ISSN: 2472-3703 DOI: 10.6911/WSRJ.202205_8(5).0038

Maritime Small Ship Detection in Complex Ocean Environment Based on Improved Yolov3

Le Ye[1, a]

[1]Institute of Logistics Science and Engineering, Shanghai Maritime University, Shanghai, 201306, China

[a]732837590@qq.com

Abstract

Maritime small ship detection is a challenge problem in computer vision. At present, YOLOv3 network is widely used for object detection, but it gets low recall rate and detection accuracy for small objects in the complex ocean environment. Addressing this problem, we improve the backbone and predicted network of YOLOv3 network for detecting maritime small ship. Firstly, we build a maritime small ship dataset including four kinds of scenes: small traffic flow and heavy traffic flow in sunny and foggy weather. Secondly, we use K-means to re-cluster the anchor box for matching the shape of maritime ship. Thirdly, we introduce spatial pyramid pooling (SPP) module and frequency channel attention (FCA) module, and redesign the structure of YOLOv3 network, called it as SPP-FCA-YOLOv3. Here SPP module is used to fuse local features with global features and enriches the expression capability of the feature maps. FCA module emphasizes important object feature and suppresses unnecessary noise. Experimental results show that proposed SPP-FCA-YOLOv3 has higher detection accuracy for maritime small ship detection, getting a 2.2% improvement in average precision compared with YOLOv3, and a 1.2% improvement in average precision as well as higher speed compared with YOLOv5.

Keywords

Ship detection; Convolutional neural network; Frequency channel attention mechanism; Spatial pyramid pooling; YOLOv3.

1. INTRODUCTION

In recent years, maritime small ship detection has become a hot topic, and it is important to find suspicious ships at a long distance and early warning. Due to the complex and changeable marine environment, the traditional object detection methods easily cause false and missed detection, and cannot meet the requirements for safe navigation of the ship. Therefore, this paper focuses on the deep learning-based object detection method to solve these problems.

With the rapid development of deep learning (He et al., 2016; Huang et al., 2017), it has become a new direction to tackle the problem of ship detection (Lin et al., 2017; Liu et al., 2018). The existing deep learning-based object detection methods can be grouped into two categories according to whether to generate regional proposals or not: one-stage and two-stage detection methods. Two-stage detection methods have an advantage in detection accuracy but relatively time-consuming, such as R-CNN (Girshick et al., 2014), Fast R-CNN (Girshick, 2015), Faster R-CNN (Ren et al., 2015), while one-stage detection methods keep a balance between the detection accuracy and speed, such as SSD (Liu et al., 2016), YOLO series (Redmon et al., 2016; Redmon et al., 2017; Redmon et al., 2018), and RetinaNet (Lin et al., 2017).

YOLO series networks are popular in object detection, and have advantages in both accuracy and speed. However, in the face of complex ocean environment, the YOLO network cannot gain satisfying results for maritime small ship detection. Addressing this problem, we introduce the SPP and FCA module to propose a SPP-FCA-YOLOv3 network.

The main contributions of our paper are as follows:

(1) We build a maritime small ship dataset for object detection, including four kinds of scenes: small traffic flow in sunny weather, heavy traffic flow in sunny weather, small traffic flow in foggy weather and heavy traffic flow in foggy weather.

(2) We introduce the SPP and FCA module to improve the backbone and predicted network of the YOLOv3 network to gain SPP-FCA-YOLOv3. The SPP module fuses local features with global features to enrich the expression capability of the feature maps. The FCA module emphasize the difference between the ships and backgrounds to highlight the semantic information of ships.

(3) Our experimental results show that the proposed SPP-FCA-YOLOv3 achieves high accuracy and speed for maritime small-sized ship detection.

The remainder of this paper is organized as follows. The related work about YOLOv3 is described in Section 2, the framework and details of our proposed method are introduced in Section 3. The dataset implementation details and the evaluation protocol are shown in Section 4. The experimental results and analysis are discussed in Section 5. Finally, Section 6 concludes this paper.

2. RELATED WORK

2.1. YOLOv3 Network

As a representative network of the YOLO series, YOLOv3 (Redmon et al., 2018) has been recently received extensive attention, and adopts the idea of regression. For a given input image, YOLOv3 network divides it into three grids with different scales of 13*13, 26*26, 52*52, and directly returns the target boundary and target category of each grid in the prediction stage. During the feature extraction stage, YOLOv3 network is further improved on the YOLOv2-based network. Specifically, it proposes a more powerful Darknet53 backbone network based on Residual Network (ResNet) (He et al., 2016), which has better performance in feature extraction. The loss function of it is calculated as following:

$$L = L_{box} + L_{cls} + L_{obj} \tag{1}$$

$$L_{box} = \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^{B} 1_{i,j}^{obj} (2 - w_i \times h_i)[(x_i - \widehat{x_i})^2 + (y_i - \widehat{y_i})^2 + (w_i - \widehat{w_i})^2 + \left(h_i - \widehat{h_i}\right)^2] \tag{2}$$

$$L_{cls} = \lambda_{class} \sum_{i=0}^{S^2} \sum_{j=0}^{B} 1_{i,j}^{obj} \sum_{c \in classes} p_i(c) \log(\widehat{p_i}(c)) \tag{3}$$

$$L_{obj} = \lambda_{noobj} \sum_{i=0}^{S^2} \sum_{j=0}^{B} 1_{i,j}^{noobj} (c_i - \widehat{c_i})^2 + \lambda_{obj} \sum_{i=0}^{S^2} \sum_{j=0}^{B} 1_{i,j}^{obj} (c_i - \widehat{c_i})^2 \tag{4}$$

where L_{box}, L_{cls}, L_{obj} represents the loss of predicted box position regression, object classification and object confidence. S^2 represents the grid size, x and y represent horizontal and vertical coordinates respectively, w and h represent the width and high respectively. c indicates the predicted value of confidence, and $\hat{c}$ indicates the label value of c confidence.

