IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 23 October 2023, accepted 12 December 2023, date of publication 15 December 2023,
date of current version 22 December 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3343429

APPLIED RESEARCH

An Automatic Detection and Counting Method for Fish Lateral Line Scales of Underwater Fish Based on Improved YOLOv5

HUIHUI YU[1,2,6], ZIMAO WANG[2,3,4,5], HANXIANG QIN[2,3,4,5], AND YINGYI CHEN[2,3,4,5]

[1]School of Information Science and Technology, Beijing Forestry University, Beijing 100083, China
[2]National Innovation Center for Digital Fishery, Beijing 100083, China
[3]Key Laboratory of Smart Farming Technologies for Aquatic Animal and Livestock, Ministry of Agriculture and Rural Affairs, Beijing 100083, China
[4]Beijing Engineering and Technology Research Center for Internet of Things in Agriculture, Beijing 100083, China
[5]College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China
[6]National Forestry and Grassland Administration Engineering Research Center for Forestry-Oriented Intelligent Information Processing, Beijing 100083, China

Corresponding author: Yingyi Chen (chenyingyi@cau.edu.cn)

This work was supported in part by the National Natural Science Foundation of China "Intelligent Identification Method of Underwater Fish Morphological Characteristics Based on Binocular Vision" under Grant 62206021, in part by the Beijing Digital Agriculture Innovation Consortium Project under Grant BAIC10-2023, and in part by the National Natural Science Foundation of China "Analysis and Feature Recognition on Feeding Behavior of Fish School in Facility Farming Based on Machine Vision" under Grant 62076244.

ABSTRACT The lateral line scales of fish are an important phenotype of fish species. As an important countable feature, the accurate and effective counting of lateral line scales is an important reference standard for breeding, determining the growth status of fish, and identifying fish species. At present, the statistical work of fish lateral line scales mainly depends on manual statistics and semi-automatic methods, which cannot meet the current needs of sustainable development and precision digital fisheries. The method based on computer vision and deep learning can provide an a real-time, efficient and non-contact method for identifying and counting fish lateral line scales. However, it is still a challenge due to the high similarity between fish scales and the variable size issues caused by the free movement of the fish. Hence, we proposed a transformer module improved YOLOv5 model (TRH-YOLOv5) for fish lateral line scale detection and counting, which focus on the high similarity of fish scales. In addition, we design a small target detection module in the head layer to address the challenge of multi-scale fish. To evaluate the effective of our method, performance of proposed model is analyzed on different type fish dataset and it is also compared with classical method including SSD, YOLOv4 and YOLOv5. Comprehensive experimental results show that the proposed model achieves fine results (e.g., 98.8% precision, 96.7% recall and 99.0% mean average precision) with relatively lower computational coat (e.g., 16.1M model size) and fast detection speed (e.g., 37 FPS) compared with the benchmark algorithm. The TRH-YOLOv5 model is also used for swimming fish video to detect fish lateral line in real-time and can be integrated into aquaculture vision system for aquaculture precision and sustainable management.

INDEX TERMS Detection, fish lateral line scales, YOLOv5, transformer, aquaculture.

I. INTRODUCTION

The lateral line scales of fish are one of the most important fish phenotypes [1]. The detection and counting of fish lateral line scales play a vital role for breeding, growth status assessment and species identification of fish in aquaculture [2], [3]. Currently, the detection and counting techniques of fish lateral line scales primarily rely on manual identification and statistical counting. These techniques always directly contact to fish and detect with naked eye [4]. These contact recognition techniques are time-consuming, labor-intensive and fish body injuries and they cannot satisfy the

The associate editor coordinating the review of this manuscript and approving it for publication was Marco Martalo.

requirements of efficient and intelligent in modern aquaculture [5], [6], [7]. Recently, the semi-automatic techniques have been applied to the identification of fish body phenotypes, e.g., body size, tail and head distribution and fish lateral line scales detection [8]. However, it is highly susceptible to external equipment, environmental interference and subjective factors such as the habits, experiences, and preferences of the collector [8], [9]. For above reasons, the development of non-contact measurement methods is urgent and necessary to replace the directly manual methods [10]. In recent years, computer vision technology and deep learning develop rapidly. These technologies have been fusion applicated in aquaculture production grading, phenotypic feature acquisition and fish size measurement [9], [11], [12]. These researches provide a well way and idea for non-contact detection of fish lateral scales and counting for aquaculture intelligent development.

In terms of fish phenotypic extraction and recognition, White et al. used image binarization method to detect the direction of the fish head and tail and used these intersecting points and information to calculate the final fish body length [13]. With the development of the optical ranging and automatic data acquisition technologies, Costa et al. measured the length and shape of northern Bluefin tunas in a sea cage using underwater binocular cameras [14]. These methods had a confidence level of up to 95% accuracy, reducing measurement costs and improving measurement speed. Compared to manual feature detection methods, the above methods have made significant improvements, but underwater environment, multi-scales and swimming status of fish add difficulties and limit the application for the traditional image processing technology [15]. Deep learning networks (DLN) emerge as a promising solution to address these challenges. For fish biomass estimation, Abinaya et al. presented a segment analysis technique based on YOLOv4 to determine the length and biomass of fish for health and growth rate during fish growing stages [16]. For fish texture phenotypic features, Maurya et al. proposed a color texture feature extraction method based on genetic optimization and a method based on transfer learning to efficiently identify fish color and texture features [17]. In addition, in the actual aquaculture production process, Banwari et al. used computer vision to predict the freshness of fish by extracting phenotypic features of fish eyes [18]. Liao et al. developed 3DShenoFish software based on deep learning to address the issue of automatic measurement of morphological features, extracting the morphological phenotype of fish from 3D point cloud data [19].

These previous researches of fish body phenotype based on computer vision mostly focuses on morphological parameter features, such as body length, head length, and tail stalk width data [5], [8], [20]. There are few researches on the phenotype recognition and counting of lateral scales. The only computer vision-based recognition methods for fish lateral scales remain in the early stages of recognition and cannot complete the counting function. The scales of fish lateral scales is roughly small in fish image and the edge of fish lateral scales is extremely similar [21]. These factors are the challenges for the fish lateral scales detection in the underwater environment [22]. In order to achieve non-contact and precision detection technique of fish lateral scales in the reality aquaculture environment, we explored the application of deep learning in animal detection using the cutting-edge object detection framework YOLOv5 and transformer self-attention mechanism. We proposed an TRH-YOLOv5 model to non-contact detect and count fish lateral scales in the complex aquaculture scene.

The contributions of this paper are as follows: (1) We integrated the transformer self-attention mechanism (TR) and added a small target detection layer on this basis to improve the overall feature extraction ability of the model, greatly improving the recognition rate of fish lateral line scales. (2) Establish a challenge dataset for fish lateral line scale detection and counting. We collected image data of fish lateral scales in different scenes and conducted effectiveness screening on experimental data. Additional, in order to avoid too single dataset image and ensure the reliability of the data, brightness, adaptive histogram equalization, enhance edges, Gaussian noise in six methods of image preprocessing, including horizontal and vertical flipping, are used to expand the original image data and ultimately construct fish lateral line scale datasets.

II. MATERIAL AND METHODS

A. DATA ACQUISITION AND PREPROCESSING

1) EXPERIMENT AND IMAGE TYPE

The fish body lateral line scale data collection was carried out at the Hebei Zhuozhou Digital Fishery Precision Technology Integration Base. The collection period is from July 19, 2022, to August 2, 2022. The collection objects are 8-23 cm crucian carp and koi fish, including a total of 18 black crucian carp, 10 red gill crucian carp, 1 pure white crucian carp, 6 red and white koi, 2 tri-color koi, and 1 yellow koi. The SONY IMX686 camera is used for data collection through shooting, and the data collection includes multi-directional, multi-species, different distance, different scene, and different size fish body lateral line scale images and real-time videos. A total of 1903 image data files and 16 video data files were collected, with image formats as JPG, resolution of 4624 pixels $\times$ 2604 pixels per image; video format is MP4, with each video having a frame width of 1920 and a frame height of 1080.

