Underwater Target Detection Based on Improved YOLOv7

Junshang Fu, Ying Tian

Abstract—Underwater target detection is an important part of marine exploration. However, in complex underwater environments due to factors like light absorption and scattering, as well as variations in water quality and clarity. These challenges result in inaccurate target feature extraction, sluggish detection speeds, and insufficient robustness in the detection methods. In order to address these issues, an enhanced YOLOv7 network (YOLOv7-SPNW-D) is proposed for underwater target detection in this study. The SPD-MP module structure replaces the MP module in the neck network to capture small targets and enhance detection accuracy. A novel NWD loss function is employed to facilitate smoother extraction of small target features. This enhances feature extraction and improves network inference speed. Additionally, incorporating a small target detection module enables the providing of more comprehensive small target information within a deep feature map. This, in turn, improves the capture of small target features in complex backgrounds, and avoids feature loss and enhancing model exactness. Through ablation experiments on the URPC dataset, it is shown that the improved YOLOv7-SPNW-D algorithm performs better than the original YOLOv7 algorithm, with the mAP50 value increased to 87.0%, proving the effectiveness of this method. In conclusion, the improved YOLOv7-SPNW-D model is more suitable for underwater marine organism target detection.

***Index Terms*—underwater target detection, marine resources, YOLOv7, small target**

I. INTRODUCTION

Our marine industry has entered a brand new stage of booming development, but the ocean is different from land areas. First of all, the Acquire optical images in complex commonly found underwater environments have serious blur, occlusion, degradation, and other problems, This greatly affect the dependability of underwater object recognition. The aim of underwater object detection [1] is to locate and identify targets in underwater scenes with high accuracy. In order to achieve this goal, the dataset can be collected using underwater equipment such as remote control operating vehicles (ROVs) [2] and autonomous underwater vehicles (AUVs) [3]. The success of this study not only makes a big step forward in underwater target detection in China, but also locates and identifies specific small scale aquaculture organisms according to the particularity of the dataset. After success, it provides convenience for the majority of seafood aquaculture merchants and facilitates aquaculture merchants to monitor aquatic products in real time in specific areas. It is convenient to detect the density of aquatic products grown, the preferred water area range, and determine whether the disease state occurs through the density in underwater, which greatly facilitates aquatic product aquaculture merchants, saves manpower and material resources, and realizes the recycling of resources. Underwater target detection instead of human underwater biological monitoring and fishing activities, deep sea research and shallow aquaculture [4] also have a significant role.

However, as we all know, various problems will be encountered in underwater scenarios. Firstly, in terms of optical images, absorption and scattering related to underwater wavelengths significantly reduce the quality of the underwater images obtained, which leads to low precision in detection of small objects, incorrect detection and missed detection. Secondly, in terms of the underwater environment, it is also affected by complex scenarios such as temperature, water source, and visibility in the ocean, which can increase the difficulty of underwater robot surveys [5]. Therefore, we have to consider the generalization capability of target detection in different environments, that is, models are trained in one domain and evaluated in another domain [6]. Finally, after obtaining optical images through remote controlled vehicles and underwater vehicles, there may also be some problems in the images, such as occlusion by plankton, overlap and blurring of different species of organisms. These image problems adversely affect performance of target detection. Classic target detection systems that rely on manual characteristic engineering include scale invariant feature transform (SIFT) [7], Histogram, Oriented Gradients (HOG) [8], etc. However, traditional target detection has certain limitations due to the use of manual features. Therefore, methods based on deep learning are gradually emerging.

Although deep learning techniques have great success in computer vision, existing target detection technology in the ocean still encounters several obstacles stemming from the intricacies of underwater scenes.

Based on the above problems, such as a fuzzy underwater environment, small target detection accuracy, and occlusion overlap, a method based on improving YOLOv7 target detection is proposed in this paper. On this basis, it is applied to marine organisms, thus supporting human underwater operations and fully locating and identifying small target organisms [9]. In this paper, the depth hierarchical processing transformation of the input pictures involves

Manuscript received October 13, 2023; revised February 28, 2024.

Junshang Fu is a postgraduate student majoring in software engineering at School of Computer Science and Software Engineering, University of Science and Technology Liaoning, Liaoning 114051, China. (e-mail: fjs2284781073@163.com).

Ying Tian is a professor of School of Computer Science and Software Engineering, University of Science and Technology Liaoning, Anshan, 114051, China. (corresponding author to provide phone: +8613898015263; e-mail: astianying@126.com)

adding an SPD module to the neck network, and applying the SPD module's molecular mapping principle to extract small targets from blurred images. This approach aims to reduce missed detections of small targets. By adding a new measure of an optimized loss function (NWD) to address image deviation, occlusion, and other issues, the speed of inference is improved. Furthermore, the addition of a tiny target detection module [10, 11] further strengthens the detection of small targets of marine organisms and enhances the feature extraction ability.

II. RELATED ALGORITHMS

YOLOv7 [12] was developed by Chien and Alexey et al. in 2022 and is a classical representative of the one-stage target detection algorithm. The current YOLOv7 model exhibits higher accuracy compared to earlier models in the YOLO series. Furthermore, when evaluated on the URPC dataset, the YOLOv7 model demonstrates superior accuracy over the recently released YOLOv8 model. The YOLOv7 model integrates E-ELAN (Extended efficient layer aggregation networks) [13], which is incorporated into the Backbone and Neck networks. A good balance between detection efficiency and accuracy is achieved through strategies such as connection-based tactical model scaling [14] and model reparameterization [15]. The YOLOv7 network mainly consists of an Input, Backbone, Neck, and Head.

