ISSN: 2414-1895 DOI: 10.6919/ICJE.202204_8(4).0023

An Efficient Ship Detection and Classification Algorithm based on YOLOv4

Weina Zhou[a], Lu Liu[b]

Shanghai Maritime University, School of Information Engineering, Shanghai, China

[a]wnzhou@shmtu.edu.cn, [b]2608508033@qq.com

Abstract

Ship detection and classification has been paid more and more attention in marine environment, since it plays an important role in marine observation, port monitoring and marine management. In order to detect and classify ships in a more accurate and faster manner, this paper proposes an improved target detection algorithm based on YOLOv4. This algorithm integrates a new designed Multi-layer Feature Fusion (MFF) module and a Multi-layer Receptive Field Block(M-RFB) module into the neck of YOLOv4. They could fully fuse the context information of the neural network and reduce the loss of feature information. The experimental results show that, the proposed algorithm can effectively solve the problem of difficult detection and low recognition rate of small ships in complex marine environment. Compared with the state of art, this algorithm could obtain higher accuracy under the complex ocean navigation conditions. And it improves the accuracy rate by about 11% relative to the original YOLOv4, while still achieving a good balance between detection accuracy and speed.

Keywords

Maritime Target Detection; Feature Fusion Module; YOLOv4.

1. Introduction

Visual target detection in maritime environments is one of important research topics in the field of computer vision. An accurate and rapid ship detection method could not only provide very important information for terminal management, port monitoring and safe navigation, but also play an important role in smuggling ships detection and marine rescuing.

In the past, many researchers have contributed and provided different methods to detect objects on the water surface. Huang et al. designed an edge detection method based on structured forest, which only demands a small training set and then obtains ship proposals by connected domain detection [1]. Zhu et al. proposed a ship detection system by manually extracting target features based on edges, corners, and color information, which can accurately detect ships and meet the requirements of real-time processing [2]. Li et al. developed an inshore ship detection method by ship head classification and body boundary determination, which needs to generate novel bow features to complete ship detection in the transformation domain of polar coordinates [3]. Although the above studies have achieved good results, the traditional methods are mostly based on the ship structure and shape for manual feature design. Even if the best nonlinear classifier is used to classify these manually designed features, the accuracy of ship detection cannot meet the practical needs.

In recent years, with the development of deep learning, the convolutional neural network (CNN) has achieved more significant performance in image classification owing to the advantages in high-level feature extraction and representation from raw data automatically. At present, the target detection algorithms based on CNN can be divided into two categories: (1) the region-based algorithms, which

ISSN: 2414-1895 DOI: 10.6919/ICJE.202204_8(4).0023

forms a two-stage algorithm represented by R-CNN [5-7]. This kind of algorithms has been used widely in ship detection due to high detection accuracy, but the detection speed is insufficient to meet the requirements of real-time detection. (2) the regression-based algorithms, which forms a one-stage algorithm represented by You Only Look Once (YOLO) [8-10] and Single Shot multi-box Detector (SSD) [11]. This kind of algorithms transforms the detection problem into a regression problem, which greatly improves the detection speed, and has a prominent advantage in real-time ship detection.

Kang et al. disclosed a contextual convolutional neural network with multilayer fusion for ship detection, which employs an intermediate layer combined with a downscaled shallow layer and an up-sampled deep layer to predict the bounding box [12]. Wang et al. designed a SSD model with transmission learning, which could detect ships in SAR images quickly and accurately [13]. Yet, the above detection methods which based on remote sensing images are difficult to accurately classify ships because the remote sensing images are mainly obtained by satellites.

Thus, Shao et al. constructed a new large-scale dataset of ships, which contains six common ship types for training and evaluating ship object detection algorithms [14]. Yan et al. proposed a new target detection algorithm based on the image of the unmanned surface vehicle(USV) in the South China Sea, which can detect ships in real ocean environment in real time by fusing DenseNet and YOLOv3 to improve the stability of detection to decrease the feature loss [15]. Z. Shao et al. put forward for the first time the real-time detection of ships by using the visual images captured by the land surveillance camera network, and put forward a significant awareness CNN framework for ship detection, which can predict categories and positions of ships [16]. Nevertheless, these methods cannot achieve good results when detecting small ships in complex background and small hull differences in a real environment, and the recognition rate of multiple-ship classification is also not ideal.

Above all, in order to solve the problem of low accuracy of small ship detection and classification in complex marine environment, it is necessary to design a fast and accurate algorithm for ship detection and classification. Thus, this paper proposes a ship detection and classification algorithm based on YOLOv4, which mainly consists of a Multi-layer Feature Fusion (MFF) module and a Multi-layer Receptive Field Block(M-RFB) module. The algorithm improves the accuracy requirements of the network by fusing the MFF and the M-RFB into the neck of YOLOv4. Experimental results indicate that the proposed algorithm has a higher accuracy rate than the previous detection algorithm.

The contributions of this paper can be summarized as follows: (1) We propose a novel method which expands the acceptance of small ships by fusing the feature layers of four different scales (13×13, 26×26, 52×52, 104×104) from backbone into the neck of YOLOv4; (2) We propose an improved ship detection algorithm based on YOLOv4. The algorithm mainly integrates MFF and M-RFB into the neck of YOLOv4 to further enhance the semantic features of the context, so as to improve the accuracy of the network for ship detection. (3) We have constructed relevant datasets that contains multiple categories from Singapore Maritime Dataset (SMD) to evaluate the proposed algorithm and compared it with the advanced ship detection algorithms.

