Journal of *Marine Science and Engineering*

MDPI

Article

An Experimental Study on Estimating the Quantity of Fish in Cages Based on Image Sonar

Guohao Zhu [1,2], Mingyang Li [1], Jiazhen Hu [1,2], Luyu Xu [1], Jialong Sun [1,3], Dazhang Li [4], Chao Dong [5], Xiaohua Huang [2,6,*] and Yu Hu [2,6,*]

1 School of Geomatics and Marine Information, Jiangsu Ocean University, Lianyungang 222001, China
2 Key Laboratory of South China Sea Fishery Resources Exploitation & Utilization, Ministry of Agriculture and Rural Affairs, South China Sea Fisheries Research Institute, Chinese Academy of Fishery Science, Guangzhou 510300, China
3 Jiangsu Marine Resources Development Research Institute, Lianyungang 222005, China
4 Zhejiang Provincial-Subordinate Architectural Design Institute, Hangzhou 310007, China
5 Key Laboratory of Marine Environmental Survey Technology and Application, Ministry of Natural Resources, Guangzhou 510300, China
6 Tropical Fisheries Research and Development Center, South China Sea Fisheries Research Institute, Chinese Academy of Fishery Science, Sanya 572018, China
* Correspondence: huangxhua@scsfri.ac.cn (X.H.); huyu@scsfri.ac.cn (Y.H.); Tel.: +86-020-3406-6940 (Y.H.)

Abstract: To address the highly demanding assessment of the quantity of fish in cages, a method for estimating the fish quantity in cages based on image sonar is proposed. In this method, forward-looking image sonar is employed for continuous detection in cages, and the YOLO target detection model with attention mechanism as well as a BP neural network are combined to achieve a real-time automatic estimation of fish quantity in cages. A quantitative experiment was conducted in the South China Sea to render a database for training the YOLO model and neural network. The experimental results show that the average detection accuracy mAP50 of the improved YOLOv8 is 3.81% higher than that of the original algorithm. The accuracy of the neural network in fitting the fish quantity reaches 84.63%, which is 0.72% better than cubic polynomial fitting. In conclusion, the accurate assessment of the fish quantity in cages contributes to the scientific and intelligent management of aquaculture and the rational formulation of feeding and fishing plans.

Keywords: cage fish; forward-looking image sonar; target recognition; quantity estimation

Citation: Zhu, G.; Li, M.; Hu, J.; Xu, L.; Sun, J.; Li, D.; Dong, C.; Huang, X.; Hu, Y. An Experimental Study on Estimating the Quantity of Fish in Cages Based on Image Sonar. *J. Mar. Sci. Eng.* **2024**, *12*, 1047. https://doi.org/10.3390/jmse12071047

Academic Editor: Sergei Chernyi

Received: 30 May 2024
Revised: 18 June 2024
Accepted: 20 June 2024
Published: 21 June 2024

1. Introduction

As a major agricultural country in the world, the development of China's agricultural economy is related to the development of the national economy [1]. As an important branch of aquaculture, fishery farming has always been an important pillar of China's agricultural economy. With the development of society, science, and technology, the level of agricultural modernization has rapidly improved, and the intelligent development of fish farming has accelerated. The monitoring and regulation of the breeding environment and the decision making of feed feeding have gradually shifted from completely relying on manual diagnosis, decision making, and adjustment to the mechanization and precision of monitoring equipment, and then to the digitalization and intelligence of the system [2].

At present, fish farming in China varies between pond and cage cultures. Among these, cage culture exhibits the highest level of intensification, with a myriad of issues arising during fish farming [3]. Fish quantity monitoring, as an important part of cage aquaculture production management, is of profound significance mainly in the following three aspects: 1. Intelligent management of aquaculture, allowing aquaculture managers to adjust the feeding amount and make fishery harvesting plans according to the fish

production; 2. Early warning of the safety of the fishnet and the breakage of the fishnet in the case of abnormal fish quantity, to repair it in time to reduce losses; 3. Facilitation of the assessment of the financial assets of the catch, rendering necessary technical conditions for achieving financial assets of fishery harvesting [4].

Given the above requirements, experts and scholars at home and abroad put forward solutions based on different monitoring methods. Baumgartner et al. [5] observed fish in artificial ponds and calculated the fish quantity and body length by software, concluding that sonar was effective in observing fish activities and obtaining quantitative information. Ding et al. [6] collected 59h underwater data by using ARIS sonar and completed the automatic processing of a large number of acoustic data through an image processing algorithm, including target extraction and counting. A remote cage monitoring system that combines light and sound with motor rotation scanning was jointly developed by the Massachusetts Institute of Technology and Woods Hole Oceanographic Institution, which can identify individual fish well to achieve safe monitoring of fishnet [7]. However, its high cost and the prolonged acoustic imaging time required by motor rotation detection (compared to the standard imaging time of 3 min) cause the repeated detection of swimming fish in cages, resulting in a large error in fish quantity estimation. Domestically, the Fishery Machinery and Instrument Research Institute of the Chinese Academy of Fishery Sciences developed a multi-angle cage monitor by optical means [8]. Because of the turbid sea water in most coastal areas of our country, except Hainan, the instrument had been limited by effectively observing a range of underwater targets and higher power consumption. Given the limitation of the above optical monitoring technology in the actual condition of cages, most of the domestic research has prioritized acoustic monitoring methods. The Shanghai Acoustics Laboratory of the Chinese Academy of Science put forward the acoustic warning tape method and the remote-operated vehicle patrol method, which were mainly used for monitoring the size of netting and fish but were less able to obtain quantity data. Xiamen University has successively developed acoustic monitoring systems based on the vertical detection method and single-beam transducer motor-rotating horizontal scanning method. The circular multi-beam scanning detection method had high requirements for the estimation of fish swimming speed, and either the underestimation or overestimation will lead to partial fish missed detection or repeated detection [9–11]. Yihan Feng et al. [12] introduced an automated method for estimating fish abundance in sonar images based on the modified MCNN (multi-column convolutional neural network), named FS-MCNN. They also proposed the multi-dilation rate fusion loss, which improved the accuracy and robustness of the model. This method improved the impact of low pixels in sonar images and blurry edges of target objects in sonar images.

The target recognition technique was indispensable for locating and counting fish in acoustic images. Since the R-CNN (Region with CNN Features) was put forward in 2014, the target detection method based on deep learning has become the main technique, instead of the traditional method [13]. Initially, the two-stage method was adopted for target detection based on deep learning, that is, the detection process was explicitly divided into two stages: candidate region selection and target region judgment, with a high detection accuracy but slow detection speed. Later, in 2016, the one-stage target detection method represented by YOLOv1 came into being. Instead of extracting candidate regions in advance, the method directly predicted the category probability and position of the output target object, which attracted more attention by greatly reducing the consumption of computing resources and improving the detection speed [14]. The YOLO series of target detection methods, along with the development of single-stage object detection, has been regarded as a typical representative of the one-stage method. Ye Zhaobing et al. [15] proposed the YOLOv3-SPP (Spatial Pyramid Pooling) underwater target detection algorithm to solve the problem of missed detection and false detection caused by unclear images and the complex underwater environment in underwater target detection. Chen Yuliang et al. [16] put forward a method for detecting and identifying underwater biological targets in shallow water based on the YOLOv3 network, aiming to overcome the low detection

accuracy of underwater biological targets in a shallow sea caused by color distortion, rough image, local overexposure, and large size difference in underwater images.

In response to the deficiencies of the aforementioned detection methods, this paper proposes a method for estimating the quantity of fish in net cage farming based on forward-looking imaging sonar. This method utilizes forward-looking sonar to generate acoustic images of aquaculture net cages, employs a YOLOv8 neural network model with an added attention mechanism to identify fish targets, and utilizes a BP neural network to invert feature data to estimate the overall quantity of fish. Quantitative detection experiments were conducted in constructed fish cages, with multiple sets of experimental results showing that the average accuracy of fish quantity assessment reached 84.63%, thereby validating the feasibility of this method. By using this method, fish farmers can gain real-time insights into the quantity of fish inside net cages during the farming process, enabling scientific aquaculture management and reducing farming risks.

2. Materials and Methods

2.1. Overall Process

On the whole, the adopted method is divided into three steps: Firstly, the image sonar is fixed on one side of the cage and observed for more than 10 min, recording sonar data and exporting it to video. Secondly, the improved YOLOv8 model is used to detect all the frames of the current video, and there is only one detection category, namely fish, Thirdly, the number of fish shoals detected in each frame of the video is sorted from the largest to the smallest, and the actual quantity of fish in the cage is estimated by using the trained neural network model according to the top 20 fish quantity. In the second step, the YOLOv8 model needs to be trained with fish sonar image data, and the neural network in the third step is trained by the mapping relationship between the previous observation data and the actual quantity.

2.2. Introduction to Image Sonar

The ARIS1800 (Adaptive Resolution Imaging Sonar) sonar used in the present study was introduced by Sound Metrics in 2012. When forward-looking sonar performs detection, the transducer at its top emits ultrasonic waves in the forward direction, and subsequently, the objects illuminated by these waves reflect them, forming echo signals. The sonar receives these signals to generate acoustic images. Typically, dividing the detection beam horizontally into multiple smaller fan-shaped beams can enhance imaging precision, with the vertical angle of each beam group remaining unchanged. Table 1 below lists the specific parameters of ARIS-1800 [17].

