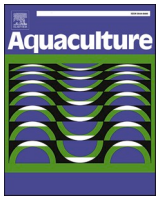
Aquaculture 595 (2025) 741554

Contents lists available at ScienceDirect

Aquaculture

journal homepage: www.elsevier.com/locate/aquaculture



Yolo-AWD+CBT: An efficient algorithm for *Micropterus salmoides* swimming behavior analysis based on multi-object tracking

Peng Xiao [a,b], Ming Chen [a,b], Guofu Feng [a,b,*], Wanying Zhai [c,d], Yidan Zhao [a,b], Yongxiang Huang [c,d]

[a] Key Laboratory of Fisheries Information, Ministry of Agriculture and Rural Affairs, Shanghai Ocean University, Shanghai 201306, PR China
[b] College of Information Technology, Shanghai Ocean University, Shanghai 201306, PR China
[c] Center for Aquacultural Breeding Research, Shanghai Ocean University, Shanghai 201306, PR China
[d] International Research Center for Marine Biosciences, Ministry of Science and Technology, Shanghai Ocean University, Shanghai 201306, PR China

ARTICLE INFO

Keywords:
Multi-object tracking
Swimming behavior
Object detection
Deep learning

ABSTRACT

In aquaculture, analyzing the swimming behavior of micropterus salmoides using multi-object tracking technology is a crucial non-contact method for obtaining data and assessing vitality. However, existing approaches suffer from high false detection and target loss rates due to issues like occlusion between individuals and their variable body shapes. Therefore, this paper proposes a novel multi-object tracking model (Yolo-AWD + CBT) for accurate and real-time tracking of micropterus salmoides swimming behavior. Our method improves upon the Yolov8n backbone network feature extraction module by incorporating an Adaptive Weight Downsampling (AWD) module to address the loss of feature information during downsampling in the original network. To tackle the challenge of variable body shapes during swimming, we replace the original loss function with XIOU, enhancing the network's ability to localize targets. For the tracking algorithm, we introduce trajectory confidence information into ByteTrack, thus improving the tracking accuracy of micropterus salmoides during swimming. Experimental results on object detection and multi-object tracking datasets demonstrate that our proposed model (Yolo-AWD + CBT) achieves a 1.07 % and 5.4 % improvement in P and ($AP_{50:95}$). In terms of tracking performance, compared to the original model, HOTA increases by 6.25 %, MOTA by 3.15 %, and IDsw decreases by 58.33 %, resulting in swimming behavior data errors within the range of −7 % to +3 %. These results indicate that our proposed multi-object tracking method can effectively track multiple targets in various scenarios and accurately capture micropterus salmoides swimming behavior data, providing technical support for non-contact vitality analysis.

1. Introduction

Swimming behavior is fundamental to the basic life activities of micropterus salmoides (Downie et al., 2020) and is a common indicator in welfare-oriented aquaculture, influenced by factors like water environment and temperature. Consequently, fish swimming behavior data can be utilized as a welfare indicator on farms to assess vitality, stress levels, and overall health status (Muñoz et al., 2020). Additionally, individual fish are highly sensitive to changes and disturbances in water quality and surrounding environmental factors, which highlights the unique nature of fish farming (Luo et al., 2017). During the breeding process, judging the physiological vitality of micropterus salmoides cannot rely on direct contact or arbitrary capture; observation is the primary method. Early fish swimming behavior monitoring mainly relied on direct human observation and manual recording (Barbedo, 2022), which was heavily influenced by subjective judgments and was time-consuming. Therefore, accurate and non-contact, high-throughput acquisition of micropterus salmoides swimming behavior data and subsequent analysis can facilitate timely detection of individual vitality and health status, providing quantitative data references for fish farming practices.

In recent years, deep learning models have gained increasing popularity in agriculture and aquaculture due to their efficient feature learning capabilities and ability to effectively capture complex patterns

* Corresponding author at: Key Laboratory of Fisheries Information, Ministry of Agriculture and Rural Affairs, Shanghai Ocean University, Shanghai 201306, PR China.
E-mail address: gffeng@shou.edu.cn (G. Feng).

https://doi.org/10.1016/j.aquaculture.2024.741554
Received 28 April 2024; Received in revised form 7 August 2024; Accepted 1 September 2024
Available online 5 September 2024

and relationships within data, offering a non-contact and efficient means for fish behavior detection (Yang et al., 2021). For instance, Zhao et al. (2016) proposed a kinetic energy model based on spatial behavioral features and biomimetic technology to monitor fish behavior in recirculating aquaculture systems. Huang et al. (2022) constructed fish school behavior graphs and employed GCN to identify specific behaviors within swimming schools. Zhou et al. (2019) utilized Convolutional Neural Networks and machine vision to detect and evaluate fish appetite. Måløy et al. (2019) introduced a deep learning-based Dual-Stream Recurrent Network (DSRN), leveraging CNN and LSTM to automatically capture the spatiotemporal behavior of salmon during swimming, enabling the prediction of feeding and non-feeding behaviors.

However, current research primarily focuses on distinguishing and identifying fish behaviors without achieving precise localization and swimming behavior analysis of individual fish. In aquaculture, the swimming behavior of each individual and the vitality it reflects should be considered. Multi-object tracking (MOT) technology enables real-time and accurate tracking of all targets within video or image sequences. By employing MOT methods, we can analyze the swimming behavior of individual fish with high throughput, providing more refined aquaculture data. Wang et al. (2022b) proposed a fish abnormal behavior tracking algorithm based on an improved YOLOv5 and SiamRPN++. Li et al. (2022) introduced CMFTNet, a network specifically designed for fish, which shares feature maps between detection and tracking tasks to achieve an end-to-end fish tracking model in complex scenarios. Arvind et al. (2019) presented a Mask R-CNN based segmented neural network for fish detection and utilized the GOTURN algorithm to track the detection results. Liu et al. (2024) designed a Transformer-based method for tracking multiple fish in tank environments.

While the aforementioned studies have achieved some success in fish behavior detection and tracking, they haven't effectively addressed the challenges posed by the multi-scale, multi-directional variations and mutual occlusions that occur when fish swim in water. This leads to a weak association between individual fish detection and tracking trajectories. For multi-object tracking of micropterus salmoides, both individual detection and trajectory association are crucial, and real-world aquaculture environments demand fast and high-throughput tracking capabilities. YOLO is an object detection algorithm that balances speed and accuracy, initially proposed by Redmon et al. (2016) in 2016. Several teams have previously utilized YOLO series models for fish behavior detection ((Hu et al., 2021), (Wang et al., 2022a), (Iqbal et al., 2022)), demonstrating its significant advantage in detection speed. Latest Improved Version of YOLOv8 further allows for different detection tasks by using the same backbone with interchangeable detection heads, facilitating simultaneous measurement of swimming speed and tail beat frequency. ByteTrack, proposed by Zhang et al. (2022), boasts a simple architecture and superior handling of occlusion and loss issues. Numerous teams have applied this tracking algorithm to fish tracking ((Zhao et al., 2024), (Qian et al., 2023)).

To accurately and efficiently obtain individual fish swimming behavior data, this paper proposes a method based on an improved YOLOv8 and an enhanced ByteTrack (Zhang et al., 2022) for real-time and stable tracking of each micropterus salmoides. The improved YOLOv8 addresses the challenges of varying positions and shapes of micropterus salmoides, enhancing the accuracy of object detection and localization. The modified ByteTrack, through improvements in Kalman filtering and the association scheme, reduces instances of track loss due to occlusion or abnormal movement. By utilizing the proposed multi-object tracking algorithm, we capture swimming behavior data of micropterus salmoides, including Swimming Ability Index (SAI), Tail Beat Frequency (TBF), and Reverse flow state, providing quantitative indicators for analyzing vitality and assessing the status of micropterus salmoides.

2. Materials and methods

2.1. Experimental fish conditions

Our experimental protocol received approval from the College of Fisheries and Life Science at Shanghai Ocean University. The experiment was conducted in strict accordance with the guidelines of the College of Fisheries and Life Science at Shanghai Ocean University. The micropterus salmoides used in this study were obtained from the research laboratory of the College of Fisheries at Shanghai Ocean University. The experiment took place at the aquaculture laboratory in the Marine Science and Technology Building at the Lingang Campus of Shanghai Ocean University (quantity: 20, body weight (BW): 9.8 ± 1.8 g). All fish were kept in a specially designed transparent device (Loligo Systems Swimming Respirometry) with a water depth of 0.2 m and a length of 0.5 m. The device has a motor that can generate water flow by adjusting the rotational speed. During measurement, the rotational speed was gradually increased from 100 RPM to 500 RPM, raising the water flow velocity from 8 cm/s to 43 cm/s. The water temperature was maintained at 27.6 ± 1.3 degrees Celsius (mean ± standard deviation). The measurement time was from 11:25 AM to 9:20 PM. Each fish was implanted with a chip for individual identification.

2.2. Overview of the methodological process

The flowchart of the micropterus salmoides swimming behavior analysis is illustrated in Fig. 1. A high-resolution camera was fixed directly above the Loligo Systems Swimming Respirometry to capture real-time video data of swimming micropterus salmoides. The captured video data is converted into a video sequence in the computer, followed by feature extraction, object detection head, and Keypoint Head processing to obtain target fish and keypoint information. Subsequently, swimming behavior data is obtained through the tracking algorithm. Based on the proposed Yolo-AWD network, feature information is extracted, and bounding boxes are used to detect fish bodies and locate key points on their bodies, obtaining fish body positions and key point coordinates. Afterward, the proposed CBT algorithm is applied to associate and track the same targets in consecutive frames of the video sequence. Finally, swimming behavior data is analyzed and acquired using the tracking results.

2.3. Data acquisition and preprocessing

In this study, video data was captured using a Hikvision (DS-U32W) 2-megapixel camera positioned 150 cm above the water tank. The image resolution was set to 1280 × 720, with a frame rate of 30 fps and an MP4 video format. All videos were recorded under fixed lighting conditions. Dataset images were extracted from the video data using FFmpeg software, with one image extracted every 10 frames to prevent excessive similarity due to short frame intervals. This study constructed three datasets for object detection training and validation (Fig. 2a), keypoint training and validation (Fig. 2b), and multi-object tracking algorithm validation (Fig. 2c).

The dataset for training the object detection and keypoint models was constructed from 1325 image frames extracted from the acquired videos. The object detection dataset was annotated using the Labelme tool, excluding instances where micropterus salmoides were occluded by more than 80 %. For the keypoint dataset, based on the physiological structure of fish (Jerry and Cairns, 1998), four points along the fish's spine (head, body, fin, tail) were annotated using the Labelimg tool. The multi-object tracking algorithm validation set, which was not used for training but rather to verify the effectiveness of our method, consisted of 5 randomly selected video segments annotated using DarkLabel 2.4.