2. YOLOv4 Algorithm

The YOLO series of algorithms is a typical end-to-end network designed for improving speed of target detection in deep learning. The core idea of algorithms is to combine target area prediction and category prediction as a regression problem. The target boundary and category probability can be directly predicted through a forward operation, which can greatly improve the detection speed. On this basis, Alexey Bochkovskiy put forward YOLOv4[17] algorithm, which has a significant improvement in detection accuracy and speed.

Compared with the previous YOLO series network framework, the YOLOv4 chooses CSPDarknet53 backbone, SPP[18] additional module, PANet[19] path-aggregation neck, and YOLOv3[20] head as the architecture of algorithm. In YOLOv4, feature maps with three different sizes of 13 × 13, 26 × 26, and 52 × 52 are applied to construct feature pyramids, which are used to detect large, medium,

ISSN: 2414-1895 DOI: 10.6919/ICJE.202204_8(4).0023

and small-sized targets respectively. Multi-scale prediction makes YOLOv4 more sensitive to weak targets and significantly boosts its detection ability. The specific structure of YOLOv4 is shown in Fig. 1.

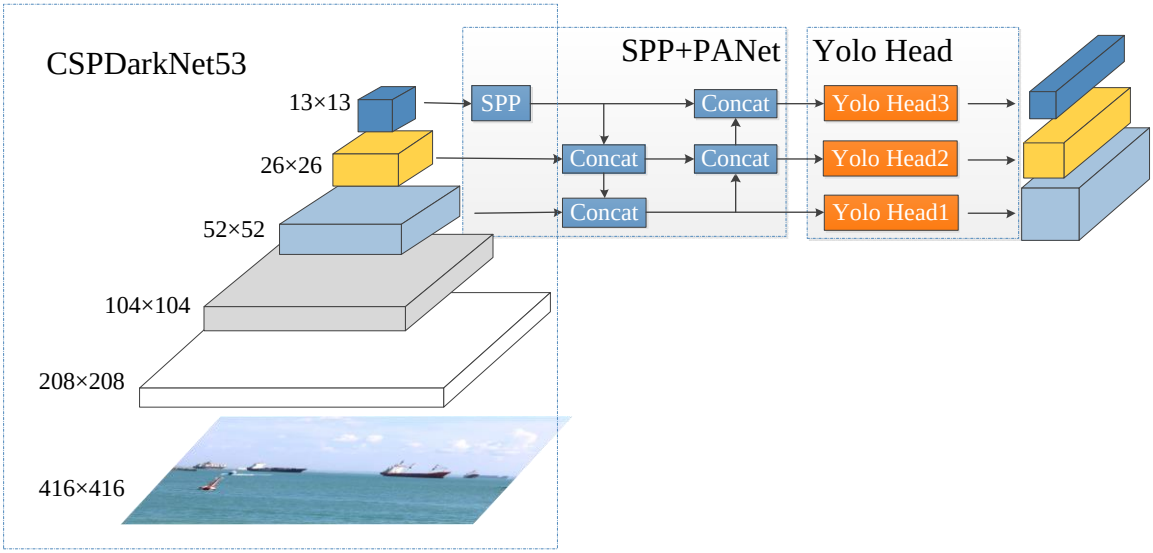

Fig. 1 The structure of YOLOv4

YOLOv4 adopts the CIOU loss function to evaluate the performance of algorithm, which takes three geometric factors into account, namely overlapping area, center distance and aspect ratio. The formula is determined as follows:

$$L_{CIOU} = 1 - IOU + \frac{\rho^2\left(b, b^{gt}\right)}{c^2} + \alpha\upsilon$$

Where IOU represents the intersection ratio of the prediction bounding box and the target bounding box, b is the center point coordinate of prediction box, b^{gt} is the center point coordinate of target bounding box, $\rho\left(b, b^{gt}\right)$ represents the euclidean distance between the center of prediction box and target bounding box, and c represents the diagonal distance of the smallest closure region that can contain both prediction box and target bounding box. $\alpha\upsilon$ is a penalty term for the aspect ratio, υ is used to measure the consistency of the aspect ratio, and α is a positive number. The specific definition is as follows:

$$\upsilon = \frac{4}{\pi^2}\left(\arctan\frac{\omega^{gt}}{h^{gt}} - \arctan\frac{\omega}{h}\right)^2, \alpha = \frac{\upsilon}{\left(1 - IOU\right) + \upsilon}$$

where ω^{gt} and h^{gt} are the width and height of the target bounding box, and ω and h are the width and height of the prediction box. If the width and height of the target bounding box are similar to those of the predicted box, then υ is 0, and the penalty term will not work. So intuitively, the function of this penalty term is to control the width and height of the predicted box to approach the width and height of the target bounding box as quickly as possible.

