Underwater Object Detection via Structural Pruning of YOLOv7

Junjie Liu TRAVIS_LAU3344@163.COM
School of Information Science and Engineering, Dalian Polytechnic University, Dalian, Liaoning, China
Zhanying Li* L_ZY1979@126.COM
School of Information Science and Engineering, Dalian Polytechnic University, Dalian, Liaoning, China
Zijun Wang W13354078397@163.COM
School of Information Science and Engineering, Dalian Polytechnic University, Dalian, Liaoning, China
Longhui Liu 18239430590@163.COM
School of Information Science and Engineering, Dalian Polytechnic University, Dalian, Liaoning, China
**Corresponding author*

Editors: Nianyin Zeng, Ram Bilas Pachori and Dongshu Wang

Abstract

This study addresses the challenge of deploying object detection models in resource-constrained underwater environments by optimizing YOLOv7 through pruning techniques.Underwater detection faces limitations due to low-light conditions, water turbidity, and mobile device constraints.The proposed method applies channel pruning to YOLOv7, strategically removing low-weight channels to reduce computational load and parameter count while maintaining accuracy.Comparative experiments evaluated pruning rates (0%, 20%, 40%, 50%, 60%, 80%) on the UPRC dataset, focusing on sea urchins, scallops, sea cucumbers, and starfish.Results showed that a 50% pruning rate achieved optimal balance: mAP increased by 2.3% (from 83.8% to 85.7%), while parameters and computations reduced to one-fourth of original values.

Keywords: YOLO, pruning, Light weight

1. Introduction

The ocean is an important part of the earth, it covers most of the area on the earth's surface, but the ocean is still very mysterious, at present, human development and exploration of the ocean is only 5%, but the Marine resources inside the ocean are very rich, which makes people's future development and exploration of the ocean becomes necessary. With the rise of Marine economy in recent years, people's demand for Marine products is also increasing year by year, which forces people to test Marine products faster than before and more accurate than before. Marine products such as sea urchins, sea cucumbers, oysters and scallops are very popular with people, and they also have very high nutritional value to supplement the nutrients that the human body needs. Therefore, in order to better explore Marine resources, people must vigorously develop Marine product testing technology, which can more effectively promote the development of Marine fisheries. However, in the complex Marine environment, there are many problems to be solved in order to accurately and quickly detect underwater targets, such as the intensity of underwater light, the speed of water flow in each body of water and the water quality, which can affect the visibility of mobile devices in the water and ultimately affect the quality of the image, greatly increasing the difficulty of underwater target detection.

Currently, with the ongoing advancements in technology, individuals are opting to utilize intelligent machines for underwater fishing activities. However, the underwater environment poses significant challenges due to its harsh conditions. Mobile devices, which are employed in such settings, have limited storage and computational capabilities. This necessitates the integration of as many functionalities as possible within these resource-constrained devices to effectively handle the complexities of the underwater domain.

Consequently, deploying large-scale object detection models on mobile devices is impractical. To address this issue, this paper introduces a pruning operation that strategically removes channels with excessively low weights from the model. By doing so, it significantly reduces both the parameter count and computational load, thereby achieving a lightweight model design while enhancing overall accuracy.

2. Materials and Methods

The algorithm presented in this paper builds upon the YOLOv7 framework, enhancing its potential for deployment on mobile devices through lightweight processing. YOLOv7 is an object detection algorithm that leverages E-ELAN to improve model efficiency by optimizing feature extraction and fusion mechanisms.

The architecture of YOLOv7 consists of three main components: Backbone, Neck, and Head.The Backbone section employs E-ELAN (Li et al., 2025) to extract multi-scale and high-semantic feature maps from input images, which are then utilized for subsequent target classification and localization tasks. By incorporating feature reuse and gradient propagation techniques, the model's capacity to represent complex targets is significantly enhanced.

