International Core Journal of Engineering Volume 8 Issue 6, 2022
ISSN: 2414-1895 DOI: 10.6919/ICJE.202206_8(6).0041

Improved YOLOv5 with Attention Mechanism for SAR Ship Target Detection in Complex Environment

Jiajia Feng[1, a], Zhijing Xu[1, b], and Zhi Liu[2, c]

[1] College of Information Engineering, Shanghai Maritime University, Shanghai 201306, China

[2] Faculty of Management, Kyushu Sangyo University, Fukuoka, Japan

[a]2211048682@qq.com, [b]zjxu@shmtu.edu.cn, [c]389512975@qq.com

Abstract

For ship target detection in synthetic aperture radar (SAR) images in complex environments, an algorithm based on improved YOLOv5 is proposed. First, a new feature pyramid network is proposed by using residual network thinking, which adds weighted cross-scale connections and reduces the loss of small target feature information; Secondly, the CBAM attention module suppresses the learning of background features and improves the feature learning ability of ship targets in complex backgrounds; Finally, the Soft-NMS linear penalty function is introduced to reduce the loss of the bounding box of the ship in the complex environment and improve the average accuracy. The experimental results show that the method proposed in this paper can effectively detect the ship target in the complex background, the detection speed is fast, and the missed detection rate is low.

Keywords

SAR; Ship Detection; YOLOv5; Feature Pyramid Network; Attention Mechanism.

1. Introduction

Synthetic Aperture Radar is an active microwave sensing system. Compared with traditional optical or infrared imaging sensors, its imaging is not affected by external environmental factors such as light and climate. It has the characteristics of all-weather, all-weather and wide range[1]. At present, it has become one of the important means of earth observation. In recent years, with the development of high-resolution SAR microwave imaging technology, SAR image target detection has become a research hotspot. Ship target detection based on SAR image can quickly and accurately detect ship targets in various scenes. It has great application value in military and civil fields. At the same time, it also puts forward higher requirements for the speed and accuracy of target detection algorithm in practical application.

Traditional ship target detection algorithms, such as CFAR[2], which is a statistical model based on contrast information, selects thresholds according to the statistical characteristics of background clutter to achieve target detection. However, CFAR has problems such as complex modeling process and long detection time, and generally only applicable to It is difficult to adapt to the changes of the complex environment in the ocean, and it cannot meet the needs of practical applications in terms of detection accuracy and detection efficiency. Convolutional neural network has become the mainstream algorithm of target detection technology due to its advantages of high precision, high efficiency and high robustness[3], so the SAR image target detection model based on deep learning method has also become a key research trend.

The target detection based on deep learning mainly includes: a two-stage detection model represented by R-CNN (Region-based CNN) and a single-stage detection model represented by YOLO (You Only

ISSN: 2414-1895 DOI: 10.6919/ICJE.202206_8(6).0041

Look Once) and SSD (Single Shot MultiBox Detector). The two-stage detection model splits the object detection work into two processes, first judging potential regions where objects may exist, and then detecting these regions. In 2014, Ross Girshick proposed the Region-based CNN target detection algorithm, and successfully applied the convolutional neural network to the image recognition task[4]. Girshick proposed Fast R-CNN[5] in 2015 and Faster R-CNN[6] in 2017 to improve the two-stage detection, improving the detection efficiency and detection accuracy. Based on Faster R-CNN, Tsung-Yi Ling proposed an improved feature pyramid network algorithm, which further improved the model detection performance[7]. On this basis, Cascade R-CNN[8] and Mask R-CNN[9] have been proposed and widely used. Compared with the two-stage detection model, the single-stage detection model cancels the region extraction process and directly predicts in the input image, which has a very high detection speed. In 2015, Joseph Redmon proposed the first single-stage target detection algorithm YOLO[10], and then proposed YOLOv2[11] and YOLOv3[12], which achieved higher detection efficiency and detection accuracy. Liu proposed the SSD algorithm on the basis of the single-stage detection algorithm[13]. On the basis of ensuring the detection rate, the accuracy has also been improved. Among the target detection algorithms based on deep learning, the YOLOv3 algorithm has been widely used in the field of ship detection because of its advantages of detection speed and accuracy at the same time. In 2020, Chen introduced a visual attention mechanism to YOLOv3 for SAR image detection in complex backgrounds[14]. By improving the saliency map generation algorithm, the background weight of the original image was weakened, and the interference of complex backgrounds on SAR image target detection was reduced. . In 2021, Wu will focus on detection difficulties such as large changes in ship target scale, close arrangement of near-shore scenes, diverse angles and directions, and complex background environments. -means anchor box clustering, rotating target loss function definition and other related technologies, designed a ship target detection model R-YOLOv3 based on a rotating rectangular frame[15]. Aiming at the problem that traditional SAR image target detection is susceptible to interference, Chen proposed an improved YOLOv3 ship detection method[16]. By adding variable convolution, ResNet50 and ShuffleNetv2 and other module designs, the efficiency of SAR image ship detection in complex scenes has been improved.

The key to target detection in SAR images is to pay more attention to the detected targets such as ships, and ignore the interference of information such as sea clutter, islands and reefs, and coastal ports, which makes detection difficult. At this stage, the SAR ship detection algorithm based on the improved YOLOv3 still has many missed detections and false detections in complex environments, and the detection results are not ideal.