The specific collection environment is shown in Fig.1, which displays the equipment, scene, and collection of three types of fish photos required for the experimental data. The first category is static lateral line scale images. These images are collected by setting different vertical distances (three distances: 15cm, 30cm, 45cm), different lighting conditions (three levels: dim, weak, and strong), different angles (two types: front and oblique side), and different vertical positions (four positions: upper left, upper right, lower left, lower right). The second category of images are fish in

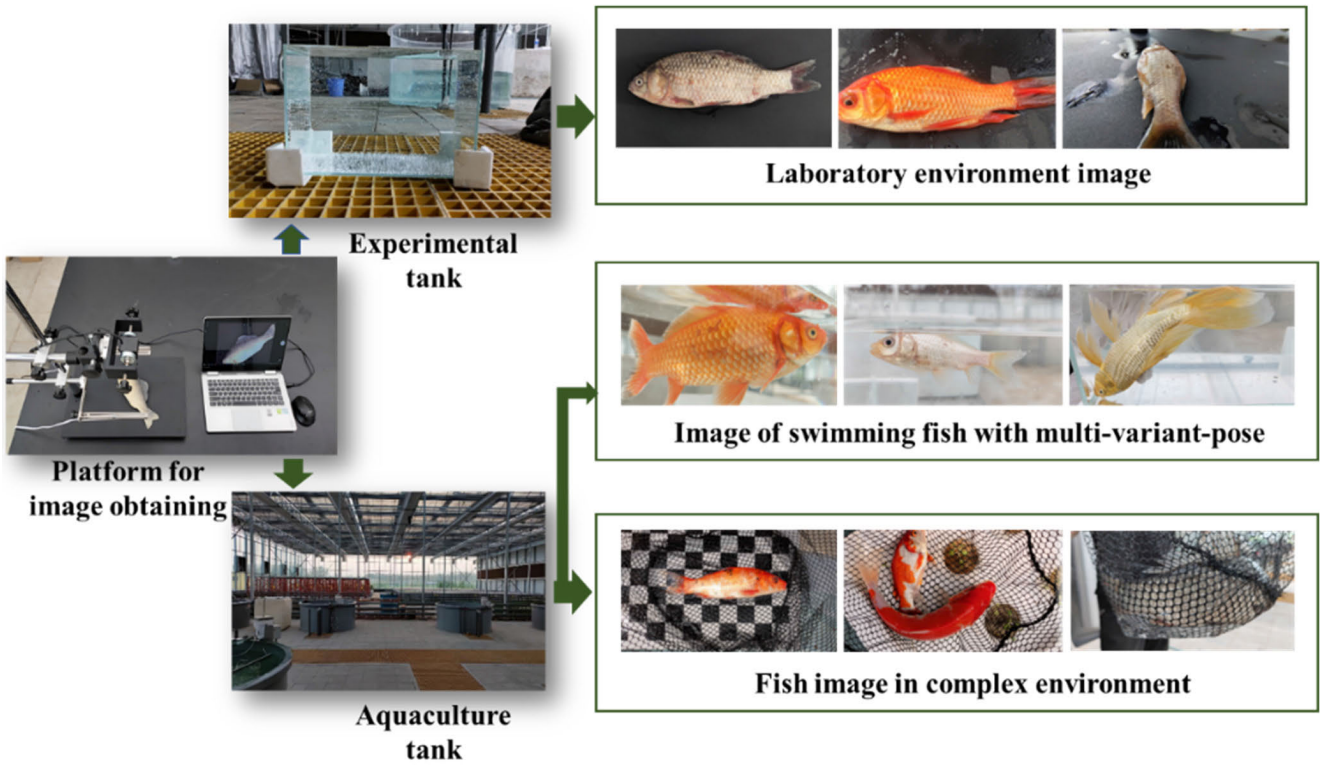


FIGURE 1. The structure of experiment system and categories of fish lateral line.

motion. Ten representative fish are randomly selected and placed in a glass tank, and images and videos of freely swimming fish are collected outside the glass tank. This category of photos is taken in various environments, such as a calibration board background, multi-fish culture pond background, fishing background, artificial holding background, and natural environment background. In order to increase the effective and rich dataset, enhance the dataset's robustness, the experiment adds data collection from multiple scenes, and this type of image is the third category of images.

2) DATASET OF FISH LATERAL LINE BASED ON DATA AUGMENTATION

In order to improve the quality of fish lateral line scale datasets, this study first cleans the collected 1903 images to remove images with high repeatability, indistinguishability, or significant distortion. In the end, 233 images are selected for the future model building. In order to avoid a simple dataset structure and reduce the impact of the dataset on the accuracy of the later fish lateral scale recognition and counting model construction. Six image pre-processing methods, including adding random brightness(Bright), adaptive histogram equalization(Ahe), enhancing edges(Edge), adding Gaussian noise(Noise), horizontal flipping(H-Flip) and vertical flipping(V-Flip) are used as the data augmentation methods in this study [23]. The pre-processed images (random selected three images) are shown in Fig.2. The top line is the initial images; the other lines are pre-processed images. Adding Random Brightness is designed to simulate the diversity of light intensity in real-world scenes, and the fact that the surface of the fish body will experience different levels of reflection and chose a method of adding random brightness for data processing. Adaptive Histogram Equalization is compensated for some of the blurring issues of the lateral line scale images and use the method of adaptive histogram equalization for data processing. Edge Enhancement used edge enhancement for data processing to increase the distinguishability of the fish body's lateral line scales in the image and the usage rate of some blurry images. Adding Gaussian Noise chose Gaussian noise for data processing to meet the varying image quality due to differences of image acquisition equipment and environmental conditions. Horizontal flip (H-Flip) is done by flipping the image 180 degrees on its vertical axis, and vertical flip (V-Flip) is done by flipping the image 180 degrees on its horizontal axis, to enhance data diversity.

Finally, a total of 1615 images and annotation files were obtained through image annotation. And it is divided into training set, testing set, and validation set in a 3:1:1 ratio, forming a scientific fish lateral scale dataset. (The data divided ratio is according the book ''Machine Learning Yearning'' of Andrew Ng and his instructional video. when the amount of data is not very large (below ten thousand), the training set, test set, and validation set can be divided into a

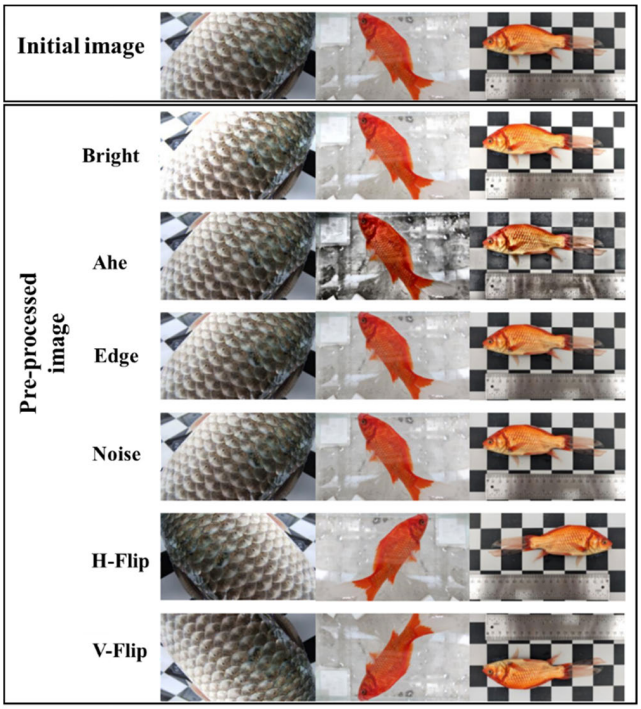


FIGURE 2. Results of different data augmentation techniques for fish lateral line image.

