Ship Detection: An Improved YOLOv3 Method

Haiying Cui, Yang Yang, Mingyong Liu, Tingchao Shi, Qian Qi

***Abstract*—YOLOv3 is the state of art detector, which performs an excellent balance in detection speed and accuracy. In this paper, an improved YOLOv3 model named YOLOv3-ship is proposed for the ship detection. The main contributions to the YOLOv3-ship consists of dimension Clusters, network Improvement and embedding of the Squeeze-and-Excitation(SE) module. The experiments results show that the detection accuracy has been significantly improved by the YOLOv3-ship.**

I. Introduction

Ship detection has great demands in civil and military fields. For example, in the civil field, ship detection can supervise transportation, marine traffics and illegal smuggling. In the military field, one can be monitored for cross-border or other illegal behaviors. However, traditional ship detection is based on naked eyes monitoring, which causes huge labor costs.

Computer-aided detection method greatly saves the labor cost and improves detection efficiency at the same time. In recent years, the computer-aided detection algorithms have been made breakthroughs both in speed and accuracy, which are mainly divided into two branches: one-stage methods and two-stage methods. The one-stage methods including YOLO [1]–[3], SSD, use a single network to performs all predictions based on the actual image, which has faster detection speed. The two-stage methods including Fast RCNN [6], Faster RCNN [7] divides the detection process into two steps, which use a network to locate and classify the target roughly and then use another network to correct the target precisely. The two-stage methods have an advantage in detection accuracy.

Yolov3 [1] is the latest version of the YOLO series. It draws on the idea of Resnet [8]and establishes a Darknet53 ConvNet as a feature extractor to improve the ability to extract features. Besides, it improves the detection defect of YOLO v1,v2 [2], [3] in small object by using multi-scale prediction method. Due to its high detection speed and accuracy, Yolov3 has achieved the a state of art level. In this paper, an improved YOLOv3 model named YOLOv3-ship is proposed for the ship detection.

The main contributions of this paper can be listed as follows:

- Determine the anchor settings for the ship dataset by kmeans++ algorithm.
- Design a convolutional neural network named Darknet-ship to solve the problem of excessive YOLOv3 parameters.
- Embed the Squeeze-and-Excitation module in YOLOv3 to increase the network's ability to extract global features.

This paper is organized as follows: the proposed improvement methods are given in the next section. Section III introduces the datasets and the experiments results. Finally, a conclusion of the proposed work and further developments are given in section IV.

II. THE IMPROVED METHOD

A. Dimension Clusters

YOLOv3 introduces anchors, a set of initial candidate boxes with fixed width and height. The settings of anchors boxes affect the detection accuracy and speed. Kmeans algorithm is selected to conduct dimension clusters in YOLOv3. However, the kmeans algorithm is sensitive to the initial points. Therefore, an improved clusters algorithm named kmeans++ is introduced to solve this problem. The distance function of K-means++ algorithm is defined as:

$$d(a,b) = 1 - IOU(a,b) \tag{1}$$

where a is the size of rectangular box, b is centroid of the rectangular box. The IOU function represents the overlapping ratio of two rectangular boxes, as shown in Fig. 1.

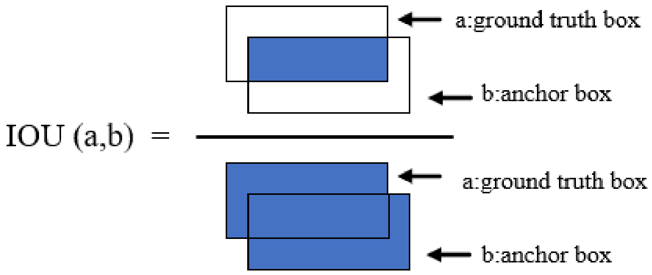


Fig. 1: The description of IOU function

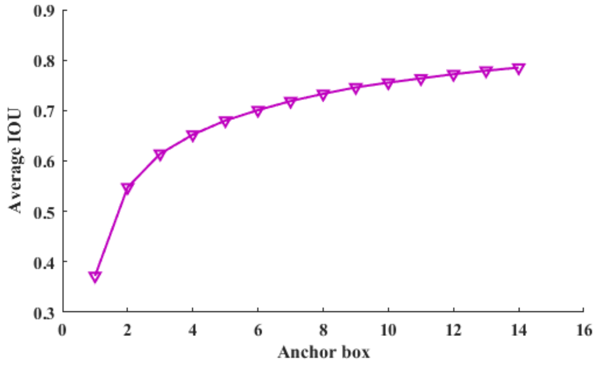


Fig. 2: The relationship between the number of anchor boxes and average IOU

Haiying Cui, Yang Yang, Mingyong Liu, Qian Qi are with the School of Marine Engineering, Northwestern Polytechnical University, Xi'an, China.
cuihaiying@nwpu.edu.cn; y_yang@126.com;
liumingyong@nwpu.edu.cn; shi_tingchao@163.com;
qiqian@nwpu.edu.cn;

