

Article

A-BFPN: An Attention-Guided Balanced Feature Pyramid Network for SAR Ship Detection

Xiuqin Li ¹, Dong Li ^{1,*}, Hongqing Liu ², Jun Wan ¹, Zhanye Chen ¹  and Qinghua Liu ³

¹ School of Microelectronics and Communication Engineering, Chongqing University, Chongqing 400044, China

² Chongqing Key Laboratory of Mobile Communications Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

³ Guangxi Key Laboratory of Wireless Wideband Communication and Signal Processing, Guilin 541004, China

* Correspondence: lid0705@cqu.edu.cn

Abstract: Thanks to the excellent feature representation capabilities of neural networks, target detection methods based on deep learning are now widely applied in synthetic aperture radar (SAR) ship detection. However, the multi-scale variation, small targets with complex background such as islands, sea clutter, and inland facilities in SAR images increase the difficulty for SAR ship detection. To increase the detection performance, in this paper, a novel deep learning network for SAR ship detection, termed as attention-guided balanced feature pyramid network (A-BFPN), is proposed to better exploit semantic and multilevel complementary features, which consists of the following two main steps. First, in order to reduce interferences from complex backgrounds, the enhanced refinement module (ERM) is developed to enable BFPN to learn the dependency features from the channel and space dimensions, respectively, which enhances the representation of ship objects. Second, the channel attention-guided fusion network (CAFN) model is designed to obtain optimized multi-scale features and reduce serious aliasing effects in hybrid feature maps. Finally, we illustrate the effectiveness of the proposed method, adopting the existing SAR Ship Detection Dataset (SSDD) and Large-Scale SAR Ship Detection Dataset-v1.0 (LS-SSDD-v1.0). Experimental results show that the proposed method is superior to the existing algorithms, especially for multi-scale small ship targets under complex background.



Citation: Li, X.; Li, D.; Liu, H.; Wan, J.; Chen, Z.; Liu, Q. A-BFPN: An Attention-Guided Balanced Feature Pyramid Network for SAR Ship Detection. *Remote Sens.* **2022**, *14*, 3829. <https://doi.org/10.3390/rs14153829>

Academic Editors: Alfonso Farina, Hing Cheung So and Shiyang Tang

Received: 6 July 2022

Accepted: 6 August 2022

Published: 8 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/>).

1. Introduction

Synthetic aperture radar (SAR) is a kind of active microwave imaging sensor with an all-day and all-weather capability to provide high-resolution images [1]. Due to the particularity of SAR images, artificial interpretation of SAR images is a time-consuming and labor-intensive process. SAR image target detection aims to automatically locate and identify specific targets from images, which has a significant application prospect in defense and civilian fields, such as target identification, object detection, ocean development and terrain classification [2–4], to name a few. As a basic marine task, SAR ship detection is of great value in maritime traffic control, maritime emergency rescue and intrusion target warning [5], which has received wide attention in recent years.

Traditional SAR ship detection methods can be divided into three types, including exploiting backscattering amplitude properties, polarization properties and geometric properties. In the first category, the constant false alarm rate (CFAR) algorithm has been widely utilized in SAR ship detections [6,7]. According to the background clutter statistical distribution model, the detection threshold is adjusted adaptively, and the most suitable threshold for ship target is obtained. However, this approach relies heavily on

human predefined distributions and is strongly influenced by background statistical regions. Therefore, under the interference of complex background, the traditional SAR image detection method is difficult to achieve accurate ship detection. In the second category, methods based on polarization properties distinguish the target and background by using the differences in scattering mechanism between ships and sea clutter [8,9]. However, it is a difficult task to build a comprehensive polarization scattering characteristic library. The last category is based on geometric properties by using the size, area, shape, and texture features [10,11], in which template matching strategy is usually adopted for ship detection. However, the object detection performance of these methods relies heavily on a library of templates previously built through expert experience. They require all SAR image pixels to be matched with the template, resulting in significant computational costs. However, the detection ability of these traditional algorithms for SAR ship targets of different scales is weak, and furthermore, it is difficult to deal with complex scenarios and small targets.

In recent years, with the rise of artificial intelligence and rapid advances in SAR imaging technologies, many researchers in the SAR field began to study the ship detection method based on deep learning (DL). At present, several SAR ship detection methods based on DL have been proposed, which are mainly divided into two groups including two-stage methods based on regional convolutional neural network (CNN) (RCNN) [12] and one-stage methods based on you only look once (YOLO) [13] and single shot detector (SSD) [14]. Compared with traditional methods, approaches based on DL have realized significant advantages of high efficiency and precision, because they enable computing models with multiple layers of processing to learn data representations with multiple layers of abstraction, which effectively improves detection accuracy. However, due to the unique imaging technology of SAR, there is a large amount of speckle noise, and ship detection in SAR images is susceptible to land facilities, islands, wave clutter, and changeable sea conditions. At the same time, due to the multi-resolution imaging mode, multi-scale, especially small targets are the characteristics of SAR images. When the small target is mapped to the final feature map, the missing rate is high due to the lack of localization refinement and classification information, which reduces the detection performance. Therefore, it is necessary to further develop the algorithm suitable for multi-scale variation, small targets with complex background.

Considering the above issues, a novel DL network model for SAR ship detection, called attention-guided balanced feature pyramid network (A-BFPN), is proposed to better exploit semantic and multilevel complementary features. First, we developed the enhanced refinement module (ERM) in the balanced feature pyramid network of the neck to emphasize or suppress visual information acquired through training and enhance feature representation. The enhanced feature map can be used to learn “which” feature map is more useful in channel dimension and “where” objects in spatial dimension may exist, to enhance the feature learning ability of the network to the region we are interested in and suppress its interference by complex background. In doing so, the ship objects can be accurately located and distinguished. Second, in order to reduce the aliasing effect of cross-scale feature fusion, this article proposes a channel attention-guided fusion network (CAFN) to optimize the aliasing feature of fusion and enhance its discrimination ability. Finally, based on the SAR Ship Detection Dataset (SSDD) [15] and Large-Scale SAR Ship Detection Dataset-v1.0 (LS-SSDD-v1.0) [16], the effectiveness and feasibility of the proposed method are verified. The experiments show that this method is superior to other detection algorithms based on CNN for SAR multi-scale ship targets.

With the exploration of DL by many researchers, CNN is widely used in natural scene object detection and has made remarkable progress. At present, object detection methods based on deep learning can be classified into two groups:

- (1) Two-stage detection methods. The representative two-stage methods include Faster-RCNN [15], feature pyramid network (FPN) [17], Mask-RCNN [18], region-based fully convolutional network (R-FCN) [19], and Cascade-RCNN [20]. The detection process includes two steps. First, a series of proposal boxes produced by the candidate

regions, then those proposal boxes are classified, and the exact position of the proposal boxes is further regressed. R-CNN [21] is the first one that applied DL to object detection. Later, inspired by SPP-Net [22], Fast-RCNN [23] proposed a pooling layer of the region of interest (RoI), which improved processing speed and detection accuracy. Subsequently, Faster-RCNN and improved Fast-RCNN use the region proposal network (RPN) instead of selective search to extract proposal. It is worth noting that the emergence of Faster-RCNN is an important milestone for two-stage detectors. It is composed of a backbone network, RPN and boundary box regression network. The introduction of RPN significantly improves detection accuracy. On the contrary, it also greatly increases the cost of testing time. Then Mask-RCNN came into being, where FCN is introduced to generate relevant mask branches, and the corresponding RoI align solution is proposed for pixel bias in RoI pooling. Based on Faster R-CNN, R-FCN greatly improves the detection speed through network shared computing. The cascade-RCNN alleviates the problem of quality mismatch in training overfitting and inference through multi-level architecture and has a good optimization effect on RPN candidate regions. In addition, in order to solve the problem of multi-scale variation in object detection, many researchers proposed different multi-scale feature extraction modules. As is known to all, FPN builds a rich multi-scale feature pyramid through single-resolution input images, where each layer of the pyramid is used to detect targets of different scales. It integrates multiple layers of feature information and has been widely adopted by subsequent algorithms. In [24], atrous spatial pyramid pooling (ASPP) captures object and image contexts on multiple scales by using atrous convolution at multiple sampling rates. In [25], AugFPN further tapped the potential of multi-scale features by integrating three simple and effective components: consistency supervision, residual feature enhancement and soft RoI selection. In [26], Libra R-CNN solves the problem of imbalance in training through balanced IoU sampling, balanced L1 loss, and balanced feature pyramid network.