Volume 8 Issue 4, 2022
DOI: 10.6919/ICJE.202204_8(4).0023

3. Improvement of YOLOv4

As mentioned above, YOLOv4 algorithm adopts the method of multi-scale feature fusion to improve the detection accuracy of small targets. However, the features information of small ships will disappear after down-sampling at the last layer of the backbone, which reduces the detection accuracy of the algorithm. Especially when the environment on the ocean surface is complicated and multiple small ships appear at the same time, the detection effect will be unsatisfactory. Therefore, in order to solve the above-mentioned problems, we propose an improved algorithm based on YOLOv4. This algorithm can improve the detection accuracy of small ships without affecting the detection speed, and the specific structure of the proposed algorithm is demonstrated in Fig. 2.

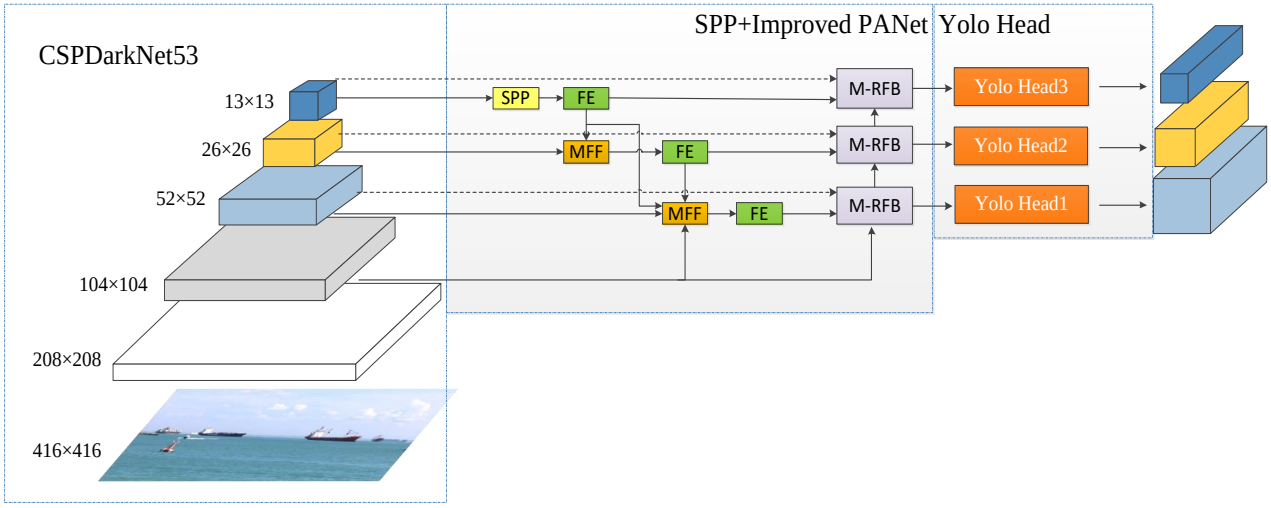

Fig. 2 The structure of Improved-YOLOv4

The backbone of algorithm still adopts the CSPDarknet53 network, in view of the advantages in high-level feature extraction. The SPP structure is also used for the 13×13 feature layer, since it significantly increases the receptive field, and separates out the most significant context features while causing almost no reduction of the network operation speed. In addition, unlike methods in YOLOv4, the neck of algorithm not only integrates four feature layers of different scales from backbone, but also merges feature extractor model (FE), MFF and M-RFB into a novel structure.

In order to solve the problem of low accuracy of small targets, [21] proposed that when the backbone network is very deep, the semantic information of small targets can be expanded by fusing feature layers of more scales. Considering the deep network structure of CSPdarknet53, and the loss of semantic information of small targets will become more and more serious as the network deepens, so we incorporate the feature scale of 104×104 into the neck of improved algorithm, which can effectively combine the semantic information of deep network with the edge feature information of shallow network.

After selecting the corresponding feature layer from the backbone, it is necessary to further extract their features. According to[22], we choose to use the FE module to extract features, which contains two bottleneck and two convolution operations, i.e., Bottleneck and Conv2D. The former is used to reduce the number of channels for less calculation, and the latter is designed to extract contextual features. The specific structure is shown in Fig. 3. In addition, the proposed algorithm also contains two important parts: the MFF module and M-RFB module. The specific details are described in section 3.1 and 3.2.

Volume 8 Issue 4, 2022
DOI: 10.6919/ICJE.202204_8(4).0023

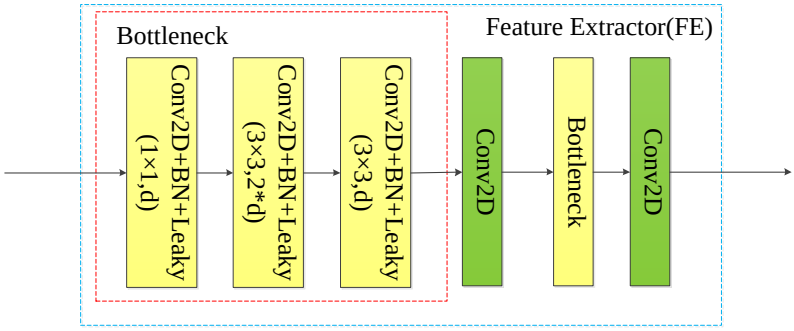

Fig. 3 The FE module

3.1 Multi-layer Feature Fusion (MFF)

In order to further improve the performance of small ship detection, the MFF module is designed to fuse the multi-layer features along the top-down path. Unlike concatenation methods in YOLOv4, the proposed MFF module recursively concatenates contextual features from adjacent or even deeper layer. In other words, the MFF module fuses the features of four adjacent scales (shallow, current, deep and deeper) of the backbone together to enrich features for better detection. The improved algorithm includes two combinations of the MFF module and the MFF-s module. The MFF module is used on the feature layer with a scale of 26×26 and its structure is demonstrated in Fig. 4.

