Smart Agricultural Technology 13 (2026) 101766

Contents lists available at ScienceDirect

Smart Agricultural Technology

journal homepage: www.journals.elsevier.com/smart-agricultural-technology

Bézierfusion: A unified fish detection and measurement paradigm based on tightly-coupled vision-geometry modeling

Lanlan Liang [a], Zhuhua Hu [a,b,*], Gaosheng Liu [a], Jie Liu [a], Zekun Deng [a], Shuangyin Liu [c], Yaochi Zhao [a], Ziqi Yang [a]

[a] School of Information and Communication Engineering, Hainan University, 570228, Haikou, China
[b] Hainan Yazhou Bay Seed Laboratory, 572000, Sanya, Hainan, China
[c] College of Artificial Intelligence, Zhongkai University of Agriculture and Engineering, 510225, Guangzhou, China

ARTICLE INFO

Keywords:
Fish length measurement
Bézier curve modelling
Binocular vision
Precision aquaculture

ABSTRACT

Accurate measurement of body length in free-swimming fish is critical for digital aquaculture, yet it remains challenging due to complex underwater environments, dynamic fish postures, weak texture features, and limited annotated data. To overcome these limitations, this paper introduces a structured detection paradigm that unifies object detection and parametric curve modeling in a single forward pass, enabling end-to-end differentiable mapping from pixels to physical measurements. This approach eliminates the need for traditional "detect-then-postprocess" workflows and offers a robust solution for measuring non-rigid objects with regular geometric shapes. To address feature extraction difficulties caused by body curvature and texture degradation, we propose a visually-geometrically coupled model. At the visual level, an Efficient Adaptive Multi-scale Receptive Field (EAMRF) module is designed to integrate RGB appearance with geometric depth cues through cross-spatial learning, enhancing feature discriminability under weak textures. At the geometric level, a sampling-endpoint composite loss function is introduced to ensure accurate global curve fitting while significantly improving the localization precision of key points such as the head and tail. Furthermore, we contribute two benchmark datasets comprising over 5000 annotated samples to mitigate data scarcity. Extensive experiments demonstrate that the proposed method achieves mAP50 values of 95.1% and 98.4% on Dataset 1 and Dataset 2, respectively, average absolute relative length measurement errors of 0.154 and 0.168, and an inference speed approximately three times faster than two-stage methods.

1. Introduction

To address the core requirement of accurately assessing fish growth in smart aquaculture, real-time measurement of fish body length is crucial [1]. In response to challenges such as deformation, occlusion, and complex underwater environments faced by traditional methods in measuring body length, binocular vision technology offers an ideal solution. This technique acquires three-dimensional data of fish and, combined with deep learning, provides accurate and interpretable length information for intelligent aquaculture, thereby enabling non-contact, rapid, and precise measurement [2].

Traditional contact measurements (pre-1990s) suffered from manual error and harmed fish [3,4]. Later 2D deep learning methods required simple backgrounds and struggled with real-world occlusion and deformation [5–8]. Now, 3D stereo vision addresses these by enabling non-contact 3D model-based measurement [9], yet practical challenges remain for its aquaculture use, as outlined in the research roadmap (Fig. 1).

1.1. Motivation and contributions

Current research on 3D fish body measurement based on deep learning has yielded notable progress. Studies [10,11] employed improved segmentation networks to extract fish masks for body-length estimation. However, these mask-based approaches often rely on relatively fixed postures and exhibit limited robustness in complex underwater scenes. To address contour fitting, [12] designed a deformable geometric model tailored for bluefin tuna. Nevertheless, such parametric models gener-

* Corresponding author.
E-mail addresses: lanlanliang@hainanu.edu.cn (L. Liang), eagler_hu@hainanu.edu.cn (Z. Hu), lgs@hainanu.edu.cn (G. Liu), 3400450560@qq.com (J. Liu), 23220854100038@hainanu.edu.cn (Z. Deng), shuangyinliu@126.com (S. Liu), zhyc@hainanu.edu.cn (Y. Zhao), 20223004596@hainanu.edu.cn (Z. Yang).

https://doi.org/10.1016/j.atech.2025.101766
Received 27 November 2025; Received in revised form 26 December 2025; Accepted 28 December 2025
Available online 5 January 2026

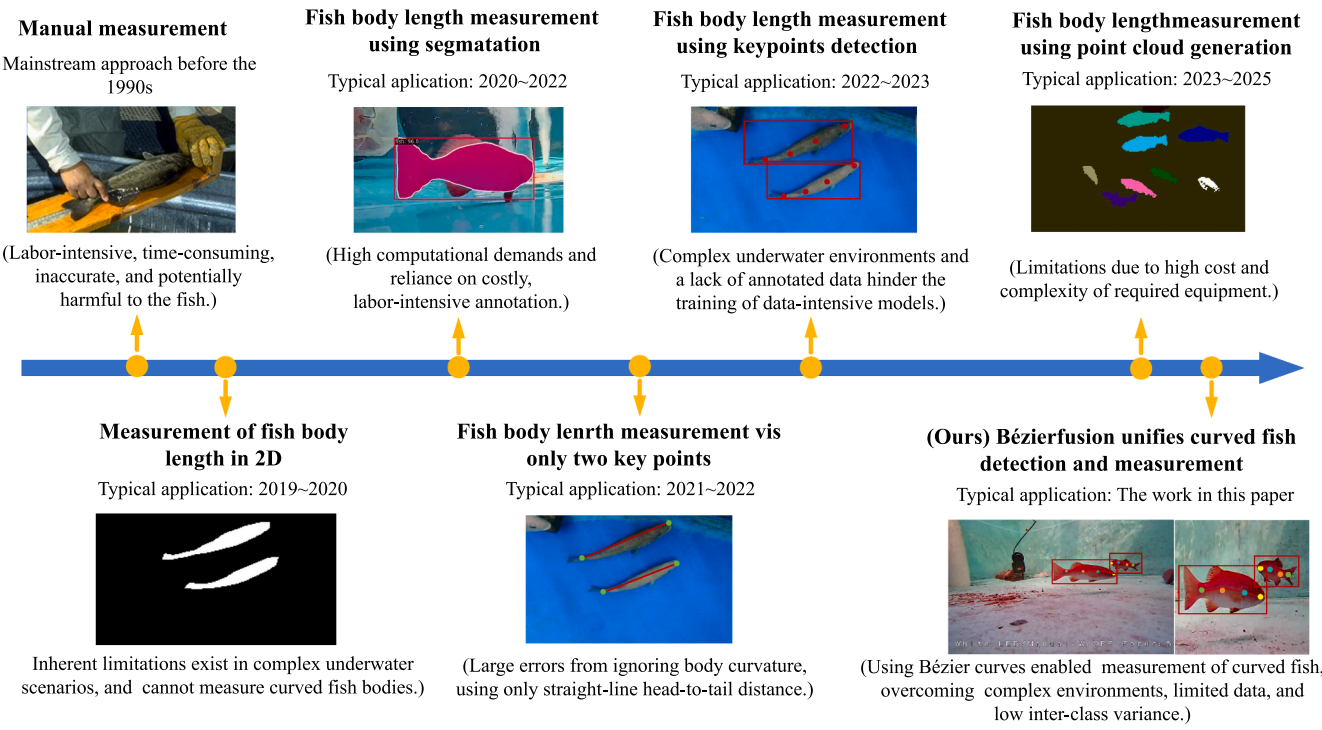

Fig. 1. The research roadmap and the location of our work.

alize poorly across species and environments. Overall, contour-based methods tend to involve high computational cost and strong posture dependence, making them difficult to deploy in real-world aquaculture settings.

In response, recent work has explored end-to-end multimodal models [13,14] and keypoint-based strategies for fish-length estimation. For instance, [15] achieved head-tail detection in commercial aquaculture cages using stereo vision, while [16] employed an improved Keypoint R-CNN in laboratory conditions. Although efficient for regularly posed fish, two-point methods cannot correct errors caused by body bending. Multi-keypoint schemes improve accuracy through posture compensation, but remain sensitive to noise, occlusion, and visual degradation in underwater environments.

In summary, binocular-vision-based underwater measurement faces three core challenges: (1) diverse fish poses and non-rigid body deformations in complex underwater environments pose difficulties for accurate measurement; (2) weak textures and indistinct visual features hinder reliable feature extraction and matching; (3) scarce public binocular datasets limit model training and generalization.

To overcome these limitations, we propose an integrated framework that combines geometric modeling, deep learning, and binocular vision. In particular, the classical cubic Bézier curve is adopted as a compact geometric representation, as it strikes a favorable balance between modeling expressiveness and computational efficiency, it provides sufficient degrees of freedom to capture typical fish body geometry while maintaining low computational complexity and stable optimization behavior, making it well suited for unified and optimized modeling within an end-to-end network. Building on this representation, the cubic Bézier curve is embedded directly into the detector to parameterize fish-body curvature, enabling simultaneous detection and curve extraction with reduced reliance on fine-grained keypoints. Furthermore, By incorporating binocular depth cues, the proposed framework achieves robust performance under degraded underwater visibility and supports accurate fish-length measurement in real aquaculture scenarios.

The primary contributions of this work are as follows.

