

Research on Underwater Object Detection Algorithm Based on Improved YOLOv5

1st Hongyi Xia

School of Information and Intelligent Science and
Technology
Hunan Agricultural University
Changshang, China
460828497@qq.com

2nd Lixin Tan*

School of Electronic Engineering
Hunan College of Information
Changshang, China
School of Information and Intelligent Science and
Technology
Hunan Agricultural University
Changshang, China
594637823@qq.com

Abstract—To address the complexities of underwater environments, poor imaging conditions, the challenges of underwater object detection, slow detection speeds, underwater target density, and underwater biodiversity, a YOLOv5-based underwater object detection method is proposed. To meet the performance and efficiency requirements of underwater robots, the ShearC3 module is introduced, which improves the YOLOv5 C3 module to reduce model parameters and network layers, thus enhancing detection speed and focusing on the most effective receptive field's central part. To tackle the issue of low-scale information loss, the PixelShuffle upsampling module is used to integrate channel dimension information for generating high-resolution images. Through multiple dataset experiments, the improved model achieves a detection accuracy of 84.3%, a 1.2% improvement over the original model, while also increasing the detection speed by 10.16%, meeting the operational needs of underwater robots.

Keywords—Underwater object detection; Improved YOLOv5; Detection algorithm

I. INTRODUCTION

Underwater bio-detection [1,2] is a popular and widely applied field, encompassing areas such as marine science, oceanography, underwater salvage, and aquaculture monitoring. However, manual processing of underwater videos and images is time-consuming, expensive, and prone to fatigue-induced errors. Therefore, employing artificial intelligence for underwater object classification is an excellent choice. The underwater environment presents several challenging factors, including complex backgrounds, lighting variations, dynamic backgrounds, and the camouflage of underwater organisms, all of which affect the accuracy of underwater object detection.

In recent years, the rapid development of deep learning has brought significant breakthroughs to the field of object detection. Object detection algorithms based on deep learning offer high detection accuracy and strong robustness. Pengfei Shi^[3] et al. proposed an improved underwater bio-detection algorithm based on Faster-RCNN^[4], utilizing ResNet^[5] as the feature extraction network and employing EIou instead of IoU for object detection. However, Faster-RCNN, as a two-stage detection algorithm, is relatively slow and has a large number of parameters, making it less suitable for underwater operations. Another single-stage algorithm,

SSD^[6], was proposed by Wei Qiang^[7] et al. using ResNet in place of VGG as the foundational network for object detection, but it still had limitations in terms of detection speed.

For the two detection methods mentioned above, they have achieved good results but still face several issues. One is the complexity of the underwater environment, with significant lighting effects, making recognition challenging and detection difficult. The second issue is the underperformance of underwater embedded devices, resulting in slow detection speeds. Therefore, this paper proposes an improved algorithm based on YOLOv5, using the introduced ShearC3 module to replace the C3 module in the original medium model prediction and utilizing the PixelShuffle module to replace the original upsampling part, thus enhancing detection speed and accuracy.

II. YOLOV5 MODEL

YOLOv5 is an object detection algorithm initially proposed by Joseph Redmon^[8] et al. It is based on regression principles. In comparison to Two-Stage models like RCNN and FasterRCNN, YOLOv5 utilizes a One-Stage detection approach, offering faster detection speeds and improved detection accuracy.

YOLOv5 primarily consists of three components, including the backbone network (Backbone), the detection layer (Neck), and the prediction layer (Head).

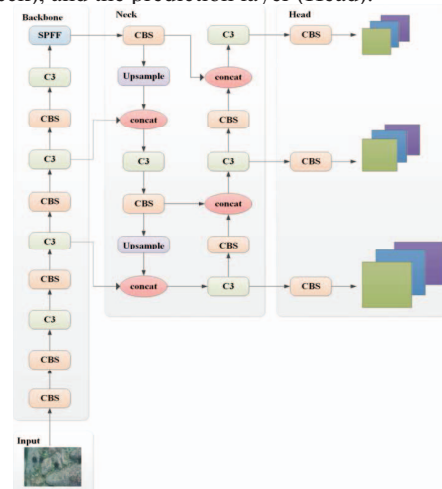


Figure 1. Structure diagram of YOLOv5 model

III. YOLOv5 NETWORK STRUCTURE IMPROVEMENT

A. YOLOv5-ShearC3 Network Structure

The original YOLOv5 model utilized the C3 network structure in its "neck" section, resulting in a larger number of parameters and slower detection speed. In practical applications, especially in embedded devices and complex environments such as underwater bio-detection, the use of large and complex models is constrained. Additionally, in the intricate underwater environment, low latency and rapid response are essential, making the research of smaller and more efficient network models significant^[9].

