Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

PPDS-Net: A Physical-Prior-Guided Dual-Stream Network for Non-contact Fish Weight Estimation from Single-View 3D Point Clouds

Rui Zhang
Beijing Information Science & Technology University
Shiyu Zhang
zhangshiyu@bistu.edu.cn

Beijing Information Science & Technology University
Shuo Li
Beijing Information Science & Technology University
Yihang Lu
Beijing Information Science & Technology University
Runbei Feng
Beijing Information Science & Technology University
Kai Lin
Beijing Academy of Agriculture and Forestry Sciences

Research Article

Keywords: Fish weight estimation, Point clouds, Deep learning, Physical priors, Multimodal fusion, Computer vision

Posted Date: November 2nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11292039/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

PPDS-Net: A Physical-Prior-Guided Dual-Stream Network for Non-contact Fish Weight Estimation from Single-View 3D Point Clouds

Rui Zhang [a,b], Shiyu Zhang [a,*], Shuo Li [a], Yihang Lu [a], Runbei Feng [a], Kai Lin [b,*]

[a] College of Instrument Science and Opto-electronics Engineering, Beijing Information Science and Technology University, Beijing 100192, China

[b] Fisheries Science Institute, Beijing Academy of Agriculture and Forestry Sciences & National Engineering Research Center for Freshwaters (Beijing) & Beijing Key Laboratory of Fisheries Biotechnology, Beijing 100068, P. R. China

Corresponding author: Shiyu Zhang, Kai Lin

E-mail: zhangshiyu@bistu.edu.cn; linkai@baafs.net.cn

Abstract

Accurate, noninvasive estimation of fish weight is critical for biomass assessment, feeding management, and health monitoring in precision aquaculture. In underwater scenes, however, freely swimming fish undergo nonrigid deformation, adopt variable postures, and often yield sparse or noisy point clouds, which limits the robustness of existing vision-based methods. This study develops PPDS-Net, a physical-prior-guided dual-stream feature fusion network for end-to-end sturgeon weight estimation from single-view point clouds. A mask-guided single-view 3D reconstruction strategy is first used to suppress background interference and generate high-quality fish point clouds from RGB-D data. An implicit geometric stream based on point-cloud learning is then fused with an explicit physical stream composed of four morphological priors: raw point count, sampling-density ratio, spatial dispersion, and topological skeleton length. To embed morphological constraints into feature learning, the framework further introduces physical-attention modulation and physical-consistency loss. Experiments using 4,175 RGB-D samples from 86 sturgeon show that the proposed method achieves an MAE of 118.31 g and a Pearson correlation coefficient of 0.9563. Compared with the baseline PointNet++ model, the proposed framework substantially improves prediction accuracy and reduces the number of epochs required to reach the 200 g error threshold from 191 to 8. PPDS-Net also retains stable performance under highly sparse point-cloud conditions. These results demonstrate that integrating physical priors with deep point-cloud learning provides an accurate, robust, and interpretable approach for underwater fish weight estimation.

Keywords

Fish weight estimation; Point clouds; Deep learning; Physical priors; Multimodal fusion; Computer vision

1. Introduction

Accurate, non-contact estimation of fish weight is a central technical requirement in precision

aquaculture, as it supports dynamic biomass assessment, feeding management, health monitoring, and improvements in farming efficiency (Li et al., 2024). Compared with manual netting and contact-based weighing, vision-based estimation can reduce labor demand while avoiding stress responses and potential injuries caused by mechanical handling (Kong et al., 2024). For practical deployment in aquaculture environments, however, a vision-based method must do more than provide non-contact measurement; it must also maintain stable performance when fish swim freely, change posture continuously, and are observed under complex imaging perturbations (Badgujar et al., 2024).

Early vision-based studies of fish weight estimation mainly relied on two-dimensional images. These approaches typically extract geometric descriptors from fish projections, including body length, body width, and projected area, and then establish regression models for body-weight prediction (Zion, 2012). For example, image-derived morphological features have been combined with neural networks or partial least squares regression to estimate fish body mass, and preprocessing strategies such as caudal-fin removal have been used to reduce interference from non-trunk regions (Hao et al., 2022; Zhang et al., 2020). Although such methods are inexpensive and easy to deploy, they are inherently limited because three-dimensional morphology is compressed into a two-dimensional projection. Fish weight is more closely related to body volume than to projected area or contour shape, both of which are sensitive to viewpoint, occlusion, and body bending. During free swimming, the same individual can present substantially different projected areas and contours across postures, which causes unstable weight estimates. Underwater light attenuation, reflection interference, and background noise further degrade 2D features, making it difficult for projection-based models to achieve high accuracy and robustness at the same time (González-Sabbagh and Robles-Kelly, 2023).

To overcome the limited ability of two-dimensional projections to represent spatial morphology, 3D reconstruction, stereo matching, and RGB-D perception have increasingly been introduced into fish phenotyping and weight-estimation studies (Zhao et al., 2024). Compared with 2D contour- or area-based methods, 3D perception provides spatial geometric information that more directly reflects fish body volume and therefore offers stronger physical interpretability for non-contact measurement of free-swimming fish. Most 3D fish weight-estimation studies in aquaculture have followed a pipeline of geometric reconstruction, explicit feature extraction, and statistical mapping. In this pipeline, the spatial structure of the fish body is first recovered through binocular vision, depth perception, or 3D reconstruction (Wang et al., 2024; Yu et al., 2022); handcrafted geometric features such as body length, perimeter, surface area, or volumetric proxies are then extracted and mapped empirically to body weight. This strategy is interpretable because the prediction is based on intuitive geometric quantities. Its performance, however, is constrained by multi-stage processing and underwater imaging degradation. Segmentation, depth recovery, reconstruction, geometric measurement, and regression form a cascade in which errors from one stage can propagate and be amplified in later stages. In addition, light attenuation, backscattering, local reflections, and weak-

texture regions can produce depth noise, point-cloud voids, and boundary distortion, making explicit geometric measurements highly sensitive to reconstruction quality, fish posture, and local occlusion (González-Sabbagh and Robles-Kelly, 2023). These challenges are particularly severe under single-view observation, where the dorsal side and occluded body regions cannot be fully captured and where bending, twisting, and nonrigid deformation during swimming are difficult to describe with simple rigid measurements. Thus, adding 3D information alone does not fully solve the stability problem in free-swimming fish weight estimation. A more important challenge is to reduce error accumulation from multi-step geometric modeling and to learn robust 3D representations that remain closely associated with fish body volume.

End-to-end point-cloud deep learning offers a different paradigm for 3D phenotypic regression because it can take unordered discrete point sets directly as input (Bello et al., 2020; Qi et al., 2017a; Qi et al., 2017b). Rather than relying on manually extracted body length, surface area, or volumetric proxies, point-cloud networks learn local geometric structures and global morphological representations from raw 3D point sets, reducing information loss caused by feature compression and multi-step modeling. In precision livestock farming, end-to-end point-cloud regression has been applied to body-size and weight-estimation tasks for broilers, pigs, beef cattle, and other farmed animals, where it has shown favorable prediction accuracy and cross-individual adaptability (Hou et al., 2023; Paudel et al., 2024; Zheng et al., 2025). These findings suggest that point-cloud learning can establish nonlinear mappings between 3D morphology and biomass-related indicators. In free-swimming fish, however, weight estimation still relies largely on explicit reconstruction and indirect regression, and end-to-end prediction from single-view fish point clouds remains insufficiently explored. Existing aquaculture studies have mainly emphasized 3D reconstruction quality or deep feature enhancement (Wang et al., 2026). Although these methods partly alleviate underwater imaging degradation and posture variation, purely data-driven models still lack explicit constraints on fish morphological regularities when learning high-dimensional features for single-view point-cloud weight estimation. Consequently, such models may learn environmental noise, sampling-density fluctuations, or posture-specific artifacts as weight cues. This risk increases when the dataset is limited, the view is incomplete, and underwater point-cloud quality is unstable. Explicit morphological priors related to fish body volume therefore need to be incorporated into end-to-end point-cloud learning so that the model retains deep representation capacity while gaining more stable scale perception and biologically meaningful constraints (Qi et al., 2017a).

To address these limitations, this study proposes a physical-prior-guided dual-stream network (PPDS-Net) for non-contact, end-to-end sturgeon weight estimation from single-view point clouds in complex underwater environments. PPDS-Net preserves the ability of point-cloud networks to capture local geometric details while converting volume-related morphological regularities into explicit physical priors. Feature modulation and physical-consistency optimization are then used to integrate knowledge-driven constraints with data-driven representation learning. Relative to explicit geometric fitting, the proposed approach avoids the information compression and error

accumulation caused by multi-stage modeling. Relative to generic data-driven point-cloud networks, it provides more stable predictions under complex postures, point-cloud sparsification, and underwater noise. The main contributions are as follows:

(1) A mask-guided single-view underwater 3D reconstruction pipeline was developed for fish bodies. By using instance segmentation and depth projection to suppress complex background interference and invalid depth measurements, the pipeline improves the cleanliness and usability of fish point clouds.