- (2) One-stage detection methods. The representative one-stage methods include YOLO [13], SSD [14], RetinaNet [27], CornerNet [28], and FCOS [29]. Different from the two-stage detector, RPN is not required, and the one-stage detector classifies and regresses the target directly at each position of the feature map. The YOLO algorithm regards detection as a regression problem and directly uses CNN to realize the whole detection process. It divides the raw image into $s \times s$ grids, and the center of each object grid is responsible for predicting the location and category of target. However, one grid center predicts only one class of objects. The SSD detects objects of different sizes on multi-scale feature maps, where anchor module and multi-scale feature extraction layer are introduced to solve the shortcomings of YOLO rugged mesh and low detection accuracy of small objects. RetinaNet overcomes the imbalance of positive and negative samples in the detection by using focus loss. More recently, some anchor-free models have been proposed that do not require prior knowledge to design anchors, including the key point-based algorithms such as CornerNet, and anchor point-based algorithms such as FCOS. Compared with the two-stage detection methods based on R-CNN, the one-stage detection methods improves detection speed, but at the cost of accuracy.

The traditional SAR ship detection algorithms are listed as follows. He et al. [30] further proposed an automatic detection method for polarization SAR ships. Kapur et al. [31] used Shannon entropy concept for image segmentation, which overcomes the problem of detecting unconnected ships with single parameter threshold segmentation algorithm. Several two-parameter CFAR algorithms use Gaussian model to build the clutter model of ocean background [32], which in many cases does not describe the clutter well. In addition, Shi et al. [33] separated the ship target from the background by extracting target features through directional gradient histogram.

However, the performance of these approaches is much lower than algorithms based on DL. Fan et al. [34] applied the full convolutional network to ship detection of polarimetric SAR images. Kang et al. [35] combined the contextual features of high-resolution

RPN and ships to improve the positioning precision of small SAR ships. Based on Libra R-CNN, Guo et al. [36] proposed a method that uses rotation Angle information balance for ship position prediction with three levels of neural network, which refer to the sample level, feature level and objective level. In addition, in our previous work [37], a new multidimensional DL network model was proposed to improve detection performance by utilizing complementary properties of spatial and frequency domains. In recent studies, Yu et al. [38] proposed CR2A-NET to solve the problem of intensive ships in ship detection tasks, which achieve high-precision detection of ships at any orientations through three parts of rotating anchor assisted detection module, rotating alignment convolution layer, and data preprocessing module. Lin et al. [39] proposed the squeeze and excitation rank (SER) Faster R-CNN, which suppressed redundant sub-feature maps through SE mechanism and rank modification to improve detection performance. In order to suppress the false alarms and capture ship target features, attention mechanism is used in some studies as well [40–43]. Cui et al. [44] proposed a new multi-scale ship detection method for SAR images based on dense attention pyramid network (DAPN), which connects feature maps from the top of the pyramid to the bottom to detect multi-scale SAR ships. Fu et al. [45] introduced a feature balancing and refinement network (FBR-Net) to balance semantically the multiple features across different levels. However, these methods do not adequately consider the characteristics of SAR images in the ship detection tasks. Therefore, this article proposes an attention-guided balanced feature pyramid network (A-BFPN).

2. Methods

In this section, the A-BFPN network for SAR ships detection is proposed, and the detailed implementation procedures are also presented. First, the motivation and overall architecture of the proposed A-BFPN are illustrated via analyzing the shortcomings of existing BFPN network model. After this, each module of A-BFPN is introduced in detail. At the end of this section, the loss function used in the training process is given.

2.1. Problem Formulation and Method Overview

It is important in SAR ship detection to solve problems such as the large proportion of small object features, the imbalance of multi-scale features and the interference of complex background. As shown in Figure 1, ships in SAR images are small in volume. Since convolutional neural network is a multi-layer structure composed of several convolutional layers, the feature information of small targets will become less rich with the increase of network depth. Therefore, the performance of the detection network for small objects is poor. In addition, due to the different resolutions and incident angles, SAR ships always present multi-scale characteristics. However, some simple feature pyramid networks pay more attention to the features of adjacent layers than to other layers, and the semantic information contained in adjacent layers is diluted. Therefore, the multi-scale features of ship targets are not fully utilized. Finally, the wide range of complex backgrounds in SAR images can lead to some false positives, which contains background clutter and inland facilities in the inshore. All these problems will lead to the limited performance of SAR ship detection.

To solve the above problems, we adopt Resnet-101 as the backbone to construct the neural network. A CNN network can extract the features of labeled data independently, which effectively avoids the traditional complex feature design and has the advantages of fast and efficient. Different from the BP neural network, the CNN network uses parameter sharing mechanism, which can achieve the purpose of DL to train the deeper network. In this work, we use deep residual network ResNet101 to extract SAR image features to simplify the training process. In addition, the residual structure in the network is called a bottleneck, as shown in Figure 2. The structure of ResNet101 extractor consists of five convolutional stages, represented by {conv1, conv2, conv3, conv4, conv5}, each of which contains several traditional convolutional layers or bottlenecks and outputs a feature of a different scale from the other stages.

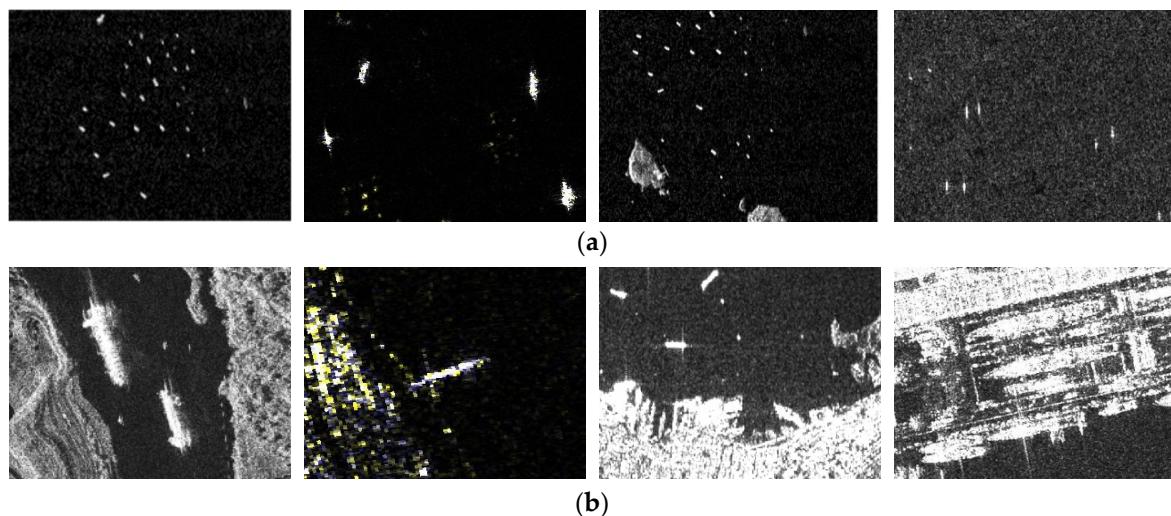


Figure 1. Examples of SAR images from the SSDD dataset. (a) Large proportions of small ships. (b) Multi-scale ships with complex background.

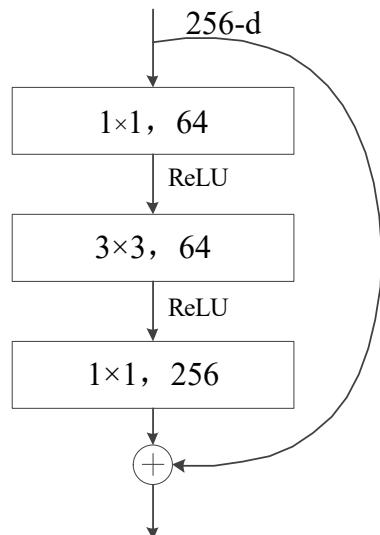


Figure 2. The architecture of a residual bottleneck.

With the increase of convolutional step size and residual network hierarchy, the spatial resolution of feature maps decreases gradually, while the number of channels increases gradually. The low-level features contain accurate location information, but lack semantic information, which is suitable for small-scale ship detection. In contrast, the high-level feature contains abundant semantic information but poor location information, indicating that it is suitable for detecting large ships. Therefore, FPN uses the information of different layers in a CNN network to obtain the features of the final combination. As presented in Figure 3, the structure of FPN can be divided into bottom-up path, top-down path, and horizontal connection. In addition, several typical FPN and its extended versions were designed to take advantage of multi-scale features, and they augment the features of the pyramid with bottom-up or top-down information paths. Unlike other FPNs that directly add and fuse features of different resolutions, BiFPN [46] improves detection accuracy and efficiency by repeatedly utilizing each feature network layer composed of top-down and bottom-up paths. However, they pay more attention to the features of adjacent resolutions than to other resolutions in the integration process, and the semantic information of adjacent layers will be diluted. Therefore, Pang et al. proposed the balanced feature pyramid network (BFPN) to efficiently utilize multi-scale features, as shown in Figure 4.