1. We construct two challenging binocular datasets covering indoor cages and outdoor ponds. To reduce annotation effort, a Bézier-curve-based scheme replaces dense masks with four control points while preserving curved-body geometry.
2. We develop Bézierfusion, an end-to-end detector that embeds a cubic Bézier representation into the YOLOv8 regression head. This integration enables a single forward pass to complete both detection and curve fitting, improving consistency and eliminating fragile post-processing.
3. To handle degraded underwater imagery, we introduce the EAMRF module. It adopts early-stage RGB channel fusion, multi-scale grouped convolution, and lightweight cross-space attention to enhance discriminative features while maintaining low computational overhead.
4. We propose a composite loss that combines Gaussian sampling along the predicted curve with endpoint supervision, balancing global shape modeling and keypoint accuracy. Leveraging stereo depth, 3D curve arc-length integration yields high-precision measurements suitable for real-time aquaculture deployment.

1.2. Paper organization

The remainder of this paper is organized as follows. Section 2 reviews related work. Section 3 describes data acquisition and preprocessing. Section 4 presents our integrated approach for fish detection, modeling, and measurement. Section 5 details the experimental setup and results. Section 6 discusses the findings and future directions. Section 7 concludes the paper.

2. Related work

2.1. Semi-automated fish body length measurement with monocular images

Deep learning-based fish body measurement achieves length prediction through automated identification and feature extraction. Early stud-

ies, such as [17] manually annotated head and tail keypoints via an interactive interface and calculated length using stereoscopic vision. [18] and [19] employed direct linear transformation to estimate the three-dimensional positions of manually marked keypoints. To improve efficiency, [20] and [21] introduced pattern recognition and template matching algorithms, which required only processing of fish bodies in a single frame. This enabled automatic matching of stereo image pairs and subsequent frames through inter-frame differences and epipolar constraints. With technological advancement, [22] proposed an automatic detection method based on a level set framework, further reducing the need for manual intervention.

However, due to limitations in labor dependency and low efficiency, semi-automatic methods are unsuitable for large-scale application, making fully automated measurement algorithms a key focus of current research.

2.2. Automated fish body length measurement with stereo vision

Automated fish body measurement using stereo vision achieves morphological analysis through 3D information fusion. Early research primarily focused on geometric modeling of contours. For instance, Elliptic Fourier analysis was applied for contour fitting [23], while an enhanced geometric model was constructed based on head and tail keypoints [24]. With technological advancement, methodologies shifted from traditional image processing to multi-view systems, such as orthogonal binocular vision for multi-angle observation [25] and 3D coordinate computation for improved measurement consistency [26].

In recent years, deep learning has propelled the field forward. A stereo matching method based on convolutional neural networks was proposed in [27], which accomplishes 3D reconstruction and size measurement of freely swimming fish via instance segmentation and video interpolation. Although such methods have not yet demonstrated significant advantages in measurement accuracy, they exhibit considerable potential in environmental adaptability and automation.

2.3. CNN-based automated measurement of fish body length

In deep learning-based fish length measurement, contour segmentation and keypoint detection represent two predominant methodologies. Contour segmentation employs networks like Mask R-CNN to extract contours integrated with stereo vision information. For instance, [28] established a measurement scheme based on 3D coordinate projection, while [29] achieved length estimation through morphological processing to obtain the fish body's principal axis. Keypoint detection locates feature points combined with stereo vision techniques. The Voskakis team realized fish measurement via multi-keypoint detection [30], and [31] adopted fully convolutional networks for target region segmentation, completing depth estimation and size calculation with semi-global block matching algorithms.

For accurate measurement of curved fish bodies, methodologies have evolved from bounding box detection to parametric modeling. However, keypoint detection is prone to localization deviations under environmental interference, while traditional parametric models exhibit limited adaptability to curve variations. These factors collectively constrain measurement accuracy in practical applications.

3. Dataset preparation

3.1. Data collection

The grouper samples (Plectropomus leopardus) used in this study were collected from the scientific aquaculture base of the College of Marine Sciences, Hainan University, located in Wenchang, Hainan. To enhance image acquisition efficiency, we developed a specialized underwater binocular measurement system (Fig. 2). The system simultaneously captures top and side views of fish bodies using underwater

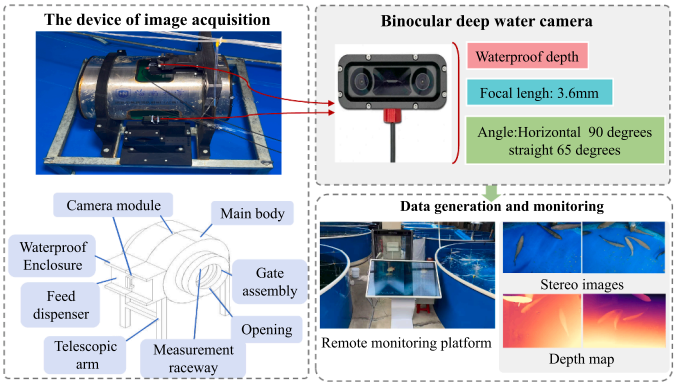

Fig. 2. Underwater binocular measurement device.

binocular cameras mounted above and beside the measurement channel. Connected to a computer via data cables, the device can be positioned at optimal shooting locations in aquaculture tanks for data collection.

3.2. Dataset description and preprocessing

Two image datasets were constructed for evaluation. Dataset 1 contains 725 RGB images (1920×1080) from an indoor grouper farm, showing blue tank backgrounds with 2036 annotated fish. Dataset 2 includes 2012 outdoor farm images with 5652 annotations, presenting challenges like weak textures, uneven lighting, and suspended particles. The datasets were divided into training/validation and test sets (Dataset 1: 580/145; Dataset 2: 1610/402), with Dataset 2 exhibiting greater complexity. Statistics are in Table 1, with examples in Fig. 3.

Data augmentation was applied to improve generalization and prevent overfitting. We used geometric transformations (random scaling and translation) to handle size and position variations. Photometric adjustments (contrast changes, Gaussian noise) simulated different lighting conditions. Blur processing mimicked water turbidity effects, collectively enhancing model robustness.

3.3. Annotation for curve fish body

For efficient annotation, we propose a Bézier curve-based strategy for fish body labeling. This approach utilizes bounding boxes combined with sparse keypoints to define fish contours, effectively replacing traditional pixel-wise masks. The method provides significant cost benefits in annotation while maintaining complete characterization of curved body shapes through compact parametric representation. Although the number of keypoints is adjustable, we employ a four-point configuration based on comprehensive analysis of fish pose diversity, as further elaborated in Section 4.1.

4. Methodology

As illustrated in Fig. 4, our proposed fish body length measurement method comprises four key stages. Initially, a custom underwater binocular imaging system captures stereo images, from which fish bodies are detected in the 2D image plane. Subsequent curve fitting yields continuous 2D pixel coordinates of each fish body. Stereo matching then generates pixel-level depth information, enabling the transformation of 2D coordinates into 3D space. Finally, body length measurement is formulated and solved as a 3D curve integration problem.

4.1. Bézier curves and fish body modeling

The dynamic bending of freely swimming fish bodies poses a challenge to modeling accuracy. To strike a balance between precision and

Table 1
Detailed information of Dataset 1 and 2. N_I represents the number of images, N_F represents the number of fish.

Dataset	Train/Validation		Test		Range of fish length (cm)	Type of scene	Resolution
	N_I	N_F	N_I	N_F			
1	580	1746	145	290	4.1–9.8	Aquaculture pond	1920×1080
2	1610	4845	402	807	13.5–18.5	Complex underwater environment	1920×1080

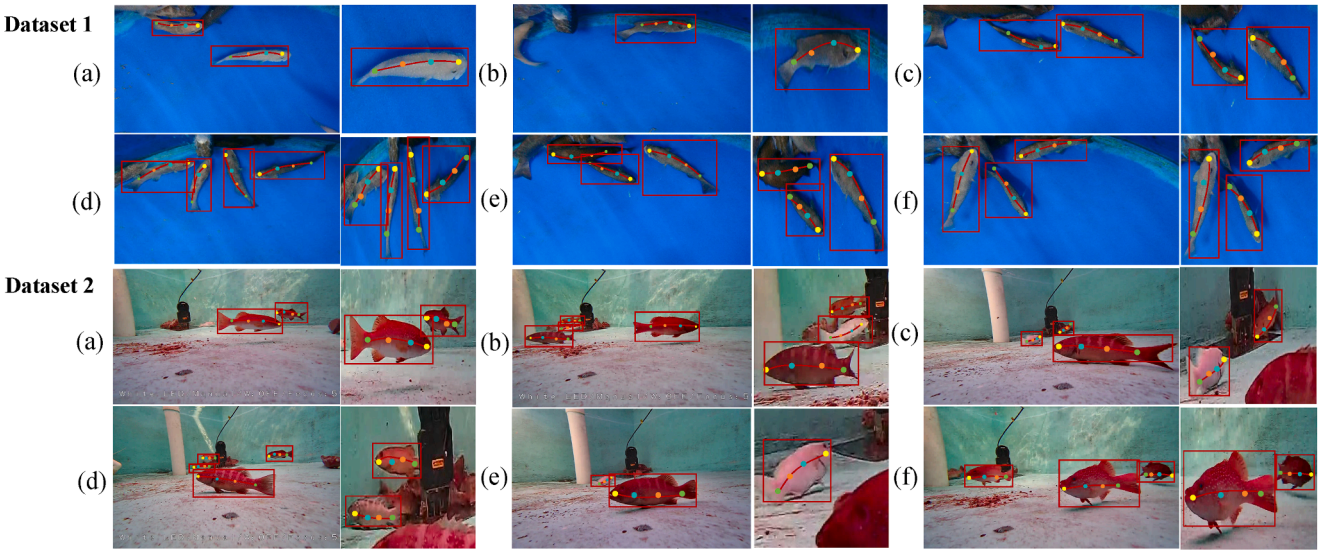

Fig. 3. Datasets examples in the underwater environment.