978-1-7281-1450-7/19/$31.00 ©2019 IEEE

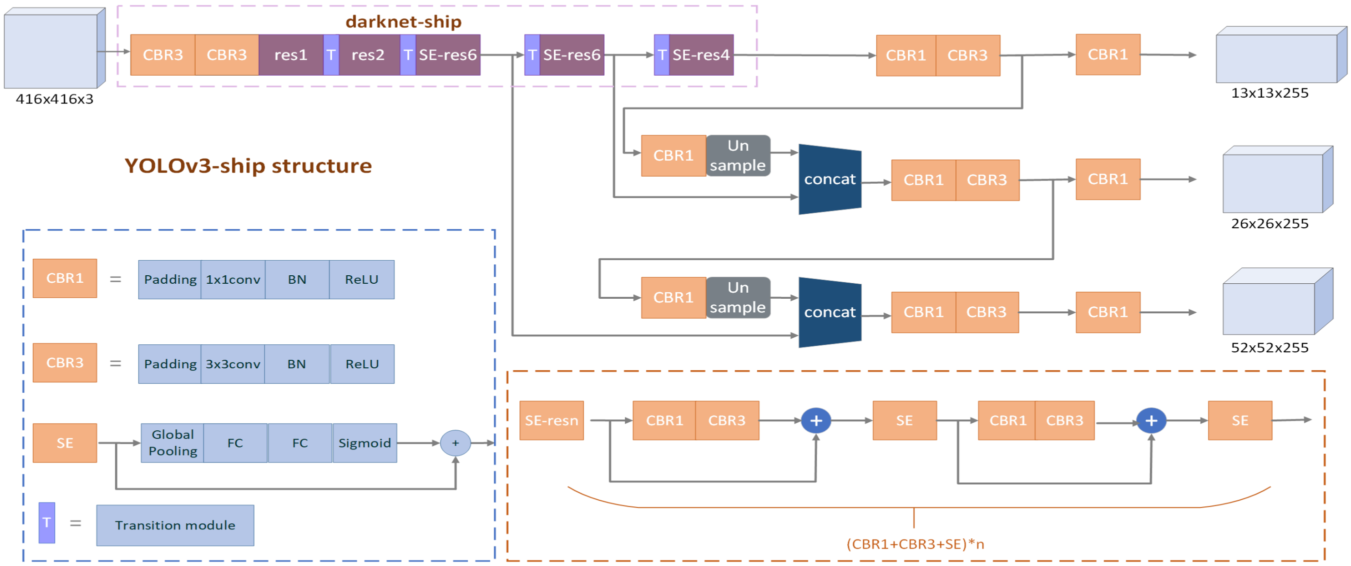


Fig. 3: The description of YOLOv3-ship structure

To determine the settings of anchor box, the relationship between average IOU and anchor box is depicted in Fig. 2. According to the inflection point method, we selects six clusters and divide up the six clusters on three scales. the corresponding sizes of six clusters are: (31, 15), (65, 26), (115, 42), (156, 28), (221, 55), (304, 104).

B. Network Structure Improvement

Table 1 The network structure of Darknet-ship.

Type	Filters	Size	Output
Conv	32	3x3 conv stride=2	208x208
Residual block(1)	32	3x3 conv stride=1	208x208
x1	64	3x3 conv stride=1	
Transition module	32	1x1 conv stride=1	104x104
	64	3x3 conv stride=2	
Residual block(2)	64	1x1 conv stride=1	104x104
x2	128	3x3 conv stride=1	
Transition module	64	1x1 conv stride=1	52x52
	128	3x3 conv stride=2	
Residual block(3)	128	1x1 conv stride=1	52x52
x6	256	3x3 conv stride=1	
Transition module	128	1x1 conv stride=1	26x26
	256	3x3 conv stride=2	
Residual block(4)	256	1x1 conv stride=1	26x26
x6	512	3x3 conv stride=1	
Transition module	256	1x1 conv stride=1	13x13
	512	3x3 conv stride=2	
Residual block(5)	256	1x1 conv stride=1	13x13
x4	512	3x3 conv stride=1	

Yolov3 establishes a Darknet53 ConvNet as a feature extractor. However, it seems too complex and redundant for ship detection, which may lead to more complex training and slower detection speed. Based on Darknet-53, a ConvNet named Darknet-ship is designed to reduce the parameters and improve the performance of network to extract features. In Darknet-ship, the 1*1 convolution kernels are used to reduce the parameters in the transition module, as defined in Table 1.

