

Article

Non-Contact Tilapia Mass Estimation Method Based on Underwater Binocular Vision

Guofu Feng *, Bo Pan and Ming Chen 

Key Laboratory of Fisheries Information, Ministry of Agriculture and Rural Affairs, Shanghai Ocean University, Hucheng Ring Road 999, Shanghai 201306, China; m210911562@st.shou.edu.cn (B.P.); mchen@shou.edu.cn (M.C.)
* Correspondence: gffeng@shou.edu.cn

Abstract: The non-destructive measurement of fish is an important link in intelligent aquaculture, and realizing the accurate estimation of fish mass is the key to the stable operation of this link. Taking tilapia as the object, this study proposes an underwater tilapia mass estimation method, which can accurately estimate the mass of free-swimming tilapia under non-contact conditions. First, image enhancement is performed on the original image, and the depth image is obtained by correcting and stereo matching the enhanced image using binocular stereo vision technology. And the fish body is segmented by an SAM model. Then, the segmented fish body is labeled with key points, thus realizing the 3D reconstruction of tilapia. Five mass estimation models are established based on the relationship between the body length and the mass of tilapia, so as to realize the mass estimation of tilapia. The results showed that the average relative errors of the method models were 5.34%~7.25%. The coefficient of determination of the final tilapia mass estimation with manual measurement was 0.99, and the average relative error was 5.90%. The improvement over existing deep learning methods is about 1.54%. This study will provide key technical support for the non-destructive measurement of tilapia, which is of great significance to the information management of aquaculture, the assessment of fish growth condition, and baiting control.

Keywords: image enhancement; binocular stereo vision; 3D reconstruction; mass estimation



Citation: Feng, G.; Pan, B.; Chen, M. Non-Contact Tilapia Mass Estimation Method Based on Underwater Binocular Vision. *Appl. Sci.* **2024**, *14*, 4009. <https://doi.org/10.3390/app14104009>

Academic Editors: Paweł Kielbasa, Anna Miernik and Akinniyi Akinsunmade

Received: 23 March 2024
Revised: 25 April 2024
Accepted: 6 May 2024
Published: 8 May 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The tilapia (*Oreochromis* sp.) has characteristics such as fast growth, easy cultivation, and rich nutritional content [1]. Accurate measurement of fish length information is crucial for estimating fish mass, bait control, and aquaculture sales [2]. Length and body weight are fundamental biological features of fish populations and can reflect individual physiological states [3]. Fish mass information is essential for various aquaculture management activities. Traditional fish mass measurement methods require capturing fish and measuring their length and weight, which can result in fish mortality, resource waste, and poor measurement consistency [4].

With the continuous development of computer vision technology, more and more researchers are investing in the field of underwater binocular vision. Non-contact fish mass estimation methods [5] have attracted increasing attention from researchers. Abinaya et al. [6] used deep learning techniques, specifically the YOLOv4 model, to detect and analyze fish head, body, and tail features. Completely visible fish (CVF) were identified by the segmentation analysis technique. The fish biomass was calculated from the estimated length using a length–mass relationship calibration curve. The method showed high accuracy and reliability in estimating fish biomass, especially in cryptic environments. Konovalov et al. [7] collected about 2500 images of golden eye bass and estimated their mass. Two instances of a convolutional neural network (CNN) were trained using LinkNet-34 architecture. A CNN that regresses the weight directly from the images was also trained to estimate the mass of the fish efficiently and automatically. That study demonstrated the

potential for automated mass estimation of golden eye bass using deep learning techniques. Fernandes et al. [8] designed the CVS to be able to efficiently distinguish the fish body from the background and fins, and the extracted fish body regions could be successfully used to predict fish weight and carcass weight. That study demonstrated the potential of utilizing computer vision and deep learning techniques for fast, accurate, and non-invasive measurements in the aquaculture industry. Yu et al. [9] improved the feature pyramid network (FPN) for Mask R-CNN and proposed an effective scheme for the accurate measurement of fish body length and body width in precision agriculture, which can achieve highly accurate measurement results in different contexts. But, in these methods, the fish were caught and placed on a platform. Such a process may cause a lot of stress to the fish, affecting its growth and even leading to its death [10]. Of course, there are researchers who conduct experiments underwater. Sanchez-Torres et al. [11] used a single underwater camera to acquire images in a controlled environment in order to shorten the measurement time and reduce fish stress. The length and mass of the fish were also estimated by image processing and regression analysis techniques. Saberioon et al. [12] developed an automated system using a Kinect as an RGB-D camera to collect depth map and top view images of 295 farmed bass of different sizes. The geometric features of the back of the fish and machine learning algorithms were utilized to estimate the mass of the fish. Hao et al.'s [13] study proposed an unsupervised fin removal method. The fish mass was estimated based on the area and the area squared after fully automated caudal fin removal with a maximum relative error of 8.46%. Although this method of mass prediction based on fish body area is efficient, it is not possible to accurately estimate the mass of the fish when its body plane is at an angle to the camera level.

This study utilized binocular stereoscopic vision technology to capture images of tilapia, eliminated the influence of uneven illumination on underwater images using the Retinex image enhancement algorithm, conducted stereo matching, segmented fish bodies using a segmentation platform, marked key point coordinates, and estimated the length of the tilapia using three-dimensional reconstruction technology. Five tilapia length–mass prediction models were established, and the one with the smallest error was selected through analysis and comparison, providing technical support for rapidly estimating the length and mass of freely swimming fish underwater. The main contributions of the work in this paper are as follows:

1. Firstly, a binocular image dataset of underwater tilapia is established, which can be used for binocular image stereo matching in combination with underwater image enhancement algorithm. It provides an effective basis for the three-dimensional reconstruction of tilapia.
2. Aiming at the problem of non-contact tilapia length estimation based on underwater binocular stereo vision, this paper proposes a tilapia length estimation method based on binocular vision, image segmentation, and key pixel marking, which is aimed at realizing efficient and low-cost non-contact tilapia length estimation.
3. A regression model for predicting the mass of tilapia body length–body mass relationship was developed. The body length was obtained by the body length estimation algorithm, and the mass of tilapia could be estimated by inputting the model. The experimental results showed that the method was highly reliable and could be used for non-contact mass estimation of underwater tilapia.

2. Materials and Methods

2.1. Experimental Setup

In this experiment, the image data were acquired from the aquaculture plant in the coastal aquaculture base of Shanghai Ocean University, and tilapia, which has a high economic and edible value, was selected to obtain the images. The underwater binocular vision system used a deepwater binocular camera model ZF-USB-02B10 (Weihai Zhifan Marine Equipment Technology Company Limited, Weihai, and China), with a pixel size of $3.75 \times 3.75 \times 10^{-6} \text{ mm}^2$, a frame rate of 30 fps, a resolution of 2560×960 pixels, a focal

length of 3.0 mm, a baseline of 60 mm, and a horizontal fixation, which was connected to a laptop computer via USB 2.0. The processor of the laptop was Intel i5-9300H 2.4 GHz; the memory was 16 G; the graphics card was GTX-1660Ti; the operating system was Windows 11; and the programming environment was Python3.8. The experimental environment is as shown in Figure 1, and the tilapia were swimming freely in a white culture tank with a size of 120 cm × 90 cm × 34 cm. The weighing was carried out using an electronic scale with an accuracy of 1 g and connected to a DC power supply.

