A new method-improved yolov5 for detection of the marine

Xingkun Li [1], Guodong Chen [1], Fengguang Jia [1] and Ziyang Wang [1]

[1] Naval Architecture and Port Engineering College, Shandong Jiaotong University, WEH 264299, China.

Abstract. Due to the particularity of underwater environment, such as uneven distribution of light, interference caused by various underwater noises to imaging, and weak target features caused by the protective colors of underwater organisms themselves, the existing target detection technology is not ideal in underwater application scenarios. In this paper, the photography function of underwater robot was used to take photos of four objects, namely sea urchins, scallops, sea cucumbers and starfish, and the parts with obvious texture characteristics were selected to make a dataset URFD with 8,000 images. In order to ensure the adaptive ability of the detection algorithm to application scenarios, we selected the yolov5 algorithm with the best comprehensive performance by testing the performance comparison of various existing networks, and optimized and improved it. In order to reduce the redundant computation of convolution operations, the fasterNet structure is introduced. By optimizing the loss function of the original network structure and reducing the coupling degree of the original detection head, yolov5 dropped 2 points from floating point operation and improved the map by 2 points on the basis of the original performance. And the training time cost is reduced by 22%. The name of the improved algorithm is yolovsfd.

Keywords: underwater object detection; yolov5; deep learning; yolovsfd.

1. Introduction

In recent years, with the continuous development of fisheries, deep learning algorithms have provided technical support for smart fisheries. Through the combination of computer vision and underwater robot, intelligent fishing can not only improve the efficiency of fishing, but also reduce the risk of artificial underwater operation[1-3]. Compared with the application scenarios on land, the underwater environment has weak light and uneven distribution, low chroma, and the influence of water flow, impurities and other factors. For underwater organisms, factors such as low foreground differentiation caused by protective colors adapted to the environment, and irregular texture caused by shapes are great tests for visual algorithms[4-6].

At present, there is no shortage of data sets of underwater related scenes to provide training resources for deep learning[7]. However, most of the existing data sets are pictures taken from a wide Angle, so the detection targets are almost all widely distributed small targets, which will cause the model to be unable to fully extract and train the features of the target[8-11]. In order to make up for this shortcoming of the underwater datasets, the underwater photography robot is used to shoot the target in shallow water area at close range, so that the target has a clear texture display[1, 5, 12]. We selected 2000 pictures with significant texture for four kinds of targets, namely sea urchin, scallop, starfish and sea cucumber, and integrated them with other original small target distribution pictures to make a dataset URFD with a total amount of 8000 pictures.

As for the network model of target detection, it is mainly divided into two-stage and single-stage network model algorithms. The classical two-stage algorithm, such as faster-rcnn, which has high accuracy, but too many operational parameters lead to slow detection speed and cannot meet the real-time requirements[13-16]. Detection single-precision algorithms include ssd and yolo series, which are characterized by high speed but low precision[17-19]. With the development of multi-scale network algorithms, yolo series has been continuously improved, and the higher version of yolo algorithm also has relatively high accuracy[20-22]. Coupled with the speed advantage, yolo series algorithm is the best choice to meet the real-time detection.

By comparing the performance of various target detection algorithms in underwater application scenarios, yolov5 is selected to improve the network model. In order to improve the underwater adaptive ability of the network model, the previous methods are to introduce attention mechanism or modify a certain network structure module, but this is often at the expense of speed to improve the accuracy[20, 22-24]. In this paper, the latest algorithm idea, fasternet, is introduced for convolutional redundant parameter calculation to ensure that the parameter number of the whole network will not be affected by the improvement of accuracy. In this way, the accuracy will not be improved at the premise of sacrifice. In this paper, by optimizing the original loss function and combining the decoupling detection head, the convergence ability of the network model is improved, the training time cost is reduced by 22%, and the map index is also improved by 2 points. We call the new algorithm yolovsfd.