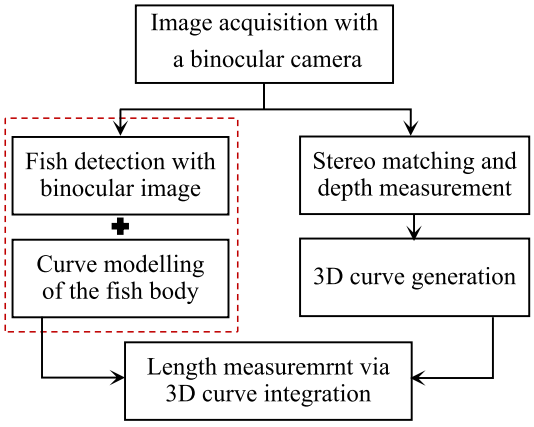

Fig. 4. The flow chart of the measurement method.

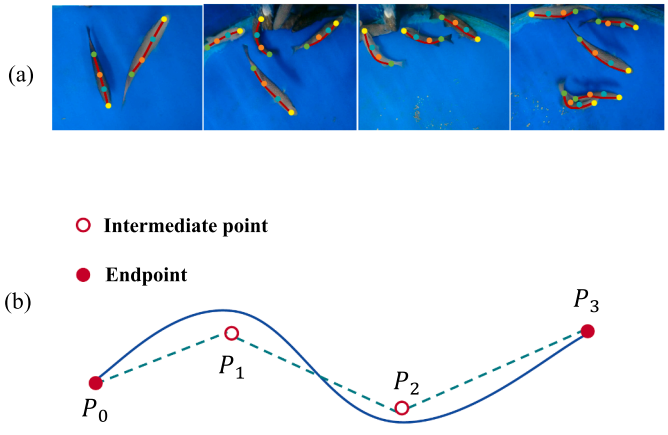

Fig. 5. Third-order Bézier curve and its application in fish body modeling. (a) Using third-order Bézier curves to model fish body in different postures. (b) Formula of third-order Bézier curve and visualization curve with control points. Hollow points are control points used to control curve morphology, solid points are curve endpoints whose length is directly determined by the endpoints.

efficiency, we employ computationally efficient Bézier curves. Specifically, the classical cubic Bézier curve requires only four control points to flexibly fit various fish body curvatures, as shown in Fig. 5(b), and is mathematically defined in Eq. (1).

$$B(t) = (1-t)^3P_0 + 3(1-t)^2tP_1 + 3(1-t)t^2P_2 + t^3P_3, \quad t \in [0,1] \tag{1}$$

where P_1 and P_2 are the two intermediate control points in the middle of the fish body. By controlling the spatial coordinates of these two points, the curved of the fish body can be precisely fitted.

4.2. Curved fish body detection network

Underwater fish detection remains challenging due to the inherent complexity of aquatic environments, where fish exhibit highly variable poses and non-rigid deformations, compounded by weak texture information. These factors severely limit the applicability of traditional keypoint detection and curve modeling approaches.

To address these challenges, we propose a joint framework that embeds Bézier curve modeling into the YOLOv8 detector for simultaneous fish detection and body curve fitting. As illustrated in Fig. 6(a), the framework extends the original YOLOv8 architecture by augmenting the regression branch with eight Bézier control point coordinates, while preserving the classification branch for fish presence detection.

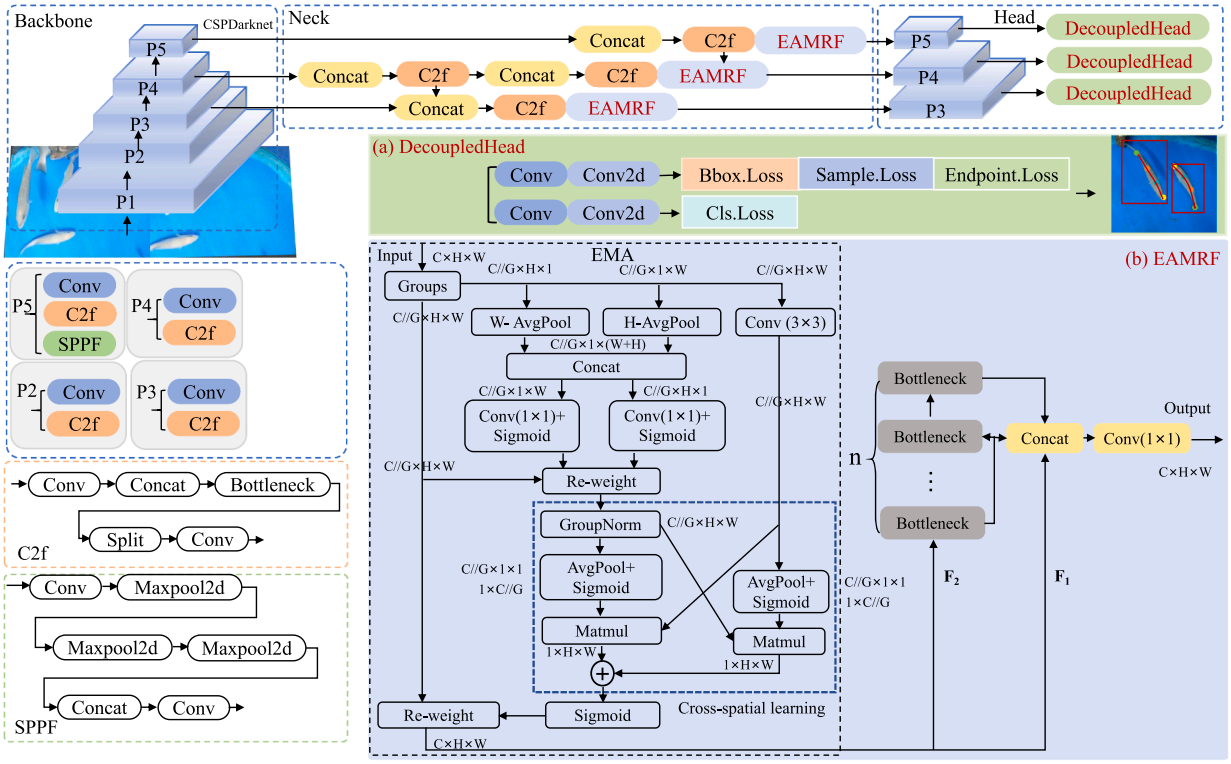

Fig. 6. A joint framework for fish detection and body curve modeling.

The backbone network employs a weight-sharing mechanism to support both tasks in an end-to-end manner.

The network architecture comprises three core components: the backbone network, the neck detection layer, and the head detector (Fig. 6). Building upon YOLOv8's anchor-free strategy and multi-scale decoupled head design, our approach integrates Bézier control point prediction into the regression branch, enabling unified detection and curve modeling.

To enhance feature representation in complex underwater environments, we further introduce the EAMRF module, which is embedded into the neck structure. This module strengthens cross-channel feature interactions and explicitly models cross-spatial dependencies, improving performance for both object detection and curve prediction. The detailed architecture and implementation of the EAMRF module are presented in Section 4.3.

4.3. EAMRF

Multi-scale features play a critical and complementary role in visual understanding: fine-scale features (e.g., edges and textures) capture local details, while coarse-scale features (e.g., object shape and spatial layout) provide global semantic context. However, the conventional C2f module, which relies primarily on local convolution operations, lacks adaptive attention mechanisms to emphasize spatially salient regions. This limitation becomes particularly pronounced in underwater fish detection, where complex spatial structures arise from varying poses, scale changes, and occlusions.

The challenge is further compounded by significant scale and spatial distribution differences among fish body parts (e.g., head, body, tail). Feature representations based on a single scale or fixed fusion strategy prove inadequate for simultaneously capturing fine-grained local keypoints and coarse-grained global structural information. To address these limitations, we propose the EAMRF module for adaptive multi-scale feature fusion.

As illustrated in Fig. 6(b), EAMRF builds upon the EMA (Efficient Multi-scale Attention) mechanism [32], which encodes spatial information through direction-decoupled one-dimensional global average pooling along height and width dimensions. This design effectively captures long-range spatial dependencies while maintaining low computational cost.

EAMRF extends EMA through a dual-branch architecture: a direct connection branch preserves original semantic information, while a recursive Bottleneck branch progressively refines features through multiple stages to model multi-scale relations. This hierarchical design enables feature propagation and cross-spatial dependency modeling across different receptive field scales. Consequently, EAMRF enhances the model's ability to perceive complex fish structures and critical regions without substantial parameter overhead, yielding notable performance gains in both detection and curve prediction tasks.

Given an input feature map $\mathbf{X} \in \mathbb{R}^{C\times H\times W}$ (where C, H, W denote channels, height, and width, respectively), obtained by concatenating multi-scale RGB features $\{\mathbf{P}_1, \mathbf{P}_2, \ldots, \mathbf{P}_5\}$ from the backbone after spatial alignment, the processing proceeds as follows.

