Method for the multi-view estimation of fish mass using a two-stage neural network with edge-sensitive module

Zeyu Jiao[1], Yingjie Cai[2], Qi Zhang[3*], Zhenyu Zhong[1]

(1. *Guangdong Key Laboratory of Modern Control Technology, Institute of Intelligent Manufacturing, Guangdong Academy of Sciences, Guangzhou 510070, China*;
2. *Department of Electronic Engineering, The Chinese University of Hong Kong, Hong Kong 999077, China*;
3. *College of Economics and Management, Northwest A & F University, Yangling 712100, Shaanxi, China*)

Abstract: The estimation of fish mass is one of the most basic and important tasks in aquaculture. Acquiring the mass of fish at different growth stages is of great significance for feeding, monitoring the health status of fish, and making breeding plans to increase production. The existing estimation methods for fish mass often stay in the 2D plane, and it is difficult to obtain the 3D information on fish, which will lead to the error. To solve this problem, a multi-view method was proposed to obtain the 3D information of fish and predict the mass of fish through a two-stage neural network with an edge-sensitive module. In the first stage, the side- and downward-view images of the fish and some 3D information, such as side area, top area, length, deflection angle, and pitch angle, were captured to estimate the size of the fish through two vertically placed cameras. Then the area of the fish at different views was estimated accurately through the pre-trained image segmentation neural network with an edge-sensitive module. In the second stage, a fully connected neural network was constructed to regress the fish mass based on the 3D information obtained in the previous stage. The experimental results indicate that the proposed method can accurately estimate the fish mass and outperform the existing estimation methods.

Keywords: fish mass, multi-view estimation, two-stage neural network, edge-sensitive module, image segmentation
DOI: 10.25165/j.ijabe.20241703.6840

Citation: Jiao Z Y, Cai Y J, Zhang Q, Zhong Z Y. Method for the multi-view estimation of fish mass using a two-stage neural network with edge-sensitive module. Int J Agric & Biol Eng, 2024; 17(3): 222–229.

1 Introduction

Timely and accurately estimating the mass of fish is an important task in fish production. In different growth stages of fish, it is necessary to rationally mix feed, control the feeding environment, and separate boxes to ensure the production of fish[1]. In traditional fish production, weighing fish one by one is very complicated and manually operated, which is neither practical nor economic. While statistical methods, such as tagged recapture, would reduce the requirements on the number of fish caught to a certain degree, the human cost is still large, and the mass of fish can only be estimated roughly inaccurately estimation. Therefore, a more convenient and accurate method to estimate fish mass is in urgent need by the fishery.

In recent years, the development of computer vision has brought a new dawn to fish production. The highly automated and non-invasive properties of computer vision methods make the identification, classification, and production estimation of fish free of manual operations, which have attracted extensive attention from both the academic and fish production industries[2]. According to the difference of input influence factors, the existing research can be divided into single- and multi-factor methods[3].

The single-factor methods are to couple the mass of the fish with one of the main characteristics of the fish, such as the length of the fish[4-8], the side area[9,10], etc. As early as 1904, Fulton[4] proposed 'Fulton's Condition Factor' and established $W=a\cdot L^b$ to represent the function between the length L and the mass W of fish. a and b are both constant parameters, which are determined by the species of fish and the growing environment, and their values are generally determined empirically. Based on Fulton's work, the researchers focused on how to obtain fish length and map the relationship between fish length and fish mass accurately. One of the most representative is the multi-view fish length estimation method proposed by Al-Jubouri et al.[11] They used front- and side-view cameras to synchronously get multi-view images of the fish. Then, by binarization of the side-view image, the position of the fish is separated from the background to obtain the contour of the fish, and then the length of the fish is obtained. Miranda and Romero[12] mainly focused on the length measurement method of swimming rainbow trout. By pre-processing the fish images, they approximated a third-order regression curve to the fish body to estimate a fish's length. Furthermore, considering that fish of the same length may have different heights, which can also significantly affect the mass of the fish, some scholars have tried to figure out the relationship between the side area and the mass of the fish. For instance, Hufschmied et al.[9] developed an automatic- and stress-free sorting device by means of digital image processing. A correlation between silhouette area and absolute body mass was applied to estimate the fish body mass with a relative average error of 5.5%. It is worth noting that the single-factor method is simple and efficient. However, the fish is moving in the complex 3D space, which may result in the 2D features of a fish, such as length and

Received date: 2021-06-18 **Accepted date:** 2021-09-13
Biographies: Zeyu Jiao, PhD, Associate Professor, research interest: expert systems, smart agriculture, Email: zy.jiao@giim.ac.cn; **Yingjie Cai**, PhD, research interest: deep learning, object detection, Email: caiyingjie@link.cuhk.edu.hk; **Zhenyu Zhong**, PhD, Professor, research interest: smart agriculture, Email: zy.zhong@giim.ac.cn.
***Corresponding author: Qi Zhang**, Associate Professor, research interest: digital agriculture, sustainable development. Room C416, School of Economics and Management, Northwest A&F University, Yangling 712100, Shaanxi, China. Tel: +86-18810077399, Email: qzhang@nwafu.edu.cn.

circumference, being affected by various factors such as morphology, position, attitude, etc., in 3D space. A single factor contains limited information, which is not enough to establish the relationship between a single factor and the mass of the fish theoretically.

To deal with the disadvantages of single-factor methods, some scholars have explored the relationship between multiple factors and fish mass, such as length, the circumference of contour, the side area, and the established several multi-factors fitting model to estimate the fish mass[13,14,15]. Viazzi et al.[13] established a Jade Perking S. Barcoo estimation system based on 2D computer vision, which builds up a relationship between the fish shape (including length, width, and side area) and mass. Then the training dataset is analyzed by regression to generate the best-matching model to accurately estimate the fish mass. Saberioon and Císař[14] obtained 8 different mass-related factors by using the infrared reflection system and estimated the fish mass based on support vector machine (SVM) and random forest (RF) algorithm. De Verdal et al.[15] utilized area, perimeter, length, height, and volume to characterize the mass information, and adopted partial least squares regression with a coefficient of determination to estimate the mass of very small sea bass larvae.