6:2:2 ratio. If the amount of data is large, the data set can be adjusted to a 98:1:1 ratio.)

B. TRH-YOLOv5 METHOD FOR DETECTION OF FISH LINE SCALES

YOLOv5 is a standard convolutional neural network that performs various convolution operations, pooling operations, and result output through a fully connected layer on input three-channel images [24], [25]. It adopts the Path Aggregation Network (PANet) structure [26], which leads to insufficient fusion of multi-scale features. YOLOv5 mainly includes three parts: the Backbone layer, the Neck layer and the Head layer. The Backbone layer continuously extracts key and general features through convolutional down sampling. The Neck network layer is used for further feature extraction [27]. It includes two parts: the left FPN and the right PAN. The FPN extracts features at different scales in the image by constructing a feature pyramid with different resolutions. The right-side PAN obtains multiscale information through a bottom-up path aggregation module. Finally, the Head layer is used for object detection and output of corresponding final detection results, converting the three feature maps extracted by the backbone network into the final target detection results. We chose the smallest YOLOv5s as the base model, which ensures detection accuracy while saving detection time and model space size.

In the fish body lateral line scale detection aspect, due to the small scale of lateral line scale targets and the small interclass differences between other scales, it increases the difficulty of lateral line scale detection and recognition for dynamic fish bodies. Therefore, a TRH-YOLOv5 fish body lateral line scale detection model is proposed in this paper. The overall structure of proposed method is illustrated in Fig.3.

To solve the problem of the extremely similarity of fish lateral line scales boundaries, the Transformer Encoder is fused with the C3 module in the eighth layer of the Backbone layer, forming a new module to further improve feature extraction capability in YOLOv5-6.1. The features obtained from the C3 module at the seventeenth layer of the Neck layer are then convoluted and up-sampled. Afterward, the same-scale features from the third layer of the backbone network are fused by concat, forming the feature map corresponding to the small object detection layer. Subsequently, the small object detection layer (Head) is built in the detection network, completing the overall construction of the model.

1) IMPROVE THE BACKBONE OF YOLOv5 BY TRANSFORMER MODULE

Transformer is a neural network model based on self-attention mechanism, which has long been applied in the field of Natural Language Processing (NLP) [28]. It is implemented through the transmission between encoder-decoder repetitive units and can be regarded as a combination of two recursive neural network substructures. The detail algorithm is described by Dai et al. [29]. Our main focus is on the left-side Encoder. In the encoder, each encoder has two sublayers: the first sublayer is a multi-head attention mechanism layer, i.e., the Multi-Head Attention module; the second sublayer is the feed-forward neural network layer MLP module. Its advantage lies in enabling the model to cover the global image better and obtain rich contextual information [30].

The most critical challenge is to effectively combine Transformer with computer vision and convert the features of processing text data in the original model to image data. Therefore, the paper simulates the application of Transformer in natural language processing tasks, and when processing images, a one-dimensional vector still needs to be input, so the input image needs to be cropped. First, assume that the input raw fish lateral line scale image size is H×W×C, where H is the image height, W is the image width, and C is the image depth. The image is cropped into nine different image patches through the Patch operation, with each image patch size being P×P, where P is a preset fixed value, i.e., the height or width of the obtained image patches. The number of patches is as Eq.(1):

$$N = H \times \frac{W}{P^2} \tag{1}$$

The lateral line scale images are cropped into nine different image patches through the Patch operation. Then each lateral line scale image patch is flattened into a one-dimensional vector through the Flatten operation. The class token and positional encoding are combined with it to generate a new vector, transforming the lateral line scale

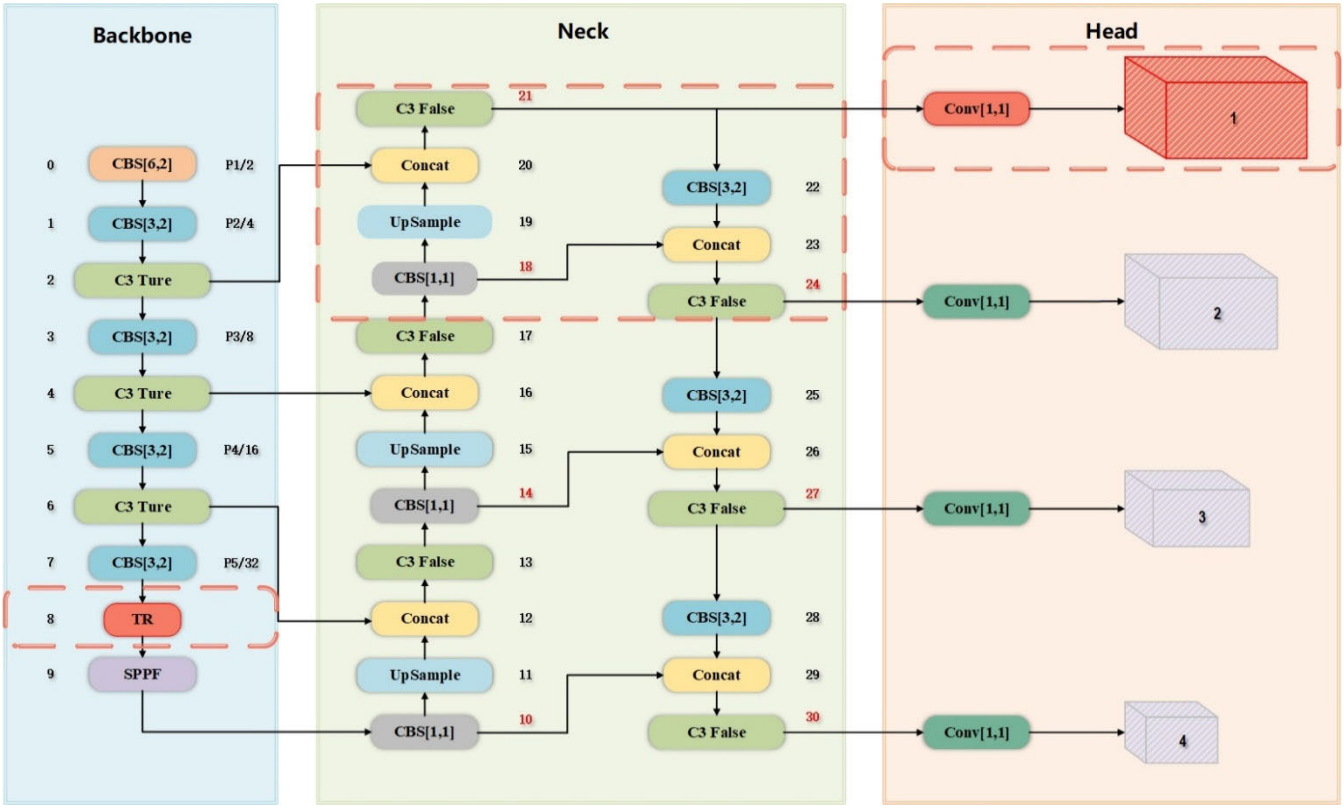


FIGURE 3. An overview of the TRH-YOLOv5. TR represents the Transformer self-attention mechanism module; H represents the small object detection layer.

image of the fish into a one-dimensional vector with category and positional features. This serves as the input for the model, achieving the integration of Transformer with visual information.