The feature representation is enhanced through the EMA attention operator, as formulated in Eq. (2).

$$\mathbf{X}' = \mathcal{E}_{\text{EMA}}(\mathbf{X}) \tag{2}$$

where $\mathcal{E}_{\text{EMA}}(\cdot)$ incorporates two key operations:

- Channel grouping divides the C channels into G groups (C/G channels per group), balancing representational diversity and computational efficiency-particularly important for high-dimensional underwater features with limited training data.
- Direction-aware pooling applies separate 1D global average pooling along height and width dimensions to encode spatial dependencies in both directions, enabling the model to capture elongated fish body structures and cross-spatial correlations.

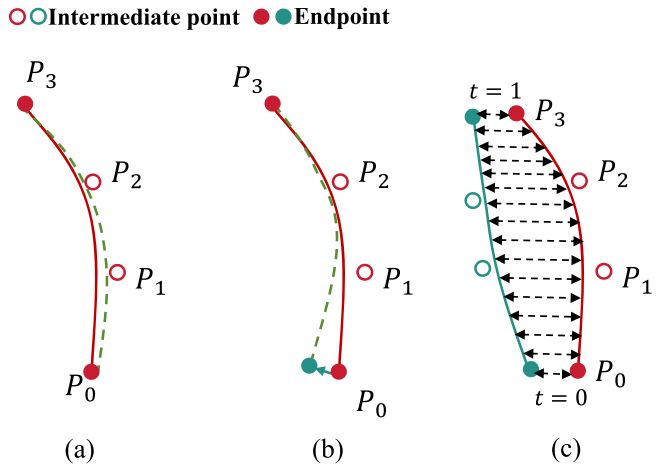

Fig. 7. Sampling loss on fish body curves. (a) Bézier curves are less affected by slight deviations of intermediate points. (b) Even minor deviations of Bézier curve endpoints (terminal control points) can cause significant positional deviations of the entire curve. (c) Sampling loss used to measure curve deviations.

The enhanced feature map $\mathbf{X}'$ is decomposed along the channel dimension, as formulated in Eq. (3):

$$\mathbf{F}_1, \mathbf{F}_2 = \text{Split}(\mathbf{X}'), \quad \mathbf{F}_1, \mathbf{F}_2 \in \mathbb{R}^{\frac{C}{2} \times H \times W} \tag{3}$$

where $\mathbf{F}_1$ is preserved through a direct connection branch to retain attention-enhanced semantic information, while $\mathbf{F}_2$ is fed into a recursive Bottleneck branch for multi-scale relation modeling.

The recursive branch progressively refines $\mathbf{F}_2$ through n stages, as formulated in Eq. (4):

$$\mathbf{F}_2^{(i)} = \mathcal{B}\left(\mathbf{F}_2^{(i-1)}\right), \quad i = 1, 2, \ldots, n \tag{4}$$

where the Bottleneck operation $\mathcal{B}(\cdot)$ is defined as Eq. (5).

$$\mathcal{B}(\mathbf{x}) = \mathbf{x} + \text{Conv}(\text{Conv}(\mathbf{x})) \tag{5}$$

with $\mathbf{F}_2^{(0)} = \mathbf{F}_2$. Each stage consists of two consecutive 3×3 convolutions followed by a residual connection, progressively expanding the receptive field. This recursive design enables deeper stages to capture coarser-scale patterns (e.g., overall body layout) while earlier stages focus on finer details (e.g., keypoint localization).

The outputs from the direct branch and all recursive stages are concatenated and projected, as formulated in Eq. (6):

$$\mathbf{Y} = \text{Conv}\left(\text{Cat}\left(\mathbf{F}_1, \mathbf{F}_2^{(1)}, \mathbf{F}_2^{(2)}, \ldots, \mathbf{F}_2^{(n)}\right)\right), \quad \mathbf{Y} \in \mathbb{R}^{C \times H \times W} \tag{6}$$

The concatenation aggregates features from multiple receptive field scales, while the 1×1 convolution fuses multi-scale information and restores the original channel dimension. This final output $\mathbf{Y}$ encodes both fine-grained local details and coarse-grained global context, making it well-suited for fish detection and keypoint regression in challenging underwater environments.

4.3.1. Sampling loss

As shown in Fig. 7, the sampling loss quantifies the average morphological deviation between the predicted Bézier curve and the ground-truth curve. Specifically, we adopt a Gaussian sampling strategy in which N points are sampled along the predicted curve, and the shortest Euclidean distance from each sampled point to the ground-truth curve is computed; the sampling loss is defined as the mean of these distances. To maintain geometric consistency with the cubic Bézier representation, we set $N = 4$, corresponding to the four control points that uniquely determine a cubic Bézier curve, as formulated in Eq. (7).

$$L_{\text{sampling}} = \frac{1}{N} \sum_{i=1}^{N} d_i \tag{7}$$

where d_i denotes the minimum Euclidean distance from the i-th sampled point to the ground-truth curve.

4.3.2. Endpoint loss

The endpoint loss is designed to explicitly constrain the positional accuracy of the Bézier curve's terminal points. Given that endpoint coordinates are directly utilized in subsequent arc-length computation, even minor positional errors can propagate and result in substantial measurement distortions, as illustrated in Fig. 7(b). To mitigate this issue, we formulate the endpoint loss as the summation of Euclidean distances between predicted and ground-truth endpoint coordinates, thereby imposing explicit penalties on positional deviations at these critical locations. The endpoint loss is mathematically expressed as Eq. (8)

$$L_{\text{endpoint}} = \|P_0^{\text{pred}} - P_0^{\text{gt}}\|_2 + \|P_3^{\text{pred}} - P_3^{\text{gt}}\|_2 \tag{8}$$

where P_0^{pred} and P_3^{pred} denote the predicted coordinates of the initial and terminal control points of the cubic Bézier curve, respectively; P_0^{gt} and P_3^{gt} represent their corresponding ground-truth annotations; and $\|\cdot\|_2$ denotes the L_2 norm (Euclidean distance).

4.3.3. Overall loss function

The overall training objective integrates both object detection and Bézier curve modeling components, which is formulated as a weighted combination in Eq. (9):

$$L_{\text{total}} = \lambda_1 L_{\text{detection}} + \lambda_2 L_{\text{sampling}} + \lambda_3 L_{\text{endpoint}} \tag{9}$$

Where $\{\lambda_1, \lambda_2, \lambda_3\}$ denote the weighting coefficients for each loss term. In the experiments, these values were set to 1, 1, and 0.1 respectively.

4.4. Three-dimensional length measurement via curve integration

Traditional fish length measurement methods rely on 2D projections or simplified geometric approximations, which fail to capture body curvature and depth variations. We propose a 3D curve integration approach that combines predicted 2D Bézier curves with stereo-vision-based depth reconstruction, enabling accurate measurement of bent fish bodies in underwater environments.

4.4.1. 3D parametric curve and arc length computation

The fish body is represented as a parametric curve in 3D Cartesian space, as shown in Eq. (10):

$$\mathbf{P}(t) = \big(x(t), y(t), z(t)\big), \quad t \in [0, 1] \tag{10}$$

where $x(t)$ and $y(t)$ are derived from the predicted 2D Bézierfusion model, and $z(t)$ is obtained from stereo depth reconstruction. The true body length L is computed as the arc length integral, with the mathematical expression shown in Eq. (11):

$$L = \int_0^1 \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2}\, dt \tag{11}$$

This formulation naturally accounts for both in-plane curvature and out-of-plane depth variations, providing geometrically accurate measurements regardless of fish posture or camera viewpoint.

4.4.2. Stereo vision and 3D reconstruction pipeline

Our system employs the TM02 waterproof stereo camera (ROROVSETTER Inc.) for passive depth sensing. The reconstruction pipeline establishes stereo correspondence through epipolar geometry, computes disparity maps via semi-global block matching (SGBM), and recovers metric depth through triangulation with calibrated camera parameters.

Table 2
System environment configuration.

Configuration items	Parameters
Operating System	Ubuntu 20.04
Processor	Intel(R) Xeon(R) Gold 6132 CPU @ 2.60GHz
Video Cards	Tesla V100-PCIE-16GB
Development Environment	Pytorch 2.0.1
CUDA Version	11.7
Cudnn Version	8.5.0
Python Version	Python 3.8

Prior to deployment, underwater recalibration compensates for refractive index changes at water-glass-air interfaces, ensuring stable performance across varying visibility and turbidity conditions.

The measurement workflow proceeds as follows: the network predicts 2D Bézierfusion control points $\{P_i^{2D}\}_{i=0}^{3}$ in stereo image pairs, stereo matching yields depth coordinates z_i for each control point, and these are unprojected to 3D space using camera intrinsics to obtain $\{P_i^{3D}\}_{i=0}^{3}$. A cubic Bézierfusion curve is constructed in 3D, as defined in Eq. (12).

$$\mathbf{P}(t) = \sum_{i=0}^{3} B_i^3(t) P_i^{3D}, \quad B_i^3(t) = \binom{3}{i} t^i (1-t)^{3-i} \tag{12}$$

The arc length is then numerically evaluated using adaptive Gaussian quadrature. This approach preserves geometric fidelity without linearization, handles arbitrary 3D orientations, and provides measurement accuracy invariant to camera distance within the operational range.