2. Yolovsfd

The yolovsfd network model, based on yolov5, uses FasterNet network structure to replace C3 modules that bottleNeck and backbone parts. In this way, the efficiency of feature processing and the overall computing speed of the network are improved[25]. In the network header, the decoupling header is used to replace the original shared branch header. In this way, the detection accuracy can be improved. In addition to the structural explicit improvement, we also improved the CIOU loss function of the original network by using NCIOU-Loss. This loss function not only improves the fitting ability of the whole network, but also greatly saves the convergence time, thus saving the time cost of model training. As shown in Figure 1:

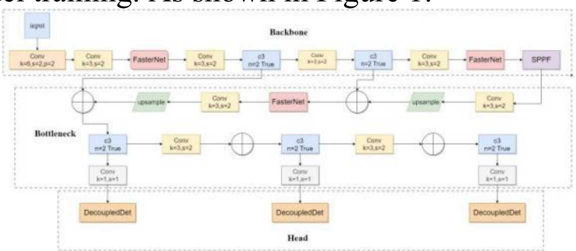

Figure. 1 yolovsfd

2.1 Optimization loss function

Yolov5 takes the center point of the bounding box, the intersection over union and the aspect ratio as the loss basis. This method has excellent convergence speed and regression performance[26].

$$\alpha = \frac{v}{1 - IOU + v} \tag{1}$$

$$v = \frac{4}{\pi^2}(\arctan\frac{w^{gt}}{h^{gt}} - \arctan\frac{w}{h})^2. \tag{2}$$

$$\mathcal{L}_{CIoU} = 1 - IoU + \frac{\rho^2(b,b^{gt})}{c^2} + \alpha v. \tag{3}$$

Where the αv is the aspect ratio of bounding box. $\rho^2(b, b^{ot})$ is the Euclidean distance between the center point of the prediction box and the ground truth. c is the diagonal distance of the smallest closure area capable of containing both the prediction box and the ground truth. However, the aspect ratio is still different from the true length and width, which may result in that even if the aspect ratio of the prediction frame and the ground truth is exactly the same, the length and width are not necessarily the same. This shortcut lowers the performance of network .

In order to solve this defect, length and width are added as the loss basis of regression box on the basis of original loss function, so as to increase the upper limit of prediction accuracy.

$$NCIOU_{Loss} = 1 - IOU + \alpha v + \frac{\rho^2(b^{at}, b)}{c^2} + \frac{\rho^2(h^{gt}, h)}{c_h^2} + \frac{\rho^2(w^{gt}, w)}{c_w^2} + \lambda \tag{4}$$

Where the $\rho^2(h^{gt}, h)$ and $\rho^2(w^{gt}, w)$ respectively are the difference of height and width between prediction box and ground truth. λ is used to avoid overfit of the network, we set it's number is 0.001.

2.2 FasterNet Moudle

The characteristic of PConv structure is that only part of the input feature image is processed by convolutional filter, which achieves lower FLOPs than conventional convolutional network and higher FLOPS than Depthwise Group convolutional method. The FasterNet module is the first feature extraction by the above PConv module, and immediately connects two 1-dimensional convolution kernels. The primary BN layer and RelU activation function are used in the intermediate modules of the two PWConv modules to ensure feature diversity and low network latency[25].