Based on the latest research results of YOLOv5, this paper improved the method. The main contributions of this paper include the following points. (1) A new feature pyramid network is proposed by using weighted cross-scale connections, which effectively improves the accuracy of ship target recognition in complex backgrounds. (2) The CBAM attention module is introduced to improve the feature expression ability and the feature learning ability of ship targets in complex backgrounds. (3) In view of the problem that the bounding box of ships will be lost and the average accuracy will be reduced in complex environments, the Soft-NMS linear penalty function is introduced. The experimental results show that the method not only has good accuracy, but also has a good detection rate in remote sensing target detection.

The rest of this paper is as follows. In Section 2, we introduced the prediction principle of YOLO and the framework of YOLOv5. In Section 3, we describe the improvement of our method in detail. Section 4 presents the experiments of the algorithm on the HRSID dataset and compares its performance with other classical algorithms. Lastly, the conclusion is shown in Section 5.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202206_8(6).0041

2. Related Work

2.1 The Principle of YOLO.

YOLO is a deep learning target detection algorithm based on regression method. Its core idea is to perform a one-time regression of the bounding box coordinates of the predicted target and the classification category probability in the output layer. The entire detection process is a single network, which can optimize the algorithm detection efficiency end-to-end.

YOLO innovatively proposes the idea of grid cell. The network first divides the input original image into S×S grids. If the coordinates of the center point of the detection target fall within a certain grid, then this The grid is responsible for detecting this target. Each grid regression B bounding boxes, each prediction box calculates 5 parameters, (x, y, w, h) and confidence. Where (x, y) is the relative value between the center point of the prediction frame and the grid boundary, w and h are the ratio of the width and height of the prediction frame to the entire image, and the confidence level includes the probability of the existence of the target to be tested and the target prediction boundary The accuracy of the box is two layers of information, where the accuracy of the target predicted bounding box is obtained by the intersection-over-union ratio (IOU) of the predicted box and the ground-truth box. The network implements an independent end-to-end network by regressing the location information and class probability of the target in the output image in one go.

YOLO can detect based on the entire image, make full use of the information on the entire image, so that the background false detection rate is low, and the detection of the location and category of the target is completed through a unified convolutional neural network, which saves a lot of time.