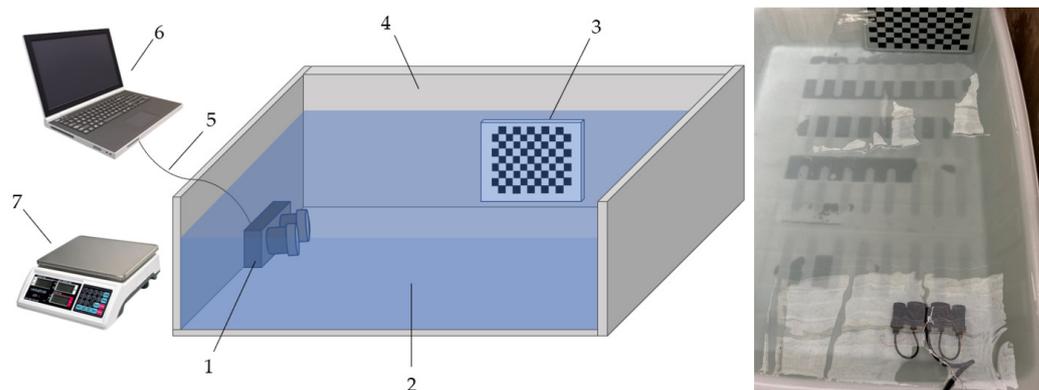


Figure 1. Experimental equipment and materials: 1, binocular camera; 2, breeding water; 3, calibration board; 4, breeding box; 5, cables; 6, laptop; 7, electronic scale.

2.2. Data Acquisition

When acquiring data, the specific steps were as follows: Water was drawn directly from the breeding pool into a white breeding box, and the tilapia were scooped out and placed on a flat surface. Their body length was measured with a tape measure once they ceased to move. The body length of the tilapia was defined as the straight-line distance from the tip of the snout to the base of the tail fin. Afterward, the tilapia, whose length had been measured, were placed on a container weighing m_1 , and the total weight was determined with an electronic scale to be m_2 . Thus, the mass of the tilapia was as follows:

$$m = m_2 - m_1 \quad (1)$$

To minimize errors, the same sample was measured three times, and the average value was taken. This study acquired the body length and mass information for approximately 50 tilapia specimens. These samples were divided into three different weight groups: M1 (0–200 g), M2 (200–400 g), and M3 (400–800 g) for analyzing tilapia at different stages of growth. Additionally, 10 tilapia from the samples were sequentially placed in the breeding box for photography. The computer software AMCap 9.08 (build 63.4) was used to acquire and filter out 1000 unobstructed images of the tilapia, and during the image acquisition process, the binocular camera was continuously moved to capture images of the swimming fish from multiple angles.

2.3. Image Processing and Analysis

The main flow of this experiment is shown in Figure 2. Firstly, underwater camera calibration was performed to obtain the parameters of the camera for aberration correction. Then, the acquired images were processed in a division of labor. On the one hand, the underwater image was enhanced as a way to eliminate some effects of uneven illumination. The enhanced image was then stereo-matched to calculate the disparity value and obtain the depth information of the image. On the other hand, the acquired data were segmented, and the purpose of segmentation was to facilitate the labeling of key points in the next step. After labeling, the pixel body length of the tilapia could be calculated. Then, the depth information and pixel body length were combined to reconstruct the tilapia in three dimensions, which was used to calculate the actual body length of the tilapia. Finally,

the relationship between the body length and the mass of tilapia was modeled. The mass information of the tilapia was obtained by inputting the estimated body length of the tilapia.

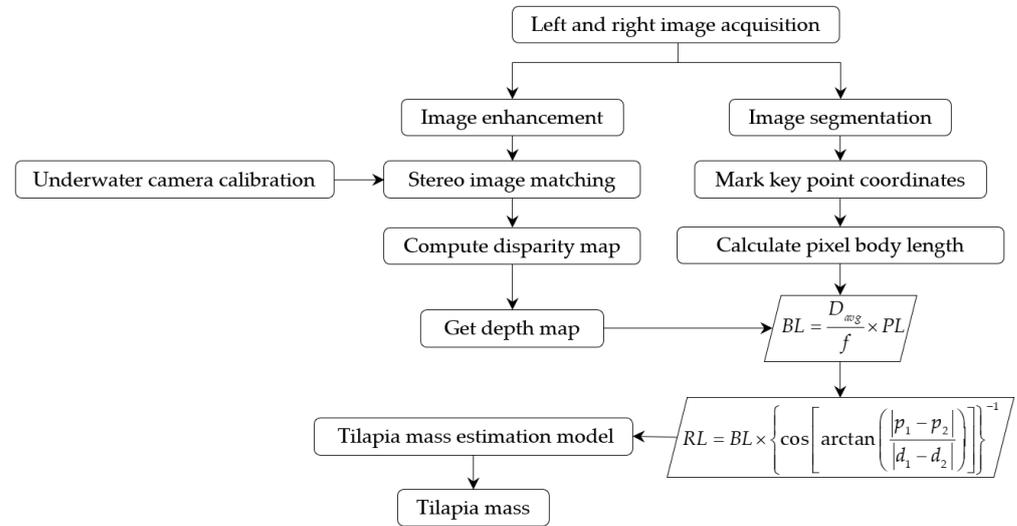


Figure 2. Flowchart for mass estimation of tilapia based on underwater binocular vision.

2.3.1. Camera Calibration

The refraction of light when shooting in underwater environments can lead to image distortion. To ensure the accuracy of the results, the binocular camera requires binocular calibration and distortion correction underwater [14]. The ZHANG’s method [15,16] was used, and a 12 × 9 square grid aluminum substrate was selected for the calibration plate. The size of each square was 30 mm × 30 mm.

Keeping the position of the camera unchanged, the position of the calibration plate was constantly changed, and several groups of images with different angles and positions were captured. The relative position of the binocular camera to the calibration board is shown in Figure 3. Forty of these images were selected and programmed to be split into separate left and right viewpoint maps and stored in separate folders. The left and right viewpoint maps were automatically calibrated using tools in Matlab2017b. The corner point detection results of the calibration board are shown in Figure 4. The optical aberrations produced by the images can be corrected using Equations (2)–(4).

$$\begin{cases} X_1 = x(1 + k_1t^2 + k_2t^4 + k_3t^6) \\ Y_1 = y(1 + k_1t^2 + k_2t^4 + k_3t^6) \end{cases} \quad (2)$$

$$\begin{cases} X_2 = x + [2p_1xy + p_2(t^2 + 2x^2)] \\ Y_2 = y + [2p_2xy + p_1(t^2 + 2y^2)] \end{cases} \quad (3)$$

$$t = \sqrt{\left(\frac{x}{f_x}\right)^2 + \left(\frac{y}{f_y}\right)^2} \quad (4)$$

where (x,y) are the coordinates of the original image. $X_1, X_2, Y_1,$ and Y_2 are the coordinates of the corrected image. f_x and f_y are the focal lengths in each direction. k_1 and k_2 are correction coefficients for radial distortion, and p_1 and p_2 are correction coefficients for tangential distortion. t is the distance from the image pixel point to the center of the image.

