

Closely arranged inshore ship detection using a bi-directional attention feature pyramid network

Hao Guo & Dongbing Gu

To cite this article: Hao Guo & Dongbing Gu (2023) Closely arranged inshore ship detection using a bi-directional attention feature pyramid network, International Journal of Remote Sensing, 44:22, 7106-7125, DOI: [10.1080/01431161.2023.2277166](https://doi.org/10.1080/01431161.2023.2277166)

To link to this article: <https://doi.org/10.1080/01431161.2023.2277166>



© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 22 Nov 2023.



Submit your article to this journal [↗](#)



Article views: 541



View related articles [↗](#)



View Crossmark data [↗](#)

Closely arranged inshore ship detection using a bi-directional attention feature pyramid network

Hao Guo^a and Dongbing Gu^b

^aInformation Science and Technology College, Dalian Maritime University, Dalian, China; ^bSchool of Computer Science and Electronic Engineering, University of Essex, Colchester, UK

ABSTRACT

The detection of inshore ships in Synthetic Aperture Radar (SAR) images is seriously disturbed by shore buildings, especially for closely arranged inshore ships whose appearance is similar when compared with detection of deep-sea ships. There are many interference factors such as speckle noise, cross sidelobes, and defocusing in SAR images. These factors can seriously interfere with feature extraction, and the traditional Fully Convolutional One-Stage (FCOS) network often cannot effectively distinguish small-scale ships from backgrounds. Additionally, for closely arranged inshore ships, missed detections and inaccurate positioning often occur. In this paper, a method of inshore ship detection based on Bi-directional Attention Feature Pyramid Network (BAFPN) is proposed. In order to improve the detection ability of small-scale ships, the BAFPN is based on the FCOS network, which connects a Convolutional Block Attention Module (CBAM) to each feature map of the pyramid and can extract rich semantic features. Then, the idea from Path-Aggregation Network (PANet) is adopted to splice a bottom-up pyramid structure behind the original pyramid structure, further highlighting the features of different scales and improving the ability of the network to accurately locate ships under complex backgrounds, thereby avoiding missed detections in closely arranged inshore ship detection. Finally, a weighted feature fusion method is proposed, which makes the feature information extracted from the feature map have different focuses and can improve the accuracy of ship detection. Experiments on SAR image ship datasets show that the mAP for the SSDD and HRSID reached 0.902 and 0.839 respectively. The proposed method can effectively improve the ship positioning accuracy while maintaining a fast detection speed, and achieves better results for ship detection under complex background.

ARTICLE HISTORY

Received 29 June 2023
Accepted 13 October 2023

KEYWORDS

Deep learning; inshore ship; object detection; Synthetic Aperture Radar (SAR)

1. Introduction

Synthetic Aperture Radar (SAR) is a kind of high-resolution imaging radar, which can obtain high-resolution radar images similar to optical photography all-day and all-

CONTACT Dongbing Gu  dgu@essex.ac.uk  School of Computer Science and Electronic Engineering, University of Essex, Colchester, Essex CO4 3SQ, UK

© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

weather under extremely low visibility weather conditions. Ship detection research using SAR images has clear advantages, which make large-coverage all-weather ship surveillance task possible. Most detection methods in SAR images inherit from the field of optical remote sensing. Detection methods can be divided into four types: (1) Template matching-based object detection; (2) Knowledge-based object detection; (3) Object-Based Image Analysis (OBIA)-based object detection; (4) Object detection based on machine learning or deep learning.

However, compared to optical images, SAR images belong to microwave imaging, which records the backscatter information of ground targets on radar beams. The essence of ship detection in SAR images is to detect ship targets based on the differences in pixel greyscale values between ship targets and sea level backgrounds. However, different types of SAR sensors inevitably have differences in their speckle noise level, incidence angle, and polarization mode. Strong speckle noise (as well as the cross sidelobes and focusing phenomena) can greatly reduce the quality of SAR images. The incidence angle mainly affects the geometric shape of ships in SAR images, and different incidence angles can lead to three phenomena: overlap, perspective shrinkage, and shadow. Different polarization methods can also lead to diversity and uncertainty in ship contours. These factors can seriously interfere with feature extraction, which in turn has a great impact on the accuracy of ship detection.

In addition, for the detection of densely arranged inshore ships, it is affected by complex backgrounds, positioning of densely arranged ships, and insufficient feature information of small

ships. These factors further pose greater challenges to the effectiveness of ship detection.

1.1. Ship missed detection caused by complex nearshore background

The most important factor affecting ship detection is the complex background near shore. The high similarity in scattering characteristics between onshore buildings and ships has led to many land targets being mistakenly detected as ships, leading to a significant increase in false alarms. In addition, drilling platforms and reefs in the ocean can also have an impact on the ship detection performance in SAR images.

1.2. Inaccurate positioning of ships caused by closely arranged ships

Closely arranged ships may appear to stick together in space in the feature display image, making it difficult to distinguish the specific positions of two closely arranged ships. In the algorithm based on key point detection, the key centre points of two closely arranged ships will be extracted to the centre positions of the two ships during the detection process. The detection results are shown in [Figure 1\(b\)](#), which often mistakenly detects tightly arranged ships as a ship.

1.3. Missed detections of small-scale ships

There are many types and quantities of nearshore ships, and at the same resolution, the scale differences presented in SAR images are significant. A small target is considered if

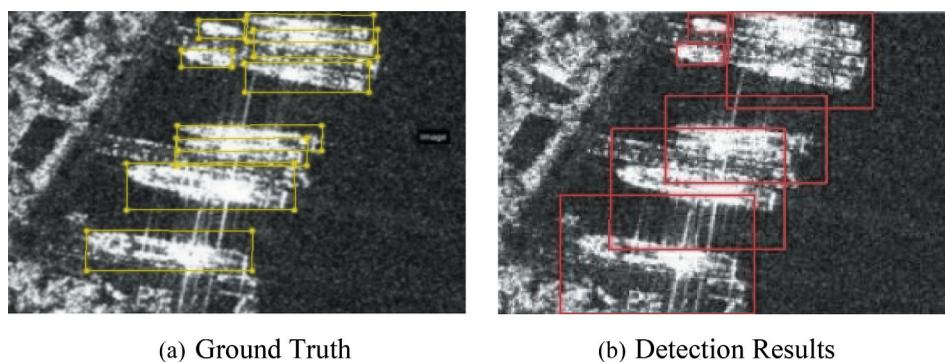


Figure 1. Inaccurate positioning of ships caused by closely arranged ships.

the anchor box area S is less than 32^2 pixels, a medium scale target is considered if the anchor box area S is less than 96^2 pixels, and a large-scale target is considered if the anchor box area S is more than 96^2 pixels (Zhuang et al. 2019). Small ships occupy less pixels in SAR images, and even only show small bright spots. Their contrast with the sea background is not strong, making them easily confused with hidden reefs on the sea, leading to false or missed detections.

A detection method of Bi-directional Attention Feature Pyramid Network (BAFPN) under Fully Convolutional One-Stage (FCOS) network structure is proposed in the paper. Firstly, the Convolutional Block Attention Module (CBAM) is connected to FPN to obtain significant features at different scales. Then, a bottom-up pyramid module is added after the improved FPN to enrich the extracted feature information. In addition, weighted fusion operation is carried out during feature fusion, and different weights are given to different feature maps for feature fusion, thus improving the feature extraction capability of the network. In this way, the method can effectively avoid serious interference from complex background and improve the detection accuracy. Finally, in terms of data processing, this paper carries out training strategies such as image enhancement and transfer learning to further improve the detection performance.

The rest of this paper is structured as follows. The related work is presented in Section 2. Section 3.1 provides a detailed description of the proposed BAFPN framework model. Section 3 introduces the experiment of this method and analyzes the experimental results. Finally, conclusions and outlooks are given in Section 4.

