Transactions on Computational and Scientific Methods | Vo. 4, No. 12, 2024
ISSN: 2998-8780
https://pspress.org/index.php/tcsm
Pinnacle Science Press

Efficient Underwater Biological Target Detection Using the YOLOv7-Tiny Algorithm

Calder Wainwright[1], Ravi Vishwanath[2] , Xi Chen[3]
[1]School of Marine Technology, University of Southampton, Southampton, UK
[2]School of Marine Technology, University of Southampton, Southampton, UK
[3]School of Marine Technology, University of Southampton, Southampton, UK
*Correspondence should be addressed to Xi Chen; chenxixi1997@gmail.com

Abstract:
The detection and recognition of underwater organisms are critical for marine resource utilization and economic development. This study introduces a YOLOv7-tiny-based target detection algorithm optimized for underwater biological target detection. Trained on the Aquarium marine organism dataset, the proposed algorithm is evaluated against YOLOv5s, demonstrating improved accuracy, recall, and average precision while significantly reducing model size and parameter count. These improvements enhance its suitability for deployment on embedded devices in complex underwater environments. Future work will focus on integrating attention mechanisms, small target detection layers, and novel loss functions to broaden the algorithm's applicability across diverse domains.

Keywords:
Deep Learning; YOLOv7; Underwater Environment Detection.

1. Introduction

Underwater resources are inseparable from human development, and the development of underwater resources has gradually gained attention in recent years, and our country has abundant sea areas to develop and utilize resources such as underwater organisms, and turn them into economy, so that they can promote our economic development. Underwater organisms, such as fish, have also become an integral part of the daily diet of human beings, as well as the focus of fishing and aquaculture. The number as well as the type of underwater organisms affects the development and utilization of marine resources, and therefore, the development of underwater resources has become a major driving force in the detection of underwater targets.

Underwater target detection tasks can be divided into two categories according to the target signal to be detected [1]: one is the use of underwater sonar technology for underwater target detection, this method is more mature development, but compared with the optical image resolution is low, and vulnerable to interference, only applicable to long-range target detection; one is based on computer vision technology for underwater target detection, optical image resolution, contains rich information, in The optical image has high resolution and contains rich information, which has outstanding advantages for underwater target detection at close range. Therefore, target detection based on computer vision technology is gradually becoming the main research direction for underwater near-range target detection and recognition.

At this stage, target detection based on computer vision technology mainly uses deep learning, and the mainstream target detection algorithms of deep learning can be divided into two categories according to the completion steps: target detection techniques based on deep learning, in terms of network structure, can be divided into one-stage framework (One-stage) and two-stage framework (Two-stage) [2]. The One-stage network approach discards the candidate region extraction step and uses only the first-stage network to perform both classification and regression tasks.

The common One-stage algorithms mainly include RCNN algorithm, YOLO series algorithms, etc. RCNN algorithms mainly include Faster-RCNN [3], Mask-RCNN [4], etc. YOLO series algorithms mainly include YOLOv3 [5], YOLOv4 [6], YOLOv5, YOLOv7 [7], etc. Shi [8] et al. proposed an underwater target detection algorithm based on YOLOv4 network and improved YOLOv4 by adding CBAM attention mechanism to the backbone network, improving the original path aggregation network PANet of YOLOv4, and enhancing the data with PredMix, and this algorithm outperformed YOLOv4 by 7.03% on the URPC2018 dataset . Qiang [9] et al. proposed the study of improved SSD-based underwater target detection algorithm to improve the SSD target detection model using depth-separated deformable convolution, so that the detection accuracy is improved. Huang [10] et al. improved the YOLOv5 algorithm, including the introduction of an improved CBAM attention mechanism, and improved the PANet of YOLOv5, which led to improved detection accuracy on the public dataset of underwater optical target detection jointly established by Dalian University of Technology and Pengcheng Laboratory.SUNG M et al [11] applied YOLO to underwater target detection, and the NOAA The method was trained and tested on an underwater dataset with a classification accuracy of 93%.CAI K et al [12] and YANG H et al [13] applied the improved YOLOV3 to underwater target detection, which further improved the target detection accuracy and speed.CHEN L et al [14] applied the improved YOLOV4 detection network to underwater, and the average accuracy of IOU threshold greater than 0.5 rate AP50 improved by 0.11 compared to the base model.