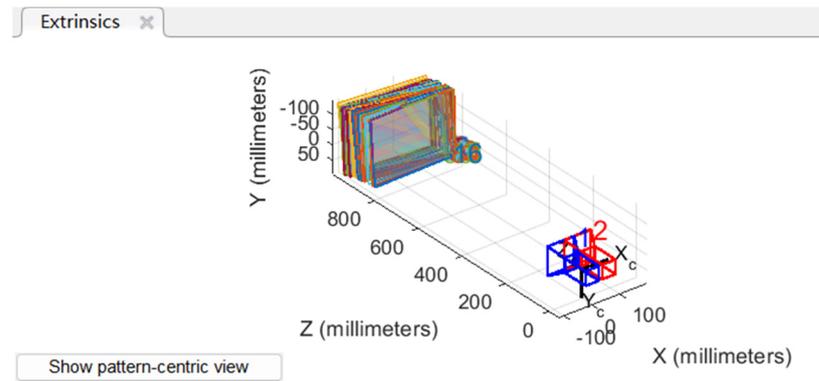


Figure 3. Relative position of binocular camera to calibration board.

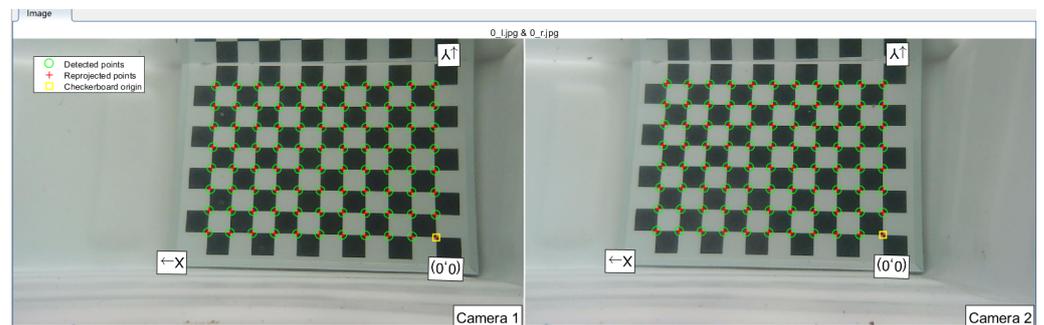


Figure 4. Calibration board corner point detection.

The histogram of the reprojection error of the Matlab2017b first calibration results is shown in Figure 5a. The maximum error in the first calibration reached 0.15, and it can be seen that several images had a large impact on the error. In order to further improve the subsequent estimation accuracy, we removed some image groups with large errors. The histograms of the reprojection errors after removal are shown in Figure 5b, and the errors were all below 0.1.

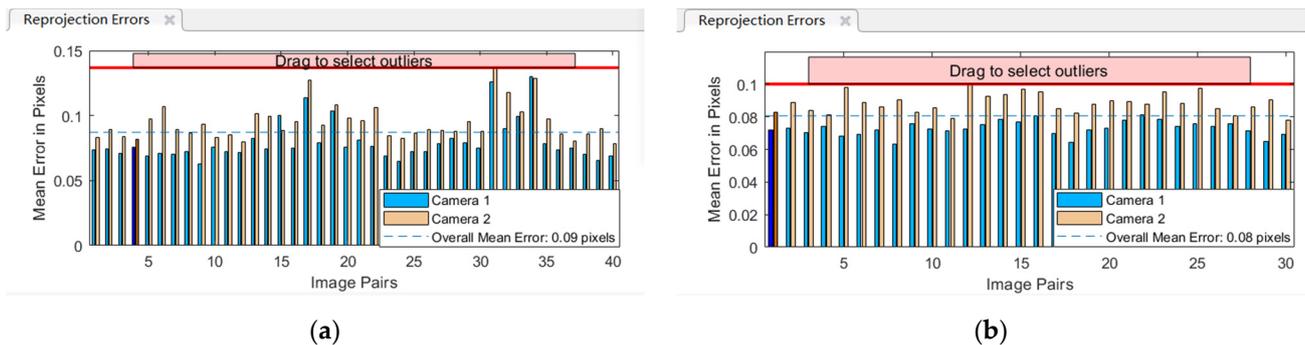


Figure 5. Comparison of reprojection error before and after processing: (a) histogram of reprojection error for first calibration; (b) histogram of reprojection error after removal of large errors.

The calculated camera parameters are shown in Table 1. It can be seen that the rotation matrix R approximates the unit matrix. The first parameter of the translation vector T represents the distance between the centers of the two cameras. The default parameter for the binocular camera baseline was 60 mm, which is within the tolerance of the error.

Table 1. Camera calibration results.

Parameter	Left Camera	Right Camera
Internal parameters	$\begin{pmatrix} 1757.9769 & 0 & 0 \\ 0 & 1739.4673 & 0 \\ 691.2410 & 350.4650 & 1 \end{pmatrix}$	$\begin{pmatrix} 1755.6439 & 0 & 0 \\ 0 & 1737.4004 & 0 \\ 639.7740 & 341.6467 & 1 \end{pmatrix}$
Radial distortion	$k_1 = 0.2938$	$k_1 = 0.2956$
	$k_2 = 0.7482$	$k_2 = 0.7320$
Rotation matrix	$R = \begin{pmatrix} 0.9999 & 0.0005 & -0.0037 \\ -0.0005 & 0.9999 & -0.0024 \\ 0.0037 & 0.0024 & 0.9999 \end{pmatrix}$	
Translation matrix	$T = [-60.0704 \ 0.0448 \ -0.2057]$	

2.3.2. Image Enhancement and Stereo Matching

Because of the uneven illumination of the underwater images obtained from the experiment, this study preprocessed the images before stereo matching. A Retinex-based image enhancement algorithm was used to enhance the underwater images, and the final image was obtained to eliminate the uneven illumination and retain the nature of the fish itself. In this study, several mainstream image enhancement methods were selected for comparison, including single-scale Retinex (SSR), multi-scale Retinex (MSR), and multi-scale Retinex with color restoration (MSRCR). A comparison of the underwater enhancement results is shown in Figure 6.