5. Experiments and analysis

5.1. Experimental environment configuration

To ensure the reproducibility of our experiments, all were conducted under a unified hardware and software configuration, detailed in Table 2. The model was trained for 500 epochs on both datasets with a fixed batch size of 8. We employed the AdamW optimizer, configured with an initial learning rate of 1e-3 and a weight decay of 5e-4. A cosine annealing strategy was adopted to adjust the learning rate, preventing the model from converging to local optima.

5.2. Evaluation metrics

We assess the proposed model using three groups of metrics: detection accuracy, keypoint localization accuracy, and length measurement accuracy. Detection performance is evaluated using Average Precision (AP), computed as the area under the Precision–Recall curve. Keypoint localization accuracy is measured using the Percentage of Correct Keypoints (PCK), where a prediction is regarded as correct if $\frac{d_{pi}}{d_p^{ref}} \leq T_k$, with d_{pi} denoting the keypoint error and d_p^{ref} the normalization factor. For length estimation, we report the Mean Absolute Relative Error (MARE), which is defined as $\text{MARE} = \frac{1}{N}\sum_{i=1}^{N} \frac{|L_m^{(i)} - L_g^{(i)}|}{L_g^{(i)}}$, where $L_m^{(i)}$ and $L_g^{(i)}$ represent the measured and ground-truth lengths for the i-th sample, respectively, and N is the total number of samples.

5.3. Comparison experiments

To comprehensively evaluate the model's performance, we selected YOLOv8 as an internal baseline and introduced two two-stage frameworks, Faster R-CNN + RTMpose and RTMDet + RTMpose, as external comparators. These methods were chosen to represent distinct performance paradigms, with the advanced RTMpose model ensuring efficient pose estimation collaboration. All experiments were conducted on the OpenMMLab [33] platform, using a unified dataset split for training and testing. The overall performance comparison is summarized in Table 3.

5.3.1. Selection of the curve modeling method

To better determine the optimal geometric representation for curved fish bodies, we systematically compared B-splines, piecewise linear curves, and Bézier curves. As shown in Table 4, the Bézier curve achieved optimal performance across all metrics. B-splines involve complex computations and offer negligible advantages under four-point constraints, while the piecewise linear method struggles to guarantee smoothness. Therefore, owing to its endpoint interpolation properties, low computational overhead, and high fitting accuracy, the Bézier curve was selected as the optimal approach and employed in all subsequent experiments.

5.3.2. Comparative evaluation of the proposed method

We further conducted quantitative and qualitative evaluations on the test sets of Dataset 1 and Dataset 2 to comprehensively verify the effectiveness of the proposed method, with the results detailed in Table 5 and Fig. 8. The experiments indicate that our method significantly outperforms the two-stage approaches in the core task of detecting curved fish bodies. This is primarily attributed to its joint optimization mechanism, which effectively avoids the mismatch between fish body detection and curved fish body prediction. More importantly, its inference speed is three times faster than that of the two-stage approaches (Table 3), substantially enhancing real-time performance. Naturally, the model still fails under extreme conditions, such as severely tilted (Fig. 8(a)) or partially occluded (Fig. 8(c)) fish bodies, which will be discussed in depth in Section 5.5.

5.4. Ablation experiments

To systematically validate the effectiveness of each innovative module, we conducted detailed ablation studies, with the results presented in Table 6 and Fig. 9.

Upon removal of the EAMRF module, the mAP50 for fish body detection on Dataset 1 dropped significantly from 93.8% to 85.6%, and the mAP50 for curved fish body detection decreased from 90.1% to 86.6%; a similar trend was observed on Dataset 2. Furthermore, removing the sampling loss caused the mAP50-95 for curved fish body detection on Dataset 1 to fall from 81.8% to 75.7% and the PCK to drop from 91.3% to 89.7%, with an even more pronounced performance decline on Dataset 2.

The endpoint loss is designed to specifically optimize the prediction accuracy of curve endpoints. Its removal led to a significant decrease in curve positioning accuracy, with the PCK for curved fish body detection on Dataset 1 dropping from 91.3% to 86.4%. This loss function is compatible with existing loss functions, enabling specialized optimization of endpoint prediction while maintaining the performance of the primary detection task.

5.5. Experiments on fish length measurement

Fig. 10 presents the measurement error distribution of the proposed method on both datasets. As illustrated in Fig. 10 (a-b) for Dataset 1 and (c-d) for Dataset 2, the relative errors approximately follow a normal distribution with means close to zero (Dataset 1: $\mu_1 = 0.031, \sigma_1 = 0.192$; Dataset 2: $\mu_2 = 0.023, \sigma_2 = 0.184$), confirming the absence of systematic bias. The MARE are 0.154 and 0.168, respectively. Notably, 87.6% and 84.0% of samples from Dataset 1 and Dataset 2, respectively, exhibit errors below 0.30, which validates the high accuracy and robustness of the method. The cumulative distribution curves (Fig. 10 (b) and (d) further demonstrate that the proposed method achieves excellent error concentration.

Although the proposed Bézierfusion framework demonstrates strong robustness across diverse underwater scenarios, performance degradation is observed under certain extreme conditions, as illustrated in Fig. 11 and systematically categorized in Table 7. The dominant error source (45.3%) arises from depth quality degradation, where

Table 3
Performance comparison of the proposed method with benchmark models.

Models	Backbone	Time(ms)	Total Loss	Params(M)	FLOPs(B)	FPS	Ref
YOLOv8	CSPDarknet	112.3	1.37	45.8	9.1	110	[34]
FasterRCNN + RTMpose	ResNet50	102.4	0.74	41.3	213.0	11	[35]
	CSPNeXt	103.5	0.74	26.7	4.1	11	[36]
RTMdet + RTMpose	CSPNeXt	149.3	1.49	54.2	79.9	13	[37]
	CSPNeXt	161.1	1.73	29.5	4.1	13	[38]
YOLOv8-seg + Skeleton Extraction	CSPDarknet	124.7	1.23	11.8	21.4	73	[39]
Mask RCNN + Skeleton Extraction	ResNet50	134.0	1.08	44.2	260	7.5	[40]
Ours	**CSPDarknet + EAMRF**	**103.4**	**0.69**	**26.6**	**82.9**	**34**	–

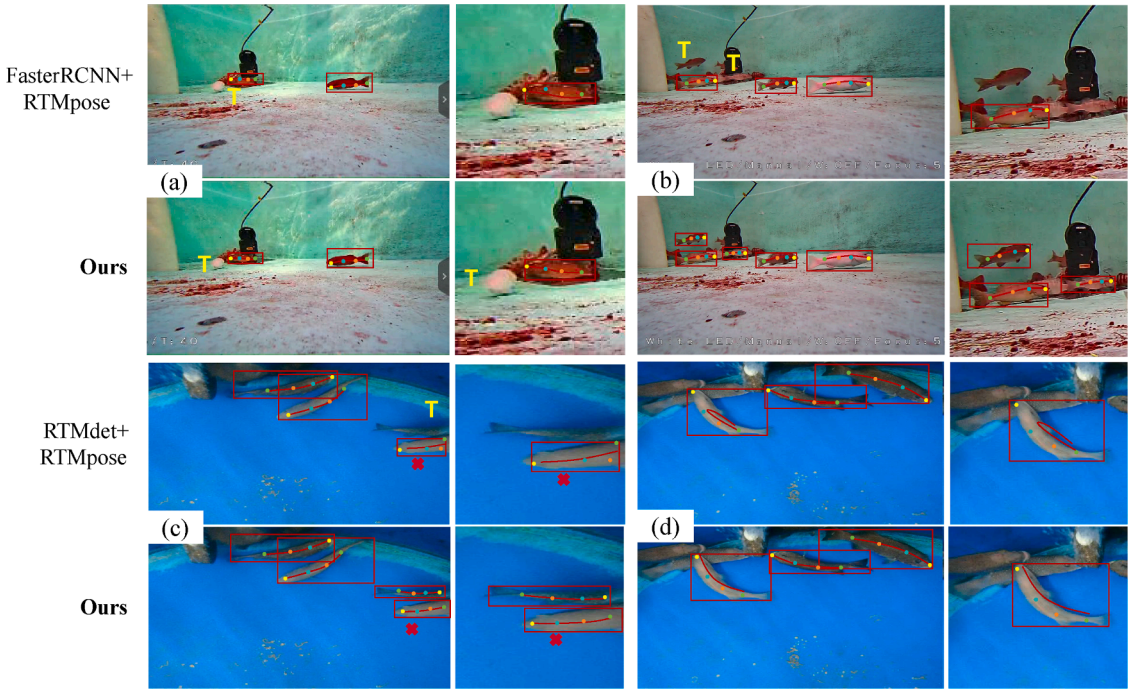

Fig. 8. Visual comparison of our method with two two-stage methods (FasterRCNN + RTMpose and RTMdet + RTMpose) on the test sets of Datasets 1 and 2. Yellow "T"s mark missed detections, and red "×"s mark recognition errors. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