(2) PPDS-Net, a dual-stream feature fusion network guided by multidimensional physical priors was proposed. A four-dimensional vector of macroscopic physical priors, including raw point count, sampling-density ratio, spatial dispersion, and topological skeleton length, is explicitly incorporated to enhance scale perception and physical interpretability under single-view and nonrigid posture conditions.

(3) A physical-consistency optimization mechanism was designed specifically for weight regression. By jointly modeling regression loss and physical-consistency loss, the proposed method improves prediction stability and generalization under point-cloud sparsification and complex underwater conditions.

2. Materials and methods

To achieve accurate, non-contact prediction of sturgeon weight in complex underwater environments, this study establishes a technical pipeline that integrates data acquisition, instance segmentation, single-view 3D reconstruction, and weight regression, as illustrated in Fig. 1. The framework first extracts RGB images and depth maps from fish videos and then obtains fish-body foreground masks through instance segmentation. Calibrated camera parameters are subsequently used to project depth values within the target region and reconstruct a single-view point cloud. In the prediction stage, PPDS-Net integrates multidimensional physical-prior features extracted from the point cloud with deep geometric representations to estimate sturgeon weight under physical-prior constraints. This design retains the end-to-end learning capacity of point-cloud-based models while incorporating prior knowledge of fish morphology, thereby balancing prediction accuracy, interpretability, and robustness to complex postures.

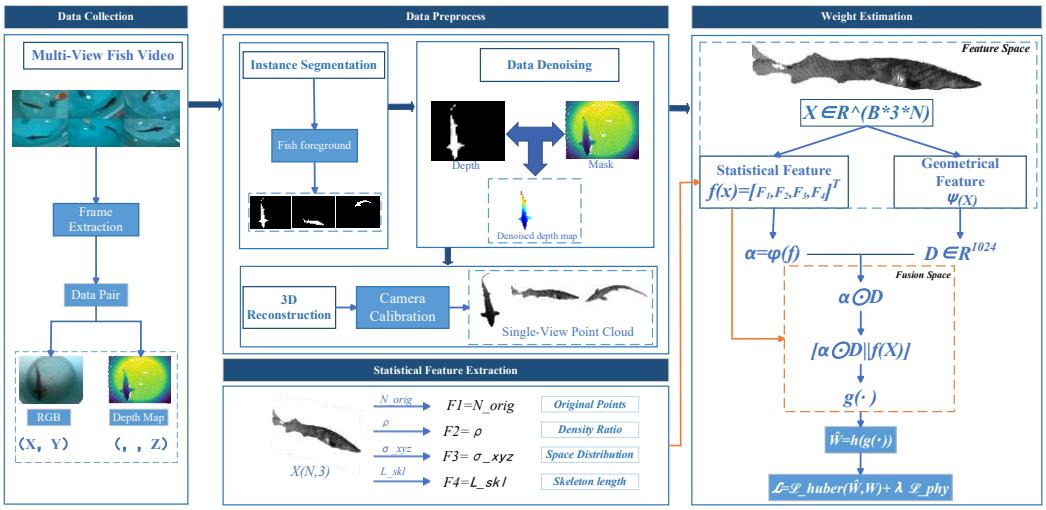


Fig. 1. Overall framework for single-view point-cloud-based sturgeon weight estimation using PPDS-Net.

2.1 Data collection

Point clouds are a fundamental data representation in 3D vision and are commonly acquired using sensors such as depth cameras and LiDAR systems. A point cloud consists of numerous discrete three-dimensional points, each containing spatial coordinate information (x, y, z); when needed, it can also be supplemented with geometric attributes such as surface normals. In this study, raw RGB-D data of sturgeon were collected using an Intel RealSense D455 depth camera.

Data were collected at the aquaculture facility of the Fisheries Research Institute, Beijing Academy of Agriculture and Forestry Sciences, and at the Huxinzhuang experimental site in Pinggu District. The experimental setup, manual weighing procedure, and data acquisition system are shown in Fig. 2. The experimental container was a large cylindrical tank approximately 1.8 m in diameter and 0.6 m in water depth. The depth camera was mounted above the tank, with the lens positioned approximately 40 cm above the water surface, and continuous acquisition was performed under top-lighting conditions. In total, 86 sturgeon were included, covering three varieties: Siberian sturgeon, Amur sturgeon, and the hybrid sturgeon "Xunlong No. 1". The image resolution was set to 640 × 480 pixels. To ensure accurate depth projection and point-cloud reconstruction, the intrinsic parameters, extrinsic parameters, and distortion coefficients of the camera were calibrated using Zhang's calibration method (Zhang, 1999).

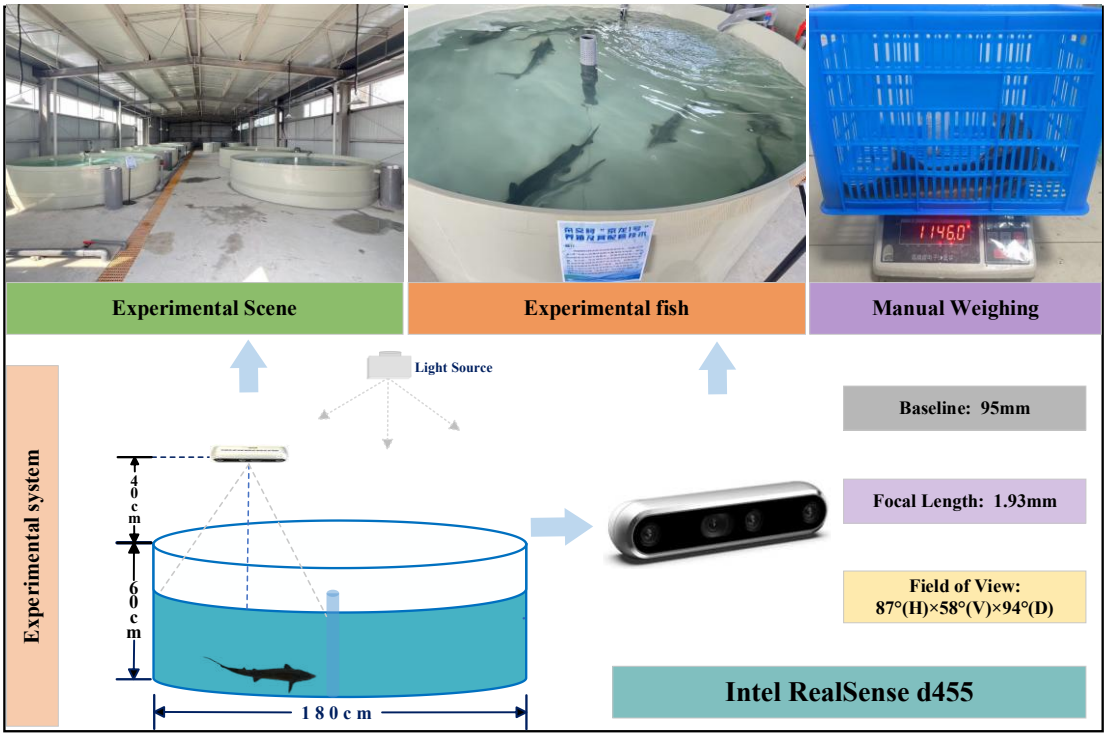


Fig. 2. The experimental system and scenarios.

Before formal data acquisition, the ground-truth weight of each fish was measured using an electronic scale and used as the supervisory label for model training. Each individual was then placed in a simulated aquaculture environment, and video-stream data were collected after its swimming behavior had stabilized. RGB images and the corresponding depth maps were recorded synchronously. To capture sufficient posture variation, one frame was extracted every 25 frames from the continuous video, producing approximately 50 sample images for each fish under different viewing angles and postures. After manual screening, 4,175 valid samples were retained. The samples for each fish were then matched one-to-one with the corresponding ground-truth weight to construct the final dataset for developing and evaluating PPDS-Net. The weight distribution of the dataset is shown in Fig. 3.

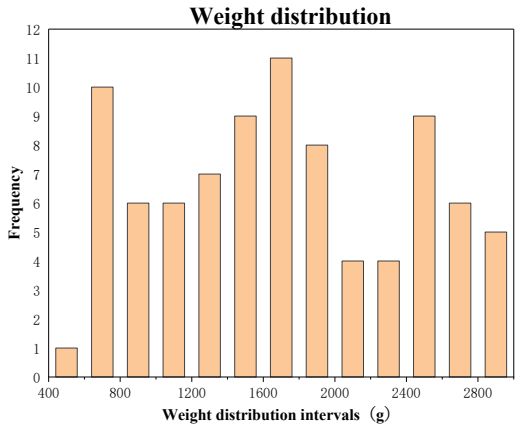


Fig. 3. The distribution of weight among the experimental fish.