C. Embedding of the Squeeze-and-Excitation Module

Squeeze-and-Excitation(SE) module is a ConvNet structure proposed by J. Hu in 2017 [4], which won the championship of the last imagenet classification competition. The SE module improves the expressive ability of network by accurately modeling the interaction between channels of convolution features.

In the last three resnet blocks in Darknet-ship, the SE module is introduced to increase the receptive field and enhance the ability of network to extract global information. The SE module and the YOLOv3-ship structure are shown in Fig. 3.

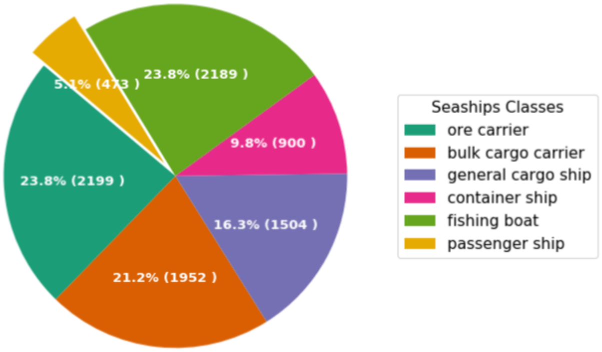
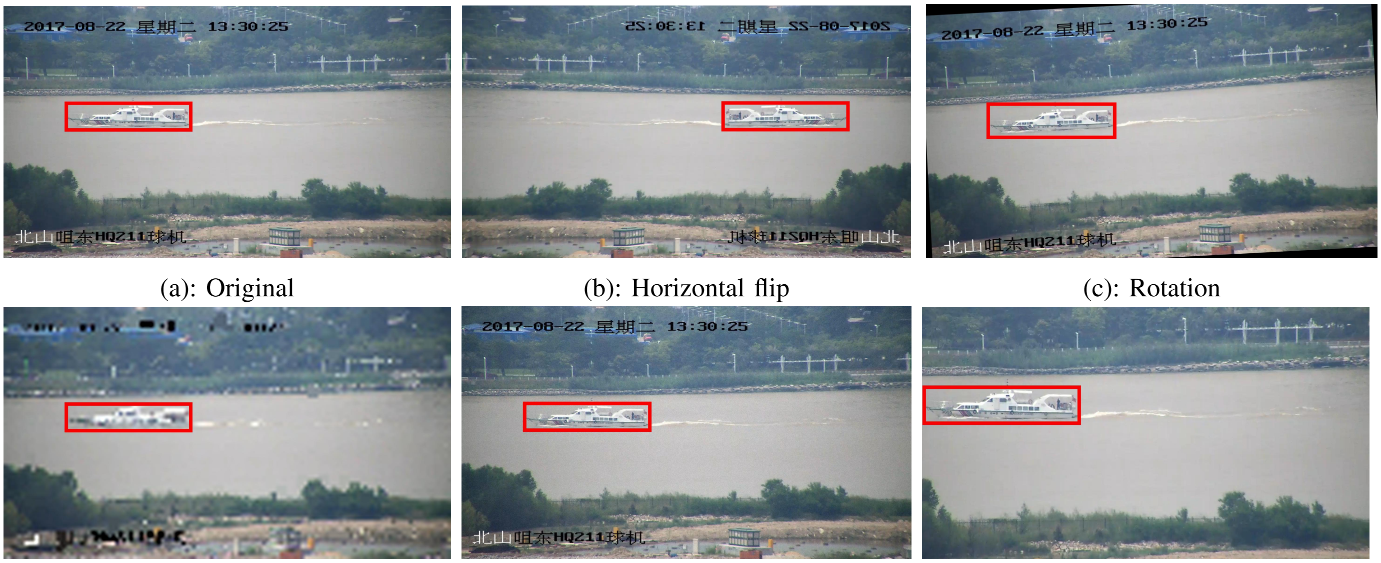


Fig. 4: The numbers of bounding boxes for each class



(a): Original (b): Horizontal flip (c): Rotation

(d): Super pixel (e): Gaussian noise (f): Scale

Fig. 5: Some samples of the augmentation methods

TABLE I: Experiment results

Model	MAP	AP(ore carrier)	AP(bulk cargo carrier)	AP(general cargo ship)	AP(container ship)	AP(fishing boat)	AP(passenger ship)	FPS
YOLOv3	0.87	0.86	0.86	0.87	0.87	0.88	0.90	33
YOLOv3-ship	0.91	0.90	0.91	0.90	0.94	0.90	0.91	31

