Proceedings of 2022 IEEE
International Conference on Mechatronics and Automation
August 7 - 10, Guilin, China

Study on Marine Fishery Law Enforcement Inspection System based on Improved YOLO V5 with UAV

Yong Lu[1], Jian Guo[1,2*]
[1]Tianjin Key Laboratory for Control
Theory& Applications
In Complicated systems and
Intelligent Robot Laboratory
Tianjin University of Technology
BinshuiXidao Extension
391,Tianjin,300384,China
Fuqiang6369@hotmail.com;
1303945356@qq.com

Shuxiang Guo[1,2,3*]
[2] Shenzhen Institute of Advanced Biomedical Robot Co., Ltd.
No.12, Ganli Sixth Road, Jihua Street, Longgang District,
Shenzhen,518100,China
*corresponding author :
jianguo@tjut.edu.cn

Qiang Fu[3]
[3]Key Laboratory of Convergence Medical Engineering System and Healthcare Technology,The Ministry
of Industry and Information Technology,School of Life Science
Beijing Institute of Technology
No.5, Zhongguancun South Street, Beijing,100081, China
*corresponding author :
guoshuxiang@hotmail.com

Jigang Xu[4*]
[4]Unit68709
Qinghai
Haidong,810700, China
*corresponding author :
xujigang216@163.com

***Abstract* –As a part of China's territory, the wide sea not only represents China's powerful territorial power, but also contains incomparably rich Marine resources.And the vast area of the sea has led to a lack of inadequate maritime supervision. The abundant fishery resources are one of the important economic sources of China, but the law enforcement of ocean and fishery has long been plagued by the contradictions among the wide sea area, long coastline, large number of fishing boats in fishing ports, difficult management and limited law enforcement. It is difficult to achieve full-time, all-directional and all-coverage control.Relying on the existing Marine and fishery law enforcement personnel, vehicles and vessels, equipment and so on, it is difficult to achieve sufficient and effective law enforcement supervision.At present, the combination of UAV technology and target detection has a wide application prospect, providing a new technical means for Marine and fishery law enforcement. In this paper, the improved YOLO V5 algorithm is configured on the Unmanned Aerial Vehicle （UAV） to further improve the work efficiency of the UAV.Compared with the initial YOLO V5, the paper changed the 3-scale feature extraction of Neck end into 4-scale feature extraction. By integrating more shallow feature information of more scales, the feature expression ability of path aggregation network was enhanced, the detection accuracy of small targets was improved, and the missed detection rate was reduced.At the same time, the improvement of boundary box detection function and non-maximum suppression NMS improves the research of target recognition and target overlapping detection.Finally, the training results show that the improved YOLO V5 can further improve the effectiveness of UAV Marine fishery law enforcement.**

***Index Terms* –UAV, Marine law enforcement, YOLO V5, Feature extraction.**

I. INTRODUCTION

In recent years, in order to maintain the order of Marine and fishery management, the ocean and fishery law enforcement agencies at all levels have strengthened publicity and cooperation[1], and the overall order of Marine and fishery management has been stable. However, it is difficult to achieve full-time, all-directional and comprehensive law enforcement supervision, and illegal activities such as "illegal fishing", "illegal farming" and "illegal reclamation" occur from time to time to rely on the existing Marine and fishery law enforcement personnel, law enforcement vehicles, law enforcement vessels and law enforcement equipment. At present, the rise of UAV technology and the application of target detection algorithms provide new technical means for marine and fishery law enforcement.

UAV are being used more and more widely in the ocean. The North Sea Technical Support Center of the State Oceanic Administration (SOA) and The Aerospace Star Technology Co., Ltd. have carried out UAV flight work for ocean monitoring and island exploration[2].It provides an important basis for maritime surveillance and law enforcement.

The North Sea Forecast Center of the State Oceanic Administration uses UAV to conduct terrain mapping and environmental exploration at sea[3]. In the sea ice monitoring, the UAV is equipped with YOLO V5 detection algorithm, which can comprehensively monitor the movement track of icebergs and ice floes, which can not only improve the effect of surveying and mapping, but also realize the safety monitoring of airlines.