Inspired by the above Vision Transformer, this model addresses the multi-scale target and the problem of being easily confused in fish lateral line scale phenotype recognition and counting. The last C3 bottleneck block in YOLOv5-6.1 version is fused with the Encoder in Transformer to form a new TR module at the last layer of the backbone network. Its structure is shown in Fig.5. This process mainly takes into account that the last C3 layer has higher semantic information, which can better cover the entire image and obtain rich context information. Particularly, it can better capture the difference information at the boundary of the lateral line scales. In contrast, the shallow semantic information obtained from the early C3 layers, if an encoder is added to each layer, will certainly result in a large number of parameters and slow calculations. Therefore, centered around the study's requirements to automatically and real-time detect lateral line scales, an encoder module was added at the high-level semantic location. Compared with the original C3 module, after integrating the Transformer Encoder, its self-attention mechanism and other characteristics enhance the model's ability to capture different feature information. It can better cover the global image and obtain rich contextual information.

2) IMPROVED NECK OF TRH-YOLOv5

Due to the different species and sizes of fish, the corresponding lateral line scale sizes are multiple scales in different data images and have many small targets. In the model training and detection process, the model goes through multiple convolution operations from bottom to top, causing some feature information loss in the shallow feature map [31]. Although the resolution of the P3 detection layer is 80 × 80 pixels, its detection capability is still limited.

To achieve better recognition and counting results for multi-scales and small lateral line scales, we designed a new small object detection layer, the P2 detection layer, to make full use of shallow semantic information as shown in the Fig.6. Its resolution is 160×160 pixels, which can be regarded as having only performed two convolution operations in the Backbone layer and containing more comprehensive shallow object feature information. In the Neck layer, two same-scale features obtained from FPN and PAN methods and P2 feature layer are merged through concat concatenation, and the final fusion result is output. When the model deals with small-scale fish lateral line scale targets, it can use the P2 detection layer for accurate detection and improve recognition and counting accuracy.

3) PERFORMANCE EVALUATION

This study uses the following seven indicators to evaluate the model: Precision, Recall, Mean Average Precision

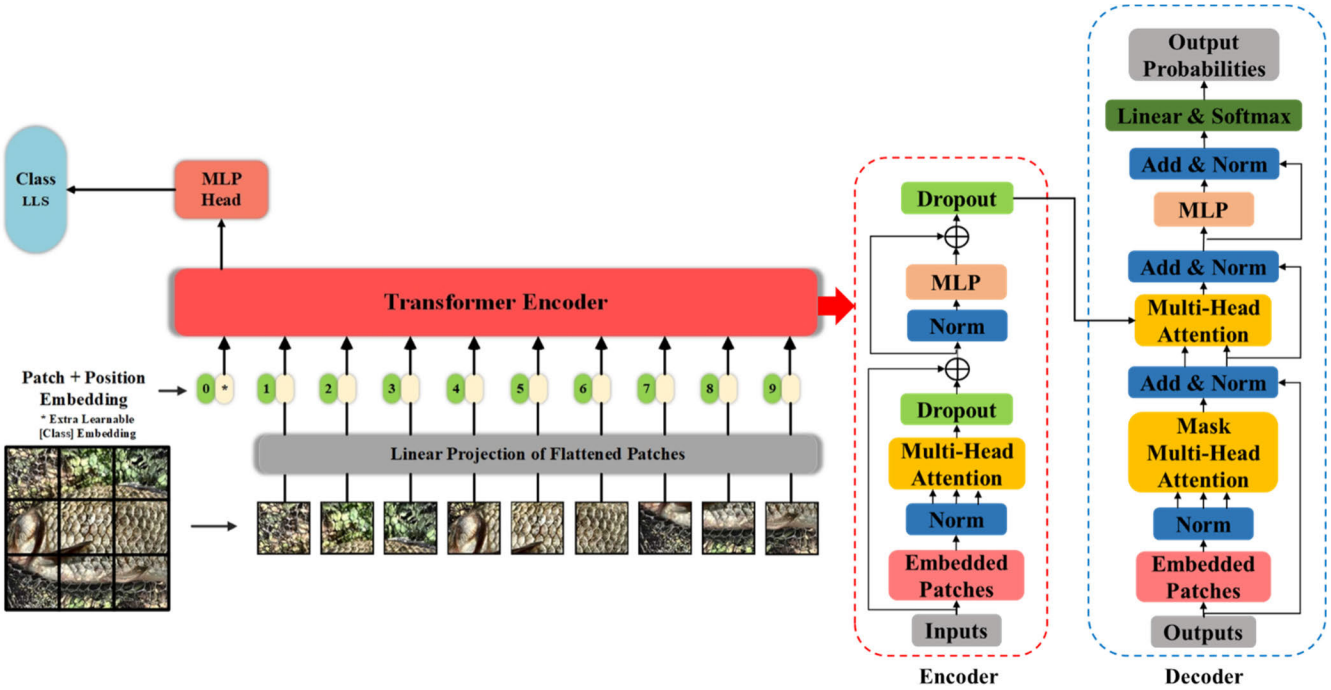


FIGURE 4. Transformer structure diagram and fish body lateral line scale image Transformer conversion principle.

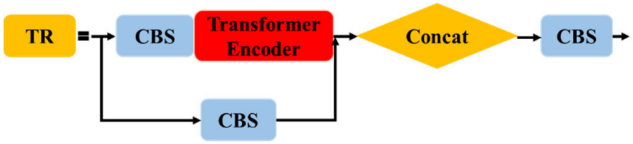


FIGURE 5. Transformer self-attention mechanism (TR module).

(mAP), Model Size, Model Parameters (Params), Floating-point Operations per Second (FLOPS), and Frames per Second for Single-frame Image Inference (FPS). The specific expressions of the evaluation indicators are as follows:

(1) Precision: The ratio of correctly predicted positive samples to all samples predicted as positive, that is, how many of the samples predicted as positive are actually positive.

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

(2) Recall: The ratio of correctly predicted positive samples to the total number of real positive samples, that is, how many lateral line scale positive samples the model can predict correctly from these samples.

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

In the above formulas, where TP, FP and FN represent true positives, false positives and false negatives.

(3) Mean Average Precision (mAP): This indicator is not an absolute measure of the model, but it can relatively reflect the performance of the model.

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

$$AP_{50:95} = \frac{1}{10}(AP_{50} + AP_{55} + \ldots + AP_{90} + AP_{95}) \tag{5}$$

$$mAP = \frac{1}{|Q_R|}\sum\nolimits_{q \in Q_R} AP(q) \tag{6}$$

$$mAP_{50:95} = \frac{1}{10}(mAP_{50} + mAP_{55} + \ldots + mAP_{90} + mAP_{95}) \tag{7}$$

In the above formulas, AP_{50} is the average accuracy at $IOU = 0.5$. The mAP_{50} is the average AP of all categories at $IOU = 0.5$. The $AP_{50:95}$ is the average accuracy at $IOU = 0.5$ to $IOU = 0.95$ with an interval of 0.05, and the $mAP_{50:95}$ is the average AP of all categories at $IOU = 0.5$ to $IOU = 0.95$ with an interval of 0.05. Among the above three indicators, the larger the value, the better the performance of the model. In this study, the mAP_{50} is used for the model evaluation.

III. RESULT AND DISSCUSION

A. NETWORK TRAINING PARAMETERS

Since many important parameters are involved in the training of the TRH-YOLOv5 object detection model, mainly including: Batch-size, learning rate, and optimizer type. Changes in these parameters will directly affect the accuracy and speed of model training. Therefore, a comparison experiment of parameter settings is conducted in this section to set the optimal parameters.

