SIN-CHN SCIENTIFIC PRESS PTE.LTD.

Molecular & Cellular Biomechanics 2025, 22(3), 1545.
https://doi.org/10.62617/mcb1545

Article

Fish fry body length measurement with improved YOLOv8n-pose and biomechanics

Zhiyan Ma[1,2,*], Xiaofei Li[1], Jiajun Wu[1]

[1] Institute of Agricultural Machinery Engineering Research and Design, Hubei University of Technology, Wuhan 430068, China
[2] Hubei Engineering Technology Research Center for the Intelligentization of Agricultural Machinery Equipment, Wuhan 430068, China
*** Corresponding author:** Zhiyan Ma, 20071017@hbut.edu.cn

CITATION

Ma Z, Li X, Wu J. Fish fry body length measurement with improved YOLOv8n-pose and biomechanics. Molecular & Cellular Biomechanics. 2025; 22(3): 1545. https://doi.org/10.62617/mcb1545

ARTICLE INFO

Received: 12 February 2025
Accepted: 26 February 2025
Available online: 28 February 2025

Abstract: Accurate measurement of fish fry body length is crucial in biomechanical research and the development of intelligent aquaculture, as it directly affects the growth, locomotion, and ecological adaptability of fish. Traditional manual methods are time-consuming, labor-intensive, and may harm fish fry. Therefore, accurate, rapid, and non-destructive measurements of large quantities of fish fry are highly important in aquaculture. This study used 20–100 mm grass carp fry (Ctenopharyngodon idella) as test subjects. An image acquisition platform was developed to obtain RGB-D data from the top view of the fry. We proposed ROS-YOLO, which replaces the original C2f module of YOLOv8n-Pose with reparameterized convolution-based shuffle one-shot aggregation (RCS-OSA) and introduces a simple attention module (SimAM) into the main feature extraction layer, to detect key body length points of fish fry. Depth information for 3D keypoint coordinate transformation was obtained through the depth map. Additionally, biomechanical principles were incorporated to study the movement patterns, muscle activity, and hydrodynamic efficiency of fish fry. High-speed cameras and motion tracking software were used to analyze swimming kinematics and dynamics, while biomechanical modeling was employed to simulate the effects of water flow on growth and development. Finally, fish fry body lengths were calculated based on keypoint coordinates. In experiments, ROS-YOLO achieved an average keypoint detection accuracy of 99.2%, with 3.97 M parameters and 125 FPS. Compared to manual measurements, the overall average error in automatic measurement results was 2.87 mm (5.85%). Therefore, the proposed method meets real-time measurement requirements for fish fry body length and provides insights into the biomechanics of fish fry growth and movement.

Keywords: grass carp fry; fry body length; attention mechanism; three-dimensional coordinates; YOLOv8n-pose; keypoint detection

1. Introduction

In aquaculture, body length of fish is important. The size of the fish not only reflects growth status but also serves as a crucial basis for feeding, grading cultivation, harvesting, selling, and estimation of the fish biomass. Body length plays a vital role in evaluating production efficiency and intelligent management during fry cultivation. In the current aquaculture industry, fry size measurements rely on manual sampling. This traditional contact-based measurement method is inefficient and can cause varying degrees of physical damage to the fry, affecting their normal growth and resulting in economic losses for fish farms [1,2]. The development of machine vision technology, has led to it being widely applied in various fields of aquaculture (such as quality grading [3], identification counting [4,5], behavior analysis [6–8], and health assessment [9,10]), making it indispensable in aquaculture.

However, obtaining size information quickly and accurately on a large number of fry, remains a problem to be solved.

Researchers worldwide have extensively explored vision-based fish size measurement methods. Yang et al. [11] applied thresholding to segment fish body images and employed the Canny algorithm to extract body contours, achieving an average relative error of 0.3% in determining measurement points. Tseng et al. [12] trained a convolutional neural network to detect fish heads and caudal forks, defining the distance between these points as body length. A pixel-to-real distance conversion factor derived from a calibration board was used to estimate fish length, resulting in an average relative error of 4.26%. Zhou et al. [13] used the SOLOv2 model to segment fish bodies and generated depth images via binocular stereovision. By combining image plane features with depth data, they reconstructed 3D fish poses and precisely estimated total length with a 2.67% average relative error. These studies demonstrate the feasibility of vision-based fish size measurement; however, strict experimental constraints on fry conditions limit its applicability in real-world aquaculture environments.

In recent years, advancements in human pose estimation have led to the widespread adoption of convolutional neural network-based keypoint detection methods for dimension measurement, yielding promising outcomes. Li and Teng [14] pioneered the use of deep learning techniques to localize livestock feature points by employing stacked hourglass networks, achieving keypoint localization in segmented images of goat and cow trunks. Wang et al. [15] developed an improved keypoint detection model, HRNet with Swin Transformer block (HRST), to detect keypoints on standing pigs, enabling non-contact measurement of pig body dimensions. Li et al. [16] proposed DSS-YOLO, a keypoint detection model based on YOLOv8n-pose, and integrated it with point cloud data processing methods to detect measurement points in 3D point clouds of Mongolian horses. This approach enabled automatic measurement of five body size parameters, including body height, body length, hip height, chest girth, and hip girth. Li et al. [17] developed a method to locate and measure continuously casting billets by integrating a Transformer with binocular vision. They employed an improved neural network to detect and extract keypoint coordinates, achieving measurement via 3D reconstruction using binocular vision.

In previous studies, the measurement environments were relatively controlled, and the targets were clearly distinguishable. However, this approach is unsuitable for keypoint detection in scenarios involving a large number of small fish fry with mutual occlusion. To address these challenges, this study proposes a new detection model, RCS-OSA-SimAM-YOLO, based on the YOLOv8-pose framework. The model is designed for keypoint detection to enable accurate fish fry size measurement.

2. Materials and methods

2.1. Dataset construction

2.1.1. Data collection

Sample data were collected from the Mechanical Building at Hubei University of Technology using the data acquisition setup depicted in **Figure 1**. A RealSense D435 depth camera was used, with an RGB image resolution of 1280 × 720 and a depth image resolution of 640 × 480 pixels, at a frame rate of 30 FPS. The experimental subjects were grass carp fry randomly sampled from a breeding pond, with body lengths ranging from 20 to 100 mm. Top-view RGB videos of fry moving through a chute in various rearing containers were recorded, each lasting 15 to 30 seconds, and saved in *.mp4 format. To enhance dataset diversity, videos were captured under various natural conditions, including sunny and cloudy days, as well as frontlight and backlight conditions. After each session, the fry were replaced, and the vertical distance between the camera and water surface was adjusted. A total of 40 RGB videos were collected under these conditions.