Data augmentation is a common technique in deep learning to prevent overfitting and enhance model robustness. Its principle is to expand the training dataset without altering the image category, making the

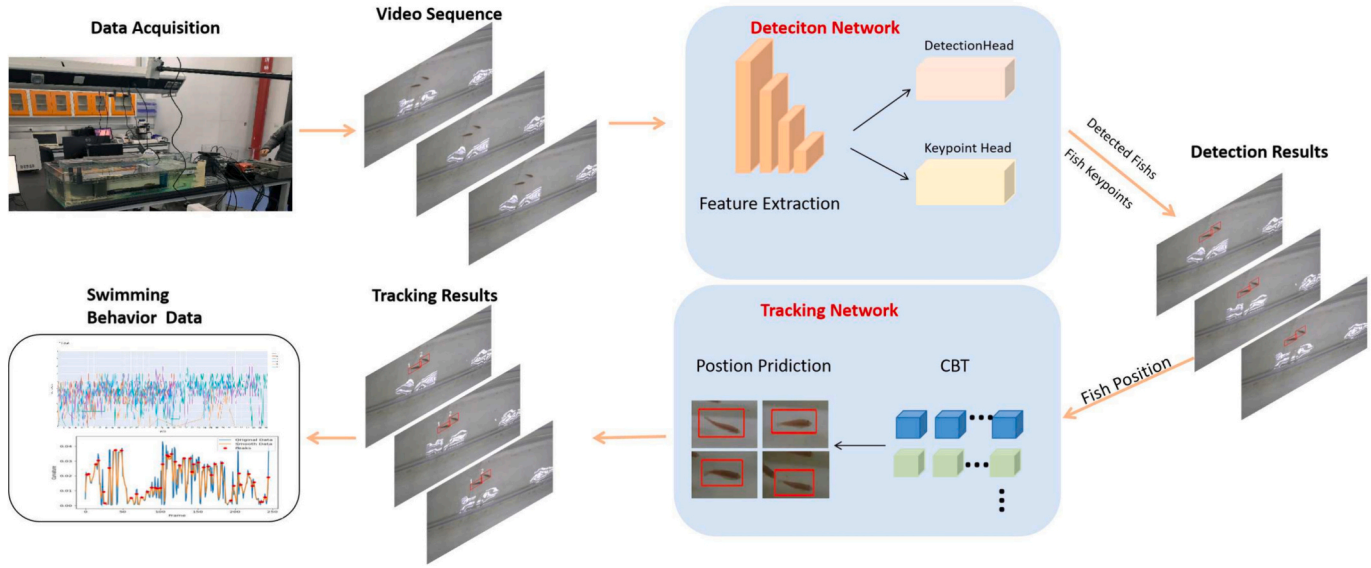


Fig. 1. Overview of the full methodological process: the real-time captured images are fed into our method to achieve fish body detection, keypoint localization, target tracking, and swimming data recording, ultimately enabling fish swimming behavior analysis.

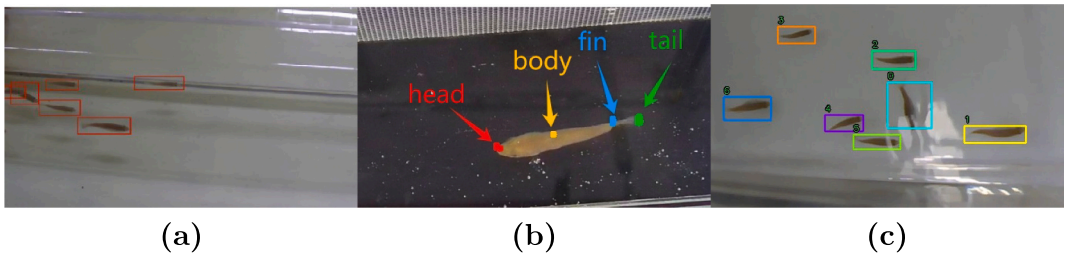


Fig. 2. Dataset annotation illustration. (a) object detection annotation: illustration of bounding box annotations for fish objects; (b) keypoint annotation: illustration of keypoint annotations, with points representing: head, body, fin, tail; (c) multi-object tracking dataset: each fish body is assigned a unique ID for tracking purposes.

dataset as diverse as possible (Song et al., 2021). Considering the influence of various factors such as water flow velocity, lighting conditions, and fish school orientation in both aquaculture and natural environments, and to ensure our method's robust performance across different settings, we applied data augmentation to the original object detection and keypoint detection datasets. This included techniques like noise addition, motion blur, CutOut, brightness adjustments, and image flipping. Specific details are shown in Table 1. After manually verifying the correctness of the augmentation results, the data was divided into training, validation, and test sets in a 7:2:1 ratio.

2.4. Swimming behavior measurement indicator

Analyzing fish swimming behavior is crucial for gaining insights into aquaculture productivity, genetic selection, and breeding strategies (Bao et al., 2018). In measuring micropterus salmoides swimming behavior, we selected three key indicators: SAI, TBF, and RFS. The SAI represents the area enclosed by the curve fitted to swimming speed and swimming time against the coordinate axes. It reflects swimming ability in terms of both swimming speed and duration, making it a valuable indicator for comparing swimming capabilities between species or different growth stages within a species (Duan et al., 2014). The TBF parameter can serve as an indicator of metabolic rate and general activity level. If accurately measured, it holds great potential for monitoring the free-swimming behavior of fish in aquaculture environments (Warren-Myers et al., 2023). The RFS of micropterus salmoides can manifest as three distinct swimming states: upstream movement, upstream stillness, or upstream retreat. These different states reflect varying levels of activity in micropterus salmoides. The calculation of SAI is shown in Eq. (1), where V represents the swimming speed of the micropterus salmoides, L is the pixel distance of the fish body movement between two consecutive frames, ratio is a scaling factor representing the ratio of actual distance to pixel distance, t_1 is the time difference between two consecutive frames, and S is the water flow speed.

$$SAI = \int_0^T Vdt \times 10^{-4} = \int_0^T \left(ratio \frac{L}{t_1} + S \right) dt \times 10^{-2}(\mathrm{m}) \quad (1)$$

TBF calculation utilizes key points to fit a curve along the fish body's spine. By calculating the curvature of the fitted curve, we can obtain the fish's body oscillation pattern. Finally, statistical analysis of curvature variations allows us to determine the tail beat frequency of the micropterus salmoides. The calculation formulas are shown in Eqs. (2)–(3). K represents the curvature of the fitted curve, $\Delta\alpha$ is the change in angle of

Table 1
Data augmentation details.

Data augmentation method	Parameter
Noise Addition	Noise Level: 0.2*255
Motion Blur	Kernel Size: 100 Pixels
CutOut	Max Cutout Size: 2 %, Number: 5–10 Random Location
Brightness Adjustment	Brightness: [0.5,1.5]
Image Flipping	90°, 180°, 270°, Horizontal Flip

the curve, and Δs is the curve length. By counting the number of peaks where the maximum curvature K exceeds a threshold, we obtain the number of tail beats (TBT). Dividing the obtained number of tail beats by the total time T yields the TBF.

$$K = \left|\frac{\Delta\alpha}{\Delta s}\right| \tag{2}$$

$$TBF = TBT/t_2 \tag{3}$$

The RFS is determined by analyzing the relative positions of the fish body's center in two adjacent frames, allowing us to identify and record the swimming state of the micropterus salmoides. Eq. (4) can be used to calculate the proportion of time spent in each swimming state. Here, t2 represents the duration of a specific swimming state, and T is the total duration of the swimming activity.

$$P(\%) = t_1/T \tag{4}$$

When quantitatively comparing swimming behavior data, it is crucial to assess the accuracy and precision of deep learning measurements. To validate the effectiveness of our proposed algorithm, we employed a manual frame-by-frame measurement approach to record ground truth data for each video segment. The accuracy of the algorithm is then analyzed by comparing the deep learning measurements to their corresponding manually measured ground truth values, evaluating the degree of proximity between the two.

2.5. Development of the proposed yolo-AWD+CBT network

Yolo-AWD + CBT is a multi-object tracking method based on the Separate Detection and Embedding (SDE) paradigm. The method (Pipeline) is divided into two parts. First, Yolo-AWD is used to detect and identify micropterus salmoides targets in each frame of the video sequence. After detecting and identifying the micropterus salmoides, the location and confidence information of the identified fish is passed to the CBT algorithm. Utilizing the target's location and confidence information, the current trajectory of the micropterus salmoides is recorded and associated with the predicted trajectory data from the previous frame. Once the matching is complete, the next frame's trajectory position is predicted based on the current trajectory and awaits association with the actual trajectory information from the next frame. Yolo-AWD and CBT are executed sequentially to complete the multi-object tracking of the micropterus salmoides. Fig. 3a displays the Yolo-AWD structure diagram, while Fig. 3b illustrates the CBT structure diagram.

In Fig. 3a, the improved Yolo-AWD model structure is primarily divided into three parts: the Backbone for extracting image features; the Neck for fusing multi-scale feature information; and the prediction Head for predicting confidence scores, classes, and bounding boxes. The Backbone of Yolo-AWD mainly consists of CBS, C2F, AWD (our improved module), and SPPF modules. The Neck part adopts a PANFPN structure, efficiently integrating features from different levels by introducing both bottom-up and top-down pathways, facilitating the transmission of low-level information to higher levels. The detection Head utilizes a Decoupled Head, adjusting the number of image channels for the P3, P4, and P5 feature maps of three different scales through CBS structures, and ultimately used for confidence and class predictions.

CBT module as shown in Fig. 3b. First, the detection results obtained from the Yolo-AWD model are divided into high-score detection box D_{High} and low-score detection box D_{Low} based on two confidence intervals, [θ_{High}, 1] and [θ_{Low}, θ_{High}]. Next, different matching strategies are employed. Initially, the tracklet T is matched with the D_{High}. Unmatched D_{High} generate new trajectories T_{New}, while the remaining trajectories T_{Remain} are matched with the D_{Low}. Finally, the Kalman filter is used to update the tracklet set T. This method utilizes the similarity between detection boxes and trajectories to eliminate background from D_{Low} while retaining D_{High}. Consequently, it enables the discovery of actual targets, such as blurry, occluded, and other challenging objects, reduces missed detection box, and improves tracklet coherency. In the multi-object tracking experiments, the θ_{High}, is set to 0.6and the θ_{Low} is set to 0.2.

