2.3. Post-Processing

The post-processing consists of two steps: decoding and non maximum suppression (NMS). Decoding is to adjust anchors to get revised prediction boxes. The output vector corresponding to each pixel in P1, P2, and P3 is shown in Figure 7, where  $t_x$ ,  $t_y$ ,  $t_w$ , and  $t_h$  represent adjustment parameters of anchors.

*Object* represents whether there is a target in the anchor. If  $object \geq 0.25$ , it is considered that there is a target in the anchor, and the anchor is set as a positive sample. If  $object < 0.25$ , it is considered that there is no target in the anchor, and the anchor is set as a negative sample. *Score* represents the score of the target corresponding to each category. Assume that the center point coordinates of the anchor are  $(c_x, c_y)$ , and the width and height of the anchor are  $p_w$  and  $p_h$  respectively. The revised prediction box can be obtained from:

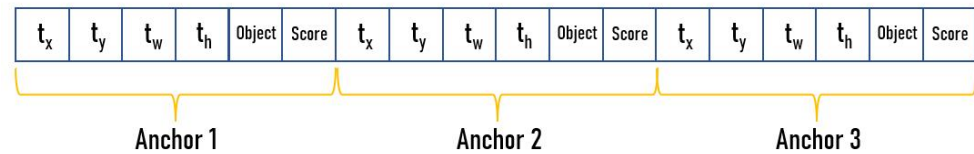
$$b_x = \sigma(t_x) + c_x, \quad (2)$$

$$b_y = \sigma(t_y) + c_y \quad (3)$$

$$b_w = p_w e^{t_w} \quad (4)$$

$$b_h = p_h e^{t_h} \quad (5)$$

where  $b_x$ ,  $b_y$ ,  $b_w$ , and  $b_h$  denote the coordinates of the center point, and width and height of the revised prediction box, respectively.  $\sigma()$  represents *Sigmoid* activation function.



**Figure 7.** Schematic diagram of the components of the output vector.

NMS can eliminate redundant prediction boxes and find the most realistic location of the target. The basic idea of NMS is simple: use the prediction box with the highest confidence to mark an object. If the intersection over union (*IoU*) of other prediction boxes and the prediction box with the highest confidence exceeds a threshold, it is considered that these boxes mark the same object and then discard the other prediction boxes. Suppose  $A$  and  $B$  are two intersecting prediction boxes, the *IoU* of  $A$  and  $B$  can be obtained from:

$$IoU = \frac{A \cap B}{A \cup B}. \quad (6)$$

In this paper, set the hyperparameter *IOU* as 0.6.

### 2.3. Network Architecture of FASC-Net

After introducing the sub-blocks used in FASC-Net, we will introduce the details and construction of FASC-Net, as shown in Figure 8 and Table 2. In Table 2,  $k$  denotes the size of convolution kernels,  $n$  denotes the number of convolution kernels, and  $s$  denotes the stride of convolution kernels. In Figure 8, AB-1 denotes ASIR-Block-1, AB-2 denotes ASIR-Block-2, and US denotes up-sampling.

**Table 2.** Specifications of the proposed FASC-Net.

Input	Operator	n	k	s	Output	Parameters
$3 \times 640 \times 640$	Focus	8	3	-	$8 \times 320 \times 320$	880
$8 \times 320 \times 320$	ASIR-Block-2	8	3	2	$8 \times 160 \times 160$	1496
$8 \times 160 \times 160$	ASIR-Block-1	8	3	1	$8 \times 160 \times 160$	1496
$8 \times 160 \times 160$	ASIR-Block-2	16	3	2	$16 \times 80 \times 80$	1768



Table 2. Cont.

Input	Operator	n	k	s	Output	Parameters
$16 \times 80 \times 80$	ASIR-Block-1	16	3	1	$16 \times 80 \times 80$	5040
$16 \times 80 \times 80$	Channel-Attention	-	-	-	$16 \times 80 \times 80$	280
$16 \times 80 \times 80$	ASIR-Block-2	32	3	2	$32 \times 40 \times 40$	6096
$32 \times 40 \times 40$	ASIR-Block-1	32	3	1	$32 \times 40 \times 40$	18,272
$32 \times 40 \times 40$	Channel-Attention	-	-	-	$32 \times 40 \times 40$	552
$32 \times 40 \times 40$	ASIR-Block-2	64	3	2	$64 \times 20 \times 20$	22,432
$64 \times 20 \times 20$	SSP	64	-	-	$64 \times 20 \times 20$	10,432
$64 \times 20 \times 20$	ASIR-Block-1	64	3	1	$64 \times 20 \times 20$	69,312
$64 \times 20 \times 20$	CBS	32	1	1	$32 \times 20 \times 20$	2112
$32 \times 20 \times 20$	Channel-Attention	-	-	-	$32 \times 20 \times 20$	552
$32 \times 20 \times 20$	Upsample	-	-	-	$32 \times 40 \times 40$	0
$32 \times 40 \times 40$	Concat (-1, 6)	-	-	-	$64 \times 40 \times 40$	0
$64 \times 40 \times 40$	ASIR-Block-1	32	3	1	$32 \times 40 \times 40$	61,056
$32 \times 40 \times 40$	CBS	32	1	1	$32 \times 40 \times 40$	1088
$32 \times 40 \times 40$	Channel-Attention	-	-	-	$32 \times 40 \times 40$	552
$32 \times 40 \times 40$	Upsample	-	-	-	$32 \times 80 \times 80$	0
$32 \times 80 \times 80$	Concat (-1, 4)	-	-	-	$48 \times 80 \times 80$	0
$48 \times 80 \times 80$	ASIR-Block-1	32	3	1	$32 \times 80 \times 80$	36,592
$32 \times 80 \times 80$	Conv	18	1	1	$18 \times 80 \times 80$	576
$32 \times 80 \times 80$	ASIR-Block-2	64	3	2	$64 \times 40 \times 40$	22,432
$64 \times 40 \times 40$	Concat (-11, 14)	-	-	-	$96 \times 40 \times 40$	0
$96 \times 40 \times 40$	ASIR-Block-1	64	3	1	$64 \times 40 \times 40$	140,768
$64 \times 40 \times 40$	Conv	18	1	1	$18 \times 40 \times 40$	1152
$64 \times 40 \times 40$	ASIR-Block-2	64	3	2	$64 \times 20 \times 20$	69,312
$64 \times 20 \times 20$	Concat (-1, 10)	-	-	-	$96 \times 20 \times 20$	0
$96 \times 20 \times 20$	ASIR-Block-1	128	3	1	$128 \times 20 \times 20$	165,472
$128 \times 20 \times 20$	Conv	18	1	1	$18 \times 20 \times 20$	2304

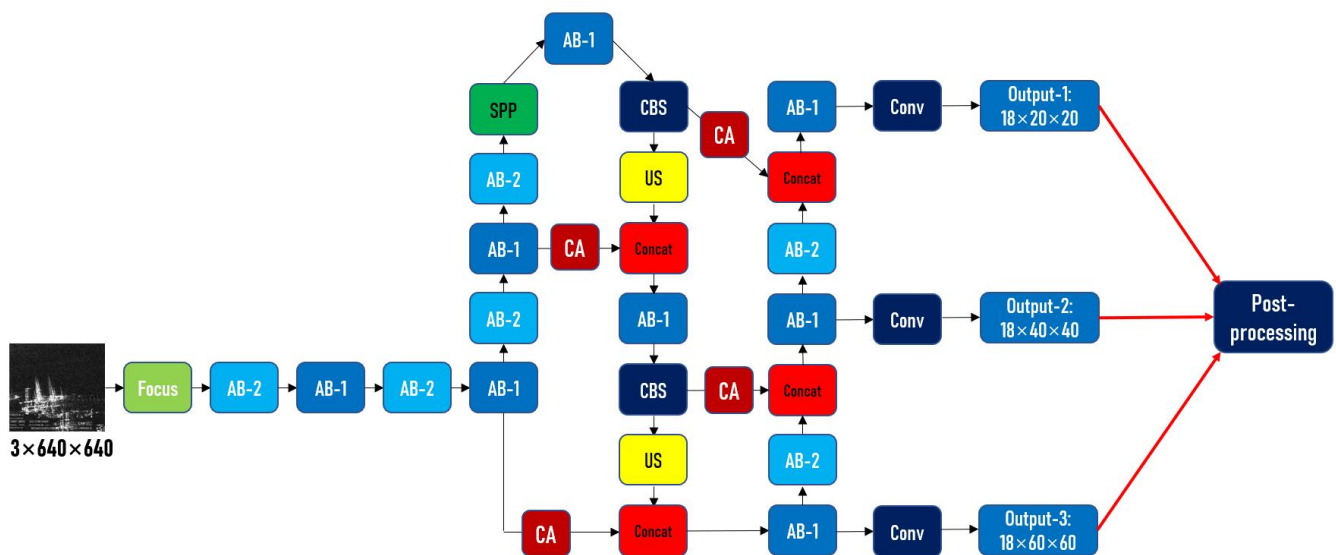


Figure 8. Illustration of the structure of the proposed FASC-Net.