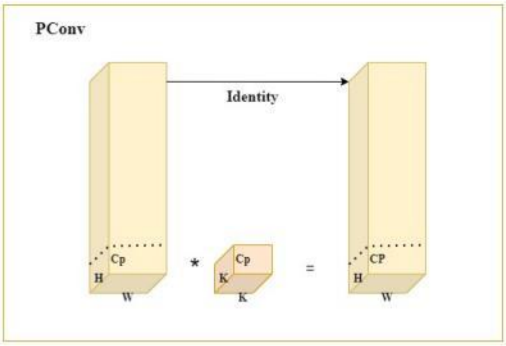

Figure. 2 PConv

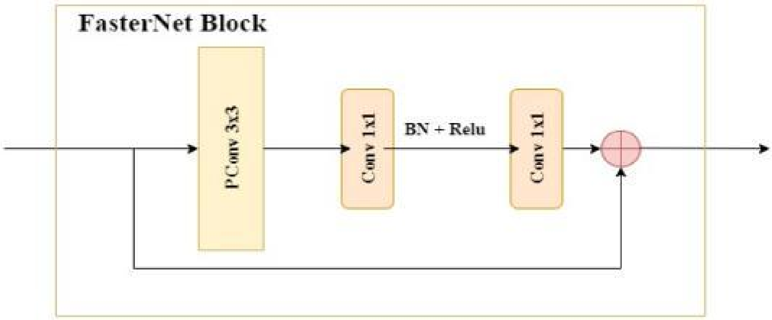

Figure. 3 FasterNet Block

2.3 Decoupled detection head

In the original structure of yolov5, the detection header is realized through classification and regression branch fusion sharing. In yolox[27], the research team uses decoupling to separate the modules of each network task, so that each function has its own branch, so as to reduce the coupling between network modules and improve the detection accuracy. The author reduces the number of convolutional modules by using mixed channel strategy. Thus, the problem of decreasing the floats of the model due to the increase of the number of parameters is alleviated[28].

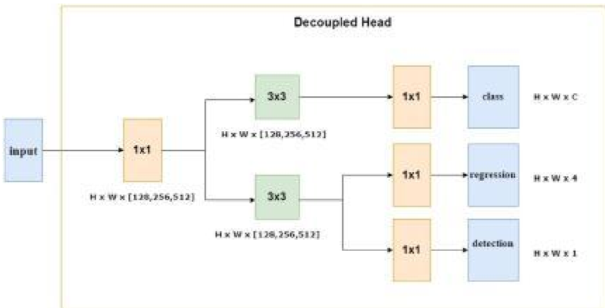

Figure 4 Decoupled detection head

3. Experiment and result analysis

3.1 Experiment environment

In this experiment environment,NVIDIA GeForce RTX 2080 was used.

3.2 Ablation experiment

Table 1. Comparison of ablation

	echinus	holothurian	scallop	starfish	Map	FLOPs	Time of train
	AP						
Yolov5-doubleHead	87.0	70.3	40.0	78.5	56.6	23.0	1.601H
Yolov5-ECIou	83.4	66.6	38.0	79.3	54.6	16.0	1.234H
Yolov5-FasterNet	79.0	60.2	23.0	86.5	52.3	13	1.405H
yolovsfd	87.5	68.2	42.0	89.2	57.2	15	1.118H

In this experiment, 50 rounds of training were conducted for each model improvement method, and the input size was 640x640. The Batchsize is 32.

Through testing different optimization components of the three network structures separately, it can be seen from the above table that the decoupling detection head can significantly improve the accuracy of the obvious original model, but it will increase the burden of floating point operation. The optimized loss function can also improve the recognition accuracy and reduce the training time. And the FasterNet module can have a low float numbers.

3.3 Experimental comparison

Table 2. cmparison of networks

	echinus	holothurian	scallop	starfish	map	FLOPs	Time of train
	AP						
FasterRcnn	88.0	65.7	35	88.9	56.7	30	1.756H
SSD	76.3	60.1	22	75.2	49.1	17	1.545H
Yolov3SPP	80.0	55.3	24	77.5	48.6	18	1.601H
Yolov5	83.4	65.6	30	88.3	55.6	16	1.434H
Yolov8	82.0	60.2	23	84.5	52.3	12	1.505H
yolovsfd	87.5	68.2	42	89.2	57.2	15	1.118H

As can be seen from the above table, before improving the network, this paper first trained and compared five original network algorithms. Although the accuracy of FasterRcnn is the highest, its floating point arithmetic is too large, which means that the real time performance is poor. Therefore, for comprehensive consideration, we chose yolov5 algorithm, and on this basis, improved the algorithm, and then proposed a yolovsfd algorithm model with higher efficiency. As can be seen from Table 3-1, yolovsfd has better recognition performance, and its real-time performance is not as good as that of Yolov8, but compared with the original yolov5, the parameter calculation of yolovsfd has also decreased by one point, which means that yolovsfd has better real-time perormance than yolov5.

3.4 Validation result

It can be seen from the test results that yolovsfd has a more ideal detection ability on sea cucumbers, starfish and sea urchins. However, due to environmental problems, feature extraction of scallops is more difficult, which makes our algorithm not good for scallop recognition. The unbalanced number of object samples and the environmental interference are serious obstacles to the model judgment,which will also be the contents for our next work.