2.5.1. Adaptive-weight downsampling module structure

In both tank environments and aquaculture settings, micropterus salmoides exhibit significant variations in body shape and frequently occlude each other during swimming, leading to challenges in accurately extracting contour features and location information, resulting in missed detections by the original network. In the study by Xu et al. (2023), it was found that the Max pooling downsampling module used in the YOLOv8 backbone for feature extraction can lead to information loss, failing to adequately represent the details and subtle contour changes present in the original image or feature map. While most architectures use Max pooling because it is fast and memory-efficient, there is still room for improvement in retaining important information in the activation maps (Stergiou et al., 2021). Furthermore, multiple downsampling operations introduce spatial distortion, resulting in blurring and deformation of target positions on the feature map. To overcome these issues, we drew inspiration from the design concepts of RFAConv (Zhang et al., 2023) and Focus (Bochkovskiy et al., 2020) to develop a novel downsampling module called Adaptive-Weight Downsampling (AWD). This module combines RFA mechanisms with group convolution, allowing the downsampling process to benefit from both local receptive field feature extraction and RFA's emphasis on the importance of different features within the receptive field, prioritizing spatial feature information. As illustrated in Fig. 4, AWD integrates group convolution and RFA methods. The upper part performs downsampling operations based on the Focus slicing concept, but we replace the time-consuming Focus slicing with a 3 × 3 group convolution with a stride of 2 to efficiently capture local receptive field features. The lower part leverages RFA to learn attention maps between global features and generate adaptive weight information for the feature maps. Finally, the downsampled features are multiplied by the attention weights and summed along the last dimension to produce the final downsampled features.

In this paper, we replace the 5 downsampling modules in the backbone and neck with AWD modules. Compared to the original downsampling modules, AWD can utilize attention maps to reduce the loss of object position and contour information while extracting semantic information during downsampling. This helps the model better capture the position information of micropterus salmoides from a global perspective, thereby improving the detection accuracy of the entire model. Yolo-AWD achieves more precise localization capabilities, enhancing the model's ability to extract feature information of micropterus salmoides in complex situations.

2.5.2. Improvement of the loss function

In object detection, bounding box regression (BBR) is a crucial step that determines the performance of object localization. However, the CIOU bounding box loss function used in YOLOv8 exists some drawbacks. CIOU Loss takes into account three important geometric factors: overlap area, center point distance, and aspect ratio, as shown in Fig. 5a. In the calculation of CIOU Loss, as defined in Eq. (5), b and b^{gt} represent the center points of the predicted bounding boxes(B) and the ground truth bounding boxes (B^{gt}).

$$L_{CIOU} = 1 - IOU + \frac{\rho^2\left(b^2 + b^{gt}\right)}{\left(w^c\right)^2 + \left(h^c\right)^2} + \alpha v \tag{5}$$

$\rho = \left\|b - b^{gt}\right\|_2$ represents euclidean distance between b and b^{gt}.c Is the diagonal length of the smallest enclosing box covering the two bounding boxes. $v = \frac{4}{\pi^2}\left(arctan\frac{w^{gt}}{h^{gt}} - arctan\frac{w}{h}\right)^2$ used to measure the consistency of the relative scale of the two rectangular boxes. $\alpha = \frac{u}{1-IOU+u}$

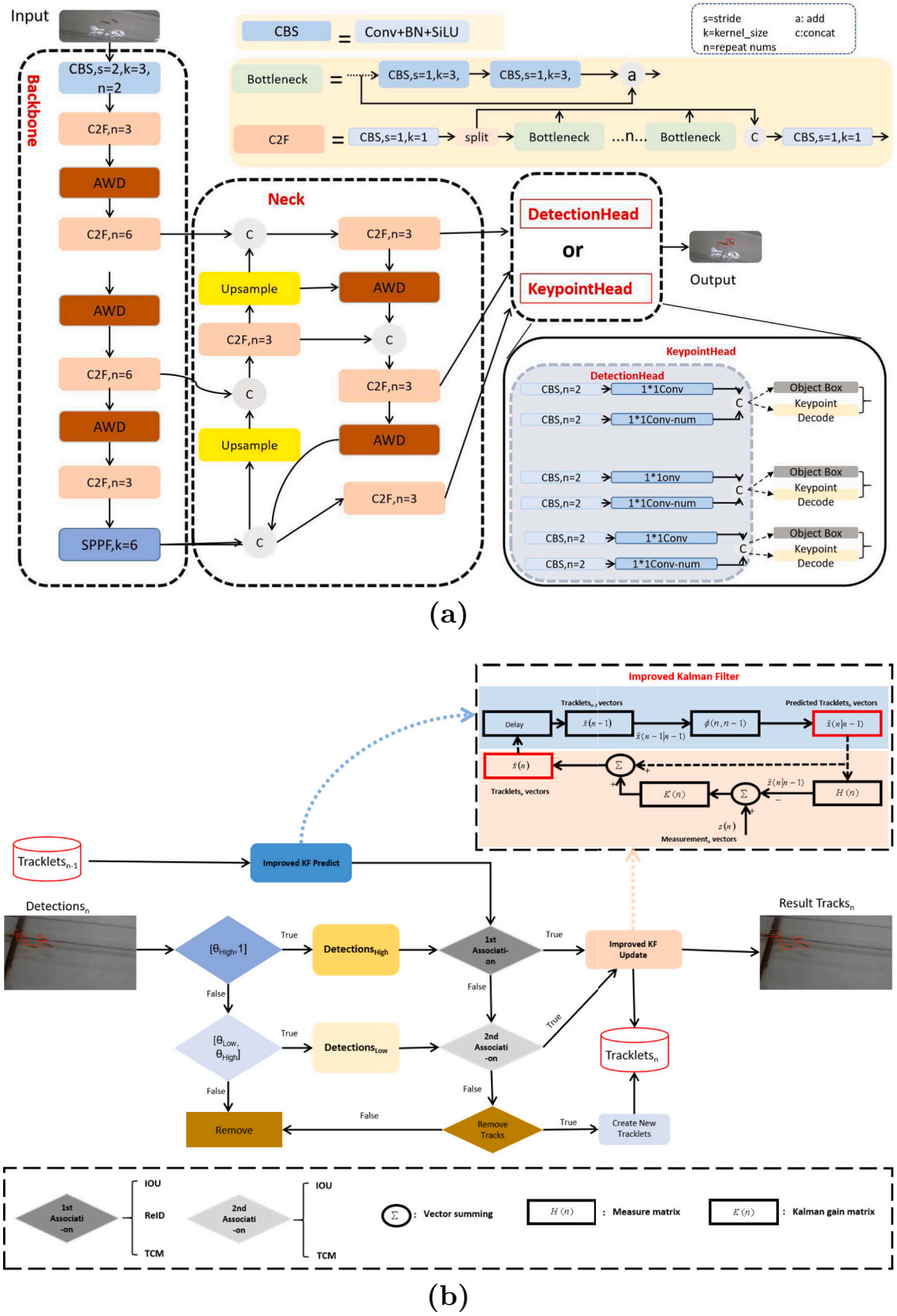


Fig. 3. Algorithm Structure Diagrams of Yolo-AWD and CBT. (a) Yolo-AWD structure diagram illustrates the process of extracting image features, fusing feature information, and utilizing head decoupling to obtain detection results; (b) CBT structure diagram showcases trajectory matching and motion prediction, achieving tracking results through data association.

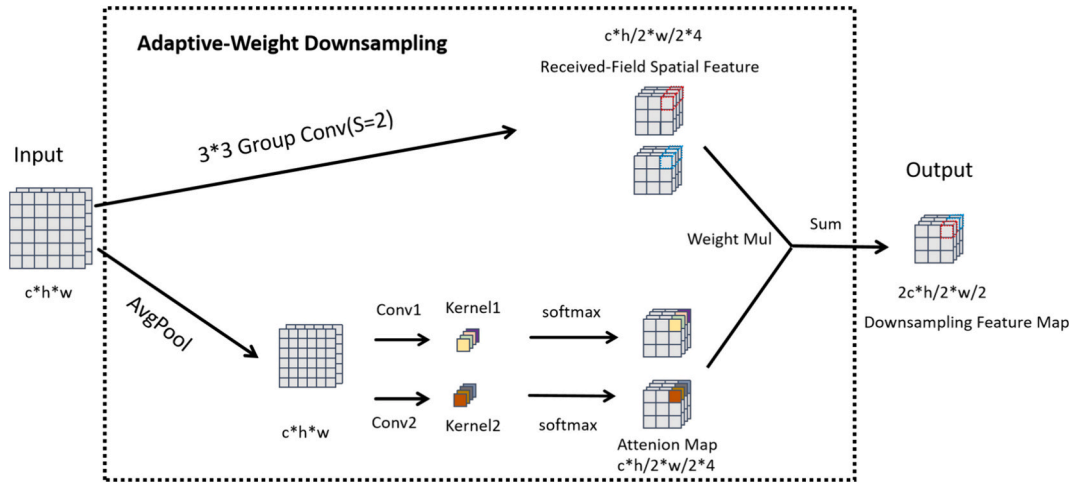


Fig. 4. The architecture of the AWD module: describing the process from the input feature map through group convolution downsampling and receptive field-based attention map generation to the weighted fusion of information, and finally, the output of the downsamplimg feature map.

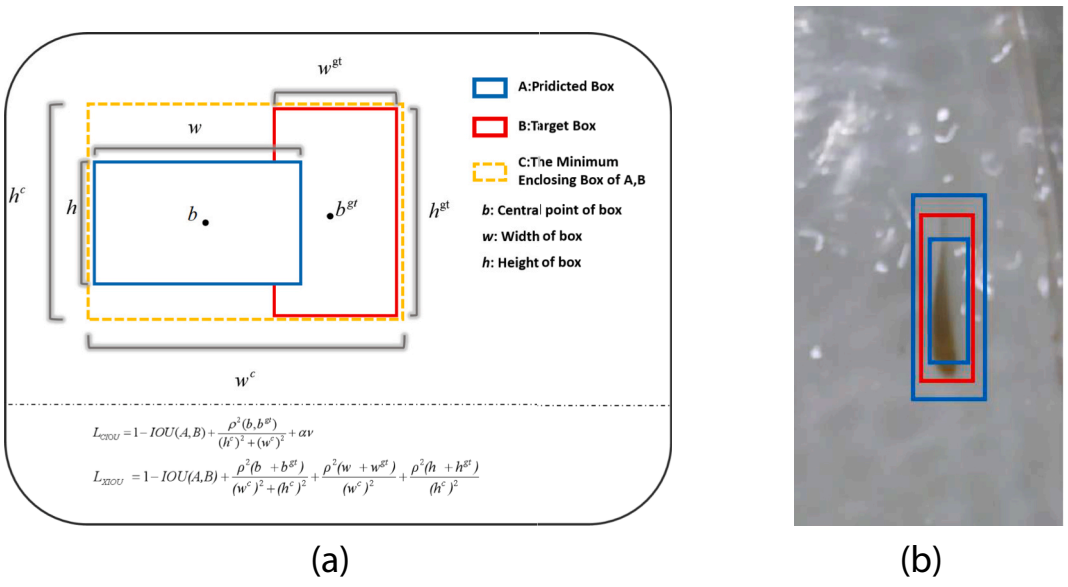


Fig. 5. Schematic diagram of Loss. (a): Calculaton of CIOU and XIOU Loss; (b) Problems existing between the ground truth box and the predicted box in CIOU, where blue represents the predicted box and red represents the ground truth box. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

represents the weight coefficient. The consideration of the aspect ratio in the regression box added in CIOU is not reasonable, as it does not reflect the correct relationship with the ground truth. As shown in Fig. 5b, the three boxes have the same center point and aspect ratio. In the loss function calculation of CIOU, the loss terms of these two predicted boxes are the same. This situation hinders the model from effectively reducing the real difference between (w, h) and (w^{gt}, h^{gt}).