2.2 The Network of YOLOv5.

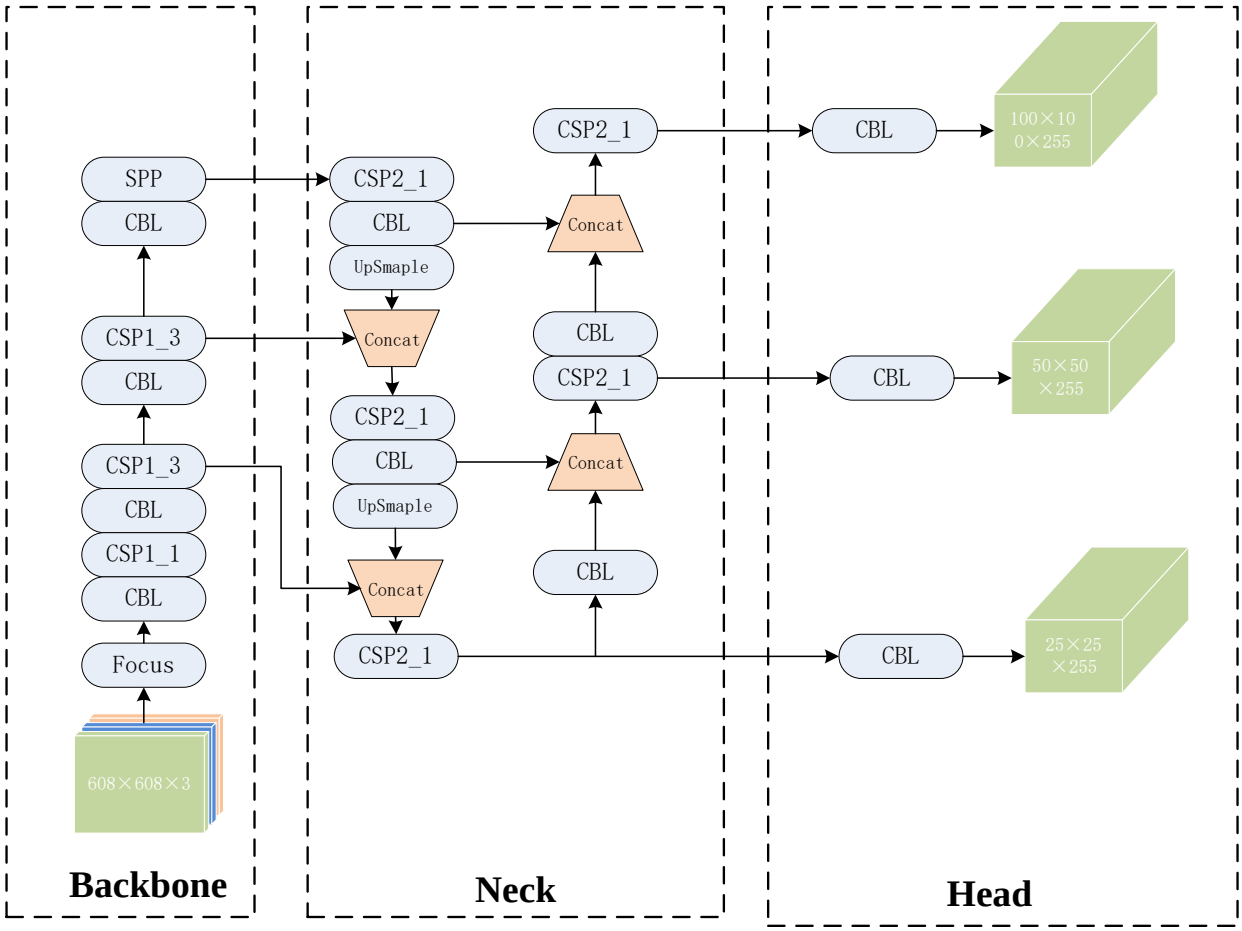


Fig. 1 Two or more references The architecture of the YOLOv5 method

ISSN: 2414-1895 DOI: 10.6919/ICJE.202206_8(6).0041

The YOLOv5 algorithm proposed by Glenn in June 2020 is a new generation algorithm of the YOLO series[17]. It has the advantages of small model size, fast detection speed, and performs well in target detection tasks. Its officially released code has 4 versions: For YOLOv5s, YOLOv5m, YOLOv5l, and YOLO5x, the model depth and feature map width gradually increase, and the target detection ability improves sequentially. In this paper, the YOLOv5 algorithm is used to detect ship targets in SAR images, and based on the original model, it is improved to improve the detection ability of targets in complex environments or small ship targets, and improve the overall detection effect of the network.

The overall network structure of the YOLOv5 algorithm includes three parts: Backbone, Neck, and Head, as shown in Fig. 1. Before being input to the Backbone part, YOLOv5 uses Mosaic for data augmentation. The Mosaic data enhancement method proposed for the first time in the YOLOv4[18] paper is different from the general distortion, flip and color gamut changes. Instead, the image is randomly cropped and randomly arranged and then stitched to increase the dimension of the dataset.

The Backbone part is the backbone network of YOLOv5 for feature extraction, mainly including Focus module, CBL module, CSP[19] structure and Spatial Pyramid Pooling (SPP) module[20]. The Focus module is a newly added part of YOLOv5. It divides the input data into 4 parts in the vertical and horizontal intervals, and performs convolution operation after splicing in the channel dimension to convert the width and height information of the image into channel information, increasing the receptive field. The loss of original feature information of the target is reduced, and the feature dimension of the image is deepened. For example, a 608×608×3 image is input into the Focus module of YOLOv5s, and the slice operation is used to convert from 3 channels to 12 channels, and then after a convolution operation of 32 convolution kernels, it finally becomes 304×304×32 Feature map, each piece of data is equivalent to twice down-sampling. The structure of the Focus module is shown in Fig. 1. The CBL module consists of convolution, batch normalization and activation function. The BN layer can improve the generalization ability of the network and prevent over-fitting. The mathematical expression of the activation function Leaky ReLU is:

$$LeakyRELU(x) = \begin{cases} x & ,if\ x \geq 0 \\ negative\_slope \times x & ,otherwise \end{cases} \quad (1)$$

It has all the advantages of the ReLu function and solves the problem of neuron death. YOLOv5 designs two different CSP structures, CSP1 and CSP2, where CSP1 is applied to the Backbone part and CSP2 is applied to the Neck part. The design of CSP1 refers to the residual structure, and its residual branch passes through the CBL, Bottleneck module, Conv and then performs the Concat operation with the branch that performs one convolution on the original data. The Bottleneck module is a classic residual structure. After the input passes through two layers of convolutional layers, the Add operation is performed with the original value to complete the residual feature transfer without increasing the output depth, reducing information loss and computation. The SPP module refers to the spatial pyramid pooling module, which uses the maximum pooling of k={1*1,5*5,9*9,13*13} for the input information, which makes the feature map smaller and simplifies the computational complexity of the network. And feature compression is performed by extracting the main features. Then perform Concat operation on the feature maps of different scales, and the output depth is the same as the input depth.

The Neck part is the feature fusion network of Yolov5, and its core is the Feature Pyramid Networks(FPN) and Path Aggregation Networks(PAN) structure[21]. From top to bottom, FPN transfers and fuses the category features of the high-level large objects to the low-level through upsampling. The PAN structure transfers the location features of the low-level large objects and the categories and locations of small objects from the features through downsampling to get. Feature map for making predictions. The FPN+PAN structure realizes the fusion and complementation of high-

ISSN: 2414-1895 DOI: 10.6919/ICJE.202206_8(6).0041

level features and low-level features, reduces the loss of low-level features, and improves the feature extraction capability of the model.

The Head part is the frame prediction structure of YOLOv5, including the prediction of boundary anchor boxes, loss function calculation and NMS non-maximum suppression. The traditional YOLO algorithm only inputs the highest layer into the Prediction part, and the network has the problem that small target features are lost after multi-layer transmission, which makes it difficult to identify. YOLOv5 inputs 3 features of different sizes into the detection layer, respectively targeting large, medium, and small-sized targets, reducing the loss of small target features and overcoming the limitation of only detecting the highest-level features. The loss function used by YOLOv5 consists of three parts, namely GIoU (Generalized Intersection over Union) loss function[22], target confidence loss function $Loss_{Obj}$ and classification loss function. GIoU is calculated by IoU (Intersection over Union), such as formula (2),

$$GIoU = IoU - \frac{|A_C - U|}{|A_C|} \tag{2}$$

Among them, A_C is the minimum enclosing rectangle area of the predicted box and the real box, and U is the area of A_C that does not belong to the two boxes. The target confidence loss function can reflect the change trend of the probability of the existence of the target in the prediction frame with the number of training steps, which is calculated by Focal loss, such as formula (3),

$$Loss_{Obj} = \begin{cases} -\alpha(1 - C_i^j)^\gamma log C_i^j, & Obj \epsilon pos \\ -(1-\alpha)(1 - C_i^j)^\gamma log(1 - C_i^j), & Obj \epsilon neg \end{cases} \tag{3}$$