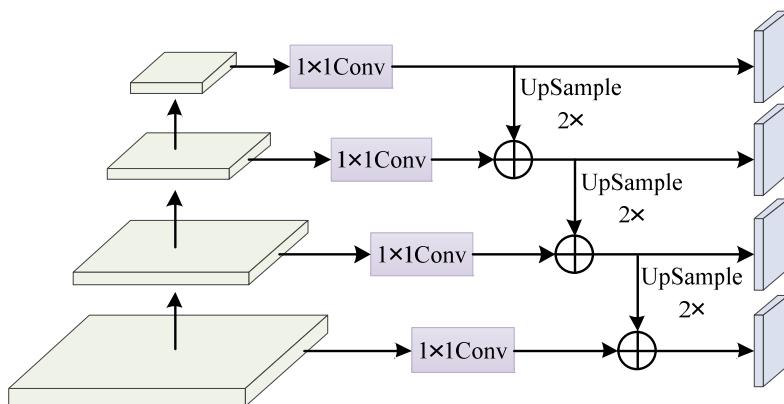


Figure 3. The structure of the feature pyramid network (FPN).

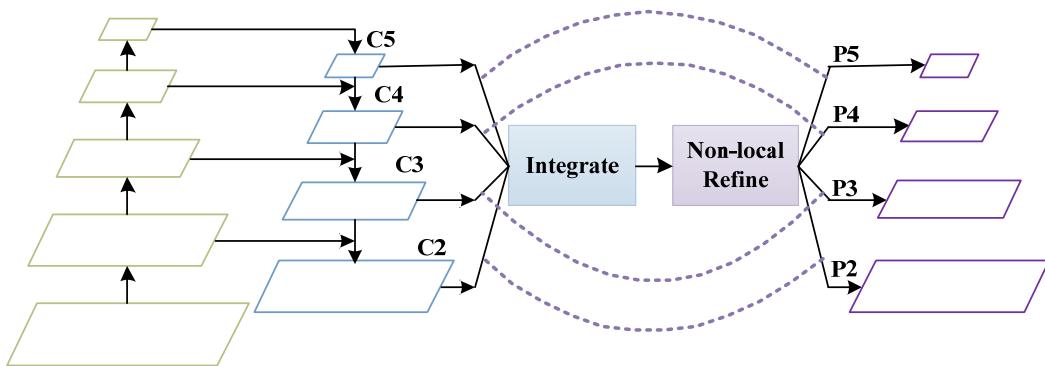


Figure 4. The structure of the balanced feature pyramid network (BFPN).

Based on FPN (Faster -RCNN), BFPN conducts rescale, integrate, refine and fuse on the features of the four levels. First, the hierarchical features extracted from FPN are adjusted to the same size as C4 using interpolation and maximum pooling, respectively. After that, the balanced semantic features can be obtained by integrating the features of all levels through mean calculation, given by

$$M = \frac{1}{L} \sum_{l=l_{\min}}^{l=l_{\max}} C_l \quad (1)$$

where the feature with resolution level l is represented as C_l . The number of multi-level features is represented as L , and the lowest and highest-grade indexes involved are represented as l_{\min} and l_{\max} , respectively.

Finally, the balanced semantic features are further refined by non-local attention, and the refined features are fused with the original features of each layer of C by direct addition.

For the complex background and the unsharp features of small objects in SAR ship detection, it is of vital importance that the detection network can enhance or suppress different channels for different tasks. However, BFPN only considers the relationship between features from the spatial dimension, and ignores the difference in each feature channel. Therefore, to overcome this issue, we develop the enhanced refinement module (ERM) to improve feature representation capabilities and make BFPN more discriminating. In addition, BFPN fuses semantic features and original features through simple superposition, resulting in serious aliasing effects and performance degradation of SAR ship detection. To mitigate the negative effects of aliasing, a straightforward solution is to exploit an attention module on a feature pyramid. We expect that attention mechanisms at different layers can learn from information at other layers. Therefore, we propose a channel attention-guided fusion network (CAFN) inspired by CBAM [40], which guides each layer of the pyramid to mitigate aliasing effects. The proposed A-BFPN contains four main components. The

backbone uses the ResNet-101, the neck includes enhanced refinement module (ERM) and channel attention-guided fusion network (CAFN), the region proposal network (RPN) and the head include classification and regression of bounding boxes, illustrated in Figure 5. Details of the algorithm are described in the following.

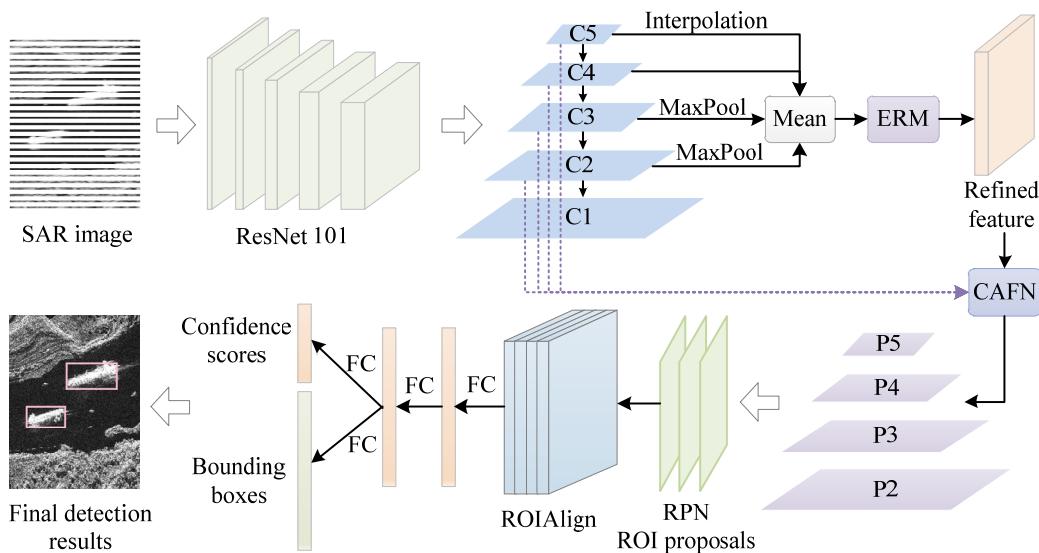


Figure 5. The framework of the proposed attention-guided balanced feature pyramid network (A-BFPN).

2.2. Enhanced Refinement Module

For the SAR ship detection task, because ship objects and complex background clutter share similar scattering characteristics, it is likely to misunderstand this background interference as ship target. In order to solve this problem, shown in Figure 6, we developed a new enhanced refinement module (ERM) to learn the dependencies between different features from the channel and spatial direction, respectively, to generate more distinctiveness features about ship objects and its background. By doing so, we enhance the expressing ability of ship objects in the features, especially for small ship targets and also restrain the false detection caused by complex background clutter.

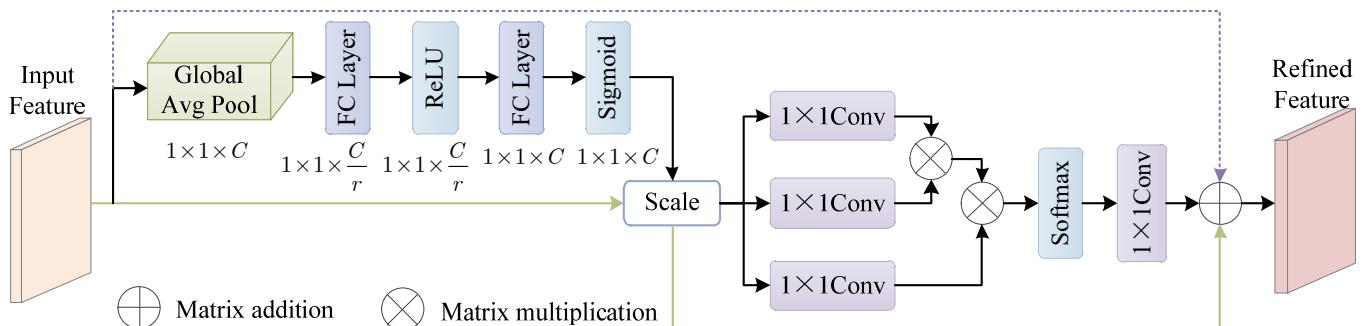


Figure 6. The enhanced refinement module (ERM) architecture in the proposed A-BFPN.