2. Related works

Inshore areas include land and sea. In SAR images, backscattering from the land and sea is generally different. These regions are usually bright and dark, respectively. SAR images are affected by the multiplicative noise known as speckle that makes the analysis and interpretation of SAR images difficult. Besides speckle, nonuniform characteristics of the signals returned from the sea surface, variety of sea states, and coastal regions complexity make accurate feature extraction a very challenging task (Modava, Akbarzadeh, and Soroosh 2018, 2019). The traditional methods of SAR ship detection have difficulty in detecting small-scale ships and avoiding the interference of inshore complex background.

The detection methods based on deep learning are considered more effective, especially the Convolutional Neural Network (CNN) evolves rapidly in the field of computer vision., and deep learning architectures include CNN, Artificial Neural Network (ANN), Recurrent Neural Network (RNN), feed backward, feed forward, Binary Long Short-Term Memory (BLSTM), and many others. The deep learning-based model performs well in generating high recognition rates with comparatively small amount of data, but its computational time is very high and most of these models are task driven (like some models such as BLSTM, RNN, feed forward, and feed backward are efficient for time related problems, CNN, VGG16, VGG32 perform better for spatial related problems) (Yasir et al. 2023).

Object detection methods based on CNNs are mainly divided into two categories: two-stage and one-stage. The representatives of two-stage object detection frameworks include R-CNN (Girshick et al. 2015), Faster R-CNN (Ren et al. 2017). The representatives of one-stage object detection frameworks include YOLO (Redmon et al. 2016) and SSD (Liu et al. 2016). However, these methods heavily rely on the final layer to predict the object detection results, which is prone to lose many high level important semantic features. In order to utilize more levels of features, it is proposed that the Feature Pyramid Network (FPN) (Liu et al. 2016) extracts the features from different levels for individual prediction, so as to achieve better detection performance. Woo et al. (2018) proposed an attention module that combines spatial attention and channel attention to refine the feature map adaptively. Liu et al. (2018) proposed a bottom-up path aggregation network, which can enhance the feature extraction effect. Tian et al. (2019) proposed the FCOS detection algorithm based on a Full Convolutional Neural (FCN) (Shelhamer, Long, and Darrell 2017) network, which is a pixel-level prediction network and can avoid the sensitive size of candidate regions. Due to its better performance, we use the FCOS as the detector in this paper.

In the field of SAR ship detection. Kang et al. (2017) combined the traditional CFAR algorithm with the Faster R-CNN algorithm. Li et al. (2017) improved the Fast R-CNN detection algorithm and applied to the ship detection in SAR image, and a dataset called SAR Ship Detection Dataset (SSDD) was provided to train and test the model. Wang et al. (2018) improved the SSD with a semantic aggregation module and an attention module to detect ship and estimate orientation simultaneously. An et al. (2018) proposed a SAR ship detection method that combines sea clutter distribution with CNN. However, most of these methods aimed at ships in deep-sea areas.

It is more difficult to detect for small ships in inshore scenes under complex backgrounds. Potdar et al. (2021) proposed a SAR ship detection method based on size invariant. The research work undertaken assists the end-user to monitor the activities of the ships, measuring their dimensions, and thus preventing potential mishaps. Zhai et al. (2016) proposed a method to perform the inshore SAR ship detection through saliency and context information. Guo and Zhou (2022) proposed a lightweight SAR ship detection model named MEA-Net for imbalanced datasets to solve the problem of large model structures, high computing resources, and poor detection results of inshore and multi-scale ship targets. Zhang et al. (2022) proposed a SAR ship detection method based on CFAR and CNN. In response to the problem of small size and tight arrangement of ship targets, Li et al. (2020) combined CReLU with the shallow network of SSD, and gradually fused shallow feature information using FPN to detect nearshore ships. In response to the

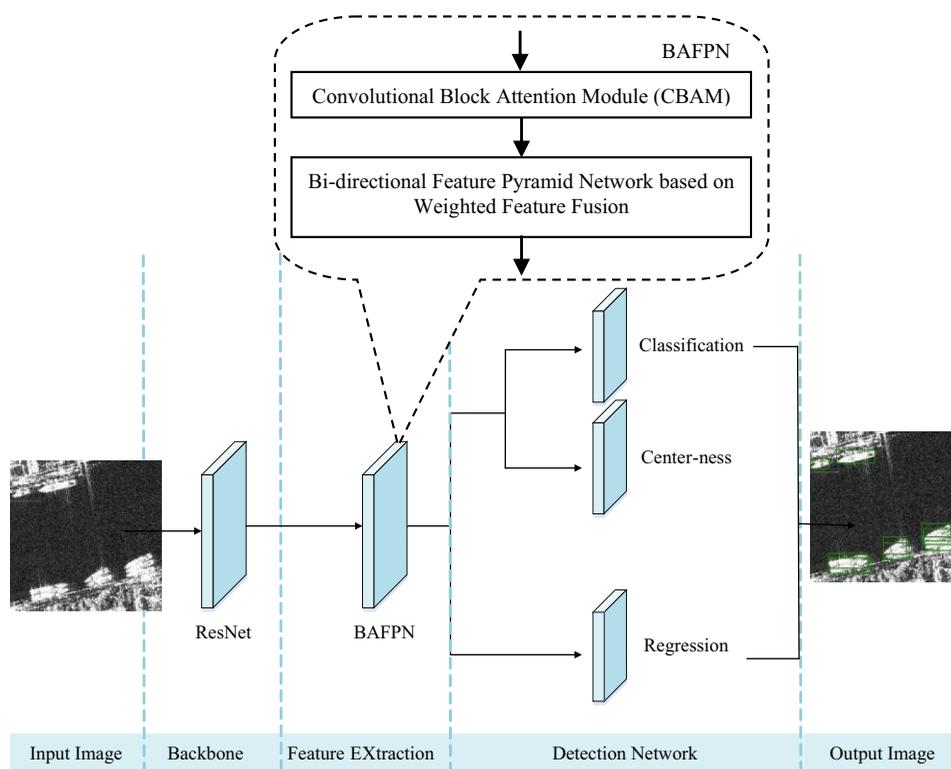


Figure 2. Workflow of proposed method.

problem of high background complexity in SAR ship detection, Song et al. (2022) proposed a rotation detection frame model based on RoI Transformer, which achieves multi-scale feature fusion and background suppression supervision. However, these inshore SAR ship detection methods need to deal with a large number of false alarms in the later period, and is not end-to-end. Recently, the anchor-free target detection methods have attracted lots of attention. They have achieved excellent performance, especially in the targets with multiscale sizes. Ma et al. (2022) proposed an anchor-free detection method and designed a key point estimation module to locate the centre point of target precisely and reduce the false alarm and missing alarm.

3. Methods

This paper proposes an inshore ship detection method for SAR images under complex background. The workflow of the method is shown in Figure 2. This method is based on the FCOS framework. FCOS solves the detection problem by predicting each pixel, similar to semantic segmentation. FCOS directly returns the object boundary box to the position for each location on the feature map, then selects the boundary box with the minimum area as its regression object and makes multi-level predictions through FPN. Next step, it adds the centre-ness branch to reduce the scores away from the boundary box of the object centre, and finally, through the Non-Maximum Suppression (NMS) algorithm,

removes the redundant borders to produce the final detection results. We use ResNet as the feature extraction network. To obtain the image object feature pyramid, a Bi-directional Attention Feature Pyramid Network (BAFPN) is proposed in the feature extraction stage. The specific steps include adding attention modules, establishing a bi-directional feature pyramid, and then fusing the features of different layers by using a weighted feature fusion mechanism. The post-processing structure of FCOS is kept unchanged in the detection stage.

3.1. Convolutional block attention module

The attention mechanism module CBAM is used in the feature extraction phase. The CBAM is used to highlight the significant features of specific scales from the spatial and channel aggregation information of the multi-level feature map. This can adaptively refine the multi-scale feature map, which is conducive to reducing error and improving the accuracy of ship detection at multiple scales.