#### 2.4. Training of FASC-Net

To improve the training effect of the proposed FASC-Net, a novel loss function is proposed for FASC-Net in the present work. We use binary cross-entropy loss to calculate the object loss and classification loss and use *GIoU* loss [36] to calculate the loss of the prediction box. *GIoU* loss can well reflect the overlap degree between the prediction box

and the ground truth box and has scale invariance. *GIoU* loss can also solve the shortcoming of *IoU* loss that the *IoU* loss is zero when the two boxes do not intersect. The total loss can be presented as follows.

$$L = \lambda_1 L_{obj}(o, c) + \lambda_2 L_{cla}(O, C) + \lambda_3 L_{GIoU}, \quad (7)$$

where  $\lambda_1$ ,  $\lambda_2$ , and  $\lambda_3$  are hyperparameters that adjust the weights of the three sub-losses.  $L_{obj}$  denotes object loss and  $L_{cla}$  denotes classification loss which can be expressed as

$$L_{obj}(o, c) = -\frac{\sum_i (o_i \ln(\hat{c}_i) + (1 - o_i) \ln(1 - \hat{c}_i))}{N}, \quad (8)$$

$$\hat{c}_i = \text{Sigmoid}(c_i), \quad (9)$$

$$L_{cla}(O, C) = -\frac{\sum_{i \in pos} \sum_{j \in cla} (O_{ij} \ln(\hat{C}_{ij}) + (1 - O_{ij}) \ln(1 - \hat{C}_{ij}))}{N_{pos}}, \quad (10)$$

$$\hat{C}_{ij} = \text{Sigmoid}(C_{ij}), \quad (11)$$

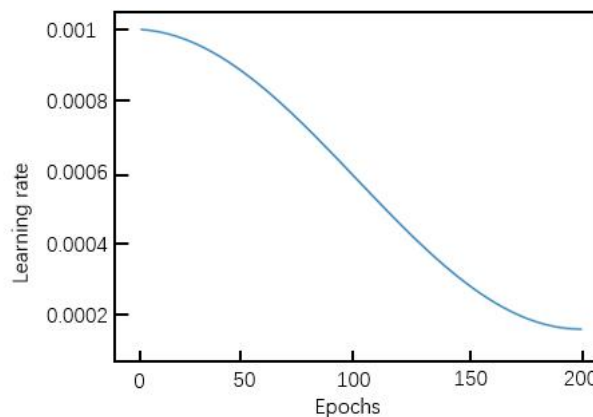
where  $o_i$  represents the *IoU* of the prediction box and ground truth box,  $c_i$  is the prediction value,  $\hat{c}_i$  is the prediction confidence obtained by the *sigmoid* function,  $N$  represents the total number of positive and negative samples, and  $O_{ij}$  indicates whether there is the  $j$ th target in the prediction box  $i$ .  $C_{ij}$  is the prediction value,  $\hat{C}_{ij}$  is the prediction confidence obtained by the *sigmoid* function,  $N_{pos}$  denotes the number of positive samples, and  $L_{GIoU}$  can be obtained by

$$L_{GIoU} = 1 - GIoU, \quad (12)$$

$$GIoU = IoU - \frac{A^c - u}{A^c}, \quad (13)$$

where  $u$  represents the area of the union of the ground truth box and prediction box,  $A^c$  represents the area of the smallest rectangle which can include both the ground truth box and prediction box.

We use Adam to update the trainable parameters in the network. The hyperparameter of Adam are set as:  $\epsilon = 10^{-8}$ ,  $\beta_2 = 0.999$ ,  $\beta_1 = 0.9$ ,  $\alpha = 0.001$ . The learning rate is set as 0.001 at the beginning, and then use cosine annealing algorithm [37] to gradually attenuate the learning rate to 0.0002, as shown in Figure 9.



**Figure 9.** Decay curve of learning rate.

Build the proposed FASC-Net by using the Pytorch1.6 framework and Anaconda3 and train it on NVIDIA GTX2070 GPU.

### 2.5. Dataset Description

Unlike many articles that only measure the detection accuracy of the network on the SSDD dataset, in order to fully assess the detection performance and generalization ability of FASC-Net, we will measure the performance indicators of the network on the following three datasets:

1. **SSDD:** SSDD [31] is made by referring to the production process of the PASCAL VOC dataset. This is because the PASCAL VOC dataset has more applications and the data format is more standardized. In the SSDD dataset, there are a total of 1160 images with  $480 \times 480$  average image size and 2456 ships from Radarsat-2, TerraSAR-X, and Sentinel-1. The SAR ship in the SSDD dataset has a resolution of 1–10 m, with HH, HV, VV, VH polarization, as shown in Figure 10. Select the image names with index suffixes 1 and 0 as the testing dataset, and the others as the training dataset. The ratio of the number of SAR images in the training dataset to the testing dataset is 8:2.
2. **SAR-Ship-Dataset:** This dataset took Sentinel-1 SAR data and Gaofen-3 SAR data as the dominant data sources. The ship slices in the dataset have diverse backgrounds and rich types, which are conducive to various SAR image applications. In the SAR-Ship-Dataset [32], there are 43,819 images with  $256 \times 256$  image size, and SAR ships have a resolution of 5 m~20 m with HH, HV, VV, and VH polarization. There is plenty of noise in these SAR images, which can effectively detect the ability of the proposed network to resist noise interference. Randomly divide the SAR images in the dataset into a training dataset and a testing dataset, the ratio of the number of SAR images in these two datasets is 8:2.
3. **HRSID:** HRSID [33] is a dataset for instance segmentation tasks, ship detection and, semantic segmentation in high-resolution SAR images. The HRSID dataset borrows from the construction process of the COCO dataset, including SAR images with different coastal ports, sea areas, sea conditions, polarization, and resolutions. This dataset contains a total of 5604 high-resolution SAR images with  $800 \times 800$  image size and 16,951 ship instances from TerraSAR-X and Sentinel-1. In the HRSID dataset, the SAR ship has a resolution of 0.1–3 m, with HH, HV, and VV polarization. These images have high resolution and complex coastal backgrounds which is able to accurately assess the ability of the proposed network to detect multi-scale targets under complex coastal backgrounds. There are 1961 and 3642 SAR images in the testing dataset and training dataset respectively. The ratio of the number of SAR images in these two datasets is 13:7.