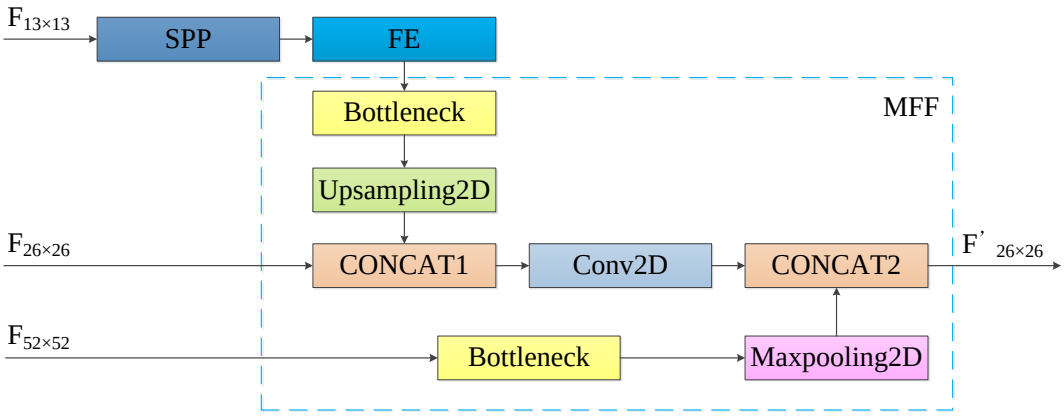

Fig. 4 MFF module with three scale feature layers

Figure 5 clearly illustrates that the MFF module integrates three adjacent scales feature layers. Its input is $F_{13\times13}$, $F_{26\times26}$, and $F_{52\times52}$, and output is $F'_{26\times26}$. The module mainly fuses multi-layer features through two concatenation operations, i.e., CONCAT1, CONCAT2. Among them, CONCAT1 connects the current feature layer(26×26) with the deep features(13×13) through SPP, FE and Upsampling2D. And CONCAT2 connects features that come from CONCAT1 with the shallow features(52×52) through downsampling. The module integrates features from three adjacent scales (shallow, current and deep) of the backbone, and the output of MFF module $F'_{26\times26}$ can be formulated as:

$$F'_{26\times26} = [F_{13\times13}, F_{26\times26}, F_{52\times52}]$$

The MFF-s module is applied on the feature layer with a scale of 52×52. Its input is $F_{13\times13}$, $F_{26\times26}$, $F_{52\times52}$ and $F_{104\times104}$, and output is $F'_{52\times52}$. The specific structure of the MFF-s module is shown in Fig. 5.

Volume 8 Issue 4, 2022
DOI: 10.6919/ICJE.202204_8(4).0023

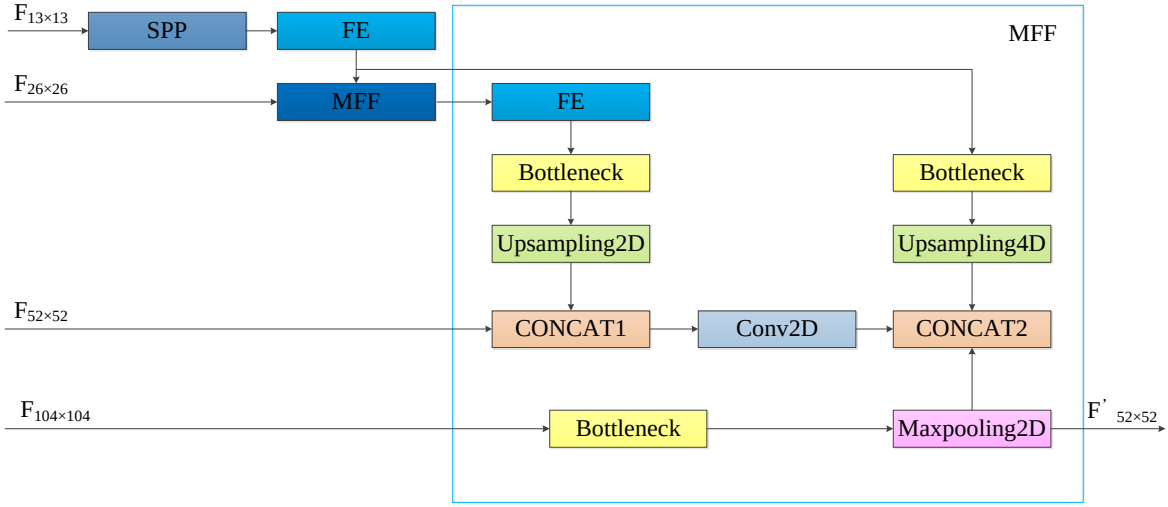

Fig. 5 MFF-s module with four scale feature layers

The MFF-s module also integrates multi-layer features through two concatenation operations, i.e., CONCAT1, CONCAT2. CONCAT1 connects the current feature(52×52) with the deep feature (26×26) that have passed through the MFF module and FE module, and CONCAT2 connects three feature layers of different scales, which are the deeper feature(13×13) through SPP, FE and Upsampling4D, current feature(52×52) through CONCAT1 and Conv2D, and shallow feature (104×104) through Maxpooling2D. The MFF-s module integrates the feature information of the four adjacent scales (shallow, current , deep and deeper), and the output of MFF module $F'_{52\times52}$ can be formulated as:

$$F'_{52\times52} = [F_{13\times13}, F_{26\times26}, F_{52\times52}, F_{104\times104}]$$