Tianjin Key Laboratory of Intelligent Remote Sensing Information Processing Technology Enterprises applied UAV remote sensing technology to marine resources supervisio[4]. The matching target detection algorithm is used for coastal and Marine supervision, island supervision, aquaculture zone supervision, ship target identification, marine pollution supervision and so on.It makes full use of UAV advantages and play a positive role in protecting the ocean.

978-1-6654-0853-0/22/$31.00 ©2022 IEEE

UAV are widely used in the ocean.In this paper, an UAV equipped with the improved YOLO V5 algorithm is used for marine fishery law enforcement inspection, which improves the law enforcement efficiency.This paper mainly consists of the following parts: The first part is the introduction, the second part is the introduction of the platform, the third part is the brief introduction of YOLO V5 algorithm, the fourth part is the introduction of its improvement, as well as the comparison of the two algorithms and the final summary.

II. The Platform of uav Inspection System

The platform uses quadrotor UAV as inspection tool.Patrol was conducted by carrying a raspberry pi equipped with camera and improved YOLO V5 algorithm on the UAV. UAV system can assist the fishery law enforcement, in suspected illegal activities or high-risk areas, in high altitude area and relatively area for the surveillance video, collecting fishery illegal evidence, accurate in detecting fishing boats and related personnel behavior whether there is any violation behavior, reducing the leak rate and rate of overlap. The law enforcement vessel can judge whether the fishing boat and relevant personnel violate laws and regulations by relying on the evidence obtained, make precise policies and quickly carry out the next administrative law enforcement. At the same time, the use of UAV with warning or commanding voice recordings can be used to call out illegal fishing boats, improving the deterrence of fishery administrative law enforcement. The system platform is shown in Figure 1.

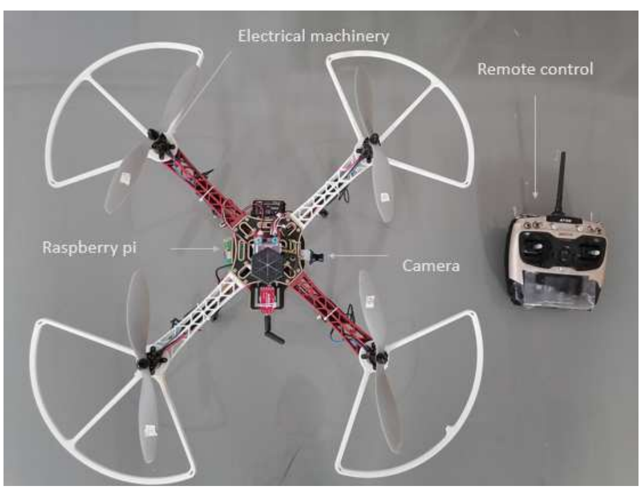

Fig. 1 UAV inspection system

The system framework of this paper is shown in figure 2. It can be seen that the Raspberry Pi system equipped with improved YOLO V5 algorithm is carried on the UAV, and the content captured is detected and analyzed by the camera, and then transmitted to the ground statio[5].The next step will be carried out according to the detected content. If illegal or irregular behaviors are detected, law enforcement personnel can remotely control the UAV to carry out warning expulsion or other administrative law enforcement action[6].

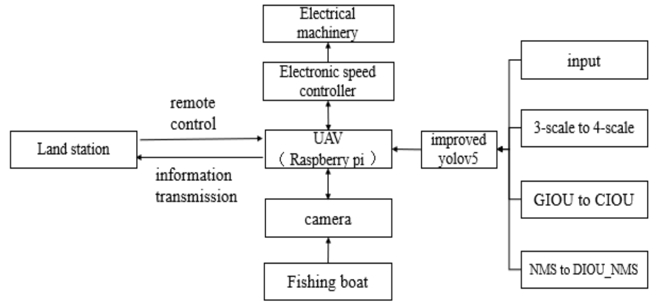

Fig. 2 System framework diagram

III. The YOLO V5 Framework

In recent years, Redmon et al. proposed a new zone-free object detection method called YOLO (You Only Look Once).As a unified, real-time detection framework, YOLO detection speed is very fast.YOLO uses a single convolutional network to predict the position and type of bounding frame directly based on the whole image. Firstly, the YOLO model of a picture is adopted to reduce the repeated detection of the same target by adopting spatial constraints, which greatly improves efficiency and can achieve real-time results. As a new member of the YOLO family, YOLO V5 is a good choice for UAV target detection. Its network structure is mainly composed of four parts: input terminal, backbone, neck and prediction.