Among them, Obj is the element in the prediction frame, α can balance the importance of positive and negative samples, γ is used to reduce the weight of easy-to-classify samples, pos is the target set, neg is the background set, and C_i^j corresponds to the i grid. The j prediction box contains the confidence of the target, and the calculation formula is as formula (4),

$$C_i^j = P_r \times IoU \tag{4}$$

Among them, P_r is the prior classification probability of the target. The idea of Non-maximum suppression is to search for and suppress local maxima. In the process of target detection, since each target will generate a certain number of candidate frames, when the target arrangement density is large in some areas of the image, it is necessary to use Non-maximum suppression reduces the impact of redundant bounding boxes on network parameter updates.

3. Proposed Method

This section mainly introduces the overall structure of the method proposed in this paper and several specific improvement measures, including bidirectional fusion feature pyramid network, the convolution block attention module and the Soft-NMS linear penalty function.

3.1 Improved FPN Network.

For the detection of small targets in complex environments, this paper proposes a bidirectional fusion feature pyramid network (BiFFPN) based on YOLOv5. The scheme of the proposed method is shown in Fig. 2, which can detect more abundant features. The scaled feature maps have better robustness when facing objects of different sizes, and improve the accuracy of object detection.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202206_8(6).0041

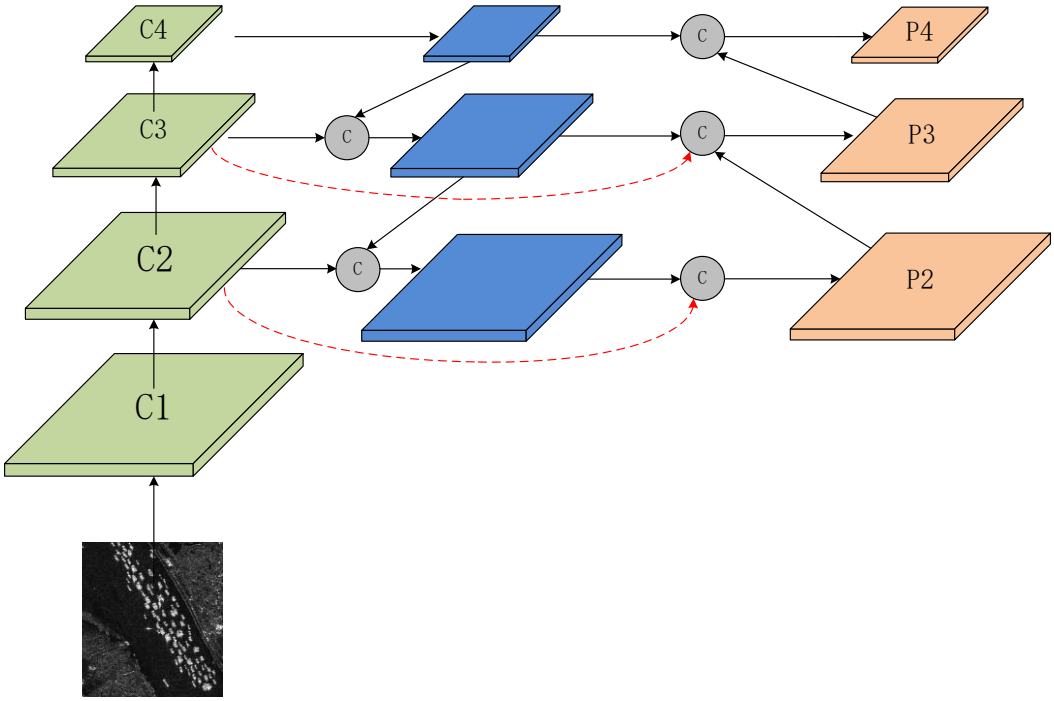


Fig. 2 The framework of the pyramid based approach

As shown in Fig. 2, the preprocessed image at the input end is sent to the backbone network for feature extraction, and then the obtained four different feature maps (C1, C2, C3, C4) are input to the feature fusion network. In order to more effectively fuse low-level information into high-level features, improve the target detection effect in complex backgrounds, and achieve multi-scale feature fusion, this paper designs a bidirectional fusion feature pyramid structure. Based on the idea of residual network, the structure adds cross-scale connections to achieve effective aggregation of multi-scale features, reduce the loss of small target feature information, and strengthen the feature extraction capability of the network. Based on the fused feature maps (P2, P3, P4), the bounding box (x, y, w, h), confidence (s), class (c) and angle class (Ac) of the target are predicted on the Head part. By iterating the loss function, the prediction results are continuously optimized.