2.2 Mask-Guided Single-View 3D Reconstruction of Underwater Fish

High-quality fish point clouds are essential for reliable weight estimation. In underwater environments, however, raw point clouds generated directly by depth cameras usually contain substantial interference from the water body, tank boundaries, and background facilities. Such interference weakens the geometric representation of the fish target and can reduce the reliability of subsequent regression. To address this problem, this study proposes a 2D mask-guided single-view 3D reconstruction method that improves the purity and usability of fish point clouds. The overall workflow is shown in Fig. 4.

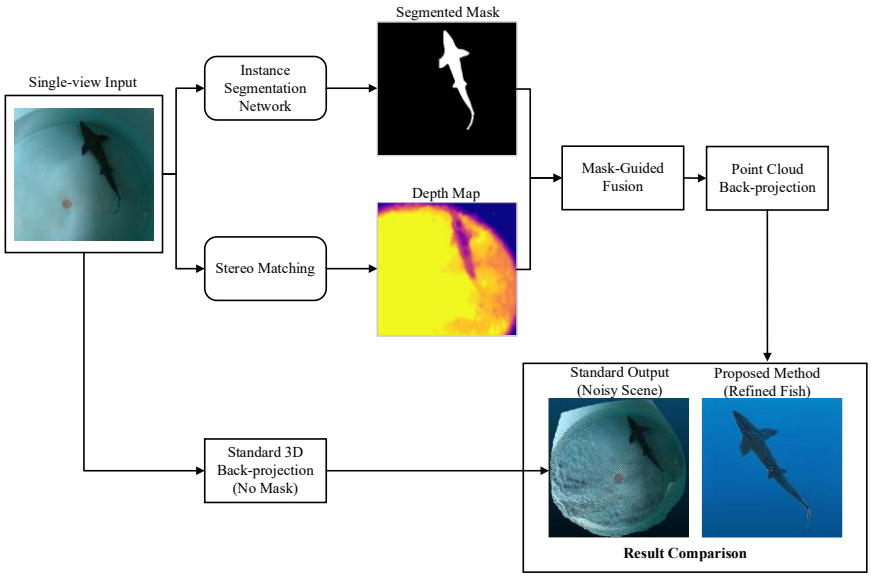


Fig. 4. The overall processing architecture of the mask-guided single-view 3D reconstruction of underwater fish.

First, a YOLOv11-based instance segmentation network was used to segment the fish body region in RGB images. Based on the collected data, an instance segmentation training set containing 2,000 annotated images was constructed. The YOLOv11n model achieved a mask mAP@0.5 of 0.98 on the test set, indicating that it could effectively separate the fish body from the complex underwater background and provide reliable foreground constraints for depth projection. Performing target separation in the 2D image domain before 3D reconstruction substantially reduces the uncertainty associated with direct geometric analysis of noisy point clouds.

After the fish body mask was obtained, it was fused pixel by pixel with the original depth map to retain only the valid depth information corresponding to the target region. Let the original depth map be denoted by D_{raw} and the binary mask by M. The target-region depth data D_{target} can then be expressed as:

$$D_{target} = D_{raw} \odot M \tag{1}$$

where $\odot$ denotes element-wise multiplication. Invalid depth values were further removed by defining the filtered depth data as follows:

$$D_{filtered}(u,v)=\begin{cases} D_{target}(u,v) & \text{if } 0 < D_{target}(u,v) < d_{max} \\ \text{NAN} & \text{otherwise} \end{cases} \tag{2}$$

where $D_{filtered}$ denotes the filtered depth data, (u,v) represents the image pixel coordinates, and $d_{\max}$ is the depth threshold.

The intrinsic matrix *K* of the Intel RealSense D455 depth camera is given by:

$$K=\begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}=\begin{bmatrix} 319.7 & 0 & 324.6 \\ 0 & 383.2 & 247.4 \\ 0 & 0 & 1 \end{bmatrix} \tag{3}$$

where f_x and f_y denote the focal lengths of the camera in pixel units along the horizontal and vertical directions, respectively, and c_x and c_y represent the coordinates of the principal point.

Based on the fused depth data, a 3D point cloud was generated using the point-cloud generation algorithm provided by the Open3D library (Zhou et al., 2018). The two-dimensional pixel coordinates (u,v) were then mapped into the three-dimensional camera coordinate system using the pinhole camera model (Hartley and Zisserman, 2004):

$$Z=\frac{D_{filtered}(u,v)}{\lambda} \tag{4}$$

$$X=\frac{(u-c_x)\times Z}{f_x} \tag{5}$$

$$Y=\frac{(v-c_y)\times Z}{f_y} \tag{6}$$

where $P=(X,Y,Z)$ denotes a 3D point in the camera coordinate system, and λ is the depth scaling factor.

To reduce local noise and outliers, the Statistical Outlier Removal (SOR) algorithm (Rusu et al., 2008) was applied to the reconstructed point cloud. For an arbitrary point p_i in the point cloud, its mean distance d_i to the $k=10$ nearest neighbors was calculated. A point was identified as an outlier when this distance deviated from the global statistical distribution by more than three standard deviations:

$$p_i \in \text{Outliers} \Leftrightarrow | d_i-\mu | > 3\sigma \tag{7}$$

where μ and σ represent the mean and standard deviation of the global average neighborhood distance, respectively.

After these processing steps, background noise introduced by direct projection of the raw depth map was substantially reduced while the main structural components of the fish body were preserved, as shown in Fig. 4. The point clouds generated by the proposed method contained

approximately 5,000–15,000 valid points on average. Compared with the conventional approach that generates point clouds directly from raw depth maps under background interference, the proposed method reduced noisy points by more than 70%. This reconstruction procedure therefore provides a stable data foundation for subsequent physical-prior extraction and weight estimation with PPDS-Net.

2.3 PPDS-Net Architecture and Physical-Prior-Guided Feature Fusion

Existing 3D point-cloud-based methods for fish weight prediction rely mainly on generic architectures such as PointNet and rarely model the biological and morphological regularities of fish explicitly. To address this limitation, this study proposes PPDS-Net, a dual-stream feature fusion and optimization network guided by multidimensional physical priors, as illustrated in Fig. 5. The framework builds an explicit physical stream and an implicit geometric stream in parallel. In this way, it preserves the ability of point-cloud networks to capture local geometric details while embedding prior physical knowledge, including macroscopic scale, spatial extent, and topological body length, into the regression process. This design improves prediction accuracy, stability, and physical interpretability under single-view observation, sparse point clouds, and nonrigid postures.

Although end-to-end point-cloud networks can learn high-dimensional implicit representations from raw unordered point sets, their feature-extraction process is usually weakly constrained with respect to macroscopic attributes such as target scale, body extension, and topological length. This weakness is especially problematic under single-view observation, where posture bending, local occlusion, and sampling-density fluctuation can easily bias the learned representation. Therefore, a low-dimensional feature vector composed of four explicit physical priors was constructed from the sturgeon point cloud P to provide an interpretable macroscopic reference for biomass-related estimation.

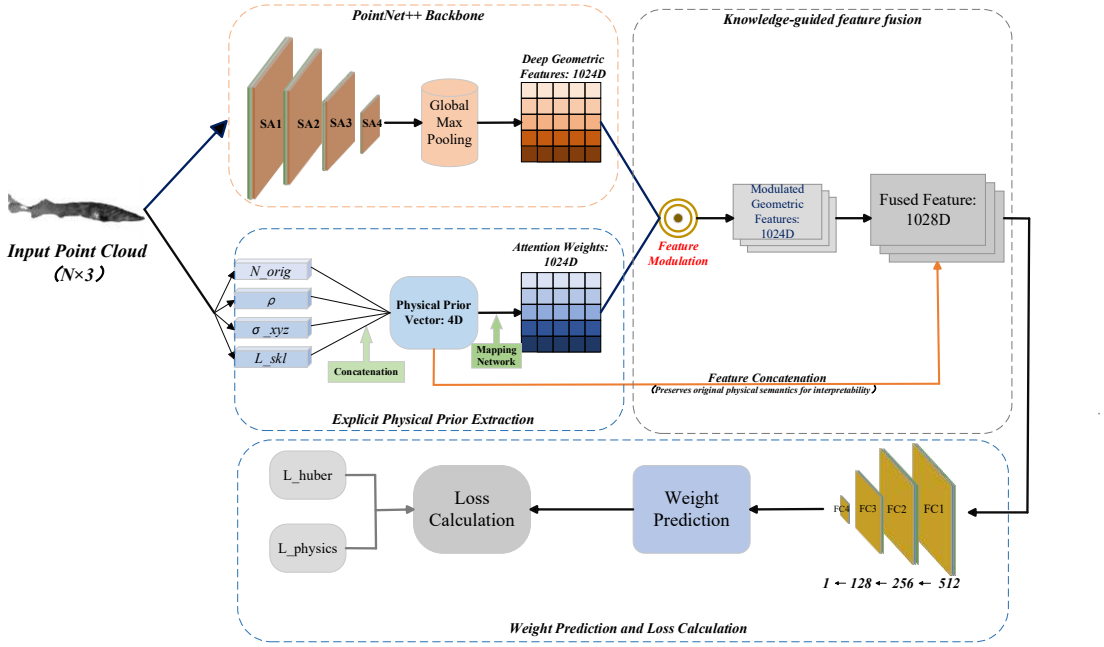


Fig. 5. Architecture of PPDS-Net, the proposed dual-stream feature fusion network guided by

multidimensional physical priors.