Table 1. Image sonar parameters of ARIS1800.

Item	Low-Frequency Mode	High-Frequency Mode
Operating frequency/MHz	1.1	1.8
Effective range/m	0.7–35	0.7–15
Resolution/mm	23	3
Maximum frame rate/second	3.5–15 frames	
Field of view (FOV)/(°)	28 × 14	
Size/cm	31 × 17 × 14	

Figure 1 is a physical diagram of ARIS1800 sonar. In the process of acoustic image generation, water reverberation, channel change, interference, and self-noise generated by target activity are usually accompanied. The non-sequential emission of the ARIS transducer elements can effectively reduce the influence of self-noise and crosstalk. As for how the ARIS system works, the transducer actively emits sound waves in the field of view according to the size of the reflected echo, thus forming acoustic images with different

light and dark characteristics. The acoustic image includes a bright area corresponding to the bottom, a bright area representing the fish target, and a dark area corresponding to the water background, as shown in Figure 2 [18].

Figure 1. ARIS1800 sonar physical diagram.

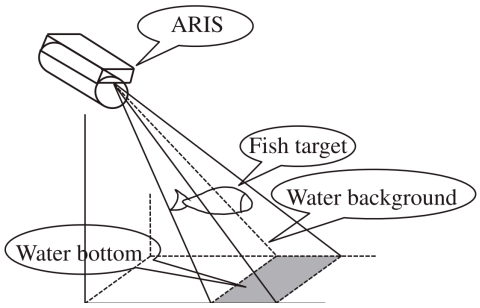


Figure 2. Schematic diagram of sonar fish detection.

2.3. Sonar Data Acquisition of Fish Quantity in Cages

The experimental data were measured in the sea area of Guishan Island, Zhuhai City, Guangdong Province, China, in March 2023 (latitude and longitude: 113.84473 and 22.12571, respectively). Figure 3 shows the satellite image of the experimental sea area, and Figure 4 shows the aerial image of the experimental base.

Figure 3. Satellite image of the experimental sea area.

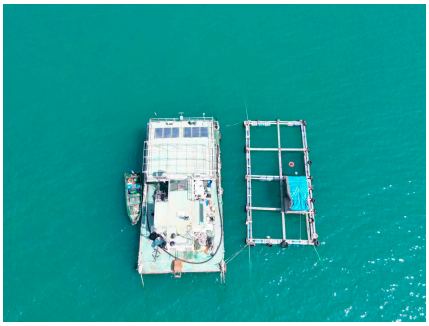


Figure 4. Aerial image of the experimental base.

The cage used in this experiment is shown in Figure 5, with the width and height of the fishnet being 6 × 3 × 4 m, respectively. During the experiment, iron blocks were tied to the four corners of the fishnet as counterweights to open the netting. The object of sonar detection was a golden pomfret with a body length of about 15 cm, which was placed in the experimental cage.

Figure 5. Experimental cage.

It can be seen from Figure 6 that the ARIS sonar is tied to a lifebuoy and floating in the water, with the sonar probe placed at a depth ranging from 30 to 40 cm and a 45-degree angle inclined to the left. The sonar was placed in the middle of the short side of the netting, and the sonar signal covered as much water space as possible. Then, the sonar was connected to a laptop computer, and the supporting software ARISFish (v2.6.3) was used for data acquisition. The upper computer software ARISFish communicates with the sonar device to receive and process the sonar-collected data. It then displays the real-time processing results graphically. A high-frequency mode was used in the sonar, that is, the frequency was 1.8 MHz, and the detection distance was set to just observe the netting on the opposite side, which was about 4.6 m.

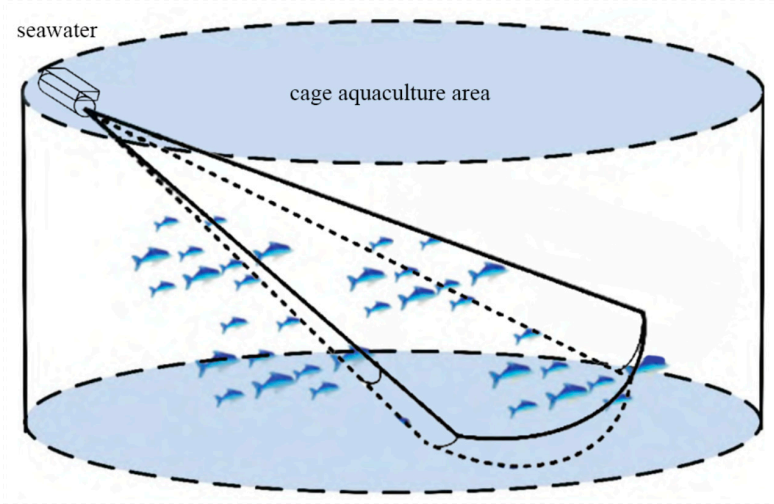
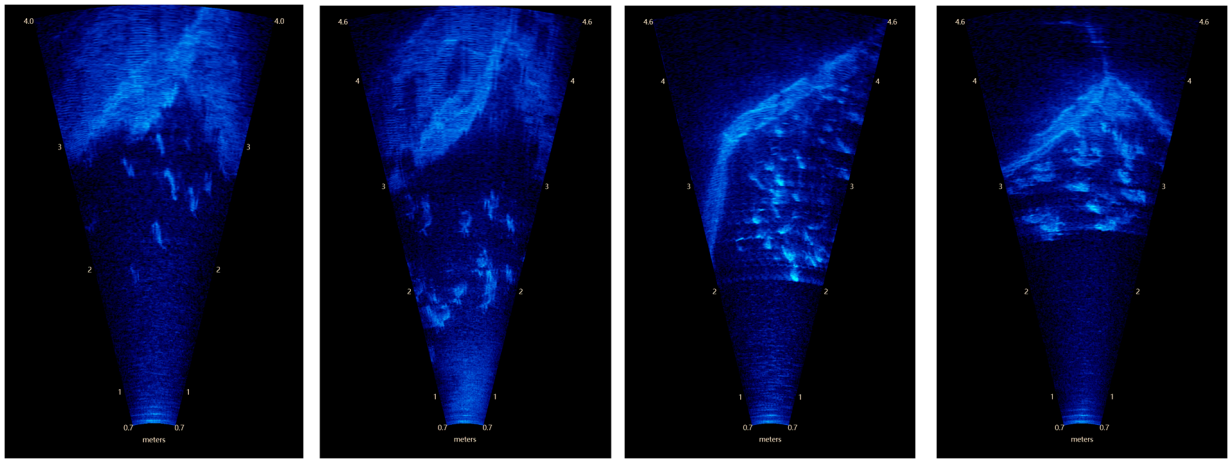


Figure 6. Schematic diagram of the sonar deployment.

The quantitative experiment was carried out with every 20 fish as the standard group, and 20, 40, 60, and 80 golden pomfrets were put into the experimental cage in turn. Each group of fish was continuously detected by sonar, and the data every 10 min were recorded as an ARIS source file, which was saved in the computer for subsequent processing.

The sonar images of different groups of fish are presented in Figure 7, from which the clear outlines of fish and netting were visible. The direction of the images was not the

same because the waves were constantly beating the sonar, causing the sonar probe to swing left and right in a certain range. Meanwhile, only a 28° sonar opening angle made it impossible for the sonar detection waves to cover the entire cage, that is, not every fish was visible, which put forward higher requirements for the next estimation method.



(**a**) (**b**) (**c**) (**d**)

Figure 7. Sonar images of different groups of fish. (**a**) Twenty fish; (**b**) forty fish; (**c**) sixty fish; and (**d**) eighty fish.

3. Recognition Algorithm

3.1. Introduction to YOLO Algorithm

YOLO is a two-step target detection model based on a neural network. Firstly, the input image is divided into S × S grids, and each grid generates B prediction frames, each of which is represented by a corresponding feature vector, generally taking S = 7 and B = 2. The feature vector is composed of: the coordinates of the center point of the corresponding prediction frame, the width and height of the prediction frame, and the confidence of the existence of the object, and each grid will generate a classification prediction feature vector. Finally, the prediction frame with high confidence and its classification are returned to the original input image [19].

By adding the feature fusion method to the feature extraction network, the algorithm adopts the backbone network of Darknet-53. The feature extraction network structure of the YOLO model is shown in Figure 8. The network is a full convolution network, which is trained and tested on the COCO dataset, and finally outputs a feature map of size 13 × 13 × 255. After the feature map is input to the target detection layer 1, position regression and classification regression are performed. Moreover, the feature map of the last layer and the feature map of the middle layer are fused by the above sampling method and input into the target detection layer 3 and the target detection layer 2, respectively, to achieve position regression and classified regression on the feature maps of multiple sizes [20].