(a) **(b)**

Figure 1. Collection environment and equipment. **(a)** scene 1; **(b)** scene 2.

The collected videos were processed to extract every 10th frame, yielding 3200 fish fry images. To prevent overfitting caused by high similarity among the images, redundant images were removed by calculating the structural similarity index (SSIM) between adjacent frames [18]. The SSIM threshold was defined as the average SSIM for each video segment. If the SSIM between two adjacent frames exceeded the threshold, only one frame was retained. The SSIM is calculated as follows:

$$\begin{cases} l(x,y) = \dfrac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \\ c(x,y) = \dfrac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \\ s(x,y) = \dfrac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \end{cases} \tag{1}$$

$$S_{SSIM}(x,y) = [l(x,y)]^{\alpha} \times [c(x,y)]^{\beta} \times [s(x,y)]^{\gamma} \tag{2}$$

where x and y represent the data from the first and second image windows respectively; $l(x,y)$, $c(x,y)$, and $s(x,y)$ are the formulas for calculating luminance, contrast, and structural similarity respectively; μ_x and μ_y are the average grayscale values of the two images; σ_x and σ_y are the grayscale standard deviations of the two images; C_1, C_2, and C_3 are constants; α, β, and γ are the weights of the different features in the SSIM calculation and were all set to 1 in this experiment.

After initial screening of similar images using SSIM, images with significant mutual occlusion among fry were manually removed to ensure that during annotation, keypoints of one fish did not overlap with the bounding box of another. Following this process, 1200 top-down fry images were obtained and split into training, validation, and test sets in a 7:2:1 ratio. The dataset was annotated using Labelme, based on key measurement points of grass carp fry, with corresponding label names listed in **Table 1**. **Figure 2** illustrates the Labelme annotation interface.

Table 1. Data annotation labels.

Label target	Fish fry	Fish head	Dorsal fin origin	Caudal fin base	Left gill	Right gill
Label	fish	0	1	2	3	4

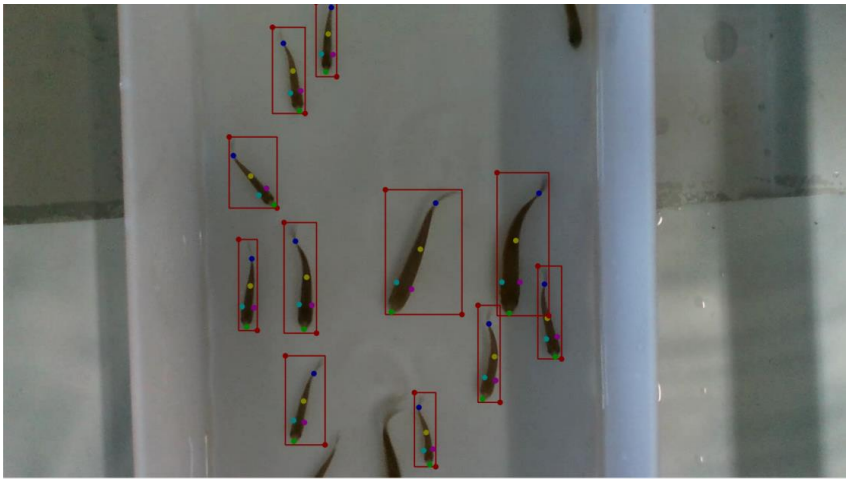


Figure 2. Schematic diagram of key feature points annotation on grass carp fry image.

Upon completion of the annotation process, a JSON file is generated and subsequently converted into a TXT file, as shown in **Figure 3**. This file contains the label information for the image, including the label category, bounding box coordinates, and keypoint coordinates. This annotation process provides precise localization of fry keypoints and classification labels, enabling high-quality data for training keypoint detection algorithms.

```
0 0.29531 0.35694 0.05625 0.15000 0.32031 0.42500 2 0.29297 0.36528 2 0.27187 0.32222 2 0.31094 0.39028 2 0.30234 0.40556 2
0 0.33750 0.14306 0.03828 0.18056 0.35000 0.22778 2 0.34219 0.14444 2 0.33125 0.08611 2 0.35156 0.18611 2 0.33984 0.19028 2
0 0.35703 0.83611 0.04688 0.18750 0.34375 0.92222 2 0.35547 0.83750 2 0.36719 0.78056 2 0.35625 0.88333 2 0.34375 0.87917 2
0 0.49922 0.89861 0.02500 0.15556 0.50469 0.97222 2 0.50000 0.90417 2 0.49141 0.85139 2 0.50859 0.93889 2 0.49766 0.94167 2
0 0.64609 0.65139 0.02813 0.19583 0.65312 0.74444 2 0.64375 0.66389 2 0.63984 0.59167 2 0.65547 0.70278 2 0.64219 0.70694 2
0 0.35078 0.57778 0.03828 0.23194 0.35547 0.68750 2 0.35859 0.57778 2 0.34531 0.50139 2 0.36484 0.63611 2 0.35000 0.63472 2
0 0.49688 0.52361 0.08984 0.26250 0.45937 0.65139 2 0.49297 0.51806 2 0.51172 0.42222 2 0.48281 0.60278 2 0.46484 0.58889 2
0 0.28906 0.59444 0.02266 0.19167 0.28828 0.68194 2 0.29219 0.59583 2 0.29375 0.53889 2 0.29688 0.64167 2 0.28203 0.63889 2
0 0.38203 0.07778 0.02422 0.15556 0.38125 0.14583 2 0.38438 0.06667 2 0.38828 0.01111 2 0.38984 0.10972 2 0.37578 0.10833 2
0 0.57422 0.73750 0.02578 0.20417 0.56719 0.83333 2 0.58125 0.74444 2 0.57422 0.67500 2 0.58281 0.79444 2 0.56953 0.78889 2
0 0.61484 0.50833 0.06094 0.30000 0.59844 0.65278 2 0.60625 0.50000 2 0.63359 0.40139 2 0.61094 0.58750 2 0.58984 0.58750 2
```

Figure 3. The content of the TXT label.

2.1.2. Data augmentation

To improve the generalizability and detection performance of the model for fish fry keypoints, this study employed various online data augmentation techniques, including random noise, random flipping, random rotation, random brightness

adjustment, and random combinations of the above methods. For example, random noise includes adding Gaussian or salt-and-pepper noise to the original image; random flipping involves flipping the original image horizontally or vertically at random; random rotation involves rotating the original image by 90° or 180° at random; random brightness adjustment involves varying the brightness of the original image; and random combination combines the aforementioned methods. **Figure 4** illustrates the effects of each image enhancement method. Each image was randomly augmented using one method, yielding 2400 fish fry images, which were then split into training, testing, and validation sets in a 7:2:1 ratio.