Compared with the single-factor methods[16,17], the multi-factor methods[18,19] factor in more information related to mass, which provides more basis for the accurate estimation of fish mass, but also brings some problems: how to choose the appropriate factors and how to determine the weight of different factors on the mass estimation. Although existing research has made great efforts how to select the characteristic factors of fish and mining conceal the relationship between these factors and the fish mass[20]. Nevertheless, the fish is moving in the complex 3D space, it is unconvincing to estimate the fish mass only by its attributes. 3D space information, e.g., position and posture, should also be taken into account.

Aiming to solve the above-mentioned problems, a two-stage model was proposed in this study that combines image segmentation and neural networks to non-invasively estimate fish mass from multi-view. First, real-time images of moving fish were captured by two vertically placed cameras simultaneously, and the fish in the side- and downward-view images were segmented by a pre-trained image segmentation model. Then, through the fusion of multi-view information, the position and attitude of the fish in 3D space are obtained. Finally, the fish mass can be predicted by feeding 3D information into the pre-trained fully connected neural network. In the first stage, the pre-trained image segmentation model is trained based on the manually annotated multi-view fish images, which are leveraged to extract the size attributes of the fish and its 3D information. In the second stage, the pre-trained neural network is constructed based on the 3D information and the corresponding fish mass, which is used to excavate the hidden connection between the 3D information and the mass. Experimental results show that the proposed method can achieve an average precision of 95.90% in the stage of fish feature extraction, and the mean absolute error of fish mass estimation is only 4.01%. Therefore, the contributions of this work can be summarized:

1) A method based on a two-stage neural network was proposed to accurately estimate the mass of fish;

2) Instead of simply extracting 2D information such as length, perimeter, and area, a multi-view feature extraction model was established to estimate the mass of fish more efficiently;

3) An edge-sensitive module was utilized to optimize the segmentation of fish edges, which makes the model perform better in the edge parts that are difficult to segment;

4) The proposed method is simple, easy, and suitable for actual fish production.

2 Materials and methods

2.1 Overall scheme

A multi-view fish estimation method based on a two-stage neural network with an edge-sensitive module was proposed in this study, which consists of three parts: data pre-processing, 3D information acquisition, and fish mass estimation. The overall scheme of the proposed method is shown in Figure 1.

2.1.1 Data pre-processing

This phase was mainly composed of two parts. One is to manually annotate the fish images collected in advance to generate the data set for training. Second, in the practical inference phase, the multi-view cameras collect real-time images of fish which are utilized to estimate the fish mass.

2.1.2 3D information acquisition

This phase corresponds to the first stage of the neural network. At this stage, the coarse position of the fish at multi-view was obtained by a pre-trained model on the training set. After further fine-tuning of the edge-sensitive module, the exact position of the fish in 3D space was acquired and can be characterized by relevant factors.

2.1.3 Fish mass estimation

This phase also includes two parts. First, the corresponding fish mass is replenished to the original training data set to form a new 3D information-mass training data set. On this basis, a fully connected neural network is established. Second, the real-time 3D information of fish is input into the neural network to estimate the fish mass.

2.2 Mass-related factors

Reviewing the existing research, the core of the fish mass m estimation methods is realized by the formula $m=\rho V$. For a certain type of fish, the density of fish ρ, can be considered as a constant within a certain range, which can be calculated from the training set. Therefore, the key to improving the accuracy of estimation is to have an accurate estimate of the volume V. Intuitively, shape features such as length, width, height, perimeter, and side area of a fish are all related to the volume of the fish to some extent. Meanwhile, it is worth pointing out that these shape features are also mutually correlated. Hence, Therefore, taking all shape features as input factors will lead to information duplication and information redundancy.

As shown in Figure 2a, assume that the position of the fish in 3D space can be represented by a cuboid. The volume of the cuboid can be calculated by Equation (1). Therefore, the volume of a fish can be approximately considered to be positively correlated with the product of the side area and the topside area divided by the length of the fish.

$$V = L \times W \times H = \frac{\text{Side}_{\text{area}} \times \text{Topside}_{\text{area}}}{L} \tag{1}$$

In addition to the aforementioned shape features that need to be taken into account, the 3D attitude of the moving fish will affect the shape feature in images, and the accuracy will be seriously influenced if attitude correction is not carried out. For the images captured by multi-view cameras, as shown in Figures 2b and 2c, the corrected side area is related to the side area in the side-view image and the deflection angle (i.e., β in Figure 2c) in 3D space, and the