The input terminal includes Mosaic data enhancement, image scaling and adaptive anchor frame. The Mosaic data enhancement method can effectively improve the detection effect of small targets, and is suitable for detection objects with small targets.

Backbone feature extraction network, including CSP structure and the newly added Focus structure.Two kinds of CSP structures are designed in YOLO V5[7-8]. The CSP1 _X structure is applied to Backbone network, and the other CSP2 _X structure is applied to Neck.In Focus structure, the key is the slicing operation and the original image of 640*640*3. First, four copies are copied, and then the four images are cut into four slices of 320*320*3 through the slicing operation. Then Concat is used to connect the four slices in depth, and the output is 320*320*12.The output of 320*320*32 is generated through the convolution layer with 32 convolution cores. Finally, the result is input to the next convolution layer through Batch_norm and Leaky_ReLU.It improves the receptive field of each point on the feature map, reduces the loss of original information and reduces the amount of calculation to accelerate the detection speed.

PANet structure is adopted in Neck structur[9-10], which is composed of FPN+PAN structure.PAN carries out up-sampling and fusion of feature images in a top-down manner. The structure of PANet is shown in the figure below:

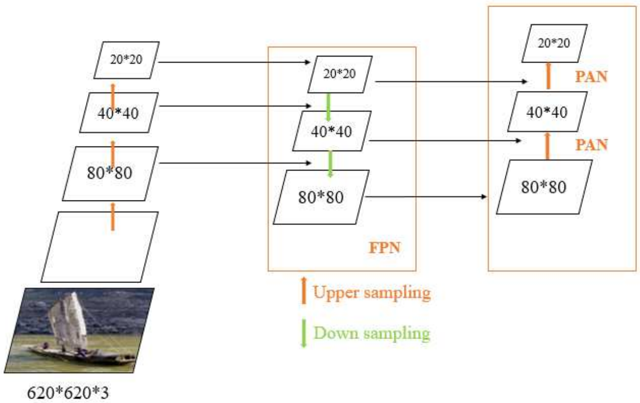

Fig. 3 PANet structure

At the output end, YOLO V5 adopts GIoU_Loss as the loss function, and adopts weighted NMS to solve the problem of inaccurate detection caused by serious occlusion.GIoU_Loss is shown in Equation (1) :

$$X_{GIoU} = 1 - IoU + \frac{\rho^2\left(b, b^{gt}\right)}{c^2} \tag{1}$$

Where b , b^{gt} represents the center point of the prediction frame and the target frame respectively, $\rho(.)$ represents the Euclidean distance, and c represents the diagonal distance of the minimum peripheral matrix that can contain both the prediction frame and the real frame.Meanwhile, YOLO V5 uses FPN+PAN structure to finally form three feature layers of different scales to predict the target to be detected. The 640*640 resolution image is taken as input, and the three feature maps of 20*20, 40*40 and 80*80 scale specifications are output. This detection method using multi-scale feature fusion has good robustness in detecting targets of different sizes in images. However, only using feature layers of three scales for prediction can not fully utilize the underlying feature information, which will lead to the loss of small target location information and is not conducive to the detection of small objects.

Aiming at the shortcomings of YOLO V5, in order to effectively improve the efficiency of Marine fishery law enforcement, the paper partially improves the YOLO V5 algorithm[11], and verifies the effectiveness of the improved part through training.

IV. Improved Parts of YOLO V5

With YOLO V5 on the UAV, multiple targets can be detected, which is one of the effective means of Marine fishery law enforcement. However, when the UAV is flying in the air, the target under the lens becomes "small" due to the height problem, so it is possible to miss some targets. Therefore, the original 3-scale feature extraction of yolo v5 is extended to 4-scale feature extraction to integrate more shallow feature information, which improves the feature expression ability of the path aggregation network[12-13], improves the accuracy of detecting small targets and reduces the missed detection rate. At the same time through the improvement of boundary box detection function and non-maximum suppression nms to improve the target recognition effect and target overlapping detection effect.To further improve the effectiveness of drone Marine fisheries law enforcemen[14-15].