To address the aforementioned issues with CIOU, we propose an improved IOU loss calculation method and introduce a more efficient loss function called XIOU, as defined in Eq. (6).

$$\begin{aligned} L_{XIOU} &= L_{IOU} + L_{dis} + L_{asp} \\ &= 1 - IOU + \frac{\rho^2(b^2 + b^{gt})}{(w^c)^2 + (h^c)^2} + \frac{\rho^2(w^2 + w^{gt})}{(w^c)^2} + \frac{\rho^2(h^2 + h^{gt})}{(h^c)^2} \end{aligned} \tag{6}$$

w^c and h^c represent the width and height of the smallest rectangle that encloses both the predicted box and the ground truth box. b, w, and h denote the predicted box and its width and height, while b^{gt}, w^{gt} and h^{gt} represent the ground truth box and its corresponding width and height. The loss function is divided into three components: IOU loss (L_{IOU}), distance loss (L_{dis}), and aspect ratio loss (L_{asp}). Through this approach, we retain the beneficial characteristics of CIOU Loss. Simultaneously, XIOU Loss directly minimizes the discrepancies between the width and height of the target box and the anchor box, leading to faster convergence and improved localization accuracy.

2.5.3. *Improved tracking algorithm*

The ByteTrack algorithm relies on the overlap between the confidence of the detection boxes and the retained tracks from the previous frame as the matching criterion. The Kalman filter algorithm it employs is typically used to model the motion of objects in a plane (Tan et al., 2022). This algorithm uses a constant velocity and linear observation model to predict and update the target tracks (Bewley et al., 2016), without considering the confidence of the tracks themselves. The reason why confidence states are useful for association is straightforward: when severe occlusion and clustering occur, the previously ignored confidence

state provides effective information to compensate for ambiguity. The confidence state of the tracks can clearly indicate the occlusion/occluded relationship between clustered targets, offering a critical clue when spatial and appearance information is lacking. From a high-level perspective, it is common sense and a key criterion that the state of the tracks should continuously change at each time step. As one of the states provided by the detector, the confidence state of the same track should also exhibit temporal continuity. Specifically, when spatial and appearance information fail due to a high overlap of multiple objects, the confidence of object tracks can provide a clear occlusion/occluded relationship, which is exactly what spatial and appearance information lack. This is because, when detecting unoccluded objects, the tracks' confidence will have a higher confidence score, whereas occluded objects impose greater demands on detector performance, leading to lower confidence scores.

Therefore, we combine confidence information with ByteTrack's methods for detecting high-confidence and low-confidence bounding boxes to propose CBT methods. When an individual fish is not occluded or only slightly occluded, the Kalman filter is an ideal model for modeling and estimating the motion tracklet within a small range of variations. In order to enhance the Kalman filter's estimation of fish motion trajectories, we extend the standard Kalman filter widely used in SORT (Bewley et al., 2016) by adding two additional states: tracklet confidence c and its velocity component $\dot{c}$. The standard Kalman filter state in SORT is shown in Eq. (7). Here, u and v represent the center of the fish, s and r respectively denote the scale (area) and aspect ratio of the object detection box. The velocity component is represented by $\dot{u}$, $\dot{v}$, and $\dot{s}$.

$$\chi = [u, v, s, r, \dot{u}, \dot{v}, \dot{s}] \tag{7}$$

Including the two newly introduced states c and $\dot{c}$, the complete state of the Kalman filter improved in CBT methods is shown in Eq. (8).

$$\chi = [u, v, s, r, c, \dot{u}, \dot{v}, \dot{s}, \dot{c}] \tag{8}$$

For the second association step involving low-confidence detections, we utilize linear prediction to estimate the matching tracklet confidence. ByteTrack has demonstrated that low detection confidence often corresponds to severe occlusions and clustering. During the onset or conclusion of occlusions, the confidence of fish objects undergoes drastic changes (suddenly decreasing or increasing). However, the Kalman filter exhibits noticeable lag when attempting to estimate sudden changes in confidence states. Nevertheless, we observe a clear directional trend in the confidence state during this period (i.e., continuous increase or decrease). Therefore, we employ a simple linear prediction based on tracklet history to address this issue. The formula for linear modeling is given by Eq. (9). where $\widehat{c}_{trk}$ represents the tracklet confidence stored in Tracklet Memory. When using the Kalman filter or linear prediction, the confidence cost calculation is computed as the absolute difference between the estimated tracklet confidence and the detection confidence, as shown in Eq. (10):

$$\widehat{c}_{trk} = \begin{cases} c_{trk}^{t-1}, & c_{trk}^{t-2} = None \\ c_{trk}^{t-1} - \left(c_{trk}^{t-2} - c_{trk}^{t-1}\right), & else \end{cases} \tag{9}$$

$$\mathbf{C}_{Conf} = |\widehat{\mathbf{c}}_{trk} - \mathbf{c}_{\mathrm{det}}| \tag{10}$$

3. Results

3.1. Experimental environment and configuration

In Section 2.3, a total of 1325 images were used for feature extraction model training and testing. All captured images were divided into training and testing sets. The experimental configuration for this study is as follows: Intel(R) Core(TM) i9-10900K CPU @ 3.70GHz + Ubuntu 20.04.4 LTS + NVIDIA GeForce RTX 3090 24G. The deep learning framework used is PyTorch 2.2, with YOLOv8 version 8.0.202 as the baseline. The input image size is 640 × 640. Training was conducted on a server via remote connection using VSCode locally. The initial learning rate was set to 0.001, and the weight decay was set at 0.0005. During the training process, we employed the Automatic Mixed Precision (AMP) training strategy commonly used in deep learning, which combines single precision (FP32) and half precision (FP16) while training the network, achieving nearly the same accuracy with the same hyperparameters as FP32. Each model was trained for a maximum of 200 epochs with a batch size of 32.

3.2. Validation evaluation indicators

To evaluate the effectiveness of our proposed algorithm and its tracking performance in different scenarios, we conducted a two-part validation process: (1) verifying the performance of the Yolo-AWD detector and (2) assessing the accuracy of the improved CBT algorithm. The detector's performance was evaluated using five metrics, including Precision (P), Recall (R), Average Precision (AP) as defined in Eqs. 11 to. 13, model size, and floating-point operations (FLOPs).

$$P = \frac{TP}{TP + FP} \tag{11}$$

$$R = \frac{TP}{TP + FN} \tag{12}$$

$$AP = \int_0^1 PRdR \tag{13}$$

Where TP (True Positives) refers to the number of correctly identified positive samples; FN (False Negatives) represents instances where the model incorrectly predicts positive samples as negative; and FP (False Positives) indicates instances where the model incorrectly predicts negative samples as positive. We can calculate different mAP values based on the IoU threshold setting. For example, AP_{50} refers to the average precision when IoU is greater than or equal to 0.5, while $AP_{50:95}$ is the average AP across IoU thresholds ranging from 0.5 to 0.95 with a step size of 0.05.

For evaluating the CBT tracking algorithm, we selected Higher Order Tracking Accuracy (HOTA) (Luiten et al., 2021), Multiple Object Tracking Accuracy (MOTA), Multiple Object Tracking Precision (MOTP), and Identification F1 (IDF1) as the multi-object evaluation metrics. HOTA introduces a higher-dimensional tracking accuracy metric, providing a more comprehensive and balanced assessment of the multi-object tracker's performance. MOTA measures the tracker's performance in detecting targets and maintaining tracks. MOTP evaluates the localization accuracy of the detections. IDF1 reflects the tracker's stability, with higher values indicating that the algorithm can accurately track targets over longer periods. The HOTA calculation formula is as follows: Eq. (14) where *DetA* represents the detection accuracy score, and *AssA* represents the association accuracy score. TP refers to the number of true positive samples; FN is the number of instances where the model predicts positive samples as negative; FP represents the number of instances where the model predicts negative samples as positive; and C is the set of points belonging to TP, based on which we can always determine a unique ground truth trajectory. A(c) represents the association accuracy, defined by Eq. (15), where TPA(c) is the accuracy of correctly associated detections, FPA(c) is the accuracy of unassociated or falsely associated predicted trajectories, and FNA(c) is the accuracy of unassociated or falsely associated ground truth trajectories.

$$\mathrm{HOTA} = \sqrt{\mathrm{DetA} \cdot \mathrm{AssA}} = \sqrt{\frac{\sum_{C \in TP} A(c)}{TP + FN + FP}} \tag{14}$$

$$A(c) = \frac{TPA(c)}{TPA(c) + FPA(c) + FNA(c)} \tag{15}$$

MOTA is calculated as shown in Eq. (16). Here, FP represents the total number of false positives in the t-th frame; FN denotes the total number of missed detections in the t-th frame; IDS refers to the number of times the target label ID switches during tracking in the t-th frame; and gt represents the number of ground truth objects observed at time t.

$$MOTA = 1 - \frac{\sum_t (FP + FN + IDS)}{\sum_t g_t} \tag{16}$$

The MOTP calculation is represented as Eq. (17). Here, $d_{i,t}$ denotes the distance between the given position and its hypothesized position in frame t, c_t represents the matching count between the target and the hypothesized position in frame t, and i denotes the current detected target.

$$MOTP = \frac{\sum_{t,i} d_{i,t}}{\sum_t c_t} \times 100\% \tag{17}$$

The IDF1 calculation is given by Eq. (18). Here, IDTP represents the total number of correctly tracked targets when the ID remains unchanged; IDFP represents the total number of incorrectly tracked targets when the ID remains unchanged; IDFN represents the total number of targets lost during tracking when the ID remains unchanged.

$$IDF1 = \frac{2IDTP}{2IDTP + IDFP + IDFN} \tag{18}$$

In addition, this paper also employs two other metrics to evaluate the model performance, including the Identity Switches (IDS) and Frames Per Second (FPS). Higher values of MOTA, MOTP, IDF1, and FPS, and lower values of IDS indicate better model performance. During the model evaluation phase, We employed Precision, Recall, mAP, Parameters and GFLOPs to evaluate the performance of the object detection model.

3.3. Experimental and prediction results

3.3.1. Yolo-AWD experimental results

After training on the training set, the evaluation metrics of the proposed Yolo-AWD network are shown in Fig. 6. The F1 score, ranging between 0 and 1, closer to 1 indicates a better balance between precision and recall in the model. Box loss and dfl loss represent the loss of object detection boxes, reflecting the disparity between predicted boxes and ground truth boxes, while cls loss refers to the loss of target categories, indicating whether the categories are correctly predicted. Smaller loss values imply more accurate predictions by the model. The Fig. 6c illustrates the performance of the model at different precision and recall levels, which is used to select the best model. As shown in Fig. 6, after 300 iterations, our model reaches its optimum. According to the trained optimal network, the detection accuracy for *Micropterus salmoides* is 98.71 %, with a recall rate of 98.2 %, and AP50: 95 reaches 69.73 %. Based on these results, our proposed model achieves high detection accuracy, covering all fish detections.