3.2 Multi-layer Receptive Field Block(M-RFB)

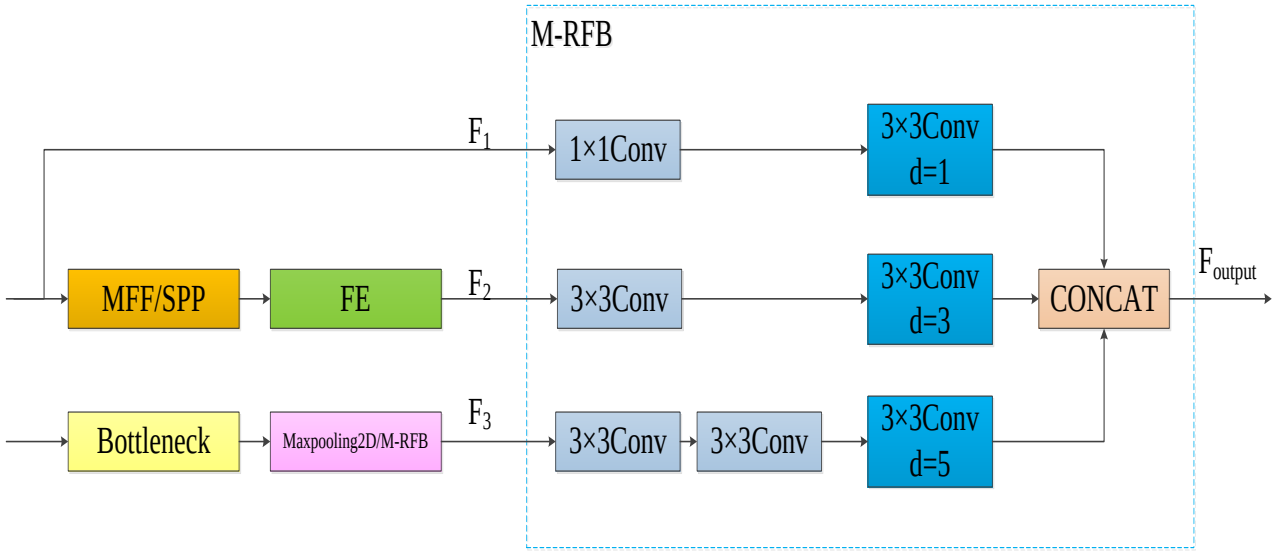

Fig. 6 The M-RFB module

If the algorithm only employs the top-down path to fuse feature maps, its accuracy on small object detection will be problematic when small targets' features gradually disappear or become single pixel at the last layer of a backbone. Therefore, the M-RFB module is designed to further expand the

ISSN: 2414-1895 DOI: 10.6919/ICJE.202204_8(4).0023

receptive field of different regions and obtain more semantic features by adding a bottom-up path. In [23], the receptive field block(RFB) mainly makes use of the principle that the size of population Receptive Field (pRF) is the function of eccentricity in some human retinotopic maps, and the pRF size increases with eccentricity in each map. The dilated convolution is designed to control the eccentricity and reconstruct it to generate the final representation. The M-RFB is mainly used to deal with the feature layer from the FE module, which introduces semantic features from the shallow scale by adding a bottom-up path. The specific details are shown in Fig. 6.

The input of the M-RFB module is F_1, F_2 and F_3 in Figure 7, where F_1 represents the current feature, F_2 represents the feature that the current feature layer comes out through the MFF module or the SPP module and the FE module, F_3 is the shallow feature through Bottleneck and Maxpooling2D or M-RFB modules. The three feature layers of F_1, F_2 and F_3 firstly reduce their number of channels by 1×1 or 3×3 convolution operations, thereby decreasing the amount of calculation. Then, the receptive field of the feature layers are expanded by dilated convolution operation with expansion rates of 1, 3 and 5. Finally, the concatenation operation is carried out, so that the shallow features can be fused with the current features to further expand the receptive field of the feature layer. The out of M-RFB module F_{output} can be formulated as:

$$F_{\text{output}} = [F_1, F_2, F_3]$$

4. Experimental Results

4.1 Datasets

The images in our datasets are mainly obtained from the Singapore Maritime Dataset (SMD) introduced by Prasad et al. [24]. SMD provides VIS and NIR videos taken by Canon 70D camera on shore (the camera is fixed on the platform) and on board (the camera is on the moving ship). It contains 81 video files, including 240,842 target tags of 10 different categories, which can train our model to accurately detect and classify ships. According to the method of dataset partition proposed in reference [25], we construct dataset from SMD. In the dataset, 4470 images are taken as training set and 1880 images as testing set. Images extracted from dataset with their corresponding label are shown in Fig. 7.