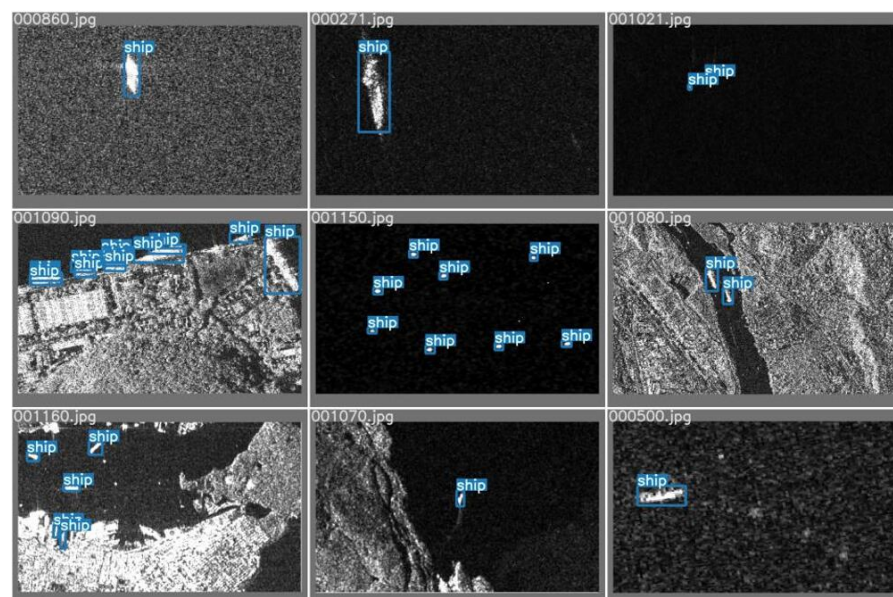
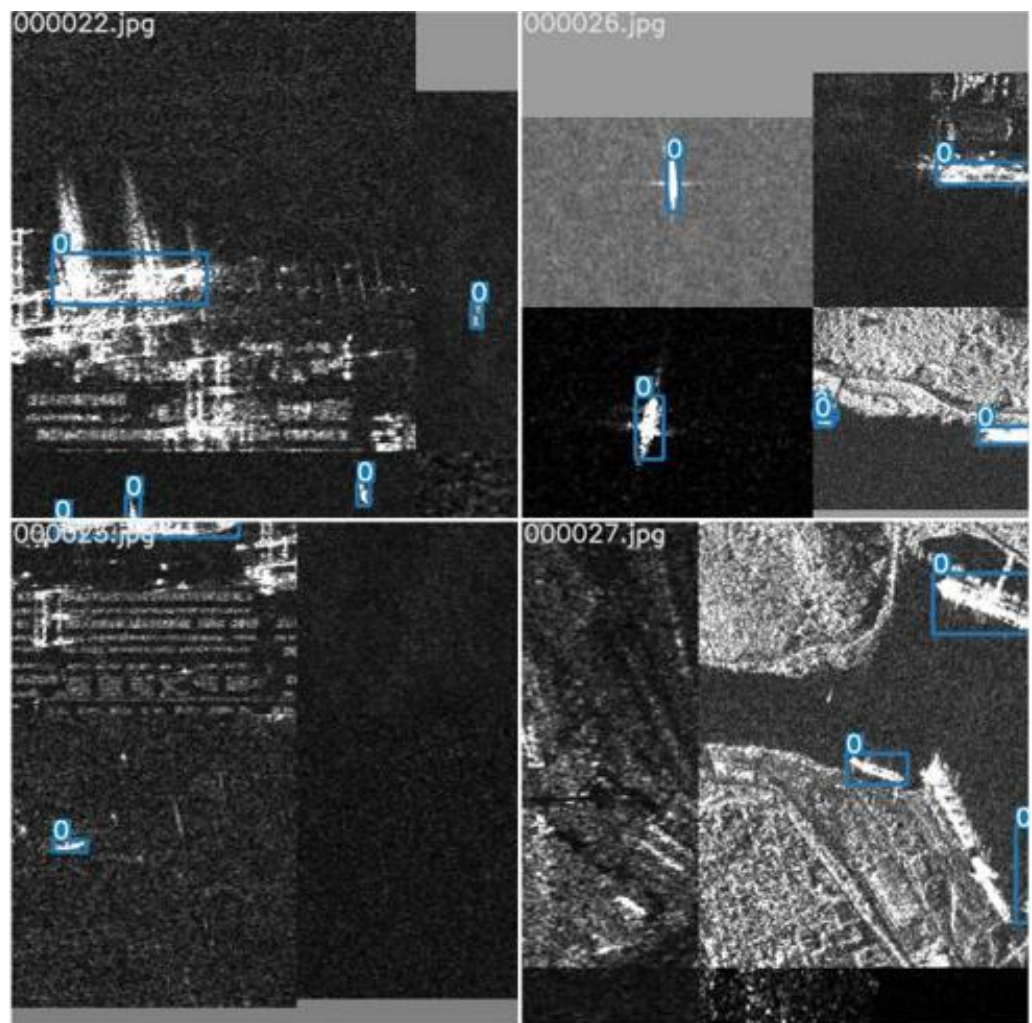


Figure 10. SAR images and labels in the SSDD dataset.

### 3. Experiments and Results

#### 3.1. Data Augmentation

Sufficient training data can greatly improve the training effect of deep learning networks. Owing to the unacceptable cost of acquiring sufficient SAR images, we adopt the Mosaic data augmentation method [30] in the training process to improve the training effect and the generalization ability of FASC-Net. Next, we will introduce the three steps of Mosaic data augmentation. (1) Read four images at a time; (2) flip, rotate, zoom and change the color gamut of the four images respectively, and place them according to the four directions; (3) combine pictures and labels. The effect of Mosaic data augmentation is shown in Figure 11. Mosaic data augmentation can greatly increase the object background, and calculate four images' data at a time when performing the batch normalization operation. In this way, the batch size does not need to be set too large, and a good training effect can be obtained with only one GPU.



**Figure 11.** The effect of Mosaic data augmentation.

#### 3.2. Evaluation Criteria

To comprehensively and quantitatively assess the detection performance, some evaluation standards of COCO dataset and PASCAL VOC dataset are adopted, including the frames per second (FPS), number of parameters, recall, precision, mean average precision (mAP), and F1 score. FPS and the number of parameters are used to measure the time cost

and memory cost of the network, respectively. Recall is the proportion of the target that the model predicts correctly among all real targets, which can be expressed as

$$Recall = \frac{TP}{TP + FN}, \quad (14)$$

where true positive ( $TP$ ) represents the number of correctly detected targets and false negative ( $FN$ ) is the number of undetected targets. Precision is the proportion of correctly predicted targets to all predicted targets. The higher the precision, the lower the false alarm rate. Precision can be expressed as

$$Precision = \frac{TP}{TP + FP}, \quad (15)$$

where false positive ( $FP$ ) represents the number of false alarms. Since precision and recall are related to each other,  $F1$  score and  $mAP$  are introduced to evaluate the overall performance of the proposed network.  $F1$  score can be calculated by

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}, \quad (16)$$

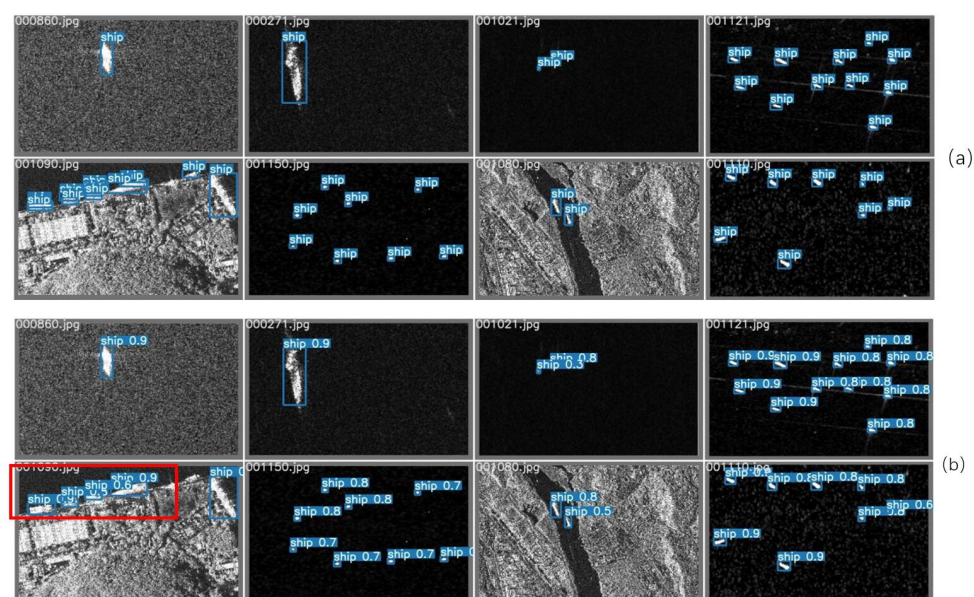
and  $mAP$  can be calculated by

$$mAP = \int_0^1 P(R)dR, \quad (17)$$

where  $R$  and  $P$  denote the single point value of recall and precision. FPS is used to measure the detection speed of the network.

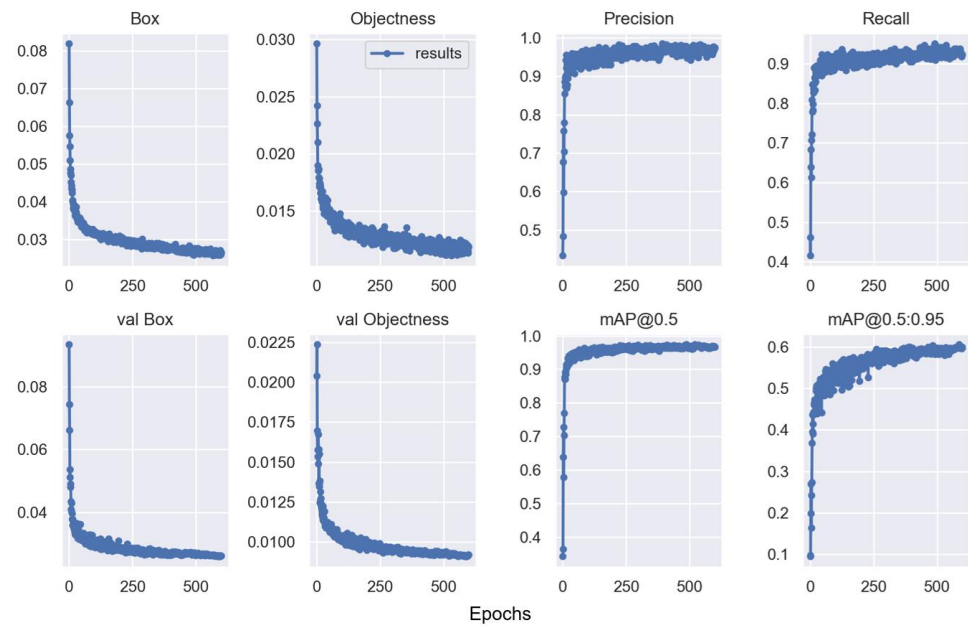
### 3.3. Detection Results on the SSDD Dataset

The visual analysis of the detection result of FASC-Net is shown in Figure 12. From Figure 12, we can see that FASC-Net has a very satisfactory detection performance on small and large targets in offshore scenes and the confidence is very high. However, some ships were not detected in inshore scenes. Because these ships are so close that a large part of the overlap occurred in the SAR image which may cause the network to detect these ships as one by mistake, as shown in the red rectangle in Figure 12.