As shown in Fig. 2 the preprocessed image at the input end is sent to the backbone network for feature extraction, and then the obtained four different feature maps (C1, C2, C3, C4) are input to the feature fusion network. In order to more effectively fuse low-level information into high-level features, improve the target detection effect in complex backgrounds, and achieve multi-scale feature fusion, this paper designs a bidirectional fusion feature pyramid structure. Based on the idea of residual network, the structure adds cross-scale connections to achieve effective aggregation of multi-scale features, reduce the loss of small target feature information, and strengthen the feature extraction capability of the network. Based on the fused feature maps (P2, P3, P4), the bounding box (x, y, w, h), confidence (s), class (c) and angle class (Ac) of the target are predicted on the Head part. By iterating the loss function, the prediction results are continuously optimized.

3.2 The Attention Mechanism.

The attention mechanism can make the neural network pay more attention to the feature information related to the ship target, while suppressing the complex background feature information, thereby improving the accuracy. This paper introduces the CBAM attention module on the BiFFPN-YOLOv5 algorithm, which can mix cross-channel and spatial information to extract informative features, as shown in Fig. 3.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202206_8(6).0041

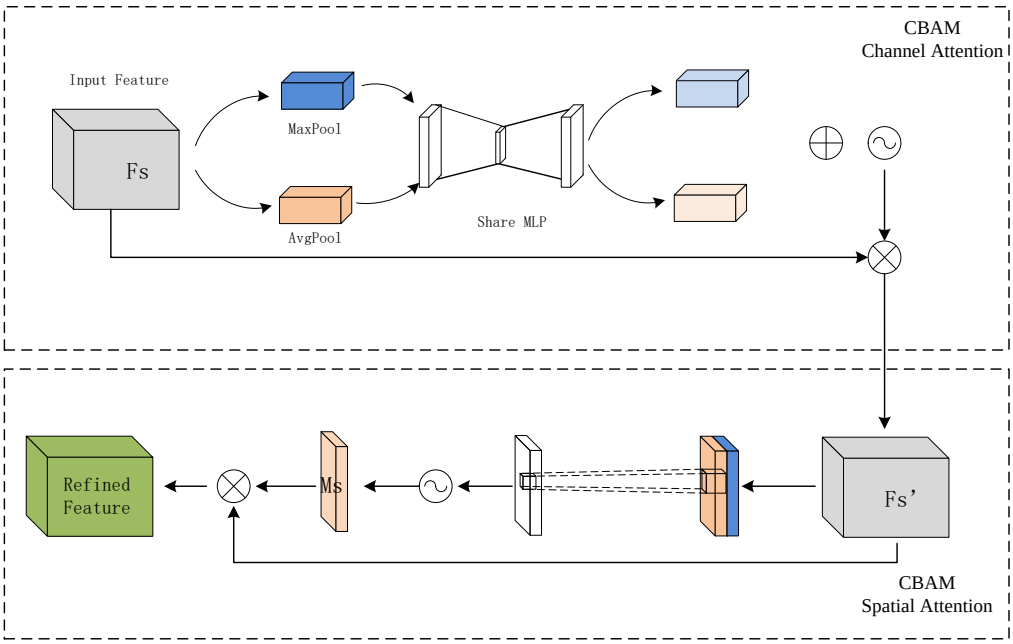


Fig. 3 The structure of the attention modules

CBAM sequentially applies the channel and spatial attention modules to realize the weight distribution of feature images in space and channel through machine learning[23], which promotes computing resources to be more inclined to focus on the target area, and achieves considerable performance while keeping the overhead small. Improve. The spatial attention module performs mean-pooling and max-pooling operations along the channel axis to aggregate the channel information of the feature maps, and concatenates the mean-pooling features and max-pooling features in the channels through standard convolutional layers to generate spatial attention feature maps. The channel attention module uses both average pooling and max pooling features to aggregate the spatial information of the feature map, which greatly improves the representation capability of the network, and then applies the shared network to generate average pooling features and max pooling features, and combines the output channels through element summation Attention feature vector.

3.3 The Soft-NMS Algorithm for Merging Bounding Boxes.

The frame prediction structure of YOLOv5 includes NMS non-maximum suppression, which is expressed as formula (5):

$$S_i = \begin{cases} S_i, & IOU(M, b_i) < u \\ 0, & IOU(M, b_i) \geq u \end{cases} \tag{5}$$

Among them, S_i is the detection score; M represents the maximum score of the detection frame; b_i represents the detection frame in the remaining detection frames; $IOU(M, b_i)$ is the IOU for calculating the two detection frames; u represents the IOU threshold. NMS retains detection boxes within its threshold, forcing the scores of detection boxes adjacent to the target box to zero, resulting in missed detections and target localization errors. This paper introduces the Soft-NMS algorithm[24] to replace the NMS algorithm. Soft-NMS considers both the score and the degree of coincidence, and sets a penalty item for the frame with a high degree of overlap with the detection frame with the highest score, to avoid the situation where the overlapping frame is deleted if it contains the target, resulting in missed detection, and the same target is not retained. Two similar detection frames, which are expressed as formula (6):

ISSN: 2414-1895 DOI: 10.6919/ICJE.202206_8(6).0041

$$S_i = \begin{cases} S_i, & IOU(M, b_i) < u \\ S_i \times (1 - IOU(M, b_i)), & IOU(M, b_i) \geq u \end{cases} \quad (6)$$

Among them, S_i is the detection score; M represents the maximum score of the detection frame; b_i represents the detection frame in the remaining detection frames; $IOU(M, b_i)$ is the IOU for calculating the two detection frames; u represents the IOU threshold. The Soft-NMS algorithm can solve the problem that the bounding box of the ship will be lost and the average accuracy will be reduced in the case of dense ships such as coastal ports.