The Backbone (backbone network) is primarily composed of convolutions, E-ELAN MPConv and SPPCSPC modules. The E-ELAN (Extended ELAN) module, building upon the original ELAN, it modifies the computational block configuration while preserving the structure of the transition layer, integrating the concepts of Expand, Shuffle, and Merge to improve the network's learning abilities without sacrificing the unaltered gradient trajectory. The SPPCSPC module prevents distortion by incorporating Integrate parallel MaxPool operations into a consecutive series of convolutions, addressing the issue of convolutional neural networks extracting redundant image features. In the MPConv, the MaxPool operation enhances the receptive field of the current feature layer, subsequently integrating it with the feature information resulting from standard convolution, thereby improving the network's generalization capabilities. Initially, the input image undergoes feature extraction in the backbone network, resulting in a feature layer that serves as the input image's feature set. Utilizing three feature layers, the backbone facilitates the construction of the subsequent network, enabling the derivation of an effective feature representation.

Neck module, traditional PAFPN structure is adopted. FPN uses three effective feature layers obtained from the backbone structure for feature fusion, achieving information integration between different feature layers and further processing the effective feature layers to construct the enhanced feature extraction network in YOLOv7. Through this series of operations, the network is able to extract more comprehensive features for object detection.

In the head section, YOLOv7 selects the IDetect head which represents three sizes of targets of large, medium and small. The YOLO Head serves as both the classifier and regressor for YOLOv7, performing both tasks simultaneously. Through Backbone and FPN, three enhanced effective feature layers can be obtained. Each layer of features possesses a width, a height, and a number of channels. we can view the feature map as a set of feature points after feature points. Each feature point has three prior frames, and multiple channel features are present in each preceding frame. In fact, the YOLO Head performs the function job of judging the characteristic points and judging whether the prior frame on the feature points has objects corresponding to them. Similar to the earlier version of YOLO, YOLOv7 also employs coupled decoupling heads where classification and regression are executed using a 1x1 convolution.

Predicted Network: YOLOv7 predictive network uses Rep structure is used to modify the number of image channels in the feature output by the head network. and then applies 1×1 convolution to predict confidence. The concept of the Rep structure drew inspiration from RepVGG [16] and a unique residuals design was introduced to facilitate the training process. During the model prediction process, the effects formed by these residual structures combinations can theoretically be simplified into a complex convolution operation, thereby reducing network complexity without sacrificing its predictive performance.

Therefore, YOLOv7 model is a typical object detection model, which has faster and more effective network and feature integration methods. Compared with the previous YOLO series version, YOLOv7 model has more accurate detection methods, more effective label matching methods and training methods, which provides a solid basis for detecting target.

In addition, the scarcity of datasets for underwater biological object recognition hinders the progress of underwater biological object detection, Obstacles exist to the research on underwater marine organism detection. In comparison to traditional detection on land, the history of underwater object detection is relatively brief. In 2017, Zhou et al. [17] by applying image enhancement techniques in the VGG16 network and utilizing the Faster R-CNN network for feature mapping processing on the URPC dataset, target detection and recognition of organisms are achieved. Moreover, Chen et al. [18] introduced a new weighted loss function called IMA in 2020, this approach improves detection performance by mitigating the impact of noise on detection. In 2021, Qiao et al. [19] A real-time underwater object classifier has been proposed specifically for the classification of underwater marine organisms, which is more suitable for underwater environments. The combined requirement for precise positioning and categorization poses significant challenges in underwater target detection tasks due to significant color deviation and limited visibility, often caused by moving acquisitions. In order to achieve better results in underwater target detection, we selected YOLOv7 for improvement based on theoretical analysis, proposed the YOLOv7-SPNW-D algorithm, optimized known issues, and demonstrated the effectiveness of this algorithm through ablation experiments and comparative experiments.