But the above methods to improve the algorithm although the accuracy has improved, but the improved model size is relatively large, but for application to embedded devices, the smaller the model size as well as detection speed, the better, so this paper introduces the latest research results of the YOLO series, and introduces the YOLOv7-tiny algorithm to underwater biological target detection and recognition, and in the complex environment in improving the accuracy at the same time, the amount of parameters is reduced, and the network model is smaller.

2. YOLOv7-tiny Network

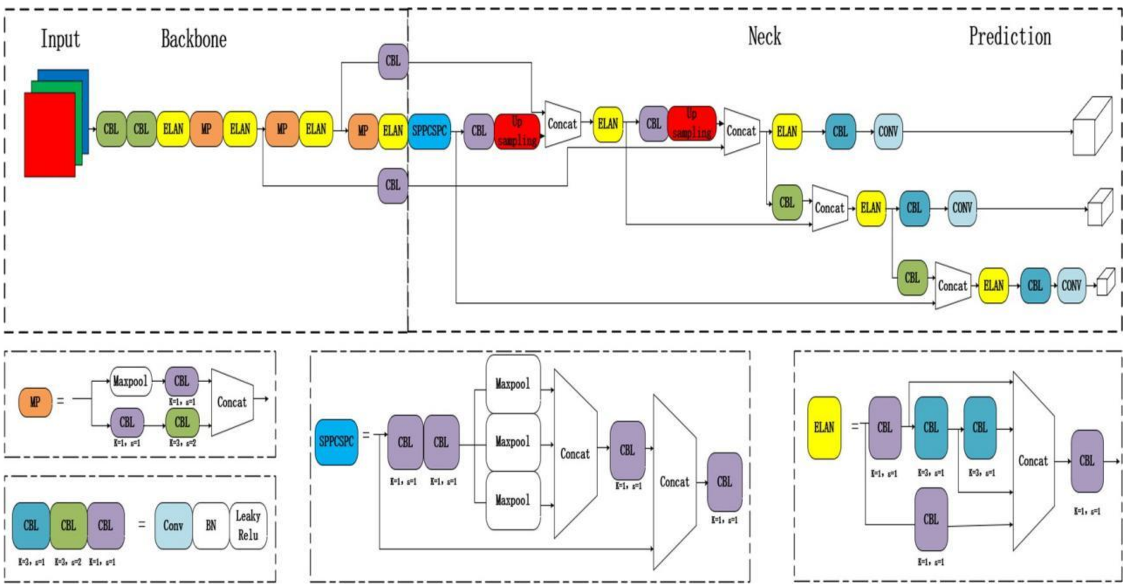

Figure 1. YOLOv7-tiny network model

Alexey Bochkovskiy's team officially released YOLOv7 target detection algorithm in July 2022, based on Pytorch deep learning framework, and the YOLOv7 network is divided into 7 versions, such as YOLOv7, YOLOv7-d6, YOLOv7x, YOLOv7-tiny, etc. Among them, YOLOv7-tiny is the YOLOv7-tiny is the version with the smallest number of network layers and parameters, and the YOLOv7-tiny network model will be used in this paper.

The network model of YOLOv7-tiny is divided into four parts, including Input, Backbone, Neck and Precision, and the network structure is shown in Figure 1.

2.1 Input

The input-side module scales the input image to a uniform pixel size so that it can meet the input size requirements of the backbone network [15].

The input module of YOLOv7-tiny extends the design of YOLOv5 and consists of Mosaic data enhancement, adaptive anchor calculation (Anchor), and adaptive image scaling; Mosaic data enhancement is performed by randomly selecting four images and reassembling them into a single image by random scaling, cropping, and placement operations. It can enrich the background of the detection target and improve the generalization of the network; adaptive anchor frame calculation can initially set the anchor frame, compare it with the labeled real frame, and iteratively update and iterate the network parameters; adaptive image scaling adds black edges to the input image to make the input image size consistent, and calculates the scale, size, and filled black edge size after scaling.