YOLOv3 network is a representative of one-stage object detection network and still has many defects though it is simple to implement. YOLOv3 simultaneously predicts the location coordinates and category information of the object, which will lead to the problem of inaccurate object localization. In addition, YOLOv3 network has to predict a total of more than 10,000 possible prediction boxes at three prediction scales. Unfortunately, only a few parts of prediction boxes contain objects while most parts only contain background information, resulting in extreme imbalance in the number of objects and backgrounds.

2.2. Attention Mechanism

Recently attention mechanism has been widely incorporated into deep learning model. Wang et al. (2017) used the Residual Attention Network to train the neural network and achieved excellent results in image classification. In the task of image captioning, Chen et al. (2017) proposed a new convolutional neural network SCA-CNN, which incorporates spatial and channel-wise attention. Hu et al. (2018) focused on the channel relationship and proposed a squeeze-and-excitation (SE) block. The SE module has significantly improved the CNN network at the cost of slightly increasing the computing cost. Nie et al. (2020) integrated channel attention and spatial attention into Mask R-CNN for ship detection and segmentation. Qin et al. (2020) designed a frequency channel attention network (FcaNet) to compensate the deficiency of feature information in existing channel attention methods.

In this paper we consider the interference problem of maritime small ship in the complex ocean environment, and introduce FCA module to propose SPP-FCA-YOLOv3 network to better distinguish object from background.

3. 3 PROPOSED METHODS

In this section, we elaborate on the architecture of the proposed SPP-FCA-YOLOv3 network for maritime small ship detection. At first, for faster and more accurate, we adjust the shape of 9 anchor boxes to better match the shape of maritime ship. Then, we redesign the backbone network and prediction network of YOLOv3 network.

3.1. Anchor Box for Ship Detection

Anchor box is a few boxes of different sizes obtained by statistics or clustering from the ground truth in the training dataset, which can avoid blind searching during the training of the model and help the model to converge quickly. The original YOLOv3 network uses K-means to cluster all samples of the training dataset to obtain the width and height of representative shapes by using VOC and COCO datasets, which do not match the shape of maritime ship. So, we use K-means to re-cluster the anchor box to meet the requirements of ship detection.

Step 1, randomly selecting k of all ground truth boxes as the center of the cluster, here we set k to 9.

Step 2, calculating the distance between each ground truth box and each cluster. As illustrated in reference (Redmon et al., 2017), using Euclidean distance in k-means will make larger boxes and generate more error than smaller boxes, so we use distance formula as following (5):

$$d(ground\ truth\ box, anchor\ box) = 1 - IoU(ground\ truth\ box, anchor\ box) \tag{5}$$

Where IoU is the ratio of the intersection of the ground truth box and the bounding box to their union counterpart, which is expressed as (6):

$$\mathrm{IoU} = \frac{area(ground\ truth\ box)\ \cap area(bounding\ box)}{area(ground\ truth\ box)\ \cup area(bounding\ box)} \tag{6}$$

Then the bounding box is divided into the closest clusters.

Step 3, recalculating the cluster center according to the ground truth boxes in each cluster.

Step 4, repeating step 2 and step 3 until the elements in each cluster are no longer changed.

As shown in Figure 1, the 9 anchor boxes are gained, namely [3, 2], [5, 2], [14, 3], [9, 4], [19, 5], [18, 8], [30, 9], [39, 12], [92, 26] which can match the shape of different maritime ships. The new anchor boxes, as an effective size prior, make the proposed network converge faster and achieve better performance.

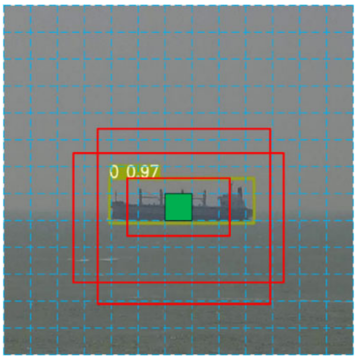


Figure 1. Different anchor boxes for ship detection

3.2. Structure of Proposed SPP-FCA-YOLOv3 Network

As shown in Figure 2, the proposed SPP - FCA -YOLOv3 network includes a backbone network and a prediction network. The backbone network is responsible for feature extraction, including 3 CBL blocks, 5 CBS blocks, 23 RES blocks, SPP module and FCA module. The prediction network consists of 8 CBL blocks stacked to form a feature pyramid structure.

We use 23 RES blocks and 5 CBS blocks alternating with each other to extract image features, where CBS block consists of convolution layer, batch normalization and Sigmoid activation function. RES block consists of 2 CBL blocks and a residual structure, while the CBL block consists of a convolutional layer, batch normalization and a Leaky ReLU activation function. The role of the CBS block is down-sampling, which enhances the learning ability of the network and preserves more information about small objects. After five times of down-sampling in the backbone network, the size of the feature map changes from 512*512*3 to 16*16*1024.