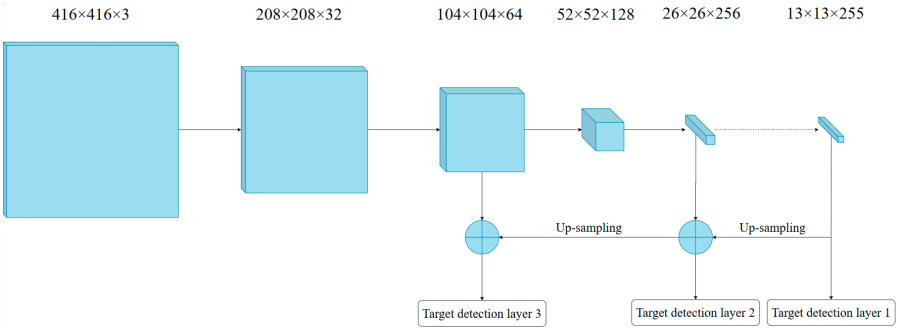


Figure 8. YOLO feature extraction network.

Considering the performance and stability of the model comprehensively, the YOLOv8 model was used for fish target detection in this study. YOLOv8 directly transforms the problem of fish detection into a regression problem. After a regression, not only the position coordinates of each fish group are generated, but also the probability of each candidate region belonging to the category is obtained.

3.2. YOLO Algorithm Improvement

On the premise of satisfying real-time performance and high detection accuracy, a target detection model based on an improved YOLOv8 algorithm is proposed. Considering that every fish is a small target in a sonar image, the core idea of the improved algorithm is to improve the network's perception ability of small target feature information [21]. Firstly, the CBAM (Convolutional Block Attention Module) [22] is improved by using the attention mechanism, and the channel-space attention module CSAM is proposed, which is lighter and can focus on the dimensional features of a small target space. The CSAM is embedded after each convolution of the backbone network to extract features. Then, a 4-fold down-sampling process is added to the YOLOv8 backbone network using 4-scale detection. After the input image is down-sampled by 4 times, a large shallow feature map is obtained. Because of the small receptive field, the feature map contains rich position information to improve the detection effect of small targets [23].

CAM is the channel attention module in CBAM. It consists of two fully connected layers to capture non-linear cross-channel interaction. However, the introduction of the fully connected layer causes a large amount of computation. Even if the channel characteristics are compressed, the parameter quantity is still proportional to the square of the number of channels [24]. For a reduced computational burden, a one-dimensional convolution with convolution kernel length k is used to achieve local cross-channel interaction by referring to the idea of ECANet, aiming to extract the dependency between channels [25]. L-CAM represents the improved lightweight channel attention module, and the convolution kernel length k is calculated by Formula (1):

$$k = \psi(C) = \left| \frac{\mathrm{lb}C}{\gamma} + \frac{b}{\gamma} \right|_{\mathrm{odd}} \tag{1}$$

where C is the number of channels of the input characteristic map, and γ and b are set to 2 and 1, respectively. "lb" means log-based binary.

SAM stands for the spatial attention module in CBAM. In this study, a new channel-spatial attention structure CSAM was constructed by using the improved L-CAM and SAM modules, as shown in Figure 9. Firstly, L-CAM and SAM were used to obtain the channel attention weight Mc and spatial attention weight M_s, respectively. Then, the map of attention M_c and M_s was extended to the size of $R^{W\times H\times C}$; W and H represent the width and height of the image, respectively; and C represents the number of channels. The sum of elements and sigmoid normalization were carried out to obtain the attention weight matrix M_{cs} based on the space and channel. The weight reflects the attention distribution in the feature map so that the model can obtain more effective features in the more accurate attention area, as shown in Formula (2):

$$M_{cs} = \mathrm{sigmoid}(M_c + M_s) \tag{2}$$

Finally, the mixed attention weight matrix M_{cs} was multiplied with the input feature map F element by element and added to the original input feature map to obtain a refined feature map F_{cs}, which was calculated as shown in Formula (3):

$$F_{cs} = F + F \otimes M_{cs} \tag{3}$$

The attention mechanism tells the model where to concentrate more calculations and improve the expressive force of the region of interest [26]. The idea of CSAM was to obtain attention weight matrices M_c and M_s from the input feature map F along the spatial dimension and the channel dimension, respectively, to improve the effective flow of feature information in the network. This module emphasizes paying attention to meaningful features, focusing on important features and suppressing invalid features in the two dimensions of channel and space. For small targets, a single feature region gains more weight and contain more effective targets. The model will place a much higher premium on learning the features of this region to extract features better with limited computing resources.

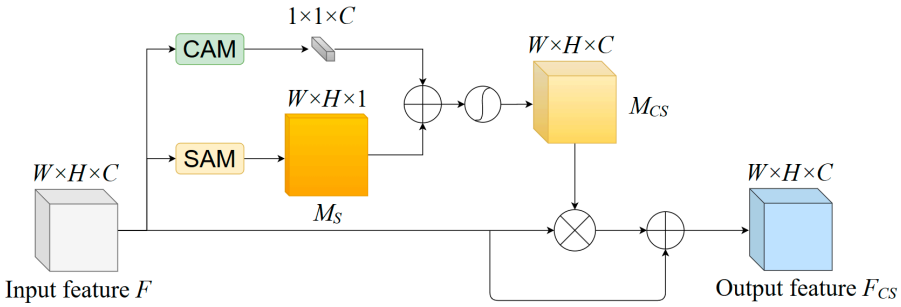


Figure 9. CSAM module.

3.3. Experimental Analysis

3.3.1. Dataset Making

One hundred sonar images of fish schools were intercepted from experimental data and processed by the MakeSense online data labeling website. There was only one labeling category, namely fish. After the completion of all labeling, the label file was exported, and each sonar image corresponded to a text file with the same name to record the labeling results. The labeled datasets were divided into two categories by random numbers, with 80 images as the training sets and 20 as the test sets.

3.3.2. Experimental Environment

The Windows 10 system was used in the experiment, with NVIDIA GeForce RTX 3070 (8 GB) as the GPU and Intel i9-12900H as the processor. The experimental environment was python3.9.13, pytorch1.13.1, and cuda11.7.

3.3.3. Evaluation Indicators

For the detection performance, the average precision (*mAP*), parameter quantity (Params), calculation quantity (GFLOPs), and speed (FPS) were used as evaluation indexes [27]. In the process of calculating *mAP*, it was necessary to calculate the average accuracy (*AP*) first, which represents the average accuracy of a category in the dataset. The calculation process is shown in Formula (4). Then, the *AP* values of different categories were averaged to obtain a *mAP*, and the calculation process was shown in Formula (5):

$$AP = \int_0^1 p(r)dr \tag{4}$$

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i \tag{5}$$

where *P* represents the precision ratio, that is, the ratio of the correct result of model recognition among all the recognized results; *r* represents the recall ratio, that is, the ratio of the correct results of model recognition to the results that need to be recognized in the dataset; *N* represents the number of categories of samples, and *N* = 1 in this study.

3.3.4. Training Process

When training the detection network model, the number of iterations was set to 300, the weight attenuation coefficient to 0.0005, the initial learning rate to 0.01, the learning rate momentum to 0.937, and the batch size to 16. As shown in Figures 10 and 11, the model triggered "Early Stopping" to stop training after 120 iterations, at which time the loss decreased to 0.6 and the mAP50 reached 73.02%.

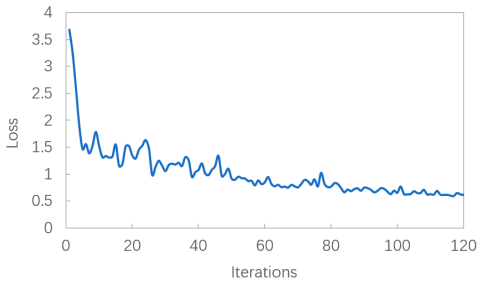


Figure 10. Training process (loss).

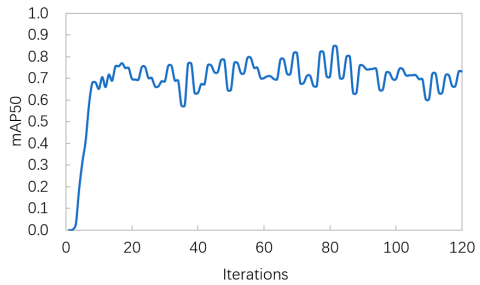


Figure 11. Training process (mAP50).

3.3.5. Ablation Experiments

To verify the effectiveness of the channel-space attention mechanism CSAM proposed in this paper, different modules were added to the YOLOv8 detection algorithm under the same experimental conditions, and the influence of each module on the performance of the detection algorithm was evaluated. The results are shown in Table 2. In the added attention module, CSAM improved the accuracy of the detection algorithm the most, which was 3.81 percentage points, while CSAM also ensured fewer parameters, less computation, and the real-time performance of the algorithm.

Table 2. Comparative results of the ablation experiments.

Models	Params/10^6	FLOPs/10^9	mAP50/%	FPS
YOLOv8	25.90	78.9	69.21	18.87
YOLOv8+CBAM	32.07	104.6	71.92	6.58
YOLOv8+CSAM(Ours)	27.20	96.5	73.02	9.72

3.3.6. Comparative Test

To verify the superiority of the improved detection algorithm, three mainstream detection algorithms were selected for comparative experiments, as shown in Table 3. When the input sizes were all set to 640 × 640 pixels, the detection accuracy of the improved detection algorithm in this paper was better than other algorithms based on ensuring real-time detection. Compared to the Faster RCNN, mAP50 increased by 18.06 percentage points, while Params and FLOPs decreased by 1.59×10^8 and 8.56×10^{10}, respectively. Compared to YOLOv5, the mAP50 of this algorithm increased by 8.24%. On the whole, the improved detection algorithm added the attention module CSAM to the backbone network, which improved the feature extraction ability of small targets and made the model better in detecting fish sonar images.