This improvement not only streamlines the computational requirements but also maintains a high level of accuracy, making it particularly suitable for resource-constrained environments such as underwater settings where real-time object detection is crucial.Based on ELAN, E-ELAN uses group convolution operations to divide feature graphs into multiple groups, and performs convolution operations separately in each group. Finally, the results obtained by each group are combined together, which can ensure feature diversity while reducing model computation. The Neck part uses PAFPN to combine image features at all levels to strengthen the detection model's ability to detect objects at different scales (Ji et al., 2023). The image features output by Backbone are optimized and multi-scale feature maps more suitable for tasks are generated. PAFPN is improved by adding adaptive feature fusion module and cross-stage dense connection on the basis of PANet. The feature map output by Backbone is spliced with shallow features, multi-branch convolution is used to improve the feature expression ability of the model, and then single-branch convolution is combined to reduce the computation amount. The ability of the model to detect small targets is enhanced by assigning dynamic weights. The Head part predicts the confidence, category of detection target, and boundary box coordinates by the feature map output from the Neck part, and improves the detection accuracy and training efficiency through dynamic label allocation and decoupling. Dynamic label allocation can dynamically assign positive and negative samples according to the IoU (Liu et al., 2022) $\times$ classification confidence between the prediction box and the true value to reduce the mismatching rate of fuzzy samples, add scattered noise, simulate light attenuation, and improve the model generalization ability. In this paper, pruning operation will be carried out on the whole model to remove the channels with lower weights, so as to improve the running speed of the whole model and realize the lightweight of the detection model (She and Li, 2024). Figure 1 presents the block diagram of YOLOv7.

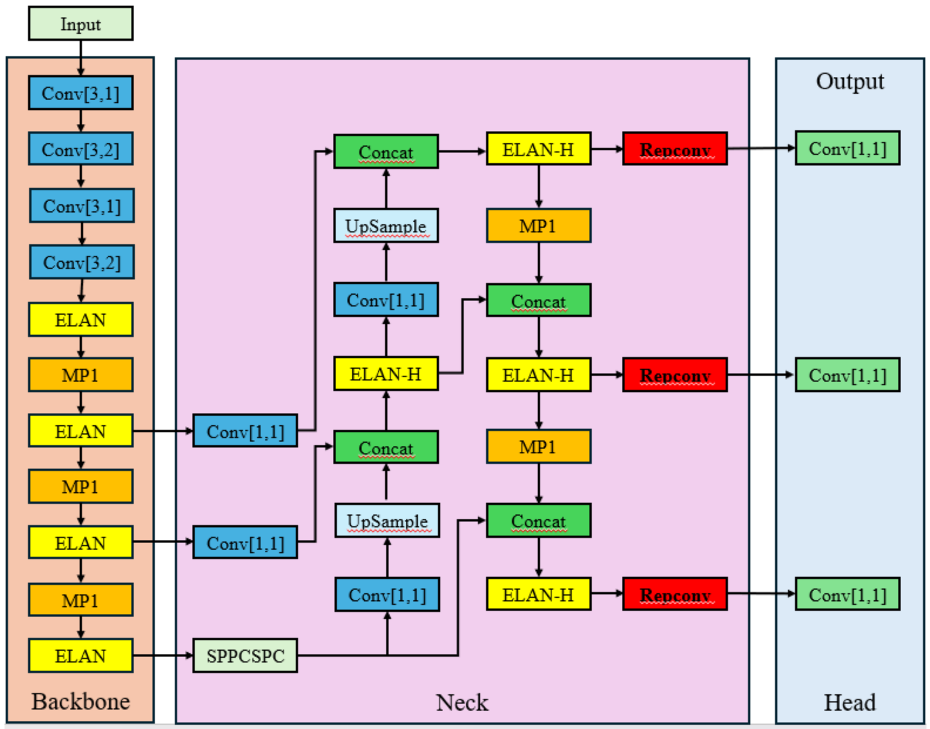


Figure 1: YOLOv7 Structure

2.1. Prune

Pruning is a technique that enables the compression and acceleration of deep learning models, making them more lightweight. Its core principle involves reducing the computational load and storage requirements by eliminating redundant parameters and structures within neural networks. This process enhances both the operational speed and training efficiency of deep learning models while improving their generalization capabilities.