A CBAM module consists of channel attention and spatial attention. As shown in Figure 3, given an intermediate feature map $F \in R^{C \times H \times W}$ as input, $M_C \in R^{C \times 1 \times 1}$ is the channel attention map and $M_S \in R^{1 \times H \times W}$ is the spatial attention map. The overall process of CBAM is as follows:

$$F_C = M_C(F) \otimes F \quad (1)$$

$$F_R = M_S(F_C) \otimes F_C \quad (2)$$

Where $F_C \in R^{C \times H \times W}$ is the feature map obtained by channel attention and \otimes denotes element-wise multiplication. F_R is the refined feature map.

The channel attention focuses on what is meaningful in the input image. It aggregates the spatial information of the feature map through the average-pooling and the max-pooling operations. Then it reduces the parameters with a multi-layer perceptron (MLP). The channel attention is computed as:

$$M_C(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) \quad (3)$$

Where *AvgPool* and *MaxPool* denote the operations of average pooling and max pooling, and σ denotes the sigmoid function.

Different from the channel attention, the spatial attention focuses on where an informative part is, which is complementary to the channel attention. It aggregates the

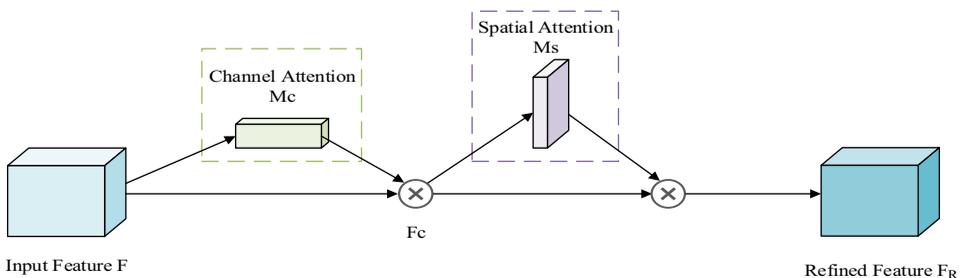


Figure 3. CBAM structure.

channel information of a feature map by using the average-pooling and max-pooling operations. Those are then concatenated and convolved by a standard convolution layer. The spatial attention is computed as:

$$M_S(F) = \sigma(f^{7 \times 7}([\text{AvgPool}(F); \text{MaxPool}(F)])) \quad (4)$$

Where σ denotes the sigmoid function and $f^{7 \times 7}$ represents a convolution operation with the filter size of 7×7 .

The CBAM improves the representation of feature maps through continuous channel attention and spatial attention. This can effectively eliminate the false alarm of various scenes in SAR images, especially the ground objects in inshore scenes. As shown in Figure 4(a), the CBAM is closely connected to the feature map, connected by the up-sampled high-level feature map, and then fused with the original feature map. The process is as follows:

$$A_i = A[U(P_{i+1})] \quad (5)$$

$$P_i = \text{conv}_{3 \times 3}[\text{conv}_{1 \times 1}(C_i) \oplus A_i] \quad (6)$$

Where A and U denote the operations of the CBAM and up-sample, respectively, and \oplus denotes the operation of concatenation.

3.2. Bi-directional feature pyramid network

Figure 4(a) shows the conventional top-down FPN. It is essentially limited by a one-way information flow. To solve this problem, PANet adds an additional bottom-up path aggregation network, as shown in Figure 4(b). The introduction of this network is to enhance the effect of feature extraction by using accurate low-level location information, thus shortening the information path between shallow and deep features.

The proposed method is improved on the basis of FCOS, which is the detection of pixel level and is very important for the extraction of shallow feature information in

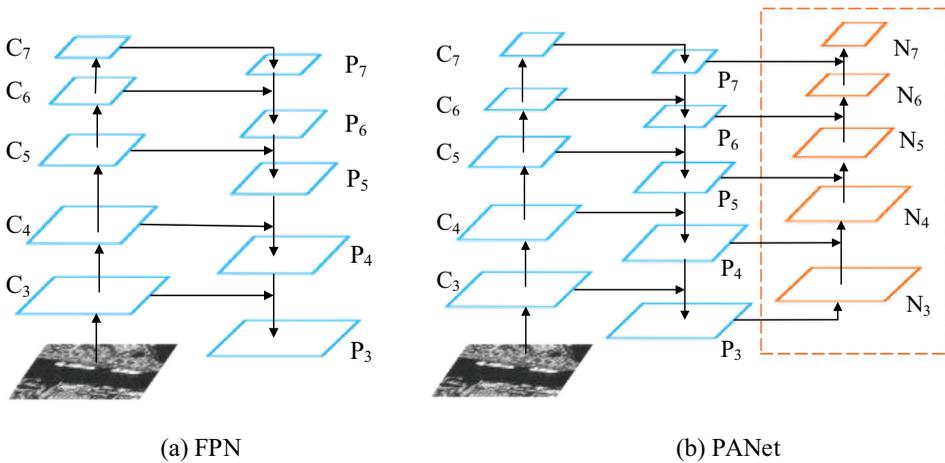


Figure 4. Two traditional features of the pyramid network structure.

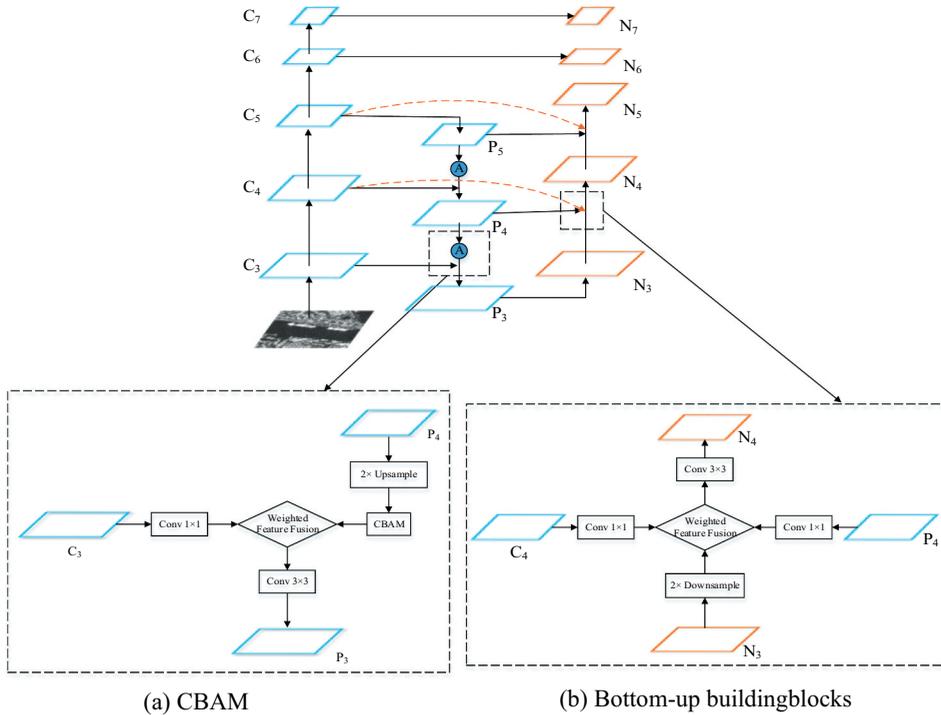


Figure 5. The operational flow of BAFPN.

the network. Inspired by PANet, the BAFPN is constructed in this paper. As shown in [Figure 5](#), behind the top-down network of the FPN, a bottom-up path is connected horizontally. Then, we add an extra connection from the original input to output the feature map if they are at the same level, in order to fuse more features without adding much cost, as shown in the orange dotted line in [Figure 5](#). By merging the global non-fuzzy features with the significant features of different scales in the local area through the horizontal connection, the feature mapping after fusion can be optimized, which can effectively reduce the tightly arranged ship detection error in complex scenes.