III. EXPERIMENTS

A. Datasets Processing

The experiment dataset comes from a new large-scale ship dataset called SeaShips [5]. The dataset consists of 31455 images(only 7000 offered) in 6 classes, including ore carrier, bulk cargo carrier, general cargo ship, container ship, fishing boat, and passenger ship. In order to avoid the problem of class imbalance, a statistical analysis of the bounding boxes ia made for each class in the dataset. As is depicted is the Fig. 4, the class of passenger ship shares the smallest proportion. Therefore, some augmentation methods including horizontal flip, rotation, super pixel, gaussian noise, scale are applied to the the bounding boxes of the passenger ship. The number of bounding boxes are increased by using one or more of the following methods for the bounding boxes of the passenger ship, the augmentation results are shown in Fig. 5.

B. Experiment Settings

The experiments have been performed on a high performance platform equipped with Nvidia GeForce GTX 1080Ti GPU, using Ubuntu OS with Core i7 7700K CPU and a RAM of 32 GB. The images are resized to 416x416 before being sent to the network. The initial learning rate is 0.0001. We divide it by 10 when training at 20000 and 30000 steps. The momentum is 0.90, the weight decay is 0.0005, and a total of 40000 steps are trained.

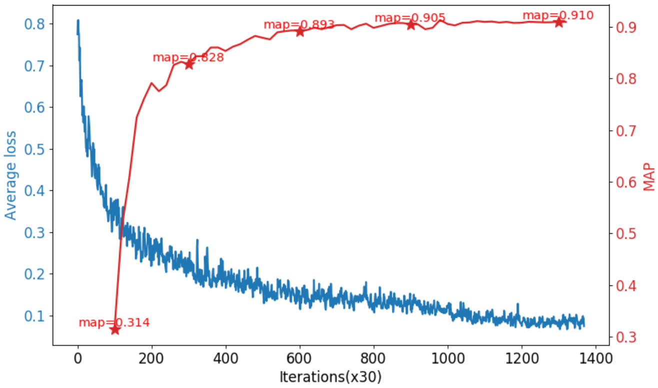
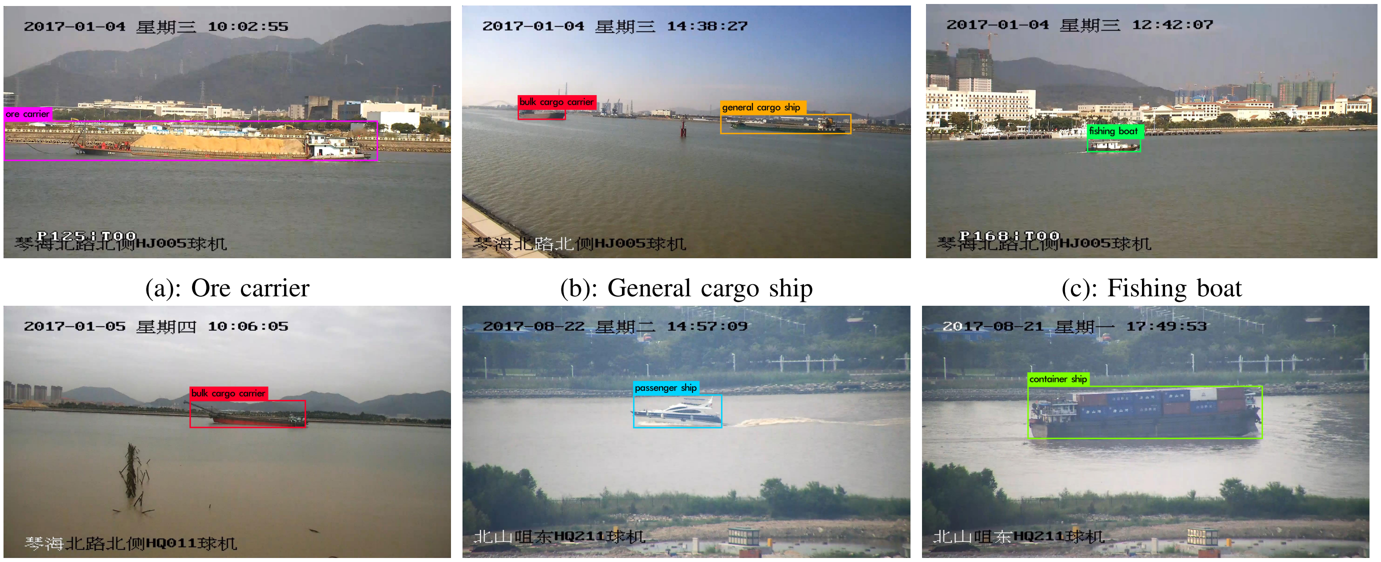