3.4 The Soft-NMS Algorithm for Merging Bounding Boxes.

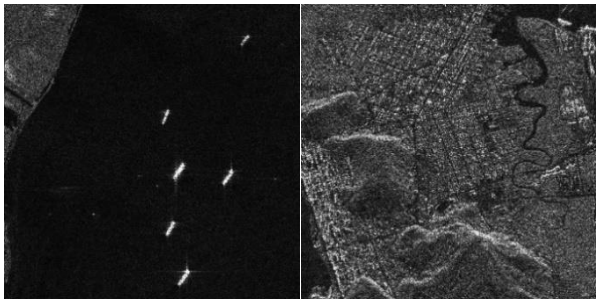
Compared with the original YOLOv5, the algorithm designed in this paper has three improvements. First, cross-scale connections are added to achieve effective aggregation of multi-scale features, reduce the loss of small target feature information, and strengthen the feature extraction capability of the network. Second, the CBAM attention module is added to the feature fusion layer. Third, the Soft-NMS algorithm is introduced to replace the NMS algorithm to reduce missed detections.

4. The Datasets and Results

In order to evaluate the improved YOLOv5 method, the experiment uses the deep learning framework Pytorch, the CPU selects the Core i7-7800K, and the GPU selects the Nvidia Geforce GTX1080Ti. In this section, we describe the dataset, experimental setup, and evaluation results. The hyperparameters of all experiments in this chapter are configured as follows: the initial learning rate of the network is 0.0001, the maximum number of iterations is 35,000, the learning rate decays at 17,500 and 26,500 iterations, the decay coefficient is 0.1, and the batch size of the model is set to 16 , the optimization algorithm adopts SGD, the optimization momentum parameter is 0.9, and the confidence threshold of soft-NMS is set to 0.45.

4.1 Experimental Training Dataset.

The dataset used in this paper is HRSID[25], a high-resolution SAR image dataset constructed by Wei et al. The dataset uses 136 panoramic SAR images from Sentinel-1B, TerraSAR-X and TanDEMX satellites, cropped and filtered out 5604 SAR images with a pixel size of 800 × 800, covering multiple polarization modes (co-polarization) , cross-polarized), with 0.5m, 1m and 3m resolutions. The dataset contains high-resolution SAR images of ships in various background environments including offshore, far-sea, and shore-based environments, with a total of 16,951 ship instances. They are marked with the MS COCO dataset's annotation format, and the SAR image information is recorded in JSON files. The annotation results are corrected using optical remote sensing images on Google Earth that are close to the time when the SAR image was taken, so they are more accurate. Preprocessing is performed before training, the image size is adjusted to 608 × 608 pixels, and the 5604 eligible images are randomly divided into training set and test set according to the ratio of 9:1.



ISSN: 2414-1895 DOI: 10.6919/ICJE.202206_8(6).0041

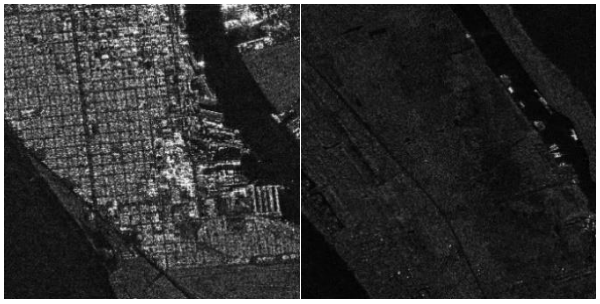


Fig. 4 Some images in the dataset

4.2 Evaluation Indicators.

The evaluation index adopts the general evaluation index in the field of target detection, mainly including the precision rate (Precision), the recall rate (Recall) and the average precision (AP) [26]. In this paper, the accuracy rate refers to the proportion of the bounding box predicted as a ship that is actually a ship, and the calculation formula is as shown in Equation (7). The ratio of the bounding boxes of all ships in the test set is calculated as formula (8):

$$P = \frac{TP}{TP+FP} \tag{7}$$

$$R = \frac{TP}{TP+FN} \tag{8}$$

In the formula, TP represents the number of ships that are correctly predicted as ships; FP represents the number of ships that are not ships and is correctly predicted to be non-ships; FN represents the number of ships But it was incorrectly predicted as the number of non-ships. But often the two cannot have both. The higher the recall rate, the lower the accuracy rate. It is not scientific to look at these two indicators separately. Therefore, AP is used as the final evaluation index of ship detection performance, and the calculation method is as follows:

$$AP = \int_0^1 P(R)dR \tag{9}$$

In the formula, P represents the precision rate, R represents the recall rate, and AP is the area under the P-R curve, which combines the two indices of precision rate and recall rate.

4.3 Results and Analysis.

Table 1. The comparison of experimental results

Numble Model	AP/%	Time(ms)
BiCBAM -YOLOv5	93.9	15.5
YOLOv5	92.9	12.3
Faster R-CNN	63.9	96
SSD	82.7	7
Improved YOLOv3[26]	87.9	4

To verify the impact of our improved method on ship recognition performance in SAR images, we conduct comparative experiments using several popular baseline methods: Faster R-CNN, SSD and YOLOv5, BiCBAM-YOLOv5. We first pre-trained the backbone of these models, and then trained and tested the four models using the same dataset. Meanwhile, to ensure training consistency, we set

ISSN: 2414-1895 DOI: 10.6919/ICJE.202206_8(6).0041

the hyperparameters and the number of training echoes for these three baseline models to be the same as BiCBAM-YOLOv5.