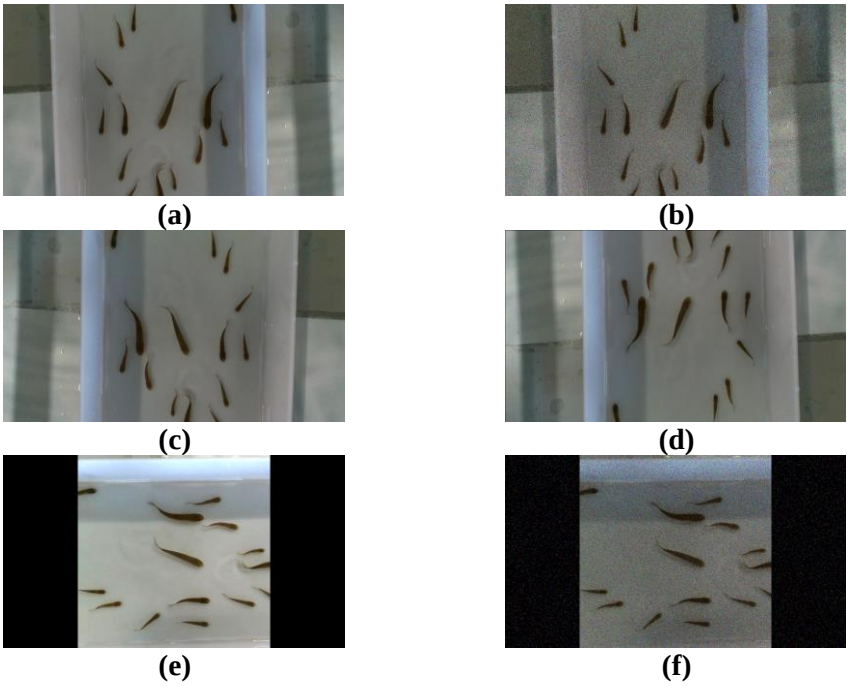


Figure 4. Schematic diagram of various image enhancement effects. **(a)** original image; **(b)** random noise; **(c)** random flip; **(d)** random angle rotation; **(e)** random brightness adjustment; **(f)** random combination.

2.2. Overall counting roadmap

The overall technical route, shown in **Figure 5**, comprises a keypoint detection model and a 3D coordinate transformation module. First, an image acquisition platform was developed to acquire RGB and depth images of the fish fry. Next, ROS-YOLO was used to detect keypoints for fish fry size measurement and project them onto synchronized depth images to obtain depth information. Finally, camera calibration was performed to obtain intrinsic and extrinsic parameters, enabling the transformation of 3D keypoint coordinates and calculation of fish fry size parameters.

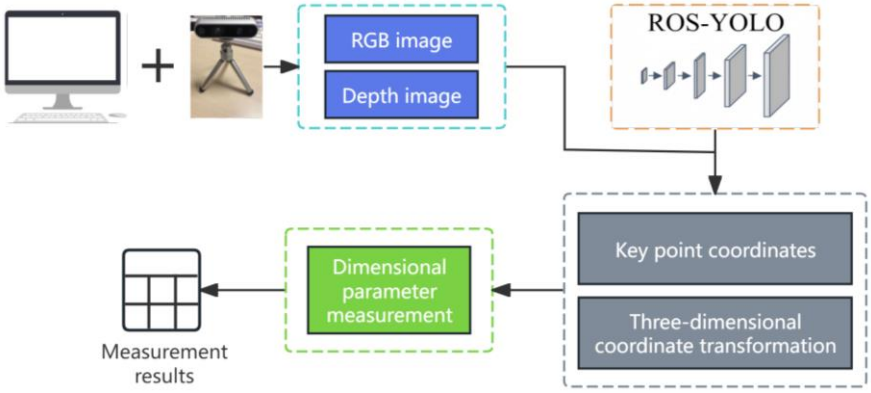


Figure 5. Technology roadmap for automatic measurement of fry size parameters.

2.3. Improvement of key point detection model

YOLOv8-Pose is a single-stage detection model within YOLOv8 that simultaneously achieves object and human pose keypoint detection. The corresponding network model is divided into three parts: backbone, neck, and detection head. The YOLOv8 algorithm has five versions: YOLOv8n, s, m, l, and x. For this study, which focused on keypoint detection and size measurement of fish fry in aquaculture, YOLOv8n-pose was selected as the base model to ensure accurate and fast detection. Because the initial YOLOv8n-pose was primarily designed for human pose estimation and thus not suitable for the identification and keypoint detection of grass carp fry, we improved YOLOv8n-pose, considering the characteristics of the self-made grass carp fry keypoint dataset, in the following two aspects to meet the requirements for keypoint detection for grass carp fry size measurements:

1) In the Backbone and Neck of the YOLOv8-pose model, the original coarse-to-fine (C2f) object detection module was replaced with the RCS-OSA [19] module, combining feature cascading with computational efficiency, to enhance the detailed extraction capability for grass carp fry keypoints and significantly reduce inference time.

2) The SimAM [20] attention mechanism was introduced into the main feature extraction layer, to generate attention weights by calculating the similarity between each pixel and its neighboring pixels in the feature map, without any additional parameters, which effectively improved grass carp fry detection accuracy and efficiency.

The improved model structure is illustrated in **Figure 6** (the improved parts are highlighted in red boxes). The optimized YOLOv8n-pose model demonstrated higher detection accuracy and efficiency in complex scenarios, was capable of precisely extracting the keypoints of fry, and could stably detect even under occluded conditions, providing reliable technical support for large-scale target detection in aquaculture environments.

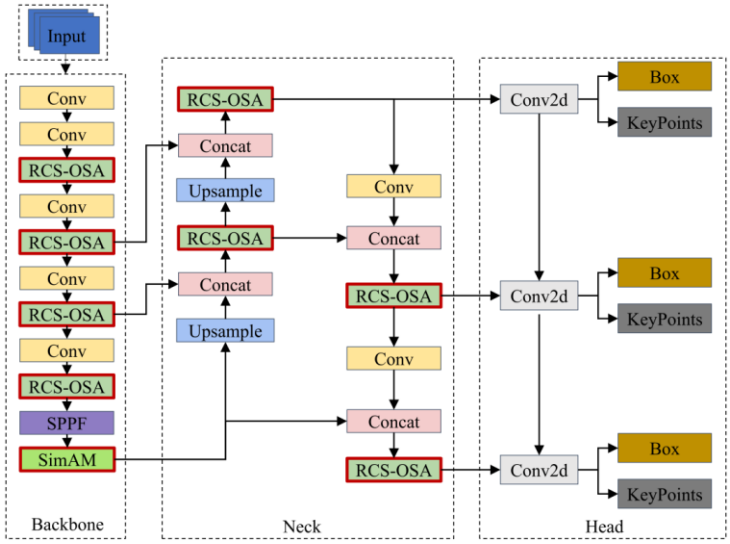


Figure 6. Improved YOLOv8-pose model.