Selecting an effective receptive field is crucial for feature extraction, target localization, and scale adaptation, as it can impact the performance and accuracy of neural networks in tasks like object detection and image recognition, consequently affecting overall accuracy^[10]. With the introduction of the ShearC3 module, the improved model exhibits a better receptive field for predicting medium-sized objects, as depicted in Figure 2. By visualizing the effective receptive field, it becomes apparent that the improved receptive field is more concentrated in the central area, thus reducing the rate of false negatives and enhancing the performance of object detection. Due to the small data set, the smaller the receptive field, the stronger the detection performance. Figure 2 provides a comparison between the original YOLOv5 model and the effective receptive field after adding the ShearC3 module.

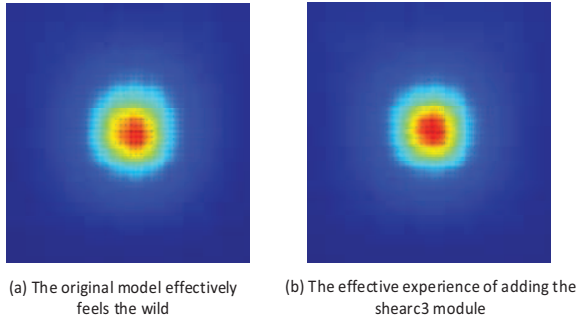


Figure 2. Comparison between the effective receptive field of the original model and the YOLOv5 ShearC3 model's effective receptive field.

The proposed trimmed YOLOv5-ShearC3 network structure is illustrated in Figure 3.

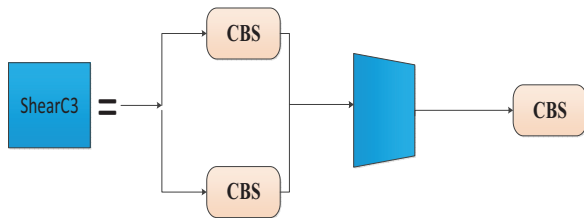


Figure 3. Structure diagram of the YOLOv5-ShearC3 model.

B. Specifications Adoption of PixelShuffle

YOLOv5 employs nearest-neighbor interpolation for upsampling to merge features from different scales.

However, this process results in information loss in the lower-scale feature maps. Therefore, in our experiments, the sub-pixel upsampling method (PixelShuffle)^[11] was introduced as a replacement for interpolation-based upsampling.

The introduction of the PixelShuffle upsampling module to YOLOv5 replaces the two linear upsampling modules in the original YOLOv5 network structure. Similar to the introduction of the ShearC3 module, a comparison was made between the effective receptive fields of the original model and the medium-sized model prediction part with the PixelShuffle module introduced, as depicted in Figure 4. Visualizing the effective receptive field revealed that the improved effective receptive field is more dispersed in the central region, resulting in a reduction in the overall effective receptive field. However, the detection accuracy remains unchanged, preserving the original level of precision.

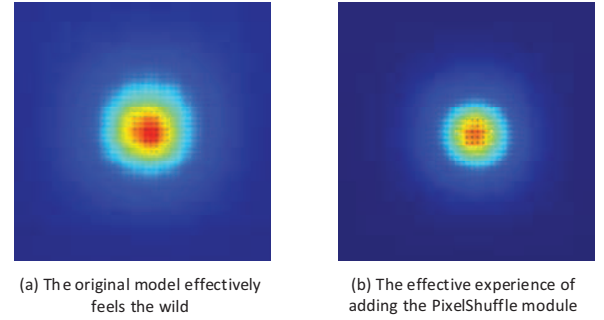


Figure 4. Comparison between the effective receptive field of the original model and the PixelShuffle effective receptive field.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A. Dataset

To validate the effectiveness of this algorithm, a dataset collected from the Underwater Robot Grasping Contest on Zhangzi Island in Dalian was used. To increase dataset diversity, Mosaic data augmentation was applied. The dataset includes images of four different types of marine life, totaling 4,161 images, which consist of sea cucumbers, sea urchins, scallops, and starfish. The dataset was divided into training, validation, and test sets in a ratio of 7:1:2.

The evaluation indexes were precision, recall and mean average precision (mAP) as the evaluation indexes, and the expressions are as follows.

$$P = \frac{TP}{TP + FP} \quad (1)$$

$$R = \frac{TP}{TP + FN} \quad (2)$$

$$AP = \int_0^1 P - R(R) d(R) \quad (3)$$

$$mAP = \left(\frac{1}{N} \right) \sum_{i=1}^N AP_i \quad (4)$$

B. Experimental Environment and Parameter

The improved algorithm based on YOLOv5 was trained and tested on a Windows 10 system with the following specifications: PC with an Intel(R) Core(TM) i5-13600K CPU @ 3.50GHz and an NVIDIA GeForce RTX 2060. The CUDA version used was 1.7. Input image size was set to 640×640 pixels, with a batch size of 8, and training was carried out for 100 epochs. The code was run in a Python environment with version 3.8.16, using PyTorch version 1.13.1.