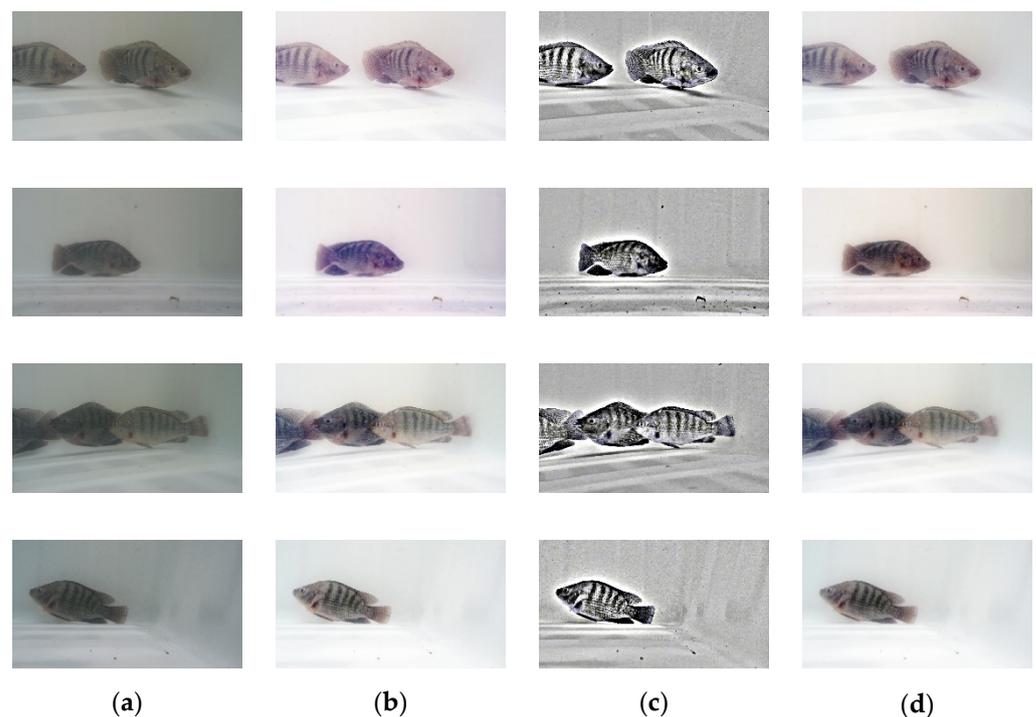


Figure 6. Underwater image enhancement results based on Retinex: (a) original image; (b) SSR; (c) MSRCR; (d) MSR.

Quantitative analysis is needed for images processed by underwater image enhancement methods. In this study, two commonly used image quality evaluation metrics, peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM), were used to quantitatively analyze and compare the performance of different algorithms.

In the application of underwater image enhancement, PSNR can be used to determine whether the enhanced image is over-enhanced. Ideally, the enhanced image should be

closer to the reality of the original scene rather than producing unnatural enhancement. A higher PSNR value means that the difference between the enhanced image and the original image is smaller, which usually indicates that the image is of better quality and does not appear to be over-enhanced. The formula for calculating the PSNR is as follows:

$$PSNR = 10 \cdot \log_{10} \left(\frac{MAX_A^2}{\frac{1}{mn} \sum_{i=0}^m \sum_{j=0}^n [A(i,j) - B(i,j)]^2} \right) \quad (5)$$

In the above formula, MAX_A denotes the maximum pixel value of the image, m and n denote the pixel value of each row and column of the image, $A(i,j)$ denotes the original image, and $B(i,j)$ denotes the enhanced processed image. The four sets of images in Figure 6 are numbered from 1 to 4. The PSNR values of different algorithms in the same environment are shown in Table 2.

Table 2. PSNR results for different algorithms.

Image	SSR	MSR	MSRCR
1	10.637	11.559	14.179
2	10.098	10.807	14.430
3	10.433	10.755	13.442
4	10.495	10.865	14.370

Table 2 shows that the peak SNR values of MSRCR algorithm were higher than those of the other algorithms for different images in the same environment, which proves that the MSRCR performed well in this index.

SSIM is a metric used to assess the similarity between images. It is determined by the covariance between two images. A larger value of SSIM indicates a higher-quality image. When two images are identical, the SSIM is equal to 1. For two images x and y , the structural similarity index between them can be calculated using a specific formula:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (6)$$

In the above equation, the mean of x is denoted by μ_x and the mean of y is denoted by μ_y . σ_x^2 denotes the variance of x ; σ_y^2 denotes the variance of y ; σ_{xy} denotes the covariance of x and y ; and c denotes a constant value. The structural similarity ranges from 0 to 1. Table 3 shows the SSIM index of different algorithms.

Table 3. SSIM results for different algorithms.

Image	SSR	MSR	MSRCR
1	0.9018	0.9218	0.6484
2	0.8991	0.9099	0.6809
3	0.8784	0.8980	0.6303
4	0.8966	0.9089	0.6393

As can be seen from Table 3, the similarity structure index of MSR algorithm was better than that of the other two algorithms. Because the image must be restored to the original image as much as possible during image stereo matching, the two evaluation indexes and the final effect of the image were synthesized. In this study, we decided to choose the MSR algorithm for underwater image enhancement before stereo matching to make the image clearer.

After completing the underwater camera calibration and image enhancement, the stereo matching algorithm can be used to match each pixel point in the left and right

cameras. The depth of each point was calculated from the disparity value and converted into a depth image. The 3D coordinates of the corresponding target points were obtained to complete the 3D reconstruction. The depth of the target point corresponding to the pixel point in the image was calculated using the similar triangle principle, as shown in Figure 7. In the figure, O_l and O_r are the optical centers of the left and right cameras, respectively, X_l and X_r are the horizontal coordinates of the pixels on the left and right imaging planes, respectively. The depth D of the target point P is related to the disparity value d by the following equation:

$$D = \frac{f \cdot B_l}{(X_l - X_r) \cdot L_p} = \frac{f \cdot B_l}{d \cdot L_p} \quad (7)$$

where D is the depth value, f is the focal length. B_l is the baseline length of the camera, and d is the disparity value. L_p is the edge length of a single pixel of the camera.

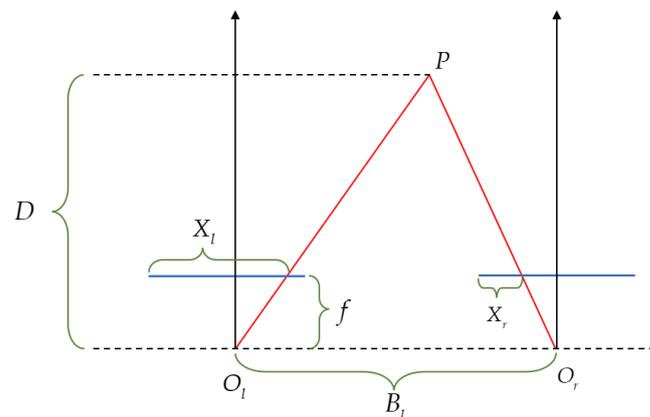


Figure 7. Binocular camera ranging model.

In this paper, we used the current mainstream semi-global block matching method (SGBM) [17,18] as the camera stereo matching algorithm, and the adopted parameters are shown in Table 4. In this paper, we used the relevant functions and methods integrated in the OpenCV library to implement the process of stereo matching. The texture filtering algorithm of the block matching algorithm was also integrated in the preprocessing of the SGBM algorithm in the OpenCV library, which helped to remove regions with low texture values. An example of the stereo matching results is shown in Figure 8. It is not difficult to observe that there is only a small void in the disparity of the fish body part of the disparity map [19], which basically meets the needs of size measurement.