Fig. 6: The loss and map versus the iterations

C. Experiment Results

We compares YOLOv3-ship with YOLOv3 to show the differences between models. The map(mean average precision) and fps(frame per second) are selected as the indicators to evaluate the performances between two models. Experiment results are listed in Table 2. It shows that the YOLOv3-ship has higher map values and same fps values, which proves that YOLOv3-ship improves the performance of the YOLOv3 by a large margin without speed drops. Comparing the average precision of each class for the two models, a conclusion can be drawn that the YOLOv3-ship improves the detection accuracy of YOLOv3 for large and medium objects, which



(a): Ore carrier (b): General cargo ship (c): Fishing boat

(d): Bulk cargo carrier (e): Passenger ship (f): Container ship

Fig. 7: Some detection results on the SeaShips based on the YOLOv3-ship

may be attributed to the modelling effects of SE module on the channels of salient objects.

Fig. 6 shows the changing trends of average loss and map during the training process. It can be found that the average loss is in a steady state of decline and eventually converges. The map maintains a high value after the middle of the iterations. In order to avoid the limitation of analysis caused by the selection of weights, the changing trends of average precision for each class is descripted in Fig. 8. Compared with other classes, we can find that the class of passenger ship converges slowly at the beginning and badly at the end, which proves that the YOLOv3-ship improves poorly comparing with the YOLOv3 in the detections of the small objects.

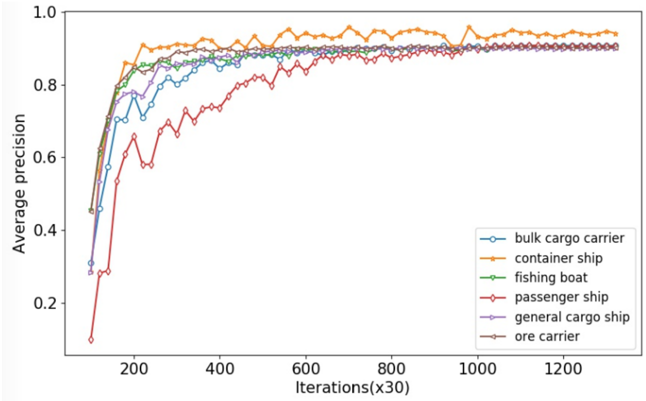


Fig. 8: The average precision for each class versus the iterations

Finally, Some detection results for each class based on the YOLOv3-ship are depicted in Fig. 7.

IV. CONCLUSIONS AND FUTURE WORKS

In this paper, an improved YOLOv3 model named YOLOv3-ship is proposed, which greatly improves the detection accuracy, especially on the large and medium-sized objects. In the future, some tricks will be attempted to improve the detection accuracy for small objects.

V. ACKNOWLEDGMENT

This paper is supported by the National Natural Science Foundation of China (NSFC, Project No.:51679201).

REFERENCES

[1] J. Redmon, and A. Farhadi, “YOLOv3: An Incremental Improvement,” Available at: https://pjreddie.com/media/files/papers/YOLOv3.pdf (Accessed 06.04.18)

[2] J. Redmon, A. Farhadi, “YOLO9000: Better, Faster, Stronger.” *Computer Vision and Pattern Recognition*, pp. 7263C7271, 2017.

[3] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, “You only look once: Unified, real-time object detection.” *IEEE conference on computer vision and pattern recognition*, pp. 779C788., 2016.

[4] J. Hu, L. Shen, and G. Sun, “Squeeze-and-excitation networks,” arXiv preprint, arXiv:1709.01507 (2017)

[5] Shao et al, “SeaShips: A Large-Scale Precisely Annotated Dataset for Ship Detection.” *IEEE TRANSACTIONS ON MULTIMEDIA*, vol. 20, no. 10, 2018.

[6] R. Girshick, “Fast R-CNN.” *IEEE International Conference, Computer Vision*, pp. 1440C1448, 2015.

[7] S.Ren, K.He, R.Girshick, et al, “Faster R-CNN: towards real-time object detection with region proposal networks.” *International Conference on Neural Information Processing Systems*, pp. 91-99., 2015.

[8] K. He, X. Zhang, S. Ren, J. Sun, “Deep Residual Learning for Image Recognition.” *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770-778, 2016.