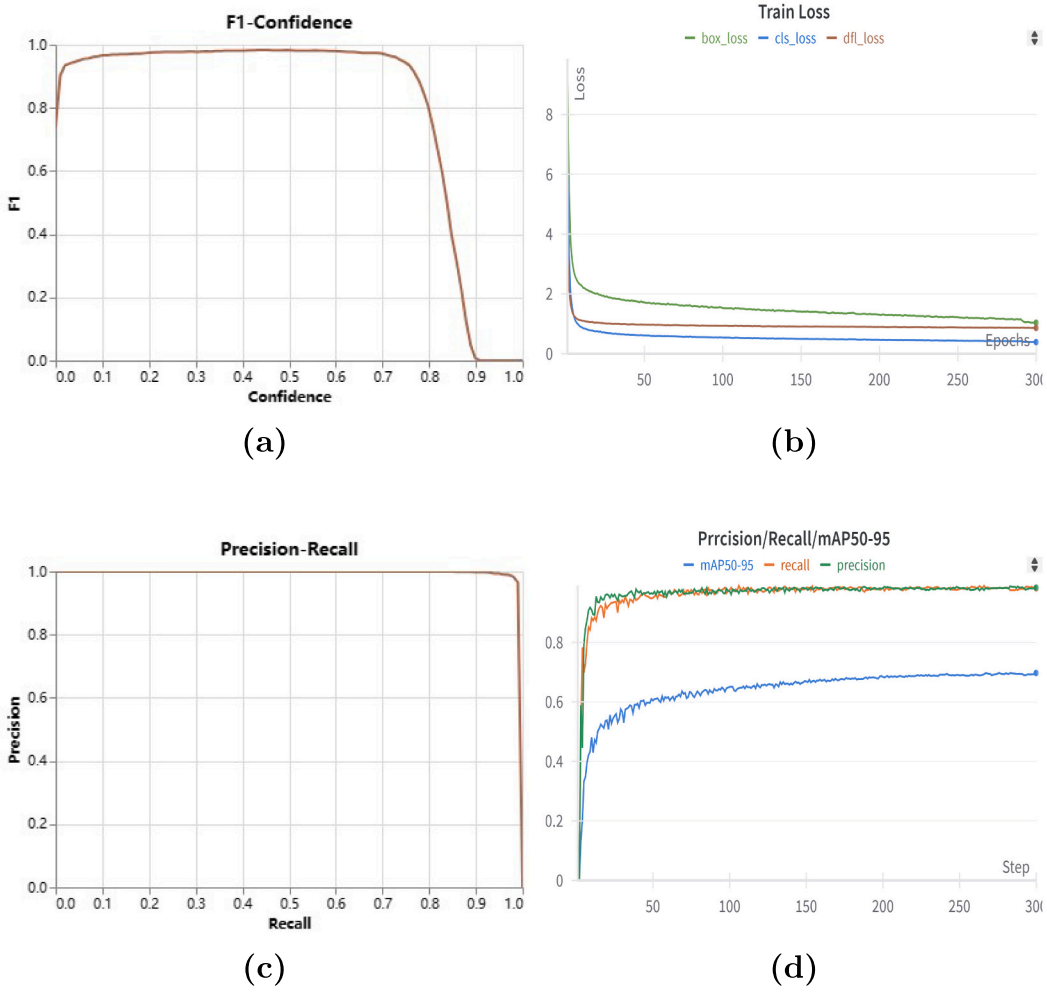


Fig. 6. Yolo-AWD Training Loss Graph. (a) The F1 score metric during the training process; (b) The value of the loss function during the training process; (c) The variation curves of accuracy and recall during training; (d) The relationship between accuracy and recall during training.

To further validate the effectiveness of our proposed improvement to the YOLOv8 algorithm, this paper conducts a comparative analysis under the same experimental conditions with Faster-RCNN (Ren et al., 2015), SSD (Liu et al., 2016), YOLOXn (YOLOX-Nano) (Ge et al., 2021), YOLOv5n (Bochkovskiy et al., 2020), YOLOv8n, and our improved Yolo-AWD. The results are shown in Table 2.

As shown in Table 2, our proposed Yolo-AWD network achieves the best performance in detecting individual *Micropterus salmoides*, with a P of 98.71 %, AP_{50} of 99.43 %, and $AP_{50:95}$ of 69.73 %. Compared to YOLOXn, although our model size and FLOPs are increased by 6.6 MB and 4G respectively, our P, R, AP_{50}, and $AP_{50:95}$ have also improved by 7.71 %, 8.7 %, 6.82 %, and 12.45 %, respectively. Compared to the original YOLOV8n, our model has achieved a reduction in size by 0.8 MB and a decrease in FLOPs by 0.6G, while simultaneously increasing Precision by 1.07 %, AP_{50} by 0.04 %, and $AP_{50:95}$ by 5.4 %. These results indicate that the proposed Yolo-AWD network for *Micropterus salmoides* detection achieves optimal performance and good detection speed.

3.3.2. CBT experimental results

To validate the effectiveness of the CBT algorithm improvement, five different scene videos were selected as test videos, and comparisons were made with state-of-the-art multi-object tracking models. The experimental results are shown in Table 3. From Table 3, it can be observed that our method achieves the highest MOTA and HOTA scores. Although slightly lower in FPS compared to the YOLOv5n + ByteTrack algorithm, overall, our tracking algorithm demonstrates better accuracy in multi-object tracking of *Micropterus salmoides*. This indicates that our method is more suitable for multi-object detection of *Micropterus salmoides*. Compared to using the original YOLOv8n and ByteTrack tracking methods, our improved method not only shows improvements in inference speed but also achieves slightly higher accuracy, demonstrating the effectiveness of our improved Yolo-AWD. Additionally, the improvements in HOTA, MOTA, and MOTP metrics demonstrate that our enhanced trajectory confidence scheme in ByteTrack effectively enhances the efficiency and accuracy of multi-object tracking of *Micropterus salmoides*.

The HOTA metric of our algorithm is 66.93 %, which represents a 6.25 % improvement over the original model (YOLOv8n + ByteTrack). This indicates that our algorithm exhibits superior overall performance in tracking trajectories, with enhancements in both the quantity and accuracy of tracked targets and their IDs. The MOTA is 3.15 % higher than the original model, demonstrating that the improved model maintains higher ID accuracy, and can correctly identify fish even under occlusion or movement. The MOTP is 0.81 % higher than the original model, indicating an improvement in the model's ability to predict the object's position. Our model more accurately predicts the position of the object at the next time step. Additionally, the IDsw decreased by 58.33 % compared to the original model, signifying that our model is better at accurately identifying and tracking targets, thereby reducing both missed and false detections. Moreover, the tracking speed of YOLO-AWD and CBT has improved from 49.7 fps to 52.3 fps compared to the original model. This indicates a slight increase in computational speed while maintaining higher accuracy and stability. The primary reasons for this improvement are the reduced computational load of the target detection network due to the use of the AWD module, and the computational advantages brought by linear matching in the second match using CBT.

Table 2
Comparison of object detection method.

Method	Model size (MB)	P(%)	R(%)	AP_{50}(%)	$AP_{50:95}$(%)	FLOPs (G)
Faster-RCNN	103.4	81.2	84.5	85.3	52.4	947.8
SSD	92.6	87.1	72.6	76.8	49.6	274.0
YOLOXn	**3.6**	90.4	89.5	92.61	57.28	**4.1**
YOLOv5n	9.56	98.7	97.92	99.4	61.16	7.7
YOLOv8n	11.4	97.64	**98.82**	99.39	64.33	8.7
Ours	10.2	**98.71**	98.20	**99.43**	**69.73**	8.1

Table 3
Comparison results of multi-object tracking algorithms.

Models	HOTA (%)	MOTA (%)	MOTP (%)	IDF1 (%)	IDsw	FPS/(f·s-1)
FairMOT	49.36	74.34	59.24	68.34	155	37.8
CMFTNet	51.12	81.35	62.45	69.2	151	34.1
YOLOXn+DeepSORT	49.56	74.23	60.27	67.78	150	40.2
YOLOXn+StrongSORT	52.44	76.35	68.45	69.32	134	45.8
YOLOXn+ByteTrack	54.37	78.27	69.17	72.45	74	50.1
YOLOv5n + DeepSORT	54.34	76.25	65.38	69.34	145	54.7
YOLOv5n + StrongSORT	55.61	**80.85**	70.54	70.02	124	51.6
YOLOv5n + ByteTrack	59.57	79.61	69.97	77.57	36	**56.6**
YOLOv8n + DeepSORT	57.92	83.24	68.56	72.34	115	45.6
YOLOv8n + StrongSORT	59.41	88.65	70.73	86.49	97	43.5
YOLOv8n + ByteTrack	60.68	87.24	72.89	88.26	24	49.7
Yolo-AWD + ByteTrack	63.22	86.26	72.83	86.51	21	50.2
YOLOv8n + CBT	62.56	88.45	73.04	89.57	17	51.6
Yolo-AWD + CBT (Ours)	**66.93**	**90.39**	**73.7**	**93.26**	**10**	52.3

Fig. 7 displays the tracking results of DeepSORT, StrongSORT, ByteTrack, and Yolo-AWD + CBT in Video 2, which was recorded in a scene with strong water flow, causing significant water wave interference and additional occlusion of *Micropterus salmoides*. Compared to the methods proposed in this paper, DeepSORT and StrongSORT exhibit varying degrees of missed detections, especially with numerous ID switches. This is because these methods directly discard low-scoring boxes during tracking, leading to more missed detections and ID switches when facing severe occlusions. Both YOLOv8 and ByteTrack have been improved in this paper, maintaining stable multi-object tracking performance even under complex light and water wave interference conditions. Compared to other multi-object tracking methods, our algorithm demonstrates superior tracking performance in complex scenes, achieving good multi-object tracking performance for *Micropterus salmoides* in different scenarios. In factory farming environments, the high density of *Micropterus salmoides* often results in occlusion, and video monitoring is easily affected by reflections caused by lighting on the water surface and interference from water flow, leading to poor target image conditions. Fig. 8 shows the performance analysis of our tracking algorithm before and after improvements under conditions of fish body occlusion and interference from water flow and lighting. Under the influence of light reflections, the fish targets appear similar in colour to the background due to reflections, and there is mutual occlusion among individuals. This causes the original model to experience prolonged periods of missed detections for multiple targets (indicated by black dashed boxes in the figure). However, Yolo-AWD + CBT successfully tracks the corresponding tilapia targets. Although both the original and improved models fail to detect the same target for 12 consecutive frames due to lighting, Yolo-AWD + CBT successfully re-identifies the target at frame 414. Overall, compared to the original model, the proposed method exhibits better tracking performance in relatively complex environments.