As shown in [Figure 5](#), ResNet-50 is utilized as the backbone, which is a 50-layer residual network that adopts shortcut connections. $\{C_3, C_4, C_5\}$ represents the multi-layer feature map of the output of the backbone network, and $\{P_3, P_4, P_5\}$ represents the feature layer generated by the top-down network. The bottom-up pathway from P_3 to P_5 is added, and $\{N_3, N_4, N_5\}$ represents the newly generated feature map corresponding to $\{P_3, P_4, P_5\}$. [Figure 5\(b\)](#) shows the operation between each building block of the bottom-up network. Each building block obtains a high-resolution feature map N_i through a horizontal connection, and after sampling it with a down-sampling operation of 2, it is fused with the feature map P_{i+1} before the horizontal connection, and the original feature map C_{i+1} . Then the fusion feature map is processed with 3×3 convolution layer, and a new

feature map N_{i+1} is generated. This is an iterative process that terminates after P5. The detailed calculation process is as follows:

$$N_i = \text{conv}_{3 \times 3}[D(N_{i-1}) \oplus \text{conv}_{1 \times 1}(C_i) \oplus \text{conv}_{1 \times 1}(P_i)] \quad (7)$$

Where D denotes the operation of down-sample, and \oplus denotes the operation of the proposed weighted feature fusion.

3.3. Weighted feature fusion

When fusing multiple input features with different resolutions, the common practice is to adjust them to the same resolution and then add them together. They treat all input features equally without distinction. However, we note that because different input feature maps have different resolutions, their contributions to the output feature map are often different. Therefore, we propose adding an additional weight for each input in the feature fusion, allowing the network to learn the importance of each feature map on its own. Based on this idea, we introduce a weighted fusion approach.

We apply SoftMax to each weight so that all the weights are normalized to a probability with the value ranging from 0 to 1, indicating the importance of each input feature map. The Softmax function is represented as follows:

$$S = \sum_i \frac{e^{w_i}}{\sum_j e^{w_j}} I_i \quad (8)$$

Where w_i and w_j denote the learning weights, and I_i represents the feature at level i . As a conceptual example, we describe the two fusion features at level 4 shown in Figure 5:

$$P_4 = \text{conv}_{3 \times 3} \left(\frac{e^{w_1} \text{conv}_{1 \times 1}(C_4) + e^{w_2} A_4}{e^{w_1} + e^{w_2}} \right) \quad (9)$$

$$N_4 = \text{conv}_{3 \times 3} \left(\frac{e^{w_1} \text{conv}_{1 \times 1}(C_4) + e^{w_2} \text{conv}_{1 \times 1}(P_4) + e^{w_3} D(N_3)}{e^{w_1} + e^{w_2} + e^{w_3}} \right) \quad (10)$$

Where C_4 is the input feature at level 4, and A_4 is the feature at level 4 after the CBAM. Respectively, P_4 is the intermediate feature at level 4 on the top-down pathway, and N_4 is the output feature at level 4 on the bottom-up pathway. All other features are constructed in a similar manner. Algorithm 1 presents the process of weighted feature fusion.

Algorithm 1 Process of weighted feature fusion

Input: C_i, A_i

Output: P_i

for $i \leftarrow 1$ to n do

$w_i^1 \leftarrow \text{Parameter}(C_i)$ // obtain the weights of the original features

$w_i^2 \leftarrow \text{Parameter}(A_i)$ // obtain the weights of attention features

$w_i^1, w_i^2 \leftarrow \text{softmax}(w_i^1), \text{softmax}(w_i^2)$ // applying the Softmax function to weights

$P_i \leftarrow \text{add}(w_i^1 C_i, w_i^2 A_i)$ // Fusion of features

end for

Return: P_i // Features after fusion

3.4. Ship detection process

The core of BAFPN is the convolutional block attention module and the Bi-directional Feature Pyramid module, both of which use the weighted feature fusion. Algorithm 2 presents the process of BAFPN.

Algorithm 2 Ship detection process using BAFPN

Input: Original ship features C_i

Output: Detected ship targets

for $i \leftarrow 1$ to n do

$A_i \leftarrow CBAM(C_i)$ // Features processed by CBAM.

$P_i \leftarrow w_1^1 C_i + w_1^2 A_i$ // Features of FPN and CBAM after weighted fusion.

$N_3 = conv_{1 \times 1}(P_3)$

$N_i \leftarrow w_2^1 C_i + w_2^2 P_i + w_2^3 upsample(N_{i-1})$ //Features through bidirectional feature pyramid and fusion

return N_i

end for

$boxlist \leftarrow BoxList(tensor(N_i))$ //obtain the initial detection bounding box list

$score \leftarrow classification(boxlist) \times centerness(boxlist)$ //obtain score for the detection bounding box

$bbox \leftarrow nms(boxlist, iou, score)$ //the final detection result after processing with nms

Return $bbox$

4. Experiments and discussion

All the experiments in this paper are programmed under the framework of Pytorch 1.1.0. The experimental platform is NVIDIA Titan RTX, the operating system is ubuntu 18.04, and the experimental environment is CUDA 10.0, Python 3.7.4. We use ResNet-50, which is pre-trained on the ImageNet18 dataset, as the initialization model. Especially, the batch size is set to 8, the learning rate is set to 0.001, the weight decay is set to 0.0001, the momentum is set to 0.9. And the maximum number of iterations is 50,000, and every 20,000 iterations is attenuated.

4.1. Datasets

The two datasets SAR Ship Detection Dataset (SSDD) (Li, Qu, and Shao 2017) and High-Resolution SAR Images Dataset (HRSID) (Wei et al. 2020) are used for training and testing, which are divided into training sets, validation sets, and test sets on a 7:1:2 scale. In addition, in order to test the effectiveness of the algorithm on the large scene SAR image, the proposed BAFPN is evaluated using Radarsat-2 SAR images from the Liugongdao Bay area of Weihai, China.

Deep learning requires sufficient training samples as support, but SAR image data is difficult to obtain in large quantities. In addition, since the imaging mechanisms of SAR image and visible image are different, a reasonable choice is needed for image enhancement strategy. To fully use the limited training data, we enhance the training data through a series of transformations. Data augmentation is beneficial to inhibit the potential overfitting and strengthen the generalization ability of models. In this paper, affine transformation, blurring and adding noise are used to enhance the dataset.

4.1.1. SSDD dataset

One of the datasets used is the exposed SAR ship dataset SSDD (Li, Qu, and Shao 2017), which imitates the POSCOL VOC dataset, with 1,160 images and 2,456 ships, and the SAR

images in the dataset are mainly from RadarSat-2, TerraSAR-X, and Sentinel-1 sensors, and the resolutions include 1 m, 3 m, 5 m, 7 m, 10 m. In addition, the ship object is small in low resolution images, and only ship targets with more than three pixels are labelled.

4.1.2. HRSID dataset

The HRSID dataset contains only one category of ships and annotates ship position information (Wei et al. 2020). Sensor types include Sentinel-1B, TerraSAR-X, and TanDEM. The dataset image size is 800×800 , the scene includes a wide variety of nearshore areas of ships and areas where ships are difficult to distinguish from clutter interference. The HRSID dataset has a total of 5604 fully polarized SAR images, with below 32×32 corresponding to small objects, with 32×32 to 96×96 corresponding to the middle object, with above 96×96 corresponding to large objects, with small ships, neutral ships, and large ships accounting for 54.5%, 43.5%, and 2% of the total number of ships in the HRSID dataset, respectively.

4.1.3. Large scene SAR image

In order to test the effectiveness of the algorithm on large scene SAR images, the proposed BAFPN is evaluated using Radarsat-2 images from the Liugongdao Bay area of Weihai, China. The pixel size of the SAR image is 3248×6130 , with the imaging time of 17 July 2013, an incidence angle of 21.10° , and a resolution of 8 m.