In the penultimate layer of the backbone network, we use the SPP module to replace the original convolutional layer, as shown in Fig.3 (d). The SPP module consists of three max-pooling layers, the stride is 1, and convolutional kernel sizes are 5*5, 9*9 and 13*13 in that order. The SPP module conveys a feature map of size 16*16*1024 to three maximum pooling layers, and finally fuses the output features of different scales to obtain a feature map of size 16*16*2048. That achieves the fusion of local features with global features and enriches the expression capability of the feature maps.

After the SPP module, we lead into a FCA module, which uses a two-dimensional discrete cosine transform (2D-DCT) to fuse multiple frequency-domain components, emphasizing the important target feature while suppressing unnecessary noise. In the prediction part, we adopt a feature pyramid structure, in which three bounding boxes are predicted in each grid of three prediction channels, and finally the detection results are obtained by filtering the bounding boxes with non-maximum suppression.

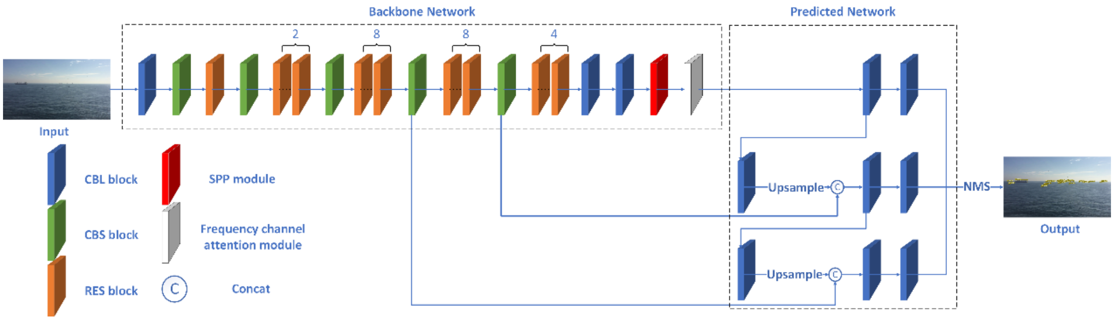
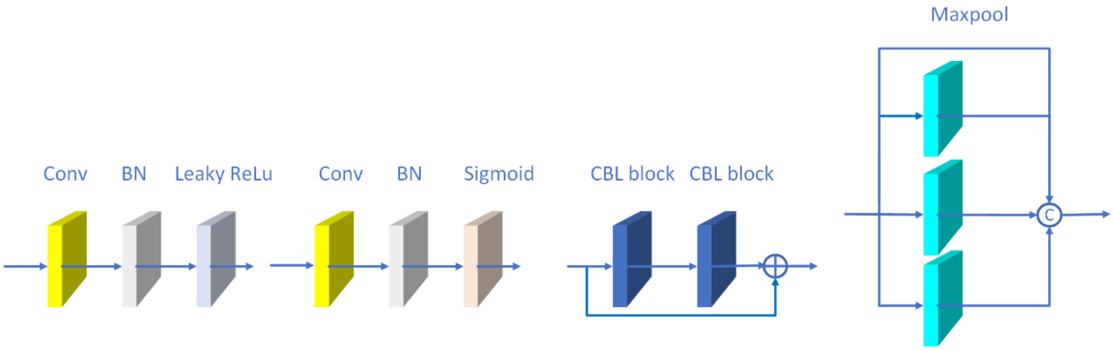


Figure 2. Structure of proposed SPP-FCA-YOLOv3 network



(a) CBL block (b) CBS block (c) RES bolck (d) SPP module

Figure 3. Block's structure

3.3. Frequency Channel Attention Mechanism

Channel attention modules are widely used in deep learning network, such as SENet (Hu et al., 2018), ECANet (Wang et al., 2020), CBAM (Woo et al., 2018), which usually use global average pooling (GAP) to get global information of each channel. GAP is a special case of two-dimensional discrete cosine transform (2D-DCT), whose result is proportional to the lowest frequency component of 2D-DCT, and loses a lot of frequency information. Although GAP is simple and efficient, it cannot well capture the rich information of input pattern. To solve this problem, we introduce a FCA (Qin et al., 2020) module using 2D-DCT into the proposed SPP-FCA-YOLOv3 network to extract different local information in the channel.

The preprocessing method GAP of channel attention uses inadequate information. So, we use 2D DCT to introduce more information to solve the problem. The FCA module (Qin et al., 2020) uses the 2D-DCT to replace the ordinary cosine transform. The two-dimensional discrete cosine transform and its inverse transform formula are as following (7) and (8):

$$f_{h,w}^{2d} = \sum_{i=0}^{H-1} \sum_{j=0}^{W-1} x_{i,j}^{2d} \cos(\frac{\pi h}{H}(i+\frac{1}{2})) \cos(\frac{\pi w}{W}(j+\frac{1}{2})) \tag{7}$$

$$x_{i,j}^{2d} = \sum_{h=0}^{H-1} \sum_{w=0}^{W-1} f_{w,h}^{2d} \cos(\frac{\pi h}{H}(i+\frac{1}{2})) \cos(\frac{\pi w}{W}(j+\frac{1}{2})) \tag{8}$$

where $f^{2d} \in R^{H\times W}$ represents the frequency spectrum of DCT, $x^{2d} \in R^{H\times W}$ represents the input of the feature map, H and W are the height and width of the feature map, $\cos(\frac{\pi h}{H}(i+\frac{1}{2}))\cos(\frac{\pi w}{W}(j+\frac{1}{2}))$ is the weights of discrete cosine transform weights.