We consider using global information to guide the network to selectively enhance features that contain useful information and suppress those that are useless. First, the global average pooling is calculated on the balanced semantic feature map to collect the global feature at channel level. Then, we perform the fully connected layer ReLU on global features to obtain the weights for different channels. Next, a sigmoid activation function is

used. In addition, the channels of the original feature map are scored at last. The above steps are calculated as follows:

$$g_{(1 \times 1 \times C)} = \text{GAP}(m_{(H \times W \times C)}) \quad (2)$$

$$g_{(1 \times 1 \times \frac{C}{r})} = \text{FC}(g_{(1 \times 1 \times C)}) \quad (3)$$

$$y_{(1 \times 1 \times C)} = \text{FC}'(\text{ReLU}(g_{(1 \times 1 \times \frac{C}{r})})) \quad (4)$$

$$\varphi_{(1 \times 1 \times C)} = \text{Sigmoid}(y_{(1 \times 1 \times C)}) \quad (5)$$

where m represents the balanced feature with $H \times W \times C$ dimensions obtained in Equation (1), $\text{GAP}(\cdot)$ represents a global average pooling function that transforms the dimension to $1 \times 1 \times C$, FC represents a fully connected layer that reduces channel dimension to C/r , the tensor $g_{(1 \times 1 \times \frac{C}{r})}$ passes through the ReLU function, and then the output tensor $y_{(1 \times 1 \times C)}$ with channel size restored to C is obtained through another fully connected layer FC' . As a result, we obtain the excitation value $\varphi_{(1 \times 1 \times C)}$ of the channel passes through the sigmoid function of the balanced feature $m_{(H \times W \times C)}$. Finally, the excitation tensor $\varphi_{(1 \times 1 \times C)}$ is broadcast and multiplied by the balanced feature $m_{(H \times W \times C)}$ to adjust its channel:

$$E_{(H \times W \times C)} = m_{(H \times W \times C)} \times \varphi_{(1 \times 1 \times C)} \quad (6)$$

In order to quickly capture the dependence between long-distance features at different positions in the feature maps, E is used as the input of non-local attention, which has higher computational efficiency and fewer stacking layers. The calculation is

$$R = h_{1 \times 1} \{ \text{soft max}[h_{1 \times 1}(E) \otimes h_{1 \times 1}(E)] \otimes h_{1 \times 1}(E) \} \oplus E \oplus m_{(H \times W \times C)} \quad (7)$$

where E represents balanced semantic features, R represents refined balanced semantic features, $h_{1 \times 1}(\cdot)$ represents the convolution operation with a convolution kernel size of 1×1 , \otimes and \oplus represent matrix multiplication and addition operations, respectively.

The final refinement integration feature is obtained by R . The differences between integration features are increased by refining the balanced semantic features. It is of vital importance to distinguish targets and background in ships detection.

2.3. Channel Attention-Guided Fusion Network

The cross-scale fusion and the skip connection are widely used to improve the performance of object detection, and the intuitive connectivity enables a full use of functionality at each level. However, there are semantic differences at all levels of feature maps. In this case, direct fusion after interpolation leads to aliasing effect, and repeated feature fusion will not only lead to more serious aliasing effect, but also bring unnecessary computational burden. Inspired by CBAM, attention mechanisms can be used to optimize the aliasing feature of fusion. Therefore, we developed a channel attention-guided fusion network (CAFN), which adaptively integrates feature information of different channels according to channel weights, illustrated in Figure 7.

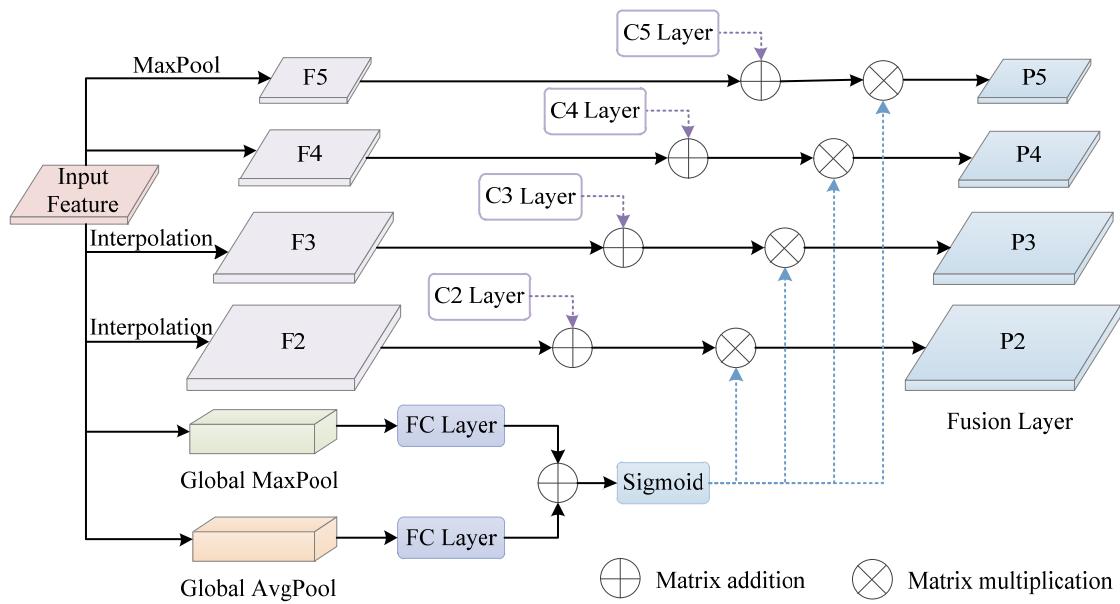


Figure 7. The channel attention-guided fusion network (CAFN) architecture in the proposed A-BFPN.

On one hand, we first use the same interpolation and maximum pooling methods for the integrated features R to recover the original feature size, and then add and fuse the original feature C respectively to obtain the preliminary fusion feature layers F' , given by

$$F' = C \oplus F \quad (8)$$

On the other hand, considering the high computational cost of performing independent attention learning at each level, and the expectation that attention mechanisms at each level can learn from information at other levels, we only extract channel weights by integrating feature map R :

$$F'' = \sigma(\text{FC}(\text{MaxPool}(R)) + \text{FC}(\text{AvgPool}(R))) \quad (9)$$

where the global maximum pool $\text{MaxPool}(\cdot)$ and global average pool $\text{AvgPool}(\cdot)$ are used to obtain two different spatial context information, respectively. Later, two descriptors are sent separately to the fully connected layer $\text{FC}(\cdot)$. As a result, the eigenvectors F'' are combined by element-by-element addition and sigmoid function σ .

The channel weight F'' is multiplied by each preliminary fusion feature layer F' to obtain the final fusion feature layer P , given by

$$P = F'' \otimes F' \quad (10)$$

In this process, the aliasing effect of feature fusion is reduced, and the levels of feature pyramid are optimized. Finally, the output features of CAFN are sent to RPN for subsequent objects detection.

2.4. Loss Function

Similar to other classical two-stage object detectors, both RPN and detection networks are optimized by using multitask loss, which is given by

$$L\{p_i, t_i\} = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(t_i, t_i^*) \quad (11)$$

where N_{cls} and N_{reg} are the numbers of minibatch samples in the training stage, p_i refer to the prediction probability of the i -th anchor, p_i^* refer to the corresponding ground truth label and it is 1 when the anchor is positive, or else p_i^* is 0, λ represents the weighting parameter.

t_i refer to the parameterized coordinate vectors, which is defined as $t_i = \{t_x, t_y, t_w, t_h\}$, where t_x, t_y are the coordinates of the center point of the prediction box and t_w, t_h are the width and height of the prediction box, and t_i^* refer to the coordinate vector of the corresponding ground truth, L_{cls} and L_{reg} refer to classification and regression losses, respectively. Cross Entropy (CE) is utilized as the classification loss, which is given by

$$L_{cls}(p_i, p_i^*) = -\log(p_i^* p_i + (1 - p_i^*)(1 - p_i)) \quad (12)$$

The regression network loss is given by

$$L_{reg}(t_i, t_i^*) = S(t_i - t_i^*) \quad (13)$$

where S represents the Smooth L1 loss, which is defined by

$$\text{Smooth}_{L_1(x)} = \begin{cases} 0.5x^2, & |x| \leq 1 \\ |x| - 0.5, & \text{otherwise.} \end{cases} \quad (14)$$

3. Results

3.1. Dataset Description and Settings

The ship target images in SSDD have multiple polarization modes, multiple resolutions, and rich scenes. The SSDD data were mainly obtained from TerraSAR-X, Sentinel-1, and Radarsat-2 sensors, including VH, HV, VV, and HH polarization modes. In addition, the resolution of SAR images ranges from 1 m to 15 m. The detailed summary of the data information is described in Table 1. SSDD dataset contains 1160 SAR images of 2358 ships, each containing an average of 2.12 ship targets. Figure 8a,b respectively represent the distribution of the height-width and height-width ratio distribution of each ship object.