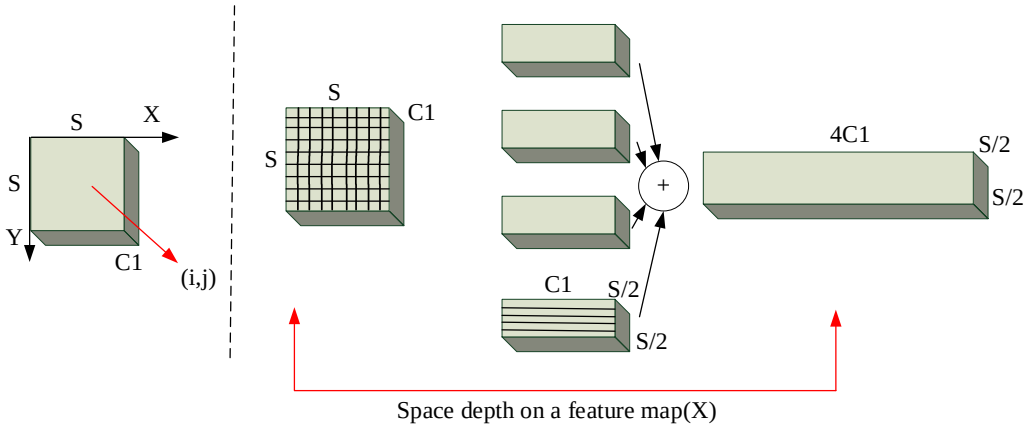

Fig. 1. SPD model

III. IMPROVEMENTS

A. MP-SPD Module

In the case of dim underwater light and blurred images, it is difficult to accurately distinguish the redundancy of small target pixels in the image in the neck network. When Conv_BN_SiLU is applied repeatedly, the original YOLOv7 structure filters out a large number of small targets and key features together. To reduce the loss of small target features, we replace Conv_BN_SiLU in the left branch of the MP structure in the sampling structure under the neck with the SPD module shown in Fig. 1 below. The SPD module achieves the spatial dimension reduction to the channel dimension by obtaining feature mappings from the upper-layer input and retains the information in the upper-layer channels, which is achieved by mapping each pixel or feature of the inputted feature representation to a channel. It maps the global spatial information of small targets to the channel dimension, achieving a transformation of the depth layer. This achieves the goal of preserving smaller target feature information in dim and fuzzy underwater scenes. The modified MP-SPD is shown in Fig. 2 below.

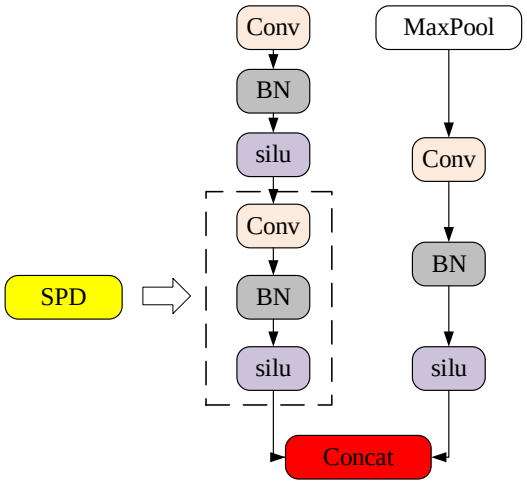

Fig. 2. MP-SPD network structure

B. Minor Target Detection Module

This article uses the underwater robotic target grasping competition (URPC) dataset, which is composed of underwater images collected by autonomous underwater vehicles (AUV) [20] at Zhangzidao ocean ranch in Dalian. As the dataset accurately reproduces the real ocean environment, some target images may not be clear or may be far away, making it difficult to train the model. The deep prediction feature layer may also fail to capture small targets.

To address these challenges, this article introduces a fourfold down sampling branch to the neck network, which is an additional up sampling module that increases image resolution. In the neck network, input resolution 160×160 fourfold down sampled feature maps are horizontally connected with 8 fold, 16 fold, and 32 fold down sampled feature maps in the PAFPN architecture to achieve multi-scale fusion.

The URPC dataset features low-pixel foreground targets that appear at a considerable distance against a vast background. Therefore, the problem of distant targets is solved by introducing an extra upsampling module in the neck network, aiming to fuse multilevel information. The output is directed to the fourth detector, which corresponds to its fourfold down sampling. This detector, combined with the other three detectors in the model, achieves multi-scale detection by altering the network hierarchy structure. This approach reduces the issue of detecting small distant targets in images and enhances the stability of the model.

The feature maps obtained through fourfold down sampling contain abundant small targets and their texture and detail information. By passing and fusing these maps, the model gains access to more comprehensive small target data, thus increasing the capture of small target features in complex backgrounds. Fig. 3 illustrates the process of generating a feature map with improved expressive power while maintaining scale matching by adding an extra layer for upsampling at the end of neck feature pyramid's sampling structure. Similarly, in the PANet structure [21], a down sampling structure is included, and robust localization features are relayed to lower levels to maintain scale matching.

Overall, the tiny target detection module enhances the effectiveness of multi scale feature fusion and bolsters the robustness of detection scales, significantly improving the model's accuracy in detecting tiny targets.

C. Normalized Wasserstein Distance

Aiming at the problem that marine organisms in URPC underwater datasets are highly responsive to slight

positional shifts in small targets compared with traditional IoU metric calculation, this paper proposes to introduce NWD (Normalized Wasserstein Distance) in the regression loss function to measure the similarity amongst the forecasted boundary frame and the real target boundary [22]. Because the original IoU metric calculates the degree of overlap, the representative methods are GIoU [23], CIoU [24], EIoU [25], etc. In addition, due to the nature of IOU, it is difficult to find a good threshold for the model to provide high-quality object detectors for samples. In this paper, because the underwater robots acquisition of samples may lead to less small target pixels in the underwater datasets, and the scale change is discrete, because the minor positional shift will lead to a significant reduction of IoU value, therefore, we use a new method to calculate the similarity of target bounding boxes and use a new metric to replace the original IOU. Through this method, the target bounding boxes are transformed into two-dimensional Gaussian distributions, and then the similarity of objects after the Gaussian distribution is measured through normalization within NWD. For the horizontal bounding region R=(c_x,c_y,w,h), where (c_x,c_y) is the center point coordinate and w and h are the width and height respectively R is represented by a normal distribution N (μ,∑), where:

$$\mu = \begin{bmatrix} c_x \\ c_y \end{bmatrix}, \quad \sum = \begin{bmatrix} \frac{w^2}{4} & 0 \\ 0 & \frac{h^2}{4} \end{bmatrix} \tag{1}$$

The wasserstein distance was then used to calculate the two Gaussian distribution distances between the bounding box R1=(c_{x1},c_{y1},w_1,h_1) and the bounding box R2=(c_{x2},c_{y2},w_2,h_2).

$$W_2^2(N_1,N_2) = \left\| \left(c_{x1}, c_{y1}, \frac{w_1}{2}, \frac{h_1}{2} \right), \left[c_{x2}, c_{y2}, \frac{w_2}{2}, \frac{h_2}{2} \right]^T \right\|_2^2 \tag{2}$$

Because this distance metric cannot directly measure the similarity between bounding box R1 and R2, it is exponentially normalized to obtain a new metric NWD:

$$NWD(N_1,N_2) = exp\left(- \frac{\sqrt{w_2^2(N_1,N_2)}}{C} \right) \tag{3}$$

Here, c is a constant related to the dataset. In comparison with the IoU rating, NWD can respond to the change of target position more smoothly. Without consideration of target overlap, the distribution similarity is quantifiable; and NWD is less sensitive to targets of different sizes and is more appropriate for assessing small-scale targets.