The structure of the FCA module is shown in Figure 4. The input image features are evenly split into N equal parts. For each feature, the frequency importance of each channel is evaluated by using a two-step heuristic criterion. Then, the frequency component with the best performance is selected as the result of the output frequency feature. After splicing N output $F^0, F^1 \dots F^{n-1}$ features, the channel correlation is fitted through the fully connected layers, and then the weight is normalized by the sigmoid function. Finally, the channel features of each frequency are obtained by multiplying the original image features.

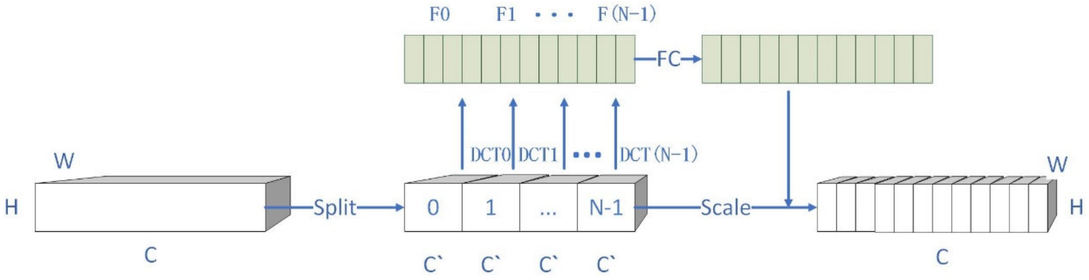
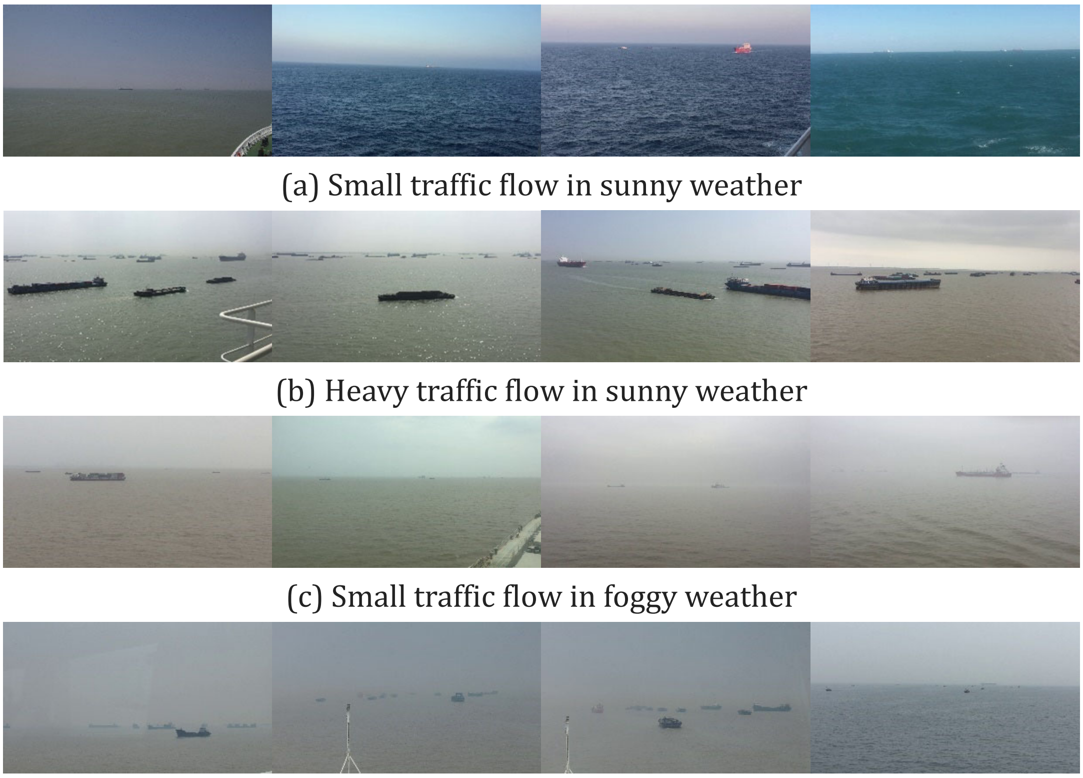


