

Article

A Method for Estimating the Injection Position of Turbot (*Scophthalmus maximus*) Using Semantic Segmentation

Wei Luo ^{1,2}, Chen Li ^{1,3} , Kang Wu ^{1,2} , Songming Zhu ^{1,2,3}, Zhangying Ye ^{1,2,3,*} and Jianping Li ^{1,2,3,*}

¹ College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou 310058, China

² Key Laboratory of Equipment and Informatization in Environment Controlled Agriculture, Ministry of Agriculture and Rural Affairs, Hangzhou 310058, China

³ Key Laboratory of Intelligent Equipment and Robotics for Agriculture of Zhejiang Province, Hangzhou 310058, China

* Correspondence: yzyzju@zju.edu.cn (Z.Y.); jpli@zju.edu.cn (J.L.)

Abstract: Fish vaccination plays a vital role in the prevention of fish diseases. Inappropriate injection positions will cause a low immunization rate and even death. Currently, traditional visual algorithms have poor robustness and low accuracy due to the specificity of the placement of turbot fins in the application of automatic vaccination machines. To address this problem, we propose a new method for estimating the injection position of the turbot based on semantic segmentation. Many semantic segmentation networks were used to extract the background, fish body, pectoral fin, and caudal fin. In the subsequent step, the segmentations obtained from the best network were used for calculating body length (BL) and body width (BW). These parameters were employed for estimating the injection position. The proposed Atten-Deeplabv3+ achieved the best segmentation results for intersection over union (IoU) on the test set, with 99.3, 96.5, 85.8, and 91.7 percent for background, fish body, pectoral fin, and caudal fin, respectively. On this basis, the estimation error of the injection position was 0.2 mm–4.4 mm, which is almost within the allowable injection area. In conclusion, the devised method was able to correctly differentiate the fish body from the background and fins, meaning that the extracted area could be successfully used for the estimation of injection position.

Keywords: turbot; vaccination; deeplabv3+; attention mechanism; measurement; aquaculture



Citation: Luo, W.; Li, C.; Wu, K.; Zhu, S.; Ye, Z.; Li, J. A Method for Estimating the Injection Position of Turbot (*Scophthalmus maximus*) Using Semantic Segmentation. *Fishes* **2022**, *7*, 385. <https://doi.org/10.3390/fishes7060385>

Academic Editor: Eric Hallerman

Received: 26 October 2022

Accepted: 9 December 2022

Published: 11 December 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 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

Fish products are rich sources of protein, vitamins, minerals, and unsaturated fats that benefit human health [1]. Additionally, the consumer demand for fish products has increased annually. In addition, fish products have a lower feed conversion ratio and higher edible ratio than chicken, pork, and beef products [2], which means that the aquaculture industry is more environmentally friendly than animal husbandry.

According to the Food and Agriculture Organization (FAO) of the United Nations [3], China has the largest production of flatfish in the world. In 2021, aquaculture production of flatfish in China was more than 122,000 tons, which was higher than that in 2020 [4]. Turbot (*Scophthalmus maximus*), a commercially valuable flatfish, is one of the most important aquaculture species in China due to its delicious flavor. In addition, the turbot is also an important aquaculture product around the globe.

Despite the gradual development of the turbot farming industry, the problem of fish diseases has largely limited its growth. Likewise, antibiotics used to solve the problem of fish diseases can cause quality problems in fish products [5]. To deal with these issues, there is a need for aquaculture companies to prevent and control fish diseases through immunization. Immunization mainly includes oral, immersion, and injection immunization [6,7]. Injection immunization has a better immune effect and a longer immune period, meaning that it is widely used in farming processes [8]. There are two types of injection

immunization: manual and mechanical [9]. However, there is no valid and efficient vaccination machine in China, where manual injection immunization is currently the main vaccination method [10]. It is laborious and time-consuming and incurs a high risk of self-injection, which is dangerous [11,12]. The injection position is the key point for designing a vaccination machine. A low immunization rate and even death could result if the injection position is incorrectly estimated [13].

Previous research on vaccination machines is mostly aimed at roundfishes, with less consideration for flatfishes. Many companies have developed related automatic vaccination machines for roundfishes, some of which can reach a maximum injection speed of 40,000 per hour [14,15] (Skala Inc., Norway, 2022; Lumeric Inc., Norway, 2022). Computer vision was used to predict the appropriate injection position of roundfishes. However, owing to the different traits between flatfishes and roundfishes, the injection position of flatfishes is more difficult to locate. Rossi Denmark Inc. [16] designed a semi-automatic vaccination machine for flatfishes, but it requires the vaccinator to adjust the injection position manually each time. In order to improve the degree of automation, Lee et al. [17] developed a vaccination machine for olive flounder (*Paralichthys olivaceus*), a flatfish, which used a vision system with a darkroom and backlight to reduce the effect of fins on injection position estimation. However, implementing a vision system within a darkroom is complicated and highly dependent on the environment. Further, relevant experiments were conducted to validate the feasibility of this method for turbot, but the results were not good.