Table 1. Information regarding the SSDD dataset.

Sensors	Polarization	Resolution	Position
TerraSAR-X			
Sentinel-1	VH, HV VV, HH	1–15 m	offshore inshore
RadarSat-2			

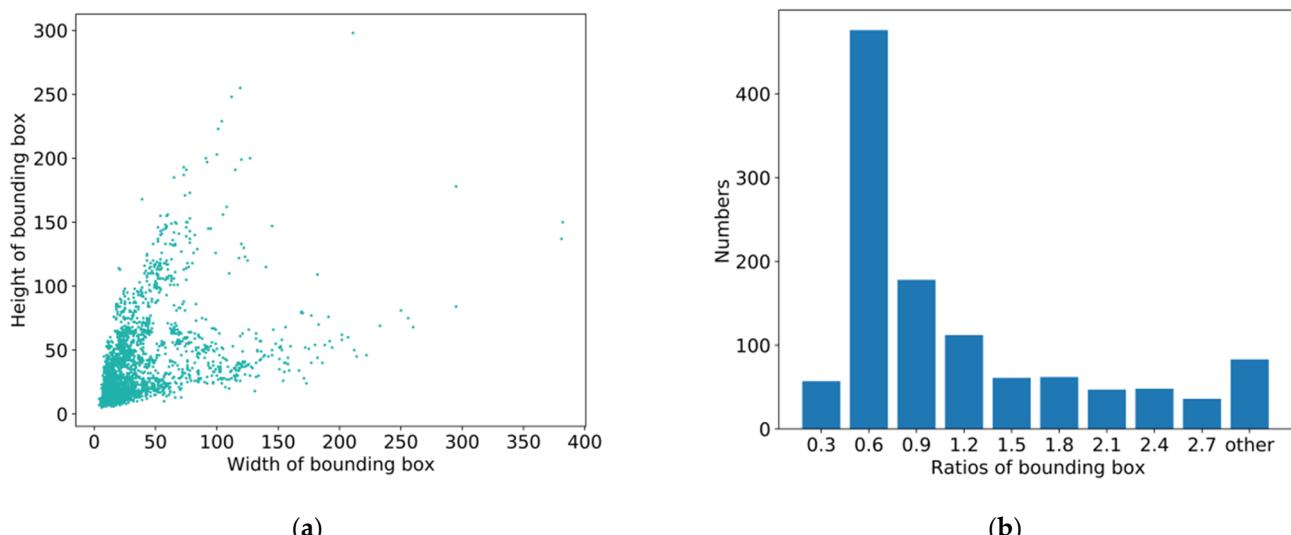


Figure 8. Distribution of the SSDD dataset. (a) Height-width distribution of bounding box. (b) Aspect ratio distribution of bounding box.

The LS-SSDD-V1.0 dataset consists of Sentinel-1 images in the interferometric wide swath mode and contains 15 large-scale SAR images with $24,000 \times 16,000$ pixels. The publisher of the dataset cut these 15 large-scale images into 9000 sub-images with 800×800 pixel. SAR ships in LS-SSDD-V1.0 are provided with various resolutions around 5 m, and VV and VH polarizations. According to the setting of the original reports in [16], there are 6000 SAR images in the training set and 3000 SAR images in the test set.

In this work, all experiments were carried out under PyTorch1.1.0 framework, and network training was carried out on computers using NVIDIA GTX1660s GPU with Ubuntu16.04 and Cuda9 and Cudn7. In order to standardize the research benchmark of SSDD, serial number SAR images ending in 0 and 9 in SSDD are unified as the test set of result analysis, and the number of samples of training set and test set are 928 and 232 respectively. The learning rate was set as 0.02, and the maximum number of iterations was 12 epochs on SSDD. Moreover, the learning rate was set as 0.002, and the maximum number of iterations was 32 epochs on LS-SSDD-V1.0. In addition, ResNet101 pre-trained on ImageNet was adopted as the initialization model.

3.2. Evaluation Criteria

Several evaluation criteria are used to quantitatively evaluate and compare the detection performance of different ship detection methods on SSDD, including the precision rate, recall rate, F1-score, and mean average precision (mAP). These criteria are obtained by four well-established components in information retrieval, true positive (TP), false positive (FP), true negative (TN), and false negative (FN). In this article, TP and TN represent the number of correctly detected ships and the number of correctly detected backgrounds, respectively. FP refers to the number of false positives, and FN represents the number of ships that have not been detected.

The precision rate refers to the proportion of correctly detected ships in all detected objects, which represents the correctness of detected objects. It is calculated as

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

The recall rate refers to the proportion of correctly detected ships in all ground truths, which represents the coverage of ground truths. It is calculated as

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (16)$$

Since the precision rate and the recall rate are mutually affecting, the F1-score and the mAP are used to evaluate the overall performance of detection methods. The F1-score is given by

$$\text{F1-score} = 2 \times \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (17)$$

and the mAP is computed as

$$\text{mAP} = \int_0^1 P(R) dR \quad (18)$$

where P and R represent the single point values of the precision rate and recall rate, respectively. PASCAL VOC's mAP is based on an intersection over union (IoU) threshold of 0.5. In contrast, mAP in MS COCO is based on IoU thresholds ranging from 0.5 to 0.95 with 0.05 intervals. AP_S, AP_M, and AP_L are used for evaluation, and these three indexes represent small, medium, and large-scale objects. They correspond to objects whose area ranges are (0, 322), (322, 962), and (962, $+\infty$).

3.3. Results on SSDD

This section evaluates the validity of the proposed method, using Faster RCNN as the baseline. Our previous work [41] has demonstrated that the detection performance of a backbone network using ResNet101 is superior to VGG16 and ResNet50. Therefore, the following experimental analyses are based on ResNet101. In addition, in order to analyze the detection performance of the proposed method in different scenes, experiments are conducted in inshore and offshore scenes respectively.

To verify the influence of different modules in this network, we conducted ablation experiments on inshore ships and offshore ships, respectively, and the results are presented in Tables 2 and 3. As shown in Table 2, due to the relatively simple background of offshore vessels and small interference, the detection performance of all methods is roughly similar, with little difference in each detection indicator. This is because the detection performance of each one has been excellent in the simple background, and only limited improvement can be achieved.

Table 2. The results of ablation studies on offshore ships.

Method	ERM	CAFN	mAP	F1-Score	P	R
BFPN			0.994	0.988	0.989	0.987
BFPN with ERM	✓		0.996	0.985	0.984	0.987
BFPN with CAFN		✓	0.993	0.987	0.991	0.983
Proposed method with ERM and CAFN	✓	✓	0.995	0.988	0.989	0.987

Table 3. The results of ablation studies on inshore ships.

Method	ERM	CAFN	mAP	F1-Score	P	R
BFPN			0.835	0.796	0.778	0.814
BFPN with ERM	✓		0.867	0.834	0.969	0.732
BFPN with CAFN		✓	0.865	0.815	0.868	0.767
Proposed method with ERM and CAFN	✓	✓	0.883	0.836	0.935	0.756

The background clutter of inshore ships is more complex than that of offshore. In addition, the wharfs and buildings on the shore cause great interference to the detections, and SAR ship detection performance generally degrades. As shown in Table 3, the indicators of different approaches significantly declined. However, by enhancing the feature representation capability and reducing the aliasing effect of fusion features, the detection performance is significantly improved. As can be seen from Table 3, the mAP, F1-score, precision rate, and recall rate are 88.3%, 83.6%, 93.5% and 75.6%, respectively, by using ERM and CAFN. Compared with the BFPN method, the mAP and the precision rate of the method are improved by about 5% and 15.7%, respectively. Compared with BFPN network, the mAP and F1-score of the network using only ERM improved by 3.2% and 3.8% respectively, and the mAP and F1-score of those using only CAFN improved by 3% and 2%, respectively. The results show that learning the dependencies between different features from the channel and spatial direction respectively obtains a more accurate distinction between target and background in the feature maps. By optimizing the fusion aliasing feature according to channel weights, the network can inhibit the false detection of background clutter, which improves the performance of ship detection under complicated background. However, due to the dense distribution of inshore ships and the close arrangement of ships, it is easy for the network to identify multiple targets as a single target, reducing the recall rate.

Table 4 shows the detection speeds and the number of model parameters for different network models, where t is the consuming time, FPS presents frames per second of detecting images, and parameters is network parameter number. As can be seen from the table, compared with BFPN, the proposed method only brings a slight computational burden.