Figure 4. Structure of frequency channel attention module

4. EXPERIMENTS

4.1. Dataset



(a) Small traffic flow in sunny weather

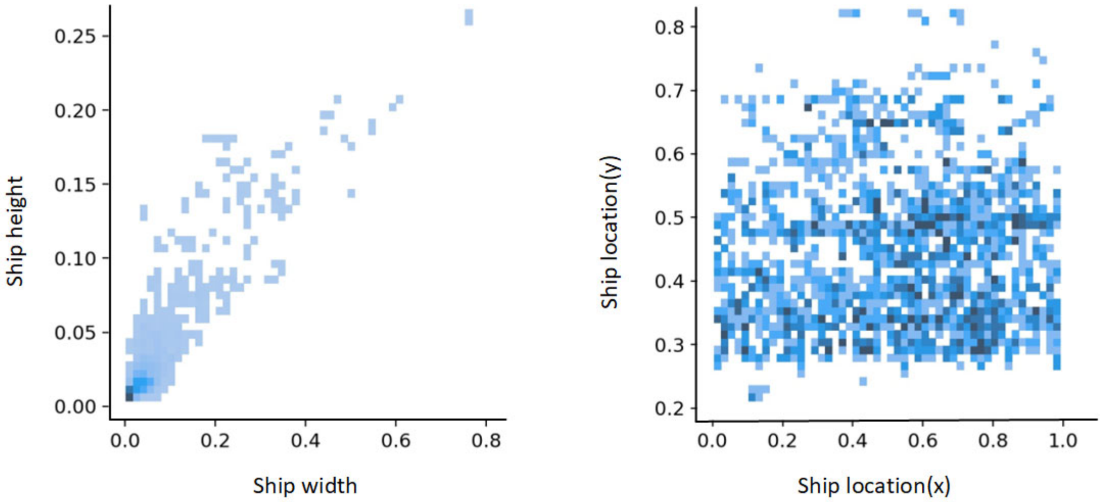
(b) Heavy traffic flow in sunny weather

(c) Small traffic flow in foggy weather

(d) Heavy traffic flow in foggy weather

Figure 5. Datasets for different scenes

Data greatly affects the performance of deep learning models. Here we construct a maritime small ship dataset with images from offshore acquisition, and the labels are manually labeled. This dataset contains four scenes: small traffic flow in sunny weather, heavy traffic flow in sunny weather, small traffic flow in foggy weather and heavy traffic flow in foggy weather. And several examples are shown in Figure 5. The resolution of the image is 1920*1080, and the label contains five information, including object category, normalized object coordinates X and Y, normalized object width W and height H respectively. The visualized analysis is shown in Figure 6. It can be seen from Figure 6 (a) that the size of the ships varies greatly, with the smallest ship being about 14*9 pixels and the largest ship being about 643*197 pixels. The distribution of ships in all images can be seen in Figure 6 (b), which shows the diversity of ships' positions in the images.



(a) Size distribution of ship (b) Location distribution of ship

Figure 6. Visualization distribution of ship. The coordinate values are normalized between 0 and 1. The change in color shades represents the change in the number of ships.

4.2. Experimental Results and Analysis

In our experiment, we randomly select 80% of dataset as training sets and the rest as test sets.

First, we use CBS block to replace the convolutional layer which plays a down-sampling role in the YOLOv3 network, and the results are listed at Table 1. It can be seen from Table 1 that the recall of the original YOLOv3 network was improved by 4% and the precision was improved by 3.9%. The AP@0.5 increases by 0.7% and AP@0.5:0.95 increases by 1.2%. The CBS block hardly increases the training cost, but reduces the loss of object features in the process of down-sampling.

Then, we replace the original convolution layer with the SPP module in the third reciprocal layer of the backbone network. That increases the parameters of the network and the training time, yet hardly affects the speed of ship detection in the test phase. Compared with only using CBS block, the AP@0.5 of our model is improved by 0.6%, the precision is improved by 5.6%, and the AP@0.5:0.95 is increased by 5.5%.

Finally, we add the FCA module after the SPP module. Compared with using CBS block and SPP module, the recall of our model is increased by 0.3%, and the precision is increased by 1.2%. This improvement is attributed to the fusion of multiple frequency components by the FCA module, the highlight of important information and the suppression of noise information.

Table 1. Experimental results about adding different modules

Methods	AP@0.5	AP@0.5:0.95	Precision	Recall	F1
YOLOv3	0.961	0.735	0.855	0.935	0.894
YOLOv3 +CBS	0.968	0.747	0.894	0.975	0.932
YOLOv3+CBS+SPP	0.974	0.802	0.950	0.963	0.956
SPP-FCA-YOLOv3	0.983	0.825	0.962	0.966	0.964

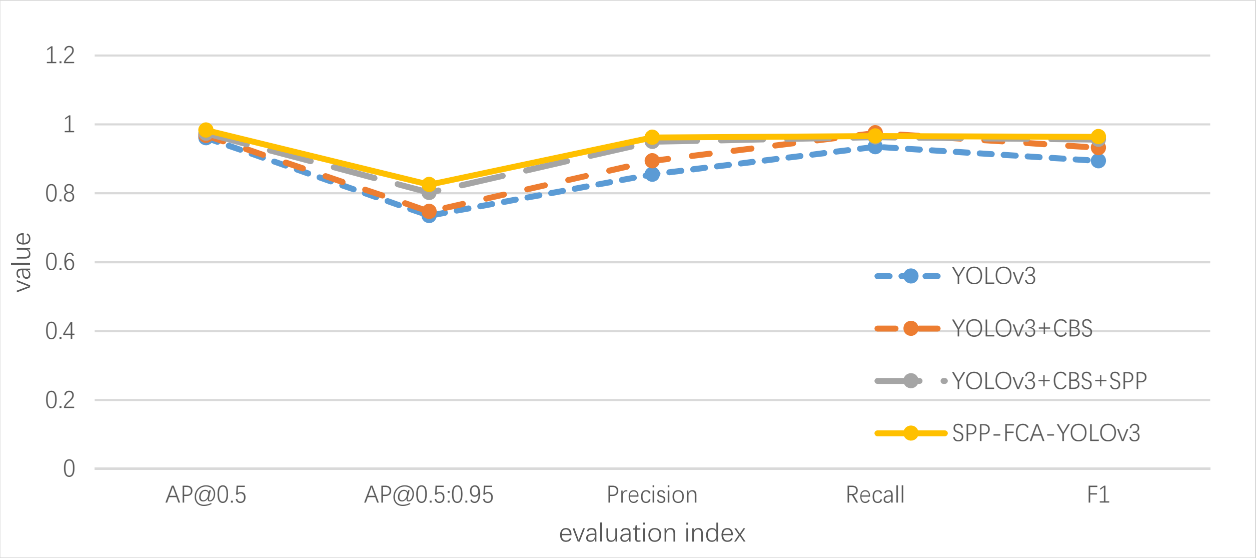


Figure 7. Performance comparison of ship detection about adding different modules