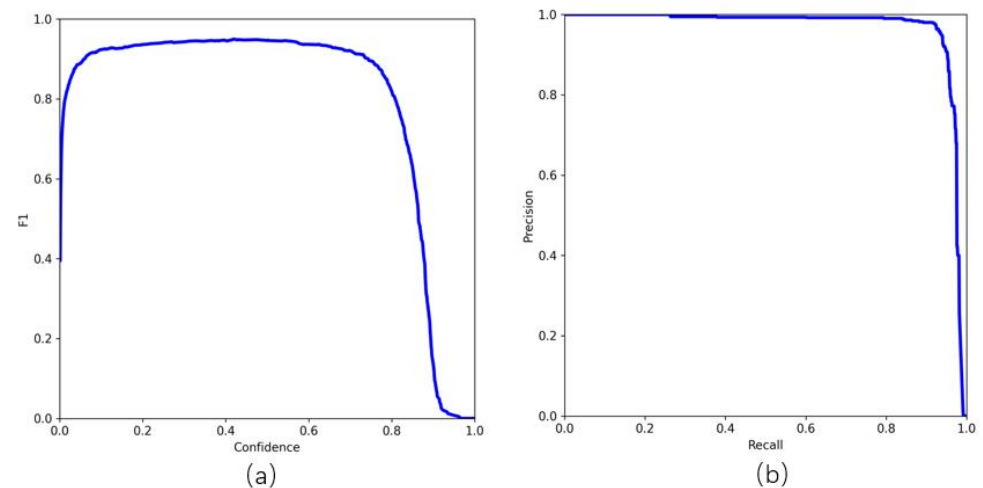


**Figure 12.** The visual analysis of the detection result of the proposed FASC-Net. (a) SAR images and labels. (b) The detection result.

The quantitative analysis of the detection result of FASC-Net is shown in Table 3, Figures 13 and 14. Figure 13 indicates the relationship between loss, precision, recall, mAP, and epoch. We can see that the proposed network has a fast convergence speed. Figure 14 shows the P–R curve and F1–Confidence curve. The area enclosed by the P–R curve and the coordinate axis is the value of mAP. From Table 3, we can see that the various evaluation indicators of FASC-Net on the SSDD dataset are very high, especially the mAP has reached an astonishing 97.4%.



**Figure 13.** The relationship between loss, precision, recall, mAP, and epoch.



**Figure 14.** The F1–Confidence curve and P–R curve. (a) F1–Confidence curve. (b) P–R curve.

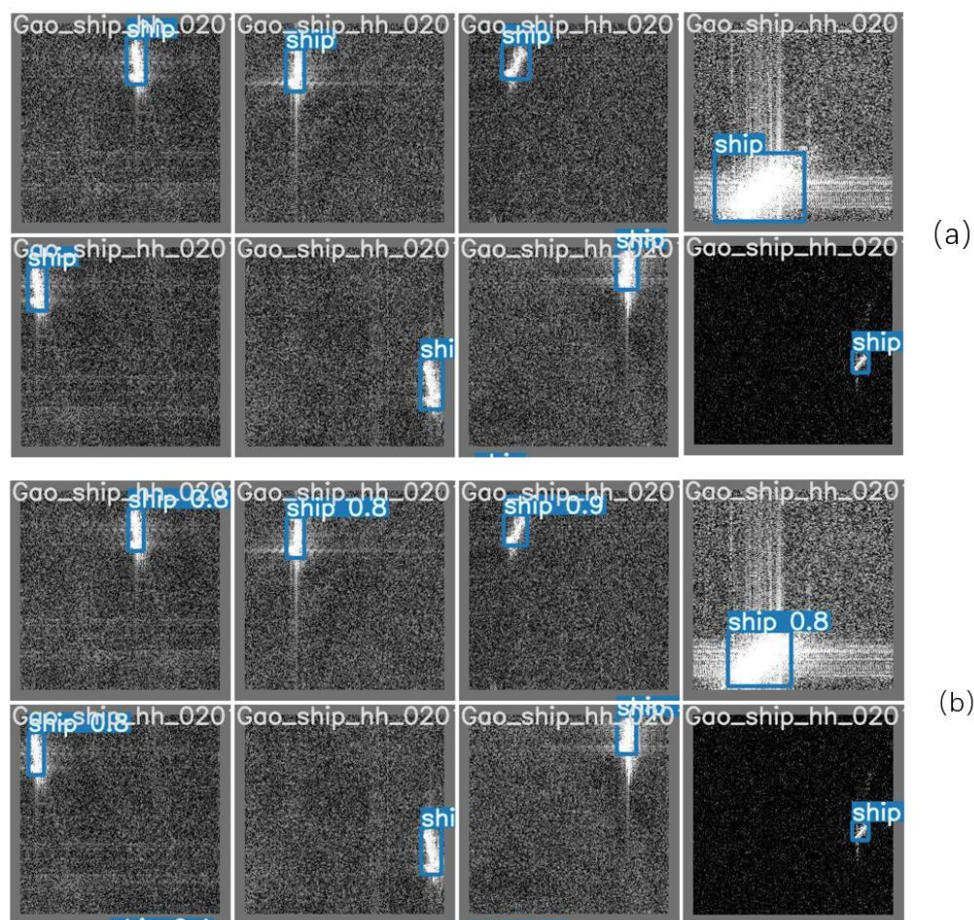
**Table 3.** Experimental results on the SSDD dataset.

Method	Dataset	Precision (%)	Recall (%)	F1 (%)	mAP (%)	FPS
FASC-Net	SSDD	95.6 ± 1.1	94.5 ± 0.4	94.9 ± 0.4	97.4 ± 0.3	42.5 ± 2.1

### 3.4. Detection Results on the SAR-Ship-Dataset

The visual analysis of the detection result of FASC-Net is shown in Figure 15. It is obvious that the SAR images in the SAR-Ship-Dataset have more noise than those in the

SSDD dataset. This increases the difficulty of ship detection, but our proposed FASC-Net can still detect ships of different sizes in this case.



**Figure 15.** The visual analysis of the detection result of the proposed FASC-Net. (a) SAR images and labels. (b) The detection result.

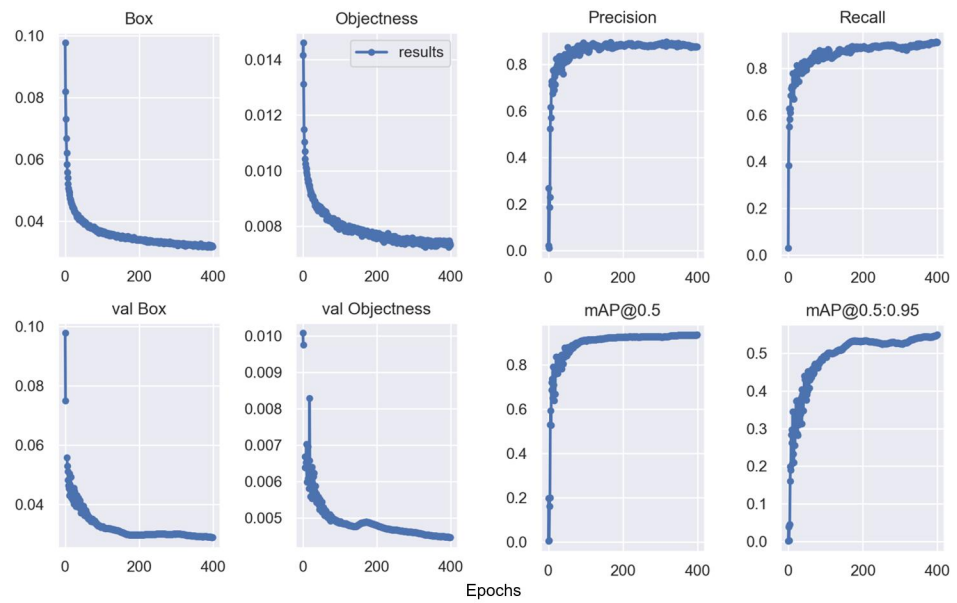
The quantitative analysis of the detection result of FASC-Net is shown in Figures 16 and 17, and Table 4, from which we can find that despite the noise and island interference, the proposed FASC-Net still achieves relatively high precision (91.1%), recall (92.2%), F1 score (92.1%), and mAP (96.1%) on the SAR-Ship-Dataset. FPS achieved an astonishing 60.4 owing to the lower image resolution in this dataset. This proves that the proposed network has a good anti-noise interference capability.