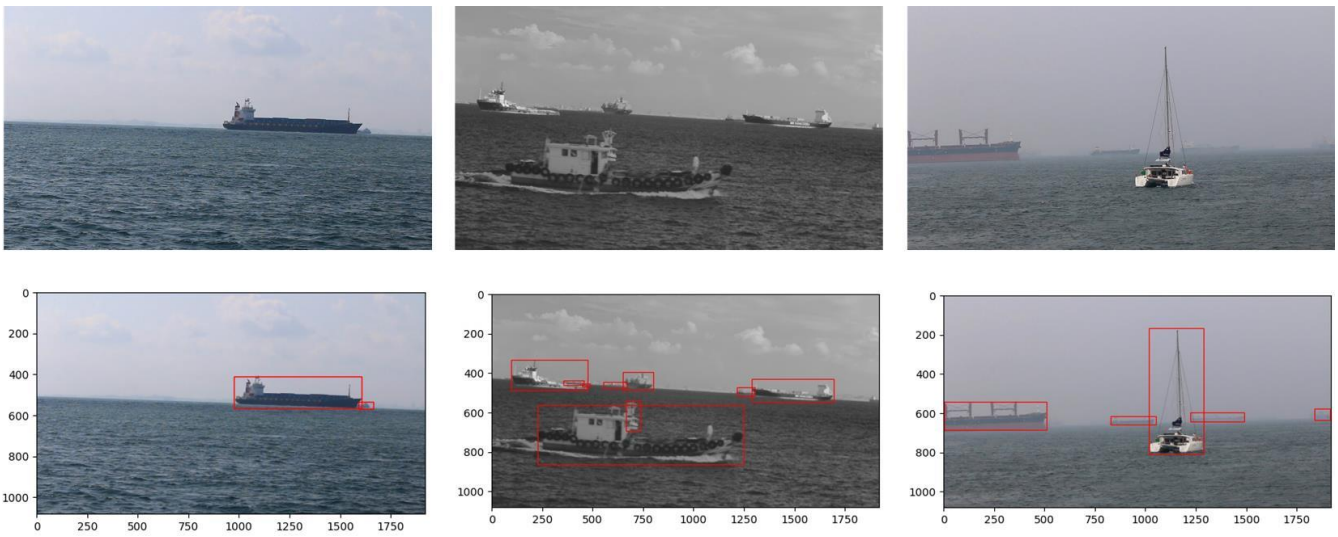

Fig. 7 Images of the first dataset

ISSN: 2414-1895 DOI: 10.6919/ICJE.202204_8(4).0023

4.2 Evaluation Index

In order to effectively evaluate the performance of the proposed algorithm for ship detection, we use average precision (AP) and mean average precision (mAP) as evaluation indexes of detection accuracy. In multi-class target detection, AP calculates the area under the precision-recall (PR) curve, and mAP is the average value of AP of multiple classes. Precision represents the proportion of samples that are correctly detected in all test results. Recall represents the proportion of samples that are correctly detected in all positive samples.

4.3 Detection Performance

Fig. 8 shows that the loss decline curves of the YOLOv4, the YOLOv4+MFF and the YOLOv4+MFF+M-RFB algorithm. The loss of all the three models decreases gradually with the increase of the epochs, and eventually converges to a low constant. After 100 epochs, the final loss of the YOLOv4+MFF+M-RFB is about 16, and the YOLOv4+MFF is about 17, while the final loss of the original YOLOv4 is about 18. It is notable that the initial value of the YOLOv4+MFF+M-RFB at the beginning of training is much lower than the other two algorithms, which illustrates that the weights of the proposed algorithm could be trained with a lower time cost. In addition, the loss curve of the YOLOv4+MFF+M-RFB is smoother with the increase of epochs times, which means that this network is stable and robust.

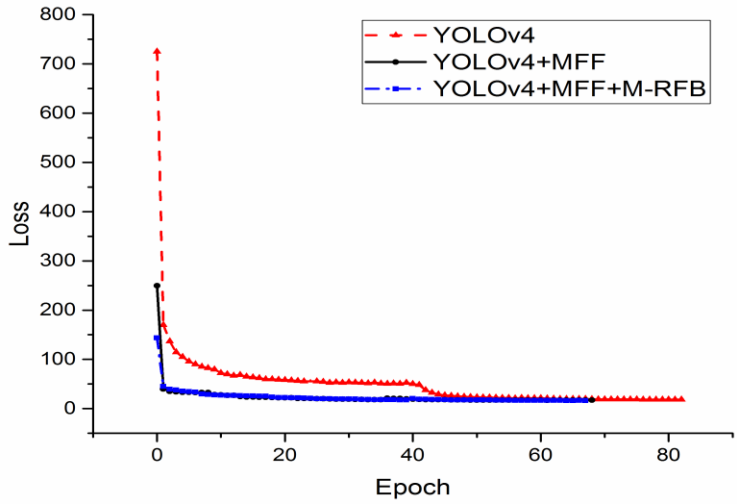

Fig. 8 Loss decline curves of three algorithms

We have taken some typical images from the dataset, and compared the detection performance of the YOLOv4 and the proposed algorithm. The detection results were illustrated in Fig. 9. The (a), (b), (c)and(d) of Fig. 9 listed the detection results achieved under different environmental conditions by implementing the YOLOv4 and the proposed algorithm. In the case of the interferences, such as camera shaking and other objects, the proposed algorithm can still accurately detect small ships in the distance (marked with a red dotted circle in Fig. 9(a),(b)). The images in Fig. 9 (c) and (d) were taken on shore, which contains more small ships, and the red dotted circles marked in figure demonstrate that the proposed algorithm can detect small ships more accurately than YOLOv4 without missing detection. These experimental results surely prove that the proposed algorithm can detect ships more accurately and robust under different environment.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202204_8(4).0023