C. Comparative Experiments on Other Networks

In order to objectively assess the effectiveness of the improved algorithm, comparative experiments were conducted with existing high-performance object recognition algorithms. The experimental results are presented in Table 1.

From Table 1, it can be observed that the improved algorithm achieved mAP(@0.5) and mAP(@0.95) of 83.3% and 48%, respectively, surpassing the performance of mainstream object detection models. Compared to the original YOLOv5s model, the improved algorithm increased mAP(@0.5) by 1.2%. Furthermore, the detection

speed of the improved algorithm was 7.2 frames per second higher than the original YOLOv5s model, reaching 78.12 frames per second, which is a 10.16% improvement. This makes the improved algorithm more suitable for real-time underwater object detection. Additionally, the improved algorithm reduced the number of parameters by 0.82 million, making it more suitable for embedded use on mobile devices.

These experimental results indicate that the ShearC3 module and the PixelShuffle upsampling module played a crucial role in obtaining irregular local pixel information. The improved model not only increased detection speed but also showed some improvement in detection accuracy. In conclusion, the enhanced network based on YOLOv5 has significant potential for real-time underwater object detection tasks, enhancing both detection speed and accuracy.

Although the recently proposed YOLOv7 shows superior performance on major datasets, it cannot show its advantages on small datasets due to its large network model, and the detection speed cannot meet the requirements of real-time performance.

TABLE I. OVERALL IDENTIFICATION PERFORMANCE EVALUATION

Model	AP(@0.5)				mAP(@0.5)	mAP(@0.95)	Parameters (M)	(frames/s)
	Sea Cucumber	Sea Urchin	Scallop	Starfish				
YOLOV3-tiny ^[12]	0.719	0.892	0.817	0.879	0.754	0.469	8.67	77.51
YoloV4-tiny	0.675	0.876	0.812	0.84	0.752	0.468	6.06	74.07
YOLOv5S	0.705	0.893	0.821	0.906	0.831	0.458	7.03	70.92
YoloV7 ^[13]	0.621	0.872	0.727	0.854	0.769	0.41	37.21	39.21
YoloV7-tiny	0.553	0.864	0.708	0.828	0.738	0.377	6.01	73.99
FasterRCNN(resnet50)	0.781	0.869	0.73	0.814	0.799	/	/	10.43
FasterRCNN(vgg)	0.803	0.881	0.736	0.829	0.812	/	/	10.66
SSD	0.504	0.729	0.443	0.641	0.579	/	/	57.69
Ours	0.731	0.904	0.828	0.91	0.843	0.464	6.88	78.12

D. Ablation Experiments

In order to validate the effectiveness of the algorithm, ablation experiments were conducted in this study to assess the impact of each module on the model's performance, and the results are presented in Table 2.

Through Table 2, it can be observed that after introducing the ShearC3 network structure separately, the parameter count decreased by 2.2%. The detection accuracy remained relatively stable, while the detection speed increased by 12.28% compared to the original network, meeting the real-time requirements for underwater detection. By employing the PixelShuffle up-sampling improvement algorithm with a bottleneck layer count of 3, the mAP@0.5 value reached 0.831, comparable to the detection accuracy of the original YOLOv5 network model.

However, it exhibited better detection performance when identifying sea cucumbers.

Therefore, simultaneously introducing the ShearC3 and PixelShuffle modules in the experiment resulted in a 1.2% improvement in detection accuracy and a 10.16% increase in detection speed compared to the original YOLOv5 model. This effectively enhanced the model's detection accuracy while ensuring the detection speed of the ShearC3 network model, meeting the performance requirements for underwater target detection.

Through the experiment, an analysis of the effective receptive field was conducted, revealing that the detection performance is optimal when the effective receptive field is smaller and more concentrated. The effective receptive field introduced by the ShearC3 and PixelShuffle modules is illustrated in Figure 5.

TABLE II. PERFORMANCE COMPARISON OF DIFFERENT MODULES

Model	bottlenecklayer				AP(@0.5)				mAP(@0.5)	mAP(@0.95)	Parameters (M)	(frames/s)
	0	1	2	3	Sea Cucumber	Sea Urchin	Scallop	Starfish				
YoloV5s	×	×	×	√	0.705	0.893	0.821	0.906	0.831	0.458	7.03	70.92
YoloV5-ShearC3	×	×	√	×	0.715	0.893	0.816	0.906	0.833	0.46	6.86	79.32
YoloV5-ShearC3	×	√	×	×	0.709	0.891	0.82	0.904	0.831	0.458	6.86	79.43
YoloV5-ShearC3	√	×	×	×	0.715	0.893	0.816	0.906	0.833	0.46	6.86	79.36
YoloV5-ShearC3	×	×	×	√	0.72	0.89	0.809	0.907	0.831	0.463	7.04	69.93
PixelShuffle ours	√	×	×	×	0.731	0.904	0.828	0.91	0.843	0.464	6.88	78.12