2.2 Bankbone

The backbone network of YOLOv7-tiny consists of several CBL, ELAN, and MP modules.

The structure of CBL is shown in Figure 2. CBL can be regarded as a way to encapsulate convolutional kernels, consisting of CONV, BN (batch normalization), and Leaky Relu (nonlinear activation function), with different colors representing different convolutional kernel sizes and convolutional kernel sliding steps. CBL makes the input feature map go through the convolutional layer, normalization layer, and activation function to get the output.

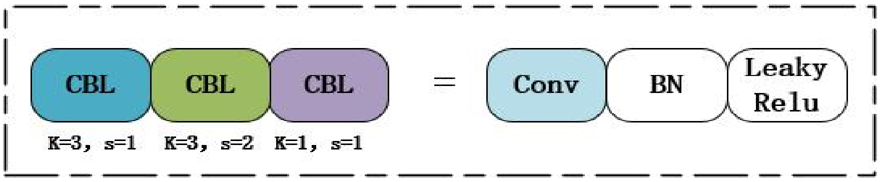

Figure 2. CBL structure

ELAN is an efficient aggregation network, and the structure is shown in Figure 3. ELAN consists of different CBL modules with Concat splicing operation, so that the design can further improve the learning ability and convergence of the network by controlling the shortest and longest gradient paths so that the information between model layers can be effectively propagated.

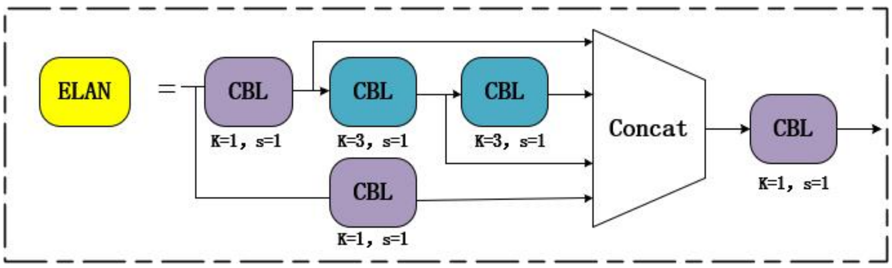

Figure 3. ELAN structure

The MP structure is shown in Figure 4. MP is composed of Maxpool (maximum pooling) and CBL in one way, through Maxpool to complete downsampling, so that the image height and width are scaled by half, and through CBL to halve the image channel; the other way is composed of CBL modules of different scales, the first CBL to halve the image channel, the second CBL to halve the image height and width, and finally the two are spliced to complete feature fusion.

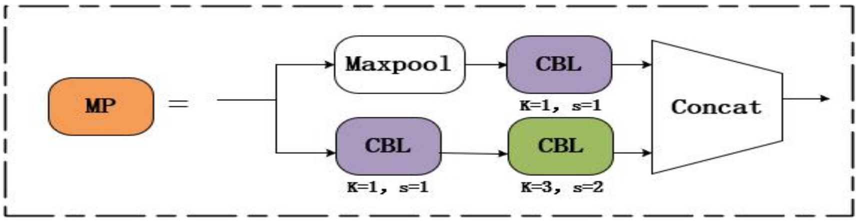

Figure 4. MP structure

2.3 Neck

The neck network of YOLOv7-tiny mainly implements feature fusion, which consists of SPPCSPC module, CBL module, ELAN module, and Concat. SPPCSPC module is the optimized SPP (Spatial Pyramid Pooling) network of YOLOv5, and the SPP structure diagram is shown in Fig. 5. The SPPCSPC structure diagram is shown in Fig. 6. SPP performs multi-scale feature fusion by using maximum pooling of 1×1, 5×5, 9×9 and 13×13. The SPPCSPC module is richer in feature extraction information compared with the previous SPP module. In addition, the YOLOv7-tiny neck network extends the neck network structure design idea of YOLOv5 by using the feature pyramid net-works (FPN) and path aggregation net-work (PAN) [16] structure design. The PAN complements the FPN by adding a bottom-up feature pyramid after the FPN to transfer the strong localization information from the bottom to the top, using the combined structure of PAN and FPN to achieve feature information fusion from different backbone layers to different detection layers to further improve the detection of dense target detection capability.