From the comparison of the experimental results, it can be seen that the algorithm in this paper has obtained the best results in the average accuracy of ship detection. In terms of detection time, compared with the traditional YOLOv5 algorithm, the algorithm in this paper requires more 3.2ms, but it still satisfies the Real-time detection features, faster detection speed, and the superiority of the algorithm in this paper is proved from the two standards of average accuracy and detection time.

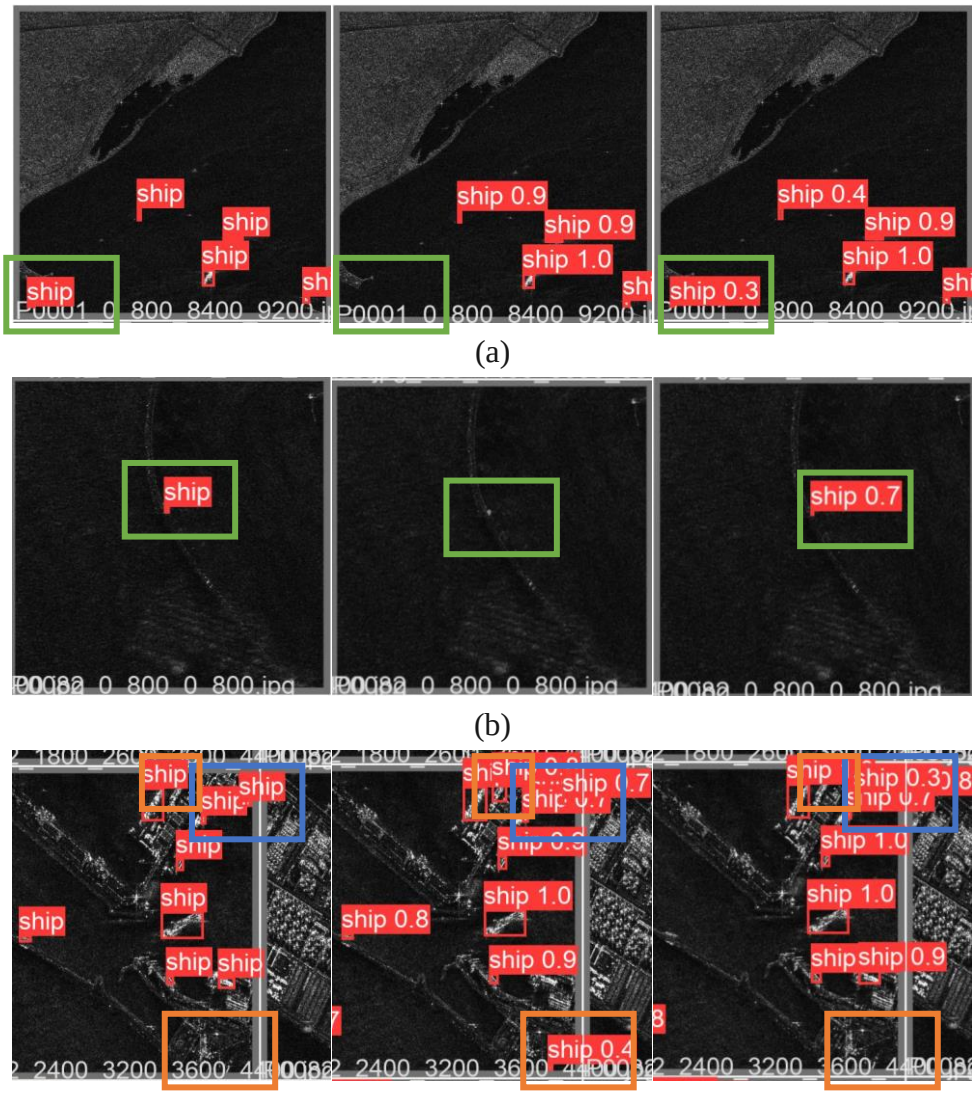


Fig. 5 Comparative experiments of BiCBAM-YOLOv5 and the traditional YOLOv5 for ship detection in nearshore (a) and ocean (b) and port (c).

Fig. 5 shows the detection results of the BiCBAM-YOLOv5 algorithm and the traditional YOLOv5 algorithm on the consent dataset respectively, where the first row is the real box, the second row is the prediction frame of the traditional YOLOv5 algorithm, and the third row is the prediction of the BiCBAM-YOLOv5 algorithm frame. Green boxes represent ships that were missed by the traditional YOLOv5 algorithm but not by the BiCBAM-YOLOv5 algorithm, orange boxes represent ships that were falsely reported by the traditional YOLOv5 algorithm but not by the BiCBAM-YOLOv5 algorithm, and blue boxes represent two All algorithms have falsely detected ships.From the ship target that can be on the same data set in Fig. 5, the prediction frame of the BiCBAM-YOLOv5

ISSN: 2414-1895 DOI: 10.6919/ICJE.202206_8(6).0041

algorithm is more in line with the real frame than the prediction frame of the traditional YOLOv5 algorithm, which further proves the detection effect of the improved algorithm, which can make the YOLOv5 algorithm better application on the SAR target detection dataset.

