WEP Warwick Evans Publishing

International Journal of Computer Science and Information Technology

ISSN: 3005-9682 (Print), ISSN: 3005-7140 (Online) | Volume 5, Number 3, Year 2025
DOI: https://doi.org/10.62051/ijcsit.v5n3.13
Journal homepage: https://wepub.org/index.php/IJCSIT/index

Freshwater Fish Detection Based on the Yolov8 Model

Xinyu Wang *

School of Geomatics and Urban Spatial Information, Beijing University of Civil Engineering and Architecture Beijing, China

*Corresponding Author: Xinyu Wang

ABSTRACT

Fish is a class of important components of river and lake ecosystems, and the judgment of their population migration and the statistics of their number is an important task in water environment regulation. Current intelligent identification methods based on deep learning models provide efficient and accurate means. In this paper, based on the Yolov8 model, we realize the recognition of freshwater fish species under the underwater complex environment. We add the attention mechanism of random masks to the model to reduce the computational complexity. A dataset containing ten freshwater fish species is also constructed from integrating web open datasets. Experimental results show that the proposed method achieves good recognition results, and the evaluation index Precision\Recall\ mAP50\ mAP50-95 is achieved.

KEYWORDS

Freshwater fish; Yolov8; Target detection; Deep learning; Water environment protection

1. INTRODUCTION

Fish is an irreplaceable component of the water ecosystem and an important part of aquatic organisms, which is of great value to the health and balance of the water ecosystem and the protection of water resources [1]. The invasion of exotic fish species greatly destroys the ecological diversity of river and lake water [2], and the establishment of an efficient and accurate fish monitoring and identification method is the key to protect the biodiversity of the water environment.

The current deep learning technology is rapidly developing and has achieved great success in tasks such as target detection, semantic segmentation, and instance segmentation of images. The YOLO series of models started with YOLOv1 (You Only Look Once) [3], which used single-stage detection and multi-scale prediction to achieve efficient real-time target detection. However, despite the fast inference speed, its ability to detect small targets is more limited.YOLOv3 significantly improves the detection accuracy and small target recognition ability by introducing three different scales of feature maps for prediction [4]. However, clustering analysis is required to determine the optimal number, size and scale of anchors before training, which leads to an increase in computation and an additional filtering process.YOLOv6 further improves the training and inference efficiency by removing the anchor mechanism [5]. YOLOv8 fuses the Anchor-Free and Anchor-Based methods, which maintains the efficient detection and extends the inference speed of YOLOv8 [6]. YOLOv8 combines Anchor-Free and Anchor-Based approaches, which extends the applicability of tasks such as instance segmentation while maintaining efficient detection. In addition, YOLOv8 introduces Distribution Focal Loss (DFL) as a loss function to improve the accuracy of bounding box prediction and reduce the dependence on manually designed anchors, making the model more flexible and robust [7].

Traditional machine learning methods rely on manually designed features and are difficult to understand image spatial contextual relationships as deep learning methods do, making fish detection in complex underwater environments more difficult. Underwater environments often have the effects of dark image tones, blurring, and occlusion, so the target detection algorithms are prone to misdetection, omission, or inaccurate localization.

Han et al. [8] proposed an underwater holographic target detection algorithm, FA-CenterNet, based on improved CenterNet and scene feature fusion, which utilizes a feature pyramid transformer to enhance the detection of fuzzy and small targets.Wang et al. [9] designed a lightweight underwater target detection model, YOLOv6-ESG, based on the YOLOv6 architecture, which is dedicated to detecting marine organisms such as sea urchins, sea stars, scallops, etc. YOLOv6 architecture, which is specialized in detecting marine organisms such as sea urchins, starfish and scallops. The model introduces lightweight ODConv, GSConv and VoVGSCSP modules to improve the detection accuracy of small targets in low-quality, low-resolution underwater images. In another study, Wang et al. [10] proposed B-YOLOX-S, a YOLOX-based underwater target detection model, which is mainly used for marine organisms detection. The method firstly combines poisson fusion and wavelet transform for image enhancement, and optimizes the network structure by combining BIFPN-S and FPN to improve the detection efficiency, while Ren et al. [11] proposed the CER-YOLOv7 underwater target detection algorithm, which employs an efficient intersection and merger (EIoU) loss function to optimize the positioning of the detection frame, thus reducing the probability of misdetections and miss-detections. In addition, Wu et al. [12] proposed the YOLOv5-fish fish detection algorithm for underwater fuzzy scenes, which utilizes the auto-MSRCR algorithm to enhance the fuzzy underwater images and introduces the Shuffle Attention mechanism to strengthen the model's focus on fish targets.