The first prior is the global projected point count F_1. When the camera intrinsics, viewing angle, and depth range are relatively fixed, the number of points in the raw point cloud reflects the scale of the visible fish-body surface in three-dimensional space to a certain extent. This feature therefore provides a basic macroscopic scale constraint for weight prediction:

$$F_1 = |\mathrm{P}| = N \tag{8}$$

where P denotes the number of points in point cloud P, and N represents the total number of points in the original point cloud.

The second prior is the network sampling-density ratio F_2. PointNet++-style networks typically require a fixed number of input points, which means that individuals of different body sizes undergo different degrees of information compression after uniform downsampling. Heavier fish may lose more local geometric detail than smaller individuals under the same fixed input size. To quantify this asymmetric information loss, the sampling-density ratio is defined as:

$$F_2 = \frac{N_{target}}{N} \tag{9}$$

where N_{target} denotes the fixed number of input points required by the network, and N represents the total number of points in the raw point cloud. This feature informs the network of the current compression ratio of the fish point cloud, helping the model recover true scale differences among samples under a fixed input size.

The third prior is point-cloud spatial dispersion, denoted as F_3, which characterizes the overall spatial extent of the fish point cloud. It is calculated as the root-mean-square distance of all points from the geometric center and therefore represents the spatial occupancy of the point cloud:

$$F_3 = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i^2 + y_i^2 + z_i^2)} \tag{10}$$

where (x_i, y_i, z_i) denotes the centered coordinates. Compared with axis-aligned bounding-box descriptors, this feature has better rotational invariance and is less sensitive to local outliers, providing a more robust description of overall fish-body extension in three-dimensional space.

The fourth prior is topology-skeleton-based biological body length F_4. Traditional PCA- or OBB-based principal-axis lengths are effective when the fish body is nearly straight, but they measure chord length rather than true arc length when the sturgeon bends markedly, which can distort volume estimation. To address this limitation, this study introduces a point-cloud skeleton extraction method based on the Laplace-Beltrami operator (Cao et al., 2010). The method iteratively contracts the point-cloud surface and collapses it into a one-dimensional topological medial axis. Its objective function is defined as:

$$E(P') = \| W_L L P' \|^2 + \| W_H (P' - P) \|^2 \tag{11}$$

where P' denotes the updated point cloud during contraction, and L is the discrete Laplacian

operator matrix. W_L and W_H are diagonal weight matrices that control the smoothness of the contraction curve and the geometric fidelity to the initial position, respectively.

Because the initial skeleton points may contain noise and local discontinuities, a cubic B-spline curve (Piegl and Tiller, 1997) was further used to fit and refine the skeleton nodes. Let the fitted curve be denoted as $C(t)$:

$$C(t)=\sum_{i=0}^{n} N_{i,3}(t)P_i \tag{12}$$

where P_i denotes the control points, and $N_{i,3}(t)$ represents the cubic B-spline basis functions. The control points are optimized using least squares so that the fitted curve remains smooth while adhering as closely as possible to the original skeleton points.

Fish body length is obtained by calculating the arc length of the skeleton curve. Let the parameter interval be defined as $t\in[a,b]$; the skeleton-curve arc length can be expressed as:

$$L_{skeleton}=\int_b^a \left\| \frac{dC(t)}{dt} \right\| dt \tag{13}$$

In practical computation, the curve is uniformly sampled into M points $\{q_i\}_{j=1}^{M}$, and the arc length is approximated through discrete integration as follows:

$$L_{skeleton}=\sum_{j=1}^{M-1} \| q_{j+1}-q_j \| \tag{14}$$

The skeleton arc length is finally used as the fourth physical-prior feature $F_4=L_{skeleton}$. Because it follows the medial axis of the fish body, this feature can better represent the physiological body length under bent postures and can mitigate scale distortion caused by posture variation. The physical-prior vector constructed in this study is therefore expressed as:

$$V_{prior}=[F_1,F_2,F_3,F_4] \tag{15}$$

Using these physical priors, PPDS-Net constructs a dual-stream feature fusion architecture consisting of an explicit physical stream and an implicit geometric stream. The explicit physical stream takes the four-dimensional prior vector V_{prior} as input and describes macroscopic biomass-related attributes of the fish body. The implicit geometric stream takes the single-view point cloud P as input and uses an improved PointNet++ backbone to extract high-dimensional geometric features, capturing local surface structure and fine-grained spatial morphology. Let the implicit geometric representation produced by PointNet++ be denoted as:

$$V_{implicit}\in R^{1024} \tag{16}$$

Instead of using conventional direct channel concatenation, this study adopts a physical-attention modulation mechanism to coordinate the two streams. The four-dimensional physical-prior vector is first projected nonlinearly by a lightweight multilayer perceptron Φ_{MLP} into an attention space with the same dimensionality as the geometric features. It is then fused with the deep

geometric representation to recalibrate feature channels under the guidance of physical knowledge. The fused feature V_{fuse} is defined as:

$$V_{fuse}=\Phi_{MLP}(Concat(V_{implicit},V_{prior})) \tag{17}$$

where $Concat(\cdot)$ denotes feature concatenation, and Φ_{MLP} consists of a linear projection, ReLU, and dropout. This module learns how different physical priors should modulate the geometric channels. The fused features are then passed to the regression head to output the predicted sturgeon weight:

$$\hat{W} = g(V_{fuse}) \tag{18}$$

where $g(\cdot)$ denotes the final regression mapping function.

In PPDS-Net, this fusion strategy converts fish morphological regularities into differentiable dynamic gating signals. As a result, the network can learn high-dimensional geometric representations while remaining aligned with fundamental physical constraints. Compared with purely data-driven feature learning, the proposed design improves adaptability to posture variation, noise perturbation, and distribution-shifted samples with only a small parameter increase, while also strengthening the physical interpretability of weight prediction.

2.4 Model Optimization and Evaluation Metrics Based on Physical Priors

To improve the physical plausibility of PPDS-Net predictions without sacrificing regression accuracy, a collaborative optimization mechanism constrained by physical priors is introduced during network training. The mechanism combines conventional supervised regression loss with physical-consistency loss, encouraging the network to learn the weight-estimation mapping while remaining consistent with fish morphological regularities. The overall optimization objective is defined as:

$$\xi_{\text{total}} = L_{\delta} + \lambda L_{physics} \tag{19}$$

where L_{δ} computes the regression error between the predicted weight $\hat{W}$ and the ground-truth weight W, $L_{physics}$ denotes the physical-consistency constraint term, and λ is a weighting coefficient that balances data fitting and physical regularization.

The Huber loss function (Meyer, 2021) is defined as:

$$L_{\delta}\left(\mathrm{y}, f\left(x\right)\right)=\begin{cases}\frac{1}{2}\left(y-f(x)\right)^{2}, |y-f(x)|\le\delta \\ \delta|y-f(x)|-\frac{1}{2}\delta^{2}, |y-f(x)|>\delta\end{cases} \tag{20}$$

where y denotes the actual value, $f(x)$ represents the predicted value, and δ is the Huber loss parameter. In this study, δ is set to 0.4. The loss is quadratic when the prediction error is small, which supports smooth convergence. When the error is large, it becomes linear, reducing the influence of outlier samples during training.

Physical-consistency loss is a key component of PPDS-Net. Under a fixed number of network

input points, fish body weight has a relatively stable physical correspondence with density-related point-cloud features. To prevent the network from relying only on empirical fitting to the sample distribution, a physical baseline model is introduced. A consistency term then constrains the regression predictions so that they do not deviate excessively from plausible estimates implied by prior knowledge. The physical baseline model is defined as:

$$\hat{W}_{\text{physics}} = \frac{k}{N_{\text{orig+}}\beta} \tag{21}$$

where N_{orig} denotes the number of points in the raw point cloud, and k and b are scaling coefficients. This formulation represents the inverse relationship between fish body weight and sampling density under fixed sampling conditions. The physical-consistency loss is defined as:

$$L_{\text{physics}} = \frac{1}{B}\sum_{\text{i=1}}^{B}(\rho_i - \frac{k}{W_i + b})^2 \tag{22}$$

where B denotes the batch size, ρ_i is the density-ratio feature of the i-th sample, W_i is the corresponding ground-truth weight, and k and b are empirical parameters used to adjust the strength of the inverse relationship and prevent numerical instability, respectively.