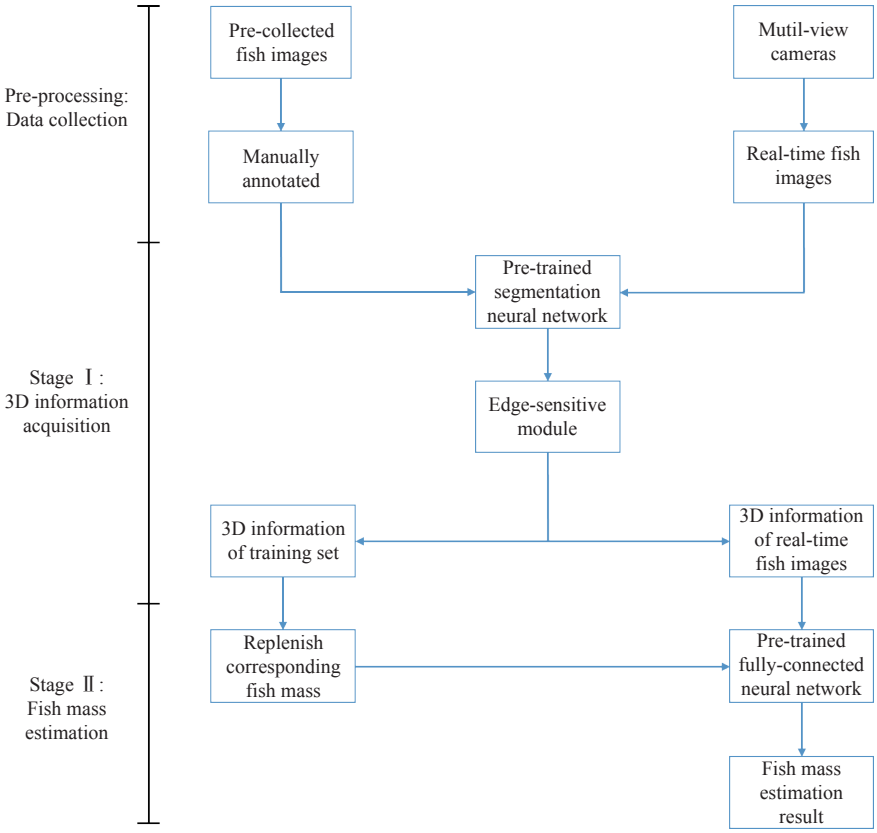
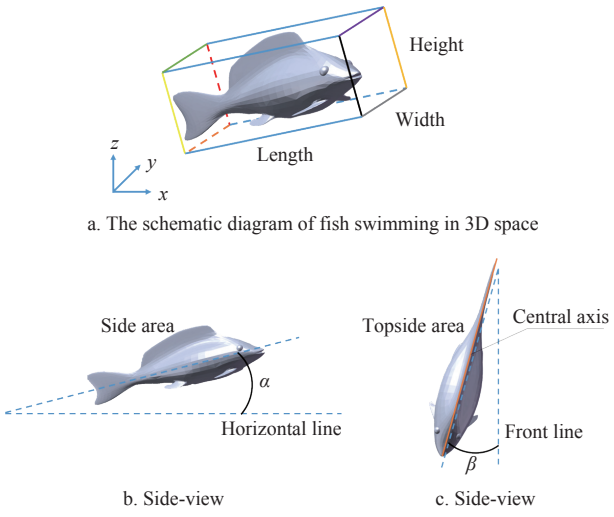


Figure 1 Overall scheme of the proposed method in this study



a. The schematic diagram of fish swimming in 3D space

b. Side-view c. Side-view

Note: α is the pitch angle, (°); β is the deflection angle, (°). Same below.

Figure 2 3D information of the swimming fish

corrected top area is related to the topside area in the downward-view image and the pitch angle (i.e., α in Figure 2b). Similarly, the true length of the fish can be obtained by aligning the length of the central axis of the fish in the downward-view image with the pitch angle.

In summary, the factors associated with fish mass in the multi-view images consist of five parts, including the side area in the side-view, the topside area and the length of the central axis in the downward view, the pitch angle α and the deflection angle β.

2.3 Experimental apparatus and data set

2.3.1 Experimental apparatus

The experimental apparatus was designed to weigh the fish non-invasively when they were swimming through a transparent pipe between two tanks, as shown in Figure 3. A centrifugal pump was deployed in Tank 2 for pumping water to ensure that the fish swim from Tank 1 to Tank 2 follow the water flow direction, as shown by the blue arrow. The interface size of the transparent pipe is 32 cm× 32 cm. The 2 cameras were placed vertically, both HIKVISION (China) webcams with 1/2.7′′ Progressive Scan CMOS sensors, which are connected to the remote model server through Wifi and can capture images with a resolution of 1280×720 pixels from different views at the same time. Note that, to ensure the accuracy of estimation, additional devices such as a sorter or bending pipe can be leveraged to make fish pass through the transparent pipeline one by one. Therefore, in this study, only the case of fish passing through the camera one by one is considered. To simplify the experiment, a fish tank of the same size was exploited as the transparent pipe to simulate the real experimental apparatus. It is important to note that the two vertically positioned webcams were pre-calibrated to ensure that images of the fish were captured simultaneously. In addition, by fixing the distance between the 2 cameras and the transparent pipe, the size of the image was proportional to the real size of the fish, which can be obtained by the camera's internal parameter matrix and external parameter matrix.

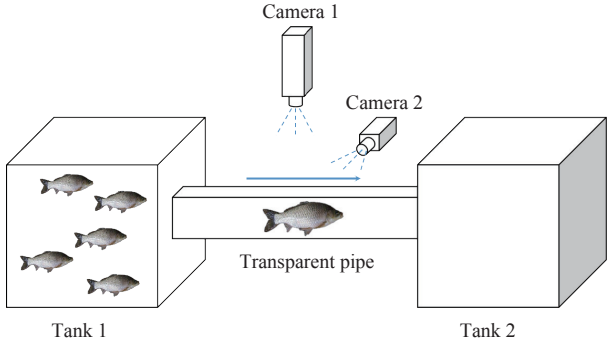


Figure 3 Schematic diagram of the experimental apparatus

2.3.2 Construction of the training set

This study took crucian carp as the experimental objective. Crucian carp is one of the most common and cultivated fish in China. Extensive farming and proper feeding of crucian carp to increase production make non-invasive fish mass estimation an urgent problem to be solved. In order to build a model to estimate the fish mass, it is necessary to construct a training set with sufficient samples to train a robust neural network. The flowchart of the construction of the training set is shown in Figure 4.