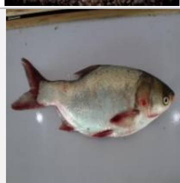
In this paper, we use the YOLOv8 model for detection of freshwater fish species, which reduces the amount of computation while enabling the model to focus on the more important target features in the image and reduce the interference of background features by chunking the input image and masking it randomly. In this paper, an underwater freshwater fish dataset containing ten fish species is constructed based on web open source data, and experiments are conducted on this dataset to analyze the results.

2. METHODS

2.1. Constructing Datasets

For common freshwater fish species, including grass carp, crucian carp, bighead carp, tilapia, black carp, cod, perch, scavenger, silver carp, catfish. Images are collected from web crawling as well as open-source datasets, and the labels are produced after comparing and confirming them with the actual species, and expert identification. The species division, quantity statistics, and image examples in the dataset are shown in Table 1 below, which contains a total of 2610 fish images of ten species, which are randomly divided into a training set test set and a validation set according to the ratio of 8:1:1.

Table 1. Examples of species, numbers and images of fish samples

Species	Numbers	Examples
grass carp	900	
crucian carp	140	
bighead carp	190	
tilapia	170	
black carp	220	
cod	230	
perch	250	
scavenger	210	
silver carp	180	
catfish	120	
Total	2610	

2.2. Image Enhancement

In fish detection tasks in underwater scenarios, the shooting equipment is often in an irregular posture and in a state of motion, and the images obtained are often blurred. Due to the limitation of water quality, the visibility is relatively low, and it is difficult to recognize the fish in the far field of view. At the same time, there is no obvious pattern of fish behavior and activities, and they will appear in front of the camera in various postures, including side, back, and top view. Therefore, in order to adapt to the actual environment, image enhancement methods such as adjusting brightness, increasing Gaussian noise, and random angle rotation are used in the form of random combination for sample data enhancement.

2.3. Model Structure

A Yolov8 structure proposed in this paper is shown in Figure 1 below. The backbone network of YOLOv8 consists of multiple CBS convolutional blocks, C2f blocks, and SPPF blocks stacked together to enhance the feature extraction capability. The CBS block consists of 1×1 convolution, batchNorm, and SiLU activation function, which is able to extract features efficiently while reducing the computational cost. The C2f block is optimized on the basis of the CSP structure by first doubling the number of channels of input feature graph by doubling the number of channels and gradually extracting deep features through multiple convolutional layers, followed by splicing the extracted features with the original input and outputting the final result through compression operation. This module can aggregate multi-scale information more efficiently and reduce the computational effort, while enhancing the expressive ability of the model. In the neck network, the model fuses different layers of features through a multi-scale feature pyramid to expand the sensory field and enhance the detection capability of small targets. The detection head part adopts a decoupled head, which performs target classification, bounding box prediction, and confidence calculation for three different sizes of feature maps, respectively, to make the detection task more accurate and efficient.

For the neck network part, this paper adds randomly mask Transformer block [13, 14] to enhance the feature extraction capability of the model. As shown in Figure 2, for the input dimension [H, W, C], the input is first divided into n feature blocks along the spatial dimension [H, W], and then randomly mask a portion of the blocks without the operation of the attention mechanism to reduce the computation.