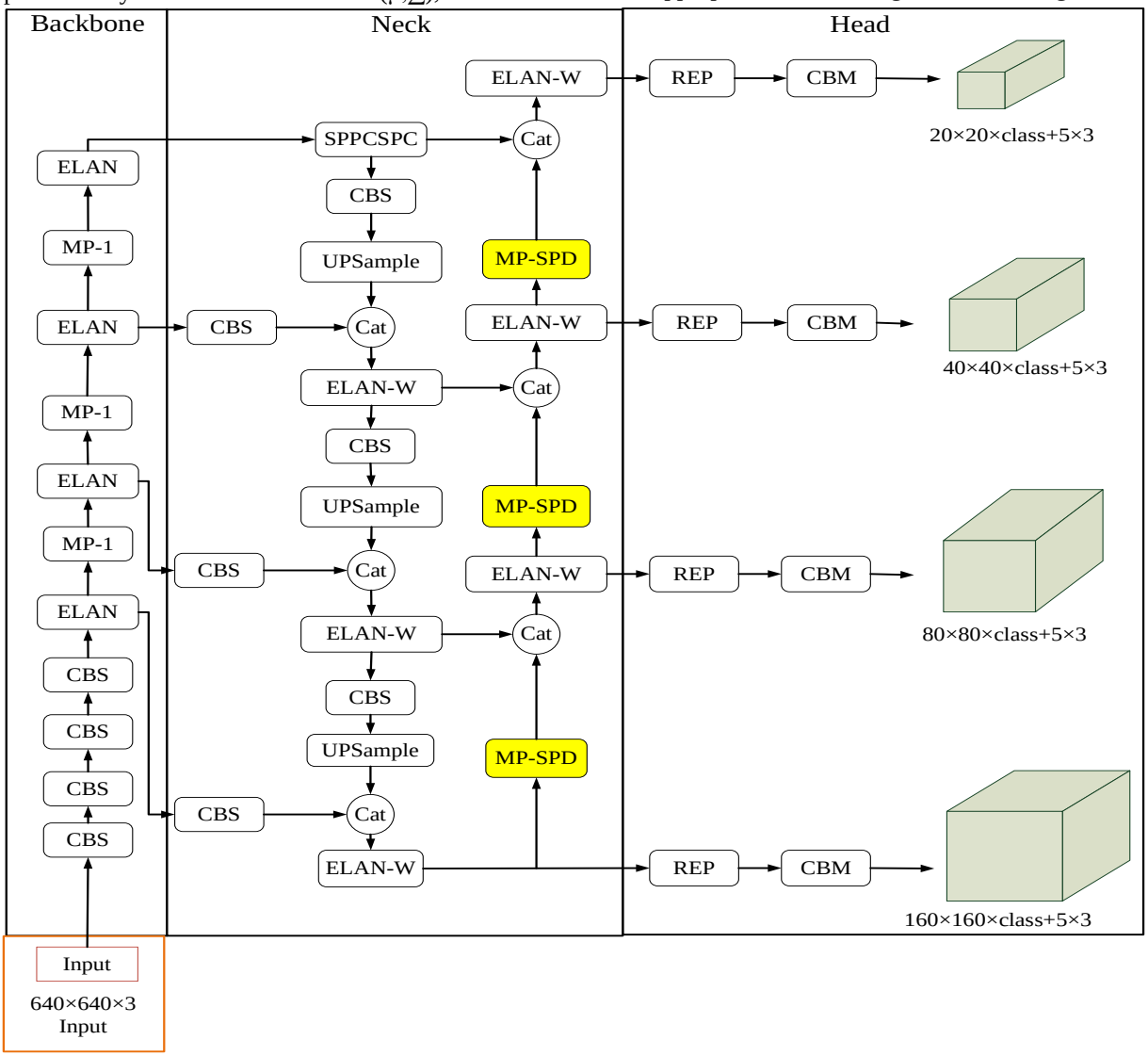

Fig. 3. Structure diagram of the YOLOv7-SPNW-D model

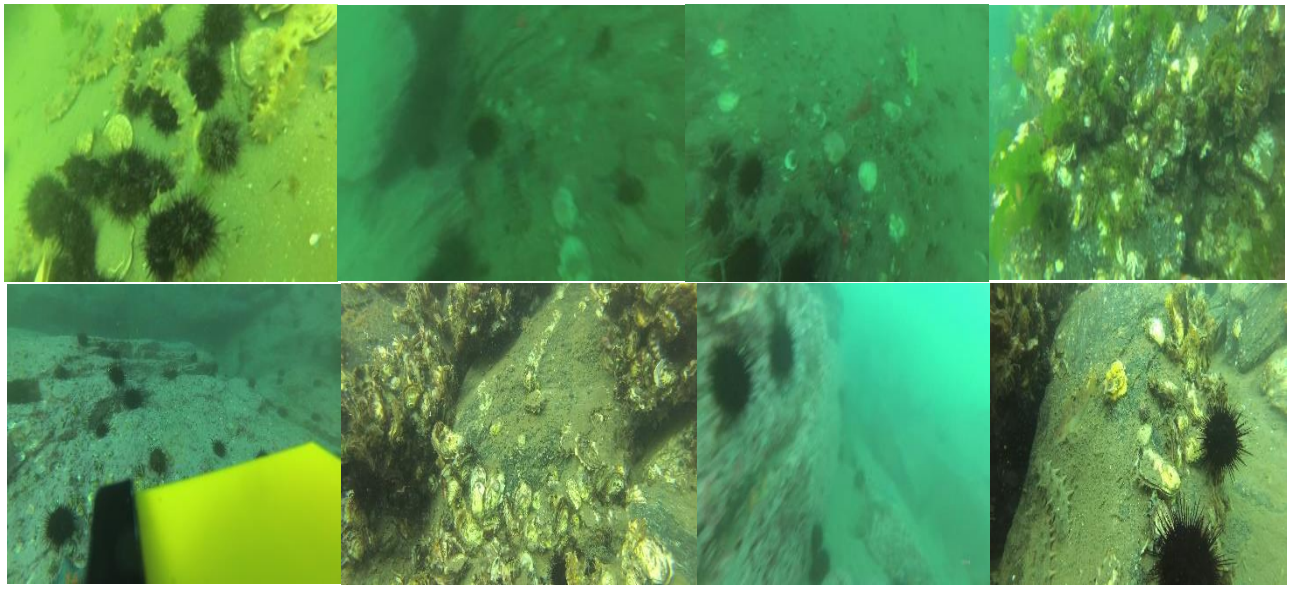

Fig. 4. The URPC marine organism dataset

The improved YOLOv7 algorithm structure replaces the MP module with the SPD-MP module in the original YOLOv7 neck network, making the neck network more stable, Increase image resolution, speeding up model reasoning, and enriching the features extracted by the network. At the same time, it adds a quadruple down sampling branch and a fourth target detection head, enhancing the effect of multi scale feature fusion without significantly increasing the computational load, improving the robustness of detection scale and the accuracy of small target detection. The loss function in the network is replaced with a new metric called the NWD loss function, replacing the original IoU loss function improving the detection capability for small objects and improve detection accuracy. The improved Yolov7 network structure is shown in Fig. 3.

IV. EXPERIMENTS

A. Experiment Environments

Hardware environment: NVIDIA GeForce RTX 4070 Ti, 12 GB video memory, Software environment Windows 10, CUDA 11.3, Pytorch 1.11.0, Python 3.8.0. Parameter setting input image resolution is 640×640, total iteration number is 300, iteration batch size is 4, optimizer is SGD, momentum is 0.937, learning rate is 0.01, weight attenuation coefficient is 0.0005, and the learning rate is updated by cosine annealing learning algorithm.

B. Datasetsets and Settings

In this paper, we selected the underwater dataset from the Underwater Vehicle Target Grasping Contest (URPC). The dataset is composed of underwater images collected by the autonomous unmanned underwater robot (AUV) at the marine pasture of Zhangzi Island, Dalian, and it accurately represents the real marine environment. Because the competition test set is not publicly available, we selected 6671 pictures containing four types of marine organisms: sea urchin, sea cucumber, sea star, and scallop. With their corresponding labels in the dataset