When implementing pruning, it's crucial to strike a balance between the pruning rate and overfitting risks. Excessive pruning can lead to underfitting, while insufficient pruning fails to significantly compress the model, resulting in minimal benefits and continued heavy computational demands.

The application of pruning technology in underwater target detection model is of great significance. The adaptability and practicality of the algorithm are improved by pruning the model size and maintaining the detection accuracy to meet the unique requirements of thin deployment in the underwater environment. Underwater images have problems of atomization, low contrast and color distortion. Traditional models require a large number of parameters to capture weak features, which leads to computational redundancy. Redundant channels or layers can be removed to preserve the weights sensitive to underwater key features and improve the efficiency of feature extraction. The scale difference of underwater targets is large, and the model needs multi-scale feature fusion, but the number of parameters is large. For example, channel pruning can optimize the redundant calculation of multi-scale branches, preferentially retain channels sensitive to multi-scale, and balance accuracy and speed. The data distribution in different waters is very different, and the large model is easy to overfit specific scenes. After pruning, the model complexity is reduced, and the overfitting risk is reduced. Combined with the lightweight domain adaptive method, the cross-scene generalization ability is enhanced. Pruning effectively addresses the challenge of performing benthic biological detection tasks under resource-constrained conditions, while preserving the accuracy and performance of the model. Loss function of channel pruning:

$$L = \sum_{(x.y)} l(f(x, W), y) + \lambda \sum_{\gamma \in \Gamma} g(\gamma) \tag{1}$$

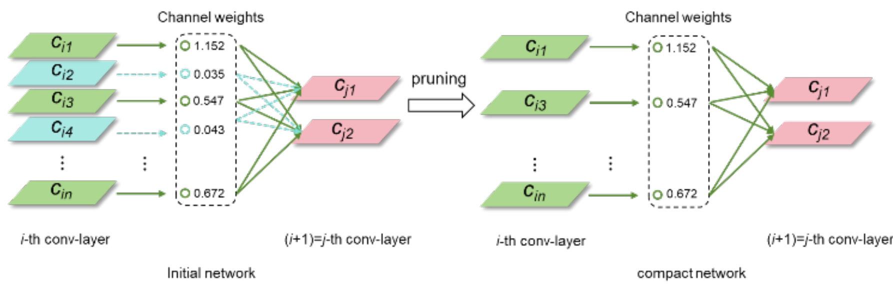


Figure 2: Channel pruning algorithm

Figure 2 is the schematic diagram of the principle of channel pruning. The object of pruning is the channel that has less influence on the model and the channel that has less weight. Pruning channel means removing all the input and output connections of the channel.

3. Experiments and Results

3.1. Experimental environment

The experimental dataset utilized in this paper is UPRC, which serves as the official dataset for the 2019 China Underwater Robot Professional Competition (Wen et al., 2022). The data set contains the species categories of sea starfish, seaweed, sea urchin, sea cucumber and scallop, but only four categories of sea urchin, sea cucumber, sea star and scallop are tested, and seaweed interference detection as interference items is not included in the final test. Therefore, the images of seaweed in the original UPRC data set were removed, and only the other four types of target detection were carried out. For the rest of the data set, we will separate the test set from the training set in a ratio of 3:7.

The experimental environment configured in this paper is as follows: the operating system selected is Windows 10 version, with an Intel Core i7-12700F central processor and an NVIDIA GeForce RTX 3080 graphics processor. The development environment chosen for the experiment is PyCharm 2022.1, while the acceleration environment is CUDA 11.1 and cuDNN 8.0.5. For programming, Python version 3.9.0 is used along with PyTorch version 1.9.0.

In the experimental model comparison, SSD, Faster-RCNN, YOLOX-s, YOLOv6-s, YOLOv7, and YOLOv8 were evaluated. In terms of accuracy, YOLOv7 achieved the highest accuracy at 83.9%, demonstrating superior performance. For parameter count and computational efficiency, YOLOX-s performed exceptionally well. However, SSD outperformed the other models in frames per second (FPS). Table 1 records the results of the comparative tests conducted on each model.