Compared with classic machine-learning methods, deep-learning methods have better performance in image segmentation [18]. Therefore, we proposed using semantic segmentation with a deep learning network (DLN) to analyze the shape of the turbot and estimate the injection position, which would be more robust and valid than the darkroom method. Some studies have been completed to segment other kinds of fish images. Fernandes et al. [19] developed an algorithm based on SegNet to extract fish body measurements and predict body weight. Liu et al. [20] processed fish images using semantic segmentation to identify fish body posture. Additionally, Li et al. [21] proposed a measurement method of dynamic fish dimensions based on the mask region convolution neural network (Mask-RCNN), which had a low error in the measurement of body length and body width. Therefore, it is demonstrated that a DLN to achieve segmentation of the fish image would be a useful development.

This research aimed to devise a new method for estimating the injection position of turbot based on semantic segmentation and attention mechanism. This work is organized as follows: Section 2 mainly describes the design of the software algorithm to segment turbot images and estimate the injection position. The results and discussion for turbot segmentation, traits measurement, and injection position estimation are presented in Section 3, and the conclusion is presented in Section 4. The proposed method has two main contributions as follows:

- In order to accurately recognize the fish body, pectoral fin, and caudal fin of turbot, the classic Deeplabv3+ network was improved by using attention modules. Moreover, the proposed Atten-Deeplabv3+ was successfully executed to calculate the BL and BW;
- Using semantic segmentation, a method for estimating the injection position of the turbot was proposed. The experiments compared the errors of the injection position to prove the efficacy of the proposed approach, which would benefit the development of turbot vaccination machines.

2. Materials and Methods

Turbots weighing 50 to 100 g were purchased from a commercial fish farm (Tianyuan, Weihai, Shandong, China). The maintenance, handling, and experiments conducted on fish during this study were carried out in strict accordance with the guidelines of the Experimental Animal Welfare Ethics Committee of Zhejiang University (no. ZJU20190079).

2.1. Image Acquisition and Datasets

A self-built device collected image data after the fish were anesthetized. To obtain stable and clear images, the device, including a Mindvision camera (Mindvision Technology Co., Shenzhen, China), a mounting bracket, and a circular light source, was built. The camera was fixed at a height of 0.25 m using a mounting bracket, and a circular light source was installed between the camera and the fish (Figure 1). Meanwhile, each fish was photographed in a similar lighting environment, so that fish of different sizes displayed similar features. Finally, the table was covered with white PU tape to match the white background of the conveyor belt of the vaccination machine. The OpenCV library in Python was used to correct the distortion of the image, and a ruler with a scale value was placed in the lens to calculate the actual length corresponding to the pixel point.

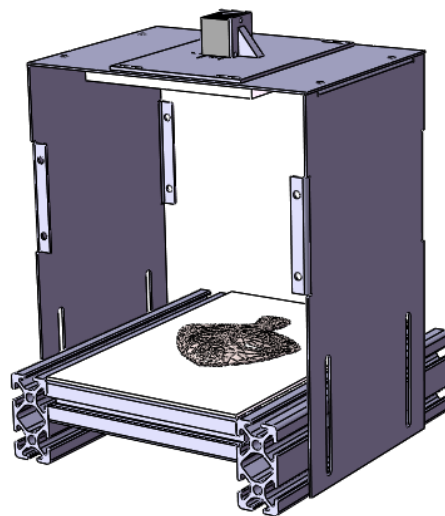


Figure 1. The device for image acquisition.

Images of 500 turbot were acquired by the above method. Each turbot was photographed once. All the turbot images were obtained using the Mindvision camera with a JPEG image format and a resolution of 1024×1024 pixels to develop an injection position estimation algorithm. The dataset was split into two parts randomly: dataset 1 contained 300 images used to train and evaluate the semantic segmentation networks. Meanwhile, it also was used to build the injection position estimation model and evaluate the performance of the BL and BW estimation algorithms. Thus, for each image, a PNG annotation file was created containing the label information of each pixel, which was marked as background, fish body, pectoral fin, and caudal fin. Image annotation was performed through the open-source Labelme annotation tool [22]. For evaluation of the semantic segmentation networks, dataset 1 was split into a training set and test set at a ratio of 6:4. Dataset 2 comprised 200 images which were used to evaluate the performance of the injection position estimation model. While taking turbot images, their morphological measurement information was also recorded manually. During the measurement process, water was removed from the turbot to reduce the impact on measurement results. To handle the uncertainty in manual measurement, each turbot was measured three times and received an average value as the final result. In addition, the theoretical injection position was annotated on the turbot images in dataset 2 by expert vaccinators.