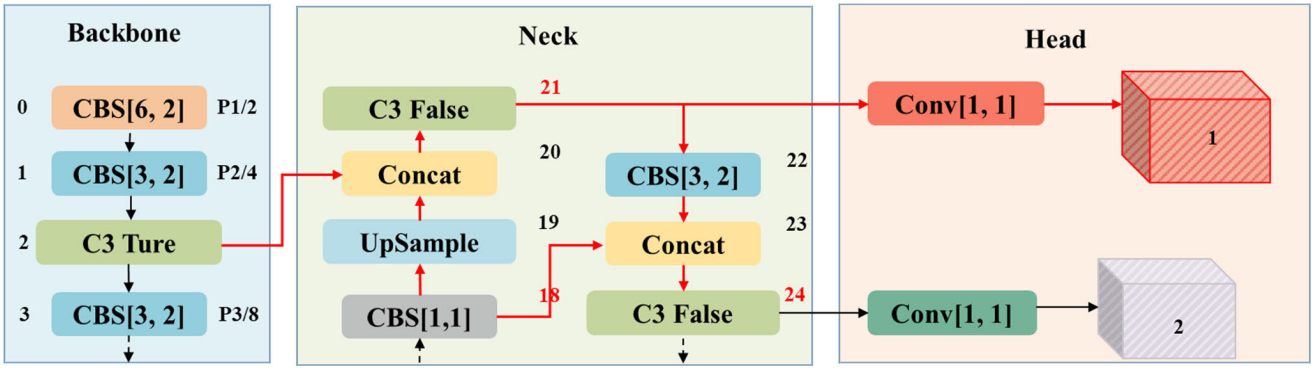


FIGURE 6. Improve head layer for small fish lateral line detection.

In order to optimize the batch size and learning rate for the proposed model, the SGD optimizer was used as the default optimizer for comparative experimental analysis. SGD is a classic optimizer that minimizes the loss function of the model by adjusting parameters through gradient descent. The advantages of SGD are simple implementation and high efficiency, but it converges slowly and is prone to local optima under some special scenarios or dataset conditions. Adam is an optimizer that approximates random gradient descent, which adjusts model parameters by maintaining the first-order and second-order momentum of the gradient and gradient square of the model. Adam has the advantage of fast convergence speed. However, the adjustment of hyperparameters increases the complexity of model construction. In the third part of this section, a comparative experiment was conducted on the training of SGD and Adam optimizer models, and it was found that the convergence speed and convergence effect of SGD optimizer were better in the process of fish lateral scale detection.

1) BATCH-SIZE SETTING

Batch-size is the number of samples selected in a single training session for a model. Its reasonable adjustment can not only reduce memory usage but also improve training speed to some extent. Therefore, a Batch-size comparison experiment is conducted. According to the control variable approach and method, while keeping the learning rate at 0.01 and using the default SGD optimizer, the model is trained for 300 epochs with Batch-size set to 4, 8, 16, and 32 respectively. The experimental results can draw the conclusion, as shown in Fig.7. It can be observed that when Batch-size is set to 4, the convergence speed of this model is the fastest, but due to its doubled training time and considering the objective conditions of hardware devices, the overall best effect is achieved when the Batch-size is set to 8.

As shown in Table 1, a faster convergence speed is shown when Batch-size is set 8, the model also performs outstandingly in evaluation indicators such as Precision (P), Recall (R), and Mean Average Precision (mAP_{50}).

TABLE 1. TRH-YOLOv5 detection results for fish lateral line with different Batch-size.

Batch-size	P	R	mAP_{50}	Training Time（h）
4	98.3%	97.0%	98.9%	47.93
8	98.8%	96.7%	99.0%	25.82
16	98.5%	97.3%	99.0%	22.09
32	98.9%	97.3%	99.1%	21.39

TABLE 2. TRH-YOLOv5 detection results for fish lateral line with different learning rate.

Learning rate	P	R	mAP_{50}	Training time（h）
0.01	98.8%	96.7%	99.0%	25.82
0.001	94.5%	90.7%	96.5%	29.60
0.0001	80.9%	66.3%	79.5%	29.06

2) LEARNING RATE

The learning rate is an important hyperparameter, and the size of its parameter value determines the step size of each training iteration, which can make the loss function converge to the minimum. Therefore, a learning rate comparison experiment is conducted. According to the control variable approach and method, while keeping Batch-size = 8 and using the default SGD optimizer, the model is trained for 300 epochs with learning rates set to 0.01, 0.001, and 0.0001 respectively, as shown in Fig.8. The experimental results can draw the conclusion: It can be observed that when the learning rate is set to 0.01, the convergence effect of the loss function of this model is the best.

As shown in Table 2, the learning rate is set to 0.01, the model achieves the best performance in comprehensive evaluation indicators such as Precision (P), Recall (R), and Mean Average Precision (mAP_{50}).

3) OPTIMIZER

Optimizers play a very important role in the process of data training and model construction. Keeping the Batch-size

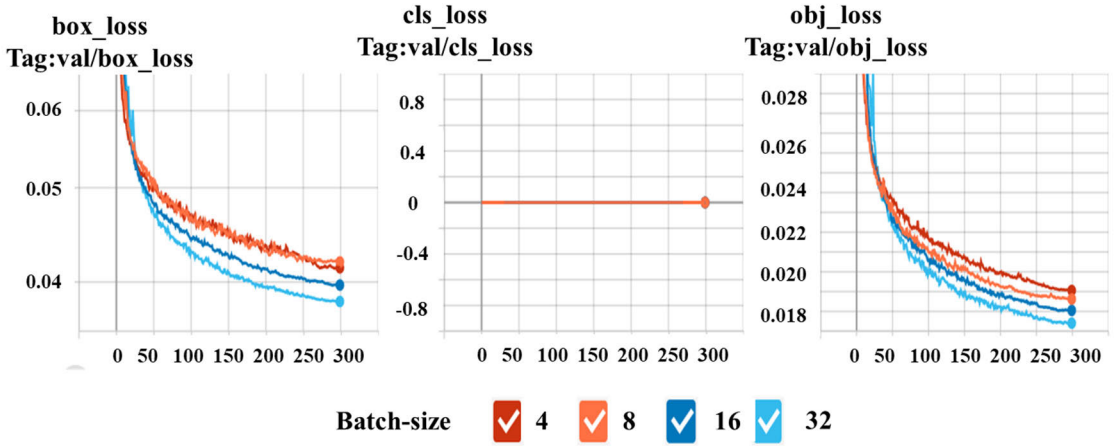


FIGURE 7. Convergence speed with different batch-size.

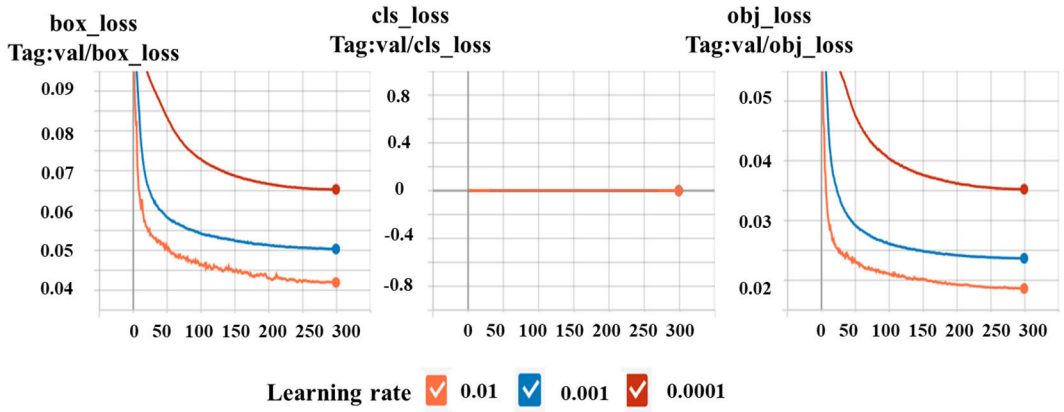


FIGURE 8. Convergence speed with different learning rate.

at 8 and the learning rate at 0.01, the default SGD optimizer and the Adam optimizer are set for model training, with the number of training rounds set to 300 epochs, as shown in Fig.9. The experimental results can draw the conclusion: When using the default SGD optimizer, the convergence speed of the model is the best, that is, the most effective.