being echinus, holothurian, starfish, and scallop, an illustration of certain exemplary images from the URPC dataset is presented in Fig. 4. To create the experimental dataset, we established a 7:1:2 ratio of training set, validation set, and test set with the training set containing 4669 pictures, the validation set containing 667 pictures, and the test set containing 1335 pictures, which were randomly divided.

C. Criteria for Assessing Model Performance

The standard evaluation metrics commonly utilized for object detection encompass Precision (P), Recall (R), Intersection over Union (IoU), Average Precision (AP), and mean Average Precision (mAP) [26], weighted harmonic average F1, parameter count (Params).

And computational complexity measured by FLOPs as comprehensive evaluation metrics for assessing the effectiveness of underwater marine organism detection. Specifically, Using *TP*, *FP*, and *FN* to indicate whether the number of marine organisms in the current URPC dataset is detected, the *AP* value is then represented by the area under the precision-recall curve, which leads to the calculation of the mean Average Precision (mAP) value. They were calculated as follows:

$$P=\frac{TP}{TP+FP} \tag{4}$$

$$R=\frac{TP}{TP+FN} \tag{5}$$

$$AP=\int_0^1 P(R)dR \tag{6}$$

$$F1=\frac{2PR}{P+R} \tag{7}$$

$$mAP=\frac{1}{n}\sum_{j=1}^{n} AP(j) \tag{8}$$

Among them, *TN* is true negative, N is the number of species of marine organisms in the dataset; *AP*(*j*) indicates the AP of class *j*. To quantify the detection speed, we employ the measure of FPS, which signifies the rate at which images can be processed within a given second. On

the other hand, the complexity of the model is reflected through the number of its parameters. Below is the specific formula used for this calculation:

$$Params=C_0\times(k_w\times k_h\times C_i+1) \quad (9)$$

The notation C_0 represents the number of output channels, C_i represents the number of input channels, while k_w and k_h respectively denote the width and height of the convolution kernel.