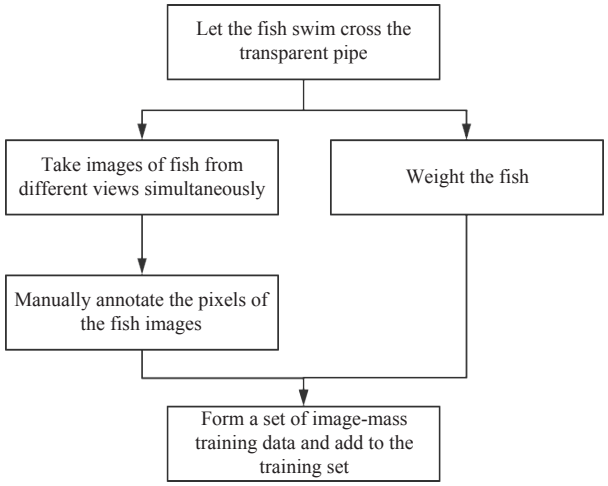


Figure 4 Flowchart of the construction of the training set

At first, the centrifugal pump in Tank 2 was leveraged to apply negative pressure, so that the water flows along the direction of Tank 1 to Tank 2, thus driving the fish to swim along the direction of the water flow. Secondly, the cameras deployed vertically along the transparent pipe capture images that contain the shape features and 3D attitude information of several instantaneous fish at the same time. Next, the fish mass was obtained by weighing the fish that had just been photographed. Finally, the training set was formed by annotating the collected multi-view images and coupling them with the measured mass. One set of data in the training set consists of 2 annotated images and 1 real weighted mass.

A total of 500 fish at different growth stages were taken into account in the construction of the training data set, with fish weights evenly distributed between 0.5 kg and 1.5 kg. Starting from 0.5 kg, there are 100 fish in each 0.2 kg interval. Furthermore, to simulate the actual swimming speed of the fish, the camera sampled at the frequency of 10 frames per second (fps) when collecting multi-view images. By changing the light intensity, different brightness images are added to the data set to enhance the model's robustness to the light. Meanwhile, images of the same fish swimming in different 3D attitudes are added to the training data set so that the model can effectively utilize 3D information from multi-view. Finally, there are 24 200 side- and downward-view images in the training set respectively, and 500 corresponding fish masses.

2.4 Proposed two-stage neural network

After completing the construction of the training set, what follows is how to extract the shape features and 3D attitude information of the fish, and how to establish the mapping relationship model between the 3D information and the fish mass. In the first stage, the neural network is mainly used to extract 3D information about fish, including Mask Region-based Convolutional Neural Networks (Mask R-CNN) with edge-sensitive module and 3D information acquisition. In the second stage, a fully connected neural network was constructed, and the 3D information obtained in the first stage was fed into the network to obtain the estimated fish mass by regression.

2.4.1 Mask R-CNN with edge-sensitive module

In order to obtain the pixel-level position of the fish in the images, image segmentation methods like Mask RCNN generally predict the fish in the low-resolution image, as a compromise between under-sampling and oversampling, and then continuously restore to the original image size through up-sampling. In this study, Mask R-CNN[21] was selected as the CNN backbone. In the process of up-sampling, bilinear interpolation and other methods can be utilized to remission the problem that the edge of the predicted region does not coincide with the actual edge to some extent, but the value inserted in the up-sampling process is only an approximation of the real value of the position. This inevitably leads to inaccurate predictions of the fish area in the images, which in turn reduces the accuracy of fish mass estimation.

Analogous sampling problems have been studied for decades in computer graphics while still pending. In image rendering, for multiple 3D objects, it is necessary to determine which objects are closer to the lens, and the renderings should be anti-aliasing. The core idea of common rendering is to calculate pixel values on an irregular subset of adaptively selected points on an image. The classic method[22] constructs a quadtree-like tree of "rays" and creates this tree for each pixel through the visible surface algorithm, which efficiently renders an anti-aliased, high-resolution image. In view of this, accurate edge prediction of image segmentation is analogous to the rendering, following the PointRend[23], an edge-sensitive module was built to act as a branch of the original image segmentation network, to solve the problem of inaccurate edge prediction of images, so as to get more accurate shape features of fish.

The architecture of the edge-sensitive module is illustrated in Figure 5. The edge-sensitive module consists of three components: 1) A point selection strategy was adopted to find the pixels that are likely to be the edges of the object; 2) A pixel-level feature extraction network was used to fuse features of different levels; 3) A multi-layer perceptron (MLP) was utilized to predict the label of the corresponding point.

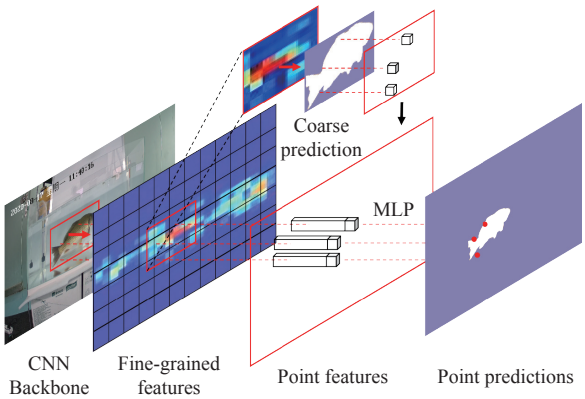


Figure 5 Illustration of the edge-sensitive module

1. Point selection strategy

Analogous to rendering, the edge is the most important thing to focus on in image segmentation. If all pixels are processed directly, it will consume a lot of computing resources and time. Hence, selecting appropriate points on the edges of fish and optimizing them during the up-sampling process is of great significance for more accurate image segmentation.

As shown in Figure 6, the output feature map of the CNN backbone is up-sampled to select N of the most uncertain (most likely on the edge) points on this denser feature map using adaptive subdivision[22]. Then, the pixel-level feature of the N points is calculated point-by-point to predict their labels. This process was repeated until the segmentation was sampled up to the desired resolution.