A. Improvement of neck end

The original yolo v5 detects the target object through feature maps of three different scales, and outputs feature maps of three scales with specifications of 20*20, 40*40 and 80*80 respectively. However, the shallow information is not fully utilized by only using the features of three scales, which will lead to the loss of some small target information.Therefore, for the improvement of YOLO V5 network, the original 3-scale detection is expanded to 4-scale, forming 4 detection branch structures, as shown in Figure 4. The input size is 640*640, and each branch extracts features from Backbone network respectively. The deep feature maps extracted from the backbone network are up-sampled in FPN and fused with the corresponding size of the early stage of the network into valid information, thus realizing concat connection operation. However, there is a problem in FPN network. It is difficult for shallow feature graph to fuse with high-level feature graph when transferring information to deep layer. Therefore, PAN network structure is implemented on the basis of FPN feature pyramid network, and a bottom-up path is added. The fusion feature images in FPN network are continued to be sampled and fused from bottom to top to achieve reverse fusion of feature images, so that richer feature information can be obtained.The bottom-up path of PAN structure is used to fuse the fused feature images in FPN network from bottom to top again and perform down-sampling operation to realize reverse fusion of feature images. Finally, independent detection is performed on the fused feature images of four scales.The improved multi-scale fusion can learn the strong location features from the shallow feature layer, and the bottom-up path of PAN structure can be fused again so that the deep features can be detected more accurately. The feature expression ability of path aggregation network (PANet) is enhanced by integrating shallow feature information of more scales, and the detection accuracy of small targets is improved and the missing rate is reduced.

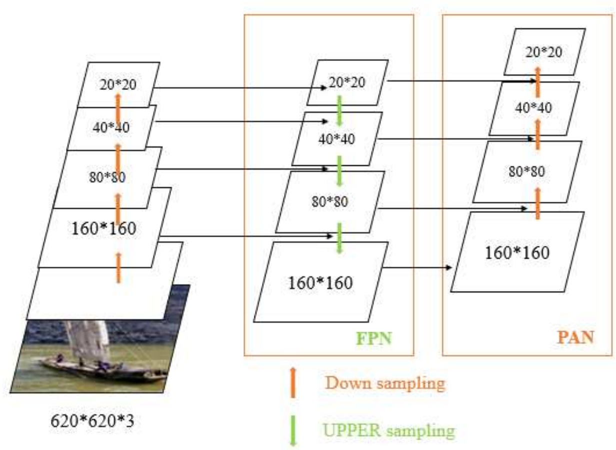

Fig. 4 The multi-scale feature fusion method of PANet was improved

B. Improvement of boundary box detection loss

GIoU_ Loss function is used in IoU_ Los function in original yolo v5 algorithm. The Loss function of target detection task is generally composed of Classificition Loss (classification loss function) and Bounding Box Regeression Loss(regression loss function)[16].Recent regression functions of target box mainly include IOU_Loss, GIOU_Loss, DIOU_Loss and CIOU_Loss.

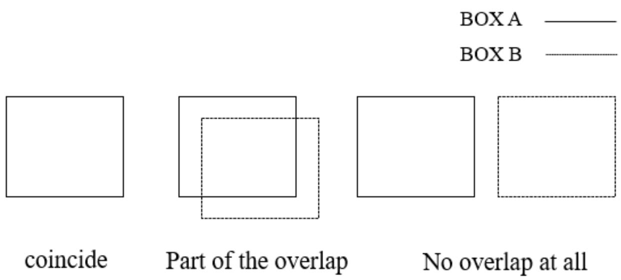

Fig. 5 IoU loss function diagram

As shown in fig 5, when the two boxes overlap completely, $X_{IoU}=1$;When the intersection of the two boxes is empty, $X_{IoU}=0$;When the two boxes overlap, the value of X_{IoU} is between 0 and 1.However, when the intersection of the two boxes is 0, no matter how far apart the two boxes are, the value of the IoU loss function is equal to 0, which cannot represent the loss in this case. GIoU derived from IoU means that the minimum closure area of the two boxes is calculated first, that is, the overlap of the two boxes.Divide the absolute value of the ratio of C to $A\cup B$ by the absolute value of C to get the specific gravity of the region in the closure region that does not belong to the two boxes. Finally, calculate the difference between IoU and specific gravity to get the value of G.In the case that the two boxes approach to overlap indefinitely $X_{GIoU}=X_{IoU}=1$.