In addition to showing a faster convergence speed when using the default SGD optimizer, the model also performs outstandingly in comprehensive evaluation indicators such as Precision (P), Recall (R), and Mean Average Precision (mAP_{50}) as shown in Table 3.

B. MODEL ABLATION EXPERIMENT

TRH-YOLOv5 is based on the YOLOv5 model. To evaluate the effectiveness of different modules in our proposed approach, some strategies are implemented on fish school feeding behavior dataset, such miss Transformer module, improved head module and both miss the two modules. Based on the above training strategy, we obtain the ablation results for fish lateral line detection.

TABLE 3. TRH-YOLOv5 detection results for fish lateral line with different optimizers.

Optimizer	P	R	mAP_{50}	Training time（h）
SGD	98.8%	96.7%	99.0%	25.82
Adam	97.6%	95.1%	98.2%	25.93

As shown in Table 4, when the fish lateral line datasets are executed on the baseline YOLOv5, the model accuracy in the validation the model precision and mAP_{50} reach 97.4% and 95.3%. To enhance the small object detection precision, the improve head layer is added to the baseline model and the precision and mAP_{50} can reach 98.2% and 98.8%. The detection result accuracy of fish lateral line is improved 3.5%. When just the transformer module is added to the baseline model (TR-YOLOv5), the precision and mAP_{50} can reach 97.3% and 95.3%. The accuracy of TR-YOLOv5 is not significant. The TR module and H module are both added to the YOLOv5, the precision increase from 97.4% to 98.8%

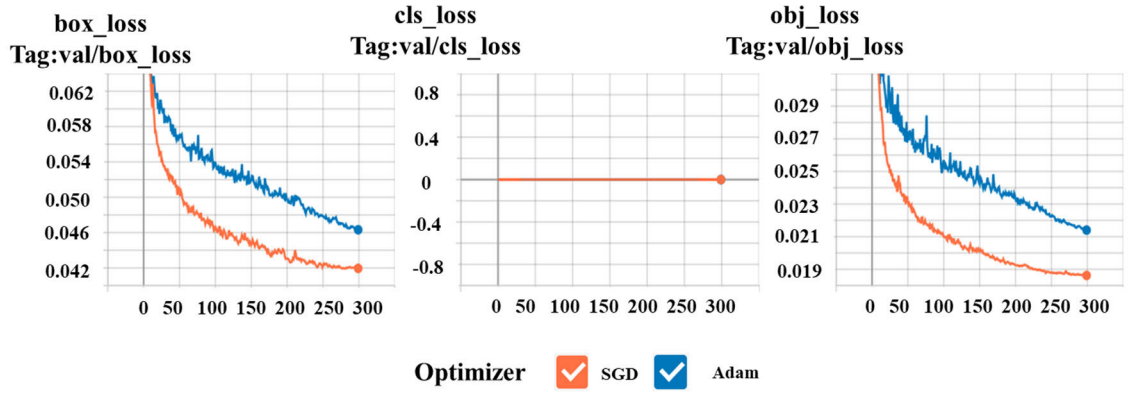
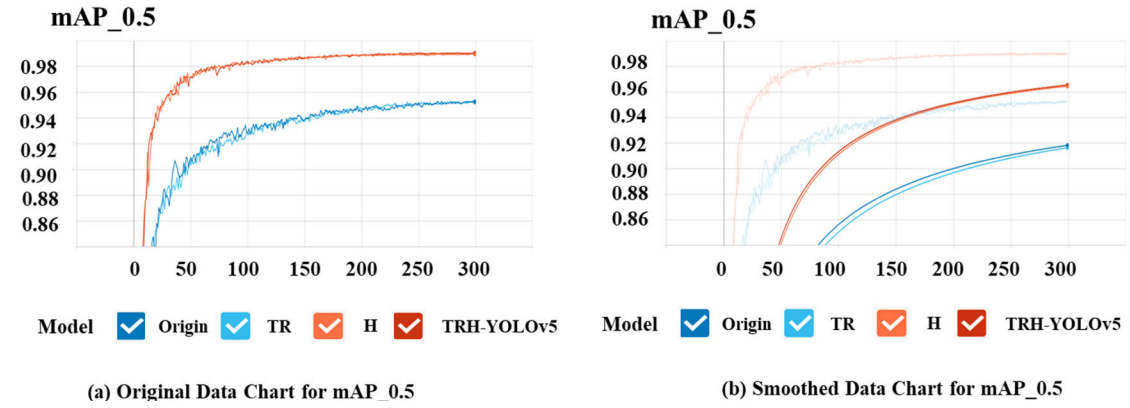


FIGURE 9. Convergence speed with different optimizers.

TABLE 4. Model improvement ablation experiment.

	Transformer (TR)	H	P	R	mAP_{50}	Size (MB)	FPS
YOLOv5s			97.4%	88.5%	95.3%	14.4	42
		√	98.2%	96.5%	98.8%	16.5	35
	√		97.3%	87.9%	95.3%	14.5	49
	√	√	98.8%	96.7%	99.0%	16.1	37



(a) Original Data Chart for mAP_0.5

(b) Smoothed Data Chart for mAP_0.5

FIGURE 10. Model training process comparison in ablation experiment. (a) The original data chart is constructed using the original data in the model building process, and (b) the smoothed chart processes the data of the original data chart for smoothing, making it easier to observe the results.

and the mAP_{50} increase from 95.3% to 99.0%. From the Size and FPS results, it is found that the TRH-YOLOv5 has a relatively smaller number of model parameters and a higher detection speed under the premise of prediction accuracy.

To explain this achievement, the accuracy curves of different models are produced and shown in Fig.10. Special, based on YOLOv5, the origin model and TR-YOLOv5 model, the accuracy curve can achieve convergence, but the final prediction result accuracy is relatively low. In comparison, the convergence speeds of TRH-YOLOv5 model and the H-YOLOv5 model are faster and the accuracies are higher than baseline model and the TR-YOLOv5 model.

C. COMPARING WITH OTHER CLASSIC DETECTION MODEL

In order to validate the reliability of proposed TRH-YOLOv5 model, our proposed model is compared with the following baseline: SSD, YOLOv4, and YOLOv5. These model are all has strong ability for multi-scale object detection. It is worth noting that the same experimental settings are presented in baseline to achieve a fair comparison.

The experiment results are illustrated in Table 5. It shows that the proposed TRH-YOLOv5 achieves the best performance on the test datasets in terms of precision and mAP_{50} with other baseline models. The SSD and YOLOv4 have the poorest performance on validation dataset. And,

TABLE 5. Model performance comparison table.