This is because we use simple structure in ERM and CAFN rather than heavy structure. Therefore, the proposed network model can achieve the improvement of SAR ship detection performance at a small computational cost.

Table 4. Detection speeds and model information.

Method	t (s)	FPS	Parameters (M)
BFPN	0.123	8.13	60.40
BFPN with ERM	0.128	7.81	60.43
BFPN with CAFN	0.131	7.63	60.43
Proposed method with ERM and CAFN	0.133	7.52	60.46

Figure 9 depicts the comparison results of P-R curves of different methods, where Figure 9a,b respectively represents the results of offshore and inshore tests. In Figure 9a, the P-R curves of these methods show little difference, indicating satisfactory performance in offshore ship detection. In Figure 9b, when the background becomes complex, there are obvious differences between these methods. ERM and CAFN are used to achieve the best detection performance, because we take full advantage of multi-scale features and reduce the aliasing effect of fusion investment. This reflects the validity of the proposed method.

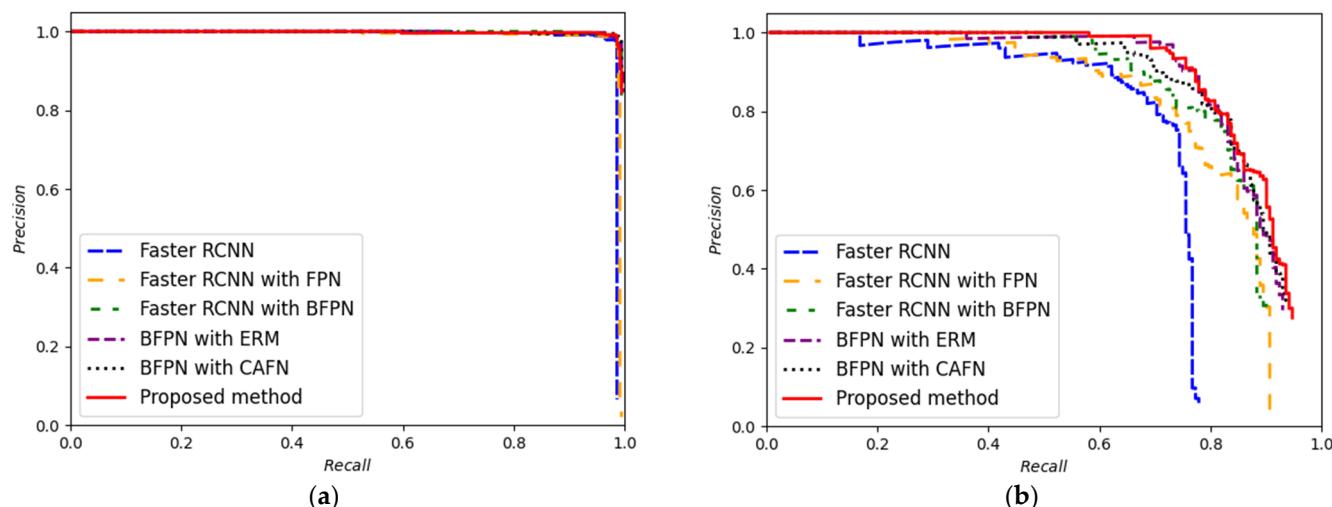


Figure 9. The P-R curves of different methods. (a) P-R curves of different methods on offshore ships. (b) P-R curves of different methods on inshore ships.

Finally, in Table 5, we provide the ship detection statistics generated by bounding box AP on SSDD. Through statistical comparison, the performance of the proposed method on bounding box AP is generally better than that of the two baseline methods. For multi-scale ships detection, the bounding box AP of detector is 57.9% in the detection of small ships in our network, which exceeds FPN and BFPN by about 3.4% and 1.6%, respectively. However, due to the relatively strict definition of large-scale SAR ships in the MS COCO evaluation index, when detecting large ships, the bounding box AP is slightly improved compared with BFPN when detecting large ships. This indicates that the proposed method is more effective in improving the detection performance of small ships.

Table 5. Ship detection statistics generated by bounding box AP on SSDD.

Method	AP	AP ₅₀	AP ₇₅	AP _S	AP _M	AP _L
FPN	0.584	0.945	0.659	0.545	0.637	0.519
BFPN	0.594	0.954	0.662	0.563	0.647	0.584
Proposed method	0.596	0.968	0.699	0.579	0.654	0.591

4. Discussion

4.1. Detection Results of Different Methods

The detection results of offshore and inshore ships in SSDD based on Faster RCNN and other methods are presented in Figures 10 and 11, respectively. In Figure 10a, the pink rectangles represent the corresponding ground truths of the SAR images. Figure 10b shows the visualized detection results of the Faster RCNN network, where detected ship targets are marked with red rectangles. In addition, we also performed experiments on the Faster RCNN with FPN network, and the detection results are shown in the purple rectangle marked in Figure 10c. Finally, Figure 10d describes the detection results of Faster RCNN using BFPN. Figure 10e displays the visualized detection results of the proposed method.

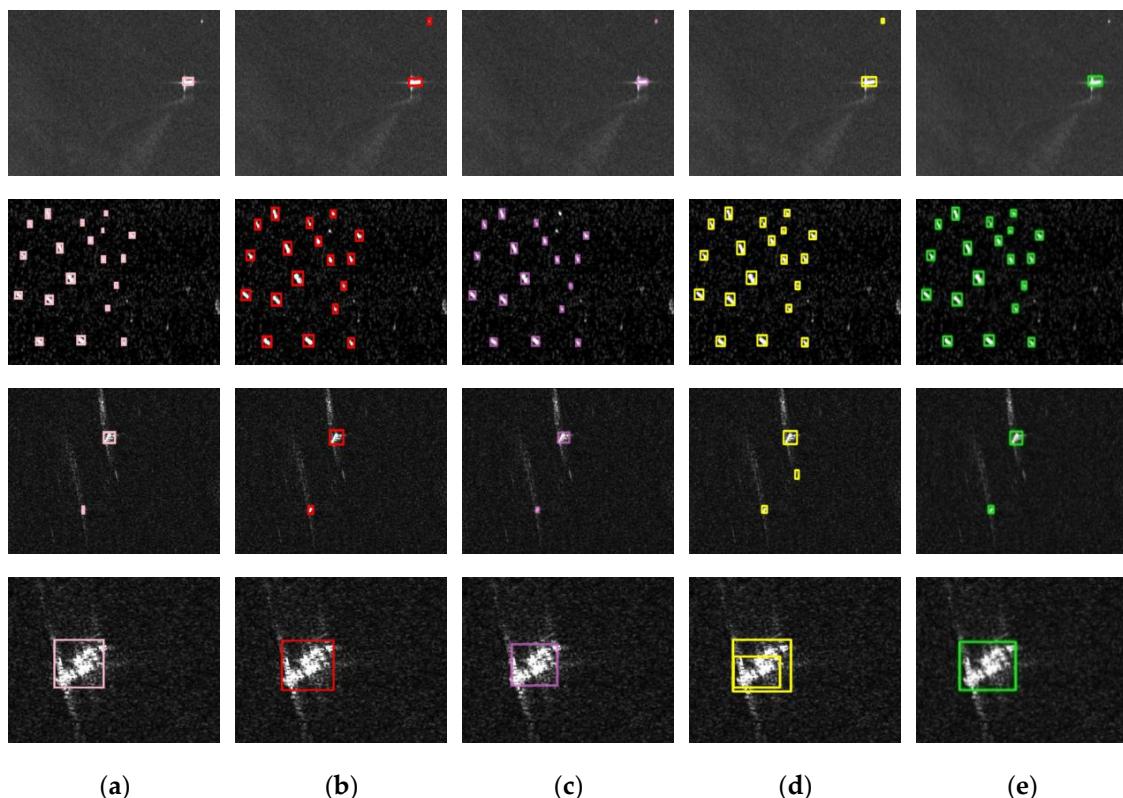


Figure 10. SAR ship detection results of different methods on offshore. (a) Ground truths. (b) Detection results of Faster RCNN. (c) Detection results of Faster RCNN with FPN. (d) Detection results of Faster RCNN with BFPN. (e) Detection results of proposed method.