From an optimization perspective, L_{huber} provides direct supervision from labeled data and ensures adequate model-fitting capacity, whereas $L_{physics}$ introduces domain-knowledge constraints that suppress representations inconsistent with biological regularities under small-sample conditions, noise perturbations, and distribution shifts. Because this constraint acts during feature learning rather than as post-processing, it improves interpretability and generalization at the level of intermediate representations.

The training code was implemented in the PyTorch deep learning framework. The AdamW optimizer (Loshchilov and Hutter, 2019) was used during training, with the weight decay coefficient set to 1×10^{-4}. To improve robustness to point-cloud perturbations, random translation and random point dropout were applied before training, and early stopping was used to mitigate overfitting. For learning-rate scheduling, a warm-up single-cycle cosine annealing strategy (Loshchilov and Hutter, 2017) was adopted and can be expressed as:

$$\eta_{\text{t}} = \eta_{\min} + \frac{1}{2}\left(\eta_{\max} - \eta_{\min}\right)\left(1 + \cos\left(\frac{T_{\text{cur}}}{T_{\max}}\pi\right)\right) \tag{23}$$

where η_{min} denotes the minimum learning rate, $\eta_{\max}$ is the initial learning rate, T_{cur} represents the current training epoch, and T_{max} is the total number of training epochs. This strategy expands the parameter search space during early training and promotes convergence and generalization through smooth decay in the later stage.

The model was evaluated using three metrics: Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Pearson correlation coefficient (p). Let the number of test samples

be n, the ground-truth weight be y_i, the predicted weight be $\hat{y}_i$, and their corresponding means be $\bar{y}$ and $\bar{\hat{y}}$. The metrics are then defined as follows:

$$\mathrm{MAE}(y,\hat{y}) = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{24}$$

$$\mathrm{MAPE}(y,\hat{y}) = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{25}$$

$$p = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}\sqrt{\sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})^2}} \tag{26}$$

3. Results

3.1 Point-Cloud Reconstruction Quality and Training Convergence Analysis

To evaluate the effectiveness of the mask-guided single-view 3D reconstruction strategy, qualitative visualization was first performed using sturgeon samples with different postures and viewing directions. As shown in Fig. 6, the original underwater images contained interference from tank boundaries, local reflections, and depth noise. After instance segmentation, the fish-body mask effectively constrained the target region and was fused with the depth map to generate fish-body point clouds. The reconstructed results show that the proposed strategy substantially suppresses background and environmental points, thereby producing clean single-view fish point clouds for subsequent feature learning and weight regression.

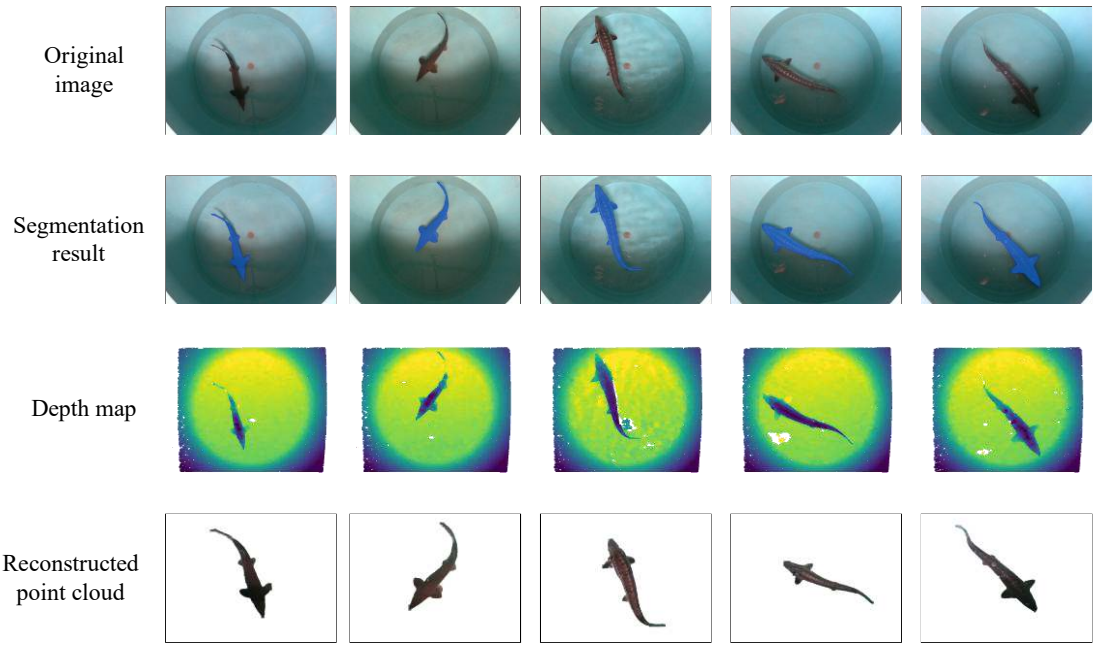
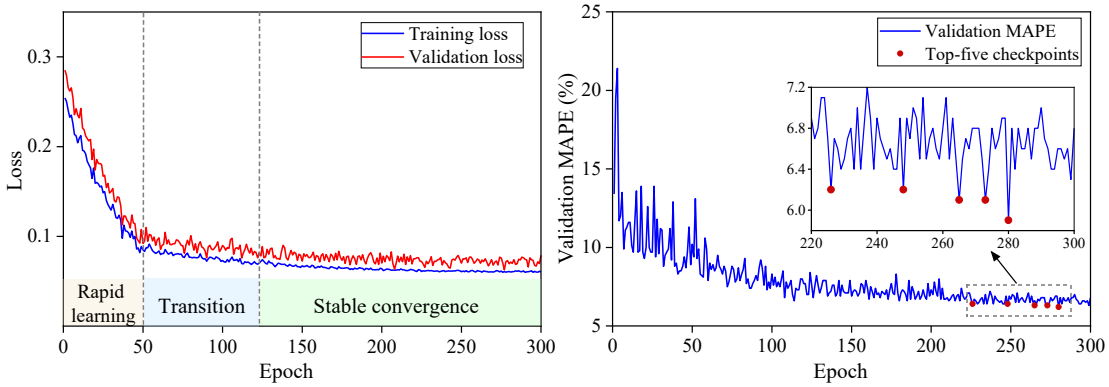


Fig. 6. Qualitative results of mask-guided single-view 3D reconstruction of underwater sturgeon. The rows show the original RGB images, segmentation results, depth maps, and reconstructed point clouds, respectively.

All experiments were implemented in PyTorch, and PPDS-Net was trained on a server equipped with an NVIDIA GeForce RTX 4090D GPU. To ensure stable and reproducible training, the initial learning rate was set to 1×10^{-3}, the batch size to 32, and the total number of training epochs to 300. A cosine annealing learning-rate scheduler with warm-up was used to help the model enter an effective optimization regime during the early stage of training and to suppress parameter oscillation through smooth decay in the later stage, thereby improving convergence quality and generalization performance.

As shown in Fig. 7(a), both training loss and validation loss decreased rapidly within the first 50 epochs and gradually stabilized after approximately 125 epochs. The two loss curves then continued to decline slowly and remained within a stable range without obvious divergence. These results indicate that PPDS-Net learned a stable mapping between single-view point-cloud morphology and sturgeon weight without severe overfitting. Under the cosine annealing learning-rate schedule, the model maintained a favorable balance between exploration and convergence, confirming the suitability of this training strategy for small-sample point-cloud regression.



(a) Training and validation loss curves (b) Validation MAPE and top-five checkpoints

Fig. 7. Training convergence and top-five checkpoint stability analysis. (a) Training and validation loss curves. (b) Validation MAPE curve and distribution of the top-five checkpoints. The inset magnifies the late-stage stable region containing the top-five checkpoints.

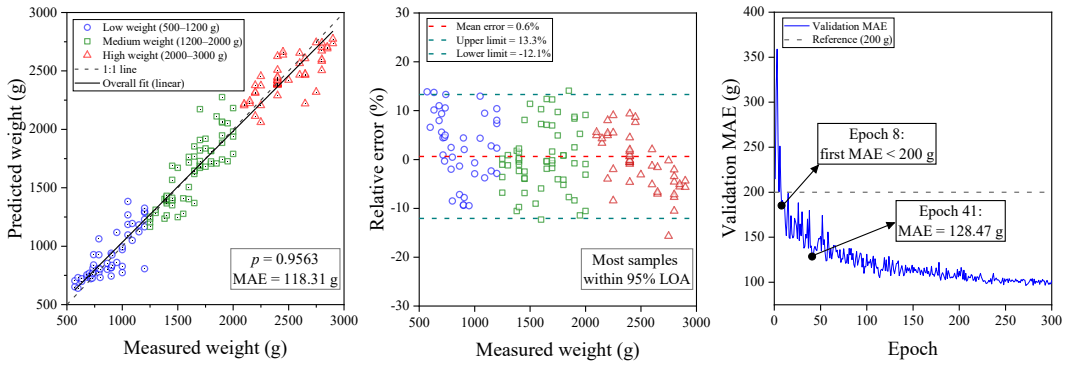
To verify convergence stability and rule out the possibility that the best checkpoint resulted from accidental fluctuation, the validation MAPE curve and the top-five validation checkpoints were analyzed, as shown in Fig. 7(b). The validation MAPE fluctuated markedly during the early training stage and then gradually decreased as training proceeded. After approximately 220 epochs, the curve entered a relatively stable low-error region. The inset further shows that the top-five checkpoints were distributed within this late-stage stable region rather than concentrated at a single isolated fluctuation point. These results indicate that the model did not converge to an isolated accidental optimum. Instead, it reached a stable convergence state during late training, demonstrating the reliability of the training configuration.