Table 4. SGBM stereo matching algorithm parameters.

Parameter	Value
Min disparity	0
Max disparity	128
Block size	3×3
Uniqueness ratio	15
P1	$8 \times 3 \times 3$
P2	$32 \times 3 \times 3$

2.3.3. Image Segmentation

Before further processing the image and estimating the fish body length, the fish body needs to be recognized and segmented as the foreground. Deep-learning-based semantic segmentation methods have better environmental adaptability than traditional image segmentation methods. The effect of environmental factors on the segmentation results can be overcome. As this paper used a white breeding tank, the difference between the tilapia and the background was more obvious. In order to simplify the processing, the

segment anything model (SAM) [20] modeling platform of Meta-AI was used to segment the acquired images like the one illustrated in Figure 9. The model has a huge amount of training data and can perform hierarchical and full segmentation of images as shown in Figure 10. The segmented fish body parts help in the next step of key point labeling and improve the accuracy of manual labeling.

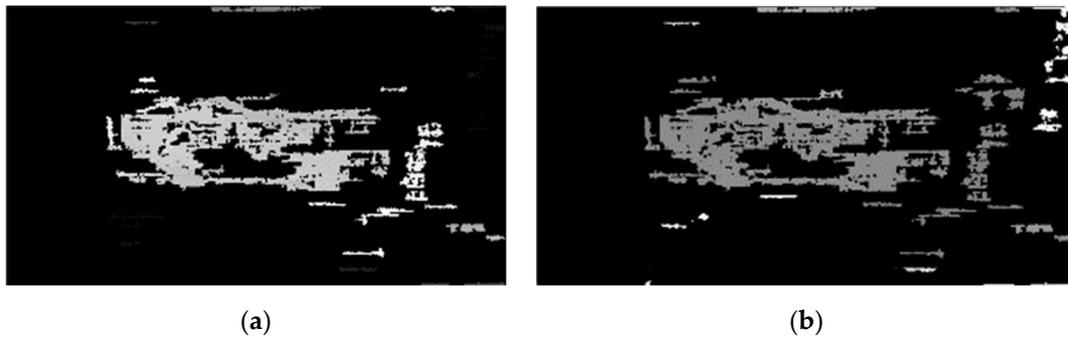


Figure 8. Maps generated by the SGBM algorithm: (a) disparity map; (b) depth map.

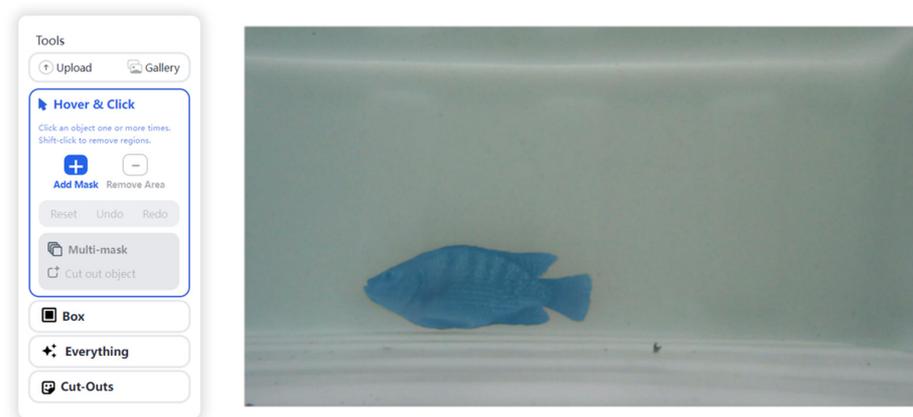


Figure 9. Image segmentation platform.

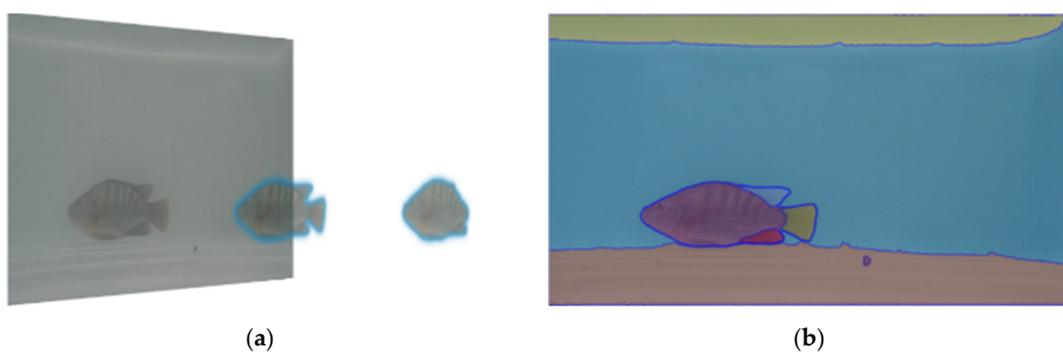


Figure 10. SAM platform segmentation results: (a) fish body layering; (b) full segmentation result.

2.3.4. Methods of Estimating Body Length

The segmented fish body image was labeled with the coordinates of the muzzle (x_1, y_1) and the base of the caudal fin (x_2, y_2), as shown in Figure 11. Since stereo matching is not good at obtaining high-density correspondences, the problem of not finding the corresponding depths on key points can occur. For such cases, we acquired the depths of multiple neighboring points of the key point. The average of the depths of these neighboring points was taken as the depth of the key point. The pixel length (PL) of the tilapia was calculated by Equation (8), and we tried to avoid acquiring the values of empty areas when

acquiring the depth information. And the depth values of five different parts of the fish body were acquired. The average value D_{avg} was taken as the depth value of the fish body in this figure, so as to minimize the error. Then, we combined the depth information with the triangle similarity principle to calculate the body length (BL) of the fish:

$$PL = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \tag{8}$$

$$BL = \frac{D_{avg}}{f} \times PL \tag{9}$$



Figure 11. Key point pixel labeling.