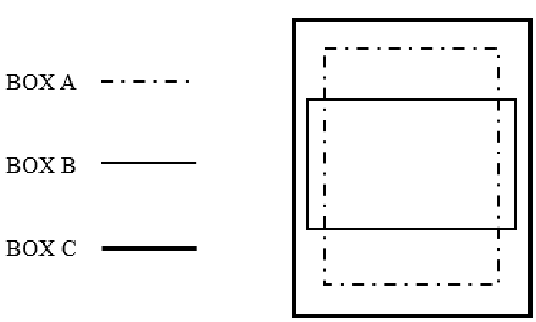

Fig. 6 GIOU loss function diagram

When IoU loss function is adopted, when two boxes do not coincide, no matter how big the gap is, the loss function is 0. Different from IoU, GIOU algorithm not only focuses on the size of the overlap area of two boxes, but also adds the non-overlap area[17], so YOLO V5 avoids the above problems.However, when GIOU_Loss contains two boxes, GIOU_Loss will degenerate into IOU_Loss and GIOU_Loss needs many iterations to converge. Considering the shortcomings of GIOU, CIOU_Loss is introduced. CIOU_Loss takes into account three important geometric factors: overlap area, center distance and aspect ratio in the regression function of the target box. The expressions are as follows:

$$X_{CIoU}=1-IoU+\frac{\rho^2\left(b,b^{gt}\right)}{c^2}+\alpha v \qquad (2)$$

Where, α is the weight, v measures the similarity of aspect ratio, b,b^{gt} respectively represent the center point of the prediction box and the target box, and the distance between them adopts the European distance ρ . c represents the minimum skew of the bounding box that can contain both the prediction box and the target box. CIOU_Loss can directly minimize the distance between the center point of the prediction frame and the real frame to accelerate convergence. Meanwhile, it can also increase the detection of loss in different scales of the real frame, and increase the length and width loss, so that the whole prediction frame will be more completely consistent with the real frame.So CIOU_Loss instead of GIOU_Loss, the effect will be better.

C. Non - maximum suppression improvement

In the final processing stage of the target detection algorithm, non-maximum suppression (NMS) algorithm is usually required to select the target box for multi-target box screening. In addition, one step of the NMS algorithm is to calculate the current maximum score box and the IOU size of other boxes. For this step, change the calculation of the IOU.Traditional NMS is mainly used to filter prediction boxes. The IOU index is used to suppress redundant prediction boxes. However, overlapping parts may cause error information.The prediction frame to suppress redundancy should consider not only the overlapping part but also the center distance between the prediction frame and the target frame.DIOU considers both.Therefore, the original network NMS is changed to DIOU, so in the network detection for overlapping multiple targets, the detection effect of DIOU is significantly better than that of the traditional NMS algorithm, as shown in the figure, and the formula is shown below.

$$k_i=\begin{cases}k_i, IoU-X_{DIoU}\left(N,B_i\right)<\varepsilon \\ 0, IoU-X_{DIoU}\left(N,B_i\right)>\varepsilon\end{cases} \qquad (3)$$

$$X_{DIoU}=\frac{\rho^2\left(b,b^{gt}\right)}{c^2} \qquad (4)$$

Where, IoU represents the intersection ratio between the prediction box and the target box, ε represents the nms threshold, and k_i represents the classification score of each different categor[18-19].