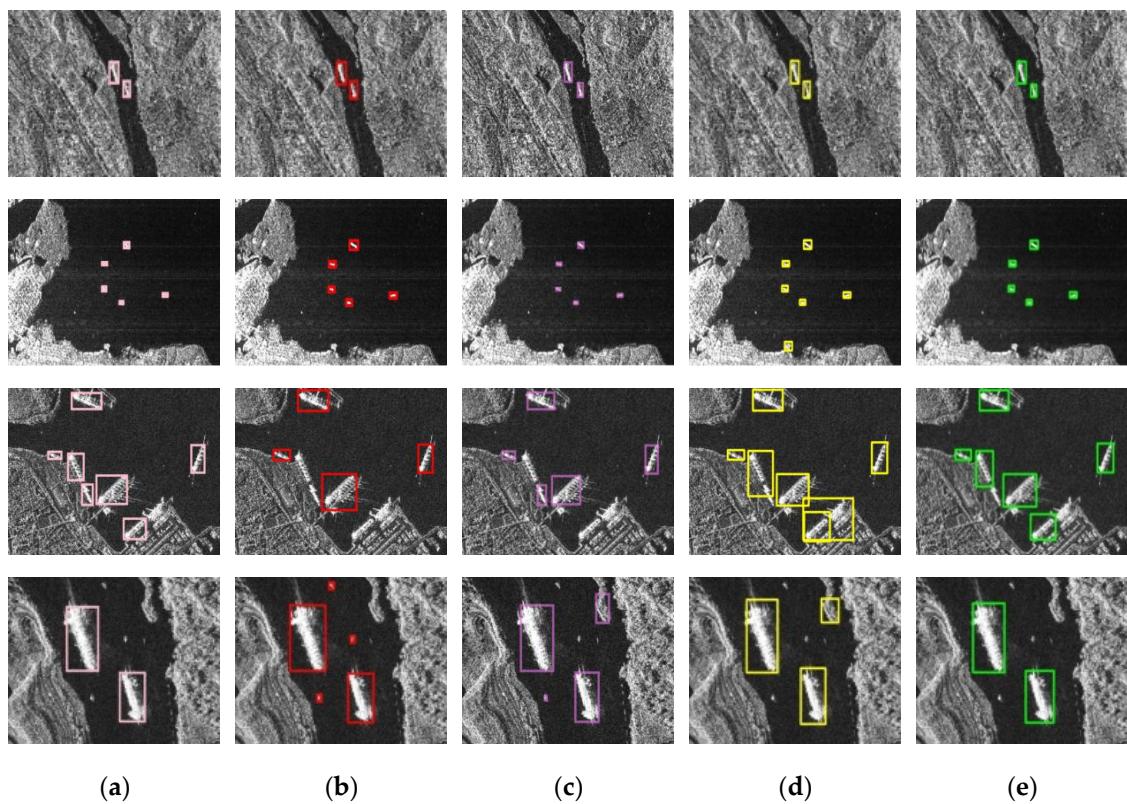


Figure 11. SAR ship detection results of different methods on inshore. (a) Ground truths. (b) Detection results of Faster RCNN. (c) Detection results of Faster RCNN with FPN. (d) Detection results of Faster RCNN with BFPN. (e) Detection results of proposed method.

From Figure 10, both compact small ships and single ships can be detected by the proposed method. Among them, in the first row, the former three methods all mistakenly detect the prominent part in the background as the target, while the detection result of the proposed method is the same as that of ground truth. In the second row, the former two methods miss dense and unsharp ships, which are effectively detected by the method using BFPN. In the last two rows, compared with BFPN method, the proposed method can effectively reduce the number of false detections. In general, compared with other approaches, the proposed method reduces the false positives and missed detection of ships in SAR ship images.

The visualized detection results of different methods on inshore ships are shown in Figure 11. Similar to Figures 10 and 11a refers to the corresponding ground truths, and Figure 11b–e presents the ships detection results of different approaches. For the inshore ships, as the environment of the ships becomes more complex, the problem of missed detection and false positives becomes more obvious. In the first row, small ships in the inshore scene can also be detected by the proposed method. In the second row, due to the great similarity between inshore buildings and ship targets, false positives are easy to occur. However, compared with the BFPN method, the proposed method correctly detects all real ship objects. As can be observed in the last two rows, even though the SAR ship target has a high similarity with the background clutter such as the inshore buildings, it can still obtain accurate detection results compared with the ground truth. Therefore, the proposed method achieves superior performance of SAR ship detection in both offshore scenes and inshore.

4.2. Comparison with the Existing Methods

We compared the proposed A-BFPN with several state-of-the-art CNN-based SAR ship detection methods on SSDD: HR-SDNet [47], DAPN [44], Quad-FPN [48], SER Faster

R-CNN [39], RIF [37], and Faster RCNN [12], FPN [17], BFPN [26]. The comparison results are shown in Table 6. It is clear from the table that A-BFPN achieves the highest detection accuracy of 96.8% in the entire scenes. In particular, in the inshore scenes, excellent performance was achieved with 88.3% mAP. By the development of ERM and CAFN, the proposed deep network model not only emphasizes the features of object areas we want to focus on and restrains unnecessary background clutter, but also enhances the representation of multi-scale semantic information. Therefore, the robust detection performance of SAR ship target is obtained.

Table 6. Comparison results of the other state-of-the-art CNN-based methods on SSDD.

Methods	Entire Scenes			Offshore Scenes			Inshore Scenes		
	mAP	P	R	mAP	P	R	mAP	P	R
Faster RCNN [12]	0.908	0.934	0.887	0.984	0.976	0.986	0.718	0.823	0.703
FPN [17]	0.945	0.955	0.876	0.988	0.976	0.981	0.810	0.869	0.697
BFPN [26]	0.954	0.962	0.893	0.994	0.989	0.987	0.835	0.778	0.814
HR-SDNet [47]	0.908	0.964	0.909	0.985	0.986	0.986	0.736	0.907	0.744
DAPN [44]	0.905	0.855	0.913	0.974	0.975	0.975	0.732	0.641	0.779
Quad-FPN [48]	0.952	0.895	0.857	0.993	0.973	0.994	0.846	0.747	0.877
SER Faster R-CNN [39]	0.915	0.861	0.922	0.982	0.968	0.983	0.745	0.663	0.790
RIF [37]	0.962	0.946	0.932	0.992	0.985	0.982	0.852	0.903	0.762
Proposed method	0.968	0.975	0.944	0.995	0.989	0.987	0.883	0.935	0.756

Finally, we verified the proposed method on the large-scale background SAR ship detection dataset LS-SSDD-V1.0. The experimental results are shown in Table 7. In the offshore scene, our method achieved the best 92.1% mAP. In the inshore scene, the proposed method achieves the best 47.1% mAP. Compared with SSDD, LS-SSDD-V1.0 is richer in background and contains a wealth of pure background images, which increases the difficulty of detection. However, for SSDD, the most studied dataset in the field of SAR image detection, the proposed method achieves the best detection results, which indicates that the proposed method has excellent robustness and generalization ability.

Table 7. Comparison results of the other state-of-the-art CNN-based methods on LS-SSDD-v1.0.

Methods	Entire Scenes			Offshore Scenes			Inshore Scenes		
	mAP	P	R	mAP	P	R	mAP	P	R
Faster RCNN [12]	0.630	0.735	0.658	0.846	0.815	0.874	0.253	0.491	0.291
FPN [17]	0.748	0.737	0.777	0.899	0.828	0.919	0.467	0.559	0.536
BFPN [26]	0.736	0.735	0.767	0.901	0.815	0.920	0.383	0.778	0.419
HR-SDNet [47]	0.688	0.849	0.705	0.883	0.875	0.899	0.348	0.760	0.378
MTL-Det [49]	0.717	-	-	0.887	-	-	0.387	-	-
SII-Net [43]	0.761	0.682	0.793	0.916	0.819	0.934	0.469	0.461	0.554
Proposed method	0.766	0.850	0.736	0.921	0.921	0.889	0.471	0.770	0.545

5. Conclusions

In this article, we propose a robust SAR ship target detection network A-BFPN for complex background clutter, multi-scale variation, and small objects. First, ERM is developed to reduce channel information loss and emphasize the features of the region of interest while reducing the weight of invalid features, which greatly alleviates the effects of complex background clutter and noise. To take full advantage of the semantic features of multi-scale context and reduce the aliasing effect of fusion features, CAFN is introduced to optimize the feature fusion network, which adaptively filters feature information of different levels according to channel weights, improving the detection performance of multi-scale ships. Experiments using SSDD and LS-SSDD-v1.0 datasets and comparisons

with the state-of-the-art methods show that the proposed A-BFPN demonstrates superior performance in SAR ship detection tasks.