The data augmentation technique was used to enlarge the image dataset, which helped to reduce over-fitting in supervised deep-learning algorithms [23]. To enhance the robustness of the turbot segmentation algorithm, this research considered many possible situations during semantic segmentation. Data augmentation, including brightness transformation, adaptive histogram equalization, random cropping, and rotation, was used. Dataset 1 was expanded to 1200 images by the above methods. Although many studies have proved that flipping can augment network performance, considering that the turbot

is asymmetrical, which would give different injection position results, flipping was not used in the data augmentation process.

2.2. Semantic Segmentation Model Architecture

Semantic segmentation, an essential method for feature extraction, is a key step in this work. Deeplabv3+ is one of the best semantic segmentation networks. Deeplabv3+ introduces a common encoder–decoder form of semantic segmentation, which uses Deeplabv3 as the encoder and the depthwise separable convolution network as the decoder [24,25]. Meanwhile, Deeplabv3+ combines atrous spatial pyramid pooling (ASPP) and the encoder–decoder form to obtain multi-scale information and fuse low-level and high-level features. It used the previous Deeplabv3 as the encoder layer and, referring to the structure of Unet [26], added a decoder layer. Therefore, an encoding and decoding structure was constructed to achieve end-to-end image semantic segmentation. The encoder layer used the aligned Xception network for feature extraction and a concatenated ASPP module to increase the receptive fields and obtain the multi-scale information without reducing the feature size. Finally, the decoder layer further fused the high-level and low-level features obtained by the encoding layer to generate the final predicted image. However, Deeplabv3+ still had some problems with poor boundary information prediction and holes in the predicted image.

As a lightweight and general module, a convolutional block attention module (CBAM) can be integrated into any CNN architecture seamlessly with negligible overheads and has been proven to play a certain role in feature extraction [27,28]. CBAM consists of channel and spatial attention modules to ensure that each branch can learn ‘what’ and ‘where’ to pay attention. The channel attention module focused on what is the meaningful information; it used two parallel max-pooling layers and an average-pooling layer, then passed through the shared multi-layer perceptron (MLP) module, and finally added the two output results and used the sigmoid activation function to obtain the weight of each channel use. Additionally, the channel attention was computed as follows:

$$M_c(F) = \sigma(\text{MLP}(\text{AvgPool}(F)) + \text{MLP}(\text{Maxpool}(F))), \quad (1)$$

where σ denotes the sigmoid function and F denotes the input feature.

The Spatial attention module focused on the informative parts, and it obtained two feature maps through the max-pooling layer and average-pooling layer, then concatenated the feature maps, and finally obtained the weight of each pixel of the entire image through 7×7 convolution and sigmoid activation function. In short, it was computed as:

$$M_s(F) = \sigma(f^{7 \times 7}([\text{AvgPool}(F); \text{Maxpool}(F)])), \quad (2)$$

where σ denotes the sigmoid function, F denotes the input feature, and $f^{7 \times 7}$ represents a convolution operation with a filter size of 7×7 .

We added the CBAM into the Deeplabv3+ network to improve network performance, which was named Atten-Deeplabv3+. Chen et al. [25] used the aligned Xception as the backbone of the Deeplabv3+ network and achieved good experimental results. However, the parameters of this network are too large, and the graphics processing unit (GPU) memory usage is high. Mobilenetv2 [29] was proposed by Google for mobile designs; its number of parameters and operations was much smaller than that of the aligned xception, but the performance of the mobilenetv2 was slightly inferior to the performance of aligned xception. Hence, taking into account the tradeoff between speed and performance, the mobilenetv2 was chosen as the backbone of the Deeplabv3+ network. The attention modules were placed after the two output feature layers of the mobilenetv2 networks to enhance the understanding of the features, as is shown in Figure 2. The image of the turbot was entered into the mobilenetv2 backbone, the high-level feature and low-level feature were generated by the mobilenetv2, and the weights of each channel and pixel

could be automatically adjusted in the training process through the attention module so as to improve the performance of the network.