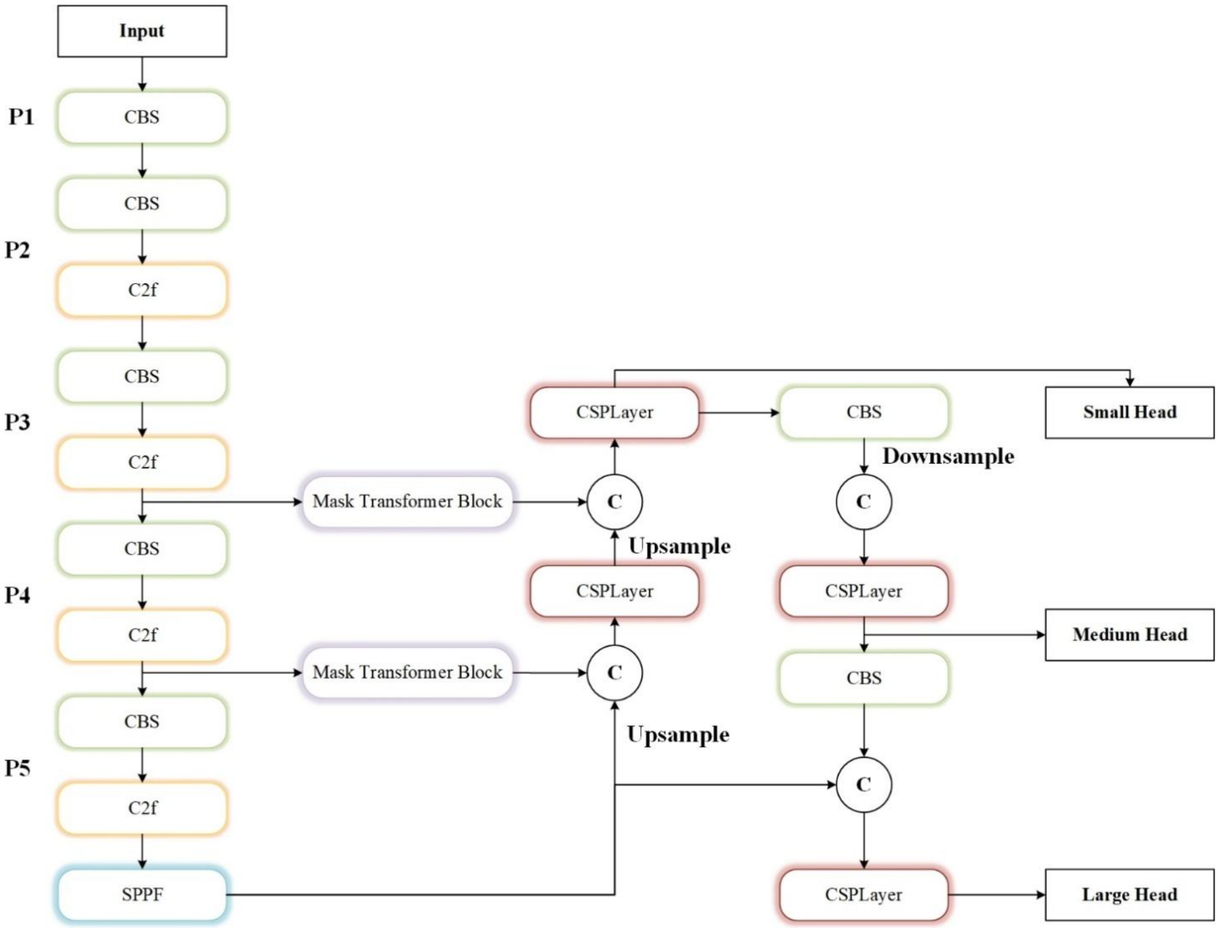


Figure 1. Overall network structure

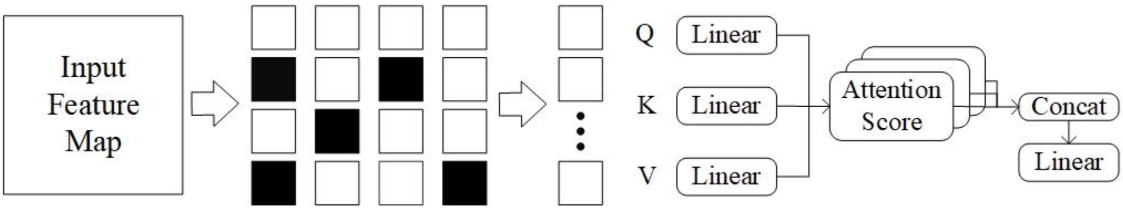


Figure 2. Mask Transformer block

2.4. Experiments

2.4.1. Evaluation metrics.

Precision indicates the proportion of targets predicted to be positive samples that are actually positive samples, and is used to measure the detection accuracy of the model. recall reflects the proportion of actual positive samples that are correctly detected, and measures the detection coverage of the model. mAP50 refers to the calculation of the average precision (AP) of all categories and averaging the values to evaluate the overall detection capability of the model when the IoU (Intersection over Union) threshold is set at 0.5. The mAP_{50} refers to calculating the AP of all categories when the IoU (Intersection over Union) threshold is set to 0.5 and finding the mean value to evaluate the overall detection capability of the model. The $mAP_{50\text{-}95}$ is further extended by averaging the APs computed in the IoU range of 0.5 to 0.95, which is a more comprehensive measure of the model detection performance at different IoU thresholds. The formula for calculating the indicator is as follows:

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

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

$$mAP = \frac{\sum_{i=1}^{K} AP_i}{K} \quad (3)$$

2.4.2. Experiment setup.

All experiments in this paper were implemented on an NVIDIA 3070ti GPU with 8G of video memory.Pytorch version 2.1.0 and CUDA version 11.8.The batch size is 24, the image size is uniformly 500 × 500 pixels, the epochs are 300 and the optimizer is SGD.