Author Contributions: All of the authors contributed significantly to the work. X.L. and D.L. proposed the method; X.L. and J.W. provided suggestions and designed the experiments. X.L. and Z.C. performed the experiments and wrote the paper, H.L. and Q.L. revised the paper. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China, under Grant 61971075 and Grant 62001062; the Opening Project of the Guangxi Wireless Broadband Communication and Signal Processing Key Laboratory, under Grant GXKL06200214 and Grant GXKL06200205; the Engineering Research Center of Mobile Communications, Ministry of Education, under Grant cqupt-mct-202103; the Natural Science Foundation of Chongqing, China under Grant cstc2021jcyj-bshX0085.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

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

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

References

- Li, Z.; Wu, J.; Huang, Y.; Sun, Z.; Yang, J. Ground-moving target imaging and velocity estimation based on mismatched compression for bistatic forward-looking SAR. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 3277–3291. [[CrossRef](#)]
- Zhang, P.; Xu, H.; Tian, T.; Gao, P.; Li, L.; Zhao, T.; Zhang, N.; Tian, J. SEFEPNet: Scale Expansion and Feature Enhancement Pyramid Network for SAR Aircraft Detection with Small Sample Dataset. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2022**, *15*, 3365–3375. [[CrossRef](#)]
- Hong, Z.; Yang, T.; Tong, X.; Zhang, Y.; Jiang, S.; Zhou, R.; Han, Y.; Wang, J.; Yang, S.; Liu, S. Multi-Scale Ship Detection from SAR and Optical Imagery Via A More Accurate YOLOv3. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2021**, *14*, 6083–6101. [[CrossRef](#)]
- Du, L.; Dai, H.; Wang, Y.; Xie, X.; Wang, Z. Target discrimination based on weakly supervised learning for high-resolution SAR images in complex scenes. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 461–472. [[CrossRef](#)]
- Brusch, S.; Lehner, S.; Fritz, T.; Soccorsi, M.; Soloviev, A.; van Schie, B. Ship surveillance with TerraSAR-X. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 1092–1103. [[CrossRef](#)]
- Robey, F.; Fuhrmann, D.; Kelly, E. A CFAR adaptive matched filter detector. *IEEE Trans. Aerosp. Electron. Syst.* **1992**, *28*, 208–216. [[CrossRef](#)]
- Qin, X.; Zhou, S.; Zou, H.; Gao, G. A CFAR detection algorithm for generalized gamma distributed background in high-resolution SAR images. *IEEE Geosci. Remote Sens. Lett.* **2013**, *10*, 806–810.
- Wang, C.; Wang, Z.; Zhang, H.; Zhang, B.; Wu, F. A PolSAR ship detector based on a multi-polarimetric-feature combination using visual attention. *Int. J. Remote Sens.* **2014**, *35*, 7763–7774. [[CrossRef](#)]
- Atteia, G.; Collins, M. On the use of compact polarimetry SAR for ship detection. *ISPRS J. Photogramm. Remote Sens.* **2013**, *80*, 1–9. [[CrossRef](#)]
- Wang, C.; Bi, F.; Chen, L.; Chen, J. A novel threshold template algorithm for ship detection in high-resolution SAR images. In Proceedings of the IEEE International Geoscience Remote Sensing Symposium (IGARSS), Beijing, China, 10–15 July 2016; pp. 100–103.
- Zhu, J.; Qiu, X.; Pan, Z.; Zhang, Y.; Lei, B. Projection shape template-based ship target recognition in TerraSAR-X images. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 222–226. [[CrossRef](#)]
- 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](#)]
- Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 27–30 June 2016; pp. 779–788.
- Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.; Berg, A. SSD: Single shot MultiBox detector. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 8–16 October 2016; pp. 21–37.
- Li, J.; Qu, C.; Shao, J. Ship detection in SAR images based on an improved faster R-CNN. In Proceedings of the Conference on SAR in Big Data Era-Models, Methods and Applications (BGSARDATA), Beijing, China, 13–14 November 2017; pp. 1–6.
- 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](#)]
- Lin, T.; Dollar, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Honolulu, Polynesia, 21–26 July 2017; pp. 936–944.

18. He, K.; Gkioxari, G.; Dollar, P.; Girshick, R. Mask R-CNN. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2980–2988.
19. Dai, J.; Li, Y.; He, K.; Sun, J. R-FCN: Object detection via region-based fully convolutional networks. In Proceedings of the Conference on Neural Information Processing Systems, Barcelona, Spain, 5–10 December 2016; pp. 379–387.
20. Cai, Z.; Vasconcelos, N. Cascade R-CNN: Delving into high quality object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 6154–6162.
21. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 580–587.
22. 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](#)]
23. Girshick, R. Fast R-CNN. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 11–18 December 2015; pp. 1440–1448.
24. Chen, L.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 834–848. [[CrossRef](#)]
25. Guo, C.; Fan, B.; Zhang, Q.; Xiang, S.; Pan, C. AugFPN: Improving Multi-Scale Feature Learning for Object Detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 12592–12601.
26. Pang, J.; Chen, K.; Shi, J.; Feng, H.; Ouyang, W.; Lin, D. Libra R-CNN: Towards balanced learning for object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 821–830.
27. Lin, T.; Goyal, P.; Girshick, R.; He, K.; Dollar, P. Focal loss for dense object detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 318–327. [[CrossRef](#)]
28. Law, H.; Deng, J. CornerNet: Detecting objects as paired keypoints. In Proceedings of the European Conference on Computer Vision, Munich, Germany, 8–14 September 2018; pp. 765–781.
29. Tian, Z.; Shen, C.; Chen, H.; He, T. FCOS: Fully convolutional one-stage object detection. In Proceedings of the IEEE International Conference on Computer Vision, Seoul, Korea, 27 October–2 November 2019; pp. 9626–9635.
30. He, J.; Wang, Y.; Liu, H.; Wang, N.; Wang, J. A novel automatic PolSAR ship detection method based on superpixel-level local information measurement. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 384–388. [[CrossRef](#)]
31. Kapur, J.; Sahoo, P.; Wong, A. A new method for gray-level picture thresholding using the entropy of the histogram. *Comput. Vis. Graph. Image Process.* **1985**, *29*, 273–285. [[CrossRef](#)]
32. Eldhuset, K. An automatic ship and ship wake detection system for spaceborne SAR images in coastal regions. *IEEE Trans. Geosci. Remote Sens.* **1996**, *34*, 1010–1019. [[CrossRef](#)]
33. Shi, Z.; Yu, X.; Jiang, Z.; Li, B. Ship detection in high-resolution optical imagery based on anomaly detector and local shape feature. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 4511–4523.
34. Fan, Q.; Chen, F.; Cheng, M.; Lou, S.; Xiao, R.; Zhang, B.; Wang, C.; Li, J. Ship detection using a fully convolutional network with compact polarimetric SAR images. *Remote Sens.* **2019**, *11*, 2171. [[CrossRef](#)]
35. 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](#)]
36. Guo, H.; Yang, X.; Wang, N.; Song, B.; Gao, X. A rotational Libra R-CNN method for ship detection. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 5772–5781. [[CrossRef](#)]
37. Li, D.; Liang, Q.; Liu, H.; Liu, Q.; Liu, H.; Liao, G. A novel multidimensional domain deep learning network for SAR ship detection. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–13. [[CrossRef](#)]
38. Yu, Y.; Yang, X.; Li, J.; Gao, X. A cascade rotated anchor-aided detector for ship detection in remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–14. [[CrossRef](#)]
39. 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.* **2019**, *16*, 751–755. [[CrossRef](#)]
40. Woo, S.; Park, J.; Lee, J.; Kweon, I. CBAM: Convolutional Block Attention Module. In Proceedings of the European Conference on Computer Vision, Munich, Germany, 8–14 September 2018; pp. 3–19.
41. 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](#)]
42. Du, Y.; Du, L.; Li, L. An SAR Target Detector Based on Gradient Harmonized Mechanism and Attention Mechanism. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 1–5. [[CrossRef](#)]
43. Su, N.; He, J.; Yan, Y.; Zhao, C.; Xing, X. SII-Net: Spatial Information Integration Network for Small Target Detection in SAR Images. *Remote Sens.* **2022**, *14*, 422. [[CrossRef](#)]
44. 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](#)]
45. Fu, J.; Sun, X.; Wang, Z.; Fu, K. An anchor-free method based on feature balancing and refinement network for multiscale ship detection in SAR images. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 1331–1344. [[CrossRef](#)]

46. Tan, M.; Pang, R.; Quoc, V. EfficientDet: Scalable and Efficient Object Detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 10778–10787.
47. 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](#)]
48. Zhang, T.; Zhang, X.; Ke, X. Quad-FPN: A novel quad feature pyramid network for SAR ship detection. *Remote Sens.* **2021**, *13*, 2771. [[CrossRef](#)]
49. 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 Observ. Remote Sens.* **2021**, *14*, 8048–8062. [[CrossRef](#)]