Table 1: Model comparison experiment

Model	mAP(%)	Params(M)	Flops(G)	FPS
SSD	54.7	26.29	62.75	**112.4**
Faster-RCNN	73.4	28.48	941.17	24.7
YOLOX-s	81.7	**8.94**	**26.76**	62.5
YOLOv-6s	82.3	18.5	45.3	82.9
YOLOv7	**83.9**	36.5	103.2	70.4
YOLOv8	83.5	11.2	28.6	81.8

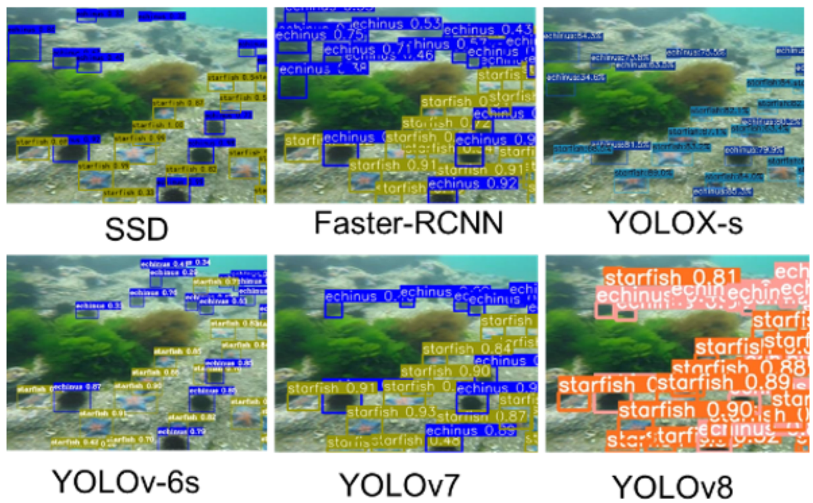


Figure 3: Model visualization comparison

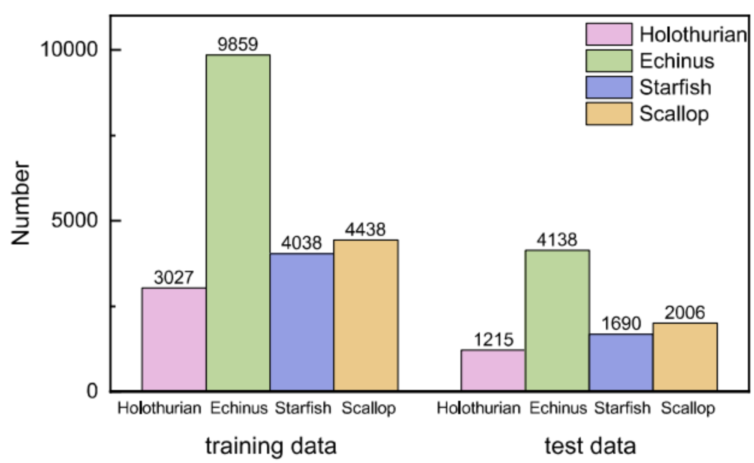


Figure 4: Number of different categories of targets

Figure 3 is a visual comparison of each model. Figure 4 shows the number of targets of different categories in the dataset used in the experiment. Table 2 shows the parameters used in the experiments of this paper.

Table 2: Experimental parameters

Parameter	Value
Resolution	640*640
Initial learning rate	0.001
Weight decay coefficient	0.0005
Batch size	16
Workers	8
Epoch	300

3.2. Evaluation metrics

If you want to compare the advantages and disadvantages of models, evaluation indicators are very effective. Parameters Precision(P), Recall(R), Average Precision(AP), mean Average Precision(mAP) and Frames Per Second (FPS) will be used to evaluate the experimental results. The following describes the significance of each parameter indicator.

$$P(\Pr ecision) = \frac{TP}{TP + FP} \tag{2}$$

Precision(P) refers to how many of the samples in which the model predicts positive results are correct, TP is True Positive, FP is False Positive.

$$R(\mathrm{Re}\, call) = \frac{TP}{TP + FN} \tag{3}$$

Recall(R) is how much of what is really positive is correctly predicted to be positive, FN is False Negative.

$$AP = \int_0^1 P(r)dr \tag{4}$$

Average Precision(AP) refers to the average accuracy of different classes.

$$mAP = \frac{1}{N}\sum_{n=1}^{N} AP_n \tag{5}$$

Mean Average Precision(mAP) refers to the average accuracy of a multi-class population (Wang et al., 2023).