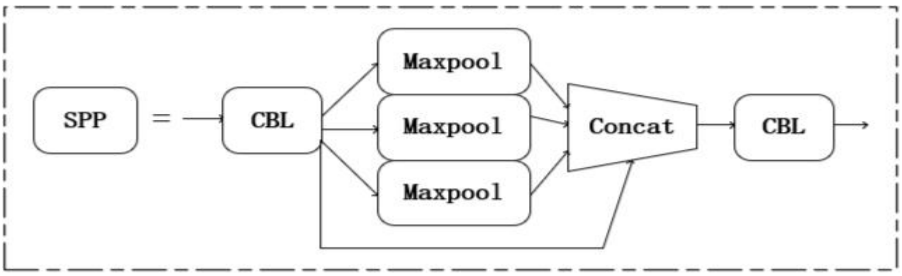

Figure 5. SPP structure

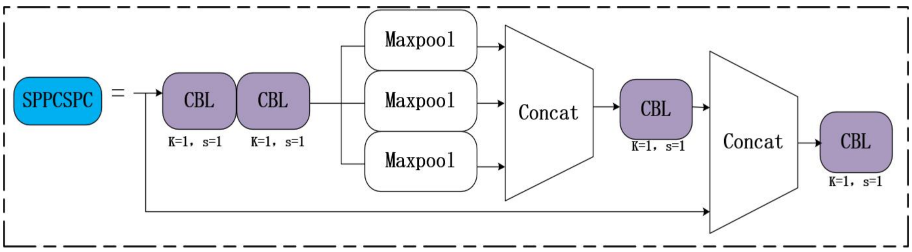

Figure 6. SPPCSPC structure

2.4 Prediction

The YOLOv7-tiny prediction layer has 3 output channels and detects large, medium and small targets after 1×1 ordinary convolutional output, and then performs confidence, category and anchor frame prediction.

In addition, the prediction layer of YOLOv7-tiny is similar to the YOLOv5 prediction layer, including the loss function and the target detection phase. The target detection stage uses NMS non-extreme suppression to filter the target box, the loss function mainly includes the loss function of Bounding box (bounding box), confidence loss, and classification loss, and the prediction layer is used to output the confidence and bounding box of the detection target. Among them, CIOU_Loss is used as the loss function of Bounding box by default in YOLOv7-tiny.

3. Experiments and Analysis

3.1 Data Set Collection

The experimental dataset in this paper uses the publicly available Aquarium marine life dataset, which has 4670 marine life images after data enhancement, and the dataset is labeled by Roboflow's team, including seven marine life images of fish, jellyfish, penguins, sharks, puffins, stingrays, and starfish, and the dataset is divided according to 96% training set, 3% validation set, and 1% test set, and the example figure is shown in Figure 7.

Figure 7. Aquarium marine life example

3.2 Experimental Settings and Model Evaluation Metrics

The experimental operating system is Windows, the CPU is Intel(R) Core(TM) CPU@3.70GHz, the GPU is GeForce GTX 1080 Ti with 11GB video memory, and the deep learning framework is Pytorch. pre-trained weights with initial learning rate of 0.001 and termination learning rate of 0.1 are used for the COCO dataset. The parameters are optimized using Stochastic Gradient Descent (SGD) optimizer.

In target detection, the following parameters are often used as evaluation indicators for the model.

$$\Pr ecision = \frac{TP}{TP + FP} \tag{1}$$

$$\operatorname{Re} call = \frac{TP}{TP + FN} \tag{2}$$

$$AP = \int_0^1 P(\mathrm{r}) d(r) \tag{3}$$

$$mAP = \frac{1}{N} \sum_{i=1}^{c \quad lass} \int_0^1 P(r) d(r) \tag{4}$$

Where Precision is the accuracy, Recall is the recall, AP is the individual category accuracy, TP is the true case (positive for both detection and true), FP false positive case (positive for detection and negative for true), FN is false negative case (negative for detection and positive for true), mAP is the average accuracy, and the number of categories in this paper is 7, which is generally calculated when IOU=0.5 mAP value, that is mAP@0.5.