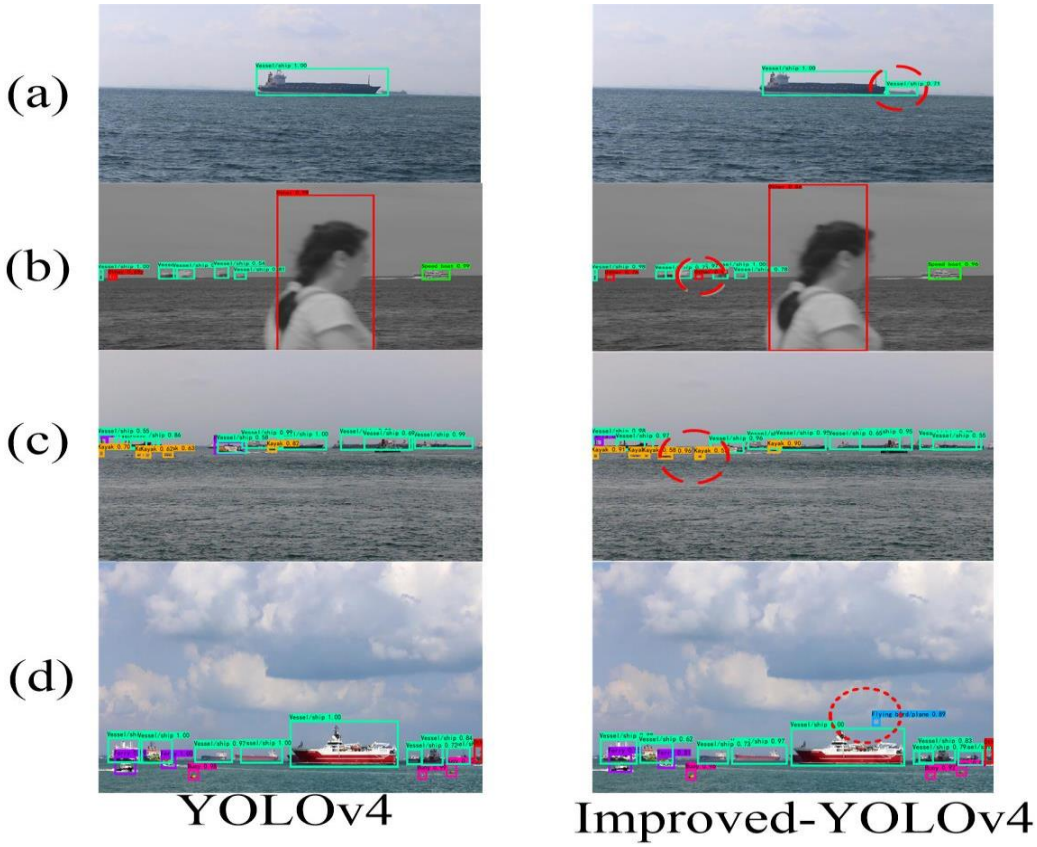

Fig. 9. Comparative detection results of Yolov4 and Improved-YOLOV4 algorithms

4.4 Comparison with State-of-the-arts

In order to fully verify the validity of the proposed algorithm and prove the good performance of this algorithm in ship detection, the proposed algorithm is also compared with other intelligent methods, such as Faster R-CNN+Resnet50, Faster R-CNN+Resnet101, SSD, SSD+MobileNet and YOLOv3 under the same dataset and adopt the same evaluation index: mAP and FPS. The test results are shown in Table 1.

Table 1. Comparison results with other detection algorithms

Model	mAP	FPS
Faster-RCNN+Resnet50	73.68%	1.09
Faster-RCNN+Resnet101	64.08%	0.80
SSD	45.65%	9.07
SSD-MobileNet	56.34%	**11.32**
YOLOv3	57.64%	10.36
YOLOv4	64.75%	9.90
Improved-YOLOV4	**76.39%**	9.22

ISSN: 2414-1895 DOI: 10.6919/ICJE.202204_8(4).0023

According to the results in the above table, in terms of average precision, the mAP value of the proposed algorithm is the highest, which is 76.39%. In terms of detection speed, the proposed algorithm is 9.22, which is slightly lower than SSD-MobileNet, but its accuracy is much higher. Compared with YOLOv4, the proposed algorithm also has a slight sacrifice in speed, but achieves higher accuracy. Therefore, the experimental results illustrate that the proposed algorithm based on YOLOv4 has a strong detection ability on Singapore Maritime Dataset, and performs well in the practical application of ship detection.