2.3.1. RCS-OSA module

The RCS-OSA module consists of RCS and OSA. The RCS is a structural reparametric convolution based on channel shuffle (**Figure 7**), whereby model performance during training and inference is optimized through channel shuffling and multi-branching structures. In the training phase, the input feature tensor is divided into two parts of the same dimension, with one part processed through multibranch structures, such as identity mapping, 11 convolutions, and 33 convolutions, to learn rich feature representations, followed by channel connection and shuffling; the other part ensures that the features are fully integrated among different channels. In the inference phase, all branches are parameterized into a layer of 33 convolutions, which significantly reduces computational complexity and memory consumption, enabling fast and efficient inference while maintaining information exchange.

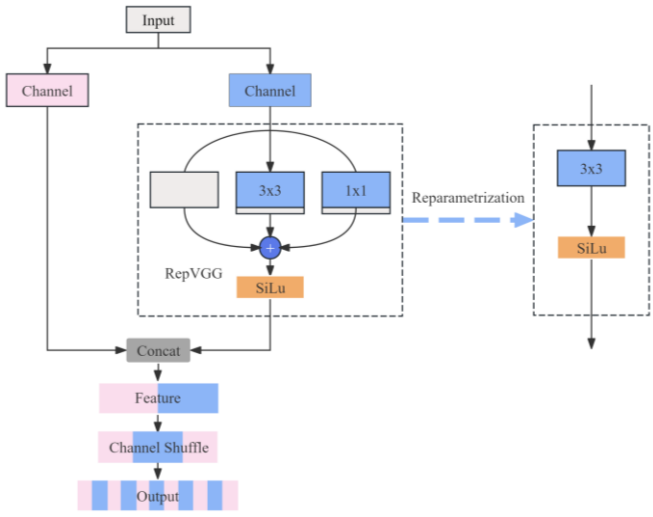


Figure 7. RCS structure diagram.

Note: RepVGG is the structural reparameterization module used during the training phase, whereas RepConv is used during model inference, and SiLu is the activation function.

The OSA module overcomes the inefficiency of dense connections in DenseNet by representing features with multiple receptive fields, aggregating them only once at the last stage, thereby enhancing speed and energy efficiency. The RCS-OSA module (**Figure 8**) integrates the RCS structure, which enhances feature reuse and cross-layer information flow by stacking RCS modules, thereby reducing computational load and memory usage. This helps the model perform high-precision, rapid inference of keypoints of fish fry sizes in different postures in real-world aquaculture environments.

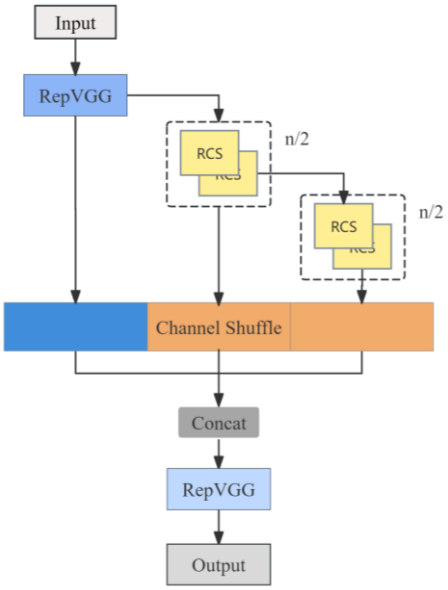


Figure 8. RCS-OSA structure diagram.
Note: n represents the number of stacked RCS modules.

2.3.2. SimAM attention mechanism

The attention mechanism can help models focus on key areas of the images, thereby enhancing model performance and achieving more effective and accurate predictions. However, most attention mechanisms typically require additional parameters, which increases model complexity and computational costs. In this study, a lightweight, parameter-free SimAM convolutional neural network attention mechanism was introduced (**Figure 9**). Unlike existing channel and spatial attention modules, the SimAM module computes 3D attention weights for feature maps by analyzing their local self-similarity, achieving this without adding any learnable parameters. This allows the model to focus on keypoint feature information acquisition without increasing the number of parameters, thereby significantly improving detection efficiency.

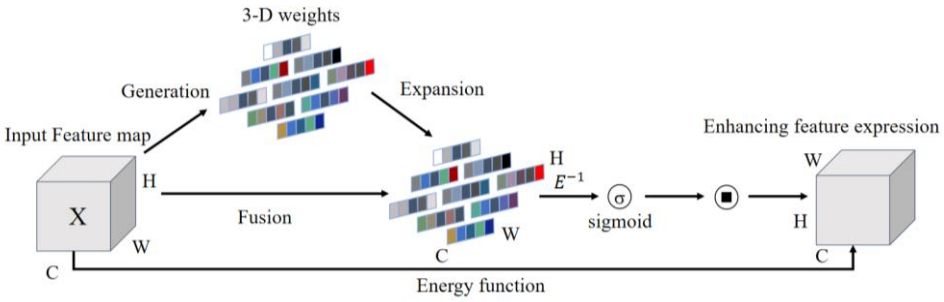


Figure 9. Structure diagram of SimAm attention mechanism.

Note: C, H, and W represent the number of channels, height, and width, respectively; sigmoid denotes the activation function.

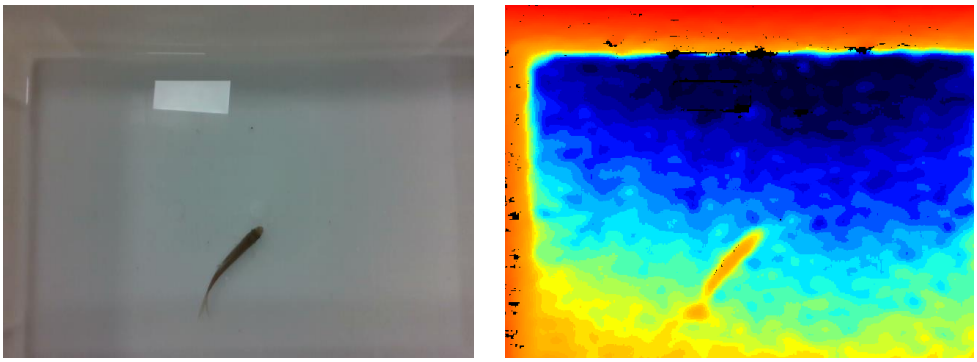
2.4. Three-dimensional coordinate transformation