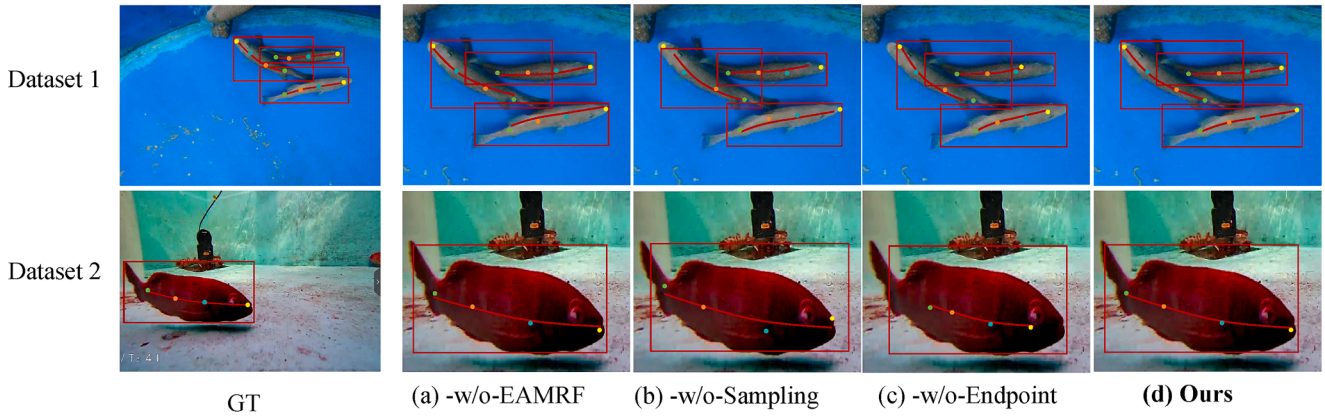

Fig. 9. Ablation study example results for key modules to enhance curve feature perception and learning, including sampling loss, EAMRF module, and endpoint loss. Upper and lower images are from Dataset 1 and Dataset 2 respectively.

Table 4
Performance comparison of different curve modeling methods.

Curve Type	Control Points	Computational Complexity	Fitting Accuracy (IoU)	Training Convergence Rounds
B-spline (3rd order)	4	O(n)	0.823	450
Polyline	4	O(n)	0.756	280
Bézier (3rd order)	4	O(n)	0.861	320

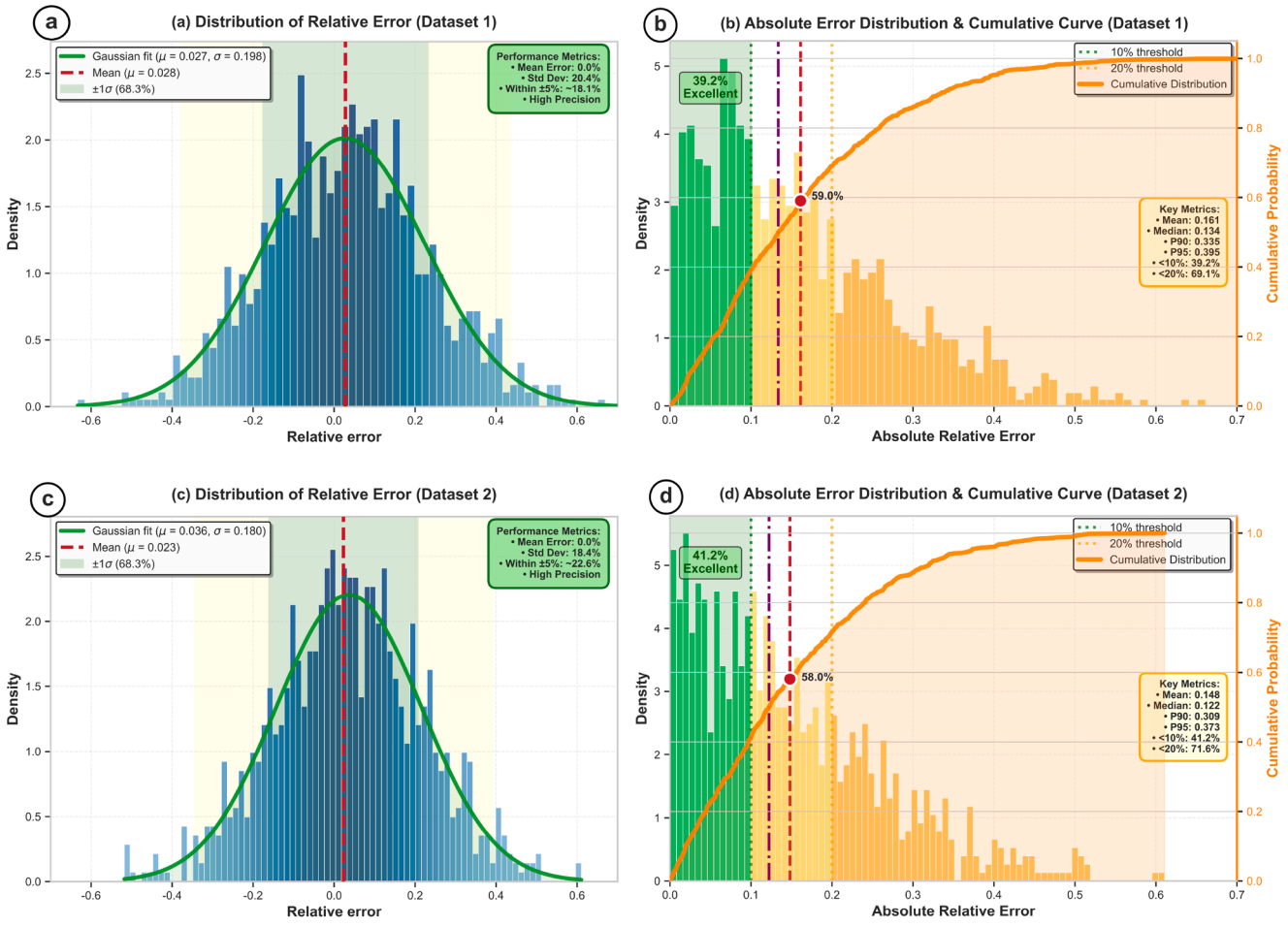

Fig. 10. Error distribution analysis of the proposed method.

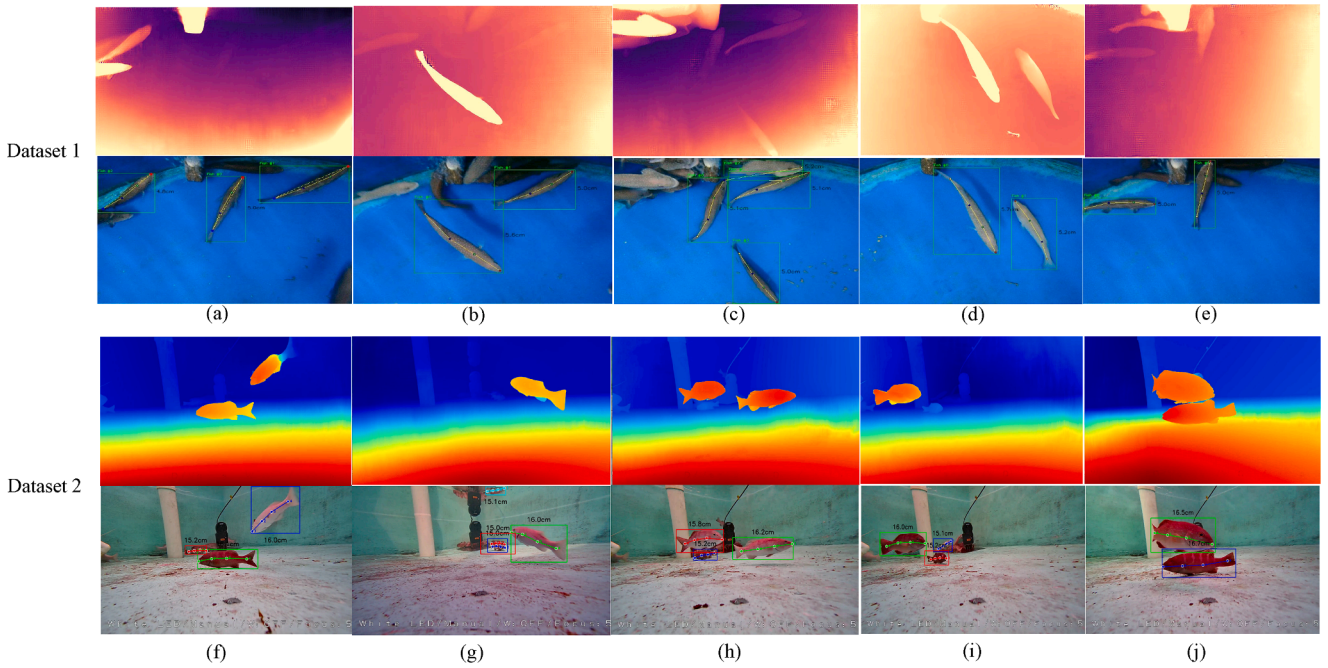

Fig. 11. Visualization of fish length measurement results.

Table 5
Results on test sets of Dataset 1 and Dataset 2. Two-stage methods reproduced results on OpenMMLab platform, best performance from three random runs.