3.2 Weight Prediction Results

On the independent test set, PPDS-Net achieved high predictive accuracy and stable performance in sturgeon weight estimation. The final model obtained an MAE of 118.31 g and a

Pearson correlation coefficient p of 0.9563, indicating that the proposed method effectively recovered the nonlinear relationship between sturgeon weight and single-view point-cloud morphology under complex underwater conditions. To further evaluate model performance across different weight ranges, prediction agreement, relative error distribution, and validation MAE convergence were analyzed, as shown in Fig. 8.

As shown in Fig. 8(a), most predictions were distributed close to the 1:1 reference line across a weight range of approximately 500–3000 g, indicating high agreement between predicted and measured values. The overall fitted line also remained close to the ideal prediction line, and no obvious one-sided bias was observed among low-, medium-, or high-weight samples. These results show that the proposed model maintained stable prediction performance across different weight groups. Compared with conventional fitting methods based on a single geometric quantity, the proposed approach uses physical priors to mitigate the effects of nonrigid posture, local occlusion, and point-cloud sparsification on scale estimation, thereby improving the representation of fish-body scale, spatial extension, and posture variation. Fig. 8(b) further shows the distribution of relative errors across measured weight. The mean relative error was 0.6%, close to zero, indicating no clear systematic overestimation or underestimation. Most samples remained within the upper and lower error limits of 13.3% and −12.1%, respectively, and the relative errors did not show a continuous increase or directional shift with body weight. These results indicate that the model achieved not only high correlation but also stable error agreement across the tested weight range. Fig. 8(c) shows the convergence process of validation MAE during training. The validation MAE decreased rapidly in the early stage, first fell below 200 g at epoch 8, and reached 128.47 g at epoch 41 before gradually entering a stable region. These results indicate that the physical-prior-guided dual-stream feature-fusion mechanism improves not only final prediction accuracy but also early-stage optimization efficiency by reducing dependence on extensive high-dimensional parameter search.



(a) Measured and predicted weight agreement (b) Relative Error Agreement Analysis (c) Convergence of Validation MAE

Fig. 8. Weight prediction performance and error agreement analysis of PPDS-Net. (a) Measured versus predicted sturgeon weight across different weight groups. The dashed line denotes the ideal 1:1 relationship, and the solid line represents the overall linear fit. (b) Relative error agreement between measured and predicted weight. The red dashed line indicates the mean bias, and the teal dashed lines indicate the upper and lower limits of agreement. (c) Validation MAE convergence during training.

3.3 Comparative and Ablation Experiments

Comparative and ablation experiments were conducted from five perspectives to evaluate the effectiveness of PPDS-Net and the contribution of its key components: traditional geometric regression baselines, local geometric aggregation and physical-prior features, normal-vector input, robustness to point-cloud sparsity, and computational efficiency compared with mainstream point-cloud networks.

3.3.1 Comparison with Traditional Geometric Regression Baselines

To verify the advantage of end-to-end point-cloud learning over traditional geometric fitting, the proposed method was compared with three handcrafted geometric regression baselines: perimeter-based linear regression, surface-area-based power regression, and length–area multivariate regression.

The three traditional geometric regression models showed limited prediction accuracy. Perimeter-based linear regression produced the largest error, with a MAPE of 32.87%, an MAE of 455.17 g, and a p of 0.5980. Surface-area-based power regression slightly improved the result, reducing MAPE to 25.08% and MAE to 395.61 g, with p increasing to 0.6501. The length–area multivariate regression model achieved the best performance among the traditional baselines, with a MAPE of 23.32%, an MAE of 343.38 g, and a p of 0.7853. Nevertheless, its prediction error was still much higher than that of PPDS-Net, which achieved a MAPE of 7.70%, an MAE of 118.31 g, and a p of 0.9563. This performance gap arises because traditional methods first compress the 3D point cloud into low-dimensional handcrafted descriptors, such as perimeter, surface area, and body length, and then approximate the weight relationship using linear, power-function, or multivariate regression. Although this strategy can capture coarse allometric trends, it loses information on local thickness distribution, volumetric expansion, and spatial structure under bent postures. In contrast, PPDS-Net learns richer three-dimensional morphological representations directly from raw point sets, resulting in higher accuracy and stability for sturgeon weight prediction.

3.3.2 Ablation Analysis of Local Geometric Aggregation and Multidimensional Physical Priors

To verify the independent contributions and joint effects of local geometric aggregation and 4D macroscopic physical-prior features, a progressive ablation study was conducted. The baseline model was an MLP architecture with only global point-wise processing capability. The local geometric aggregation module and different physical priors were then introduced progressively. The results are presented in Table 1.

Table 1 Ablation Results of Local Geometric Aggregation and Multidimensional Physical Priors.

ID	Input	Aggregation	MAE (g)	MAPE (%)	p
A	XYZ	No	286.27	19.42	0.7334
B	XYZ	Yes	232.60	10.95	0.8743

C	XYZ+4D	No	218.47	13.94	0.8907
D(PPDS-Net)	XYZ+4D	Yes	118.31	7.70	0.9563
E	XYZ+F1	Yes	142.64	9.02	0.9339
F	XYZ+F2	Yes	140.23	8.96	0.9377
G	XYZ+F3	Yes	183.42	12.84	0.8932
H	XYZ+F4	Yes	149.02	9.44	0.9409

The results indicate strong complementarity between the local geometric aggregation module and the physical-prior features. The baseline model (A), which used only a global MLP without physical priors, produced the largest error, with an MAE of 286.27 g. Introducing either local aggregation (B) or 4D physical priors (C) alone substantially improved performance. Combining the two components in the full PPDS-Net configuration (D) further improved accuracy, reducing the MAE to 118.31 g and the MAPE to 7.70%, with a p of 0.9563. These findings indicate that the 4D physical priors provide a macroscopic biomass-related scale reference, whereas local geometric aggregation captures fine-grained structural information caused by individual variation and nonrigid postures. Both components are therefore necessary.

Further analysis of individual priors shows that, once local aggregation is available, F_1 and F_2 contribute most strongly to the accuracy improvement, suggesting that statistical scale information is closely related to body weight under fixed acquisition distance and viewing conditions. By contrast, F_3 and F_4 mainly provide complementary constraints related to spatial extent and topological length. The full PPDS-Net configuration (D) achieved the best performance, confirming the value of multidimensional physical priors for weight prediction in nonrigid targets such as sturgeon.

3.3.3 Effects of Normal-Vector Features and Physical Priors

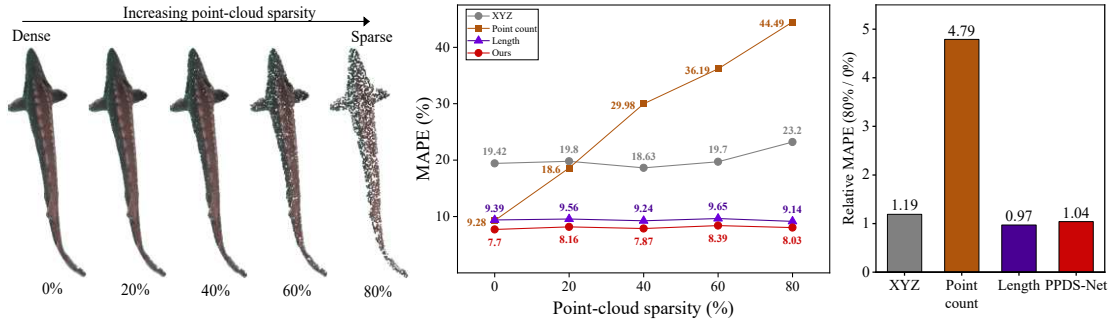
Normal vectors are commonly used in 3D point-cloud processing as explicit descriptors of local curvature and surface detail. To test whether this additional microscopic geometric information could further improve weight regression, normal-vector features were added to both the pure coordinate input (XYZ) and the optimal physical-prior configuration (XYZ + 4D).