The value range of P, R, AP, and mAP is [0,1]. The larger the value, the better the effect.

FPS (Frames Per Second) quantifies the number of image frames a model can process per second under specified hardware configurations. Higher FPS values directly correlate with enhanced real-time capabilities, as governed by the Nyquist-Shannon sampling theorem, which mandates that the detection system's FPS must exceed twice the maximum frequency of target motion to prevent temporal aliasing.

3.3. Comparison of pruning rate

In this experiment, comparative studies were conducted by varying pruning rates (0%, 20%, 40%, 50%, 60%, 80%) to evaluate target detection accuracy, parameter count, and computational load.

As shown in the table, the detection accuracy for sea urchins and scallops initially increased from 88.5% and 83.9% (0% pruning) to peak values of 90.6% and 85.3% at a 50% pruning rate. Further increasing the pruning rate to 80% reduced their accuracy to 88.2% and 70.1%, respectively. Similar trends were observed for sea cucumbers and starfish, though their peak accuracies occurred at different pruning rates: sea cucumbers achieved maximum accuracy (79.2%) at 40% pruning, while starfish peaked at 88.8% with 20% pruning. The highest mAP value was attained at a 50% pruning rate.

Both parameter count and computational load decreased significantly with higher pruning rates. The parameter count reduced from 36.47 million to 1.43 million, and computational load dropped from 103.1 GFLOPs to 4.1 GFLOPs. At the 50% pruning rate, these metrics were reduced to 8.56 million parameters and 25.2 GFLOPs, representing approximately one-quarter of their original values. Given the balance between maintained accuracy and substantial reductions in model complexity, the optimal pruning rate for this experiment was determined to be 50%. Table 3 records the accuracy, parameter count and computational complexity of the experimental model at different pruning rates.

Table 3: The precision, parameter number and calculation amount of each pruning rate

Pruning ratio (%)	AP (%)				mAP (%)	Params(M)	Flops(G)
	Echinus	Scallop	Holothurian	Starfish			
0	88.5	83.9	75.4	87.4	83.8	36.47	103.1
20	90.4	84.2	78.0	**88.8**	85.4	23.24	65.5
40	90.1	84.7	**79.2**	88.3	85.6	13.12	36.8
50	**90.6**	**85.3**	78.3	88.6	85.7	8.56	25.2
60	90.2	83.5	77.0	87.6	84.5	5.78	16.2
80	88.2	70.1	60.1	79.0	74.4	1.43	4.1

Figure 5 shows the visual comparison of the model under different pruning rates. Figure 6 shows the comparison of the model's effects before and after pruning.

4. Conclusions

The application of pruning operations on the YOLOv7 model results in a substantial reduction of its parameter count and computational load. Such optimization proves particularly critical for autonomous underwater vehicles (AUVs) operating in dynamic benthic environments, where real-time detection of low-illumination marine organisms and avoidance of moving obstacles require both algorithmic responsiveness and energy sustainability, particularly within intricate and dynamic underwater environments.

Empirical comparisons reveal that at a pruning rate of 0.5, the model's accuracy peaks, exhibiting an improvement of 2.3% relative to an unpruned model (characterized by a pruning rate of 0). Concurrently, this specific pruning rate diminishes both the model's parameter count and computational requirements to merely one-fourth of their initial values. Consequently, the model achieves a lightweight design while concurrently enhancing the speed of target detection processes.