4.2. Comparative analysis

4.2.1. Inshore docked ship detection

Ships docked near shore are susceptible to the influence of onshore buildings during the detection process, which is extremely challenging for the detection task of the model. Figure 6 shows the detection results of several algorithms on the ships docked near shore in the nearshore scenario. Overall, the SSD algorithm and Faster R-CNN algorithm have poor detection performance for the ships docked near shore. Among them, the detection of SSD algorithm is easily affected by container interferences on shore, leading to missed detection. Compared to the SSD algorithm, the CenterNet and MEA-Net perform relatively well in detecting inshore ships. However, from the image corresponding to CenterNet in Figure 6(a) and mAP in Figure 6(c), it can be seen that this algorithm is still prone to missed detections for nearshore docked and adjacent ships. Our algorithm adds a channel attention module, which enables the network model to actively learn and add different weights to each channel. During the detection process, it is not affected by the nearshore background and can accurately locate the position of ship targets.

4.2.2. Closely arranged shi

Figure 7 shows the visualization results of the SSD algorithm, CenterNet algorithm, and our algorithm for densely arranged ship detection at estuaries. Due to the large size of densely arranged ship targets in the images, both the SSD algorithm and CenterNet algorithm have shown good detection performance for this type of ship targets, which can accurately locate the ship targets. However, for densely arranged ships with smaller storage, the SSD algorithm still has missed detections and poor detection performance.

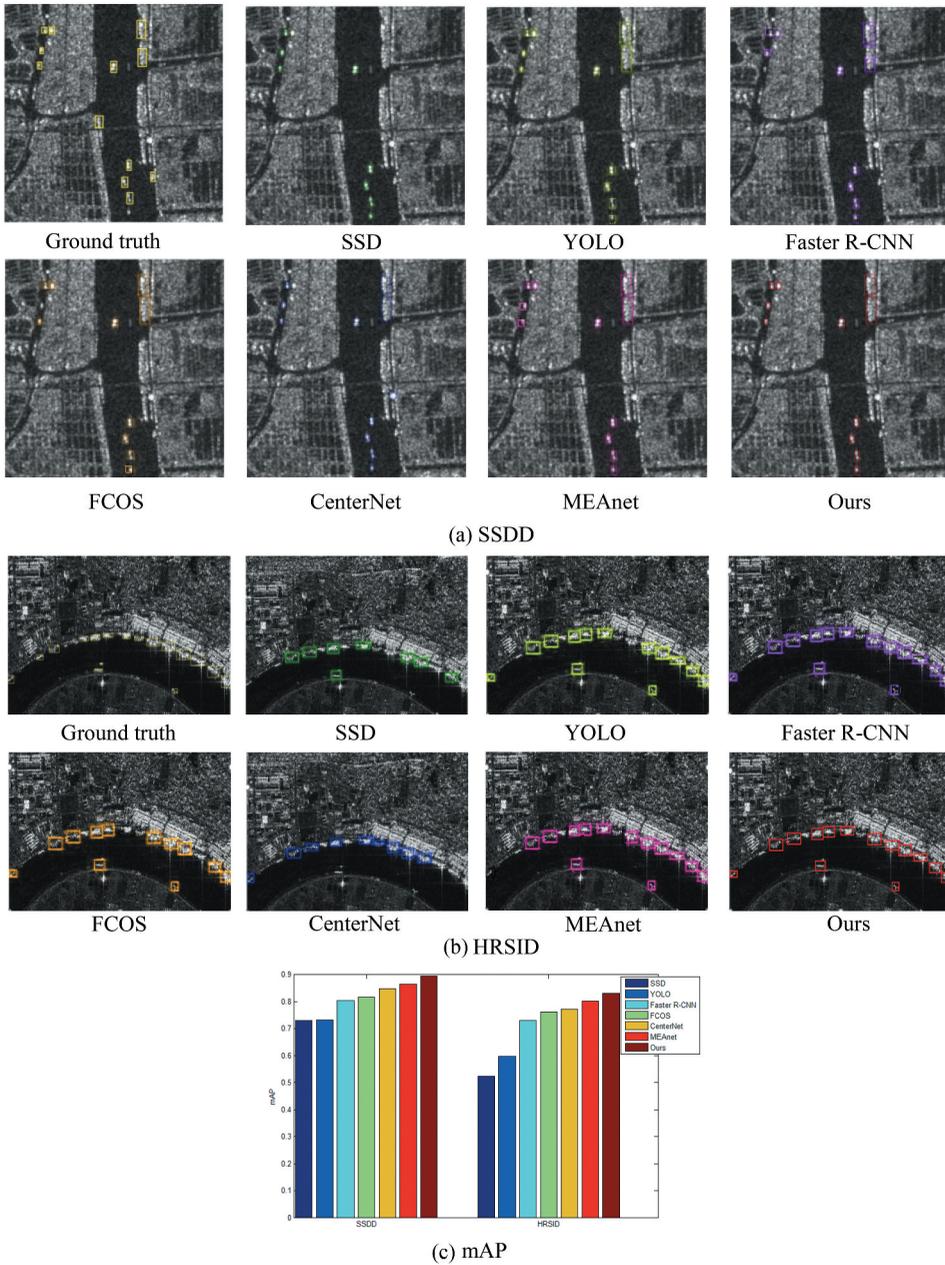
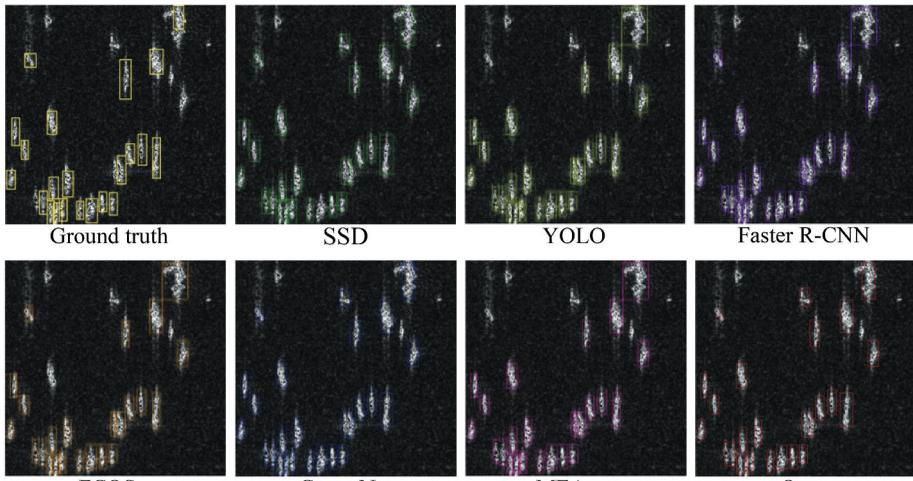
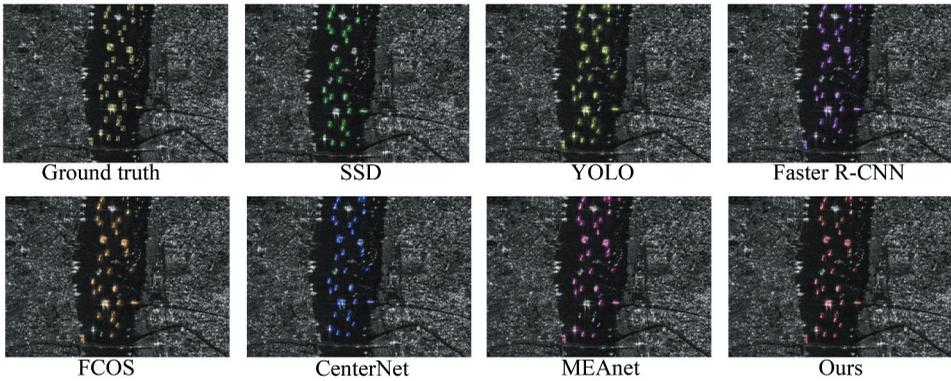


Figure 6. Detection results on different datasets.