In practice, it is difficult to obtain a high-quality image of the fish at a perfect angle (fish body parallel to the mirror) [21,22]. In most cases, the body of the fish is not parallel to the mirror. As shown in Figure 12, the angle between the fish body and the mirror is a . We need to acquire the depth information of the fish’s muzzle and the base of the caudal fin. Their pixel coordinates are combined for 3D reconstruction of the fish body. In the figure, p_1 and p_2 are the pixel coordinates of the fish’s muzzle and caudal fin base in the x-axis, respectively. d_1 and d_2 are the depth values of the first two, respectively. Finally, the trigonometric function is combined to further calculate the real length (RL) of the fish body:

$$RL = BL \times \left\{ \cos \left[\arctan \left(\frac{|d_1 - d_2|}{|p_1 - p_2|} \right) \right] \right\}^{-1} \tag{10}$$

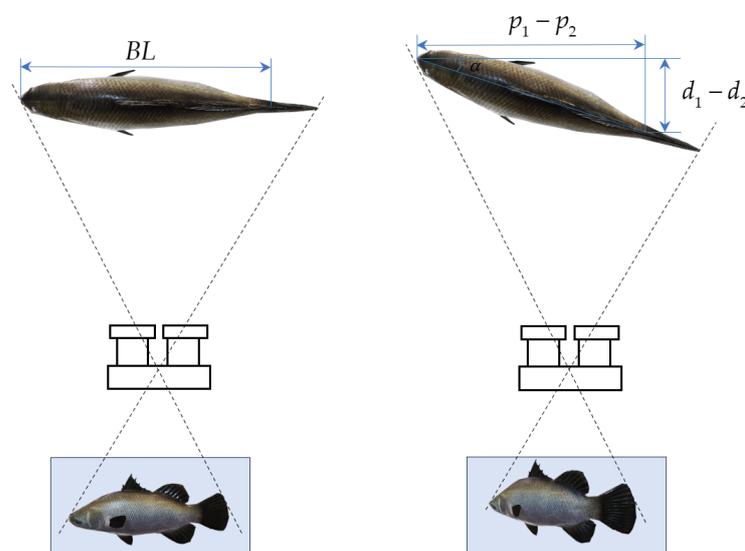


Figure 12. Effect of fish distance and angle variation on pixel body length.

3. Results

3.1. Fitting Results

In this study, a mass estimation model for tilapia was developed based on the body length of the tilapia. Five regression models were established: linear regression model, exponential regression model, logarithmic regression model, power function regression model, and quadratic term regression model. We manually measured the length and mass of 50 tilapia and fitted a model to these data. Categorized by different mass groups, the results were obtained as shown in Table 5. M denotes the mass of the fish, x denotes the actual body length of tilapia derived from Equation (10), and R^2 denotes the coefficient of determination of the model.

Table 5. Comparison of fittings of different models for tilapia mass.

Name	M1		M2		M3	
	Equation (M=)	R^2	Equation (M=)	R^2	Equation (M=)	R^2
Linear	$1.427x - 99.372$	0.89	$2.565x - 260.41$	0.90	$6.939x - 1261.2$	0.89
Exponential	$1.639 \times e^{0.028x}$	0.99	$47.527 \times e^{0.008x}$	0.91	$27.002 \times e^{0.012x}$	0.86
Logarithmic	$142.08\ln(x) - 603.06$	0.78	$558.96\ln(x) - 2709.1$	0.89	$1886.9\ln(x) - 9947.3$	0.90
Power	$3 \times 10^{-5}x^{2.996}$	0.99	$0.016x^{1.828}$	0.91	$2 \times 10^{-5}x^{3.124}$	0.87
Quadratic	$0.014x^2 - 1.861x + 71.347$	0.99	$0.047x^2 - 18.075x + 1991.7$	0.93	$-0.072x^2 + 46.135x - 6564.2$	0.91

From the results of fitting these five models, it can be seen that the correlation between fish length and mass was high. The overall coefficient of determination was greater than 0.78. The exponential, quadratic term, and power model had the highest coefficients of determination of 0.99 for the M1 mass group, but the coefficients of determination for all three types of models decreased as the fish grew. In contrast, the logarithmic model's coefficient of determination increased with fish mass. Overall, the quadratic term model had the highest coefficient of determination in each mass group and was superior to the other models. Therefore, the quadratic term model was designated as the best fitting model. The results of the quadratic term fitting for the three mass groups are shown in Figure 13.

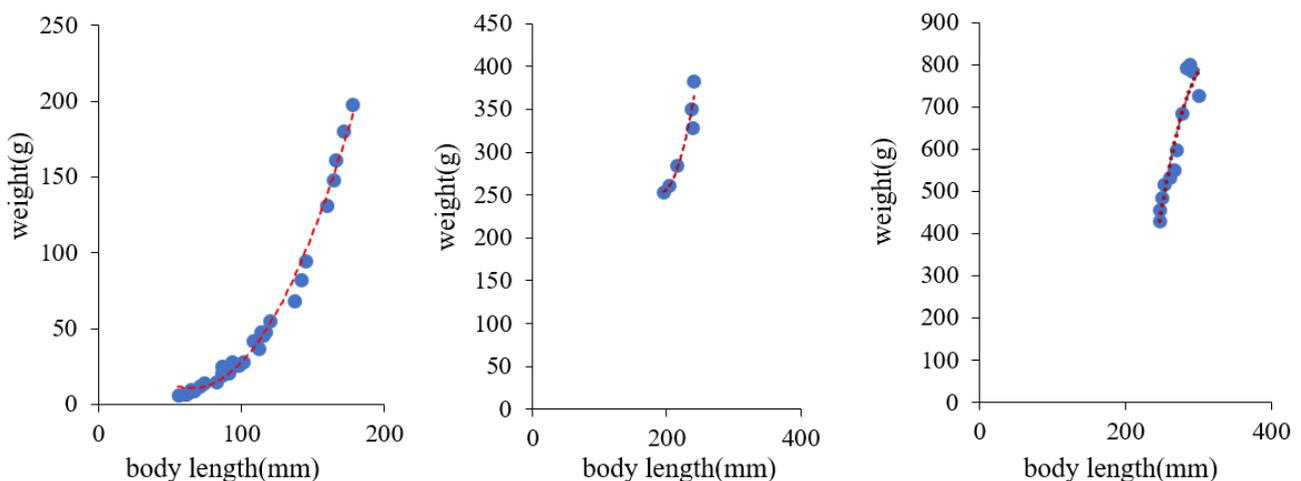


Figure 13. Quadratic term fit plots for different mass groups of tilapia.

3.2. Tilapia Mass Estimation

To further validate the stability of the model, we used the body lengths of the acquired tilapia to input directly into the model. Different mass groups can be interpreted as tilapia at different growth stages. Comparisons were made by selecting the linear, power function, and quadratic term models. The estimated values of the corresponding mass groups for the three models are shown in Table 6. The evaluation metrics were root mean square error

(RMSE), mean absolute error (MAE), and mean relative error (MRE), and their equations are as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (11)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (12)$$

$$\text{MRE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (13)$$

In the above equation, n is the number of samples. y_i denotes the true value. \hat{y}_i denotes the estimated value.

Table 6. Error analysis of tilapia mass estimation.

Group	Name	R ²	RMSE/g	MAE/g	MRE/%
M1	Linear	0.89	18.11	15.65	73.16
	Power	0.99	15.82	7.41	13.35
	Quadratic	0.99	6.62	4.97	12.14
M2	Linear	0.90	18.84	17.20	5.74
	Power	0.91	19.88	18.75	6.06
	Quadratic	0.93	22.16	17.19	5.34
M3	Linear	0.89	60.99	54.90	9.40
	Power	0.87	59.16	53.11	8.90
	Quadratic	0.91	56.21	43.38	7.25

Comparing the three tilapia mass estimation models, the RMSE ranged from 6.62 to 60.99 g. The lowest of these was the quadratic term model estimation results in M1, and the highest was the linear model estimation results in M3, with MAE ranging from 4.97 to 54.90 g. The lowest was still the quadratic term model estimation result in M1 and the highest was still the linear model estimation result in M3. It is not difficult to find that the linear model had poorer estimation results with R^2 ranging from 0.89 to 0.90. Removing the group with the smallest mass, the MRE ranged from 5.74% to 9.40%. And the power function model performed better than the linear model in the small mass estimation group. The R^2 was 0.87 to 0.99 and the MRE was 6.06% to 8.90%. The best prediction was made by the quadratic term model with an R^2 of 0.91 to 0.99, and the MRE was also minimized to 5.34% to 7.25%.