3.3.3. Swimming behavior analysis experimental results

The Yolo-AWD + CBT algorithm proposed in this paper can accurately and quickly identify and track micropterus salmoides in the video (Fig. 9a). The tracked micropterus salmoides are enclosed in rectangular boxes in the image, with each individual assigned a unique ID. Despite the small differences in shape and size between each fish, the motion trends and states of the same fish between adjacent frames are similar.

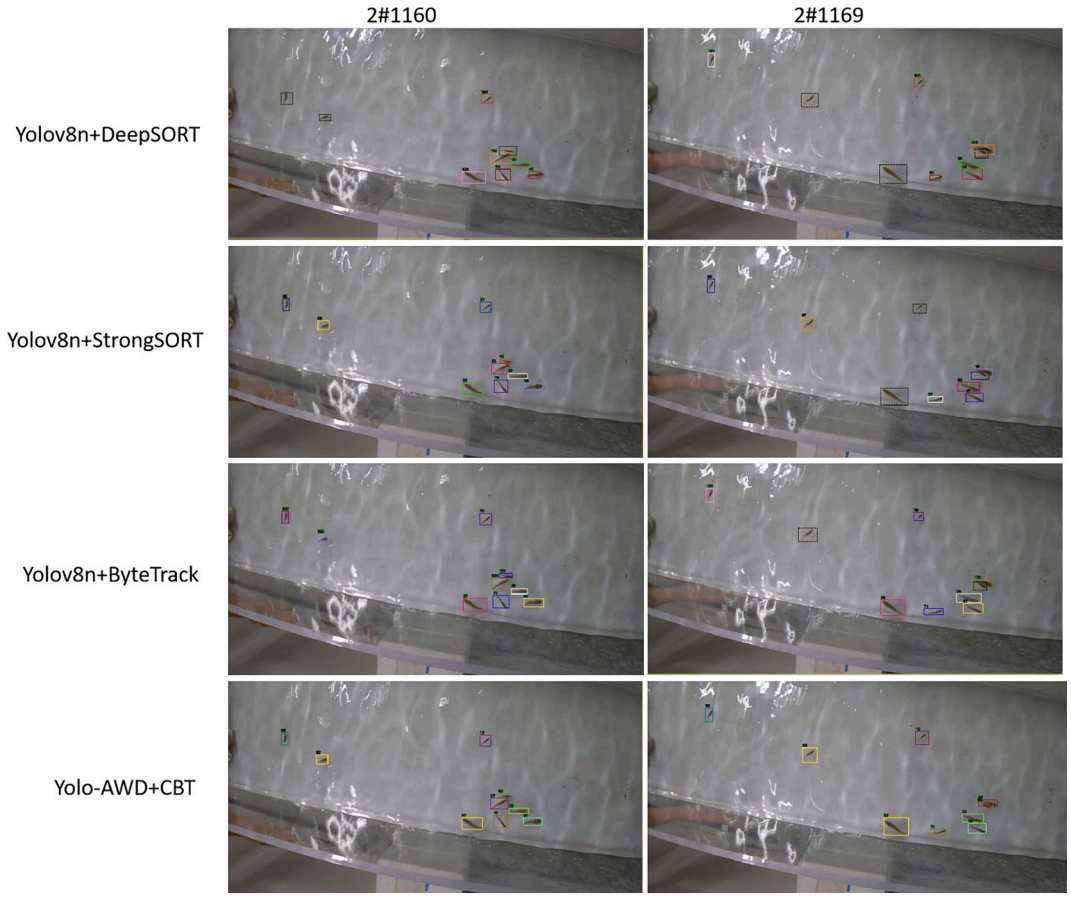


Fig. 7. The comparison of tracking results among DeepSORT, StrongSORT, ByteTrack, and Yolo-AWD + CBT is shown in the figure. The black dashed boxes in the image represent missed or false tracked targets.

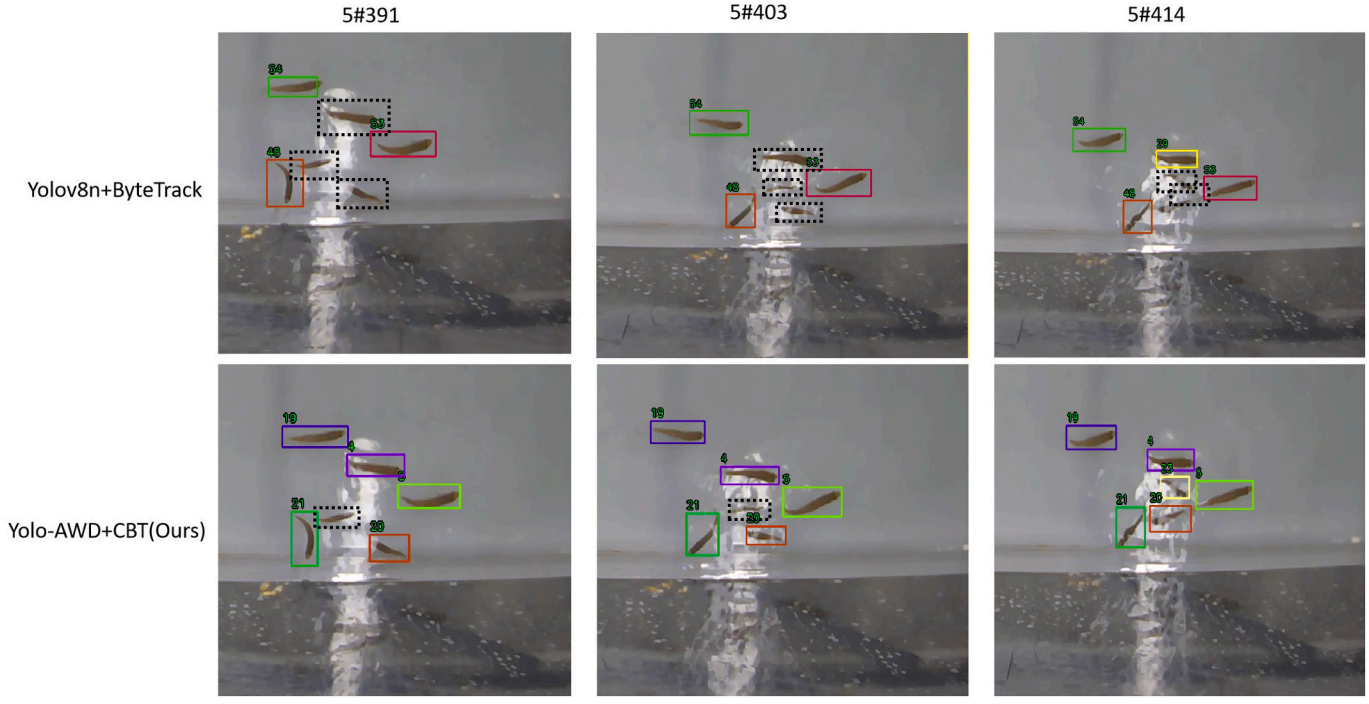


Fig. 8. The comparison between YOLOv8n + ByteTrack and Yolo-AWD + CBT under the influence of lighting and water waves is shown in the figure. The black dashed boxes in the image represent missed or false tracked targets.

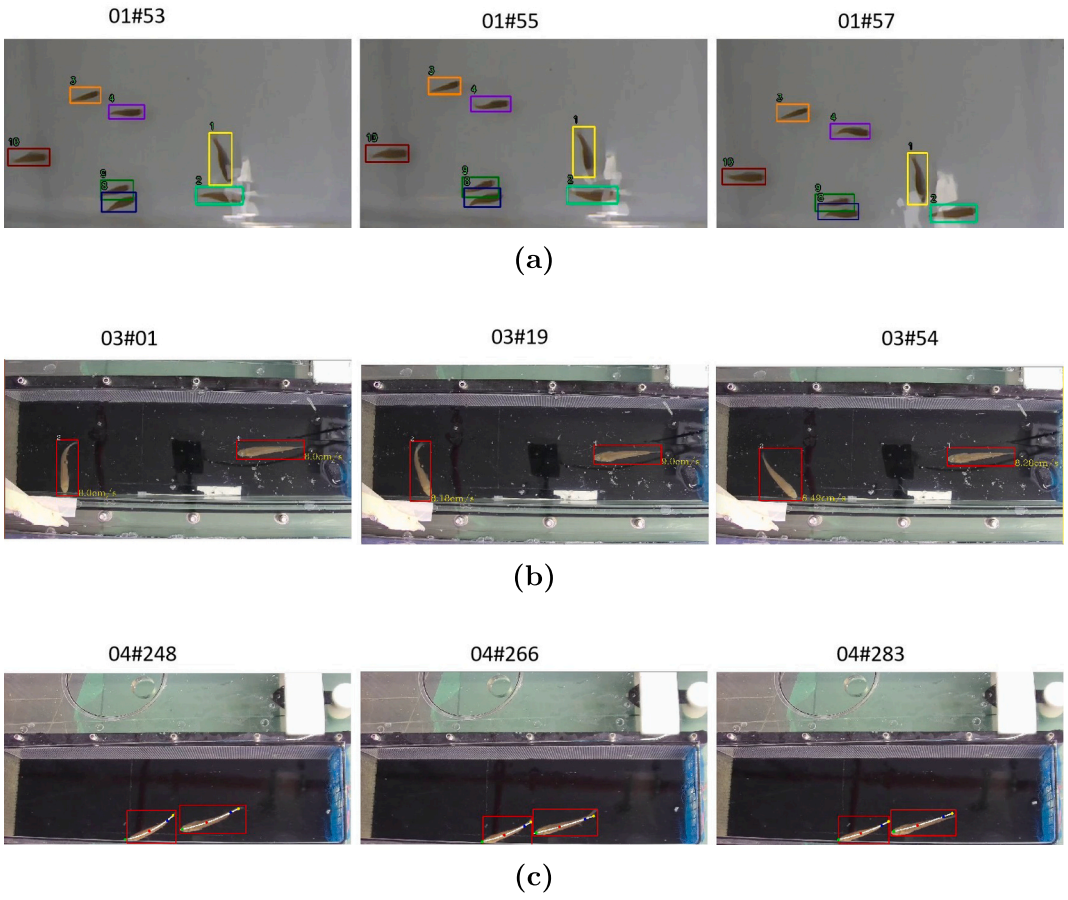


Fig. 9. Swimming behavior analysis results. (a) Effectiveness of target tracking; (b) Visualization of swimming speed measurement; (c) Visualization of fish body curvature estimation.