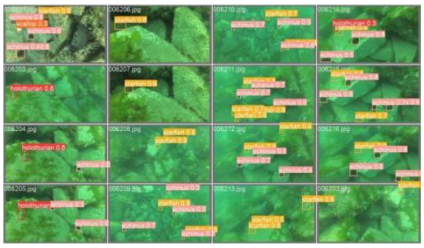

Figure. 5 The result of detection

4. Conclusion

We use the latest object detection network structure FasterNet and decoupledDt, improve the original yolov5 network structure, and propose yolovsfd algorithm. yolovsfd not only improves the recognition accuracy of the network model to the underwater target, but also improves the detection rate of the algorithm, ensuring the real-time detection of the algorithm. In terms of loss function, by optimizing CIOU loss function, we improved the limitations brought by the original aspect ratio, and added longer than wider as a loss sentence, which is that the convergence speed is increased during network training. Compared with the original algorithm, yolovsfd saved 22% of the training time cost. This paper also uses other network algorithms to compare performance with yolovsfd, and yolovsfd shows better performance.

References

[1] T. Guo, Y. Wei, H. Shao, B. Ma, Research on Underwater Target Detection Method Based on Improved MSRCP and YOLOv3, in: 2021 IEEE International Conference on Mechatronics and Automation (ICMA), 2021, pp. 1158-1163.

[2] W. Hao, N. Xiao, Research on Underwater Object Detection Based on Improved YOLOv4, in: 2021 8th International Conference on Information, Cybernetics, and Computational Social Systems (ICCSS), 2021, pp. 166-171.

[3] M.S.A.B. Rosli, I.S. Isa, M.I.F. Maruzuki, S.N. Sulaiman, I. Ahmad, Underwater Animal Detection Using YOLOV4, in: 2021 11th IEEE International Conference on Control System, Computing and Engineering (ICCSCE), 2021, pp. 158-163.

[4] G. Divyabarathi, S. Shailesh, M.V. Judy, Object Classification in Underwater SONAR Images using Transfer Learning Based Ensemble Model, in: 2021 International Conference on Advances in Computing and Communications (ICACC), 2021, pp. 1-4.

[5] F. Han, J. Yao, H. Zhu, C. Wang, Marine Organism Detection and Classification from Underwater Vision Based on the Deep CNN Method, Mathematical Problems in Engineering, 2020 2020 1-11.

[6] A. Jesus, C. Zito, C. Tortorici, E. Roura, G. De Masi, Underwater Object Classification and Detection: first results and open challenges, in: OCEANS 2022 - Chennai, 2022, pp. 1-6.

[7] C. Liu, H. Li, S. Wang, M. Zhu, D. Wang, X. Fan, Z. Wang, A Dataset and Benchmark of Underwater Object Detection for Robot Picking, in: 2021 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), 2021, pp. 1-6.

[8] A.F. Ayob, K. Khairuddin, Y.M. Mustafah, A.R. Salisa, K. Kadir, Analysis of Pruned Neural Networks (MobileNetV2-YOLO v2) for Underwater Object Detection, in: Proceedings of the 11th National Technical Seminar on Unmanned System Technology 2019, 2021, pp. 87-98.

[9] S. Cai, G. Li, Y. Shan, Underwater object detection using collaborative weakly supervision, Computers and Electrical Engineering, 102 2022.

[10] H. Chen, J. Lin, L. Zhuge, X. Xia, Underwater Image Restoration and Target Detection Based on Monocular Depth Estimation, in: 2021 China Automation Congress (CAC), 2021, pp. 5597-5601.

[11] R. Chen, Z. Fu, Y. Huang, E. Cheng, X. Ding, A Robust Object Segmentation Network for UnderWater Scenes, in: ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2022, pp. 2629-2633.

[12] J.C. Hipolito, A. Sarraga Alon, R.V. Amorado, M.G.Z. Fernando, P.I.C. De Chavez, Detection of Underwater Marine Plastic Debris Using an Augmented Low Sample Size Dataset for Machine Vision System: A Deep Transfer Learning Approach, in: 2021 IEEE 19th Student Conference on Research and Development (SCOReD), 2021, pp. 82-86.