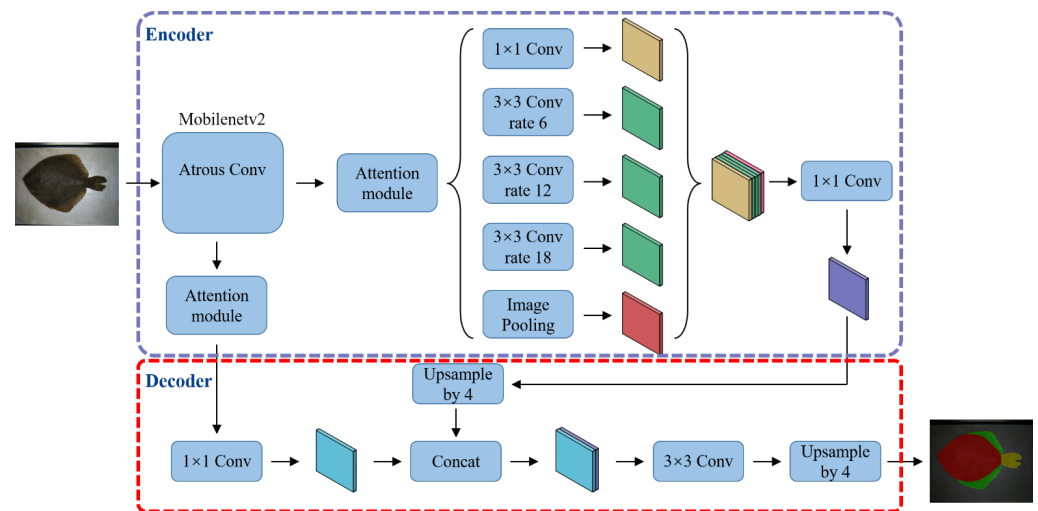


Figure 2. The network structure of Atten-Deeplabv3+.

2.3. BL and BW Estimation Algorithm

The Atten-Deeplabv3+ network was trained in Section 2.2, and segmented turbot images were acquired in Section 2.1, thereby obtaining pixel information on the image. The fish body, pectoral fin, and caudal fin were distinguished based on pixel information. The existence of segmentation errors (e.g., hollow parts, misidentified points, etc.) may lead to obtaining an incorrect estimation site. Therefore, a post-processing method was carried out on the image segmented by the Deeplabv3+ network. The opening and closing operations were used to ensure that the hollow part in the segmented image and possibly small, misidentified points were removed. The corrected images were used to calculate the BL and BW using the following algorithm:

1. Calculate the center of gravity of the fish body and caudal fin
2. Calculate the distance of gravity center of the caudal fin and the point of the fish body contour, and determine the point with maximum distance to the tip of the fish mouth as the coordinate origin
3. The line connecting the center of gravity of the fish body and the center of gravity of the caudal fin is used as the x -axis, with the positive direction of the x -axis from the center of gravity of the fish body to the center of gravity of the caudal fin. Rotate the x -axis 90 degrees around the origin to obtain the y -axis. The slope of the x -axis (k_x) and y -axis (k_y) can be calculated by Equation (3) below
4. Traverse the contours of the fish body above and below the x -axis, respectively, and determine the longest distance between the point and the x -axis. The body width of the turbot can be found by adding the two results. Then, traverse the contours of the caudal fin to find the nearest point to the coordinate origin, and the distance between the nearest point and the coordinate origin is the body length.

$$k_x = \frac{C_y - B_y}{C_x - B_x}, k_y = -\frac{1}{k_x}, \quad (3)$$

where (C_x, C_y) are the coordinates of the gravity center of the caudal fin; and (B_x, B_y) are the coordinates of the gravity center of fish body.

2.4. The Injection Position Estimation Model

The appropriate injection position of a turbot is near the abdominal cavity and has no distinguishing features, which is hard to recognize for machines. It is always recognized

by expert vaccinators. Therefore, it is necessary to establish an injection position model instead of identifying the injection position manually. However, the total length (including the caudal fin) and the total width (including the pectoral fins) of the turbot have a poor relationship with the injection position of the turbot because the caudal fin is likely to bend when the turbot is placed in the visual system, resulting in a large error between the measured and actual total length of the turbot. On the other hand, the opening and closing state of turbot fins will greatly affect the measurement of the total width. As a result, the total length and width of the turbot have a poor correlation with the injection position.