Finally, in this study, the mass of 10 tilapia whose images were acquired was estimated. After image enhancement, stereo matching, body length estimation, and mass fitting, we calculated the mass of tilapia. The results of the comparison between the manual measurements and the estimated mass of the final tilapia are shown in Figure 14. The R^2 was 0.99, and the MRE was 5.90%. No significant difference was found between the artificially measured mass and the estimated mass, so the test proved that the method can be used for the mass estimation of tilapia.

Although the experiments are affected by the water body environment with some differences in the results, we compared them with some existing methods in order to show the superiority of our approach. The methods we compared are the area- and area-squared-based mass estimation model [13] and the deep learning method using YOLOv4 DLN [6]. As can be seen in Table 7, comparing the other two methods, the MRE of our method reached 5.90%, which is an improvement of 1.54% compared to the second place. The coefficient of determination of our method reached 0.99, which is an improvement of 0.05 compared to the deep learning method of YOLOv4 DLN. Thus, our method performed better than the first two.

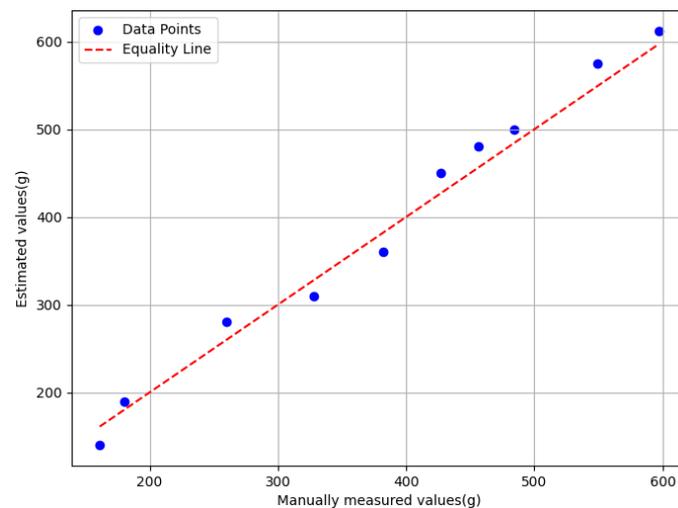


Figure 14. Comparison between estimated and manually measured fish mass.

Table 7. Comparison of results of different methods for estimating fish mass.

Method	MRE/%	R ²
Based on area and area squared	8.46	0.99
YOLOv4 DLN	7.44	0.94
Our method	5.90	0.99

4. Discussion

In some previous studies using fish body length to estimate fish mass, it was necessary to manually mark the fish contour. However, such marking is not only less accurate but also time-consuming and laborious [23,24]. Moreover, the free swimming of the fish in the water increases the difficulty of this operation. There are also combinations of features [25] to estimate fish mass and the use of fish body area [26] to estimate mass. In this study, in order to improve the accuracy of labeling the mouth and tail end of the fish, the image is segmented first. The body part of the fish is accurately segmented, and then the segmented image is labeled. And underwater image enhancement is performed before image stereo matching to further improve the matching accuracy. With the rapid development of deep learning, the SAM model released by Meta-AI has been able to perform semantic segmentation without the researcher providing any training samples. The segmentation in this study was directly based on this platform. For the segmented fish body, we can more clearly label the coordinates of the fish's muzzle and caudal fin base. And use the pixel coordinates combined with the depth information to estimate the body length of the fish more accurately.

The experimental results show that the mass of the tilapia can be accurately estimated using the single characteristic of fish body length. However, as the tilapia grows it may lead to larger errors. In this study, it was found that, at the later stage of tilapia growth, the growth rate of the body length of the tilapia was lower than the growth rate of its own mass. At this point, a single body length can no longer accurately estimate the mass of tilapia. It is necessary to consider adding other characteristic parameters such as body height and body width to improve the estimation of tilapia weight to further improve the accuracy of tilapia weight estimation.

Although this study created a dataset containing images of tilapia, the dataset was primarily collected under farm tank conditions. These experimental conditions resulted in a homogenous experimental context and fish species limitations. In the future, in order to enhance the general applicability of the method, it could be tested in more diverse natural water environments. A wider range of species of fish and their images under different background conditions could be collected to verify the applicability and generalizability

of the system. In this paper, the coordinates of the key points of the fish's muzzle and caudal fin base are extracted using manual labeling, and there is some human error in the calculation of the fish's body length. In the subsequent research, the model can be built using neural networks with deep learning methods. More complex datasets can be trained so as to achieve automatic detection and labeling of the endpoints of the snout and the base of the caudal fin for fully automated fish quality estimation. Currently, the system relies on a laptop computer as the platform for image processing and computation, and the images are captured by a binocular camera. Given the relatively large size of the laptop, the integration of the system components and the binocular camera into a smaller embedded device can be considered in the future. In this way, the size and weight of the system can be reduced, making it more suitable for use in practical application scenarios. When acquiring image depth information, some advanced devices can be used, such as the ZED depth camera, which can avoid the complex calibration and ranging process.

5. Conclusions

In this paper, a non-contact tilapia mass estimation method based on underwater binocular vision is proposed. Firstly, a platform for acquiring underwater tilapia images was built. The underwater tilapia images were acquired by a binocular vision system, and the underwater camera was calibrated, and the aberration was corrected. The image was preprocessed using a Retinex-based image enhancement algorithm, combined with stereo matching to obtain depth information. Image segmentation was performed on the acquired tilapia images, and the coordinates of key points were labeled. Thus, the three-dimensional reconstruction of the underwater tilapia was realized. The estimation model between body length and mass of tilapia was established, thus realizing the mass estimation of non-contact tilapia. The MRE of the tilapia mass estimation model developed in this paper ranged from 5.34% to 12.14%. Experimental results show that the method proposed in this paper has high accuracy and stability. The R of the estimated value of the mass estimation model to the manually measured value was 0.99, and the MRE was 5.90%. The mass estimation of underwater tilapia can be carried out effectively. The tilapia mass estimation model established by this study can estimate the mass of tilapia without contact, which improves the measurement efficiency. It provides technical support for the rapid estimation of body length and mass of underwater free-swimming tilapia.

Author Contributions: Conceptualization, G.F., M.C. and B.P.; methodology, B.P.; software, B.P.; validation, G.F. and B.P.; resources, G.F.; data curation, G.F. and B.P.; writing—original draft preparation, G.F. and B.P.; writing—review and editing, G.F. and B.P.; supervision, G.F. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Key Research and Development Program of China (2022YFD2400800) and Jiangsu Modern Agricultural Industry Key Technology Innovation Planning, No. CX (20) 2028.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Experimental data related to this paper can be requested from the authors by email if any researcher is in need of the dataset, email: gffeng@shou.edu.cn.