**Table 4.** Experimental results on the SAR-Ship-Dataset.

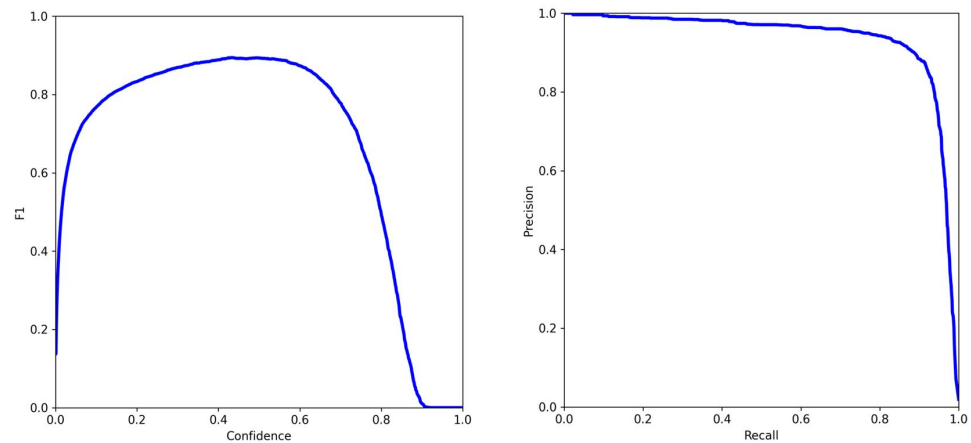
Method	Dataset	Precision(%)	Recall(%)	F1(%)	mAP(%)	FPS
FASC-Net	SAR-Ship-Dataset	91.1 ± 0.9	92.2 ± 0.6	92.1 ± 0.5	96.1 ± 0.4	60.4 ± 2.6

### 3.5. Detection Results on the HRSID Dataset

The visual analysis of the detection result of FASC-Net is shown in Figure 18. It is obvious that the SAR images in this dataset have more complex coastal backgrounds and higher resolution than those in the SSDD dataset and SAR-Ship-Dataset. This greatly increases the difficulty of detection, resulting in some false alarms in inshore scenes, as shown in the red rectangle in Figure 18.



**Figure 16.** The relationship between loss, precision, recall, mAP, and epoch.



**Figure 17.** The F1–Confidence curve and P–R curve. (a) F1–Confidence curve. (b) P–R curve.

The quantitative analysis of the detection result of FASC-Net is shown in Figures 19 and 20, and Table 5, from which we can find that despite the complex coastal background and island interference, the proposed FASC-Net still achieves relatively high precision (88.3%), recall (80.2%), F1 score (84.1%), mAP (88.3%), and FPS (24.5) on the HRSID dataset. This proves that the proposed network can detect ships of different scales under complex coastal backgrounds.

**Table 5.** Experimental results on the HRSID dataset.

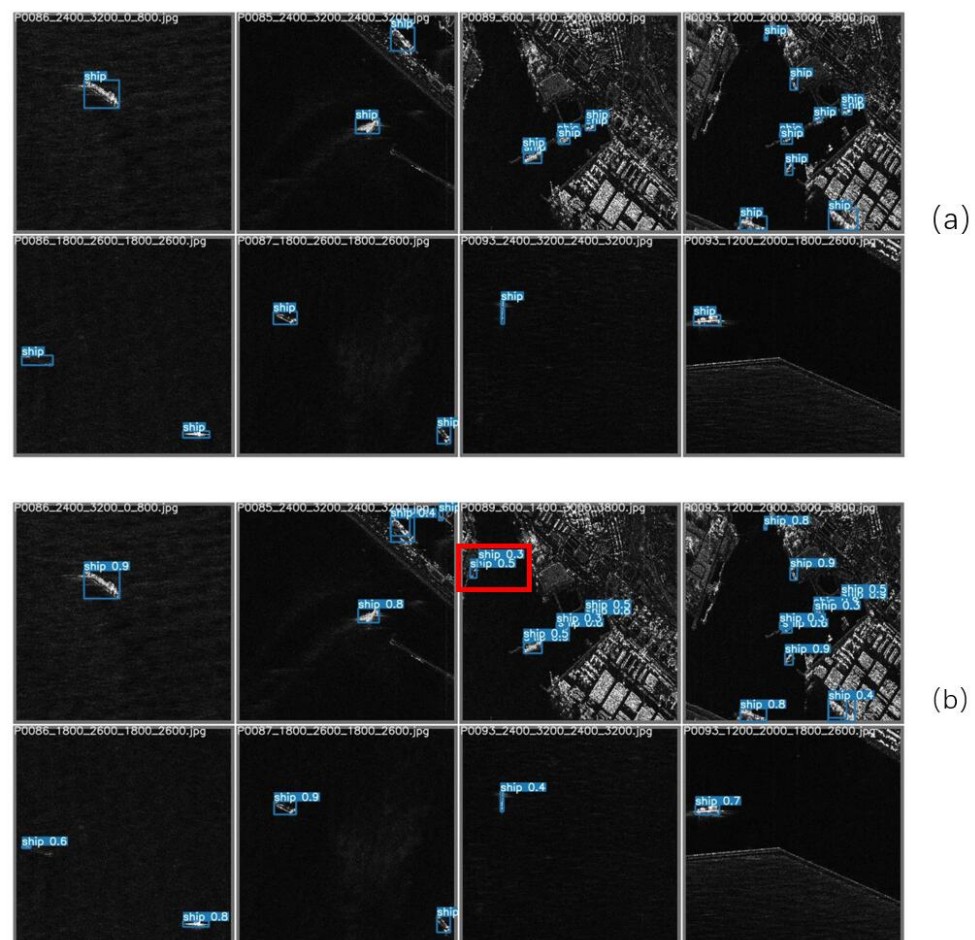
Method	Dataset	Precision(%)	Recall(%)	F1(%)	mAP(%)	FPS
FASC-Net	HRSID	88.3 ± 1.2	80.2 ± 0.5	84.1 ± 0.4	88.3 ± 0.5	24.5 ± 1.6

### 3.6. Methods Comparison

As shown in Table 6 and Figure 21, our FASC-Net is compared with six other SAR ship detection methods. These methods are Faster R-CNN [17], SSD [13], YOLO-V4 [30], DAPN [24], HR-SDNet [25], and Quad-FPN [26]. Figure 21 is the Pd–Pf curves of different methods on SSDD, where detection probability Pd and false alarm probability Pf is defined by [38]. In Table 6, Para represents the number of parameters and MS represents model size. In all comparative experiments on the three datasets, the mAPs and FPSs of the proposed



FASC-Net are the highest among those methods, and the F1 scores of FASC-Net are also ranked in the top three. The most important thing is that the proposed FASC-Net has only 0.6 million parameters, which is far lower than other methods. Compared with SSDD and SAR-Ship-Dataset, the performance of the proposed method degrades on the HRSID dataset. That is because compared with the SSDD dataset and the SAR-Ship dataset, the SAR images in the HRSID dataset have more complex coastal backgrounds and higher resolution which greatly increases the difficulty of detecting ships near the coast.



**Figure 18.** The visual analysis of the detection result of the proposed FASC-Net. (a) SAR images and labels. (b) The detection result.

Although the detection performance of the proposed network has declined on the HRSID dataset, it is still better than the other six methods. The detection performance of FASC-Net is superior to YOLO-V4 and HR-SDNet a little, yet the parameters of HR-SDNet are too many and the detection speed is too slow, it is difficult to apply in a wide range. The number of parameters of YOLO-V4 is much less than that of HR-SDNet, and the detection speed is much faster than it, but it still does not meet the requirements of transplanting to FPGA and DSP. Fortunately, the number of parameters of FASC-Net is only 1/18 that of YOLO-V4, and the detection speed is twice that of YOLO-V4, which enables the proposed FASC-Net to be well transplanted to FPGA and DSP. The results show that FASC-Net has a very low time cost and memory cost, which fully proves the rapidity and lightness of the proposed network.