D. Experimental Results and Analysis of the URPC Dataset

The experimental evaluation of the detection performance of the proposed YOLOv7-SPNW-D model was conducted on the URPC dataset, focusing on the generation of P-R curves. As demonstrated in Fig. 5. The outcomes of the enhanced model exhibit an improvement in the detection efficiency across all target categories, and the mAP of this model is calculated to be 87.0%. The performance of the proposed YOLOv7-SPNW-D model was evaluated using a confusion matrix, where each column indicates the predicted distribution of each class and each row represents the actual distribution of each class in the dataset, as illustrated in Fig. 6. From the analysis of Fig. 6, it can be seen that the prediction accuracy rates of "echinus", "holothurian", "scallop" and "starfish" are 95.0%, 87.0%, 70.0% and 93.0%, respectively, indicating that this model has high accuracy.

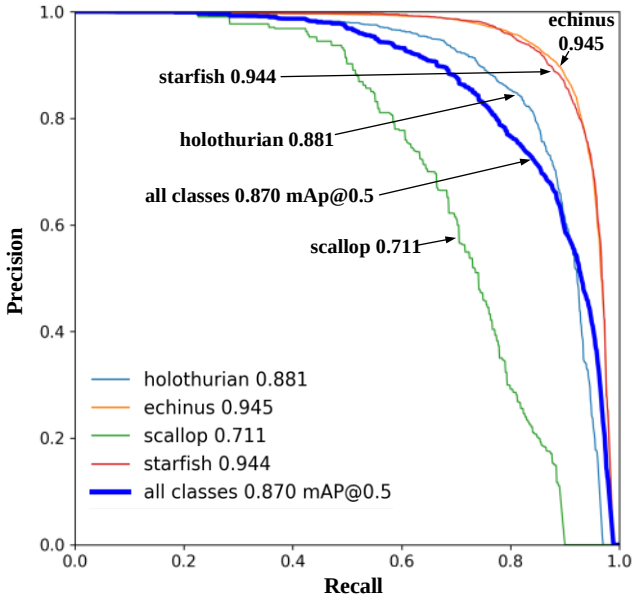

Fig. 5. The precision recal curve of the YOLOv7-SPNW-D model on the URPC dataset

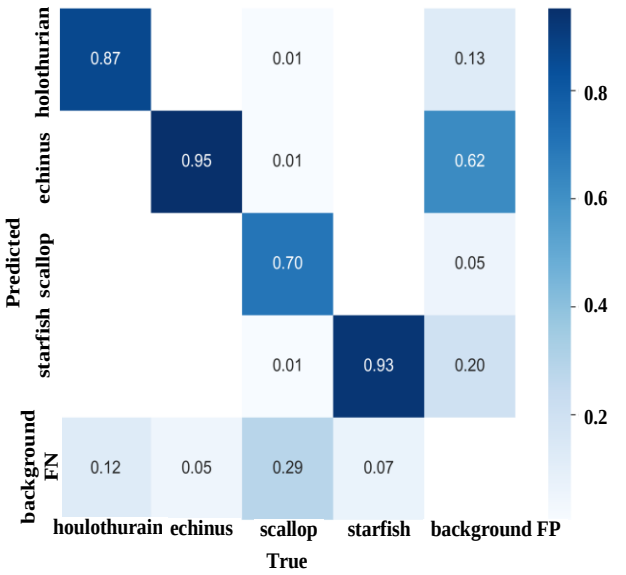

Fig. 6. The confusion matrix of the YOLOv7-SPNW-D model on the URPC dataset

E. Experimental Results and Analysis of the URPC Dataset

In order to further prove the superiority of the proposed YOLOv7-SPNW-D model, the training and testing were performed on the URPC dataset, and its mean Average Precision (mAP) and other evaluation indicators were compared with the currently popular YOLOv5s, YOLOv6n, YOLOv7, YOLOv8l and comparison results with other target detection models are presented in Table I. From the table, it can be seen that the mAP of YOLOv7-SPNW-D model is 1.9% higher than YOLOv7 and 4.1%, 4.5%, and 4.8% higher than YOLOv6, YOLOv5s, and YOLOv8, respectively, which is superior to other detection algorithms. The experimental outcomes demonstrate the practical superiority of this approach in underwater target recognition.

TABLE I
THE PERFORMANCE OF EACH NETWORK ON URPC DATASET

Methop	Precision	Recall	mAP@0.5	mAP@0.9
YOLOv5s	85.7%	75.5%	82.2%	64.8%
YOLOv6n	85.1%	76.1%	82.5	63.7%
YOLOv7	84.5%	76.7%	85.1%	64.8%
YOLOv8l	88.2%	74.2%	82.9%	67.1%
YOLOv7-SPNW-D	85.6%	79.2%	87.0%	68.5%

TABLE II
EVALUATION OF MODEL PERFORMANCE IMPROVEMENT THROUGH ABLATION STUDIES ON THE URPC DATASET.