The keypoint coordinates of fry detected by ROS-YOLO are 2D coordinates, which need to be converted into 3D coordinates in the camera coordinate system to achieve automatic measurement of fry size. Through the camera align library function, the captured RGB image is aligned and calibrated with the depth map (as shown in **Figure 10**), and the extracted key points are mapped to the depth map to obtain the depth values at these pixel key points. Subsequently, the optical center coordinates and focal length of the camera's configured stream are acquired using the Intel RealSense camera SDK. Based on Equation (3), the transformation from 2D coordinates to 3D coordinates of key points is realized. The body length of the fry can then be represented as the sum of two distances: the distance between the head point and the dorsal fin starting point L_{ab}, and the distance between the dorsal fin starting point and the caudal fin base point L_{bc}. Equation (4) shows the Euclidean distance calculation formula between two coordinate points.

$$\begin{cases} Z = depth \\ X = \dfrac{(u - c_x)Z}{f_x} \\ Y = \dfrac{(v - c_y)Z}{f_y} \end{cases} \tag{3}$$

where (X, Y) represents the three-dimensional coordinates in the camera coordinate system; (u, v) represents the pixel coordinates; "depth" represents the depth value of the pixel point; (c_x, c_y) represents the coordinates of the camera optical center in the pixel coordinate system; f_x and f_y represent the focal lengths of the camera.

$$L_{12} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2} \tag{4}$$



(a) **(b)**

Figure 10. The aligned color image and depth image. **(a)** color image; **(b)** depth image.

3. Test experiments and results analysis

3.1. Experimental environment

This research model was trained under the Windows 10 operating system with an Intel Core i5-13400F CPU processor, a GPU with 8 GB of video memory RTX4060Ti, using CUDA 12.1 for acceleration, the deep learning framework of PyTorch, and Python version 3.9. The experimental hyperparameters are listed in **Table 2**.

Table 2. Experimental hyperparameters.

Parameter	Value
Image size	640 × 640
Epoch	100
Batch size	32
Learn rate	0.01
Momentum	0.937

3.2. Evaluation index setting

To evaluate the performance of the fish–fry keypoint detection model, both prediction boxes and predicted keypoints were assessed. Average precision (AP) was calculated using object keypoint similarity (OKS), and mean average precision-keypoint (mAP-kp) was derived from AP as the keypoint evaluation index. Precision and recall were used as target identification evaluation indices. FPS was adopted as the evaluation index for the inference speed of the model, and the number of parameters was used to evaluate model size. The mAP was calculated, as follows:

$$mAP = \frac{\sum P\delta\left(OKS_p > T\right)}{\sum P\,1} \tag{5}$$

To evaluate the results, MAE and MRE were used as accuracy evaluation indices for fry size measurement and calculated as in Equations (6) and (7), respectively:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{x}_i - x_i| \tag{6}$$

$$MRE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{\hat{x}_i - x_i}{x_i}\right| \tag{7}$$

where i is the fry label; n is the total number of test fry; $\hat{x}_i$ and x_i are divided into automatic and manual measurements of i.

3.3. Test results and analysis

3.3.1. Comparison of test results of different models

To validate the performance of the ROS-YOLO model, comparative experiments were conducted with the YOLOv7-tiny-face [21], YOLOv8n-Pose, and ROS-YOLO models using a self-built fry keypoint test dataset. The experimental results are shown in **Table 3**. From the comparison in **Table 3**, it is evident that the ROS-YOLO model outperforms other network models across all evaluation metrics. In terms of measurement accuracy, ROS-YOLO achieved a mAP of 99.2%, compared to 91.8% for YOLOv7-tiny-face and 97.2% for YOLOv8n-Pose. This improvement is attributed to the added RCS-OSA module, which enhances localization capabilities in ambiguous regions through multi-scale feature fusion, and the SimAM attention mechanism, which strengthens the model's focus on fry keypoints, thereby significantly improving keypoint localization precision. Regarding detection speed, the improved ROS-YOLO model achieves 125 FPS, surpassing both YOLOv8n-Pose and YOLOv7-tiny-face. This is due to the RCS-OSA module's structural re-parameterization during inference, which simplifies the architecture into a single 3 × 3 convolutional layer, substantially reducing computational complexity and streamlining the inference process. Consequently, under these enhancements, ROS-YOLO delivers superior training results compared to the original model and the YOLOv7-tiny-face model.

Table 3. Comparison tests for different models.

Model	P/%	R/%	mAP%	Params	FPS
YOLOv7-tiny-face	86.5	88.8	91.8	7.84	92.5
YOLOv8n-Pose	96.3	93.3	97.2	3.08	115
ROS-YOLO	97.2	99.8	99.2	3.97	125

Figure 11 shows a comparison of the fry recognition and keypoint prediction results from the aforementioned models. Overall, none of the models missed any fry in their detection results. However, YOLOv7-tiny-face performed poorly in keypoint prediction, with significant deviations observed in some keypoints. YOLOv8n-Pose achieved good results in predicting the dorsal fin and gill points but showed larger errors in predicting some head and tail points. In contrast, the improved ROS-YOLO model demonstrated significantly better performance in predicting all keypoints compared to the other two models, particularly in areas with low visibility such as

the head and tail edges. These results indicate that the proposed improvements in ROS-YOLO lead to superior performance in fry keypoint prediction.