5. Conclusion

This paper proposes an improved algorithm based on YOLOv4 for ship detection in complex marine environments. The proposed algorithm mainly integrates the MFF module and the M-RFB module into the neck of YOLOv4, so that YOLOv4 can fuse features at multiple scales, fully enhance context semantic features and expand the receptive field, which improves the detection accuracy of small ships. The real images obtained from the Singapore Maritime Dataset were used to train the algorithm, and compared with the advanced algorithms such as YOLOv4 and Faster R-CNN to evaluate the performance of the proposed algorithm. Experimental results show that the proposed algorithm is superior to YOLOv4 and has better overall performance.

Acknowledgments

This paper was financially supported by the National Natural Science Foundation of China (Grant No. 52071200, 61404083) and State Key Laboratory of ASIC & System (2021KF010).

References

[1] J. Huang, Z. Jiang, H. Zhang, B. Cai and Y. Yao, "Region proposal for ship detection based on structured forests edge method," 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Fort Worth, TX, 2017, pp. 1856-1859.

[2] Zhu Qiuyu, Jiang Yilong and Chen Bo, "Design and implementation of video-based detection system for wharf ship," IET International Conference on Smart and Sustainable City 2013 (ICSSC 2013), Shanghai, 2013, pp. 493-496.

[3] S. Li, Z. Zhou, B. Wang and F. Wu, "A Novel Inshore Ship Detection via Ship Head Classification and Body Boundary Determination," in IEEE Geoscience and Remote Sensing Letters, vol. 13, no. 12, pp. 1920-1924, Dec. 2016.

[4] Liu L , Ouyang W , Wang X , et al. Deep Learning for Generic Object Detection: A Survey[J]. International Journal of Computer Vision, 2019, 128(11).

[5] R. Girshick, J. Donahue, T. Darrell and J. Malik, "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation," 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, 2014, pp. 580-587.

[6] R. Girshick, "Fast R-CNN," 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, 2015, pp. 1440-1448.

[7] S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, pp. 1137-1149, 1 June 2017.

[8] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 779-788.

[9] J. Redmon and A. Farhadi, "YOLO9000: Better, Faster, Stronger," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017, pp. 6517-6525.

[10] Redmon J , Farhadi A . YOLOv3: An Incremental Improvement[J]. 2018.

[11] Liu W , Anguelov D , Erhan D , et al. SSD: Single Shot MultiBox Detector[C]// European Conference on Computer Vision. Springer International Publishing, 2016.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202204_8(4).0023

[12] Kang M , Ji K , Leng X , et al. Contextual Region-Based Convolutional Neural Network with Multilayer Fusion for SAR Ship Detection[J]. Remote Sensing, 2017, 9(8):860-.

[13] Y. Wang, C. Wang, H. Zhang, C. Zhang and Q. Fu, "Combing Single Shot Multibox Detector with transfer learning for ship detection using Chinese Gaofen-3 images," 2017 Progress in Electromagnetics Research Symposium - Fall (PIERS - FALL), Singapore, 2017, pp. 712-716.

[14] Z. Shao, W. Wu, Z. Wang, W. Du and C. Li, "SeaShips: A Large-Scale Precisely Annotated Dataset for Ship Detection," in IEEE Transactions on Multimedia, vol. 20, no. 10, pp. 2593-2604, Oct. 2018.

[15] Li Y , Guo J , Guo X , et al. A Novel Target Detection Method of the Unmanned Surface Vehicle under All-Weather Conditions with an Improved YOLOV3[J]. Sensors, 2020, 20(17):4885.

[16] Z. Shao, L. Wang, Z. Wang, W. Du and W. Wu, "Saliency-Aware Convolution Neural Network for Ship Detection in Surveillance Video," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 30, no. 3, pp. 781-794, March 2020.

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

[18] C. Wang, H. Mark Liao, Y. Wu, P. Chen, J. Hsieh and I. Yeh, "CSPNet: A New Backbone that can Enhance Learning Capability of CNN," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle, WA, USA, 2020, pp. 1571-1580.

[19] K. He, X. Zhang, S. Ren and J. Sun, "Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 37, no. 9, pp. 1904-1916, 1 Sept. 2015.

[20] S. Liu, L. Qi, H. Qin, J. Shi and J. Jia, "Path Aggregation Network for Instance Segmentation," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, 2018.

[21] Cao C , Wu J , Zeng X , et al. Research on Airplane and Ship Detection of Aerial Remote Sensing Images Based on Convolutional Neural Network[J]. Sensors, 2020, 20(17):4696.

[22] P. -Y. Chen, J. -W. Hsieh, C. -Y. Wang and H. -Y. Mark Liao, "Recursive Hybrid Fusion Pyramid Network for Real-Time Small Object Detection on Embedded Devices," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle, WA, USA, 2020, pp. 1612- 1621.

[23] Liu S , Huang D , Wang Y . Receptive Field Block Net for Accurate and Fast Object Detection[C]// European Conference on Computer Vision. Springer, Cham, 2018.

[24] Prasad, Dilip. Singapore Maritime Dataset. 2016. https://sites.google.com/site/dilipprasad/ home/ singapore-maritime-dataset.

[25] D. K. Prasad, D. Rajan, L. Rachmawati, E. Rajabaly, and C. Quek, "Video Processing from Electro-optical Sensors for Object Detection and Tracking in Maritime Environment: A Survey," IEEE Transactions on Intelligent Transportation Systems (IEEE), 2017.