4.3. Compared with YOLOv3

Further, in order to compare the proposed SPP-FCA-YOLOv3 with the YOLOv3 network, we test them on four sub-datasets respectively. The results are shown in Tables 2, 3, 4 and 5. In Table 2, for small traffic flow in sunny weather, both the YOLOv3 network and the SPP-FCA-YOLOv3 network perform well. In Table 3, for heavy traffic flow in sunny weather, the SPP-FCA-YOLOv3 network shows a 1.7% improvement in recall. In Table 4, for small traffic flow in foggy weather, the SPP-FCA-YOLOv3 network showed a 2.0% improvement in precision, and a 1.1% improvement in F1. In Table 5, for heavy traffic flow in foggy weather, the recall of SPP-FCA-YOLOv3 network is improved by 0.6% and precision is increased by 0.2%, while overall AP@0.5 is still improved by 0.8%.

In summary, the SPP-FCA-YOLOv3 network shows better than YOLOv3 under the interference of foggy environment, which also suggests that our model is more robust in complex scenes.

Table 2. Detection results for small traffic flow in sunny weather

Methods	AP@0.5	AP@0.5:0.95	Precision	Recall	F1
YOLOv3	0.981	0.811	0.977	0.980	0.978
SPP-FCA-YOLOv3	0.985	0.834	0.984	0.981	0.982

Table 3. Detection results for heavy traffic flow in sunny weather

Methods	AP@0.5	AP@0.5:0.95	Precision	Recall	F1
YOLOv3	0.937	0.718	0.852	0.906	0.878
SPP-FCA-YOLOv3	0.945	0.751	0.898	0.923	0.910

Table 4. Detection results for small traffic flow in foggy weather

Methods	AP@0.5	AP@0.5:0.95	Precision	Recall	F1
YOLOv3	0.950	0.733	0.916	0.908	0.912
SPP-FCA-YOLOv3	0.963	0.741	0.936	0.911	0.923

Table 5. Detection results for heavy traffic flow in foggy weather

Methods	AP@0.5	AP@0.5:0.95	Precision	Recall	F1
YOLOv3	0.908	0.706	0.842	0.898	0.869
SPP-FCA-YOLOv3	0.916	0.724	0.844	0.904	0.873

4.4. Compared with Other Networks

In this section, we compare the SPP-FCA-YOLOv3 with some common object detection methods. The results are shown in Figure 8 and Table 6. As can be seen, the proposed SPP-FCA-YOLOv3 is higher than all other methods in terms of AP and recall, and slightly lower than the YOLOv5 network in detection precision, but faster than it. Compared with the YOLOv3 network, the SPP-FCA-YOLOv3 network decreases slightly in detection speed, but owns higher precision and recall. Meanwhile, over 2% improvement in the AP@0.5 metric and nearly 9% improvement in AP@0.5:0.95, as well as a 7% improvement in F1 are shown in Table 6.