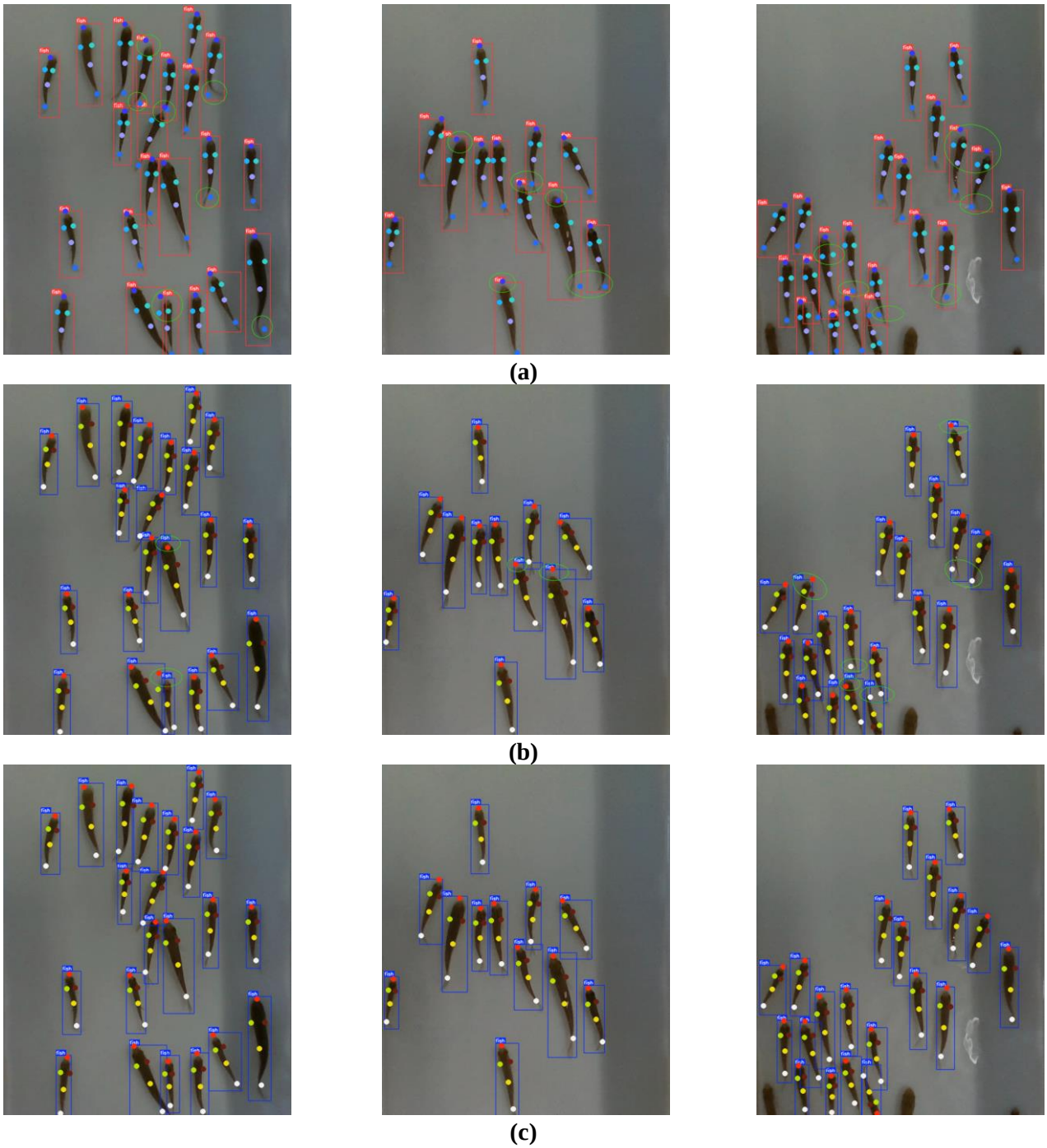


Figure 11. Comparison of identification and keypoint prediction between different network models. **(a)** YOLOv7-lite-s; **(b)** YOLOv8n-pose; **(c)** ROS-YOLO.

3.3.2. Ablation test

To verify that the proposed method has certain advantages in detecting the keypoints of fish fry size, ablation experiments were conducted on the improved parts. All experiments were conducted under the same environmental conditions and parameters. Considering the real-time detection and deployment of key points in fish

fry, the evaluation indicators focused not only on average accuracy but also on changes in the number of model parameters and processing speed. The experimental results are presented in **Table 4**.

Table 4. Ablation test results.

Model	Precision/%	Recall/%	Average precision/%	Parameters/M	FPS/s^{-1}
1	96.3	93.3	97.2	3.08	115
2	97.5	98.7	99.1	3.97	133
3	97.1	98.8	98.0	3.15	130
4	96.7	98.7	98.6	3.42	129
5	95.1	99	97.8	3.09	128
6	96.9	99.3	98.8	3.08	137
7	97.2	99.8	99.2	3.97	125

Note: Model 1 represents the original YOLOv8n-Pose. Model 2 represents the introduction of the RCS-OSA module. Model 3 represents the introduction of the CBAM attention mechanism. Model 4 represents the introduction of the CAFM attention mechanism. Model 5 represents the introduction of the CA attention mechanism. Model 6 represents the introduction of the SimAm attention mechanism. Model 7 represents the introduction of both the RCS-OSA module and the SimAm attention mechanism.

Analysis of **Table 4** reveals that the improved ROS-YOLO model shows enhancements across all evaluation metrics compared to the original YOLOv8n-Pose model. The original Model 1 achieved a mAP of 97.2%, but it exhibited limitations in detecting occluded or blurred keypoints. Model 2, which introduced the RCS-OSA module, enhanced feature reuse and multi-scale aggregation, allowing for more effective capture of subtle keypoints, increasing mAP by 1.9%, albeit with a 28.9% increase in parameter count. Models 3, 4, 5, and 6 were experimental groups incorporating different attention mechanisms. Model 3, which introduced the CBAM attention mechanism requiring the concatenation of channel and spatial attention modules, showed an improvement in accuracy over Model 6, which introduced SimAM, but had a 0.6 percentage point higher recall rate than Model 6. Model 4, which introduced the CAFM attention mechanism, relied on complex feature fusion strategies, increasing computational load, whereas SimAM achieved efficient feature enhancement through local similarity calculations, outperforming CAFM in FPS and parameter count. Model 5 introduced the CA attention mechanism focusing solely on the channel dimension, neglecting spatial information, while SimAM captured both channel and spatial features through 3D weights, outperforming CA in both mAP and FPS. Model 7, which incorporated both the RCS-OSA module and the SimAM attention mechanism, achieved an accuracy of 97.2% and a recall rate of 99.8% in object detection, representing improvements of 0.9 and 6.5 percentage points, respectively, over the original model. In keypoint detection, the average precision was 99.2%, a 2 percentage point increase over the original model. Although the parameter count of the improved model increased compared to the original, the detection speed improved by 8.7%, better meeting the requirements for real-time keypoint detection.

3.3.3. Fry size measurement results and error analysis

A fry body length measurement experiment was conducted to verify and analyze the accuracy of the automatic fry size measurement method proposed in this study. Each time, a fry was randomly scooped out of the breeding pond. After manually measuring the size (the length from the snout to the base of the caudal fin), the top-view RGB and depth images of the fry passing through were collected, using the device shown in **Figure 1**, and the length was measured as indicated. Each fish was measured 10 times, and a total of 10 fish were measured. **Table 5** shows the mean absolute error and mean relative error of the measurement results for each fish after improvement. Statistically, the MAE of the overall measurement results was 2.87 mm, and MRE was 5.85%. **Figure 12** shows the box plot of the absolute error between the automatic and manual measurements before and after model improvement. In **Figure 12**, the overall absolute error before model improvement is greater than that after improvement, and both the median and mean before improvement are greater than those after improvement. The absolute error after improvement was within the range of −0.1 to −7.3 mm.