Table 3. Comparative experimental results of the different detection algorithms.

Models	Size/Pixel	Params/10^6	FLOPs/10^9	mAP50/%	FPS
Faster RCNN	640 × 640	186.3	182.1	54.96	2.00
SDD	640 × 640	23.8	188.0	53.00	2.86
YOLOv5	640 × 640	7.2	16.5	64.78	18.01
YOLOv8+CSAM(Ours)	640 × 640	27.2	96.5	73.02	9.72

3.3.7. Comparison of the Detection Images

A comparison between the algorithm in this paper and YOLOv8 in detecting fish sonar images without an attention mechanism is presented in Figure 12. Figure 12a–c show our algorithm and Figure 12d–f show YOLOv8 in this paper. The upper and lower parts correspond to the same frame image. It can be seen that the model can distinguish the fish from the netting, and the detected fish was selected by the red identification box. By comparing Figure 12a and Figure 12d, it can be observed that the algorithm in this paper has detected the leftmost small fish, but YOLOv8 has not, and instead mislabeled the rightmost blackfish. Comparing with Figure 12b and Figure 12e, it can also be observed that the algorithm in this paper recognized one more small fish than the original algorithm. Figure 12c turns on the label and confidence display, and it can be seen that the average confidence of fish identification was higher than 80%, which shows that the neural network model can identify fish well.

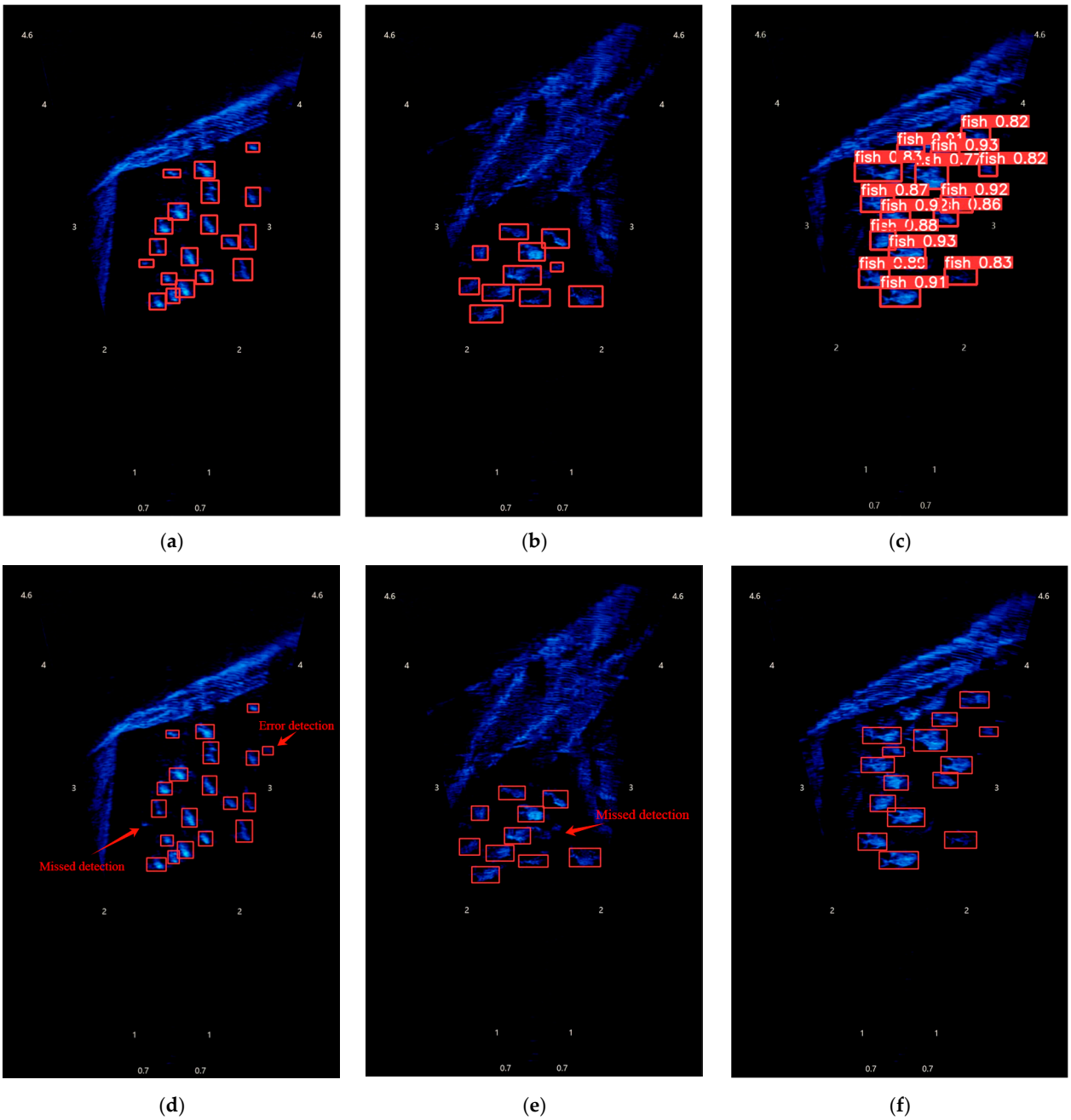


Figure 12. Comparison of detection images. (**a**–**c**) our algorithm; (**d**–**f**) YOLOv8.

4. Data Fitting

4.1. Introduction to the BP Neural Network

In this study, the estimation of the detected fish quantity to the actual quantity was a nonlinear mapping problem. Neural networks boast strong applicability in dealing with nonlinear mapping and are considered an effective method of data fitting and widely used [28].

The BP (back propagation) neural network is a widely used algorithm at present. The training steps are: initializing the weights and thresholds of each layer, inputting sample data in the input layer, and finally outputting the results in the output layer after calculation in the hidden layer. In the process of the forward transmission of each layer, the current layer only affects the adjacent next layer. If the results of the output layer do not meet the expected output value, the error with the expected value will be propagated back to

the network, so that the error function will decrease along the negative gradient direction [29].

The BP neural network includes one input layer, one or more hidden layers, and one output layer. The basic topological structure of the BP neural network (taking one hidden layer as an example) is shown in Figure 13.

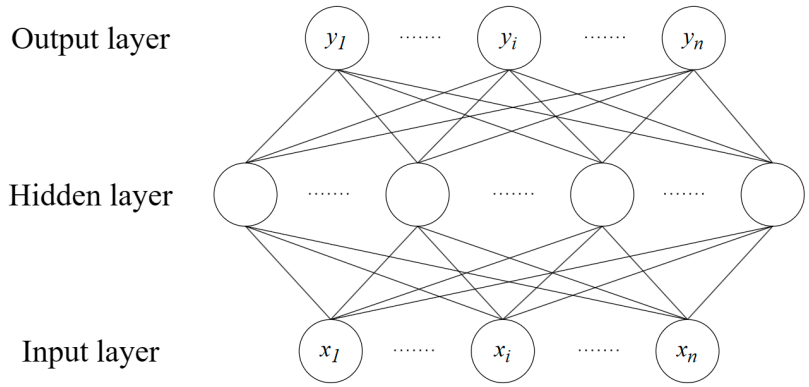


Figure 13. Topology of the neural network.

In neurons, the input acts on another function after a series of weighted summations, and this function is the activation function here. The function of the activation function in a neural network is to transform multiple linear inputs into nonlinear relationships, to achieve the mapping function from linear to nonlinear. The definition of a sigmoid function is shown in Formula (6) [30].

$$Sigmoid(x) = \frac{1}{1 + e^{-x}} \tag{6}$$

4.2. Experimental Analysis

4.2.1. Training Data

To automatically obtain the fish quantity in the cage, human subjective factors and manual intervention should be minimized. In this paper, all the images collected by sonar were selected for target recognition and detection. Sonar data were divided into four groups: 20 fish, 40 fish, 60 fish, and 80 fish, and each group had 10 continuous detection videos with a frame rate of 15 frames per second. Seven videos from each group were randomly selected as the fitting data, and the remaining three were used as detection data. These 40 sonar videos were detected by this algorithm, and the identification data of each frame was saved as a text file.

Each 10-minute video had nearly 10,000 images. If all such vast data were used for fitting, it would not only be a vast amount of calculation but also make it difficult for the algorithm to learn the key features of the data. Considering that the goal is to obtain the fish quantity in the cage, and there was a certain mapping relationship between the quantity of fish in the detection image and the actual quantity, the amount of fish detected in a single frame in each video was sorted from large to small in this paper, taking the top 30 fish quantity detected. The statistical results are shown in Figure 14.