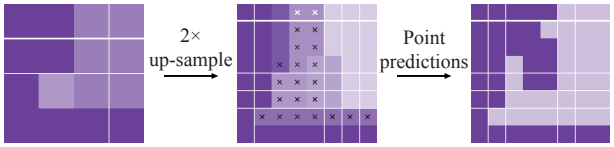


Figure 6 Schematic diagram of adaptive subdivision for edge-sensitive module

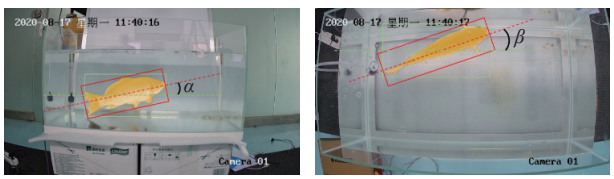
2. Pixel-level feature extraction and point predictions

The fine-grained features are obtained through CNN backbone, which contains the global information of the image, such as the shape and texture of the whole fish. However, as described before, it does not contain local region-specific information, which is a prerequisite for accurate edge recognition. Inspired by the fusion of multiple features with different resolutions in References [23] and [24], bilinear interpolation was leveraged to obtain high-resolution region of interests (RoI), and the features of points on edge were coarsely predicted, which contains more local features. By fusing fine-grained features with coarse prediction, a concatenated feature vector containing both global and local information is fed into a Multi-Layer Perceptron (MLP) to predict a more accurate label of the points. This layer-by-layer up-sample method predicts from coarse to accurate ensures that the predicted boundary can more accurately fit the edge of the fish, thus ensuring that the side area of the fish is more accurate than that of the original Mask RCNN architecture[21].

2.4.2 3D information acquisition.

Images collected by a single camera only contain the shape features of the fish, while images collected by cameras from multi-views at the same time can be exploited to extract 3D attitude information such as the pitching and deflection of the fish.

Based on the Mask RCNN architecture, the bounding box (green box in Figure 7) parallel to the pipeline can be output while getting the segmentation of the fish in the image. Meanwhile, the set of the fish's outline points can be considered as a convex roll, and the minimum area rectangle (red box in Figure 7) for the segmentation of fish can be obtained through the rotating calipers algorithm[25]. The pitch angle α can be obtained by calculating the angle between the two rectangles in the side-view image. The deflection angle β and the length L of fish can be calculated through the angle between the two rectangles and the length of the minimum area rectangle in the downward-view image respectively.



a. Side-view b. Downward-view

Note: The solid line is used to represent the peripheral outline of the fish, and the dashed line is used to represent the central axis of the fish.

Figure 7 Sample images of side- and downward-view

2.4.3 Fish mass estimation neural network

After taking images of the fish from multiple views, the shape features (Side_area, Topside_are, L) and the 3D attitude (α, β) of the fish were captured. With these fish mass-related factors obtained, the next step is to model the mapping between these factors and the fish mass. According to Equation (1), the fish mass is positively correlated with these five factors, which can be expressed as Mass $\propto$(Side_are, Topside_area, $1/L$, α, β). Here, a fully connected fish mass estimation neural network was constructed to realize the estimation of the fish mass. The network, as shown in Figure 8, is a multi-layer feed-forward model trained by the backpropagation algorithm, which has strong nonlinear mapping ability. According to Kolmogorov's theorem[26], the fish mass estimation neural network consists of three parts: input, hidden, and output layer. Since there are 5 input factors, the number of hidden layer nodes is 10, and the output is the predicted fish mass.

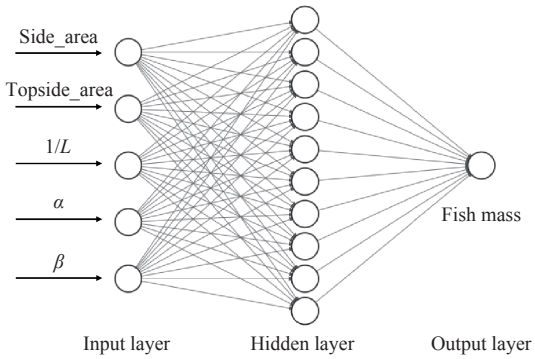


Figure 8 Structure of fish mass estimation neural network

3 Results and discussion

3.1 Experiment design

For the two-stage neural network with an edge-sensitive module proposed in this study, the performance of each stage network was evaluated separately.

3.1.1 Image segmentation validation experiment.

In this part, the experimental results of the first stage of the proposed method are given, including the difference between the predicted area and the manually annotated ground truth at different thresholds.

3.1.2 Edge-sensitive module comparison experiment

In this part, the proposed edge-sensitive module was evaluated independently to verify that the proposed module can indeed improve the fish's image segmentation results. These results indicate the 3D information of the fish can be acquired accurately.

3.1.3 Fish mass estimation comparison experiment

In this part, the proposed method was compared with the existing methods, including PCA-CF-BPNN[3], Length-Weigh[5], Multiple-factor-Weight[13], and LinkNet[27].

Throughout the experiments, the fishes between 0.5 kg and 1.5 kg swim across the transparent pipe in different attitudes and are utilized to evaluate the performance of the proposed two-stage network. In the image segmentation validation experiment and edge-sensitive module comparison experiment, the parameters of each layer of the ResNet50 CNN backbone are pre-trained on ImageNet[28]. The initial learning rate is 10^{-3}, and the decay rate is set as 10^{-1} per 2000 iterations. In the fish mass estimation comparison experiment, the learning rate was set to 10^{-2}, the activation function of the hidden layer was the Rectified Linear Unit (ReLU) function and the activation function of the output layer was the linear

function. All the experiments are conducted on an Intel i7-6700 CPU at 4.0 GHz with 16 GB RAM and 8 Nvidia P100 GPU with 16 GB memory. The programming language was Python 3.6 and the integrated development environment was Anaconda 3.