Acknowledgments: The authors are very grateful to Wang Yaohui of Nantong Longyang Aquatic Co. in Jiangsu Province, China, for providing us with the experimental data and the data collection site.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Priyadarshini, B.; Xavier, K.M.; Nayak, B.B.; Dhanapal, K.; Balange, A.K. Instrumental quality attributes of single washed surimigels of tilapia: Effect of different washing media. *Lwt* **2017**, *86*, 385–392. [[CrossRef](#)]
2. Zhou, C.; Lin, K.; Xu, D.; Chen, L.; Guo, Q.; Sun, C.; Yang, X. Near infrared computer vision and neuro-fuzzy model-based feeding decision system for fish in aquaculture. *Comput. Electron. Agric.* **2018**, *146*, 114–124. [[CrossRef](#)]
3. Andersen, K.H.; Jacobsen, N.S.; Farnsworth, K.D. The theoretical foundations for size spectrum models of fish communities. *Can. J. Fish. Aquat. Sci.* **2016**, *73*, 575–588. [[CrossRef](#)]
4. Shafry, M.; Rehman, A.; Kumoi, R.; Abdullah, N.; Saba, T. A new approach in measuring fish length using fish length from digital images (FiLeDI) framework. *Int. J. Phys. Sci.* **2012**, *7*, 607–618.
5. Li, Y.; Huang, K.; Xiang, J. Measurement of dynamic fish dimension based on stereoscopic vision. *Trans. Chin. Soc. Agric. Eng.* **2020**, *36*, 220–226.
6. Abinaya, N.; Susan, D.; Sidharthan, R.K. Deep learning-based segmental analysis of fish for biomass estimation in an occulted environment. *Comput. Electron. Agric.* **2022**, *197*, 106985. [[CrossRef](#)]
7. Konovalov, D.A.; Saleh, A.; Efremova, D.B.; Domingos, J.A.; Jerry, D.R. Automatic weight estimation of harvested fish from images. In Proceedings of the 2019 Digital Image Computing: Techniques and Applications (DICTA), Perth, Australia, 2–4 December 2019; pp. 1–7.
8. Fernandes, A.F.; Turra, E.M.; de Alvarenga, É.R.; Passafaro, T.L.; Lopes, F.B.; Alves, G.F.; Singh, V.; Rosa, G.J. Deep Learning image segmentation for extraction of fish body measurements and prediction of body weight and carcass traits in Nile tilapia. *Comput. Electron. Agric.* **2020**, *170*, 105274. [[CrossRef](#)]
9. Yu, C.; Hu, Z.; Han, B.; Dai, Y.; Zhao, Y.; Deng, Y. An intelligent measurement scheme for basic characters of fish in smart aquaculture. *Comput. Electron. Agric.* **2023**, *204*, 107506. [[CrossRef](#)]
10. Li, D.; Hao, Y.; Duan, Y. Nonintrusive methods for biomass estimation in aquaculture with emphasis on fish: A review. *Rev. Aquac.* **2020**, *12*, 1390–1411. [[CrossRef](#)]
11. Sanchez-Torres, G.; Ceballos-Arroyo, A.; Robles-Serrano, S. Automatic measurement of fish weight and size by processing underwater hatchery images. *Eng. Lett.* **2018**, *26*, 461–472.
12. Saberioon, M.; Císař, P. Automated within tank fish mass estimation using infrared reflection system. *Comput. Electron. Agric.* **2018**, *150*, 484–492. [[CrossRef](#)]
13. Hao, Y.; Yin, H.; Li, D. A novel method of fish tail fin removal for mass estimation using computer vision. *Comput. Electron. Agric.* **2022**, *193*, 106601. [[CrossRef](#)]
14. Huang, K. Research and Implement of Machine Vision Based Underwater Dynamic Fish Size Measurement Method. Master's Thesis, Zhejiang University, Hangzhou, China, 2021.
15. Zhang, Z. Flexible camera calibration by viewing a plane from unknown orientations. In Proceedings of the Seventh IEEE International Conference on Computer Vision, Corfu, Greece, 20–25 September 1999; pp. 666–673.
16. Chi, D.; Wang, Y.; Ning, L.; Yi, J. Experimental research of camera calibration based on ZHANG'S method. *J. Chin. Agric. Mech.* **2015**, *36*, 287–289+337. [[CrossRef](#)]
17. Hirschmuller, H. Stereo processing by semiglobal matching and mutual information. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *30*, 328–341. [[CrossRef](#)]
18. Lee, Y.; Park, M.-G.; Hwang, Y.; Shin, Y.; Kyung, C.-M. Memory-efficient parametric semiglobal matching. *IEEE Signal Process. Lett.* **2017**, *25*, 194–198. [[CrossRef](#)]
19. Ttofis, C.; Kyrkou, C.; Theocharides, T. A hardware-efficient architecture for accurate real-time disparity map estimation. *ACM Trans. Embed. Comput. Syst. TECS* **2015**, *14*, 1–26. [[CrossRef](#)]
20. Kirillov, A.; Mintun, E.; Ravi, N.; Mao, H.; Rolland, C.; Gustafson, L.; Xiao, T.; Whitehead, S.; Berg, A.C.; Lo, W.-Y. Segment anything. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 2–6 October 2023; pp. 4015–4026.
21. Muñoz-Benavent, P.; Andreu-García, G.; Valiente-González, J.M.; Atienza-Vanacloig, V.; Puig-Pons, V.; Espinosa, V. Enhanced fish bending model for automatic tuna sizing using computer vision. *Comput. Electron. Agric.* **2018**, *150*, 52–61. [[CrossRef](#)]
22. Dunbrack, R. In situ measurement of fish body length using perspective-based remote stereo-video. *Fish. Res.* **2006**, *82*, 327–331. [[CrossRef](#)]
23. Shafait, F.; Harvey, E.S.; Shortis, M.R.; Mian, A.; Ravanbakhsh, M.; Seager, J.W.; Culverhouse, P.F.; Cline, D.E.; Edgington, D.R. Towards automating underwater measurement of fish length: A comparison of semi-automatic and manual stereo-video measurements. *ICES J. Mar. Sci.* **2017**, *74*, 1690–1701. [[CrossRef](#)]
24. Atienza-Vanacloig, V.; Andreu-García, G.; López-García, F.; Valiente-González, J.M.; Puig-Pons, V. Vision-based discrimination of tuna individuals in grow-out cages through a fish bending model. *Comput. Electron. Agric.* **2016**, *130*, 142–150. [[CrossRef](#)]

25. 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. [[CrossRef](#)]
26. Shi, C.; Zhao, R.; Liu, C.; Li, D. Underwater fish mass estimation using pattern matching based on binocular system. *Aquac. Eng.* **2022**, *99*, 102285. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.