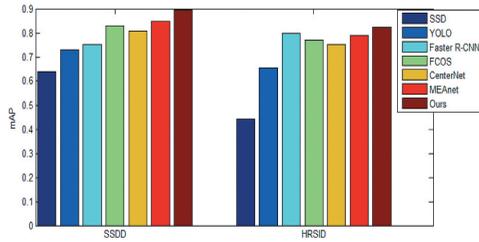
On the contrary, the CenterNet algorithm has improved the detection performance compared to SSD. However, the CenterNet algorithm is easy to recognize two adjacent ships as one ship, indicating that for small-scale ship targets, it is still necessary to strengthen the feature extraction methods, and distinguishing adjacent arranged ships is still a key concern. Our algorithm enhances the feature extraction to obtain the contextual information. From mAP in [Figure 7\(c\)](#), it can be seen that the detection results



(a) SSDD



(b) HRSID



(c) mAP

Figure 7. Closely arranged ships in deep sea.

of our algorithm can effectively and accurately locate the size of small ships, which is more accurate for detecting densely arranged ship targets compared with other algorithms. This to some extent demonstrates the generalization ability of the proposed method.

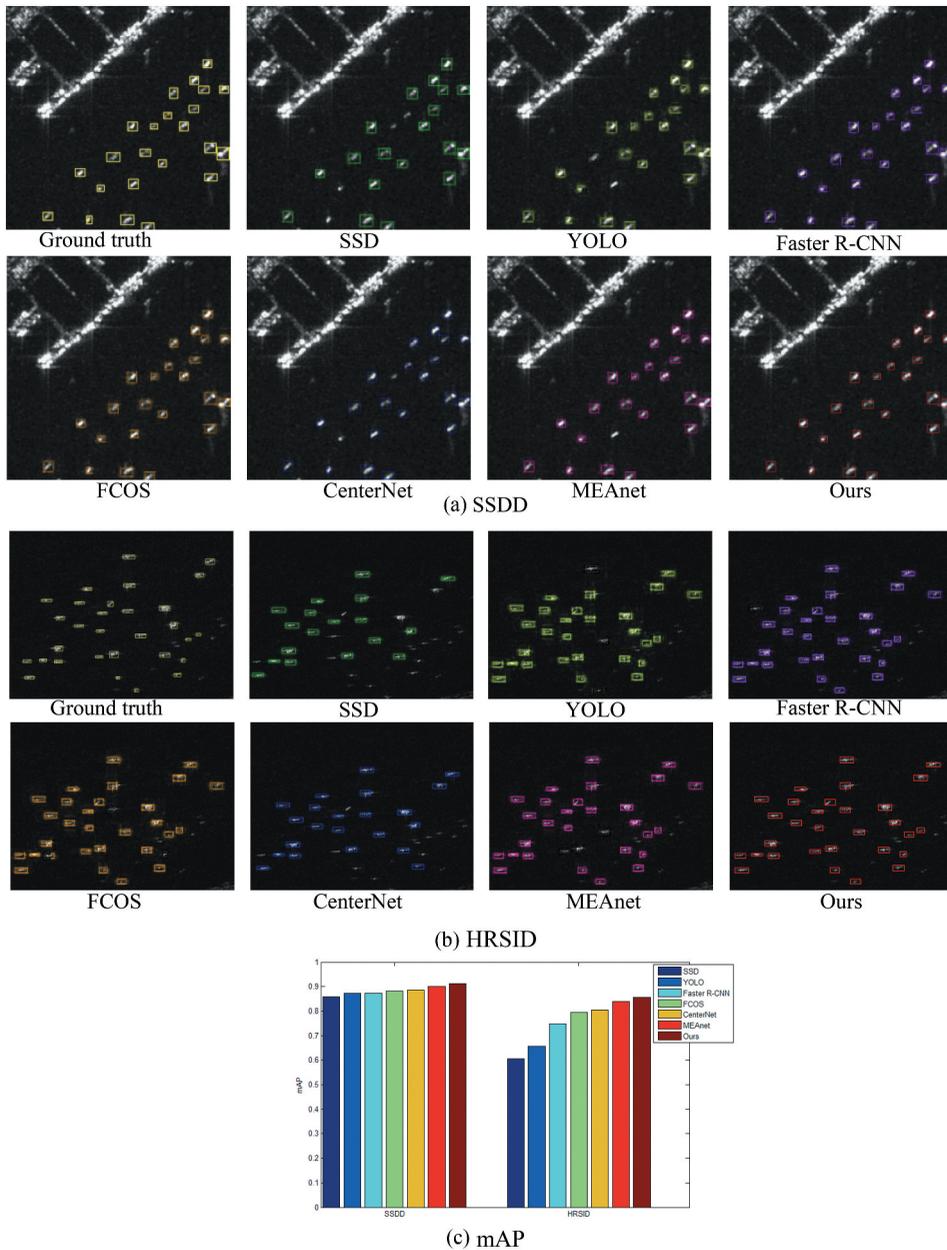


Figure 8. Small ship scenarios at deep sea.

4.2.3. Small ship detection

Figure 8 shows the effectiveness of several mainstream algorithms for detecting small ships. The SSD algorithm and CenterNet algorithm have poor performance in detecting small ships in the HRSID and SSDD. During the feature extraction process, the feature information of small ship targets is lost in multiple convolution processes, making it easy to miss detection during the detection phase. The algorithm in this article adds a feature fusion approach to better

Table 1. Comparison between the proposed method and other methods with SSDD dataset.

Model	Backbone	mAP	Training time(s)	Testing time(s)	Memory(MB)
SSD	VGGNet	0.7569	1.103	0.023	1189
YOLO-v3	DarkNet-53	0.7883	1.153	0.053	1769
Faster R-CNN	ResNet-50	0.8170	2.159	0.169	1375
FCOS	ResNet-50	0.8074	0.446	0.051	1164
CenterNet	ResNet-50	0.8033	0.457	0.042	1289
MEA-Net	CSPDarknet53	0.8747	0.695	0.078	1534
Ours	ResNet-50	0.9022	0.487	0.049	1467

Table 2. Comparison between the proposed method and other methods on the HRSID dataset.

Model	Backbone	mAP	Training time(s)	Testing time(s)	Memory(MB)
SSD	VGGNet	0.5330	1.103	0.023	1189
YOLO-v3	DarkNet-53	0.6384	1.153	0.030	1769
Faster R-CNN	ResNet-50	0.7570	2.159	0.104	1375
FCOS	ResNet-50	0.8074	0.446	0.031	1164
CenterNet	ResNet-50	0.7197	0.457	0.026	1289
MEA-Net	CSPDarknet53	0.8158	0.695	0.097	1534
Ours	ResNet-50	0.8386	0.487	0.028	1467

integrate shallow and deep feature information. In the feature extraction stage, richer ship target features are extracted, and the detection results are improved. The algorithm in this article can achieve accurate positioning results for both large-scale and small-scale targets.

4.3. Overall comparative analysis

In order to better explain the detection performance of the proposed algorithm, we compare the proposed method with other detection algorithms. The quantitative comparison with the SSDD dataset is shown in Table 1. As it can be seen from the table, compared with the SSD, MEA-Net and Faster R-CNN of different backbone networks, this method has the highest accuracy in the SSDD dataset. The training time of each image is faster than that of the other algorithms, and although it is slower than the SSD, it can still be used for real-time detection.

Further experiments are conducted on the HRSID dataset, comparing the algorithm with the current ship target detection algorithms based on deep learning. The experimental results are shown in Table 2. From it we can see that the proposed algorithm has an accuracy of 83.9%, which is the highest.