3.2 Evaluation metrics

3.2.1 Image segmentation evaluation metrics

In the first stage, the Intersection-over-Union (IoU) is exploited to evaluate the performance of the image segmentation network. The IoU is calculated by Equation 2, which is a dimensionless parameter between 0 and 1, representing the degree of coincidence between the predicted area and the ground truth. When the predicted area coincides exactly with the ground truth, the IoU will be 1; otherwise, when there is no coincidence, the IoU will be 0. The greater the value of IoU, the more accurate the image segmentation results will be.

$$IoU = (P \cap G)/(P \cup G) \tag{2}$$

Following the evaluation metrics of MicroSoft COCO Detection Evaluation[29], the effect of the model under different thresholds is analyzed. Table 1 lists the metrics that indicate the performance of the image segmentation network. Since the pixels occupied by the smallest fish in the images are greater than 32^2, APsmall and ARsmall were not evaluated in the COCO standard Evaluation. Simultaneously, the fish pass through transparent pipes one by one driven by the centrifugal pump, so the values of maxDets=1, maxDets=10, and maxDets=100 are equal, so only the indicators were given under the condition of maxDets=1.

Table 1 Metrics that represent the performance of image segmentation network

Metrics		Criterion
AP	AP for all cases	AP at IoU=0.50:0.05:0.95 (primary challenge metric)
	APIoU=0.50	AP at IoU=0.50 (PASCAL VOC metric[30])
	APIoU=0.75	AP at IoU=0.75 (strict metric)
AP Across Scales	APsmall	AP for small objects: area<32^2
	APmedium	AP for medium objects: 32^2<area<96^2
	APlarge	AP for large objects: area>96^2
Average Recall (AR)	ARmaxDets=1	AR was given 1 detection per image
	ARmaxDets=10	AR was given 10 detections per image
	ARmaxDets=100	AR was given 100 detections per image
AR Across Scales	ARsmall	AP for small objects: area<32^2
	ARmedium	AR for medium objects: 32^2<area<96^2
	ARlarge	AR for large objects: area>96^2

3.2.2 Fish mass estimation evaluation metrics

According to Equations (3)-(5), the fish mass estimation results are evaluated by the mean absolute error (MAE), root mean square error (RMSE), and coefficient of determination (R^2).

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \tilde{y}_i| \tag{3}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \tilde{y}_i)^2} \tag{4}$$

$$R^2 = \frac{\left(\sum_{i=1}^{N}(y_i - \bar{y}_i)(\tilde{y}_i - \hat{y}_i)\right)^2}{\sum_{i=1}^{N}(y_i - \bar{y}_i)^2 \sum_{i=1}^{N}(\tilde{y}_i - \hat{y}_i)^2} \tag{5}$$

where, y_i is the ground truth mass of the fish, y_i is the predicted mass of the second-stage network, $\bar{y}_i$ is the mean ground truth mass, $\hat{y}_i$ is the mean predicted mass, and N is the number of fish in the test set.

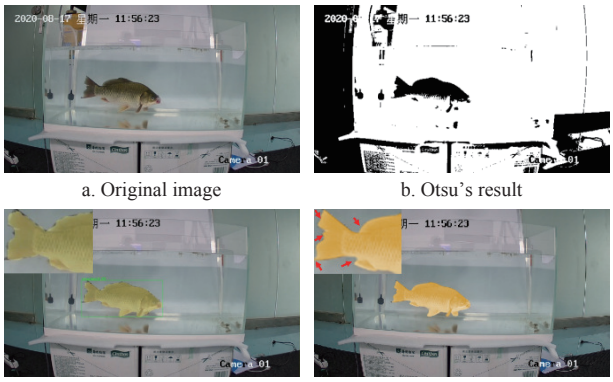
3.3 Experiment results

3.3.1 Image segmentation results

The experimental results of image segmentation are demonstrated in Table 2 and Figure 9 (Section 3.4) visualizes the image segmentation results of the Otsu, model without (w/o) edge-sensitive module and model with sensitive module.

Table 2 Experimental results of image segmentation

Metrics		w/o edge-sensitive	Proposed method in this study
AP	@[IoU=0.50:0.95 | area=all]	0.868	0.959
	@[IoU=0.50 | area=all]	1.000	1.000
	@[IoU=0.75 | area=all]	1.000	1.000
	@[IoU=0.50:0.95 | area=medium]	0.859	0.941
	@[IoU=0.50:0.95 | area=large]	0.870	0.962
AR	@[IoU=0.50:0.95 | area=all]	0.881	0.974
	@[IoU=0.50:0.95 | area=medium]	0.876	0.943
	@[IoU=0.50:0.95 | area=large]	0.883	0.975



a. Original image b. Otsu's result

c. Without edge-sensitive module d. With edge-sensitive module

Figure 9 Visualization results of different methods

3.3.2 Fish mass estimation results

The experimental results of fish mass estimation are demonstrated in Table 3, the proposed method in this study was compared with several existing methods. The experimental results indicate that the proposed two-stage network with an edge-sensitive module outperforms the existing methods. It should be noticed that the downward arrow indicates that the smaller the value, the better the model performs. The upward arrow indicates that the larger the value, the better the model performs.

Table 3 Comparison of different fish mass estimation methods

Methods	MAE	RMSE	R^2
PCA-CF-BPNN	0.3333	0.4057	0.0004
Length-Weight	0.2001	0.2312	0.6099
Multiple-factor-Weight	0.1497	0.1728	0.7367
Linknet	0.1260	0.1453	0.8102
w/o edge-sensitive of the proposed method in this study	0.0589	0.0684	0.9451
Proposed methods in this study with edge-sensitive	0.0401	0.0463	0.9738

3.4 Discussion

The experimental results show the performance of the proposed

two-stage model. In this subsection, the influence of different factors on the performance of the two-stage model was discussed in detail, including the edge-sensitive module, the fish mass-related factors, and the size of the fish.

3.4.1 The effect of Edge-sensitive module

As shown by Table 2, when the IoU threshold is less than 0.75, the absence of edge sensitive module has little impact on the performance of the image segmentation network, the predicted results can roughly cover the position of fish. As visualized in Figures 9c and 9d, while IoU=0.50:0.95 holds, the AP is reduced to 0.868 and the AR will be 0.881, which indicates that the Mask RCNN is not subtle enough in detail, especially in edge areas. The AP and AR at IoU=0.50:0.95 will be increased to 0.959 and 0.974 respectively with the adoption of the edge-sensitive module, which is mainly due to better optimization of the edge. This also leads to more accurate calculations of the side area and topside area of the fish.