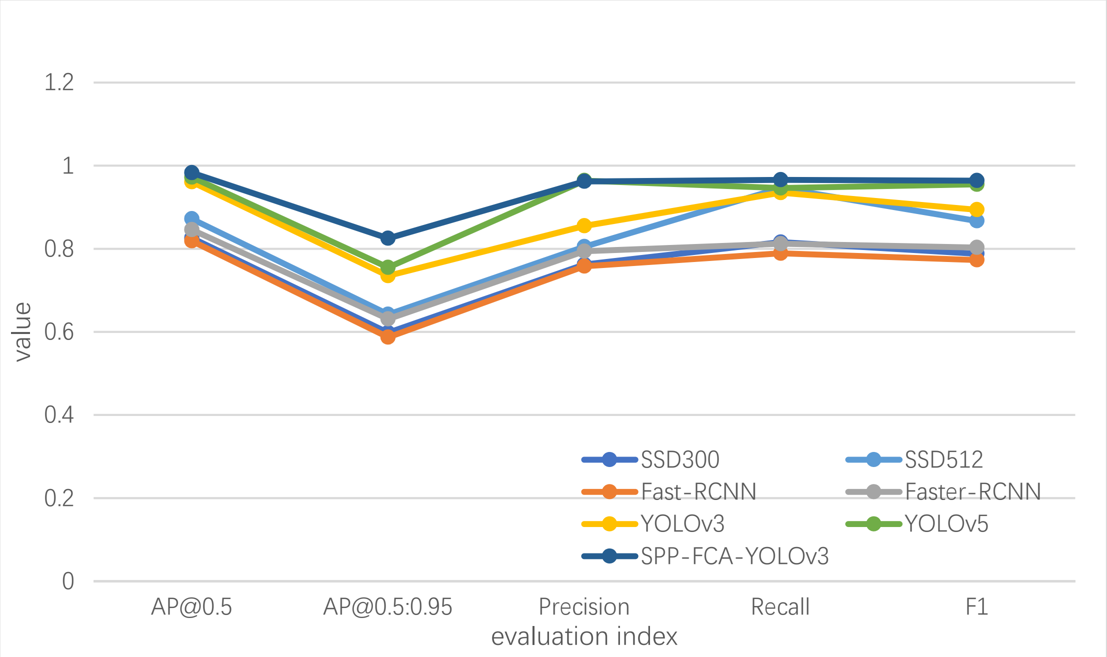


Figure 8. Performance comparison of ship detection in different methods

TABLE 6. Comparison of different methods

Methods	AP@0.5	AP@0.5:0.95	Precision	Recall	F1	FPS
SSD300	0.826	0.598	0.762	0.816	0.788	59
SSD512	0.872	0.642	0.805	0.946	0.867	28
Fast-RCNN	0.819	0.587	0.758	0.789	0.773	0.6
Faster-RCNN	0.846	0.631	0.794	0.812	0.803	8
YOLOv3	0.961	0.735	0.855	0.935	0.894	66
YOLOv5	0.972	0.755	0.964	0.946	0.955	53
SPP-FCA-YOLOv3	0.983	0.825	0.962	0.966	0.964	62

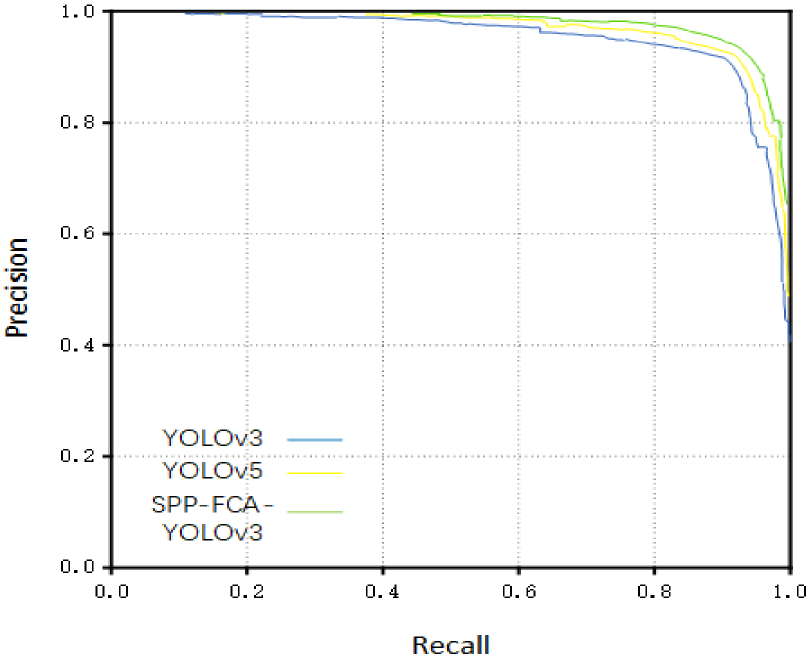
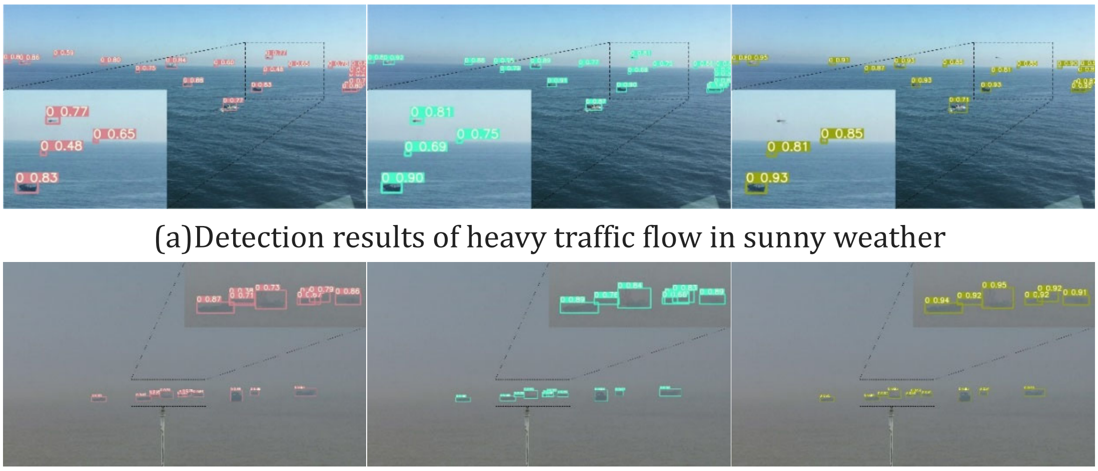


Figure 9. P-R curves of different methods

Figure 9 shows the P-R curves of the YOLOv3 network, the YOLOv5 network and the SPP-FCA-YOLOv3 network. It can be seen that the range of P-R curve corresponding to the SPP-FCA-YOLOv3 network is the largest, which means it has the highest AP value. Figure 10 gives the detection results of YOLOv3, YOLOv5 and SPP-FCA-YOLOv3 on different scenes. As shown in Figure 10 (a), on the heavy traffic flow in sunny weather, both the YOLOv3 network and the YOLOv5 network incorrectly detect the dragonfly in the image as a ship, while the proposed FCA-SPP-YOLO network avoids the false detection and moreover has a higher confidence score. As shown in Figure 10 (b), on the heavy traffic flow in foggy weather, two ships in the YOLOv3 network generate duplicate detections, and one ship in the YOLOv5 network generates duplicate detections. However, the FCA-SPP-YOLO network introduces the FCA module to depress the negative impact of foggy weather and does not generate duplicate detections. Figure 10 further demonstrates the superiority of the proposed SPP-FCA-YOLOv3 network.