Serial No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
Actual quantity	20	20	20	20	20	20	20	40	40	40	40	40	40	40	60	60	60	60	60	60	60	80	80	80	80	80	80	80
Quantity detected 1	11	13	13	15	14	12	14	18	19	18	20	19	18	20	16	17	20	24	23	27	20	23	24	28	25	24	27	26
Quantity detected 2	11	12	13	13	14	12	13	18	18	18	20	19	18	20	15	17	20	20	22	27	19	22	23	27	25	23	26	26
Quantity detected 3	11	12	12	12	13	11	12	17	18	18	20	18	18	19	15	16	19	20	22	25	19	22	23	27	24	22	25	25
Quantity detected 4	10	11	12	11	12	11	12	17	17	18	19	18	18	19	15	16	17	19	22	24	19	22	23	26	24	22	25	24
Quantity detected 5	10	11	11	11	12	11	12	17	17	18	19	18	18	19	15	16	17	19	20	24	19	22	22	24	24	22	25	23
Quantity detected 6	10	11	11	11	11	11	12	17	17	18	19	17	18	19	15	16	17	19	20	24	18	22	22	24	24	22	25	23
Quantity detected 7	10	11	11	11	11	11	11	17	17	18	18	17	18	19	15	16	17	19	19	24	18	21	22	24	24	22	24	23
Quantity detected 8	10	11	11	11	11	10	11	16	17	18	18	17	18	19	15	16	17	18	19	24	18	21	22	23	24	22	24	23
Quantity detected 9	10	10	11	11	11	10	11	16	17	18	18	17	17	19	14	16	17	18	19	24	18	20	21	23	23	21	24	23
Quantity detected 10	10	10	11	11	11	10	11	16	17	18	18	17	17	18	14	16	17	18	19	24	18	20	20	23	23	21	24	23
Quantity detected 11	10	10	11	11	11	10	11	16	16	17	18	17	17	18	14	16	16	18	19	23	17	20	20	22	23	21	24	23
Quantity detected 12	10	10	11	11	11	10	11	16	16	17	18	17	17	18	14	15	16	17	19	23	17	20	20	22	23	21	24	23
Quantity detected 13	10	10	10	11	11	9	11	16	16	17	18	17	17	18	14	15	16	17	19	23	17	20	20	21	23	21	24	22
Quantity detected 14	10	10	10	11	11	9	10	16	16	17	18	16	17	18	14	15	16	17	18	23	17	19	20	21	23	21	24	22
Quantity detected 15	10	10	10	11	11	9	10	16	16	17	18	16	16	18	14	15	16	17	18	23	17	19	20	21	22	21	24	22
Quantity detected 16	10	9	10	11	11	9	10	16	16	17	18	16	16	17	14	15	16	17	18	23	17	19	19	21	22	21	24	22
Quantity detected 17	10	9	10	11	11	9	10	16	16	17	18	16	16	17	14	15	16	17	18	23	17	19	19	21	22	21	24	22
Quantity detected 18	10	9	10	11	11	9	10	16	16	17	18	16	16	17	13	15	16	17	18	23	16	19	19	21	22	21	24	22
Quantity detected 19	10	9	10	10	11	9	10	16	16	17	18	16	16	17	13	15	16	17	18	22	16	19	19	20	22	21	21	22
Quantity detected 20	10	9	10	10	10	9	10	15	16	17	17	16	16	17	13	15	16	17	18	22	16	18	19	20	22	21	21	22
Quantity detected 21	9	9	10	10	10	9	10	15	16	17	17	16	16	17	13	15	16	17	18	22	16	18	19	20	21	21	21	22
Quantity detected 22	9	9	10	10	10	9	9	15	16	17	17	16	16	17	13	15	16	17	18	22	16	18	18	20	21	21	21	20
Quantity detected 23	9	9	10	10	10	9	9	15	16	17	17	16	16	17	13	15	16	17	18	22	16	18	18	20	21	21	21	19
Quantity detected 24	9	9	10	10	10	9	9	15	16	16	17	16	16	17	13	15	15	17	18	22	16	18	18	20	21	21	21	19
Quantity detected 25	9	9	10	10	10	9	9	15	16	16	17	16	15	17	13	15	15	17	18	22	16	18	18	20	21	20	21	19
Quantity detected 26	9	9	10	10	10	9	9	15	16	16	17	16	15	17	13	15	15	17	17	22	16	18	18	20	20	20	21	19
Quantity detected 27	9	9	10	10	10	9	9	15	16	16	17	16	15	17	13	15	15	17	17	22	16	18	18	20	20	20	21	19
Quantity detected 28	9	9	10	10	10	9	9	15	16	16	17	16	15	17	13	15	15	17	17	22	15	18	18	20	20	20	21	19
Quantity detected 29	9	9	10	10	10	9	9	15	15	16	17	16	15	17	13	15	15	17	17	22	15	18	18	19	20	20	21	19
Quantity detected 30	9	9	10	10	10	9	9	15	15	16	17	15	15	17	13	15	15	17	17	22	15	18	18	19	20	20	21	19

Figure 14. Statistical diagram of the maximum quantity detected.

4.2.2. Evaluation Indicators

In the process of neural network training, the error between the predicted or fitted data and the measured data can be expressed by the *MSE* (mean square error), as shown in Formula (7):

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 \quad (7)$$

In Equation (7), "n" represents the data quantity, "Y_i" represents the measured data, and "$\hat{Y}_i$" represents the predicted or fitted data based on the neural network model.

4.2.3. Training Process

The top 10, top 20, and top 30 fish abundance detected were input into the network for training, and the fitting target was the corresponding actual quantity. After comparative experiments, the best effect parameters were the top 20 fish quantity detected in fitting, and the best number of neurons in the hidden layer was 30. Bayesian regularization was used for training. There were 28 groups of data, 85% of which were randomly selected as training data and the remaining 15% as test data.

Based on the above parameters, the neural network was trained, and the training results are shown in Figures 15 and 16. Figure 15 shows the change in the sample mean square error. After 45 training operations, the MSE of the training group produced the best result, with a value of 85.1716. Figure 16 shows the prediction errors of the training group and the test group, in which the vast majority of sample errors were between −12 and 12, with positive numbers indicating that the prediction was greater than the actual quantity and negative numbers indicating that the prediction was lower than the actual quantity.

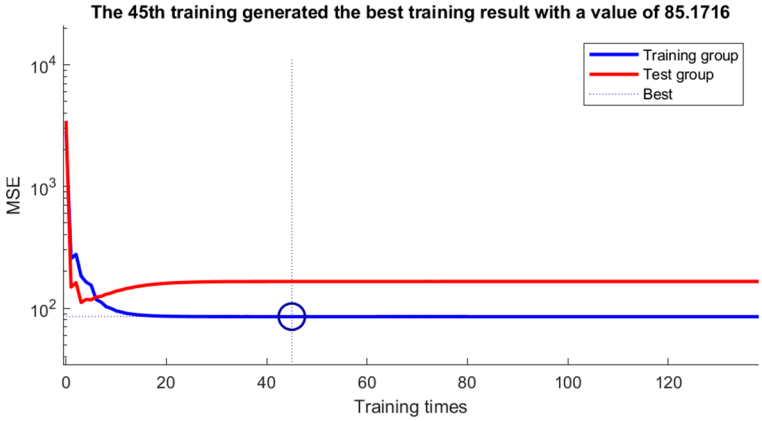


Figure 15. The sample mean square error.

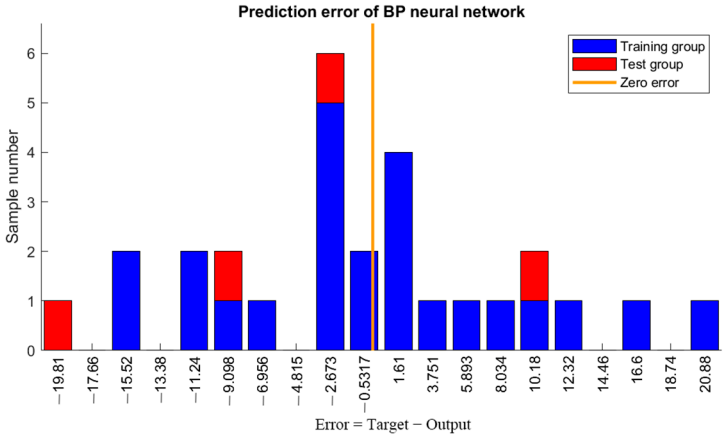


Figure 16. Prediction error of the training group and the test group.

The learning results of the BP neural network are shown in Figure 17. The regression results of the training group, the test group, and all data, that is, the fitting degree between the output value and the target value, are shown in these three small graphs. As it can be seen from the figure, most of the data are concentrated near the diagonal, and some data are far away, and the fitting results are all above 0.82, indicating that the fitting effect is relatively good.

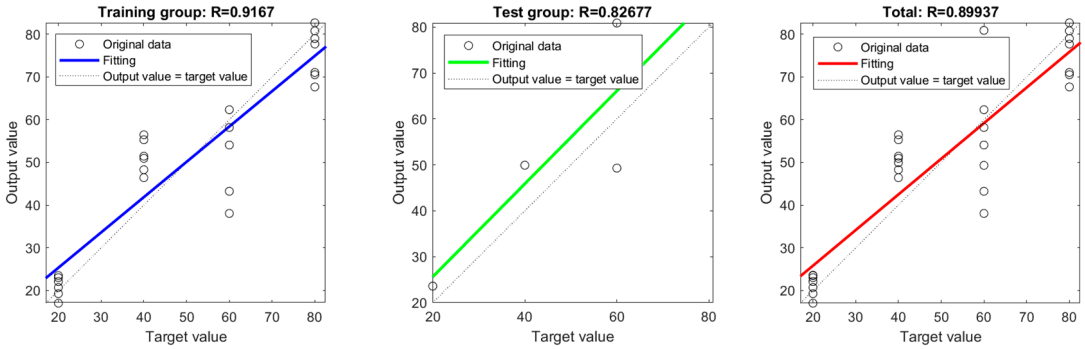


Figure 17. Regression results of the BP neural network model.