3.4.2 The effect of fish mass-related factors

In this study, 5 mass-related factors are leveraged to extract the 3D information of the fish, while the existing methods generally use 1[5], 3[13], and 14[3] selected features to characterize fish mass, including Area, Perimeter, Aspect ratio, Ratio of equivalent ellipse axis, etc. Table 4 lists the influence of the number of selected features on the final mass estimation results. It should be noted that, in order to ensure that all methods are compared under the same experimental conditions, the selected features in these methods are extracted on the basis of our first stage image segmentation network, rather than the shape features extraction methods in these existing researches. The experimental results show that more features are not necessarily beneficial, actually, too many features may lead to more errors and thus affect the mass estimation results.

Table 4 Comparison of the number of selected features

Number of features	MAE	RMSE	R^2
1	0.2015	0.2316	0.6000
3	0.1501	0.1734	0.7375
14	0.0601	0.0694	0.9426
Proposed method in this study	0.0401	0.0463	0.9738

3.4.3 Limitation

It should be noted that the fish mass estimation involved in this study is based on visual methods. It is undeniable that when a large number of fish swim by at the same time, the performance of the model will be seriously affected by occlusion and misidentification. This problem can be addressed by placing a series of bends pipes or deploying a fish sorter in front of the mass estimator, which ensures that only one fish is in sight at a time. However, since this is only a problem that can be solved by engineering techniques, it will not be discussed in detail in this study.

3.5 Real case analysis

The method proposed in this study was deployed on a freshwater aquaculture farm in Qianjiang county, Hubei province, China. The fish mass estimation device deployed is shown in Figure 10. 2 cameras were respectively fixed in the box made of metal shell. The fish in the fishpass swim through the device one by one and the corresponding images are collected from multiple views.

Since weighing the fish is cumbersome, the method was tested in a 1-acre crucian carp pond containing 534 crucian carp with a total mass of 648.56 kg. It took 2897 s to weigh the fish one by one. Table 5 lists the difference in estimation error and time cost between the proposed method and the tagged recapture method.

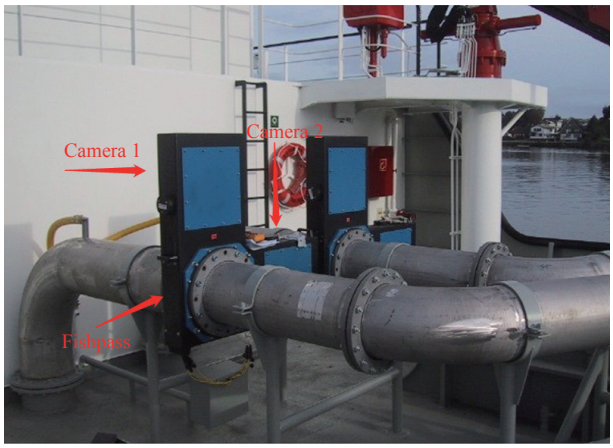


Figure 10 Schematic diagram of the real case

Table 5 Difference between the proposed method and the tagged recapture method

Index	Method	
	Tagged recapture	Proposed method of this study
Estimation mass/kg	436.05	671.67
Error/%	32.77	3.56
Estimated number	323	536
Total Time/s	1489	447
Operational details	Tagged 47 fish for the first time, totaling 60.16 kg 62 were caught again, with 9 tagged, totaling 86.80 kg Average mass: (60.16+86.80)/109≈1.35 kg	Diameter: 200 mm Water speed: 2.5-3.0 m/s Speed: 2-4 t/h

4 Conclusions and future work

A multi-view fish mass estimation method based on a two-stage neural network with an edge-sensitive module was proposed in this study, which is of great significance for scientific aquaculture and can be adopted for proper feeding, feeding environment control, and separate boxes in different growth stages of fish to ensure the production.

In the study, it was found that using multiple cameras to capture 3D information of fish, i.e., the area and corresponding angle of the fish in a certain view, outperforms relying solely on the external characteristics of the fish. This is mainly because the fish is swimming in 3D space, so the 2D image alone will lose a lot of information, including the thickness of the fish and the angle of the fish swimming, which may lead to an error in the mass estimation. In essence, the pitch angle and deflection angle mentioned are exploited to correct the attitude of the fish. With the help of this angle information, the fish can be normalized to the same attitude and angle, thus reducing the influence of swimming on the mass estimation.

In addition, by using the edge-sensitive module, the ability of the original Mask R-CNN frame to process the edge details of the image has been significantly improved, which is proved by the numerical results and visualization results of the experiment. The optimization of the edge makes the appearance of the fish more accurate, that is, the side area and top area of the fish can be obtained more accurately. Since the acquired area is the basis of the second phase mass estimation, small errors may be magnified in the second phase, which also makes the edge-sensitive module indispensable in the actual deployment. The ablation experiment

attempts to remove the edge-sensitive module and reveal that it does affect the overall performance of the model.

A case study in a real fishery scene fully demonstrates that the method proposed in this study can accurately estimate the quality of fish and save a lot of labor and time costs in practice. Despite all the efforts, the proposed method can only be applied in specific lighting, angles, and other environments, and the method with strong generalization is the main focus and direction of future work.

Acknowledgements

This research was funded by Guangdong Provincial Natural Science Foundation General Project (Grant No. 2023A1515011700), GuangDong Basic and Applied Basic Research Foundation (Grant No. 2022A1515110007), the Guangdong Provincial Natural Science Foundation General Project (Grant No. 2023A1515012869), GDAS' Project of Science and Technology Development (Grant No. 2022GDASZH-2022010108).