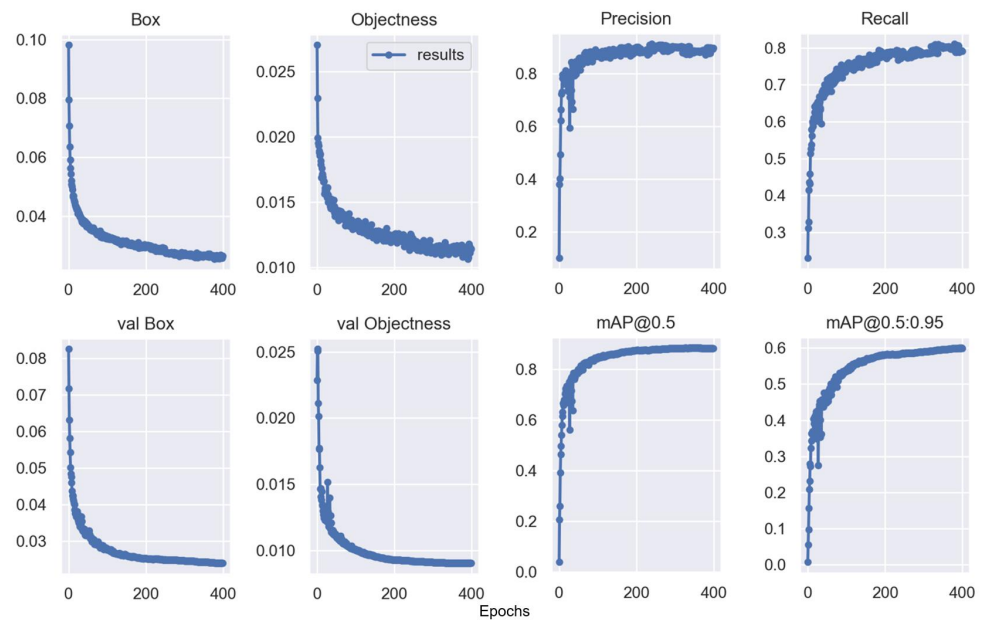


Figure 19. The relationship between loss, precision, recall, mAP, and epoch.

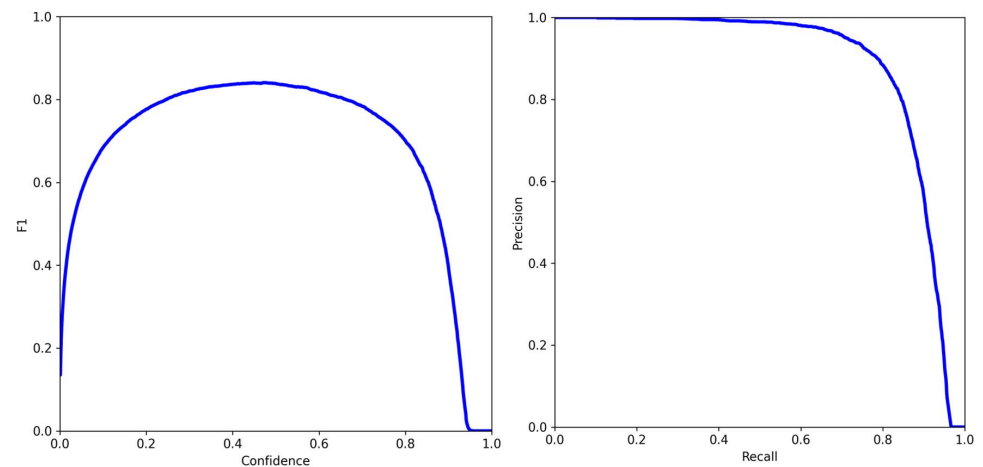


Figure 20. The F1–Confidence curve and P–R curve. (a) F1–Confidence curve. (b) P–R curve.

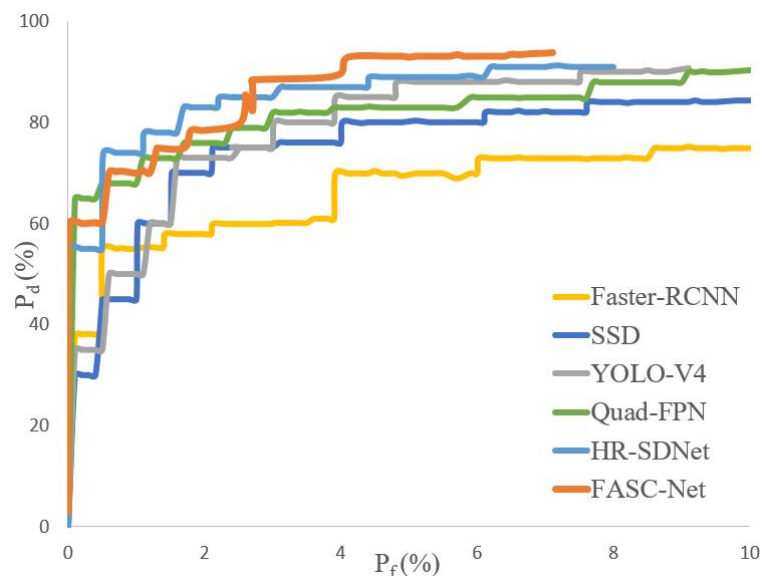
Table 6. The experimental results and number of parameters of different methods (the best is bold and the second-best is underlined). Para represents the number of parameters, and MS represents model size.

Methods	SSDD					SAR-Ship-Dataset					HRSID					Para (10 <sup>6</sup> )	MS (MB)
	P	R	F1	mAP	FPS	P	R	F1	mAP	FPS	P	R	F1	mAP	FPS		
Faster R-CNN	87.1	90.5	88.8	89.7	11.9	93.3	86.9	89.9	91.7	23.7	81.5	82	81.8	80.7	10.2	41.3	178.54
SSD	86.2	92.3	89.2	92.3	16.1	86.8	93.6	90.1	92.2	30.5	80.1	83	81.6	81.51	13.8	14.1	58.22
YOLO-V4	96.9	95.9	<b>96.4</b>	96.3	<u>22.7</u>	93.1	90.5	91.8	93.1	<u>32.6</u>	85.9	83.3	<u>84.6</u>	87.7	<u>14.5</u>	11.2	40.27
DAPN	85.6	91.4	88.4	90.6	<u>12.2</u>	87.3	93.4	90.3	91.9	21.5	83.4	80.5	<u>81.9</u>	81.9	13	22.5	90.73
HR-SDNet	95.1	93	94.1	96.8	8.3	93.3	92.1	<b>92.7</b>	92.3	8.9	86.9	81.2	83.9	<u>88.2</u>	6.8	74.8	265.44
Quad-FPN	89.5	95.8	92.6	95.3	11.4	77.6	96.1	85.9	<u>94.4</u>	23	88	87.3	<b>87.7</b>	86.1	13.4	15.2	60.52
FASC-Net	95.6	94.5	<u>94.9</u>	<b>97.4</b>	<b>42.5</b>	91.1	92.2	<u>92.1</u>	<b>96.1</b>	<b>60.4</b>	88.3	80.2	84.1	<b>88.3</b>	<b>24.5</b>	<b>0.64</b>	<b>1.47</b>

### 3.7. Generalization Ability Verification

In this section, we selected two large-scene Sentinel-1 SAR images from the LS-SSDD-V1.0 dataset [39] for actual ship detection, which verified the good generalization ability of the proposed FASP-Net. Table 7 shows the details of the two large-scene Sentinel-1 SAR

images. GT represents the total number of ships in the image. In follow experiments in this section, networks are not trained, but directly load the weights that were previously trained on other datasets. Then use those networks loaded with weights to detect ships in the new SAR image to measure the performance of the network in actual application scenarios. First, the size of the two large-scene SAR images was adjusted to  $24,000 \times 16,000$ . Then, due to the limited GPU memory, the large-scene SAR images were cut into  $800 \times 800$  sub-images. After that, those  $800 \times 800$  sub-images were input into FASP-Net trained on the HRSID dataset for the actual SAR ship detection. Finally, the sub-images output by the network are spliced into large-scene SAR images.



**Figure 21.** Pd–Pf curves of different methods on SSDD.

**Table 7.** Details of the two large-scene Sentinel-1 SAR images.

No.	Place	Time	Polarization	GT	Resolution	Image Size
Image 1	Tokyo Port	20 June 2020	VV	536	5 m × 20 m	25,479 × 16,709
Image 2	Singapore Strait	6 June 2020	VV	760	5 m × 20 m	25,650 × 16,768

Figures 22 and 23 show the visualized SAR ship detection results of FASC-Net on the two large-scene Sentinel-1 SAR images. It can be seen from Figures 22 and 23 that these two large-scene Sentinel-1 SAR images have very complex coastal backgrounds and a large number of multi-scale ships. The proposed FASC-Net can accurately detect most small ships, with very few missing detections and false alarms, as shown in the yellow boxes in the figures.

The detection performance of FASC-Net is compared with the six representative methods (CFAR, Faster R-CNN, SSD, YOLO-V4, HR-SDNet, and Quad-FPN), as shown in Table 8. Because there are a lot of large-scale backgrounds, small ships and pure backgrounds in the LS-SSDD-V1.0 dataset, the performance of all ship detection methods has declined on these two images. Nevertheless, our FASC-Net also achieves the highest mAP and F1. In Image 1, FASC-Net achieves an accuracy of 79.8% mAP and 78.5% F1, superior to the second-best Quad-FPN (79.8% mAP > 74.4% mAP, 78.5% F1 > 76.4% F1); in Image 2, FASC-Net achieves an accuracy of 78.6% mAP and 78% F1, superior to the second-best Quad-FPN (78.6% mAP > 74.9% mAP, 78% F1 > 76.2% F1). This once again proves that the proposed FASC-Net can accurately detect multi-scale ships under complex backgrounds and also proves the good generalization ability of FASC-Net.