Models	Dataset 1					Dataset 2				
	Fish detection		Curved fish body detection			Fish detection		Curved fish body detection		
	mAP50	mAP50-95	mAP50	mAP50-95	PCK	mAP50	mAP50-95	mAP50	mAP50-95	PCK
FasterRCNN + RTMpose	94.2%	63.1%	88.7%	82.1%	93.1%	95.4%	75.4%	96.8%	87.9%	89.3%
RTMdet + RTMpose	92.1%	65.9%	60.3%	85.4%	91.2%	93.6%	79.5%	94.7%	79.8%	78.6%
Ours	**95.1%**	**67.7%**	**92.3%**	**88.7%**	**95.0%**	**98.4%**	**73.5%**	**94.1%**	**88.9%**	**91.3%**

Table 6
Quantitative results of ablation studies on Dataset 1 and Dataset 2.

Models	Ablations	Dataset 1					Dataset 2				
		Fish detection		Curved fish body detection			Object detection		Curved fish body detection		
		mAP50	mAP50-95	mAP50	mAP50-95	PCK	mAP50	mAP50-95	mAP50	mAP50-95	PCK
Ours	-w/o-Sampleloss	87.3%	72.5%	86.3%	75.7%	89.7%	88.6%	71.0%	87.5%	73.5%	75.1%
	-w/o-EAMRF	85.6%	74.5%	86.6%	79.0%	87.3%	86.0%	68.6%	87.6%	76.3%	78.6%
	-w/o-Endpointloss	92.6%	69.3%	87.4%	79.3%	86.4%	91.5%	67.5%	86.5%	76.2%	73.6%
	full	**93.8%**	**71.5%**	**90.1%**	**81.8%**	**91.3%**	**92.0%**	**70.5%**	**86.9%**	**77.8%**	**77.2%**

Table 7
Systematic analysis of measurement errors.

Error Type	Proportion	Typical Scenario	Error Characteristics	Mitigation Strategy
Depth Quality	45.3%	Fig. 11(c)(e)(f)	Depth jumps (>5mm/pixel), edge blur, noise interference	Multi-frame fusion, depth post-processing
Fish Posture	28.7%	Fig. 11(c)(j)	High curvature bending (R<2L), extreme tilt (>45°)	Multi-view measurement, adaptive weighting
Environmental Factors	18.2%	Complex background	Uneven illumination, background interference	Image enhancement, attention mechanism
Algorithm Inherent	7.8%	Theoretical limitations	4-point Bézier approximation, network prediction uncertainty	Data augmentation, ensemble learning

depth discontinuities and edge blur—exemplified in Fig. 11(c), (e), and (f)—introduce measurement artifacts when depth gradients exceed 5 mm/pixel. Fish posture variations account for 28.7% of errors, as severe body bending ($R < 2L$) or S-shaped deformations, shown in Fig. 11(c) and (j), exceed the representational capacity of the four-control-point Bézier approximation, leading to arc length underestimation. Environmental factors contribute 18.2% of errors, including specular reflections, uneven illumination, and partial occlusions caused by overlapping individuals or equipment interference, as observed in Fig. 11(e), (g), and (h). The remaining 7.8% stems from inherent algorithmic limitations in curve parameterization and network prediction uncertainty. Notably, these failure modes can be effectively mitigated through multi-frame fusion strategies, which reduce the error standard deviation by 32% and ensure measurement reliability under practical deployment conditions.

6. Discussions and outlook

The Bézierfusion framework effectively addresses key challenges in underwater fish measurement while offering broader methodological implications. The proposed “structured detection” paradigm, which integrates visual perception with parametric geometric modeling, provides a generalizable strategy for non-rigid objects exhibiting regular morphological patterns-such as elongated organisms, plant stems, and constrained human poses-and thus holds considerable potential for cross-domain applications.

This work also underscores the pressing need for large-scale, publicly accessible underwater stereo-vision datasets. Current benchmarks remain limited in scale, species diversity, and annotation richness, constraining both model generalization and real-world applicability. Future efforts will focus on constructing standardized datasets spanning diverse aquaculture scenarios, enriched with multimodal annotations including behavioral, pose, and environmental metadata, to facilitate systematic advances in underwater perception research.

Although the four-control-point Bézier curve demonstrates strong representational capability for typical fish poses, its capacity to capture highly complex S-shaped deformations remains constrained by the inherent expressiveness of the geometric primitive rather than the detection paradigm itself. Future extensions may incorporate adaptive curve orders or alternative representations such as B-splines to enhance flexibility while preserving differentiability.

From an application perspective, the parameterized 3D Bézier curves serve as compact morphological descriptors amenable to high-throughput phenotypic analysis, enabling automatic extraction of traits such as body depth and head geometry. This capability positions Bézierfusion as a versatile platform supporting aquaculture breeding, health monitoring, and species identification beyond simple length measurement.

Finally, the framework’s real-time performance and robustness render it well-suited for deployment on autonomous underwater platforms. Future research will explore multi-view fusion strategies to mitigate depth-estimation errors and occlusion artifacts, thereby advancing reliable perception-decision integration for intelligent aquaculture and marine ecological monitoring systems.

7. Conclusion

This paper has introduced Bézierfusion, a novel framework that redefines underwater fish measurement through a structured detection paradigm. By establishing a tightly-coupled vision-geometry model, we have demonstrated that the integration of differentiable Bézier curves into a detection network enables direct, end-to-end, and highly accurate inference of fish length from binocular images. The proposed EAMRF module and composite loss function provide the theoretical and methodological underpinnings for robustness in challenging environments and precision in measurement.

The framework proves its practical value through high accuracy, real-time speed, and successful application in real-world scenarios, supported by two newly contributed datasets. More importantly, it transcends the specific task of length measurement, establishing itself as a foundational step towards automated phenotyping platforms and intelligent underwater perception. We believe the paradigm of structured detection paves the way for a new class of vision systems that seamlessly blend learning with geometric reasoning. It should also be noted that the proposed framework assumes access to reasonably consistent binocular imagery, and its performance may degrade when stereo cues or image quality are severely compromised, which outlines the practical conditions under which the method is most effective.

CRediT authorship contribution statement

Lanlan Liang: Writing - original draft, Methodology, Software, Formal analysis, Visualization; **Zhuhua Hu:** Conceptualization, Writing - review & editing, Project administration, Funding acquisition; **Gaosheng Liu:** Conceptualization, Supervision; **Jie Liu:** Software, Investigation, Validation; **Zekun Deng:** Validation, Data curation; **Shuangyin Liu:** Resources, Supervision; **Yaochi Zhao:** Resources, Supervision; **Ziqi Yang:** Investigation, Data curation.

Ethical Statement

The data collection process in this study was primarily observational and non-invasive, utilizing underwater stereo cameras to capture images of free-swimming groupers (Plectropomus leopardus). No surgical procedures, anesthesia, or euthanasia were performed on any experimental animals. The study was conducted at the scientific aquaculture base of Hainan University. All animal care and handling procedures (including manual measurements for ground truth validation) were conducted in strict accordance with the institutional guidelines for animal welfare and the ethical standards of Hainan University. Every effort was made to minimize stress and disturbance to the fish during the experiment.

Data availability

The datasets generated during the current study are available in the Mendeley Data repository, https://doi.org/10.17632/b249zyfswk.1 [41].

Declaration of competing interest

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

Acknowledgment

This research was supported by the Hainan Seed Industry Laboratory (B23H10004), the National Natural Science Foundation of China (62161010), and the National Key Research and Development Program of China (2022YFD2400504). The authors would like to thank the referees for their constructive suggestions.

References

[1] L. Chunhong, W. Zhiyong, L. Yachao, Z. Zhenzuo, L. Jiawei, X. Chen, D. Rongxiang, L. Daoliang, D. Qingling, Research progress of computer vision technology in abnormal fish detection, Aquacult. Eng. 103 (2023). https://doi.org/10.1016/J.AQUAENG.2023.102350

[2] Z. Minggang, S. Pingfeng, Z. Hao, S. Yang, In-Water fish body-Length measurement system based on stereo vision, Sensors 23 (14) (2023). https://doi.org/10.3390/S23146325

[3] J.A. Lines, R.D. Tillett, L.G. Ross, D. Chan, S. Hockaday, N.J.B. McFarlane, An automatic image-based system for estimating the mass of free-swimming fish, Comput. Electron. Agric. 31 (2) (2001) 151–168. https://doi.org/10.1016/S0168-1699(00)00181-2

[4] L.H. Stien, S. Bratland, I. Austevoll, F. Oppedal, T.S. Kristiansen, A video analysis procedure for assessing vertical fish distribution in aquaculture tanks, Aquacult. Eng. 37 (2) (2007) 115–124. https://doi.org/10.1016/j.aquaeng.2007.03.002

[5] G. Rafael, P. Ricard, Q. Josep, T. Alexander, G. Nuno, R. Shale, V. Håvard, L. Kristoffer, Automatic segmentation of fish using deep learning with application to fish size measurement, ICES J. Mar. Sci. 77 (4) (2019) 1354–1366. https://doi.org/10.1093/icesjms/fsz186

[6] P. Hercos Alexandre, P.A. Carolina, J.M. del Favero, A. Zuchi Nagila, F. Teixeira Túlio, E.A. Albuquerque Fábio, H.L. de Queiroz, Length-weight relationships of ornamental fish species from amanã lake, amanã reserve, amazonas, brazil, J. Appl. Ichthyol. 37 (6) (2021) 985–988. https://doi.org/10.1111/JAI.14217

[7] Z. Hu, R. Li, X. Xia, C. Yu, X. Fan, Y. Zhao, A method overview in smart aquaculture, Environ. Monit. Assess. 192 (8) (2020) 493.