V. Comparation of The Two Algorithms Before And After Improvement

A. The experimental set up

In order to compare the advantages and disadvantages of the two algorithms, we need to establish experimental analysis. Firstly, we found 500 pictures of fishing boats from the internet, and then used Labelimg image annotation tool to mark the images. After the data set is acquired, training is performed and then comparative analysis is performed. In addition, the specific process of UAV patrol is shown in Fig 7. A camera-equipped UAV (shown in Figure 1) patrols through the air, and the camera takes pictures of the sea surface. Raspberry Pi detect and analyze the content captured by the camera, and then transmitted to the ground station. According to the detected content, the staff determines whether there is any illegal or irregular behavior of fishing vessels. If there are violations, law enforcement officials will use the UAV to enforce the law. If not, it keeps the UAV to patrol.

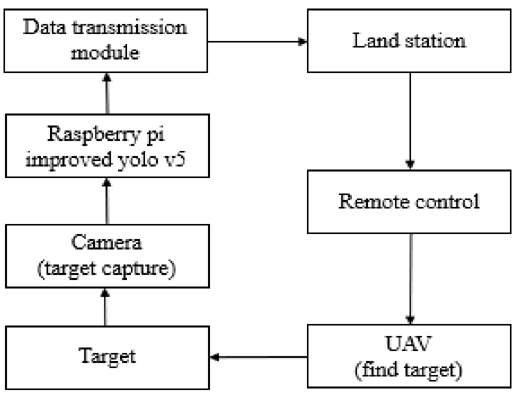

Fig. 7 Detection process

B. Experimental results and analysis

Through data training, we can get a comparison between the two algorithms for image detection in the same state. The following image is the result of image detection under original YOLO V5. The image detection results are as follows.

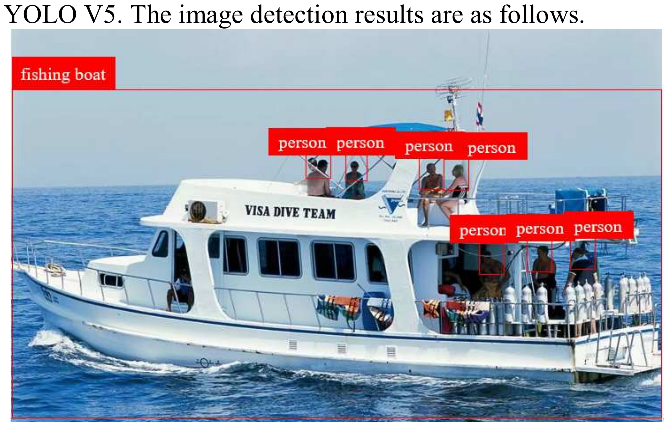

Fig. 8 Small target detection results

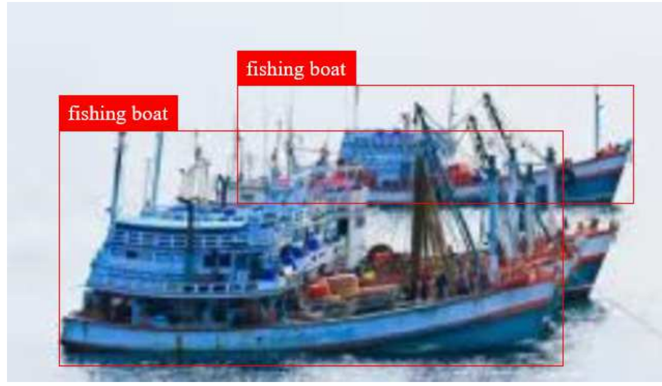

Fig.9 Overlapping target detection results

The following two pictures are the detection results of the target after improved YOLO V5. The image detection results are as follows：