4.4. Ship detection performance on different scenarios

The ship detection performance of different scenarios in the HRSID dataset are shown in Figure 9. Figure 9(a) shows the results of ship target detection in complex backgrounds. It can be seen that the proposed algorithm can detect ships affected by background interference. Figure 9(b) shows the results of ships with spatially dense distribution at the river mouth. The ships are densely distributed and adjacent ships are densely arranged with a large number. It can be seen that the proposed algorithm can achieve good detection results for the ships docked at estuaries, whether sparsely or densely distributed in space. Especially, in Figure 9(c), the blurred background is

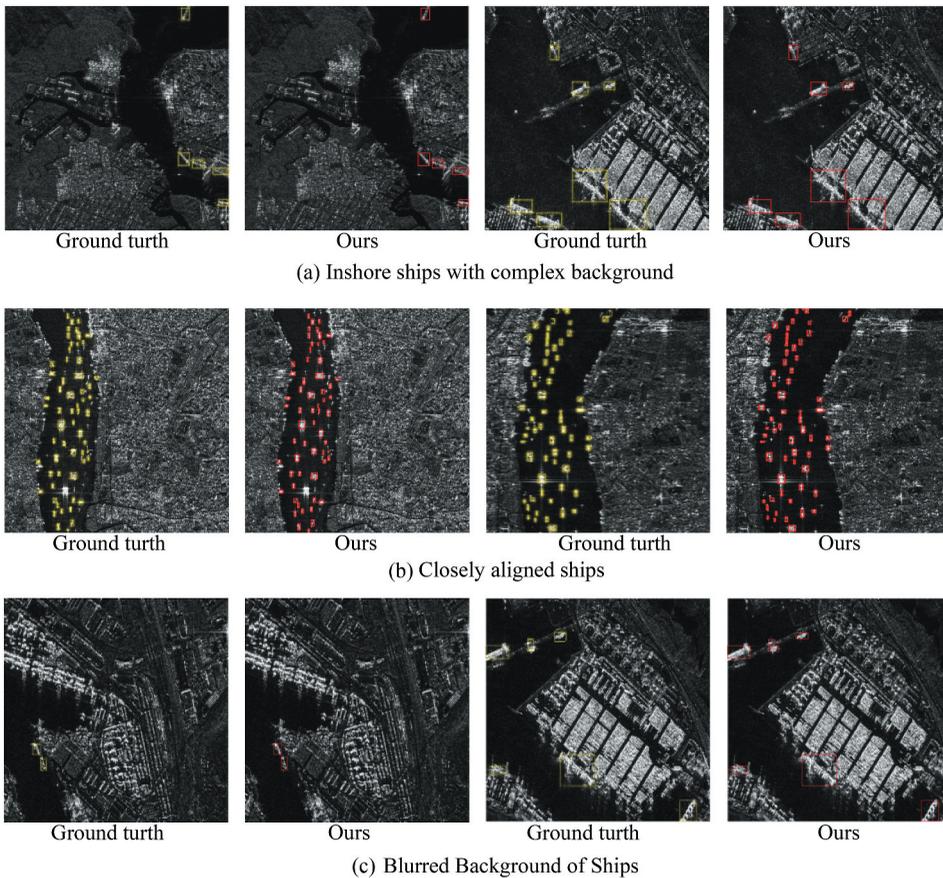


Figure 9. Results on the HRSID dataset.

affected by sea clutter. In this case, the ship features are blurry and cannot distinguish between the blurred background and the ship target. It may be difficult to distinguish whether it is the target or the background with the naked eye for the first and third detection images in Figure 9(c). However, the proposed algorithm can still detect the ship targets with significant scale differences and background interference.

4.5. Testing on large scene SAR images

We annotate the Radarsat-2 data from the Weihai Bay area using LabelImg. Due to the fact that some suspected ship targets in SAR images only have a few pixels, this article considers the targets larger than 10 pixels as ships and annotates them. The incident angle of 21.10° , and we select the HV channel of Radarsat-2 for testing. Figure 10 shows the ship detection results in the RadarSat-2 fully polarized data image. These images are local images of the Weihai region. From Figure 10, it can be seen that firstly, overall in scenarios with complex backgrounds and high levels of noise, most ships can be accurately detected in both nearshore and deep-sea scenes, indicating that the proposed method has strong robustness. Secondly, in complex nearshore scenarios, the information

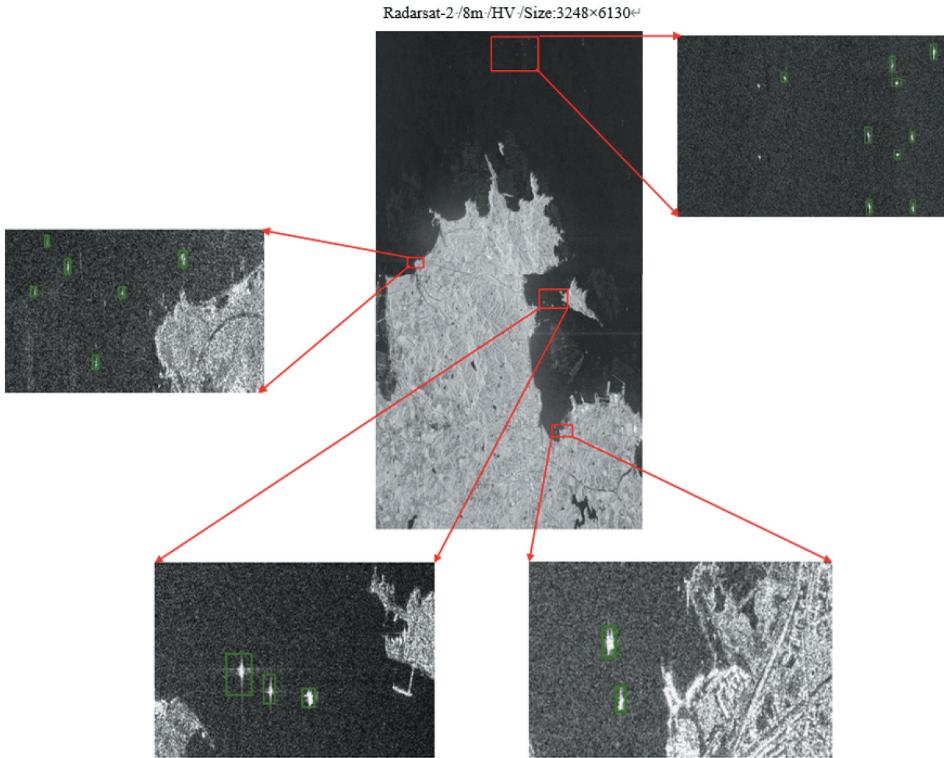


Figure 10. The ship detection results in the RadarSat-2 fully polarized data image.

features of these ship targets are usually similar to those of buildings and other interfering objects on land. However, our algorithm can still effectively detect the target, indicating that our algorithm is effective for nearshore ship targets.

4.6. Ablation analysis

In order to evaluate the detection performance of the different modules added in this paper on SAR ship images, four tests are conducted, as shown in Table 3. The FCOS+CBAM represents the model after adding the CBAM, the FCOS+BAFPN represents the model after adding the bi-directional attention feature pyramid, and the FCOS+BAFPN+Weight represents the model after the weighted feature fusion. Through the experiments, it can be seen that the CBAM has a more significant effect on the SAR ship target detection results,

Table 3. The accuracy of each module is evaluated and compared.

Model	mAP
FCOS	0.8075
FCOS+CBAM	0.8697
FCOS+BAFPN	0.8877
FCOS+BAFPN+Weight	0.9046

mainly because the original algorithm for the feature extraction is ambiguous and could mistakenly identify multiple ship as one target. The model after adding the CBAM is more accurate for the significant feature extraction of the ship. Moreover, with the addition of different modules, the positioning information for ship targets is more accurate, and the ship targets with different distributions can be effectively distinguished, so that the final detection accuracy is higher.