2.4.3. Results.

As shown in the Table 2 below, the overall accuracy evaluation of the metrics on the test set Precisio, Recall, mAP_{50} and mAP_{50-95} reached 0.901, 0.878, 0.924 and 0.858, respectively. the best performer was the silver carp, with a mAP_{50} of 0.995, and the worst one was the black carp mAP_{50} This is mainly due to the fact that the black carp image samples were mostly taken from different individual fish with slight differences between individuals. On the other hand, silver carp are mostly taken from multiple viewpoints of the same fish, resulting in a variety of gesture forms, which can be better adapted to random fish activities during the detection process.

Table 2. Accuracy evaluation results

Species	Precision	Recall	mAP_{50}	mAP_{50-95}
grass carp	0.956	0.971	0.982	0.962
crucian carp	0.986	0.934	0.987	0.979
bighead carp	0.935	0.760	0.877	0.799
tilapia	0.921	0.981	0.987	0.962
black carp	0.677	0.791	0.790	0.688
cod	0.908	0.754	0.922	0.819
perch	0.964	0.933	0.956	0.931
scavenger	0.886	0.846	0.917	0.784
silver carp	0.985	0.998	0.995	0.995
catfish	0.793	0.815	0.825	0.665
all	0.901	0.878	0.924	0.858

The change process and accuracy change of the loss function are shown in Figure3. It mainly includes the localization loss (box_loss), which is used to measure the error between the prediction box and the labeled box; the classification loss (cls_loss) calculates whether the classification corresponding to the anchor box is correct. It can be seen that the loss function gradually decreases and stabilizes during the training process.

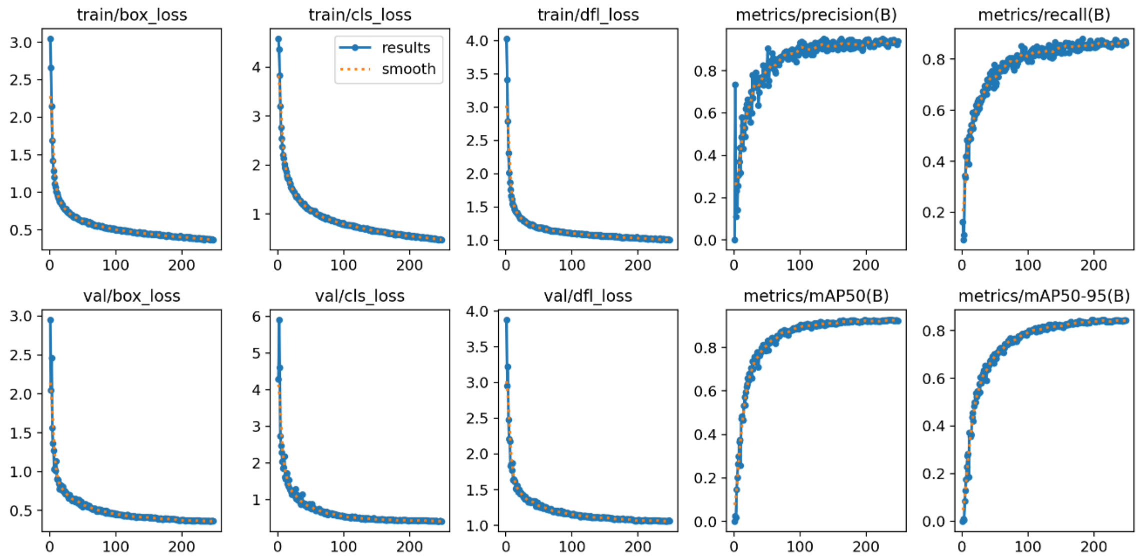


Figure 3. Loss function change curve

We test on an open source underwater fish motion target detection dataset. The dataset consists of 355 images, which are synthesized to form an underwater fish motion video at 20 photos per frame. Figure 4 shows the real-time extraction results obtained from the model trained by the method in this paper. The target can be accurately localized and classified during the complex swimming process of the fish. The fronts and backs of the fish, and the occlusion phenomenon between each other did not affect the recognition effect of the model.