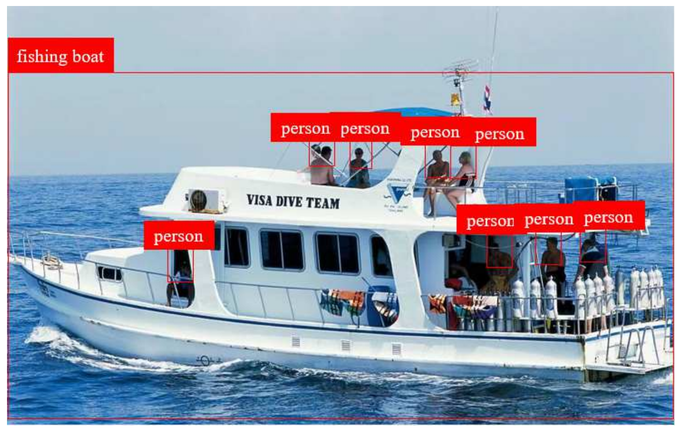

Fig. 10 Small target detection results

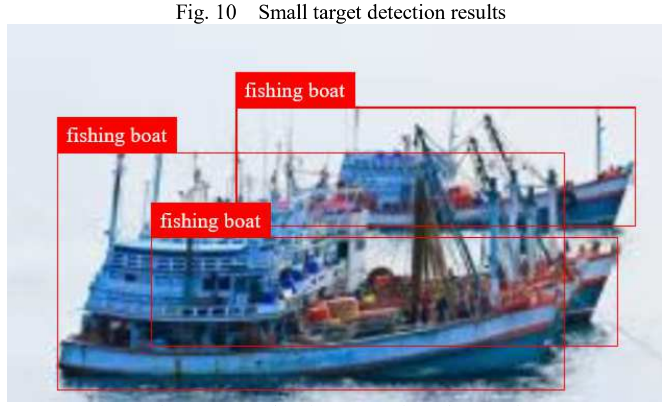

Fig. 11 Overlapping target detection results

Through training, we have obtained target detection in the same state by the two algorithms before and after the improvement. According to the comparison between Figure 8 and Figure 9, it can be seen that the improved detection algorithm can detect small targets not detected in the original figure.According to the comparison between Figure 10 and 11, it can be seen that the improved detection algorithm can detect the overlapping targets in the original figure.In addition, the data pairs of the two algorithms before and after improvement are shown in Table 1. Although the recall rate did not improve, the precision and map increased by 4.4% and 2.9% respectively..Precision is a measure of the percentage of results that are correct for all positive samples.It represents the total number of targets correctly predicted by the model in the total number of predicted targets. Recall measures the percentage of all positive samples that are predicted to be correct. It represents the total number of targets correctly predicted by the model in the total number of predicted targets. Ap is used to calculate the average accuracy of a single class model.

TABEL 1
COMPREHENSIVE INDEX TEST RESULTS

Model	P%	R%	Ap%
Yolov5	94.5%	91.7%	91.3%
Improved yolov5	98.9%	91.7%	94.2%

VI. CONCLUSIONS

In order to solve the problem of Marine fishery law enforcement, the law enforcement has been trapped in the contradiction between wide sea area, long coastline, large number of fishing boats in fishing ports, difficult management and limited law enforcement power.In this paper, the improved YOLO V5 algorithm is used to assist law enforcement personnel in Marine law enforcement inspection.By changing the initial 3-scale feature extraction of YOLO V5 Neck end into 4-scale feature extraction, the shallow feature information of more scales is integrated, the feature expression ability of path aggregation network is enhanced, the detection accuracy of small targets is improved, and the missed detection rate is reduced.At the same time, the improvement of boundary box detection function and non-maximum suppression nms improves the research of target recognition and target overlapping detection.Finally, the training results show that compared with the original algorithm, although recall is not improved, precision and map are improved by 4.4% and 2.9% respectively, which can further improve the efficiency of UAV marine fishery law enforcement.

ACKNOWLEDGMENTS

This research is supported by National Natural Science Foundation of China (61703305), Key Research Program of the Natural Science Foundation of Tianjin (18JCZDJC38500) and Innovative Cooperation Project of Tianjin Scientific and Technological (18PTZWHZ00090).

REFERENCE

[1] Ruochen An, Shuxiang Guo, Yuanhua Yu, Chunying Li, Tendeng Awa, "Multiple Bio-Inspired Father-Son Underwater Robot for Underwater Target Object Acquisition and Identification", Micromachines 2022, Vol.13, No.1, DOI: 10.3390/ mi13010025, 2021.

[2] Li Zhongqiang, Tang Wei, Zhang Zhen, Liu Tingting.Ocean development and management, Vol. 31, No. 7, pp. 42-44, 2014.

[3] Lin Sen, Tao Huihui, Wu Yongjian, Yuan Kelei.Engineering and Construction, Vol. 35, No. 1, pp. 47-48, 2021.