4.2.4. Fitting Test

The BP neural network was used to estimate the top 20 fish quantity of the three tests in each group of test data, and the fitting results are shown in Table 4. Error number = total fitting quantity-actual quantity, error percentage = absolute value of error quantity/actual

quantity; the average error was the average of all error percentages and average accuracy = 1 − average error.

Table 4. The statistical results of the method in this paper on the test dataset.

Serial No.	Actual Quantity	Maximum Quantity Detected	Fitting Total Quantity	Error Quantity	Error Percentage/%	Precision Percentage/%
1	20	12	17.68	−2.32	11.60	88.4
2		14	26.63	6.63	33.13	66.87
3		13	24.48	4.48	22.40	77.6
4	40	19	44.91	4.91	12.27	87.73
5		16	35.59	−4.41	11.02	88.98
6		17	37.87	−2.13	5.34	94.66
7	60	22	52.56	−7.44	12.40	87.6
8		23	68.46	8.49	14.10	85.9
9		22	60.99	0.99	1.66	98.34
10	80	21	74.42	−5.78	6.97	93.03
11		13	58.00	−22.00	27.50	72.5
12		18	59.19	−20.81	26.01	73.99
				Average	15.37	84.63

It can be seen from Table 4 that the algorithm in this paper had a high accuracy in fitting the sonar image data of 20, 40, and 60 groups and achieved a single-digit error. However, when fitting the sonar data of 80 fish, the error was large, and the quantity sequence detected was small, resulting in a large error of about 27%. The manual inspection of the detection videos with serial numbers 11 and 12 showed that there were few fish in the sonar images. It was speculated that the sonar probe shook badly during this period due to heavy sea waves, and the swimming trajectory of the fish was different from the usual one; so, the data detected by the sonar did not reflect the real situation in the cage. The solution can be to observe in multiple periods, obtain multiple groups of sonar image data and carry out target recognition and detection, eliminate detection sequences with too large data differences, and then estimate by a neural network. The obtained data were more objective and more realistic after averaging.

4.2.5. Data Fitting and Comparison

The commonly used data fitting methods are linear fitting and polynomial fitting. Because they can only deal with one-to-one mapping relationships, it is necessary to extract key data from the detection sequence [31]. In this paper, the quantity of fish detected in a single image in each video was sorted from large to small, and the maximum quantity of fish detected, the average of the top 10 fish quantity, and the average of the top 20 fish quantity were statistically analyzed.

The training data and neural network fitting were the same. Firstly, linear fitting and cubic polynomial fitting were carried out for these three statistical data, and the results are shown in Figure 18. The upper left corner of each small graph shows the fitting formula and fitting coefficient R^2, which reflect the overall accuracy of the model, that is, the fitting degree. The closer its value is to 1 shows that the model accurately reflects the changes in the observed data, and the better the reliability of the data. It can be seen from the figure that the R^2 of cubic polynomial fitting is greater than the corresponding linear fitting, and the fitting results of the average of the top 10 fish quantity in the two fitting methods are better than those of the maximum quantity of detected and the average of the top 20 fish quantity.

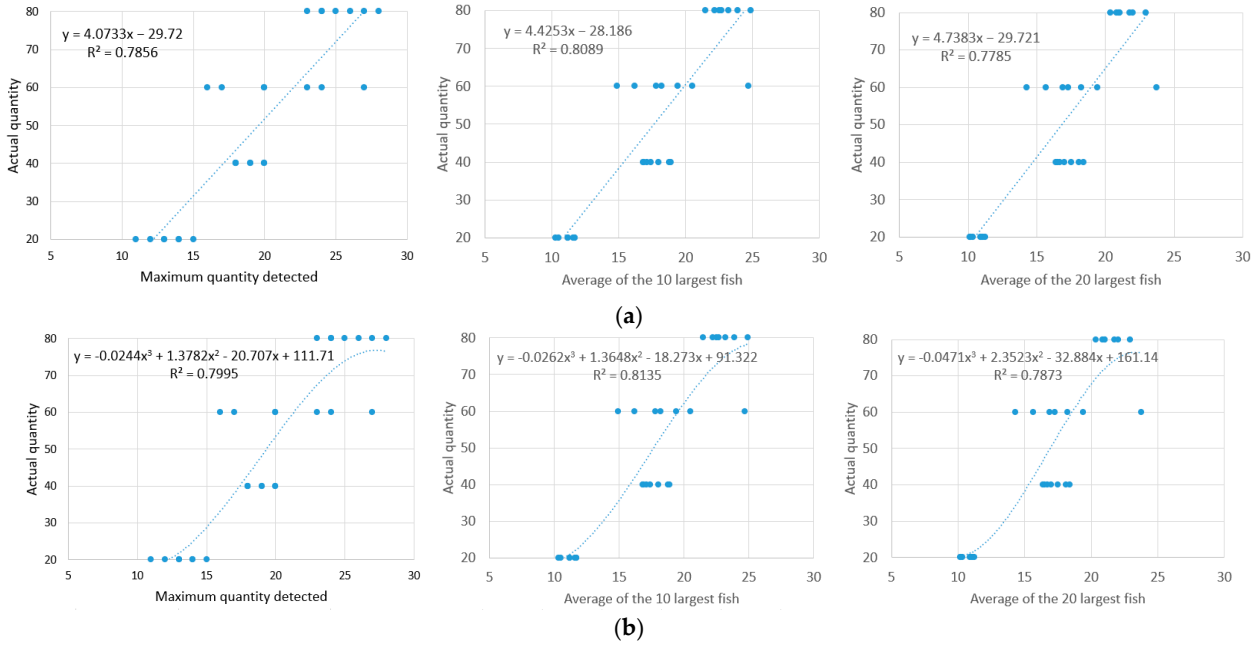


Figure 18. Comparison of the data fitting results. (**a**) Linear fitting; (**b**) cubic polynomial fitting.

When the polynomial fitting was performed on the average of the top 10 fish quantity, the fitting results of the quadratic, cubic, and quartic polynomials are compared as shown in Figure 19. It can be seen that the best fitting result of the quartic polynomial was $R^2 = 0.8387$, but the highest term was too high, which leads to a better effect on sample data, but the effect of test data will decline, that is, there will be over-fitting. Generally, the highest term was not higher than three times when the polynomial fitting was used.

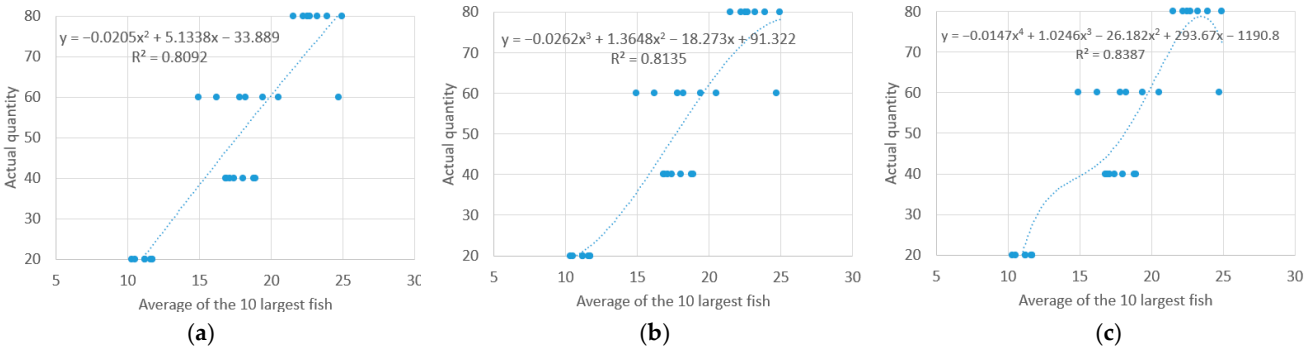


Figure 19. Comparison of the fitting results of higher order polynomials. (**a**) Quadratic polynomial; (**b**) cubic polynomial; and (**c**) quartic polynomial.

To sum up, the cubic polynomial was selected to fit the fish quantity in the comparison test, and the equation is shown in Formula (8):

$$y = -0.0262x^3 + 1.3648x^2 - 18.273x + 91.322 \quad (8)$$

where x was the average of the top 10 fish abundance detected, y was the estimated fish quantity in the cage, and R^2 of the equation was 0.8135, which can be understood as the theoretical accuracy of data fitting as about 81.35%.

Formula (8) was used to estimate the average of the top 10 fish quantity of the three tests in each group of test data, and the fitting results are shown in Table 5. It can be seen that the average accuracy is 83.58%, which is lower than the 84.63% of neural network fitting.

Table 5. The statistical results of high-order polynomial fitting on the test dataset.