[References]

[1] Zhao J, Bao W J, Zhang F D, Ye Z Y, Liu Y, Shen M W, et al. Assessing appetite of the swimming fish based on spontaneous collective behaviors in a recirculating aquaculture system. Aquacultural Engineering, 2017; 78(PartB): 196–204.

[2] Zion B. The use of computer vision technologies in aquaculture - A review. Computers and Electronics in Agriculture, 2012; 88: 125–132.

[3] Zhang L, Wang J P, Duan Q L. Estimation for fish mass using image analysis and neural network. Computers and Electronics in Agriculture, 2020; 173: 105439.

[4] Fulton T W. The rate of growth of fishes. Twenty-second Annual Report, 1904; 3(3): 326–446.

[5] Sanchez-Torres G, Ceballos-Arroyo A, Robles-Serrano S. Automatic measurement of fish weight and size by processing underwater hatchery images. Engineering Letters, 2018; 26(4): 461–472.

[6] Froese R, Tsikliras A C, Stergiou K I. Editorial note on weight-length relations of fishes. Acta Ichthyologica et Piscatoria, 2011; 41(4): 261–263

[7] Froese R, Thorson J T, Reyes R B. A Bayesian approach for estimating length-weight relationships in fishes. Journal of Applied Ichthyology, 2013; 30(1): 78–85.

[8] Venerus L A, Villanueva Gomila G L, Sueiro M C, Bovcon N D. Length-weight relationships for two abundant rocky reef fishes from northern Patagonia, Argentina: *Sebastes oculatus* Valenciennes, 1833 and *Pinguipes brasilianus* Cuvier, 1829. Journal of Applied Ichthyology, 2016; 32(6): 1347–1349.

[9] Hufschmied P, Fankhauser T, Pugovkin D. Automatic stress-free sorting of sturgeons inside culture tanks using image processing. Journal of Applied Ichthyology, 2011; 27(2): 622–626.

[10] Gümüş B, Balaban M O. Prediction of the weight of aquacultured rainbow trout (*Oncorhynchus mykiss*) by image analysis. Journal of Aquatic Food Product Technology, 2010; 19(3-4): 227–237.

[11] Al-Jubouri Q, Al-Nuaimy W, Al-Taee M, Young I. An automated vision system for measurement of zebrafish length using low-cost orthogonal web cameras. Aquacultural Engineering, 2017; 78(PartB): 155–162.

[12] Miranda J M, Romero M. A prototype to measure rainbow trout's length using image processing. Aquacultural Engineering, 2017; 76: 41–49.

[13] Viazzi V, Van Hoestenberghe S, Goddeeris B M, Berckmans D. Automatic mass estimation of Jade perch *Scortum barcoo* by computer vision. Aquacultural Engineering, 2015; 64: 42–48.

[14] Saberioon M, Císař P. Automated within tank fish mass estimation using infrared reflection system. Computers and Electronics in Agriculture, 2018; 150: 484–492.

[15] de Verdal H, Vandeputte M, Pepey E, Vidal M-O, Chatain B. Individual growth monitoring of European sea bass larvae by image analysis and microsatellite genotyping. Aquaculture, 2014; 434: 470–475.

[16] Ault J S, Luo J G, A reliable game fish weight estimation model for atlantic tarpon (*Megalops atlanticus*). Fisheries Research, 2013; 139: 110–117.

[17] Costa C, Antonucci F, Boglione C, Menesatti P, Vandeputte M, Chatain B. Automated sorting for size, sex and skeletal anomalies of cultured seabass using external shape analysis. Aquacultural Engineering, 2013; 52: 58–64.

[18] Balaban M O, Chombeau M, Gümüş B, Cirban D. Determination of volume of alaska pollock (*Theragra chalcogramma*) by image analysis. Journal of Aquatic Food Product Technology, 2011; 20(1): 45–52.

[19] Wang W J, Xu J Y, Lyu Z M, Xin N H. Weight estimation of underwater *Cynoglossus semilaevis* based on machine vision. Transactions of the CSAE, 2012; 28(16): 153–157. (in Chinese)

[20] Odone F, Trucco E, Verri A. A trainable system for grading fish from images. Applied Artificial Intelligence, 2001; 15(8): 735–745.

[21] He K M, Gkioxari G, Dollár P, Girshick R. Mask R-CNN. In: 2017 IEEE International Conference on Computer Vision, 2017; pp.2961–2969.

[22] Whitted T. An improved illumination model for shaded display. In: Proceedings of the 6th annual conference on Computer Graphics and Interactive Techniques, 1979; 13(2): 807419.

[23] Kirillov A, Wu Y X, He K M, Girshick R. Pointrend: Image Segmentation as Rendering. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020; pp.9799–9808.

[24] Lin T Y, Dollár P, Girshick R, He K M, Hariharan B, Belongie S. Feature pyramid networks for object detection. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition, 2017; Honolulu: 2117–2125.

[25] Toussaint G T. Solving geometric problems with the rotating calipers. In: Proceeding of IEEE Melecon'83, Athens: IEEE, 1983; 8p.

[26] Kůrková V. Kolmogorov's theorem and multilayer neural networks. Neural Networks, 1992; 5(3): 501–506.

[27] Konovalov D A, Saleh A, Efremova D B, Domingos J A, Jerry D R. Automatic weight estimation of harvested fish from images. In: 2019 Digital Image Computing: Techniques and Applications (DICTA), Peth: IEEE, 2019; pp.1–7.

[28] Deng J, Dong W, Socher R, Li L J, Li K, Li F F. Imagenet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami: IEEE, 2009; pp.248–255.

[29] Lin T Y, Maire M, Belongie S, Hays J, Perona P, Ramanan D, Dollár P, Zitnick C L. Microsoft CoCo: Common objects in context. In: European Conference on Computer Vision (ECCV 2014), Springer, 2014; pp.740–755.

[30] Hoiem S K, Divvala J H. Hays, Pascal VOC 2008 challenge. In: PASCAL Challenge Workshop in ECCV, Citeseer, 2009.