Normal-vector features improved the pure XYZ baseline, reducing the MAE from 286.27 g to 189.63 g and the MAPE from 19.42% to 12.93%, confirming their usefulness for local curvature learning. However, when normal vectors were combined with the 4D physical priors, performance declined, and the MAE increased from 118.31 g to 141.96 g, while the MAPE increased from 7.70% to 8.97%. This result indicates that, under the current hardware and acquisition conditions, normal vectors do not provide an effective complement to the physical priors. Point clouds generated by the consumer-grade RealSense D455 still contain quantization noise and surface ripple artifacts. Because normal-vector estimation relies on local differential operations, it can amplify high-frequency noise and interfere with the learning of robust macroscopic physical features. Under the data-quality conditions of this study, the noise sensitivity introduced by normal vectors outweighs

the benefit of additional curvature information. Therefore, the final PPDS-Net model adopts the more robust XYZ + 4D configuration and discards normal-vector input.

3.3.4 Robustness Analysis under Point-Cloud Sparsification

In real underwater measurement environments, changes in target distance, water turbidity, local reflection interference, and incomplete depth recovery can cause different degrees of point-cloud sparsification. To evaluate robustness under unstructured degradation, a random point-dropout strategy was used to construct degraded test sets with sparsity levels ranging from 0% to 80%. As shown in Fig. 9(a), the fish point clouds became progressively sparser as the sparsity level increased, and local structure and surface continuity were gradually weakened.



(a) Progressive point-cloud sparsification (b) Robustness under point-cloud sparsification (c) MAPE ratio at 80% sparsity

Fig. 9. Robustness analysis under point-cloud sparsification. (a) Progressive sparsification of fish point clouds from 0% to 80% using random point dropout. (b) MAPE trends of different feature configurations under increasing point-cloud sparsity. (c) MAPE ratio at 80% sparsity relative to the original point-cloud condition.

The performance of four feature configurations was then compared: XYZ, Point count, Length, and PPDS-Net. As shown in Fig. 9(b), the feature configurations showed markedly different sensitivities to point-cloud sparsity. Point count achieved a relatively low MAPE of 9.28% under the original condition, indicating that point number can provide a useful scale cue under fixed acquisition settings. However, its error increased rapidly as sparsity increased, reaching 44.49% at 80% sparsity. Fig. 9(c) further shows that the MAPE of Point count at 80% sparsity was 4.79 times that under the original condition, indicating that absolute point count is highly sensitive to random point loss.

By contrast, Length remained relatively stable across sparsity levels, with MAPE ranging from 9.14% to 9.65%, showing much higher stability. The XYZ baseline showed consistently higher errors, with MAPE increasing from 19.42% to 23.20%, indicating limited robustness when only coordinate input was used under sparse point-cloud conditions. PPDS-Net achieved the lowest MAPE at all sparsity levels, with errors remaining between 7.70% and 8.39%. Even at 80% sparsity, PPDS-Net maintained a MAPE of 8.03%, corresponding to a relative MAPE ratio of 1.04. These findings show that fusing statistical scale features with topological geometric features can prevent prediction collapse when a single information source fails, thereby improving the fault tolerance and deployment potential of the system in complex underwater environments.

3.3.5 Accuracy–Efficiency Comparison with Mainstream Point-Cloud Networks

To evaluate the feasibility of practical deployment in aquaculture, the proposed method was benchmarked against mainstream point-cloud deep learning architectures. All models were trained and tested on the same NVIDIA RTX 4090D hardware platform with identical hyperparameter settings. Beyond final prediction accuracy, the evaluation included first-epoch MAE, the number of epochs required to reach the 200 g error threshold, parameter count, GFLOPs, and single-sample inference time. The results are shown in Table 2.

Table 2 Comparison of Prediction Performance and Computational Efficiency of Different Point-Cloud Regression Networks.

Method	First-epoch MAE (g)	Epochs to 200 g	Params (M)	GFLOPs	Time (ms)	MAPE (%)	p
PointNet++(Qi et al., 2017b)	452.26	191	1.75	1.90	129.99	15.71	0.8277
PointConv(Wu et al., 2019)	508.39	116	19.56	1.18	62.39	12.17	0.9264
PCT(Guo et al., 2021)	527.65	Not reached	1.35	0.74	84.06	22.67	0.6909
DGCNN(Wang and Solomon, 2021)	504.74	109	1.92	4.33	83.89	15.28	0.8814
PointNext(Qian et al., 2022)	480.19	112	15.18	1.31	116.52	18.69	0.7640
PPDS-Net	241.58	8	1.82	1.90	113.28	7.70	0.9563

As shown in Table 2, PPDS-Net achieved the best overall balance among prediction accuracy, convergence speed, and computational cost. Mainstream point-cloud models generally produced large errors at the early stage of training and required many epochs to enter an effective optimization regime. By contrast, PPDS-Net reduced the first-epoch MAE to 241.58 g and lowered the error below 200 g by the 8th epoch, demonstrating markedly faster convergence. In terms of final prediction accuracy, PPDS-Net achieved the lowest MAPE of 7.70% and the highest p of 0.9563. This improvement indicates that the explicit 4D physical priors provide useful geometric inductive bias from the beginning of optimization and reduce blind search in the high-dimensional parameter space.

In terms of computational complexity, PPDS-Net added only 0.07 M parameters to PointNet++, while GFLOPs remained nearly unchanged. Despite this small overhead, the final MAPE decreased from 15.71% to 7.70%, indicating that physical priors substantially improved prediction quality without imposing a meaningful computational burden. Although PointConv had a slight advantage in inference time, its parameter size reached 19.56 M and its MAPE remained higher than that of the proposed method. Overall, the proposed method achieves higher accuracy and faster convergence while retaining a lightweight architecture, demonstrating strong practical value for

engineering applications.

4. Discussion

Estimating the weight of free-swimming fish from single-view point clouds is challenging because incomplete observation geometry, fish posture variation, and underwater imaging degradation weaken the correspondence between point-cloud representation and true body volume. The results show that PPDS-Net, by embedding four physical priors—raw point count, sampling-density ratio, spatial dispersion, and topological skeleton length—into a dual-stream point-cloud regression architecture, provides stable scale constraints beyond local geometric features. These constraints improve the consistency of body-volume perception. Therefore, for 3D phenotypic regression tasks with limited observations and small sample sizes, morphological priors that have a clear relationship with the target variable can improve both estimation accuracy and physical interpretability.

Compared with previous work, existing vision-based 3D weight-estimation methods for free-swimming fish can be grouped into two broad categories. The first category relies on explicit geometric measurements after binocular reconstruction and establishes statistical mappings using features such as perimeter and surface area. These methods are physically interpretable but usually require multi-step processing, including segmentation, depth recovery, 3D reconstruction, and geometric measurement; errors can therefore accumulate throughout the cascade. The second category includes end-to-end multimodal fusion methods, such as the RGB-D dual-stream Transformer framework (Wang et al., 2026). These methods use complementary texture and depth information to improve representation capacity, but their discriminative basis remains mainly data-driven and provides limited explicit constraints on fish morphological regularities. PPDS-Net addresses a more demanding single-view point-cloud input setting. It neither relies on low-dimensional regression after explicit geometric compression nor depends on additional modalities to improve robustness. Instead, it constrains point-cloud feature learning directly through physical priors, achieving a better balance between end-to-end representation capacity and physical consistency.

The experimental results also clarify the mechanisms underlying the performance of PPDS-Net. The comparison with traditional statistical regression shows that a small set of handcrafted geometric quantities cannot fully describe the complex volume variations of freely swimming fish. End-to-end point-cloud learning, by contrast, preserves richer 3D morphological information. The ablation experiments further show that local geometric aggregation and 4D physical priors provide complementary benefits: the former improves the representation of local morphological structures, whereas the latter supplies macroscopic constraints that are sensitive to global scale and robust to posture variation. The faster convergence and stable performance under high sparsity indicate that physical priors not only improve final accuracy but also reduce optimization uncertainty when observation quality degrades. Normal vectors did not provide additional benefit when combined with 4D physical priors, which suggests that, under the current depth-quality conditions, local

differential descriptors are more likely to amplify noise than macroscopic physical features directly related to body volume.

This study has several limitations. First, single-view observation inevitably causes information loss under self-occlusion, severe body bending, and multi-fish interference, and the current framework cannot fully eliminate this limitation. Second, the Intel RealSense D455 obtains depth information through conventional stereo matching. Under underwater refraction, scattering, and low-texture conditions, depth-map quality remains limited, which affects point-cloud reconstruction and is subsequently propagated to weight regression. Third, the physical priors incorporated into PPDS-Net are macroscopic morphological features associated with fish body volume rather than parameterized descriptors derived directly from biomechanical mechanisms. Their transferability across species, growth stages, and imaging conditions therefore requires further validation. Future research should improve underwater depth recovery, introduce temporal or multi-view supervision, and optimize physical-prior construction using larger cross-species datasets to enhance model generalization and application reliability.