For the above reasons, the semantic segmentation network in Section 2.2 was proposed to remove the tail and fin of the fish body so as to reduce the recognition error of the body length and body width of turbot and improve the recognition accuracy of the theoretical injection position of turbot. Moreover, the turbot's injection position and morphological features were recorded, as shown in Figure 3. In Figure 3, x denotes the distance on the x -axis from the tip of the turbot mouth to the injection position, and y denotes the distance on the y -axis from the tip of the turbot mouth to the injection position.

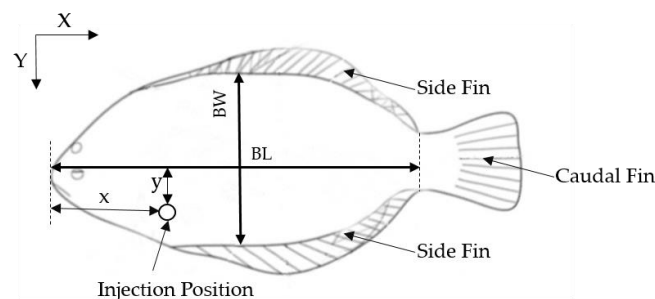


Figure 3. The measurement method for turbot body-shape parameters.

Given all this, an injection position model could be established based on dataset 2, which was fully annotated. The injection position was marked by a circle, and the appropriate injection position was within this circle. A linear regression model was used to analyze the correlation between the center of the circle and the morphological features of the turbot, followed by modeling a linear polynomial equation as shown in Equations (4) and (5):

$$x = aBL + b, \quad (4)$$

$$y = cBW + d, \quad (5)$$

where a , b , c , and d are the regression coefficients.

Since the camera coordinate system is quite different from the coordinate system we established, the translation and rotation of the coordinate system need to be modified through Equations (6) and (7) to ensure that the two coordinate systems overlap.

$$\theta = \begin{cases} \tan^{-1} k_x, & C_y > B_y, C_x > B_x \\ -\tan^{-1} k_x, & C_y < B_y, C_x > B_x \\ \pi - \tan^{-1} k_x, & C_y > B_y, C_x < B_x \\ \pi + \tan^{-1} k_x, & C_y < B_y, C_x < B_x \end{cases} \quad (6)$$

$$\begin{cases} X = x + A \cos \theta - B \sin \theta \\ Y = y + A \sin \theta + B \cos \theta \end{cases} \quad (7)$$

where (A, B) is the origin of the coordinates in the coordinate system that we established; (x, y) is the injection position in the coordinate system that we established; (X, Y) is the injection position in the camera coordinate system.

2.5. Experimental Setup

The platform for training the segmentation network included a desktop computer equipped with an Intel Core i9-9940X CPU, 16 GB of RAM, and four NVIDIA GTX 2080Ti 11 GB GPUs running on the Ubuntu 20.04 system. The software tools included CUDA11.1 (NVIDIA Corp., Santa Clara, CA, USA), cuDNN8.0.4 (NVIDIA Corp., Santa Clara, CA, USA), Pycharm 2021 (JetBrains Co., Ltd., Prague, Czech Republic), Pytorch1.2.0 (Meta Platform Inc., Menlo Park, CA, USA), Python3.6 [30], and OpenCV3.2.0 [31].

The platform for debugging the injection position estimation algorithm included a laptop computer equipped with an Intel Core i5-1135 CPU, 16 GB of RAM, and an NVIDIA MX 450 2 GB GPU, running on a Windows 10 64-bit system. The software tools were as noted above.

To compare the performance of different semantic segmentation networks on turbot image segmentation in this research, Unet, PSPnet [32], Deeplabv3+, and Atten-Deeplabv3+ were used to process turbot images. The backbone network of each network used mobilenetv2 to ensure lightweight. The initial weights of mobilenetv2 are derived from pre-trained weights on the PASCAL VOC2012 dataset [33], containing 21 classes with 27.4 k objects in 11.5 k images. The network's input size was 1024×1024 . In addition, the networks were trained for 100 epochs, with a mini-batch size of eight samples and data shuffled after every epoch. The parameters of the backbone mobilenetv2 were frozen in the first 50 epochs and unfrozen in the last 50 epochs. The optimization algorithm was a stochastic gradient descending with a momentum of 0.9, an initial learning rate of 5×10^{-3} , and a weight decay of 5×10^{-4} . Considering the large difference in the number of pixels in each part of a turbot, weighted cross-entropy loss was used to enhance the network performance.

2.6. Performance Evaluation

The performance of the turbot segmentation network trained in this study was assessed using IoU, mean intersection over union (mIoU), pixel accuracy (PA), and average segmentation speed, defined by the following formulas:

$$IoU = \frac{TP}{TP + FP + FN}, \quad (8)$$

$$PA = \frac{TP + TN}{TP + FP + FN + TN}, \quad (9)$$

$$mIoU = \sum_{i=0}^k \frac{TP}{TP + FP + FN}, \quad (10)$$

where k is the number of segmentation classes; TP is true positive; TN is true negative; FP is false positive; FN is false negative.