Model	SPD-MP	NWD	Quadruple	AP(echinus)	AP(holothurian)	AP(starfish)	AP(scallop)	mAP
YOLOv7	×	×	×	94.0%	86.4%	94.1%	66.0%	85.1%
	√	×	×	94.4%	86.4%	94.3%	66.6%	85.4%
	√	√	×	93.8%	87.5%	93.8%	69.4%	86.1%
	√	√	√	94.5%	88.1%	94.4%	71.1%	87.0%

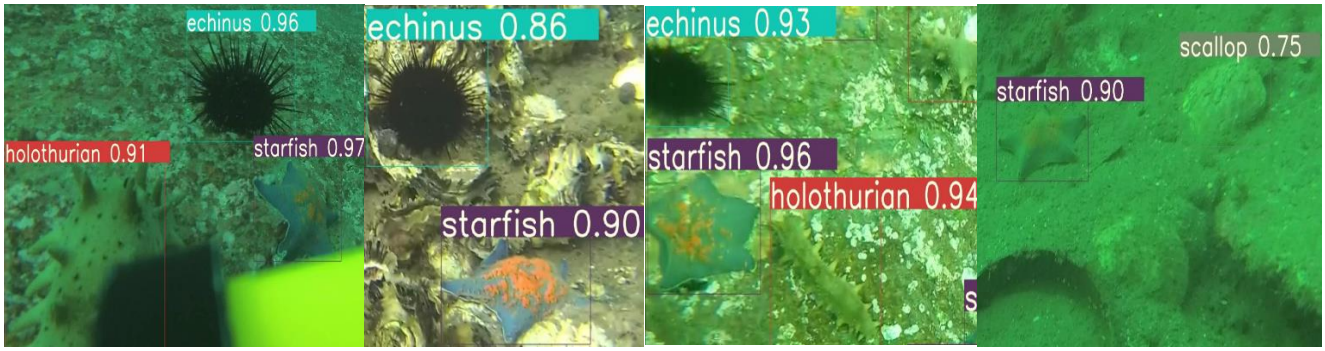

Fig. 7. The inference result on YOLOv7

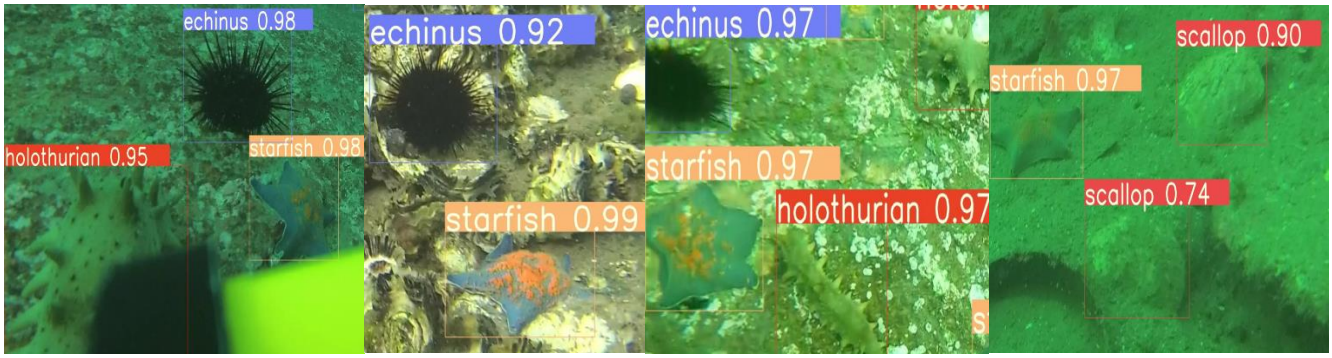

Fig. 8. The inference result on YOLOv7-SWNW-D

F. *Ablation Experiments of the URPC Dataset*

The impact of various enhancement techniques on model performance through ablation experiments. In this paper, the designed SPD-MP module is first added to the YOLOv7 model to observe its AP value and mAP value, and then the CIoU loss function of the new measure of NWD loss function is added to the YOLOv7 model to observe its AP value and mAP value. Finally, the small target detection head is introduced to observe its AP value and mAP value. The final experimental the findings are presented in Table II.

As evident from Table II, we can observe the use of SPD-MP module increases the mAP value by 0.3%. On this basis, the new loss function metric NWD-CIoU is used to increase the mAP of this model by 0.7%. Finally, mAP was also raised by 0.9 on the basis of preliminary experiments by introducing a fourfold lower sampling target detection head.

G. *Inference Result*

Fig. 7 and 8 display the inference outcomes on the URPC test set, showcasing the results generated by YOLOv7 and YOLOv7-SPNW-D, respectively.

This can be clearly seen in Fig. 8. The confidence of echinus, starfish, holothurian and scallop was significantly improved in the YOLOV7-SPNW-D model, and it can be seen from Fig. 7 and 8 that scallop had missed detection in the YOLOv7 model. The missed scallop was detected in the YOLOv7-SPNW-D model, and it can be inferred that the YOLOv7-SPNW-D not only provides higher confidence, but also provides higher accuracy when detecting underwater organisms.

V. CONCLUSION

In this study, we introduce an enhanced version of YOLOv7, named YOLOv7-SPNW-D. To enhance the detection of smaller targets, we have incorporated a small target detection head module and an SPD module into the network architecture. Furthermore, we have introduced a new metric known as the NWD loss function, which aims to enhance detection performance, correct positional deviations of small targets, and address the impact of underwater occlusion and blurring on accuracy. These enhancements collectively contribute to the overall resylts improvement of the model.

To assess the effectiveness of YOLOv7-SPNW-D, we conducted experiments using the URPC dataset and compared its performance with other popular target detection algorithms. The results demonstrate that YOLOv7-SPNW-D surpasses current state-of-the-art models in terms of robustness and detection accuracy in challenging underwater environments, therefore, this model has a wider potential for development.

REFERENCES

[1] Y. Guo, H. Li, and P. Zhuang, "Underwater image enhancement using a multiscale dense generative adversarial network," *IEEE J. Ocean. Eng*, vol. 45, no. 3, pp. 862-870, Jul. 2020.