Despite these improvements, there are still some shortcomings in the current model, such as lower detection accuracy for sea cucumbers. Due to their similar shapes to coral reefs, sea cucumbers can easily be missed during detection. To address this issue, algorithms that enhance precision

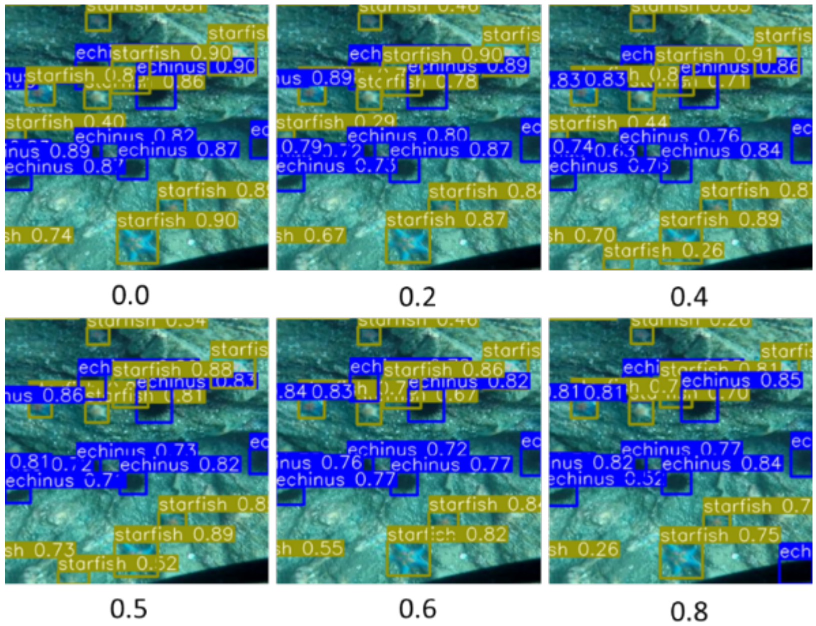
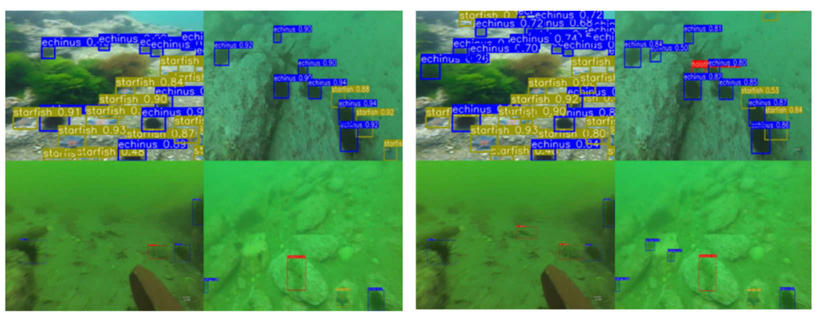


Figure 5: Pruning visual comparison



Pre pruning After pruning

Figure 6: Comparison before and after pruning

could be added; however, this would likely increase the model's size and further raise requirements for mobile devices. Future research needs to optimize the model further, balancing improved detection accuracy with resource consumption for better practical application outcomes.

References

Z. Ji, Y. Wu, X. Zeng, et al. Lung nodule detection in medical images based on improved yolov5s. *IEEE Access*, 11:76371–76387, 2023. doi: 10.1109/ACCESS.2023.3296530.

Y. Li, J. Han, H. Li, et al. Branch-yolo: An efficient object detector for thin structure objects like pantograph. *Digital Signal Processing*, 2025:105121, 2025. doi: 10.1016/j.dsp.2025.105121.

Daiyao Liu, Linfeng Xu, Heqian Qiu, Yue Liang, Haoyang Cheng, and Jingjing Xie. Scale-guided adaptive sampling for general object detection algorithms. In *2022 IEEE 8th International Conference on Computer and Communications (ICCC)*, pages 1840–1845, 2022. doi: 10.1109/ICCC56324.2022.10065860.

Y. She and H. Li. Monocrystalline silicon crystal line detection based on the improved yolox-tiny algorithm. *Journal of Intelligent Manufacturing*, 2024. doi: 10.1007/s10845-023-02312-z.

Chien-Yao Wang, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023.

G. Wen, F. Liu, X. Luo, et al. Underwater target detection based on modified yolov5. In *2022 5th International Conference on Intelligent Autonomous Systems (ICoIAS)*, pages 7–12. IEEE, 2022. doi: 10.1109/ICoIAS56028.2022.9931224.