(a)Detection results of heavy traffic flow in sunny weather

(b)Detection results of heavy traffic flow in foggy weather

Figure 10. The comparison results of different methods are as follows YOLOv3, YOLOv5, SPP-FCA-YOLOv3.

5. CONCLUSIONS

In this paper, we propose a novel network SPP-FCA-YOLOv3 for maritime small ship detection in the complex ocean environment. Firstly, the K-means method is used to re-cluster the anchor box to match the shape of the ship. Then, the CBS block is utilized to replace the convolutional layer which plays a down-sampling role for reducing the loss of small object features. Next, we use the SPP module to enrich the expression ability of the feature maps. Meanwhile, the FCA module is introduced to improve the detection performance of small object ships under the interference of different background environments. The experimental results demonstrate the efficiency of SPP-FCA-YOLOv3.

REFERENCES

[1] Chen C, Zhang Y, Lv Q, et al. 2019. Rrnet: A hybrid detector for object detection in drone-captured images. Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops.

[2] Chen L, Zhang H, Xiao J, et al. 2017. Sca-cnn: Spatial and channel-wise attention in convolutional networks for image captioning. Proceedings of the IEEE conference on computer vision and pattern recognition, 5659-5667.

[3] Deng C, Wang M, Liu L, et al. 2021. Extended feature pyramid network for small object detection. IEEE Transactions on Multimedia.

[4] Girshick R, Donahue J, Darrell T, et al. 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. Proceedings of the IEEE conference on computer vision and pattern recognition, 580-587.

[5] Girshick R. 2015. Fast r-cnn. Proceedings of the IEEE international conference on computer vision, 1440-1448.

[6] He K, Zhang X, Ren S, et al. 2016. Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition, 770-778.

[7] Hu J, Shen L, Sun G. 2018. Squeeze-and-excitation networks. Proceedings of the IEEE conference on computer vision and pattern recognition, 7132-7141.

[8] Huang G, Liu Z, Van Der Maaten L, et al. 2017. Densely connected convolutional networks. Proceedings of the IEEE conference on computer vision and pattern recognition, 4700-4708.

[9] Kisantal M, Wojna Z, Murawski J, et al. 2019. Augmentation for small object detection. arXiv preprint arXiv:1902.07296.

[10] Kong T, Yao A, Chen Y, et al. 2016. Hypernet: Towards accurate region proposal generation and joint object detection. Proceedings of the IEEE conference on computer vision and pattern recognition, 845-853.

[11] Li J, Liang X, Wei Y, et al. 2017. Perceptual generative adversarial networks for small object detection. Proceedings of the IEEE conference on computer vision and pattern recognition, 1222-1230.

[12] Lin H, Shi Z, Zou Z. 2017. Fully convolutional network with task partitioning for inshore ship detection in optical remote sensing images. IEEE Geoscience and Remote Sensing Letters, 14(10), 1665-1669.

[13] Lin T Y, Maire M, Belongie S, et al. 2014. Microsoft coco: Common objects in context. European conference on computer vision. Springer, Cham, 740-755.

[14] Lin T Y, Goyal P, Girshick R, et al. 2017. Focal loss for dense object detection. Proceedings of the IEEE international conference on computer vision, 2980-2988.

[15] Liu W, Ma L, Chen H. 2018. Arbitrary-oriented ship detection framework in optical remote-sensing images. IEEE geoscience and remote sensing letters, 15(6), 937-941.

[16] Liu W, Anguelov D, Erhan D, et al. 2016. Ssd: Single shot multibox detector. European conference on computer vision. Springer, Cham, 21-37.

[17] Nie X, Duan M, Ding H, et al. 2020. Attention mask R-CNN for ship detection and segmentation from remote sensing images. IEEE Access, 8, 9325-9334.

[18] Qin Z, Zhang P, Wu F, et al. 2020. Fcanet: Frequency channel attention networks. arXiv preprint arXiv:2012.11879.

[19] Redmon J, Divvala S, Girshick R, et al. 2016. You only look once: Unified, real-time object detection. Proceedings of the IEEE conference on computer vision and pattern recognition, 779-788.

[20] Redmon J, Farhadi A. 2017. YOLO9000: better, faster, stronger. Proceedings of the IEEE conference on computer vision and pattern recognition, 7263-7271.

[21] Redmon J, Farhadi A. 2018. YOLOv3: An incremental improvement. arXiv preprint arXiv:1804.02767.

[22] Wang F, Jiang M, Qian C, et al. 2017. Residual attention network for image classification. Proceedings of the IEEE conference on computer vision and pattern recognition, 3156-3164.

[23] Wang Q, Wu B, Zhu P, et al. 2020. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. Conference on Computer Vision and Pattern Recognition (CVPR). IEEE.

[24] Woo S, Park J, Lee J Y, et al. 2018. Cbam: Convolutional block attention module. Proceedings of the European conference on computer vision (ECCV), 3-19.

[25] Zoph B, Cubuk E D, Ghiasi G, et al. 2020. Learning data augmentation strategies for object detection. European Conference on Computer Vision, Springer, Cham, 566-58.