In Fig. 5 (c), the BiCBAM-YOLOv5 algorithm has false positives when detecting some ships in the port with small and dense targets, but the false positives are improved compared to the traditional YOLOv5 algorithm. The results show that the method still has shortcomings and needs to be further improved.

5. Conclusion

This paper proposes a ship target detection method based on improved YOLOv5 in SAR images. Firstly, the feature fusion network is improved, cross-layer connections are added, and the feature extraction ability of the network is improved, thereby improving the detection accuracy of the model. Secondly, the PSCBAM attention mechanism is introduced to improve the feature learning ability of the target area of the ship, so that the model pays attention to priority. The important information is suppressed and the environmental information is suppressed, and the target detection effect in the complex scene area is finally improved. Finally, the Soft-NMS linear penalty function is introduced to replace the NMS algorithm to reduce the missed detection rate and improve the detection accuracy. The experimental results show that the improved BiCBAM -YOLOv5 detection algorithm performs well on the HRSID dataset, with a AP of 93.9%, an increase of 1% compared with the traditional YOLOv5 algorithm, and a detection speed of 64.5FPS, demonstrating the effectiveness of the model.

References

[1] Zhang Y H , Ma H T , Yu Z X . Application of the method for prediction of the failure location and time based on monitoring of a slope using synthetic aperture radar[J]. Environmental Earth Sciences, 2021, 80(21).

[2] Pappas O , Achim A , Bull D . Superpixel-Level CFAR Detectors for Ship Detection in SAR Imagery[J]. IEEE Geoscience and Remote Sensing Letters, 2018.

[3] Cao D , Chen Z , Gao L. An improved object detection algorithm based on multi-scaled and deformable convolutional neural networks[J]. Human-centric Computing and Information Sciences, 2020, 10(1):14.

[4] Girshick R , Donahue J , Darrell T , et al. Region-Based Convolutional Networks for Accurate Object Detection and Segmentation[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015, 38(1):142-158.

[5] Girshick R. Fast R-CNN[J]. Computer Science, 2015.

[6] Ren S , He K , Girshick R , et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6): 1137-1149.

[7] Lin T Y , Dollar P , Girshick R , et al. Feature Pyramid Networks for Object Detection[J]. IEEE Computer Society, 2017.

[8] Cai Z , Vasconcelos N . Cascade R-CNN: Delving into High Quality Object Detection[J]. 2017.

[9] He K , Gkioxari G , P Dollár, et al. Mask R-CNN[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017.

[10] Redmon J , Divvala S , Girshick R , et al. You Only Look Once: Unified, Real-Time Object Detection[J]. IEEE, 2016.

[11] Redmon J , Farhadi A . YOLO9000: Better, Faster, Stronger[J]. IEEE, 2017:6517-6525.

[12] Redmon J , Farhadi A . YOLOv3: An Incremental Improvement[J]. arXiv e-prints, 2018.

[13] Liu W , Anguelov D , Erhan D , et al. SSD: Single Shot MultiBox Detector[J]. Springer, Cham, 2016.

[14] Chen Ququ. Research on SAR image target recognition method based on deep learning [D]. Hefei University of Technology, 2020. DOI: 10.27101/d.cnki.ghfgu.2020.000173. (In Chinese).

[15] Wu Keyi. Research on ship target detection method based on YOLO [D]. China University of Geosciences (Beijing), 2021. DOI: 10.27493/d.cnki.gzdzy.2021.000601. (In Chinese).

ISSN: 2414-1895 DOI: 10.6919/ICJE.202206_8(6).0041

[16] Chen Dong, Ju Yanwei. SAR image ship target detection based on semantic segmentation [J/OL]. Systems Engineering and Electronic Technology: 1-10 [2022-04-02]. http://kns. cnki.net/kcms/detail/11. 2422.TN.20210910.1450.008.html. (In Chinese).

[17] Song, Q.; Li, S.; Bai, Q.; Yang, J.; Zhang, X.; Li, Z.; Duan, Z. Object Detection Method for Grasping Robot Based on Improved YOLOv5. Micromachines 2021, 12, 1273. https://doi.org/10.3390/ mi1211127 3.

[18] Bochkovskiy A , Wang C Y , Liao H . YOLOv4: Optimal Speed and Accuracy of Object Detection[J]. 2020.

[19] Wang C Y , Liao H , Wu Y H , et al. CSPNet: A New Backbone that can Enhance Learning Capability of CNN[C]// 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE, 2020.

[20] He K , Zhang X , Ren S , et al. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 37(9):1904-16.

[21] Liu S , Qi L , Qin H , et al. Path Aggregation Network for Instance Segmentation[J]. IEEE, 2018.

[22] H Rezatofighi, Tsoi N , JY Gwak, et al. Generalized Intersection Over Union: A Metric and a Loss for Bounding Box Regression[C]// 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2019.

[23] Woo S , Park J , Lee J Y , et al. CBAM: Convolutional Block Attention Module[J]. Springer, Cham, 2018.

[24] Bodla N , Singh B , Chellappa R , et al. Improving Object Detection With One Line of Code[J]. 2017.

[25] Wei S ,Zeng X , Qu Q , et al. HRSID: A High-Resolution SAR Images Dataset for Ship Detection and Instance Segmentation[J]. IEEE Access, 2020, 8:1-1.

[26] Zhang J, et al. Improved YOLOv3 SAR image ship target detection [J]. Signal Processing, 2021, 37(9):10.