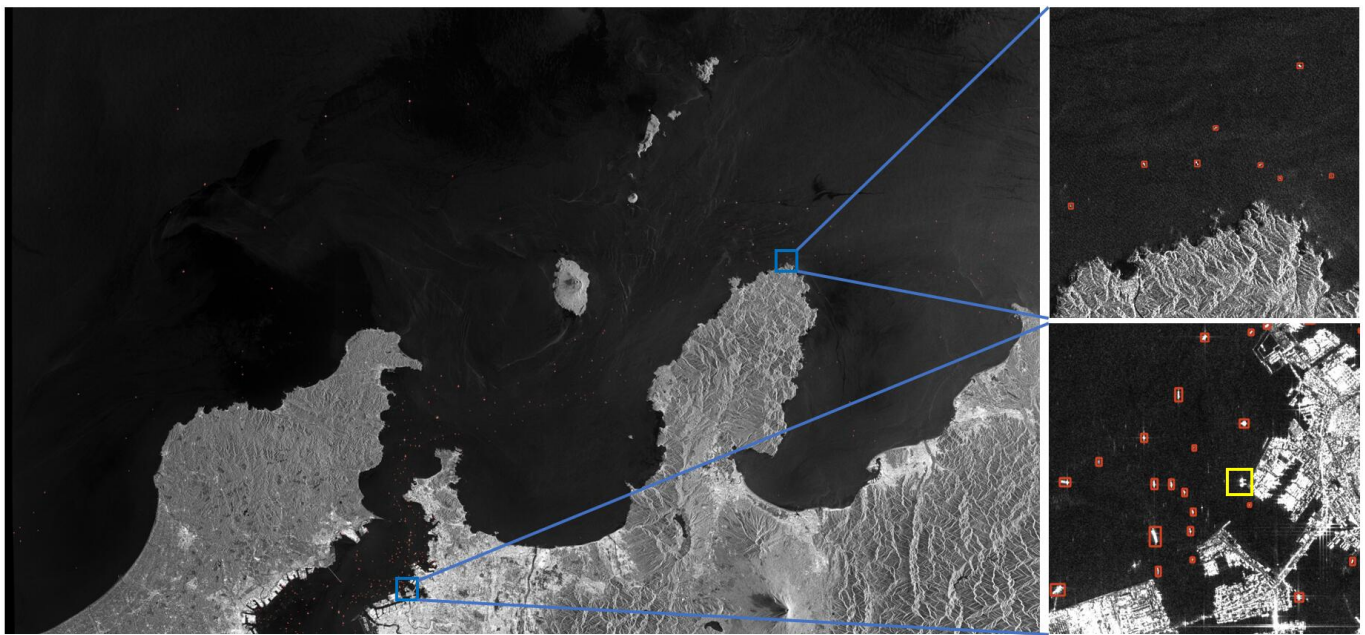


Figure 22. The visualized SAR ship detection result on Image 1.



Figure 23. The visualized SAR ship detection result on Image 2.

Table 8. The experimental results of different methods on the two large-scene SAR images.

Networks	Image 1				Image 2			
	P	R	F1	mAP	P	R	F1	mAP
CFAR	60.2	75.6	67	-	62.7	75	68.3	-
Faster R-CNN	77.7	73.7	75.6	73.7	74.8	77.2	76	74.3
SSD	67.5	62.1	64.7	61.4	70	61.7	65.6	64.8
YOLO-V4	76.9	74.1	75.5	73.9	80.9	72.3	76.4	75.3
HR-SDNet	84.9	68.8	76	68.8	85.8	70.1	77.1	70.5
Quad-FPN	78.5	74.4	76.4	74.4	79.5	73.2	76.2	74.9
FASC-Net	88.7	70.4	78.5	79.8	84.5	72.4	78	78.6

#### 4. Discussion

In this section, we will discuss the role of ASIR-Block (AB), Focus-Block (FB), SPP-Block (SB), and CAPE-Block (CB) through ablation studies. Referring to the ablation studies in literature [38], four variants of FASC-Net is designed as follows:

1. FPNet: Composed by traditional convolutional layers and an FPN-Block, FPNet has the same width and length network as FASC-Net. The traditional convolution layers are used to downsample and extract features. Same data augmentation methods are used in the training methods and loss function as FASC-Net.
2. A-FPNet: We get A-FPN by replacing all the traditional convolutional layers of FPNet with ASIR-Blocks. We can evaluate the effect of ASIR-Blocks by comparing the parameters and performance of FPNet and A-FPNet.
3. FA-FPNet: After replacing the first ASIR-Block-2 used for down-sampling in A-FPNet with a Focus-Block, we get FA-FPNet. The effect of Focus-Block can be testified by comparing the parameters and performances of FA-FPNet and A-FPNet.
4. FAS-FPNet: FAS-FPNet is obtained by adding an SPP-Block into FA-FPNet. The effect of SPP-Block can be testified by comparing the performances of FA-FPNet and FAS-FPNet and the effect of the CAPE-Block can be testified by comparing the performances of FAS-FPNet and FASC-Net.

The number of parameters, experimental results and detailed configurations of those variants are shown in Table 9.

**Table 9.** The number of parameters, experimental results, and detailed configurations of those variants.

Networks	Configurations				Results			Para (10 <sup>6</sup> )
	AB	FB	SB	CB	F1(%)	mAP(%)	FPS	
FPNet					88.1 ± 0.5	93.1 ± 0.3	20.2 ± 1.1	3.72
A-FPNet	✓				88.2 ± 0.6	92.6 ± 0.6	50.5 ± 1.6	0.27
FA-FPNet	✓	✓			89.3 ± 0.7	92.9 ± 0.6	52.1 ± 1.8	0.25
FAS-FPNet	✓	✓	✓		90.7 ± 0.6	94.8 ± 0.5	49.3 ± 1.9	0.27
FASC-Net	✓	✓	✓	✓	94.9 ± 0.4	97.4 ± 0.3	42.5 ± 2.1	0.64

**ASIR-Block:** The first and second rows of Table 9 indicate that the mAP of A-FPNet has dropped by 0.9% compared to FPNet, but the number of parameters of A-FPNet is only 1/14 of FPNet, and the FPS of A-FPNet is 2.5 times that of FPNet. This proves that ASIR-Block can extract some stable features of the target with few parameters for target detection and accelerate the detection speed of the network. That is because the Channel-Shuffle mechanism and depthwise convolution in ASIR-Block work together to extract more detailed features with fewer parameters.

**Focus-Block:** It is obvious from the second and third rows of Table 9 that compared with A-FPNet, the mAP and FPS of FA-FPNet increased by 0.4% and 1.6 respectively. It seems that the improvement effect is not very obvious, because the image resolution in the SSDD dataset is not high enough. Focus-Block can quickly down-sample high-resolution remote sensing images without losing information. Focus-Block does not show obvious advantages since the image resolution in the SSDD dataset is 480 × 480. When applied to higher resolution remote sensing images, Focus-Block can show more powerful performance.

**SPP-Block:** The third and fourth rows of Table 9 show that compared with FA-FPNet, the mAP of FAS-FPNet increased by 2.2%, and its FPS decreased by 2.8%, which proves that SPP-Block can extract the most important context features without affecting the detection speed. Because SPP-Block uses Maxpool operations of different kernel sizes for feature fusion to increase the receptive field of feature maps, this way of feature fusion has little effect on the running speed of the entire network, the performance is significantly improved owing that it can separate out the most significant context features.

CAPE-Block: Compared with FAS-FPNet, the mAP of FAS-FPNet increased by 2.7%, and its FPS decreased by 5.8 while adding some parameters. This shows that upgrading FPN-Block to CAPE-Block can indeed shorten the path of information transmission, and use the precise positioning information stored in low-level features to enhance the detection ability. Although the addition of a bottom-up enhancement path and channel attention mechanism slightly reduced the detection speed, the detection performance has been greatly improved.

## 5. Conclusions

To solve the problems of multi-scale ship feature differences, complex background interference and excessive network parameters in SAR ship detection, this paper proposes a fast and lightweight detection network for multi-scale SAR ship detection under complex backgrounds. Many experiments have been carried out on the SSDD dataset, SAR-Ship-Dataset, and HRSID dataset to evaluate the proposed FASC-Net. The following conclusions can be obtained:

1. Compared with the other six excellent methods, the fast and lightweight FASC-Net achieves higher detection performance on the three datasets and has fewer parameters.
2. The detection performance improves remarkably when upgrading FPN-Block to CAPE-Block, because it can shorten the path of information transmission and make full use of the precise positioning information stored in low-level features.
3. Using ASIR-Net as the backbone of the target detection network can significantly decrease the number of parameters while keeping the detection accuracy.
4. Compared with L2 loss, mean square error and Glou loss can better reflect the overlap of the two boxes and improve the training effect of the network.