To evaluate the estimation accuracy of the injection position, the following formula was introduced to calculate the error of the injection position:

$$Error = \sqrt{(E_x - A_x)^2 + (E_y - A_y)^2}, \quad (11)$$

where (E_x, E_y) is the coordinates of the estimated injection position; (A_x, A_y) is the coordinates of the center of annotated injection position.

3. Results and Discussion

3.1. Semantic Segmentation

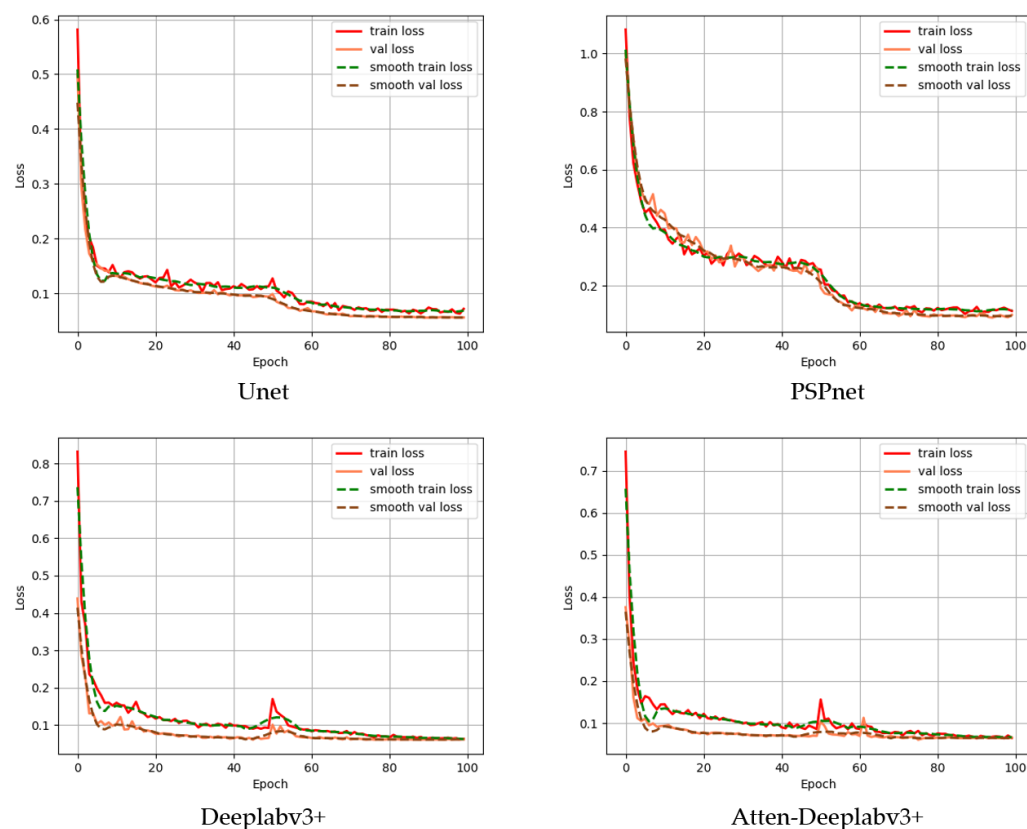
In Tables 1 and 2, the performance of the proposed method was compared with these of other semantic segmentation networks. Each network was trained three times on the training sets of dataset 1 and evaluated on the test set of dataset 1. The loss curves for each network during training are shown in Figure 4.

Table 1. The IoU results for semantic segmentation networks.

Networks	IoU			
	Background	Fish Body	Pectoral Fin	Caudal Fin
Unet	99.2%	94.4%	81.1%	90.3%
PSPnet	99.2%	94.2%	82.0%	89.9%
Deeplabv3+	99.3%	95.5%	82.9%	90.5%
Atten-Deeplabv3+	99.3%	96.5%	85.8%	91.7%

Table 2. Comparison of results among semantic segmentation networks.