5. Conclusion

This study proposed PPDS-Net, a physical-prior-guided dual-stream network for non-contact sturgeon weight estimation from single-view point clouds in complex underwater environments. The overall pipeline first obtains high-quality fish point clouds through mask-guided single-view 3D reconstruction, after which PPDS-Net performs weight regression through an implicit geometric stream and an explicit physical stream. Four physical priors, including raw point count, sampling-density ratio, spatial dispersion, and topological skeleton length, are incorporated into the end-to-end regression process. The experimental results show that PPDS-Net outperforms traditional geometric regression methods and several mainstream point-cloud networks in prediction accuracy, convergence speed, and robustness to sparse point clouds. These findings demonstrate the effectiveness of incorporating physical priors into deep point-cloud learning for underwater fish weight estimation.

Despite these findings, the proposed method remains constrained by the depth accuracy of consumer-grade RGB-D cameras and by fixed-range acquisition conditions. The generalization ability of PPDS-Net under more complex open-water environments, multi-scale variations, and stronger optical disturbances still requires further validation. Future work will focus on high-fidelity point-cloud reconstruction, robustness evaluation in open scenarios, and engineering deployment in real aquaculture environments to further improve the application potential of the proposed method in intelligent aquaculture.

Acknowledgments

The authors are grateful for the financial support provided by the National Natural Science Foundation of China and the Beijing Natural Science Foundation.

Author contributions

Rui Zhang: Writing – original draft, Formal analysis, Conceptualization. Shiyu Zhang: Writing – review & editing. Shuo Li: Formal analysis, Data curation. Yihang Lu: Data curation. Runbei Feng: Validation, Data curation. Kai Lin: Writing – review & editing, Funding acquisition.

Funding

This research was supported by the Research and Application of Healthy Aquaculture Techniques for Important Cultured Fish Species in Beijing [KJCX20261207], Young Backbone Teacher Support Plan of Beijing Information Science & Technology University [YBT202506], National Natural Science Foundation of China [32202996], Science and Technology Innovation Capacity Building Program of Beijing Academy of Agriculture and Forestry Sciences [KJCX20230814], Beijing Academy of Agriculture and Forestry Science Youth Science Foundation [QNJJ202401], and Beijing Natural Science Foundation [6232013].

Data availability

The datasets generated and/or analysed during the current study are available from the corresponding author on request.

Declarations

Competing interests

The authors declare no competing interests.

References

Badgujar, C.M., Poulose, A., Gan, H., 2024. Agricultural object detection with You Only Look Once (YOLO) Algorithm: A bibliometric and systematic literature review. Comput. Electron. Agric. 223, 109090. https://doi.org/10.1016/j.compag.2024.109090

Bello, S.A., Yu, S.S., Wang, C., Adam, J.M., Li, J., 2020. Review: Deep learning on 3D point clouds. Remote Sens. 12, 1729. https://doi.org/10.3390/rs12111729

Cao, J., Tagliasacchi, A., Olson, M., Zhang, H., Su, Z., 2010. Point cloud skeletons via laplacian based contraction. In: Proceedings of the 2010 Shape Modeling International Conference, 187–197. https://doi.org/10.1109/smi.2010.25

González-Sabbagh, S.P., Robles-Kelly, A., 2023. A survey on underwater computer vision. ACM Comput. Surv. 55, Article 268. https://doi.org/10.1145/3578516

Guo, M.H., Cai, J.X., Liu, Z.N., Mu, T.J., Martin, R.R., Hu, S.M., 2021. PCT: Point cloud transformer. Comput. Vis. Media 7, 187-199. https://doi.org/10.1007/s41095-021-0229-5

Hao, Y., Yin, H., Li, D., 2022. A novel method of fish tail fin removal for mass estimation using computer vision. Comput. Electron. Agric. 193, 106601. https://doi.org/10.1016/j.compag.2021.106601

Hartley, R., Zisserman, A., 2004. Multiple View Geometry in Computer Vision, 2nd ed. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9780511811685

Hou, Z., Huang, L., Zhang, Q., Miao, Y.S., 2023. Body weight estimation of beef cattle with 3D deep learning model: PointNet++. Comput. Electron. Agric. 213, 108184. https://doi.org/10.1016/j.compag.2023.108184

Kong, M.R., Li, B.B., Zhang, Y.H., Liu, C.H., Li, D.L., Duan, Q.L., 2024. Non-intrusive mass estimation method for crucian carp using instance segmentation and point cloud processing. Comput. Electron. Agric. 226, 109445. https://doi.org/10.1016/j.compag.2024.109445

Li, D.L., Yu, J.X., Du, Z.Z., Xu, W.K., Wang, G.X., Zhao, S.L., et al., 2024. Advances in the application of stereo vision in aquaculture with emphasis on fish: A review. Rev. Aquac. 16, 1718-1740. https://doi.org/10.1111/raq.12919

Loshchilov, I., Hutter, F., 2017. SGDR: stochastic gradient descent with warm restarts. In: International Conference on Learning Representations.

Loshchilov, I., Hutter, F., 2019. Decoupled weight decay regularization. In: International Conference on Learning Representations.

Meyer, G.P., 2021. An Alternative Probabilistic Interpretation of the Huber Loss. In: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 5257-5265. https://doi.org/10.1109/CVPR46437.2021.00522

Paudel, S., de Sousa, R.V., Sharma, S.R., Brown-Brandl, T., 2024. Deep learning models to predict finishing pig weight using point clouds. Animals 14, 31. https://doi.org/10.3390/ani14010031

Piegl, L., Tiller, W., 1997. B-Spline Basis Functions, In: Piegl, L., Tiller, W. (Eds.), The NURBS Book. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 47-79.

Qi, C.R., Su, H., Mo, K., Guibas, L.J., 2017a. PointNet: Deep learning on point sets for 3D classification and segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 77-85. https://doi.org/10.1109/CVPR.2017.16

Qi, C.R., Yi, L., Su, H., Guibas, L.J., 2017b. PointNet++: Deep hierarchical feature learning on point sets in a metric space. In: Advances in Neural Information Processing Systems, 5099-5108.

Qian, G., Li, Y., Peng, H., Mai, J., Hammoud, H.A.A.K., Elhoseiny, M., et al., 2022. PointNeXt: Revisiting PointNet++ with improved training and scaling strategies. In: Advances in Neural Information Processing Systems, 23192-23204.

Rusu, R.B., Marton, Z.C., Blodow, N., Dolha, M., Beetz, M., 2008. Towards 3D Point cloud based object maps for household environments. Rob. Auton. Syst. 56, 927-941. https://doi.org/10.1016/j.robot.2008.08.005

Wang, G., Hao, Y., Li, D., 2026. Multimodal estimation of fish weight via two-stream transformer and adaptive RGB-D fusion. Comput. Electron. Agric. 241, 111248. https://doi.org/10.1016/j.compag.2025.111248

Wang, G., Li, X., Yu, J.X., Xu, W.K., Akhter, M., Ji, S.Y., et al., 2024. Stereo matching and 3D reconstruction with NeRF supervision for accurate weight estimation in free-swimming fish. Comput. Electron. Agric. 225, 109255. https://doi.org/10.1016/j.compag.2024.109255

Wang, Y., Solomon, J.M., 2021. Object DGCNN: 3D object detection using dynamic graphs. In: Advances in Neural Information Processing Systems, 20745-20758.

Wu, W., Qi, Z., Fuxin, L., 2019. PointConv: Deep convolutional networks on 3D point clouds. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 9613-9622. https://doi.org/10.1109/CVPR.2019.00985

Yu, X.N., Wang, Y.Q., Liu, J.C., Wang, J., An, D., Wei, Y.G., 2022. Non-contact weight estimation system for fish based on instance segmentation. Expert Syst. Appl. 210, 118403. https://doi.org/10.1016/j.eswa.2022.118403

Zhang, L., Wang, J., Duan, Q., 2020. Estimation for fish mass using image analysis and neural network. Comput. Electron. Agric. 173, 105439. https://doi.org/10.1016/j.compag.2020.105439

Zhang, Z., 1999. Flexible camera calibration by viewing a plane from unknown orientations. In: Proceedings of the Seventh IEEE International Conference on Computer Vision, 666-673. https://doi.org/10.1109/ICCV.1999.791289

Zhao, Y., Qin, H., Xu, L., Yu, H., Chen, Y., 2024. A review of deep learning-based stereo vision techniques for phenotype feature and behavioral analysis of fish in aquaculture. Artif. Intell. Rev. 58, 7. https://doi.org/10.1007/s10462-024-10960-7

Zheng, H., Ma, C., Liu, D., Huang, J.D., Chen, R., Fang, C., et al., 2025. Weight prediction method for individual live chickens based on single-view point cloud information. Comput. Electron. Agric. 234, 110232. https://doi.org/10.1016/j.compag.2025.110232

Zhou, Q.-Y., Park, J., Koltun, V., 2018. Open3D: A modern library for 3D data processing. ArXiv abs/1801.09847, https://doi.org/10.48550/arXiv.1801.09847

Zion, B., 2012. Review: The use of computer vision technologies in aquaculture - A review. Comput. Electron. Agric. 88, 125–132. https://doi.org/10.1016/j.compag.2012.07.010