Serial No.	Actual Quantity	Average of the Top 10 Fish Quantity Detected	Fitting Total Quantity	Error Quantity	Error Percentage/%	Precision Percentage/%
1		12.6	25.35	5.35	26.74	73.26
2	20	10.5	19.59	−0.41	2.03	97.97
3		11.3	21.30	1.30	6.52	93.48
4		15.3	37.39	−2.61	6.52	93.48
5	40	14.5	33.44	−6.56	16.40	83.6
6		16.7	44.77	4.77	11.92	88.08
7		19	57.12	−2.88	4.80	95.2
8	60	21.8	70.14	10.14	16.90	83.1
9		18.2	52.88	−7.12	11.86	88.14
10		20.7	65.49	−14.51	18.14	81.86
11	80	17.3	48.01	−31.99	39.98	60.02
12		18	51.80	−28.20	35.24	64.76
				Average	16.42	83.58

5. Discussion

5.1. Comparison of the Fish Quantity Estimation Methods

The traditional methods to obtain the fish abundance in cages are the mark–recapture method, fish finder measurement, annular underwater acoustic multi-beam detection, and others [32]. Due to the impossibility of conducting comparison experiments in the same environment, the instruments and equipment used in various methods vary. At present, there are few reports on the estimation algorithm and estimation accuracy of fish abundance in cages. In this paper, a comparison table of the different estimation methods was made based on previous studies by scholars, which is shown in Table 6.

Table 6. Comparison of the different estimation methods.

Methods	Equipment Used	Precision	Advantages	Disadvantages
Mark–recapture method [33]	Fishing net, stain	Large discrete interval	No electronic equipment is needed	Low precision, time-consuming, and laborious, affecting the growth of fish
Fish finder measurement [34]	Fish detector	About 50%	Low equipment cost	Low accuracy, fish density, sometimes vast errors
Annular underwater acoustic multi-beam detection [35]	Annular multi-beam detector	60%-70%	Wide detection angle, high precision	Expensive equipment, difficult layout
The method in this study	Image sonar	About 84%	High precision, automatic measurement, simple layout	Expensive equipment

It can be seen from Table 6 that the forward-looking image sonar used in the method presented in this paper is more expensive than the equipment used in previous methods, and the average purchase unit price is USD 30,000, but the layout is relatively simple. After the sonar is installed, the data can be obtained and processed automatically, and the estimated fish abundance in the cage can be obtained without additional manual intervention. Compared to the traditional methods, the accuracy of this method is significantly improved, reaching about 84%.

As a high-definition image sonar, ARIS1800 is widely used in fishery. Both at home and abroad, image sonar is mainly used in the study of fish behavior, rather than in the assessment of fish quantity. This paper makes a very meaningful attempt to evaluate the quantity of fish in cages by using the imaging characteristics of ARIS1800, based on fixed detection and prediction methods. ARIS1800 can display the size, shape, and position of fish in the cage with high-definition images. It eliminates the limitation of traditional fish

finders only being able to assess fish quantities by target strength, achieving a higher credibility.

5.2. Error Analysis

The estimation of fish quantity in cages is a major challenge in fish acoustics research, which is influenced by various factors: complicated and changeable ocean factors, such as wind and waves and tidal currents in aquaculture areas; feeding, sailing, and other interferences; transducer reverberation blind area and strong sea-floor reflection; obscuration of beam detection by fish in dense schools; some fish swim close to the wall, which make the fish echo and the net echo overlap and difficult to distinguish; and repeated detection caused by swimming fish [36]. These uncontrollable factors cause the data collected by sonar at different times to be inconsistent, and in turn, the estimated neural network model has inevitable errors, affecting the final estimated quantity of fish.

Figure 20 is a histogram of the estimation and error of the fitting test in Table 4. The error of the neural network estimation in groups 1–10 is relatively small, and the prediction results in groups 11 and 12 are affected by the large wave fluctuation. By observing the cages in different periods, it was possible to estimate the average value of multiple groups of data to reduce the error.

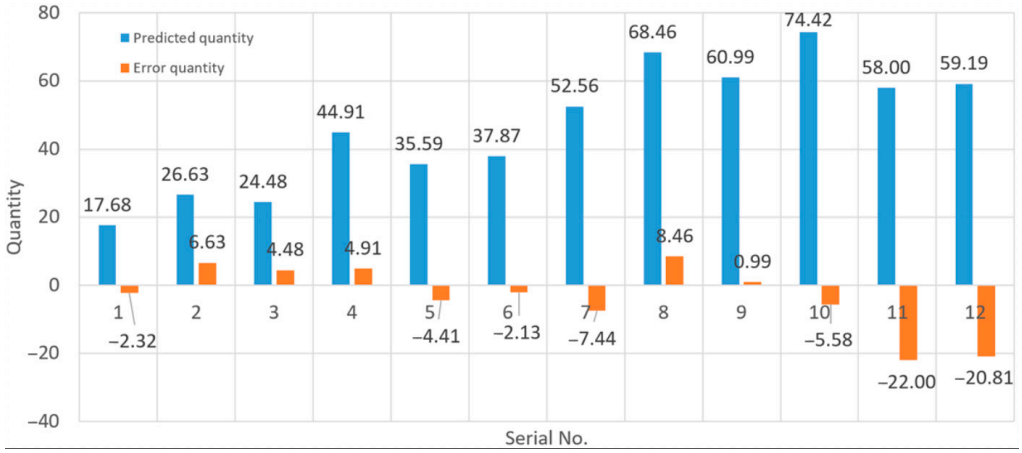


Figure 20. Estimation and error bar chart of the fitting test.

Given the measurement results of 80 fish, three additional observation data at different times were selected in this paper, and a new test dataset was formed together with the three data in the previous fitting test. The fish quantity was estimated by using this method, and the results are shown in Table 7.

Table 7. The statistical results of the method in this paper across multiple time periods of datasets.

Serial No.	Actual Quantity	Testing Time	Fitting Total Quantity	Error Quantity	Error Percentage/%	Precision Percentage/%
1	80	9:00–9:15	74.42	−5.78	6.97	93.03
2		9:30–9:45	58.00	−22.00	27.50	72.5
3		10:00–10:15	59.19	−20.81	26.01	73.99
4		13:00–13:15	82.32	2.32	2.9	97.1
5		15:00–15:15	72.45	−7.55	9.44	90.56
6		17:00–17:15	69.45	−10.55	13.19	86.81
		Average	69.31	−10.69	14.33	85.67

Table 7 reveals the fluctuations in the data measured and predicted in different periods, but only the two groups of data with serial numbers 2 and 3 have a deviation of 20, with the other groups having an error of less than 10. The average value of six groups of

data prediction was 69.31, the error was −10.69, and the average accuracy was 85.67%, which was significantly improved by 5.83 percentage points compared to the average accuracy of 79.84% of data only using the same period.

6. Conclusions

This paper proposes a method for estimating the quantity of fish in net cages based on forward-looking imaging sonar. The method first investigates the YOLO neural network model and makes improvements for underwater fish identification tasks. An attention mechanism is introduced into the YOLO model construction, allocating more computing power to focus on small targets, thereby enhancing the performance of the region of interest, especially for small targets. Through ablation experiments, the addition of the CSAM module is shown to improve the accuracy of the detection algorithm by 3.81 percentage points, and compared to the YOLOv5, the improved algorithm in this paper increases the mAP50 by 8.24 percentage points. Subsequently, quantitative detection experiments for 80 oval damselfish are conducted in the constructed fish cages. Due to the limited visual angle of the sonar, the experiments are conducted by deploying the sonar on one side of the net cage and continuously observing to obtain video images. The improved and trained YOLOv8 model is used to detect fish shoals in sonar images. The detection quantity results are sorted from large to small, and the quantity of fish in the net cage is estimated based on the top 20 maximum counts using a trained BP neural network. Multiple experimental results show that the average accuracy of fish quantity assessment reaches 84.63%, validating the feasibility of this method.

Through research on detection methods, target identification, and quantity inversion of fish in net cages, a new method for estimating the quantity of fish in net cage farming based on imaging sonar has been developed. This method achieves a high-precision assessment of fish quantity in net cage farming, providing technical support for the development of intelligent equipment for net cages in China.

Nevertheless, there are still the following problems in this research method, which need to be improved in future research:

1. In the part of fish target recognition, the background of the image is not removed in advance, and the netting in the background fluctuates with the waves. In some cases, fish will swim against the netting, and the two are mixed in the sonar image, which will affect the fish recognition effect of the YOLO model and make the recognition quantity fluctuate [37];
2. The YOLO target detection model and neural network prediction model used in this method are highly dependent on training data. For this reason, quantitative fish data collection should be carried out under the condition that the cage size and sonar layout are consistent before practical application. The above two models can only be applied to the fish quantity prediction after learning the collected data. As for the simplification of the model training process and the production of general datasets, further in-depth research is needed;
3. The quantitative experiment in this paper was carried out in a small fishing raft, and it is planned to be applied to a large deep-sea cage in the future. With the increase in the cage scale and the quantity of fish, the density of fish will increase obviously, and more fish will overlap and block each other. In theory, when detecting training data, the situation of fish occlusion is roughly the same as that when estimating the quantity, and the neural network will be relatively accurate when fitting the total quantity. However, as to whether the actual prediction effect can meet the precision of a small-scale quantitative experiment, it still needs to be tested.