Due to the shaking of ships during the voyage and the deviation of the flight trajectory of airborne SAR, significant movement errors appear in radar raw data. If a standard imaging algorithm is directly used to process the raw data, it will cause serious defocusing problems. In the future, we will try to add an autofocus module to the image preprocessing part of our network. The autofocus module combines standard imaging algorithms with effective motion compensation [40,41]. Doing so will further enhance the capabilities of our network in practical applications.

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

**Funding:** The Science and Technology Research Project of Higher Education of Hebei Province (Grant No. QN2019069), Chongqing Key Lab of Computer Network and Communication Technology (CY-CNCL-2017-02), and Guangxi Colleges and Universities Key Laboratory of Intelligent Processing of Computer Images and Graphics Project (No.GIIP1806).

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

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The python code of FASC-net is available online at [https://github.com/jack8zhou/FASC\\_Net](https://github.com/jack8zhou/FASC_Net) (accessed on 7 November 2021).

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

## References

1. Yang, M.; Guo, C.; Zhong, H.; Yin, H. A curvature-based saliency method for ship detection in SAR images. *IEEE Geosci. Remote Sens. Lett.* **2020**, *18*, 1590–1594. [[CrossRef](#)]
2. Lin, H.; Chen, H.; Jin, K.; Zeng, L.; Yang, J. Ship detection with superpixel-level Fisher vector in high-resolution SAR images. *IEEE Geosci. Remote Sens. Lett.* **2019**, *17*, 247–251. [[CrossRef](#)]

3. Chen, J.; Chen, Y.; Yang, J. Ship detection using polarization cross-entropy. *IEEE Geosci. Remote Sens. Lett.* **2009**, *6*, 723–727. [[CrossRef](#)]
4. Shirvany, R.; Chabert, M.; Tournet, J.Y. Ship and oil-spill detection using the degree of polarization in linear and hybrid/compact dual-pol SAR. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2012**, *5*, 885–892. [[CrossRef](#)]
5. Tello, M.; López-Martínez, C.; Mallorqui, J.J. A novel algorithm for ship detection in SAR imagery based on the wavelet transform. *IEEE Geosci. Remote Sens. Lett.* **2005**, *2*, 201–205. [[CrossRef](#)]
6. Tello, M.; Lopez-Martinez, C.; Mallorqui, J.; Bonastre, R. Automatic detection of spots and extraction of frontiers in SAR images by means of the wavelet transform: Application to ship and coastline detection. In Proceedings of the 2006 IEEE International Symposium on Geoscience and Remote Sensing, Denver, CO, USA, 31 July–4 August 2006; pp. 383–386.
7. Leng, X.; Ji, K.; Xing, X.; Zhou, S.; Zou, H. Area ratio invariant feature group for ship detection in SAR imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 2376–2388. [[CrossRef](#)]
8. 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](#)]
9. 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.
10. Park, K.; Park, J.J.; Jang, J.C.; Lee, J.H.; Oh, S.; Lee, M. Multi-spectral ship detection using optical, hyperspectral, and microwave SAR remote sensing data in coastal regions. *Sustainability* **2018**, *10*, 4064. [[CrossRef](#)]
11. Dai, J.; Li, Y.; He, K.; Sun, J. R-fcn: Object detection via region-based fully convolutional networks. *Adv. Neural Inf. Process. Syst.* **2016**, *30*, 379–387.
12. 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](#)]
13. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. Ssd: Single shot multibox detector. In *European Conference on Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 21–37.
14. 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.
15. 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.
16. Jiao, J.; Zhang, Y.; Sun, H.; Yang, X.; Gao, X.; Hong, W.; Fu, K.; Sun, X. A densely connected end-to-end neural network for multiscale and multiscale SAR ship detection. *IEEE Access* **2018**, *6*, 20881–20892. [[CrossRef](#)]
17. Li, J.; Qu, C.; Shao, J. A ship detection method based on cascade CNN in SAR images. *Control Decis.* **2019**, *34*, 2191–2197.
18. 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](#)]
19. Chang, Y.L.; Anagaw, A.; Chang, L.; Wang, Y.C.; Hsiao, C.Y.; Lee, W.H. Ship detection based on YOLOv2 for SAR imagery. *Remote Sens.* **2019**, *11*, 786. [[CrossRef](#)]
20. Lin, Z.; Ji, K.; Leng, X.; Kuang, G. Squeeze and excitation rank faster R-CNN for ship detection in SAR images. *IEEE Geosci. Remote Sens. Lett.* **2018**, *16*, 751–755. [[CrossRef](#)]
21. Wang, Y.; Wang, C.; Zhang, H.; Dong, Y.; Wei, S. Automatic ship detection based on RetinaNet using multi-resolution Gaofen-3 imagery. *Remote Sens.* **2019**, *11*, 531. [[CrossRef](#)]
22. 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](#)]
23. Zhang, X.; Wang, H.; Xu, C.; Lv, Y.; Fu, C.; Xiao, H.; He, Y. A lightweight feature optimizing network for ship detection in SAR image. *IEEE Access* **2019**, *7*, 141662–141678. [[CrossRef](#)]
24. 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](#)]
25. Wei, S.; Su, H.; Ming, J.; Wang, C.; Yan, M.; Kumar, D.; Shi, J.; Zhang, X. Precise and robust ship detection for high-resolution SAR imagery based on HR-SDNet. *Remote Sens.* **2020**, *12*, 167. [[CrossRef](#)]
26. Zhang, T.; Zhang, X.; Ke, X. Quad-FPN: A Novel Quad Feature Pyramid Network for SAR Ship Detection. *Remote Sens.* **2021**, *13*, 2771. [[CrossRef](#)]
27. Rostami, M.; Kolouri, S.; Eaton, E.; Kim, K. Deep transfer learning for few-shot SAR image classification. *Remote Sens.* **2019**, *11*, 1374. [[CrossRef](#)]
28. Zhang, X.; Huo, C.; Xu, N.; Jiang, H.; Cao, Y.; Ni, L.; Pan, C. Multitask Learning for Ship Detection from Synthetic Aperture Radar Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 8048–8062. [[CrossRef](#)]
29. Iandola, F.N.; Han, S.; Moskewicz, M.W.; Ashraf, K.; Dally, W.J.; Keutzer, K. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5 MB model size. *arXiv* **2016**, arXiv:1602.07360.
30. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. Yolov4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934.
31. 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.
32. Wang, Y.; Wang, C.; Zhang, H.; Dong, Y.; Wei, S. A SAR dataset of ship detection for deep learning under complex backgrounds. *Remote Sens.* **2019**, *11*, 765. [[CrossRef](#)]

33. Wei, S.; Zeng, X.; Qu, Q.; Wang, M.; Su, H.; Shi, J. HRSID: A high-resolution SAR images dataset for ship detection and instance segmentation. *IEEE Access* **2020**, *8*, 120234–120254. [[CrossRef](#)]
34. He, K.; Zhang, X.; Ren, S.; Sun, J. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 1904–1916. [[CrossRef](#)]
35. Lin, T.Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2117–2125.
36. Rezatofighi, H.; Tsoi, N.; Gwak, J.; Sadeghian, A.; Reid, I.; Savarese, S. Generalized intersection over union: A metric and a loss for bounding box regression. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 658–666.
37. Loshchilov, I.; Hutter, F. Sgdr: Stochastic gradient descent with warm restarts. *arXiv* **2016**, arXiv:1608.03983.
38. Zhang, T.; Zhang, X.; Shi, J.; Wei, S. HyperLi-Net: A hyper-light deep learning network for high-accurate and high-speed ship detection from synthetic aperture radar imagery. *ISPRS J. Photogramm. Remote Sens.* **2020**, *167*, 123–153. [[CrossRef](#)]
39. Zhang, T.; Zhang, X.; Ke, X.; Zhan, X.; Shi, J.; Wei, S.; Pan, D.; Li, J.; Su, H.; Zhou, Y.; et al. LS-SSDD-v1. 0: A deep learning dataset dedicated to small ship detection from large-scale Sentinel-1 SAR images. *Remote Sens.* **2020**, *12*, 2997. [[CrossRef](#)]
40. Chen, J.; Xing, M.; Yu, H.; Liang, B.; Peng, J.; Sun, G.C. Motion Compensation/Autofocus in Airborne Synthetic Aperture Radar: A Review. *IEEE Geosci. Remote Sens. Mag.* **2021**, *2*–23. [[CrossRef](#)]
41. Moreira, A.; Prats-Iraola, P.; Younis, M.; Krieger, G.; Hajnsek, I.; Papathanassiou, K.P. A tutorial on synthetic aperture radar. *IEEE Geosci. Remote Sens. Mag.* **2013**, *1*, 6–43. [[CrossRef](#)]