Networks	PA				Training Time
	Background	Fish Body	Pectoral Fin	Caudal Fin	
Unet	99.6%	97.3%	89.7%	94.4%	1 h 50 min
PSPnet	99.6%	97.1%	90.1%	94.3%	2 h
Deeplabv3+	99.6%	97.7%	90.7%	95.3%	2 h 16 min
Atten-Deeplabv3+	99.7%	98.2%	92.8%	96.2%	2 h 27 min

**Figure 4.** The loss curves for semantic segmentation networks.

In general, these networks had good recognition for the background and fish body but poor recognition for the pectoral and caudal fins. Compared with the other three networks, the method we proposed has better recognition results in each class. Furthermore, the mIoU results of the above networks were 91.3%, 91.3%, 92.0%, and 93.3%, respectively. The size of the parameters of Atten-Deeplabv3+ is 22.37 MB, which increased only by 0.19 MB compared with that of Deeplabv3+. The detection speed of A for 1024×1024 -resolution images is 22 FPS, while that of Deeplabv3+ is 24 FPS. This is sufficient for the use of vaccination machines, although the detection speed decreased by 2 FPS. In addition, the proposed network had the longest training time, which was acceptable.

Atten-Deeplabv3+ greatly improved the recognition of the pectoral fin, and the recognition of the fish body and caudal fin also improved slightly. This shows that embedding two attention modules behind the backbone network can help improve network performance, especially in the prediction of pectoral fins. The edge of the pectoral fin predicted by Deeplabv3+ is relatively tortuous. After embedding the attention module, the overall edge becomes smoother, and the recognition results are closer to the ground truth than other segmentation networks, as shown in Figure 5.

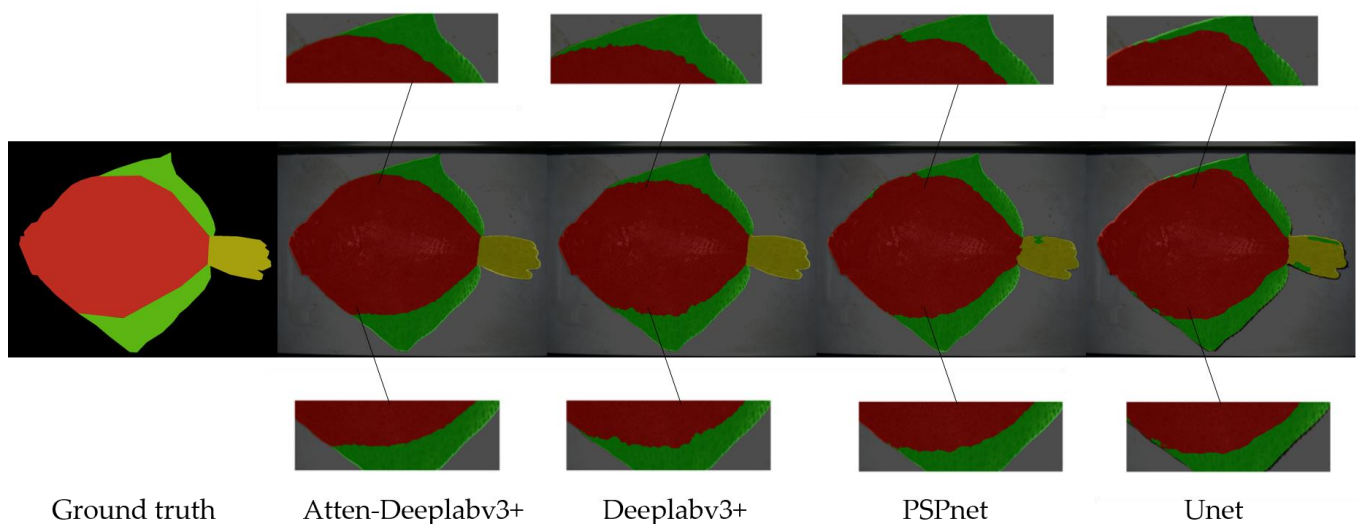


Figure 5. Segmentation results of smooth edges.

At the same time, the hollow and misrecognition phenomena of the caudal fin were also improved, as shown in Figure 6. This was beneficial for the location of the center of gravity in the algorithm of Section 2.3. However, the IOU of the pectoral and caudal fin was still not very high overall. The possible reason is that there is not a clear dividing line between a fish body and a fin, which made it hard to recognize which part was the fin and which part was the fish body for annotators. Hence, some annotation errors occurred, which would negatively affect the segmentation results. Similarly, the small dataset may also cause this problem.