In addition, this paper also measures the algorithmic model using the model size as well as the computational volume GFLOPs (1 billion floating point operations).

3.3 Training Process

The model is trained using YOLOv7-tiny network for 300 rounds. As can be seen from Figure 8(a) below, the Bounding box loss value decreases rapidly in the first 25 epochs of training; the Bounding box loss curve decreases with a gentle slope from 25 to 250 epochs of training; after 250 epochs of training, the Bounding box loss curve tends to converge smoothly and flatly. As in Fig. 8(b), after about 25 epochs, Precision floats slightly and basically stabilizes at 80%. As shown in Fig. 8(c), Recall floats slightly and stabilizes at 70% after about 25 epoch. As shown in Fig. 8(d), mAP@0.5 also remains at more than 75%.

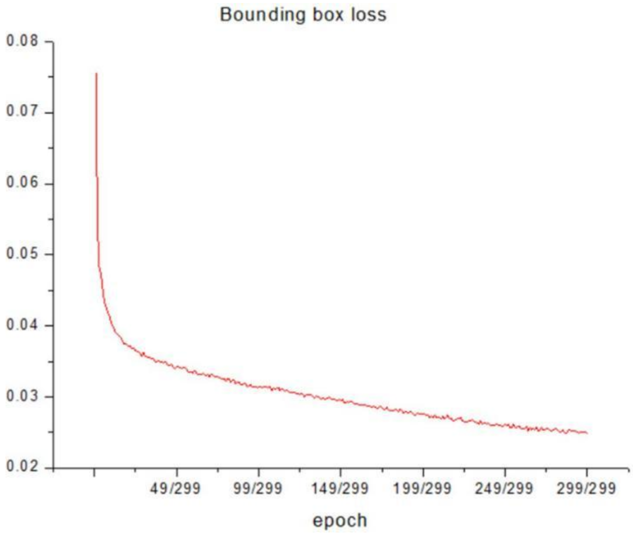

Figure 8(a). Bounding box loss

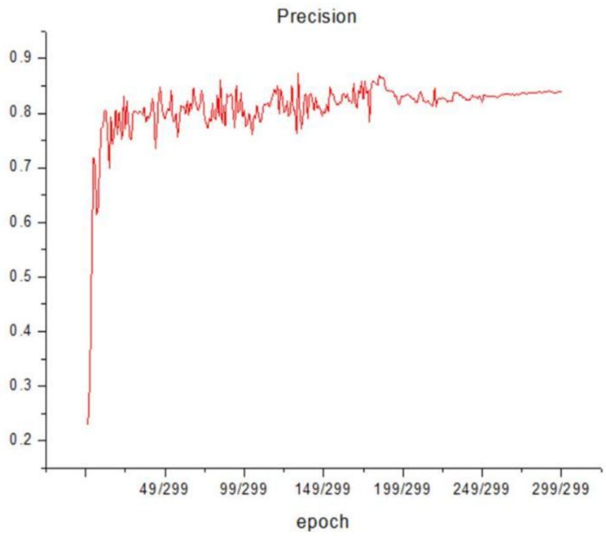

Figure 8(b). Precision

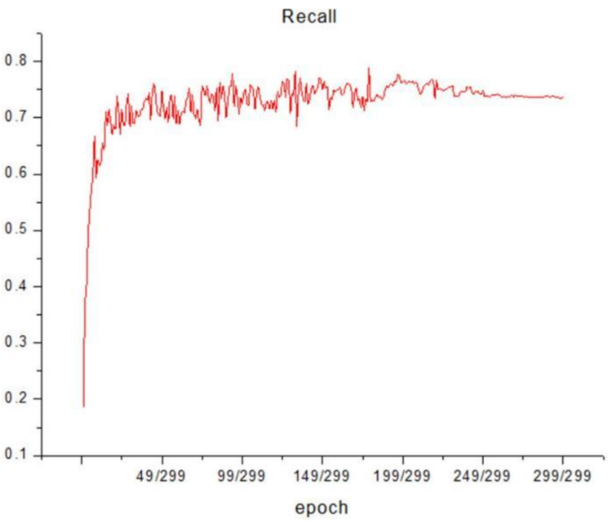