[2] J. J. Leonard and A. Bahr, "Autonomous underwater vehicle navigation," in *Springer Handbook of Ocean Engineering*. 2016, pp. 341-358.

[3] J. C. Kinsey, M. R. Eustice, and L. L. Whitcomb, "A survey of underwater vehicle navigation: Recent advances and new challenges," in *Proc. IFAC Conf.* Manoeuver Control Mar. Craft, vol. 88, 2006, pp. 1-12.

[4] M. J. Er, J. Chen, Y. Zhang, "Research challenges, recent advances, and popular datasets in deep learning-based underwater marine object detection: A review," *Sensors*, vol. 23, no. 4, p. 1990, Feb. 2023, doi: 10.3390/s23041990.

[5] J. Wang, J. Du, J. Zhuang, "CA-GAN: Class condition attention GAN for underwater image enhancement," *IEEE Access*, vol. 8, pp. 130719-130728, 2020.

[6] H. Liu, P. Song, and R. Ding, "Towards domain generalization in underwater object detection," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Piscataway, NJ, USA, Oct. 2020, pp. 1971-1975.

[7] D. G. Lowe, "Distinctive image features from scale-invariant key points," *Int. J. Comput. Vis*. vol. 60, no. 2, pp. 91-110, Nov. 2004.

[8] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recogniton (CVPR)*, San Diego, CA, USA, Jun. 2005, pp. 886-893.
[9] M. Shobana, V. S. Meha, and D. Neeraj, et al., "AI Underwater drone in protection of waterways by relating design thinking framework," in *Proc. Int. Conf. Comput. Commun. Inform. (ICCCI)*, Coimbatore, India, Jan. 2023, pp. 1-4, doi:10.1109/ICCCI56745.2023.10128399.
[10] X. Zhu, L. Yu, X. Zhao, et al., "TPH-YOLOv5: Improved YOLOv5 Based on Transformer Prediction Head for Object Detection Drone captured Scenarios." In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, ICCVW2021, Montreal, BC, Canada, 11-17 October 2021; pp. 2778-2788.
[11] Z. Liu, H. Mao, C. Wu, et al., "A ConvNet for the 2020s." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, CVPR 2022, New Orleans, LA, USA, 18-24 June 2022; pp. 11966-11976.
[12] C. Y. Wang, A. Bochkovskiy, H. Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state of the art for real time object detectors." *arXiv* 2022, arXiv: 2207.02696.
[13] P. Gao, J. Lu, H. Li, et al., "Container: Context aggregation network." *arXiv* 2021, arXiv: 2106.01401.
[14] P. Dollar, M. Singh, "Girshick, R. Fast and accurate model scaling." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Nashville, TN, USA, pp. 19-25 June 2021.
[15] P. K. A. Vasu, J. Gabriel, J. Zhu, et al., "An improved one millisecond mobile backbone." *arXiv* 2022, arXiv:2206.04040.
[16] X. Ding, X. Zhang, J. Han, et al., "Repvgg: Making vgg style convnets great again." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Nashville, TN, USA, 19-25 June 2021; pp. 13733-13742.
[17] H. Zhou, H. Huang, X. Yang, et al., "Faster R-CNN for marine organism detection and recognition using data augmentation." In *Proceedings of the International Conference on Video and Image Processing*, Singapore, 27-29 December 2017; pp. 56-62.
[18] L. Chen, J. Dong, H. Zhou, et al.,"Underwater object detection using Invert Multi Class AdaBoost with deep learning." In *Proceedings of the 2020 International Joint Conference on Neural Networks (IJCNN)*, Glasgow, UK, 19-24July 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 1-8.
[19] W. Qiao, M. Khishe, S. Ravakhah, "Underwater targets classification using local wavelet acoustic pattern and Multi-Layer Perceptron neural network optimized by modified Whale Optimization Algorithm." *Ocean Eng*. 2021, 219, p. 108415.
[20] A Sahoo, S. K. Dwivedy, P. S. Robi, "Advancements in the field of autonomous underwater vehicle." *Ocean Engineering*, 2019.
[21] S. Liu, L. Qi, H. Qin, et al., "Path Aggregation Network for Instance Segmentation." In *Proceedings of the* 2018 *IEEE Conference on Computer Vision and Pattern Recognition*, CVPR 2018, Salt Lake City, UT, USA, 18-22 June 2018; pp. 8759-8768.
[22] J. Wang, C. Xu, W. Yang, et al., "A normalized Gaussian Wasserstein distance tiny object detection." *arXiv preprint arXiv*: 2110. p. 13389, 2021.
[23] H. Rrzatofighi, N. Tsoi, J. Y. Gwak, et al., "Generalized intersection over union: A metric and a loss for bounding box regression," *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2019, pp. 658-666.
[24] Z. Zheng, P. Wang, W. Liu, et al., "Distance-IoU loss: Faster and better learning for bounding box regression," *Proceedings of the AAAI conference on artificial intelligence*. 2020, 34(07), pp. 12993-13000.
[25] Y. F. Zhang, W. Ren, Z. Zhang, et al., "Focal and efficient IOU loss for accurate bounding box regression." *Neurocomputing*, 2022, 506, pp. 146-157.
[26] C. Dewi, and R. C. Chen, "Deep Learning for Advanced Similar Musical Instrument Detection and Recognition," *IAENG International Journal of Computer Science*, vol. 49, no.3, pp. 880-891, 2022.