[4] Zhang Haolin, Xu Xiaoxu. "Application of UAV remote sensing technology in Marine resources supervision". Information Technology and Informatization, No. 07, pp. 231-233, 2020.

[5] Liang Zheng, Shuxiang Guo, Shuoxin Gu, "The communication and stability evaluation of amphibious spherical robots", Microsystem Technologies, Vol.25, No.7 pp.2625-2636, DOI: 10.1007/s00542-018-4223-5, 2018.

[6] Shuxiang Guo, Handong Cheng, Jian Guo,Jiang Xu "A Novel Target Tracking System for the Amphibious Robot based on Improved Camshift Algorithm", Journal of Bionic Engineering, DOI: 10.1007/s42235-021-00073-0, 2021.

[7] Tian Feng, Jia Haopeng, Liu Fang. "Improve YOLOv5 oilfield operation site safety dress small target detection". Computer system application, Vol.31, No.3, pp.159-168,DOI: 10.15888 / j.carol carroll nki. Csa. 008359, 2022.

[8] Liu Yunxiao, Wang Jingdong, Huang Yuqiu, Zhao Ruochen. "MTCNN optimization and its application in road vehicle detection ". Journal of optoelectronic technology, Vol.33, No.3, pp.196-204, DOI: 10.19453 / j.carol carroll nki. 1005-488 - x. 2019.03.008, 2019.

[9] Dou Qilong, Yan Mingzhong, Zhu Daqi.Marine "Small target detection based on YOLO- V5 satellite-based SAR image". Applied Science and Technology, Vol.48, No.6, pp.1-7, 2021.

[10] Woo S, Hwang S, Kweon IS. Stair Net: "Top-down semanticaggregation for accurate one shot detection". Proceedings of 2018 IEEE Winter Conference on Applications of ComputerVision. Lake Tahoe: IEEE, 2018. pp.1093–1102.

[11] Ruochen An, Shuxiang Guo, Yuanhua Yu, Chunying Li, Tendeng Awa, "Multiple Bio-Inspired Father-Son Underwater Robot for Underwater Target Object Acquisition and Identification", Micromachines 2022, Vol.13, No.1, DOI: 10.3390/ mi13010025, 2021.

[12] Zheng Qiu mei, Wang Lulu, Wang Feng-hua. "Small target detection in traffic scenes based on improved convolutional Neural network".Computer Engineering,Vol.46, No.6, pp.26-33, 2020.

[13] Ma Xiao, Shao Limin, Jin Xin, Xu Guanlei. "Research progress of ship target recognition technology". Science & technology review, Vol.37, No.24, pp.65-78, 2019.

[14] Guo Xingang, Zhang Peidong, Liang Jinming, Wang Shuai.Improve "Faster - R - CNN target detection method". Journal of changchun university of technology, Vol.9, No.5, pp.474-480, DOI: 10.15923 / j.carol carroll nki cn22-1382 / t, 2020

[15] Duan Jingya, Li Bin, Dong Chao, Tian Lianfang. "The classification of ship target detection algorithm based on YOLOv2". Computer engineering and design, Vol.9, No.6, pp.1701-1707,DOI: 10.15923 / j.carol carroll nki cn22-1382 / t., 2020

[16] Zhang Qilin, Lin Qingping, Xiao Lei. "Improved YOLOv5 aerial image recognition algorithm". Yangtze River Information Communication, Vol.34, No.3, pp.73-76, 2021.

[17] Rezatofighi H, Tsoi N, Gwak J Y, et al. Generalized intersection over union: a metric and a loss for bounding box regression//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA, 2019.

[18] Huiming Xing, Liwei Shi, Kun Tang, Shuxiang Guo, Xihuan Hou, Yu Liu, Huikang Liu, Yao Hu, "Robust RGB-D Camera and IMU Fusion-based Cooperative and Relative Close-Range Localization for Multiple Turtle-Inspired Amphibious Spherical Robots", Journal of Bionic Engineering, Vol. 16, No.3, pp. 442-454, 2019.

[19] Everingham M, van Gool L, Williams CKI, et al. The Pascal visual object classes (VOC) challenge. International Journal of Computer Vision,Vol.88, No.2, pp.303-338, DOI:10.1007/s11263-009-0275-4, 2010.