Figure 8(c). Recall

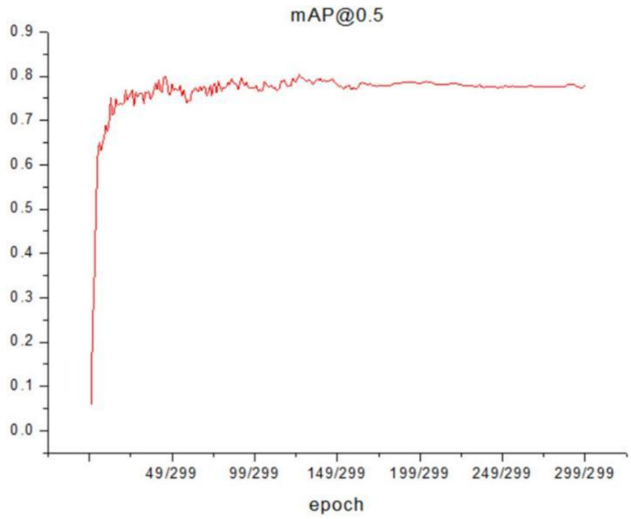

Figure 8(d). mAP@0.5

3.4 Results and Analysis

After the training of the network model, the obtained trained best.pt weight file is tested. Figure 9 shows the visualization of some detection results, with YOLOv5s detection plot at the bottom and YOLOv7-tiny detection effect plot at the top. It can be seen that YOLOv5s shows individual false and missed detection cases, and YOLOv7-tiny detection has a higher confidence score for each target classification, and in complex backgrounds, the multi-target detection case also works better than YOLOv5s.

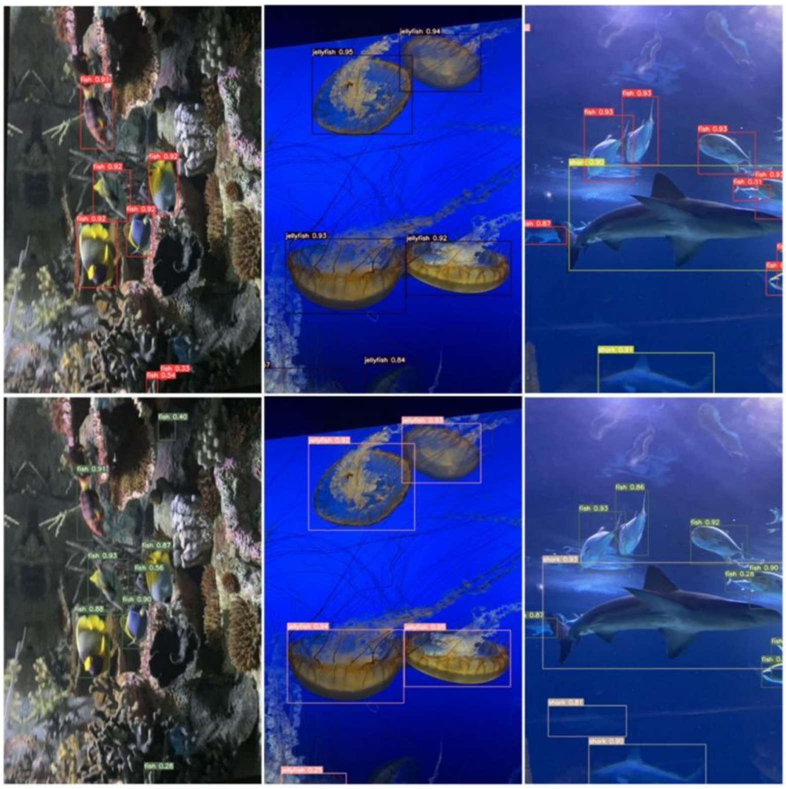

Figure 9. YOLOv7-tiny and YOLOv5s detection comparison

The experimental results using the YOLOv5s algorithm compared with the algorithm in this paper are shown in Table 1. the accuracy, recall, and average precision of YOLOv7-tiny are improved, and the model size and computation are reduced relative to YOLOv5s.