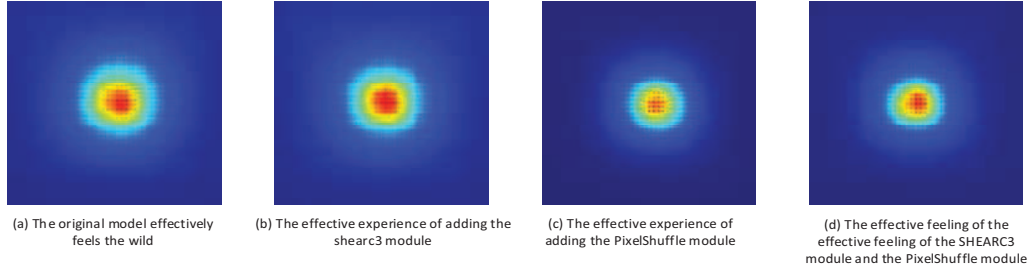


Figure 5. Comparison between the effective Receptive field of the original model and the effective Receptive field of each module

V. CONCLUSION

Underwater object detection is of paramount importance in marine environments and aquaculture. We proposed an underwater object detection algorithm based on YOLOv5 to address challenges such as complex underwater environments, vague target features, and slow detection speeds. By introducing the ShearC3 network structure, we reduced the number of parameters, improved detection speed, while maintaining detection accuracy close to the original YOLOv5 network model. Replacing the nearest-neighbor interpolation upsampling module of YOLOv5-ShearC3 with the PixelShuffle upsampling module increased image resolution and enhanced feature information without increasing computational complexity. The improved model achieved an mAP(@0.5) of 84.3%, representing a 1.2% improvement. This enhancement demonstrated the model's superior performance in terms of accuracy and detection speed, surpassing the original model and meeting the requirements for both precision and real-time underwater object detection. This approach is applicable for monitoring marine environments and underwater object detection, with broad prospects for application.

VI. ACKNOWLEDGMENT

This project is supported by Hunan Provincial Teaching Master Talent Support Program: 2019 Hunan Provincial Teaching Master Project (Xiang Jiao Tong [2019] No. 261).

REFERENCES

- [1] Y. Z. Ye, X. H. Duan, and C. Zhao, "Research on Improved YOLOv3-SPP for Underwater Object Detection," *Computer Engineering and Applications*, vol. 59, no. 06, pp. 231-240, 2023.
- [2] D. Cai and B. J. Fan, "Underwater Object Detection Method Based on Spatial Feature Selection," *Information and Control*, vol. 51, no. 02, pp. 214-222, 2022. doi: 10.13976/j.cnki.xk.2022.1597.
- [3] P. Shi, X. Xu, J. Ni, Y. Xin, W. Huang, and S. Han, "Underwater Biological Detection Algorithm Based on Improved Faster-RCNN," *Water*, vol. 13, no. 17, p. 2420, 2021. doi: 10.3390/w13172420.
- [4] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137-1149, 1 June 2017. doi: 10.1109/TPAMI.2016.2577031.
- [5] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *IEEE*, 2016. doi: 10.1109/CVPR.2016.90.
- [6] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single Shot MultiBox Detector." In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds) *Computer Vision - ECCV 2016*. ECCV 2016. Lecture Notes in Computer Science(), vol 9905. Springer, Cham. doi: 10.1007/978-3-319-46448-0_2.
- [7] Q. Wei, Y. He, Y. Guo, B. Li, and L. He, "Exploring Underwater object detection Algorithm Based on Improved SSD," *Xibei Gongye Daxue Xuebao*, Aug. 2020. EDP Sciences. doi: 10.1051/jnwpu/20203840747.
- [8] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," *Computer Vision & Pattern Recognition*. IEEE, 2016. doi: 10.1109/CVPR.2016.91.
- [9] Z. Hao, M. Lei, and W. Song, "Safety Helmet Detection Method Based on Lightweight Deep Learning Model," *Science Technology and Engineering*, vol. 22, no. 14, pp. 5659-5665, 2022.
- [10] W. Gao, "Research on the Effective Receptive Field of Deep Convolutional Neural Networks," *Harbin Institute of Technology*, 2021. doi: 10.27061/d.cnki.ghgdu.2021.001905.
- [11] W. Shi, J. Caballero, F. Huszar, J. Totz, A. P. Aitken, R. Bishop, Z. Wang, "Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). doi: 10.1109/c00vpr.2016.207.