Model	P	R	mAP_{50}	Size (MB)	Params	FLOPS	FPS
SSD	-	37.9%	45.7%	90.6	-	-	8
YOLOv4	59.5%	94.9%	95.0%	244.4	-	-	14
YOLOv5s	97.4%	88.5%	95.3%	14.4	7012822	15.8	42
TRH-YOLOv5	98.8%	96.7%	99.0%	16.1	7672808	26.6	37

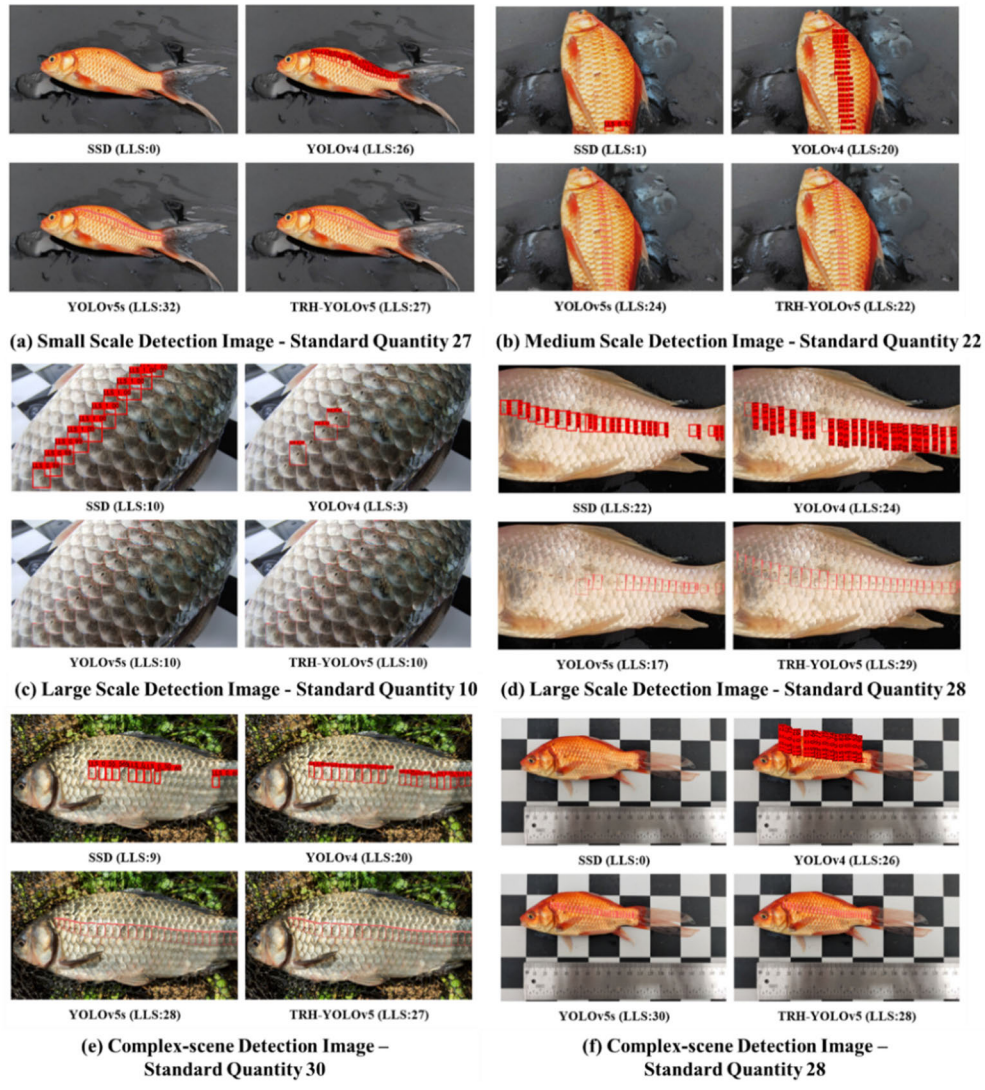


FIGURE 11. Comparative diagram of the detection effects of various models under different sizes and scenarios.

these two models have the largest model size and lowest detection speed. Compared with the baseline YOLOv5 model, precision of proposed TRH-YOLOv5 model increases from 97.4% to 98.8%, and the mAP_{50} increases from 95.3% to 99.0%. It shows the outstanding performance of the proposed model in detection accuracy and robust ability.

In order to verify the effectiveness of the proposed method in detecting the lateral line of fish in actual scene images, the paper tests the four models on selected images. Images of

fish at small scales, medium scales, large scales and complex scenarios were selected as verification images. The detection results are shown in Fig.11. In Fig.11 (a) and (b), for small and medium scale images, we can see that the SSD algorithm is almost ineffective, the YOLOv5 method has over-detection, and the method proposed in the paper detects correctly and counts accurately in all cases. In Fig. 11(c) and (d), the SSD detection of large scale different fish images is still not satisfactory, but the method proposed in this paper still correctly detects and counts all the images. In the re-examination scene of Fig.11 (e) and (f), the standard number of lateral line scales is 30 and 28. The method proposed in the paper did not accurately detect the entire image in (e), but achieved a detection accuracy of over 90%. Overall, the method proposed in the paper performed the best in different scenarios, different sizes, and different types of fish images for lateral line scale detection.

IV. CONCLUSION

The proposed TRH-YOLOv5 model aims to realize non-contact detection and counting of the fish lateral line scales in the actual aquaculture environment. In this study, based on the basic YOLOv5 model, a new target detection model was built to address the problem of the automatic identification and counting of fish body lateral line scales. The paper leverages the Transformer's self-attention mechanism as well as the fundamental principle and benefits of small target detection layer, thereby enhancing the recognition capability of easily confused small targets and significantly improving the accuracy of fish body lateral line scales' identification and counting. The experimental results showed that the TRH-YOLOv5 model performs best when the learning rate is set at 0.01, the Batch-size is set at 8, and the default SGD optimizer is used. While maintaining the same model size and detection speed, the detection accuracy mAP_{50} increased by 3.7%, and the recall rate R improved by 8.2%, demonstrating superior detection performance and identification accuracy for fish body lateral line scales. As tested on multiple datasets of image collections, the identification counting rate can reach 99%, showing excellent results. Therefore, the model built in this study is efficient and significant in terms of automatic recognition and counting of fish body lateral line scales.

At present, the proposed model has integrated into the smart fishery breeding platform, enabling efficient automatic recognition and counting of fish body lateral line scales in local images, local videos, and real-time videos, basically fulfilling the actual needs of fishery research and aquaculture. When compared with previous research, the use of this system avoids touching the fish, significantly minimizing damage to the fish. Furthermore, in terms of the identification accuracy and counting precision of fish body lateral line scales, this system performs even higher and more stable.

Currently, the proposed detection method mainly solves the problem of fish lateral line scale detection in static and dynamic scenes, including complex backgrounds and dynamic fish bodies. However, there are still issues with uneven underwater illumination in many aquaculture conditions. In the future, we will continue to study on the accurate extraction of fish body features such as lateral line scales under different lighting conditions, thereby further improving the generalizability and robustness of our model algorithms. Additionally, the quick movement of the fish can cause motion blur in the image, making it more difficult to detect feature edges, hence the fast speed of fish is yet another challenge restricting accurate extraction of detailed phenotypic features of the fish body and is needed to continue studying.