Table 5. Average absolute error and average relative error of body length measurements for different numbers of grass carp fry after model improvement.

Number	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
MAE/mm	2.9	1.6	3.1	3.5	2.8	2.9	2.0	2.9	3.2	3.8
MRE/%	5.28	4.79	5.01	5.86	5.67	7.90	7.40	5.14	4.91	6.55

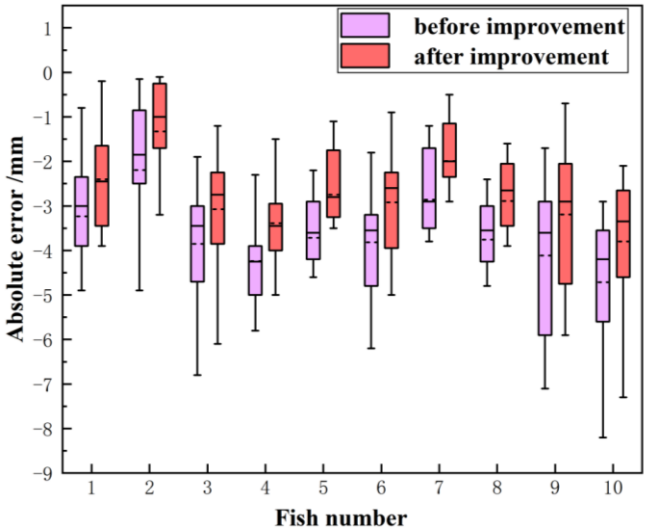


Figure 12. Boxplot of absolute errors in body length measurements of different fry before and after model improvement.

Note: Each boxplot corresponds to the calculation results of the same fish; the central rectangle of the boxplot represents the interquartile range of the measurement results; horizontal and dashed lines within the central rectangle represent the median and mean of the calculation results, respectively; and top and bottom horizontal lines represent the maximum and minimum values of the measurement results.

Compared to the method proposed by Yang et al. [11], which uses thresholding to segment fish images and the Canny algorithm to extract fish contours (MAE = 0.3%), and the SOLOv2 + binocular vision approach proposed by Zhou et al. [13] (MAE = 2.67%), both methods operate in relatively simple environments with highly

visible measurement targets. Tseng et al. [12] proposed a CNN-based method with calibration boards, which relies on manual calibration (MAE = 4.26%) and is unsuitable for detecting size keypoints in large groups of small, overlapping fry. In contrast, the ROS-YOLO method proposed in this study achieves an MAE of 5.85% but enables fully automated detection, with an error increase of only 1.2 mm in occluded scenarios. These results demonstrate that our method offers significant advantages in real-time performance and automation, making it well-suited for large-scale aquaculture applications.

During the experiments, it was observed that the error values measured by the algorithm exhibited significant fluctuations in some cases. This is primarily due to excessive bending of the fish tail fins and mutual occlusion between fish, leading to deviations in keypoint localization. As shown in **Figure 13**, when the fry's tail is excessively bent, representing the body length solely based on the connection between the head point, the starting point of the dorsal fin, and the base point of the tail fin results in substantial errors. In dense fish populations, mutual occlusion between fish occurs frequently. When the head or tail points of a fish are completely occluded by another fish, the incomplete feature information of the occluded fish causes prediction inaccuracies in the algorithm. Such occlusion issues increase the false detection rate from 1.2% in single-fish scenarios to 6.8%, representing a major source of measurement error. Therefore, in future research, we plan to address these anomalies by capturing data under more diverse conditions, increasing the number of measurement keypoints, and introducing a tracking module to correct measured dimensions, thereby improving the accuracy of body length measurements.

(a) **(b)**

Figure 13. Error analysis diagram of fry size measurement. **(a)** Fish body curvature error diagram; **(b)** Fish body occlusion error diagram.

4. Discussion

The integration of Improved YOLOv8n-Pose (ROS-YOLO) and biomechanics-inspired principles in this study has demonstrated significant advancements in the accurate, non-destructive, and real-time measurement of fish fry body length, while also providing novel insights into the biomechanical aspects of fish fry growth and movement. The proposed ROS-YOLO model achieved an impressive keypoint detection accuracy of 99.2%, with a computational efficiency of 125 FPS, making it highly suitable for large-scale aquaculture applications. The incorporation of reparameterized convolution-based shuffle one-shot aggregation (RCS-OSA) and the simple attention module (SimAM) significantly enhanced the model's ability to

detect keypoints under varying conditions, ensuring robust performance even in complex environments. The use of RGB-D data for 3D coordinate transformation further improved measurement precision, with an overall average error of only 2.87 mm (5.85%) compared to manual measurements. These results underscore the potential of computer vision and machine learning in revolutionizing traditional aquaculture practices.

From a biomechanical perspective, this study provides valuable insights into the relationship between fish fry body length and their movement patterns, muscle activity, and hydrodynamic efficiency. The integration of high-speed motion tracking and biomechanical modeling revealed that body length significantly influences swimming kinematics, such as tail beat frequency and body curvature, as well as hydrodynamic forces acting on the fish fry. These findings align with previous studies highlighting the importance of body morphology in determining swimming efficiency and energy expenditure in aquatic organisms. Furthermore, the biomechanical analysis of environmental factors, such as water flow and temperature, demonstrated their impact on fish fry growth and movement. For instance, simulations using computational fluid dynamics (CFD) showed that optimal water flow conditions can enhance muscle development and reduce energy expenditure during swimming, which is critical for improving growth rates and survival in aquaculture settings.

The practical implications of this study are twofold. First, the ROS-YOLO model provides a scalable and efficient solution for real-time body length measurement, reducing the labor and time associated with traditional methods. Second, the biomechanical insights gained from this research can inform the design of aquaculture systems that optimize environmental conditions for fish fry growth and health. For example, adjusting water flow rates or tank designs based on biomechanical principles could minimize stress and energy expenditure, leading to healthier and faster-growing fish populations.

5. Conclusions

To achieve automatic noncontact measurements of body length of farmed fry, this study constructed a fry image acquisition platform and proposed an automatic measurement method for fry body length based on binocular vision and the ROS-YOLO model. The conclusions are as follows:

1) Based on the YOLOv8n-Pose model, improvements were made by replacing the C2f module with RCS-OSA module in both the Backbone and Neck networks, and the SimAm attention mechanism was introduced into the main feature-extraction layer. The improved ROS-YOLO model showed enhancements in various evaluation metrics compared with the original model, with accuracy and mAP reaching 97.2% and 99.2%, respectively, and an FPS of 125 (S-1). Using the detected measurement points, body length measurements were completed based on binocular vision, with an overall MAE of 2.87 mm and MRE of 5.85%. The results indicate that the ROS-YOLO model can satisfy the real-time detection of key points and body length measurement requirements in actual aquaculture environments.