5. Conclusions and outlooks

In this paper, a new SAR image ship detection algorithm based on the FCOS framework is proposed, and the proposed algorithm is evaluated on the two datasets SSDD and HRSID. Aiming at the difficulty of detecting SAR ship targets in complex background, we apply the pixel-by-pixel detection method to the image detection of SAR ships. We add the CBAM in the network structure, which highlights the significant features of the target, and makes the algorithm effectively distinguish the tightly arranged ships. At the same time, this paper proposes a bi-directional attention feature pyramid network, which enhances the accuracy of the algorithm positioning function. In addition, a weighted feature fusion is introduced in the algorithm to improve the robustness for SAR ship targets in complex background. It should be pointed out that with the increase of SAR image data, in the future research work, we will further optimize the neural network structure to enhance the robust performance of the network, and improve the accuracy and efficiency of target detection.

The imaging quality directly determines the effectiveness of ship detection. Unlike optical images, SAR images vary depending on the incident angle, wavelength, and polarization method. Adil et al. (2022) conducted a study based on the L-band UAVSAR airborne dataset, which showed that the HV polarization provides the largest target-to-clutter ratio at lower incidence angles, while the HH polarization should be preferred at higher angles of incidence. Similarly, the incident angle of Radarsat-2 data (C-band) from the Weihai Bay area is 21.10° . We found that better ship detection results can be obtained based on HV channel. However, considering the effectiveness of FPN and attention mechanism in optical image object detection, overall, as long as the sample space is abundant enough, this method has a great possibility of expanding to the L-band. However, further experimental verification is needed in the future.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

- Adil, M., A. Buono, F. Nunziata, E. Ferrentino, D. Velotto, and M. Migliaccio. 2022. "On the Effects of the Incidence Angle on the L-Band Multi-Polarisation Scattering of a Small Ship." *Remote Sensing* 14 (22): 5813. <https://doi.org/10.3390/rs14225813>.
- An, Q., Z. Pan, and H. You. 2018. "Ship Detection in Gaofen-3 SAR Images Based on Sea Clutter Distribution Analysis and Deep Convolutional Neural Network." *Sensors* 18 (2): 334. <https://doi.org/10.3390/s18020334>.

- Girshick, R., J. Donahue, T. Darrell, and J. Malik. 2015. "Region-Based Convolutional Networks for Accurate Object Detection and Segmentation." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 38 (1): 142–158. <https://doi.org/10.1109/TPAMI.2015.2437384>.
- Guo, Y., and L. Zhou. 2022. "MEA-Net: A Lightweight SAR Ship Detection Model for Imbalanced Datasets." *Remote Sensing* 14 (18): 4438. <https://doi.org/10.3390/rs14184438>.
- Kang, M., X. Leng, Z. Lin, and K. Ji. 2017. A Modified Faster R-CNN Based on CFAR Algorithm for SAR Ship Detection. In *2017 International Workshop on Remote Sensing with Intelligent Processing (RSIP)* Shanghai, China: 1–4. IEEE.
- Li, J., C. Qu, and J. Shao. 2017. Ship Detection in SAR Images Based on an Improved Faster R-CNN. In *2017 SAR in Big Data Era: Models, Methods and Applications (BIGSAR DATA)* :1–6.
- Liu, W., D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Y. Fu, and A. C. Berg. 2016. Ssd: Single shot multibox detector. In *Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I* 14: 21–37. Springer International Publishing.
- Liu, S., L. Qi, H. Qin, J. Shi, and J. Jia. 2018. Path Aggregation Network for Instance Segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* Salt Lake City, UT, USA : 8759–8768.
- Li, F. H., K. P. Zhou, and T. C. Han. 2020. "Improved SSD Ship Target Detection Based on CReLU and FPN." *Journal of Instrumentation* 41 (4): 183–190.
- Ma, X., S. Hou, Y. Wang, J. Wang, and H. Wang. 2022. "Multiscale and Dense Ship Detection in SAR Images Based on Key-Point Estimation and Attention Mechanism." *IEEE Transactions on Geoscience and Remote Sensing* 60:1–11. <https://doi.org/10.1109/TGRS.2022.3141407>.
- Modava, M., G. Akbarizadeh, and M. Soroosh. 2018. "Integration of Spectral Histogram and Level Set for Coastline Detection in SAR Images." *IEEE Transactions on Aerospace and Electronic Systems* 55 (2): 810–819. <https://doi.org/10.1109/TAES.2018.2865120>.
- Modava, M., G. Akbarizadeh, and M. Soroosh. 2019. "Hierarchical Coastline Detection in SAR Images Based on Spectral-Textural Features and Global-Local Information." *IET Radar, Sonar & Navigation* 13 (12): 2183–2195. <https://doi.org/10.1049/iet-rsn.2019.0063>.
- Potdar, A., P. Kingrani, R. Motwani, T. Shetty, and S. Chopra. 2021. Size Invariant Ship Detection Using SAR Images. In *Proceedings of International Conference on Computational Intelligence, Data Science and Cloud Computing: IEM-ICDC 2020* Kolkata, India: 185–195. Springer Singapore.
- Redmon, J., S. Divvala, R. Girshick, and A. Farhadi. 2016. 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 :779–788.
- Ren, S., K. He, R. Girshick, and J. Sun. 2017. "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks." *IEEE Transactions on Pattern Analysis & Machine Intelligence* 39 (6): 1137–1149. <https://doi.org/10.1109/TPAMI.2016.2577031>.
- Shelhamer, E., J. Long, and T. Darrell. 2017. "Fully Convolutional Networks for Semantic Segmentation." *IEEE Transactions on Pattern Analysis & Machine Intelligence* 39 (4): 640–651. <https://doi.org/10.1109/TPAMI.2016.2572683>.
- Song, J. Y., T. T. Li, and J. Tian. 2022. "Cross Modal Domain Adaptive SAR Image Ship Detection and Recognition." *Journal of Huazhong University of Science and Technology (Natural Science Edition)* 50 (11): 107–113.
- Tian, Z., C. Shen, H. Chen, and T. He. 2019. Fcos: Fully convolutional one-stage object detection. In *Proceedings of the IEEE/CVF international conference on computer vision* Seoul, Korea :9627–9636.
- Wang, J., C. Lu, and W. Jiang. 2018. "Simultaneous Ship Detection and Orientation Estimation in SAR Images Based on Attention Module and Angle Regression." *Sensors* 18 (9): 2851. <https://doi.org/10.3390/s18092851>.
- Wei, S., X. Zeng, Q. Qu, M. Wang, H. Su, and J. Shi. 2020. "HRSID: A High-Resolution SAR Images Dataset for Ship Detection and Instance Segmentation." *Institute of Electrical and Electronics Engineers Access* 8:120234–120254. <https://doi.org/10.1109/ACCESS.2020.3005861>.
- Woo, S., J. Park, J. Y. Lee, and I. S. Kweon. 2018. Cbam: Convolutional Block Attention Module. In *Proceedings of the European conference on computer vision (ECCV)* Munich, Germany :3–19.

- Yasir, M., W. Jianhua, X. Mingming, S. Hui, Z. Zhe, L. Shanwei, A. T. I. Colak, et al. 2023. "Ship Detection Based on Deep Learning Using SAR Imagery: A Systematic Literature Review." *Soft Computing* 27 (1): 63–84. <https://doi.org/10.1007/s00500-022-07522-w>.
- Zhai, L., Y. Li, and Y. Su. 2016. "Inshore Ship Detection via Saliency and Context Information in High-Resolution SAR Images." *IEEE Geoscience and Remote Sensing Letters* 13 (12): 1870–1874. <https://doi.org/10.1109/LGRS.2016.2616187>.
- Zhang, G., Y. Zhao, X. Chen, B. Li, J. Wang, and D. Liu. 2022. "SAR Image Target Recognition Technology Based on CFAR and CNN." *Electrooptic and Control* 29 (7): 119–125.
- Zhuang, S., P. Wang, B. Jiang, G. Wang, and C. Wang. 2019. "A Single Shot Framework with Multi-Scale Feature Fusion for Geospatial Object Detection." *Remote Sensing* 11 (5): 594. <https://doi.org/10.3390/rs11050594>.