Author Contributions: Conceptualization, G.Z. and Y.H.; methodology, M.L.; software, J.H.; validation, L.X.; investigation, D.L.; resources, Y.H.; data curation, G.Z.; writing—original draft preparation, G.Z.; writing—review and editing, C.D., X.H., Y.H. and J.S.; supervision, Y.H.; project

administration, X.H.; funding acquisition, X.H. and Y.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received funding from the Major Science and Technology Plan of Hainan Province (Grant No. ZDKJ2021013), Hainan Province Science and Technology Special Fund (Grant No. ZDYF2021XDNY305, ZDYF2023XDNY066), Central Public-interest Scientific Institution Basal Research Fund, CAFS (Grant No. 2023TD97), Central Public-interest Scientific Institution Basal Research Fund, South China Sea Fisheries Research Institute, CAFS (No. 2022TS06), and Project supported by Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai) (No. SML2023SP204).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article.

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

References

1. Garcia Serge, M.; Rosenberg Andrew, A. Food security and marine capture fisheries: Characteristics, trends, drivers and future perspectives. *Philos. Trans. R. Soc. B* **2010**, *365*, 2869–2880.
2. Yu, J.; Yan, T. Analyzing Industrialization of Deep-Sea Cage Mariculture in China: Review and Performance. *Rev. Fish. Sci. Aquac.* **2023**, *31*, 483–496.
3. Huan, X.; Shan, J.; Han, L.; Song, H. Research on the efficacy and effect assessment of deep-sea aquaculture policies in China: Quantitative analysis of policy texts based on the period 2004–2022. *Mar. Policy* **2024**, *160*, 105963.
4. Kleih, U.; Linton, J.; Marr, A.; Mactaggart, M.; Naziri, D.; Orchard, J.E. Financial services for small and medium-scale aquaculture and fisheries producers. *Mar. Policy* **2013**, *37*, 106–114.
5. Baumgartner, L.J.; Reynoldson, N.; Cameron, L.; Stanger, J. Assessment of a Dual-Frequency Identification Sonar (DIDSON) for Application in Fish Migration Studies. *Fish. Final Rep.* **2006**, *84*, 1449–1484.
6. Shahrestani, S.; Bi, H.; Lyubchich, V.; Boswell, K.M. Detecting a nearshore fish parade using the adaptive resolution imaging sonar (ARIS): An automated procedure for data analysis. *Fish. Res.* **2017**, *191*, 190–199.
7. Guan, M.; Cheng, Y.; Li, Q.; Wang, C.; Fang, X.; Yu, J. An Effective Method for Submarine Buried Pipeline Detection via Multi-sensor Data Fusion. *IEEE Access* **2019**, *7*, 125300–125309.
8. Qiu, Z.W.; Jiao, M.L.; Jiang, T.C.; Zhou, L. Dam Structure Deformation Monitoring by GB-InSAR Approach. *IEEE Access* **2020**, *8*, 123287–123296.
9. Liu, Y.; Wang, R.; Gao, J.; Zhu, P. The Impact of Different Mapping Function Models and Meteorological Parameter Calculation Methods on the Calculation Results of Single-Frequency Precise Point Positioning with Increased Tropospheric Gradient. *Math. Probl. Eng.* **2020**, *35*, 9730129.
10. Sun, P.; Zhang, K.; Wu, S.; Wang, R.; Wan, M. An investigation into real-time GPS/GLONASS single-frequency precise point positioning and its atmospheric mitigation strategies. *Meas. Sci. Technol.* **2021**, *32*, 115018.
11. Cai, J.; Zhang, Y.; Li, Y.; Liang, X.S.; Jiang, T. Analyzing the Characteristics of Soil Moisture Using GLDAS Data: A Case Study in Eastern China. *Appl. Sci.* **2017**, *7*, 566.
12. Feng, Y.; Wei, Y.; Sun, S.; Liu, J.; An, D.; Wang, J. Fish abundance estimation from multi-beam sonar by improved MCNN. *Aquat. Ecol.* **2023**, *57*, 895–911.
13. Viswanatha, V.; Chandana, R.K.; Ramachandra, A.C. Real-Time Object Detection System with YOLO and CNN Models: A Review. *arXiv* **2022**, arXiv:2208.00773.
14. He, S.; Lu, X.; Gu, J.; Tang, H.; Yu, Q.; Liu, K.; Ding, H.; Chang, C.; Wang, N. RSI-Net: Two-Stream Deep Neural Network for Remote Sensing Imagesbased Semantic Segmentation. *IEEE Access* **2022**, *10*, 34858–34871.
15. Ye, Z.B.; Duan, X.H.; Zhao, C. Research on Underwater Target Detection by Improved YOLOv3-SPP. *Comput. Eng. Appl.* **2023**, *59*, 231–240.
16. Chen, Y.L.; Dong, S.J.; Zhu, S.K. Detection of underwater biological targets in shallow water based on improved YOLOv3. *Comput. Eng. Appl.* **2023**, *59*, 190–197.
17. Guo, H.; Li, R.; Xu, F.; Liu, L. Review of research on sonar imaging technology in China. *Chin. J. Oceanol. Limnol.* **2013**, *31*, 1341–1349.
18. Shen, W.; Peng, Z.; Zhang, J. Identification and counting of fish targets using adaptive resolution imaging sonar. *J. Fish Biol.* **2023**, *104*, 422–432.
19. Kang, C.H.; Kim, S.Y. Real-time object detection and segmentation technology: An analysis of the YOLO algorithm. *JMST Adv.* **2023**, *5*, 69–76.
20. Wang, Z.; Zhou, D.; Guo, C.; Zhou, R. Yolo-global: A real-time target detector for mineral particles. *J. Real-Time Image Process.* **2024**, *21*, 85.

21. Lü, H.; Xie, J.; Xu, J.; Chen, Z.; Liu, T.; Cai, S. Force and torque exerted by internal solitary waves in background parabolic current on cylindrical tendon leg by numerical simulation. *Ocean Eng.* **2016**, *114*, 250–258.
22. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
23. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-cam: Visual explanations from deep networks via gradient-based localization. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 618–626.
24. Wei, X.; Wang, Z. TCN-attention-HAR: Human activity recognition based on attention mechanism time convolutional network. *Sci. Rep.* **2024**, *14*, 7414.
25. Cui, Z.; Wang, N.; Su, Y.; Zhang, W.; Lan, Y.; Li, A. ECANet: Enhanced context aggregation network for single image dehazing. *Signal Image Video Process.* **2023**, *17*, 471–479.
26. Guo, M.H.; Lu, C.Z.; Liu, Z.N.; Cheng, M.M.; Hu, S.M. Visual attention network. *Comput. Vis. Media* **2023**, *9*, 733–752.
27. Zhu, G.; Shen, Z.; Liu, L.; Zhao, S.; Ji, F.; Ju, Z.; Sun, J. AUV dynamic obstacle avoidance method based on improved PPO algorithm. *IEEE Access* **2022**, *10*, 121340–121351.
28. Liu, J.; Yu, L.; Sun, L.; Tong, Y.; Wu, M.; Li, W. Fitting objects with implicit polynomials by deep neural network. *Optoelectron. Lett.* **2023**, *19*, 60–64.
29. Zhang, J.; He, X. Earthquake magnitude prediction using a VMD-BP neural network model. *Nat. Hazards* **2023**, *117*, 189–205.
30. Nabizadeh, E.; Parghi, A. Artificial neural network and machine learning models for predicting the lateral cyclic response of post-tensioned base rocking steel bridge piers. *Asian J. Civ. Eng.* **2024**, *25*, 511–523.
31. Guan, M.; Li, Q.; Zhu, J.; Wang, C.; Zhou, L.; Huang, C.; Ding, K. A method of establishing an instantaneous water level model for tide correction. *Ocean Eng.* **2019**, *171*, 324–331.
32. Zhang, X.; Xu, X.; Peng, Y.; Hong, H. Centralized Remote Monitoring System for Bred Fish in Offshore Aquaculture Cages. *Trans. Chin. Soc. Agric. Mach.* **2012**, *43*, 178–182+187.
33. Lin, W.Z.; Chen, Z.X.; Zeng, C.; Karczmarski, L.; Wu, Y. Mark-recapture technique for demographic studies of Chinese white dolphins—Applications and suggestions. *Acta Theriol. Sin.* **2018**, *38*, 586–596.
34. Garg, R.; Phadke, A.C. Enhancing Underwater Fauna Monitoring: A Comparative Study on YOLOv4 and YOLOv8 for Real-Time Fish Detection and Tracking. In *Artificial Intelligence and Sustainable Computing*; Pandit, M., Gaur, M.K., Kumar, S., Eds.; ICSISCET 2023. Algorithms for Intelligent Systems. Springer, Singapore, 2024.
35. Connolly, R.M.; Jinks, K.I.; Shand, A.; Taylor, M.D.; Gaston, T.F.; Becker, A.; Jinks, E.L. Out of the shadows: Automatic fish detection from acoustic cameras. *Aquat. Ecol.* **2023**, *57*, 833–844.
36. Li, D.; Du, L. Recent advances of deep learning algorithms for aquacultural machine vision systems with emphasis on fish. *Artif. Intell. Rev.* **2022**, *55*, 4077–4116.
37. Maki, T.; Horimoto, H.; Ishihara, T.; Kofuji, K. Tracking a Sea Turtle by an AUV with a Multibeam Imaging Sonar: Toward Robotic Observation of Marine Life. *Int. J. Control. Autom. Syst.* **2020**, *18*, 597–604.