Table 1. YOLOv5s algorithm and this paper's algorithm

Networks	Precision	Recall	Model Size/MB	Calculated volume /GFLOPs
YOLOv5s	0.824	0.722	0.763	14.6
YOLOv7-tiny	0.84	0.78	0.78	12.3

4. Conclusion

In order to achieve target detection and recognition of underwater organisms, this paper proposes a YOLOv7-tiny based target detection algorithm for underwater organisms, and introduces the YOLOv7-tiny target detection algorithm, which is trained on the Aquarium marine organism dataset. And compared with YOLOv5s algorithm, the experimental results show that the accuracy, recall, and average precision of this paper's algorithm are improved, and the model size and number of parameters are relatively reduced. The next step should be considered to improve YOLOv7-tiny by adding attention mechanism, small target detection layer, and changing the loss function, so that its model can be applied to more fields.

References

[1] Y. Guo, X.K. Wang, and H.P. Xu: A review of sonar image-based underwater target detection, identification and tracking research (Control and Decision), Vol. 33 (2018) No. 5, p.906-922.

[2] G.X. Shen, Y. Wei, L.S. Liu, et al. An improved YOLOv5 algorithm for pedestrian fall detection (Mini-computer systems), Vol. 9 (2023), p.1-9.

[3] S.Q. Shen, K.M. He, J. Sun, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks (IEEE transactions on pattern analysis and machine intelligence), Vol. 39 (2017), p.6.

[4] K.M. He, G. Georgia ,G. Ross, et al. Mask R-CNN, IEEE transactions on pattern analysis and machine intelligence, Vol. 42 (2020), 2-6.

[5] Huang, G., Shen, A., Hu, Y., Du, J., Hu, J., & Liang, Y. (2024). Optimizing YOLOv5s Object Detection through Knowledge Distillation Algorithm. arXiv preprint arXiv:2410.12259.

[6] Y. Wang, L. Wang, Y. Jiang, et al. Detection of SelfBuild Data Set Based on YOLO4 Network, IEEE3rd International Conference on Information Systems and Computer Aided Education (Dalian ,China, 2020), 640-642.

[7] C.Y. Wang, A. Bochkocskiy, H.Y.M. Liao: YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors(2022).

[8] P.Y. Shi, S. Han, J.J. Ni, et al. Combine data to enhance and improve YOLOv4's underwater target detection algorithm, Journal of Electronic Measurement and Instrumentation, Vol. 36 (2022) No.4, p.113-121.

[9] W. Qiang, L.Y. He, Y.J. Guo, et al. Research on underwater target detection algorithm based on improved SSD (Journal of Northwestern Polytechnical University), Vol. 38 (2020) No.4, p.747-757.

[10] Y.H. Huang, X.Y. Gao, C.D. Huang, et al. Research on underwater target detection algorithm based on FAttention-YOLOv5 (Microelectronics and Computers), Vol. 39 (2022), No. 6, p.60-68.

[11] M. Sung, S.C. YU, Y. Girdhad. Vision based real-time fish detection using convolutional neural network(Acta Oceanologica Sinica), Vol. 17 (2017), No. 3, p.1-6.

[12] K. Cai, X. Miao, W. Wang, et al. A modified YOLOv3 model for fish detection based on MobileNetv1 as backbone (Aquacultural Engineering), Vol. 91 (2020), No.2, p.102-117.

[13] H. Yang, P. Liu, Y.Z. Hu, et al. Research on underwater object recognition based on YOLOv3 (Microsystem Technologies), Vol. 27 (2021), No. 4, p.837-1844.

[14] L. Chen, M. Zheng, S. Duan, et al. Underwater Target Recognition Based on Improved YOLOv4 Neural Network(Electron-ics), Vol. 10 (2021), No. 14, p.1634.

[15] L. Cheng, C. Jing: Rotational target detection in X-ray images based on improved YOLOv7 (Journal of Graphics), Vol. 1 (2022), No. 1, p.1-12.

[16] S. Liu, L. Qi, H. Qin, et al. Path Aggregation Network for Instance Segmentation, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, (2018), p.8759-8768.