REFERENCES

[1] E. D. DeLamater and W. R. Courtenay, "Variations in structure of the lateral-line canal on scales of teleostean fishes," *Zeitschrift Für Morphologie der Tiere*, vol. 75, no. 4, pp. 259–266, 1973.
[2] J. Mogdans and H. Bleckmann, "Coping with flow: Behavior, neurophysiology and modeling of the fish lateral line system," *Biol. Cybern.*, vol. 106, nos. 11–12, pp. 627–642, Dec. 2012.
[3] J. F. Webb and J. B. Ramsay, "New interpretation of the 3-D configuration of lateral line scales and the lateral line canal contained within them," *Copeia*, vol. 105, no. 2, pp. 339–347, Jul. 2017.
[4] N. M. Roberts, C. F. Rabeni, and J. S. Stanovick, "Distinguishing centrarchid genera by use of lateral line scales," *North Amer. J. Fisheries Manage.*, vol. 27, no. 1, pp. 215–219, Feb. 2007.
[5] P. Risholm, A. Mohammed, T. Kirkhus, S. Clausen, L. Vasilyev, O. Folkedal, Ø. Johnsen, K. H. Haugholt, and J. Thielemann, "Automatic length estimation of free-swimming fish using an underwater 3D range-gated camera," *Aquacultural Eng.*, vol. 97, May 2022, Art. no. 102227.
[6] L. Yang, Y. Liu, H. Yu, X. Fang, L. Song, D. Li, and Y. Chen, "Computer vision models in intelligent aquaculture with emphasis on fish detection and behavior analysis: A review," *Arch. Comput. Methods Eng.*, vol. 28, no. 4, pp. 2785–2816, Jun. 2021.
[7] X. Yu, Y. Wang, J. Liu, J. Wang, D. An, and Y. Wei, "Non-contact weight estimation system for fish based on instance segmentation," *Expert Syst. Appl.*, vol. 210, Dec. 2022, Art. no. 118403.
[8] D. Li and L. Du, "Recent advances of deep learning algorithms for aquacultural machine vision systems with emphasis on fish," *Artif. Intell. Rev.*, vol. 55, no. 5, pp. 4077–4116, Jun. 2022.
[9] H. Wang, S. Zhang, S. Zhao, J. Lu, Y. Wang, D. Li, and R. Zhao, "Fast detection of cannibalism behavior of juvenile fish based on deep learning," *Comput. Electron. Agricult.*, vol. 198, Jul. 2022, Art. no. 107033.
[10] C. Shi, Q. Wang, X. He, X. Zhang, and D. Li, "An automatic method of fish length estimation using underwater stereo system based on LabVIEW," *Comput. Electron. Agricult.*, vol. 173, Jun. 2020, Art. no. 105419.
[11] J. H. Christensen, L. V. Mogensen, R. Galeazzi, and J. C. Andersen, "Detection, localization and classification of fish and fish species in poor conditions using convolutional neural networks," in *Proc. IEEE/OES Auto. Underwater Vehicle Workshop (AUV)*, Nov. 2018, pp. 1–6.
[12] C. Schellewald, A. Stahl, and E. Kelasidi, "Vision-based pose estimation for autonomous operations in aquacultural fish farms," *IFAC-PapersOnLine*, vol. 54, no. 16, pp. 438–443, 2021.
[13] D. J. White, C. Svellingen, and N. J. C. Strachan, "Automated measurement of species and length of fish by computer vision," *Fisheries Res.*, vol. 80, nos. 2–3, pp. 203–210, Sep. 2006.
[14] C. Costa, A. Loy, S. Cataudella, D. Davis, and M. Scardi, "Extracting fish size using dual underwater cameras," *Aquacultural Eng.*, vol. 35, no. 3, pp. 218–227, Oct. 2006.
[15] G. Li, X. Liu, Y. Ma, B. Wang, L. Zheng, and M. Wang, "Body size measurement and live body weight estimation for pigs based on back surface point clouds," *Biosyst. Eng.*, vol. 218, pp. 10–22, Jun. 2022.
[16] N. S. Abinaya, D. Susan, and R. K. Sidharthan, "Deep learning-based segmental analysis of fish for biomass estimation in an occulted environment," *Comput. Electron. Agricult.*, vol. 197, Jun. 2022, Art. no. 106985.
[17] R. Maurya, A. Srivastava, A. Srivastava, V. K. Pathak, and M. K. Dutta, "Computer aided detection of mercury heavy metal intoxicated fish: An application of machine vision and artificial intelligence technique," *Multimedia Tools Appl.*, vol. 82, no. 13, pp. 20517–20536, May 2023.

[18] A. Banwari, R. C. Joshi, N. Sengar, and M. K. Dutta, "Computer vision technique for freshness estimation from segmented eye of fish image," *Ecolog. Informat.*, vol. 69, Jul. 2022, Art. no. 101602.
[19] Y. Liao, "3DPhenoFish: Application for two- and three-dimensional fish morphological phenotype extraction from point cloud analysis," *Zool. Res.*, vol. 42, no. 4, pp. 492–501, 2021.
[20] J. Jurado-Molina, C. H. Hernández-López, and C. Hernández, "Evaluation of fish density influence on the growth of the spotted Rose snapper reared in floating net cages using growth models and non-parametric tests," *Ciencias Marinas*, vol. 49, pp. 1–15, Feb. 2023.
[21] H. Bleckmann and R. Zelick, "Lateral line system of fish," *Integrative Zool.*, vol. 4, no. 1, pp. 13–25, Mar. 2009.
[22] E. P. Voronina and D. R. Hughes, "Types and development pathways of lateral line scales in some teleost species," *Acta Zoolog.*, vol. 94, no. 2, pp. 154–166, Apr. 2013.
[23] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *J. Big Data*, vol. 6, no. 1, pp. 1–48, Dec. 2019.
[24] L. Huang, C. Chen, J. Yun, Y. Sun, J. Tian, Z. Hao, H. Yu, and H. Ma, "Multi-scale feature fusion convolutional neural network for indoor small target detection," *Frontiers Neurorobotics*, vol. 16, May 2022, Art. no. 881021.
[25] Y. Chen, H. Liu, L. Yang, H. Yu, D. Li, S. Mei, and Y. Liu, "A lightweight detection method for the spatial distribution of underwater fish school quantification in intensive aquaculture," *Aquaculture Int.*, vol. 31, no. 1, pp. 31–52, Feb. 2023.
[26] K. Wang, J. H. Liew, Y. Zou, D. Zhou, and J. Feng, "PANet: Few-shot image semantic segmentation with prototype alignment," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 9196–9205.
[27] X. Zhai, H. Wei, Y. He, Y. Shang, and C. Liu, "Underwater sea cucumber identification based on improved YOLOv5," *Appl. Sci.*, vol. 12, no. 18, p. 9105, Sep. 2022.
[28] K. Han, Y. Wang, H. Chen, X. Chen, J. Guo, Z. Liu, Y. Tang, A. Xiao, C. Xu, Y. Xu, Z. Yang, Y. Zhang, and D. Tao, "A survey on vision transformer," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 1, pp. 87–110, Jan. 2023.
[29] Z. Dai, B. Cai, Y. Lin, and J. Chen, "UP-DETR: Unsupervised pre-training for object detection with transformers," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 1601–1610.
[30] X. Zhu, S. Lyu, X. Wang, and Q. Zhao, "TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops (ICCVW)*, Oct. 2021, pp. 2778–2788.
[31] H. Wang, Y. Jin, H. Ke, and X. Zhang, "DDH-YOLOv5: Improved YOLOv5 based on double IoU-aware decoupled head for object detection," *J. Real-Time Image Process.*, vol. 19, no. 6, pp. 1023–1033, Dec. 2022.

HUIHUI YU received the Ph.D. degree in agricultural information technology research from China Agricultural University, Beijing, China, in 2018. She is a Lecturer of computer science and technology with the School of Information Science and Technology, Beijing Forestry University, Beijing. She has worked in the research domains of agricultural information acquisition and processing. She has published six technical articles in the information processing.

ZIMAO WANG received the master's degree from the College of Information and Electrical Engineering, China Agricultural University. His research direction is feature extraction of swimming fish in aquaculture. The fish lateral scales detection is his main study content.

HANXIANG QIN is currently pursuing the Ph.D. degree with the College of Information and Electrical Engineering, China Agricultural University. He is studying on identification and analysis method of shrimp feeding state based on deep learning and has published related articles. His main research interests include computer vision and deep learning in aquaculture application.

YINGYI CHEN received the Ph.D. degree in agricultural information technology research from China Agricultural University, Beijing, China, in 2008. He is a Professor of computer science and technology and the Deputy Director of the Department of Computer Engineering, College of Information and Electrical Engineering, China Agricultural University. He has worked in the research domains of agriculture information process, e.g., water quality prediction model of aquaculture, fish detection model, and fish behavior analysis model. He has published over 50 technical articles.