[13] W. Ge, J. Sun, Y. Xu, H. Zheng, Real-time Object Detection Algorithm For Underwater Robots, in: 2021 China Automation Congress (CAC), 2021, pp. 7703-7707.

[14] H. Huang, H. Zhou, X. Yang, L. Zhang, L. Qi, A.-Y. Zang, Faster R-CNN for marine organisms detection and recognition using data augmentation, Neurocomputing, 337 2019 372-384.

[15] A. Jalal, A. Salman, A. Mian, M. Shortis, F. Shafait, Fish detection and species classification in underwater environments using deep learning with temporal information, Ecological Informatics, 57 2020.

[16] S. Li, B. Pan, Y. Cheng, X. Yan, C. Wang, C. Yang, Underwater Fish Object Detection based on Attention Mechanism improved Ghost-YOLOv5, in: 2022 7th International Conference on Intelligent Computing and Signal Processing (ICSP), 2022, pp. 599-603.

[17] A. Amin, S. Bahnasy, K. Elghamry, A. Samir, A. Emad, M.S. Darweesh, A.A. El-Sherif, Real-Time Fish Detection Approach on Self-Built Dataset Based on YOLOv3, in: 2021 3rd Novel Intelligent and Leading Emerging Sciences Conference (NILES), 2021, pp. 41-46.

[18] M.S. Asyraf, I.S. Isa, M.I.F. Marzuki, S.N. Sulaiman, C.C. Hung, CNN-based YOLOv3 Comparison for Underwater Object Detection, Journal of Electrical & Electronic Systems Research, 18 2021 30-37.

[19] L. Chen, F. Zhou, S. Wang, J. Dong, N. Li, H. Ma, X. Wang, H. Zhou, SWIPENET: Object detection in noisy underwater scenes, Pattern Recognition, 132 2022.

[20] H. Fan, D. Zhu, Y. Li, An Improved Yolov5 Marine Biological Object Detection Algorithm, in: 2021 2nd International Conference on Artificial Intelligence and Computer Engineering (ICAICE), 2021, pp. 29-34.

[21] H. Li, C. Zhang, N. Wan, Q. Chen, D. Wang, D. Song, An Improved Method for Underwater Image Super-Resolution and Enhancement, in: 2021 IEEE 4th International Conference on Electronics Technology (ICET), 2021, pp. 1295-1299.

[22] S. Zhao, J. Zheng, S. Sun, L. Zhang, An Improved YOLO Algorithm for Fast and Accurate Underwater Object Detection, Symmetry, 14 2022.

[23] W. Chen, B. Fan, Underwater Object Detection With Mixed Attention Mechanism And Multi-Enhancement Strategy, in: 2020 Chinese Automation Congress (CAC), 2020, pp. 2821-2826.

[24] X. Guan, R. Li, P. Hu, A Novel Detection Method for Weak Underwater Acoustic Targets Based on RUSBoost, in: 2020 IEEE 9th Joint International Information Technology and Artificial Intelligence Conference (ITAIC), 2020, pp. 2257-2261.

[25] J. Chen, S.-h. Kao, H. He, W. Zhuo, S. Wen, C.-H. Lee, S.-H.G. Chan, Run, Don't Walk: Chasing Higher FLOPS for Faster Neural Networks–Supplementary Material–.

[26] Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, D. Ren, Distance-IoU loss: Faster and better learning for bounding box regression, in: Proceedings of the AAAI conference on artificial intelligence, 2020, pp. 12993-13000.

[27] L. Han, F. Li, H. Yu, K. Xia, Q. Xin, X. Zou, BiRPN-YOLOvX: A weighted bidirectional recursive feature pyramid algorithm for lung nodule detection, Journal of X-Ray Science and Technology, 2023 1-17.

[28] C. Li, L. Li, H. Jiang, K. Weng, Y. Geng, L. Li, Z. Ke, Q. Li, M. Cheng, W. Nie, YOLOv6: A single-stage object detection framework for industrial applications, arXiv preprint arXiv:2209.02976, 2022.