2) In addition, this study has certain limitations. The dataset used consists of grass carp fry from the Yangtze River. Although the model demonstrates good experimental performance, this restricts the generalization ability of the proposed method to other species or regions. The approach may show suboptimal effectiveness for different types of fry under varying conditions. Further validation is required to assess the universality of the ROS-YOLO model for other fish species with distinct body shapes and sizes. Moreover, the dataset was collected under controlled environmental conditions, which cannot fully represent the diversity found in natural aquaculture environments. While biomechanical analysis provides valuable insights, more detailed experimental data on muscle activation and tissue mechanics are needed to refine the model further. Therefore, future research will employ transfer learning training and integrate open-source datasets to expand data resources and verify applicability to other fish species, thereby enhancing model versatility. Concurrently, we will construct a three-dimensional environmental dataset incorporating varying water turbidity levels, multi-angle lighting conditions, and diverse background interference to strengthen the model's discriminative capability against environmental disturbances. By systematically implementing cross-validation and collaborating with aquaculture farms across different regions to collect data on species such as tilapia, salmon, and catfish, we aim to further enhance the model's robustness, comprehensively assess its generalization capability, and improve the accuracy and applicability of the research findings.

Author contributions: Conceptualization, XL and ZM; methodology, XL and JW; software, XL; validation, XL, ZM and JW; formal analysis, XL; investigation, XL; resources, JW; data curation, XL; writing—original draft preparation, XL; writing—review and editing, XL; visualization, ZM; supervision, ZM; project administration, ZM; funding acquisition, ZM. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Hubei Province, China; Research on the Theory of Intelligent Perception and Key Technologies of Autonomous Behavior for Mobile Robots in Uncertain Environments (Grant number 2023AFA037).

Ethical approval: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Research Ethics and Technology Safety Committee of Hubei University of Technology (Approval No. HBUT20250001). The research is in compliance with the Regulations on the Management of Experimental Animals in Hubei Province and the Charter of the Research Ethics and Technology Safety Committee of Hubei University of Technology. All procedures were carried out in accordance with relevant ethical standards.

Conflict of interest: The authors declare no conflict of interest.

References

1. Li Z, Zhao Y, Yang P. Research review on fish body length measurement based on machine vision (Chinese). Transactions of the Chinese Society for Agricultural Machinery. 2021; 52: 207-218.
2. Zhao S, Zhang S, Liu J, et al. Application of machine learning in intelligent fish aquaculture: A review. Aquaculture. 2021; 540: 736724. doi: 10.1016/j.aquaculture.2021.736724

3. Islamadina R, Pramita N, Arnia F, et al. Estimating fish weight based on visual captured. In: Proceedings of the 2018 International Conference on Information and Communications Technology (ICOIACT); 06-07 March 2018; Yogyakarta, Indonesia.
4. Tu X, Qian C, Liu S. Research on identification and counting method of Turbot fry based on ResNet34 model. Fishery Modernization. 2024; 51(1): 90-97.
5. Zhang S, Yang X, Wang Y, et al. Automatic Fish Population Counting by Machine Vision and a Hybrid Deep Neural Network Model. Animals. 2020; 10(2): 364. doi: 10.3390/ani10020364
6. Chen C, Du Y, Zhou C. Study on fish feeding behavior recognition technology based on support vector machine (Chinese). Jiangsu Academy of Agricultural Sciences. 2018; 46: 226-229. doi: 10.15889/j.issn.1002-1302.2018.07.057
7. Liu X, Zhang C. Study on fish tracking based on embedded image processing system (Chinese). Jiangsu Academy of Agricultural Sciences. 2018; 46: 203-207.
8. Zhang C, Chen M. Research status and outlook of fish feeding behavior based on computer vision (Chinese). Jiangsu Academy of Agricultural Sciences. 2020; 48: 31-36.
9. Zhang H, Zhang C, Wang R. Freshness recognition of small yellow croaker based on image processing and improved DenseNet network (Chinese). South China Fisheries Science. 2024; 20: 133-142.
10. Issac A, Dutta MK, Sarkar B. Computer vision based method for quality and freshness check for fish from segmented gills. Computers and Electronics in Agriculture. 2017; 139: 10-21. doi: 10.1016/j.compag.2017.05.006
11. Yang C, Xu J, Lu W. Computer vision-based body size measurement and weight estimation of large yellow croaker. Journal of Chinese Agricultural Mechanization. 2018; 39: 66-70.
12. Tseng CH, Hsieh CL, Kuo YF. Automatic measurement of the body length of harvested fish using convolutional neural networks. Biosystems Engineering. 2020; 189: 36-47. doi: 10.1016/j.biosystemseng.2019.11.002
13. Zhou J, Ji B, Ni W. Noncontact method for the accurate estimation of the full length of Takifugu rubripes based on 3D pose fitting. Transactions of the Chinese Society of Agricultural Engineering. 2023; 39: 154-161.
14. Li K, Teng G. Study on Body Size Measurement Method of Goat and Cattle under Different Background Based on Deep Learning. Electronics. 2022; 11(7): 993. doi: 10.3390/electronics11070993
15. Wang X, Wang W, Lu J, et al. HRST: An Improved HRNet for Detecting Joint Points of Pigs. Sensors. 2022; 22(19): 7215. doi: 10.3390/s22197215
16. Li M, Su L, Zhang Y. Automatic measurement of Mongolian horse body based on improved YOLOv8n-pose and 3D point cloud analysis. Smart Agric. 2024; 6: 91-102.
17. Li T, Xu S, Shi Y. Continuous casting slab model positioning and measurement based on binocular vision and Transformer. Journal of Central South University. 2024; 55: 1312-1322.
18. Huang Z, Xu A, Zhou S. Key point detection method for pig face fusing reparameterization and attention mechanisms. Transactions of the Chinese Society of Agricultural Engineering. 2023; 39: 141-149.
19. Kang M, Ting C, Ting FF. RCS-YOLO: A fast and high-accuracy object detector for brain tumor detection. In: Proceedings of the 26th International Conference; 8–12 October 2023; Vancouver, BC, Canada.
20. Tang Z, Hou X, Huang X, et al. Domain Adaptation for Bearing Fault Diagnosis Based on SimAM and Adaptive Weighting Strategy. Sensors. 2024; 24(13): 4251. doi: 10.3390/s24134251
21. Durve M, Orsini S, Tiribocchi A, et al. Benchmarking YOLOv5 and YOLOv7 models with DeepSORT for droplet tracking applications. The European Physical Journal E. 2023; 46(5). doi: 10.1140/epje/s10189-023-00290-x