[8] C. Yu, X. Fan, Z. Hu, X. Xia, Y. Zhao, R. Li, Y. Bai, Segmentation and measurement scheme for fish morphological features based on mask R-CNN, Inf. Process. Agriculture 7 (4) (2020) 523–534.

[9] C. Yu, Z. Hu, B. Han, P. Wang, Y. Zhao, H. Wu, Intelligent measurement of morphological characteristics of fish using improved U-Net, Electronics 10 (12) (2021) 1426.

[10] C. Yu, Z. Hu, B. Han, Y. Dai, Y. Zhao, Y. Deng, An intelligent measurement scheme for basic characters of fish in smart aquaculture, Comput. Electron. Agric. 204 (2023).

[11] B. Han, Z. Hu, Z. Su, X. Bai, S. Yin, J. Luo, Y. Zhao, Masklac R-CNN for measuring morphological features of fish, Measurement 203 (2022).

[12] P. Muñoz-Benavent, G. Andreu-García, J.M. Valiente-González, V. Atienza-Vanacloig, V. Puig-Pons, V. Espinosa, Enhanced fish bending model for automatic tuna sizing using computer vision, Comput. Electron. Agric. 150 (2018) 52–61.

[13] J. Zeng, M. Feng, Y. Deng, P. Jiang, Y. Bai, J. Wang, A. Qu, W. Liu, Z. Jiang, Q. He, Z. Wang, P. Xu, Deep learning to obtain high-throughput morphological phenotypes and its genetic correlation with swimming performance in juvenile large yellow croaker, Aquaculture 578 (2024).

[14] R.F. Ainsworth, L.H. Vickers, J.D. Bolland, M.J. Taylor, J.P. Harvey, R.A.A. Noble, I.G. Cowx, A.D. Nunn, Two decades of body length measurements of larval and juvenile fish populations in english rivers, Sci. Data 11 (1) (2024) 1271. https://doi.org/10.1038/S41597-024-04127-W

[15] N. Tonachella, A. Martini, M. Martinoli, D. Pulcini, A. Romano, F. Capoccioni, An affordable and easy-to-use tool for automatic fish length and weight estimation in mariculture, Sci. Rep. 12 (1) (2022) 15642.

[16] Y. Deng, H. Tan, M. Tong, D. Zhou, Y. Li, M. Zhu, An automatic recognition method for fish species and length using an underwater stereo vision system, Fishes 7 (6) (2022) 326.

[17] E. Harvey, M. Cappo, M. Shortis, S. Robson, J. Buchanan, P. Speare, The accuracy and precision of underwater measurements of length and maximum body depth of southern bluefin tuna (thunnus maccoyii) with a stereo–video camera system, Fish. Res. 63 (3) (2003) 315–326.

[18] T. Tanaka, R. Ikeda, Y. Yuta, K. Tsurukawa, S. Nakamura, T. Yamaguchi, K. Komeyama, Annual monitoring of growth of red sea bream by multi-stereo-image measurement, Fisheries Science 85 (6) (2019) 1037–1043.

[19] S. Torisawa, M. Kadota, K. Komeyama, K. Suzuki, T. Takagi, A digital stereo-video camera system for three-dimensional monitoring of free-swimming pacific bluefin tuna, thunnus orientalis, cultured in a net cage, Aquat. Living Resour. 24 (2) (2011) 107–112.

[20] RosenShale, JørgensenTerje, Hammersland-WhiteDarren, H. Christian, Deepvision: a stereo camera system provides highly accurate counts and lengths of fish passing inside a trawl, Can. J. Fish. Aquat.Sci. 70 (10) (2013) 1456–1467. https://doi.org/10.1139/cjfas-2013-0124

[21] F. Shafait, E.S. Harvey, M.R. Shortis, A. Mian, M. Ravanbakhsh, J.W. Seager, P.F. Culverhouse, D.E. Cline, D.R. Edgington, Towards automating underwater measurement of fish length: a comparison of semi-automatic and manual stereo–video measurements, ICES J. Mar. Sci. 74 (6) (2017) 1690–1701.

[22] M. Ravanbakhsh, M.R. Shortis, F. Shafait, A. Mian, E.S. Harvey, J.W. Seager, Automated fish detection in underwater images using shape-Based level sets, The Photogrammetric Record 30 (149) (2015) 46–62. https://doi.org/10.1111/phor.12091

[23] C. Costa, A. Loy, S. Cataudella, D. Davis, M. Scardi, Extracting fish size using dual underwater cameras, Aquacult. Eng. 35 (3) (2006) 218–227. https://doi.org/10.1016/j.aquaeng.2006.02.003

[24] R. Tillett, N. McFarlane, J. Lines, Estimating dimensions of free-Swimming fish using 3D point distribution models, Comput. Vision Image Understanding 79 (1) (2000) 123–141. https://doi.org/10.1006/cviu.2000.0847

[25] Q. Al-Jubouri, W. Al-Nuaimy, M. Al-Taee, I. Young, An automated vision system for measurement of zebrafish length using low-cost orthogonal web cameras, Aquacult. Eng. 78 (2017) 155–162. https://doi.org/10.1016/j.aquaeng.2017.07.003

[26] C. Shi, Q. Wang, X. He, X. Zhang, D. Li, An automatic method of fish length estimation using underwater stereo system based on labVIEW, Comput. Electron. Agric. 173 (2020). https://doi.org/10.1016/j.compag.2020.105419

[27] N.A. Ubina, S.-C. Cheng, C.-C. Chang, S.-Y. Cai, H.-Y. Lan, H.-Y. Lu, Intelligent underwater stereo camera design for fish metric estimation using reliable object matching, IEEE Access 10 (2022) 74605–74619.

[28] H. Yu, H. Song, L. Xu, D. Li, Y. Chen, SED-RCNN-BE: A SE-Dual channel RCNN network optimized binocular estimation model for automatic size estimation of free swimming fish in aquaculture, Expert Syst. Appl. 255 (2024) 124519.

[29] R. Garcia, R. Prados, J. Quintana, A. Tempelaar, N. Gracias, S. Rosen, H. Vågstøl, K. Løvall, Automatic segmentation of fish using deep learning with application to fish size measurement, ICES J. Mar. Sci. 77 (4) (2020) 1354–1366.

[30] D. Voskakis, A. Makris, N. Papandroulakis, Deep learning based fish length estimation. an application for the mediterranean aquaculture, in: OCEANS 2021: San Diego–Porto, IEEE, 2021, pp. 1–5.

[31] R. Cheng, C. Zhang, Q. Xu, G. Liu, Y. Song, X. Yuan, J. Sun, Underwater fish body length estimation based on binocular image processing, Information 11 (10) (2020) 476.
[32] D. Ouyang, S. He, G. Zhang, M. Luo, H. Guo, J. Zhan, Z. Huang, Efficient multi-Scale attention module with cross-Spatial learning, in: ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2023, pp. 1–5. https://doi.org/10.1109/ICASSP49357.2023.10096516
[33] K. Chen, J. Wang, J. Pang, Y. Cao, Y. Xiong, X. Li, S. Sun, W. Feng, Z. Liu, J. Xu, et al., MMDetection: Open mmlab detection toolbox and benchmark, (2019) arXiv:1906.07155.
[34] D. Maji, S. Nagori, M. Mathew, D. Poddar, YOLO-Pose: Enhancing YOLO for Multi Person Pose Estimation Using Object Keypoint Similarity Loss, 2022, https://arxiv.org/abs/2204.06806.
[35] J. Luo, L. Zhu, L. Li, P. Hong, Robot visual servoing grasping based on top-Down keypoint detection network, IEEE Trans. Instrum. Meas. 73 (2024) 1–11. https://doi.org/10.1109/TIM.2023.3335521
[36] C. Lyu, W. Zhang, H. Huang, Y. Zhou, Y. Wang, Y. Liu, S. Zhang, K. Chen, Rtmdet: an empirical study of designing real-time object detectors, (2022) arXiv:2212.07784.
[37] T. Jiang, P. Lu, L. Zhang, N. Ma, R. Han, C. Lyu, Y. Li, K. Chen, Rtmpose: real-time multi-person pose estimation based on mmpose, (2023) arXiv:2303.07399.
[38] C. Geng, A. Wang, C. Yang, Z. Xu, Y. Xu, X. Liu, H. Zhu, Application of improved YOLOv8n-seg in crayfish trunk segmentation (2024).
[39] D. Cao, H. Yang, C. Guo, Y. Cheng, W. Zhang, M. Shi, X.-Q. Xia, A real-time detection and non-destructive warning method for zebrafish body surface anomalies based on improved YOLOv8 framework, bioRxiv (2025) 2025–2104.
[40] K. Huang, Y. Li, F. Suo, J. Xiang, Stereo vison and mask-RCNN segmentation based 3D points cloud matching for fish dimension measurement, in: 2020 39Th Chinese Control Conference (CCC), IEEE, 2020, pp. 6345–6350.
[41] J. Liu, L. Liang, Z. Hu, Stereo Fish Dataset for BézierFusion: 3D Curve Estimation in Aquaculture, 2025, https://doi.org/10.17632/b249zyfswk.1