By matching similar trajectory states between consecutive frames using computer vision, effective tracking of micropterus salmoides can be achieved. Fig. 9b and Fig. 9c respectively show the visualization of micropterus salmoides swimming speed and tail-beating frequency obtained through the CBT algorithm, with selected frames from the video for display. When calculating the swimming speed and tail-beating curvature in each frame, the computer not only draws them in the image but also automatically records the ID, speed, and curvature values for each fish at the current time, saving them to an Excel file. This greatly saves time compared to manual observation and measurement.

Due to the replacement of the original downsampling module with the AWD module in the entire model, the computational cost of downsampling during feature extraction is reduced, leading to an improvement in detection speed. When performing multi-object tracking of micropterus salmoides, the algorithm achieves a frame rate of 52.3f·s^{-1}, meeting the requirements for real-time acquisition of micropterus salmoides swimming behavior data from videos.

In this study, the frame rate of the video data is 30 frames per second, so the time interval between each two frames is $T = \frac{1}{30}$ seconds. By measuring the length of the swimming pool as 46 cm, which corresponds to a pixel distance of 1276 pixels in the video, we calculate the ratio as $\frac{46}{1276} = 0.0384$. L represents the pixel distance of the same target fish between two frames, and S represents the water flow velocity generated by the water pump. Using Eq. (1), we can obtain the corresponding SAI value for each individual micropterus salmoides (Fig. 10a). Additionally, according to Eq. (4), we can obtain the RSF for each individual (Fig. 10b), where '-1' represents backward movement against the flow, '0' represents stationary against the flow, and '1' represents forward movement against the flow. By comparing the proportion of states, we can evaluate the swimming ability or vitality among individuals.

When calculating the Tail-Beat Frequency (TBF) of micropterus salmoides, we obtain the coordinates of the four key points of the fish spine and fit them into a quadratic curve. To ensure the curve fits the micropterus salmoides well, we set the x-coordinate with a step size of 1 and draw the curve according to the fitted quadratic curve. Subsequently, we use differential calculus to calculate the length (Δs) and angle change ($\Delta \alpha$) of the spine curve. By inputting them into Eq. (2), we can determine the bending curvature of the fish body at that moment. Then, by using a filtering system to remove systematic errors and counting the number of curvature peaks, we can represent the Tail-Beat Times (TBT), as shown in Fig. 10c. We then use the counted peaks and the time T in Eq. (3) to calculate the average tail-beat frequency of the entire video, as shown in Fig. 10d.

As shown in Fig. 10, we can observe that ID12 has the highest SAI value, indicating strong overall swimming ability. IDs 19 and 20 have the highest proportion of backward movement against the flow, suggesting that these two fish have poor swimming ability against the current and exhibit lower vitality when faced with external stimuli. ID15 and ID12 have similar SAI values, but ID15 has a higher TBF, indicating that ID15 exhibits more vitality when swimming against the current.

To evaluate the accuracy of our algorithm in obtaining micropterus salmoides swimming behavior data, we compared the average values obtained by manual frame-by-frame measurements with those obtained by the algorithm. Table 4 compares the average values of swimming behavior data measured by our Yolo-AWD + CBT model with those recorded manually. The errors in SAI and PSF measured by our Yolo-AWD + CBT model are relatively low, with deviations of only +0.85 %, +2.12 %, −1.36 %, and − 1.28 %, indicating excellent performance of our model in detecting swimming action data, meeting the requirements

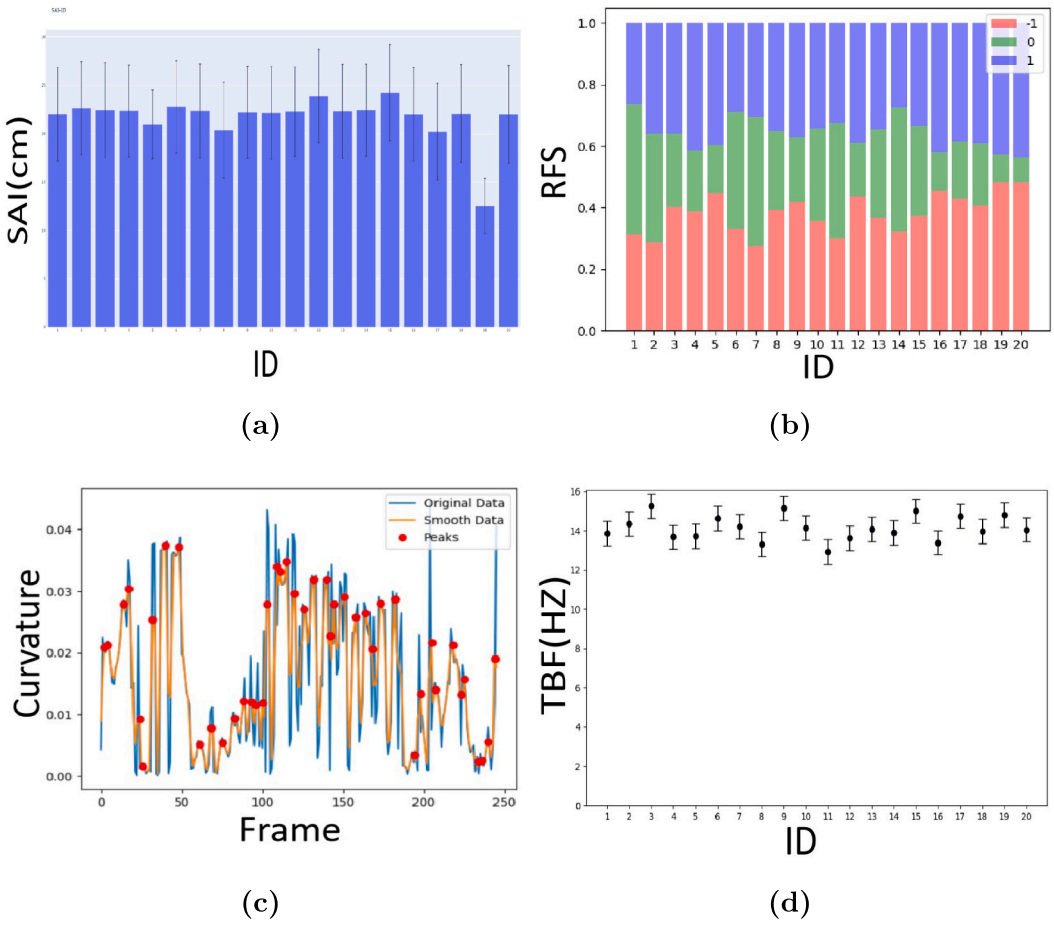


Fig. 10. Results of micropterus salmoides swimming behavior data: (a) Bar chart of SAI-ID errors during swimming; (b) Comparison chart of RFS-ID, where '0' represents stationary against the flow, '1' represents forward movement against the flow, and '-1' represents backward movement against the flow; (c) Record of *Micropterus salmoides* body curvature, with the blue line representing the raw data, the yellow line representing the data after filtering by the filter, and the red dots indicating the peak values; (d) Scatter plot of TBF-ID errors. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 4
Comparison of swimming behavior data acquisition methods.

Data acquisition method	SAI(m)	TBF (HZ)	Pstatic (%)	Pforward (%)	Pbackward (%)
Manual	21.94	14.66	37.46	26.12	36.42
Yolo-AWD + CBT	22.13	13.75	38.27	25.77	35.96
Error rate	+0.85 %	−6.62 %	+2.12 %	−1.36 %	−1.28 %

of aquaculture. However, there is a − 6.62 % error in TBF measurement, which may be due to inaccurate key point detection and positioning when the micropterus salmoides's tail-beat frequency is too fast, resulting in less obvious peaks when calculating TBT, leading to a lower TBF than the actual data.

In summary, Yolo-AWD + CBT can rapidly and fairly accurately obtain swimming behavior data of micropterus salmoides, providing quantitative technical support for the assessment of micropterus salmoides physiological activity and significantly saving manpower and material costs.

At various stages of micropterus salmoides growth and in different water quality environments, swimming behavior can largely reflect their growth, vitality, and breeding conditions. Real-time monitoring of micropterus salmoides swimming behavior during aquaculture allows for timely adjustments to farming strategies and the detection of any abnormal occurrences. Analyzing micropterus salmoides swimming behavior data through computer vision not only saves a significant amount of manpower and resources, reduces aquaculture costs, and improves farming efficiency but also enables precise monitoring of micropterus salmoides growth conditions, facilitating targeted farming plans. However, traditional manual methods of measuring swimming behavior are neither real-time nor error-proof. In laboratory conditions, acquiring behavior data either involves expensive and complex behavior analysis systems or requires sensors to be installed on fish bodies, which can affect the normal physiological status of the fish.

Therefore, developing a system that can track and analyze fish swimming data in real-time not only helps aquaculture quickly obtain various movement data of individual fish but also provides a non-contact, high-throughput means of analyzing swimming behavior in laboratory environments. Such a system can provide a new, more efficient way to evaluate and analyze fish or other aquatic animal products, which is beneficial for both aquaculture and laboratory environments. The micropterus salmoides swimming behavior analysis in this paper is based on YOLOv8n target detection and ByteTrack tracking. Additionally, we added the AWD module to Yolov8n, which reduces feature information loss during downsampling by combining adaptive weights and group convolution, thereby preserving more information during feature extraction. This effectively improves the accuracy of micropterus salmoides detection. Furthermore, the use of the improved XIOU LOSS partially eliminates detection errors caused by differences in aspect ratios among fish populations. This is because the XIOU LOSS function separately considers the width and height of the regression box during

the calculation process. Therefore, during model training, the aspect ratio of predicted boxes is not restricted, increasing the model's generalization ability and improving the positioning accuracy of micropterus salmoides. Compared to other established models, the high accuracy and high recall rate of our model demonstrate its effectiveness (Table 2). The original ByteTrack tracking algorithm, when faced with severe occlusion, retains low-score detection boxes, leading to significant tracking errors and target loss during tracking. Therefore, to address this issue, we extended the Kalman filter algorithm, incorporated trajectory confidence into matching, and used linear matching for predictions with low trajectory confidence to reduce trajectory loss, significantly improving target tracking accuracy (Table 3).

Based on the above improvement strategies, our proposed Yolo-AWD + CBT network can obtain micropterus salmoides swimming behavior data in real-time, including SAI, TBF, RFS, with errors between −7 % and + 3 % compared to manually measured data (Table 4), and processing speed reaching 52.3 frames per second. Comparative experiments with the original tracking model (YOLOv8 + ByteTrack) demonstrate that our proposed improvement strategies effectively improve the accuracy of micropterus salmoides tracking during swimming. These results fully demonstrate that our Yolo-AWD + CBT network based on multi-object tracking methods can effectively obtain and analyze swimming behavior data during micropterus salmoides aquaculture. This technology is crucial for rapidly analyzing fish exhibiting potentially abnormal swimming behaviors to facilitate appropriate farming measures.