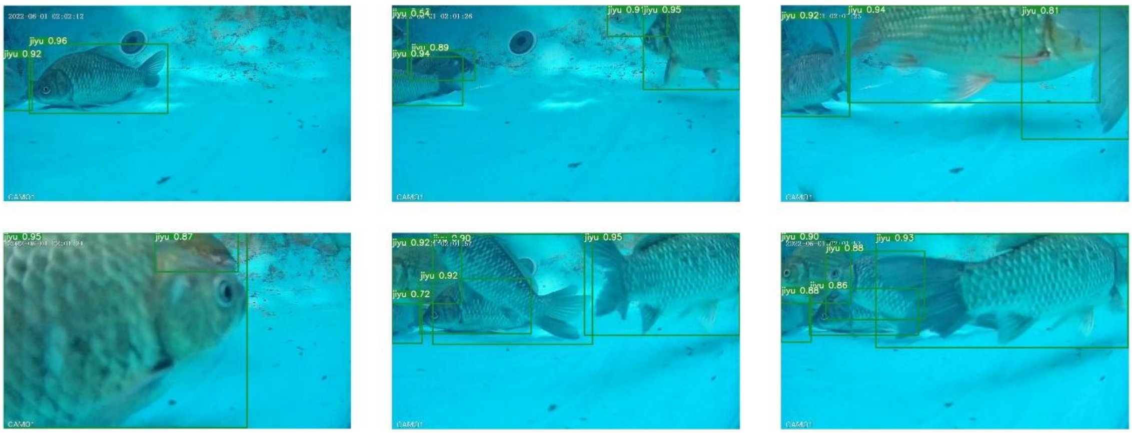


Figure 4. Fish test results

3. CONCLUSION

In this paper, a dataset of freshwater fish species was first constructed and the data were enhanced by adjusting brightness, adding gaussian noise and combining random angle rotation. Subsequently, a fish intelligent recognition model was constructed based on the Yolov8 model and an improved self-attention, which can be adapted to the recognition and statistics of freshwater fish species in underwater environments. Finally, on the test set, the evaluation metric Precision\Recall\ mAP50\ mAP50-95 reached 0.901\ 0.878\ 0.924\ 0.858.

REFERENCES

[1] Grabowska, Joanna, J. Kotusz, and A. Witkowski. "Alien invasive fish species in Polish waters: an overview." Folia Zoologica 59(2010):73-85.

[2] Shuai, Xinhui Zhao, Tian. "Biological invasions undermine the functional diversity of fish community in a large subtropical river." Biological invasions 20.10(2018).

[3] Redmon, Joseph, et al. "You Only Look Once: Unified, Real-Time Object Detection." Computer Vision & Pattern Recognition IEEE, 2016.

[4] Redmon, Joseph, and A. Farhadi. "YOLOv3: An Incremental Improvement." arXiv e-prints (2018).

[5] Li, Chuyi, et al. "YOLOv6: A single-stage object detection framework for industrial applications." arXiv preprint arXiv:2209.02976 (2022).

[6] Yaseen, M. "What is YOLOv8: An in-depth exploration of the internal features of the next-generation object detector. arXiv 2024." arXiv preprint arXiv:2408.15857.

[7] Li, Xiang, et al. "Generalized focal loss: Learning qualified and distributed bounding boxes for dense object detection." Advances in neural information processing systems 33 (2020): 21002-21012.

[8] Han, Yanling, et al. "Underwater Holothurian target-detection algorithm based on improved CenterNet and scene feature fusion." Sensors 22.19 (2022): 7204.

[9] Wang, Jing, et al. "YOLOv6-ESG: A lightweight seafood detection method." Journal of Marine Science and Engineering 11.8 (2023): 1623.

[10] Wang, Jun, et al. "B-YOLOX-S: a lightweight method for underwater object detection based on data augmentation and multiscale feature fusion." Journal of Marine Science and Engineering 10.11 (2022): 1764.

[11] Ren, Liqiu, et al. "An Underwater Target Detection Algorithm Based on Attention Mechanism and Improved YOLOv7." Computers, Materials & Continua 78.2 (2024).

[12] Wu, Fei, et al. "Fish target detection in underwater blurred scenes based on improved YOLOv5." Ieee Access 11 (2023): 122911-122925.

[13] Vaswani, Ashish, et al. "Attention is all you need." Advances in neural information processing systems 30 (2017).

[14] He, Kaiming, et al. "Masked autoencoders are scalable vision learners." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022.