However, like most tasks, our task also has some limitations that need to be considered. For example, environmental factors can impact the tracking efficiency of our model because elements such as water waves, bubbles, and reflections generated during aquaculture cause instability and distortion in the images, as well as additional occlusions. Continuously changing lighting conditions can affect the appearance of the targets, while a cluttered background may lead to false positives. Additionally, the varying body sizes of micropterus salmoides bass at different growth stages can interfere with Yolo-AWD's detection, adversely affecting CBT's trajectory association and tracking performance. To address these challenges, we need to expand our image dataset of micropterus salmoides bass to include more growth stages, body sizes, and various complex scenes. Moreover, to mitigate the effects of environmental changes, we plan to employ more advanced data augmentation techniques. This will involve simulating different lighting conditions, weather scenarios, and background complexities during training to enhance the robustness and generalization ability of the model. Lastly, in swimming behavior detection, we are currently only estimating speed in the 2D direction, whereas in reality, micropterus salmoides move in 3D space. Next, we will explore multi-target multi-camera tracking (MTMCT) to record and analyze 3D swimming behavior, We plan to incorporate depth information from RGB-D cameras or stereo vision systems to analyze behaviors in three dimensions. For multi-target multi-camera tracking (MTMCT), we will attempt to use multiple cameras to observe from top and front angles, leveraging scene understanding and contextual information to enhance tracking accuracy. We will implement and test different cross-camera data association algorithms to achieve this goal. We hope that in the future, through these studies, we can expand the applicability of our methods and contribute to the development of more effective and robust tracking algorithms for micropterus salmoides in various environments.

4. Summary and conclusions

In summary, in this paper, we developed a CNN-based multi-object tracking technique for analyzing the swimming behavior of micropterus salmoides. The micropterus salmoides detection network is built on the YOLOv8 architecture. In target detection, we improved both detection accuracy and computational speed by incorporating the AWD module mechanism and XIOU algorithm. In the tracking method, we expanded the Kalman filtering in ByteTracker for tracking and added a linear matching strategy to associate fish trajectories. Ablation experiments validated the enhancement of these modules. Comparative analysis between the swimming behavior data obtained by our algorithm and real data indicates that the Yolo-AWD + CBT algorithm achieves rapid and accurate analysis of micropterus salmoides swimming behavior data. In the future, this research can be extended to analyze the motion behavior of other aquatic organisms and actively improve tracking under different aquatic conditions. Overall, the Yolo-AWD + CBT model efficiently acquires and analyzes swimming behavior data of micropterus salmoides, providing a highly effective and cost-efficient method for a broader analysis of fish motion behavior in the future.

Author statement

We the undersigned declare that this manuscript is original, has not been published.

before and is not currently being considered for publication elsewhere.

We confirm that the manuscript has been read and approved by all named authors and that there are no other persons who satisfied the criteria for authorship but are not listed. We further confirm that the order of authors listed in the manuscript has been approved by all of us.

We understand that the Corresponding Author is the sole contact for the Editorial process. She is responsible for communicating with the other authors about progress,submissions of revisions and final approval of proofs.

CRediT authorship contribution statement

Peng Xiao: Formal analysis, Methodology, Software, Validation, Visualization, Writing – original draft. **Ming Chen:** Validation, Writing – review & editing. **Guofu Feng:** Methodology, Formal analysis, Writing – review & editing. **Wanying Zhai:** Investigation, Validation, Data curation. **Yidan Zhao:** Validation, Software. **Yongxiang Huang:** Data curation.

Declaration of competing interest

The authors declare that we have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

This research was supported by the National Key Research and Development Program of China (2022YFD2400800).

References

Arvind, C., Prajwal, R., Bhat, P.N., Sreedevi, A., Prabhudeva, K., 2019. Fish detection and tracking in pisciculture environment using deep instance segmentation. In: TENCON 2019-2019 IEEE Region 10 Conference (TENCON). IEEE, pp. 778–783.

Bao, Y., Ji, C., Zhang, B., Gu, J., 2018. Representation of freshwater aquaculture fish behavior in low dissolved oxygen condition based on 3d computer vision. Mod. Phys. Lett. B 32, 1840090.

Barbedo, J.G.A., 2022. A review on the use of computer vision and artificial intelligence for fish recognition, monitoring, and management. Fishes 7, 335.

Bewley, A., Ge, Z., Ott, L., Ramos, F., Upcroft, B., 2016. Simple online and realtime tracking. In: 2016 IEEE International Conference on Image Processing (ICIP). IEEE, pp. 3464–3468.

Bochkovskiy, A., Wang, C.Y., Liao, H.Y.M., 2020. Yolov4: optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934. https://arxiv.org/abs/2004.10934.

Downie, A.T., Illing, B., Faria, A.M., Rummer, J.L., 2020. Swimming performance of marine fish larvae: review of a universal trait under ecological and environmental pressure. Rev. Fish Biol. Fish. 30, 93–108.
Duan, Y., Zhang, X., Liu, X., Thakur, D.N., 2014. Effect of dissolved oxygen on swimming ability and physiological response to swimming fatigue of whiteleg shrimp (litopenaeus vannamei). J. Ocean Univ. China 13, 132–140.
Ge, Z., Liu, S., Wang, F., Li, Z., Sun, J., 2021. Yolox: exceeding yolo series in 2021. arXiv preprint arXiv:2107.08430. https://arxiv.org/abs/2107.08430.
Hu, J., Zhao, D., Zhang, Y., Zhou, C., Chen, W., 2021. Real-time nondestructive fish behavior detecting in mixed polyculture system using deep-learning and low-cost devices. Expert Syst. Appl. 178, 115051.
Huang, J., Yu, X., Chen, X., An, D., Zhou, Y., Wei, Y., 2022. Recognizing fish behavior in aquaculture with graph convolutional network. Aquac. Eng. 98, 102246.
Iqbal, U., Li, D., Akhter, M., 2022. Intelligent diagnosis of fish behavior using deep learning method. Fishes 7, 201.
Jerry, D., Cairns, S., 1998. Morphological variation in the catadromous australian bass, from seven geographically distinct riverine drainages. J. Fish Biol. 52, 829–843.
Li, W., Li, F., Li, Z., 2022. Cmftnet: multiple fish tracking based on counterpoised jointnet. Comput. Electron. Agric. 198, 107018.
Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y., Berg, A.C., 2016. Ssd: single shot multibox detector. In: Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14. Springer, pp. 21–37.
Liu, Y., Li, B., Zhou, X., Li, D., Duan, Q., 2024. Fishtrack: multi-object tracking method for fish using spatiotemporal information fusion. Expert Syst. Appl. 238, 122194.
Luiten, J., Osep, A., Dendorfer, P., Torr, P., Geiger, A., Leal-Taixé, L., Leibe, B., 2021. Hota: a higher order metric for evaluating multi-object tracking. Int. J. Comput. Vis. 129, 548–578.
Luo, G., Chen, J., Wang, H., Wang, Y., et al., 2017. Application of computer vision in aquaculture. In: Animal Husbandry and Feed Science (Inner Mongolia), 38, pp. 91–92.
Måløy, H., Aamodt, A., Misimi, E., 2019. A spatio-temporal recurrent network for salmon feeding action recognition from underwater videos in aquaculture. Comput. Electron. Agric. 167, 105087.
Muñoz, L., Aspillaga, E., Palmer, M., Saraiva, J.L., Arechavala-Lopez, P., 2020. Acoustic telemetry: a tool to monitor fish swimming behavior in sea-cage aquaculture. Front. Mar. Sci. 7, 645.
Qian, Z.M., Chen, X., Jiang, H., 2023. Fish tracking based on yolo and bytetrack. In: 2023 16th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), pp. 1–5. https://doi.org/10.1109/CISP-BMEI60920.2023.10373254.
Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You only look once: unified, real-time object detection. Proc. IEEE Conf. Comput. Vis. Pattern Recognit. 779–788.
Ren, S., He, K., Girshick, R., Sun, J., 2015. Faster r-cnn: towards real-time object detection with region proposal networks. Adv. Neural Inf. Proces. Syst. 28.
Song, Z., Zhou, Z., Wang, W., Gao, F., Fu, L., Li, R., Cui, Y., 2021. Canopy segmentation and wire reconstruction for kiwifruit robotic harvesting. Comput. Electron. Agric. 181, 105933.
Stergiou, A., Poppe, R., Kalliatakis, G., 2021. Refining activation downsampling with softpool. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 10357–10366.
Tan, C., Li, C., He, D., Song, H., 2022. Towards real-time tracking and counting of seedlings with a one-stage detector and optical flow. Comput. Electron. Agric. 193, 106683.
Wang, H., Zhang, S., Zhao, S., Lu, J., Wang, Y., Li, D., Zhao, R., 2022a. Fast detection of cannibalism behavior of juvenile fish based on deep learning. Comput. Electron. Agric. 198, 107033.
Wang, H., Zhang, S., Zhao, S., Wang, Q., Li, D., Zhao, R., 2022b. Real-time detection and tracking of fish abnormal behavior based on improved yolov5 and siamrpn++. Comput. Electron. Agric. 192, 106512.
Warren-Myers, F., Svendsen, E., Føre, M., Folkedal, O., Oppedal, F., Hvas, M., 2023. Novel tag-based method for measuring tailbeat frequency and variations in amplitude in fish. Anim. Biotelem. 11, 12.
Xu, G., Liao, W., Zhang, X., Li, C., He, X., Wu, X., 2023. Haar wavelet downsampling: a simple but effective downsampling module for semantic segmentation. Pattern Recogn. 143, 109819.
Yang, X., Zhang, S., Liu, J., Gao, Q., Dong, S., Zhou, C., 2021. Deep learning for smart fish farming: applications, opportunities and challenges. Rev. Aquac. 13, 66–90.
Zhang, Y., Sun, P., Jiang, Y., Yu, D., Weng, F., Yuan, Z., Luo, P., Liu, W., Wang, X., 2022. Bytetrack: multi-object tracking by associating every detection box. In: European Conference on Computer Vision. Springer, pp. 1–21.
Zhang, X., Liu, C., Yang, D., Song, T., Ye, Y., Li, K., Song, Y., 2023. Rfaconv: innovating spatital attention and standard convolutional operation. arXiv preprint arXiv:2304.03198. https://arxiv.org/abs/2304.03198.
Zhao, J., Gu, Z., Shi, M., Lu, H., Li, J., Shen, M., Ye, Z., Zhu, S., 2016. Spatial behavioral characteristics and statistics-based kinetic energy modeling in special behaviors detection of a shoal of fish in a recirculating aquaculture system. Comput. Electron. Agric. 127, 271–280.
Zhao, H., Cui, H., Qu, K., Zhu, J., Li, H., Cui, Z., Wu, Y., 2024. A fish appetite assessment method based on improved bytetrack and spatiotemporal graph convolutional network. Biosyst. Eng. 240, 46–55.
Zhou, C., Xu, D., Chen, L., Zhang, S., Sun, C., Yang, X., Wang, Y., 2019. Evaluation of fish feeding intensity in aquaculture using a convolutional neural network and machine vision. Aquaculture 507, 457–465.