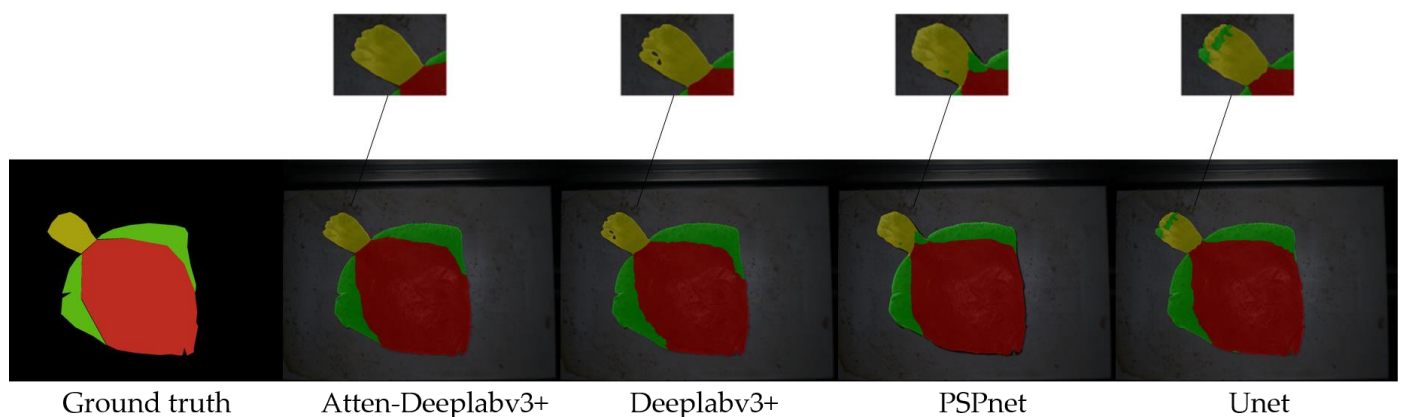


Figure 6. Segmentation results for hollow improvement.

There has been a number of experiments completed to remove fins using deep learning. Fernandes et al. [19] developed an algorithm based on SegNet to extract the fish body and fin areas of Nile tilapia (*Oreochromis niloticus*). The best IoU results of fish body and fin areas are 90% and 64%. Li et al. [21] proposed a method of dynamic fish dimension determination

based on the Mask-RCNN to segment the fish body of grass carp (*Ctenopharyngodon idellus*) with an IoU of 81%. Therefore, the network proposed in this work could achieve relatively good segmentation results, being able to segment each part of the turbot well.

To verify the robustness of our method, we used some flatfish images on the network for recognition, and the results are shown in Figure 7. On the one hand, if the image background had less interference, the boundary between the fin and the fish body was obvious, which was closer to the state of the turbot in the training set, and our method had better recognition results. On the other hand, the recognition effect is worse when there is more interference in the background. Furthermore, our method had good recognition results on olive flounder with a pure background. In general, this method only had good recognition results for turbot photographed by the image acquisition device in Section 2.1, while it had poor recognition results for images of other flatfishes and with other backgrounds.

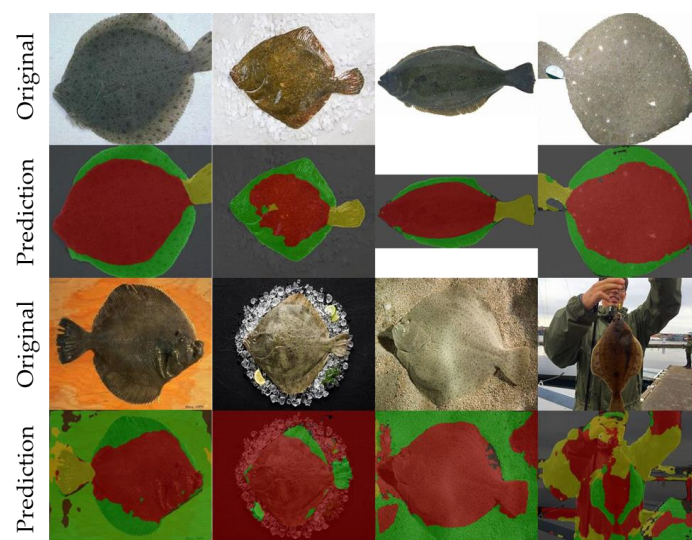


Figure 7. The segmentation results of internet images.

3.2. The Performance of BL and BW Estimation Algorithm

Figure 8 shows the absolute error of BL and BW in dataset 1. The maximum value of the absolute error of BL was 3.6 mm, the mean absolute error value was 1.07 mm, the maximum value of the absolute error of BW was 4.6 mm, and the mean absolute error value was 1.7 mm. The maximum relative error of BL was 2.7%, and the mean relative error was 0.7%. The maximum relative error of BW was 5.3%, and the mean relative error was 1.6%. In general, the accuracy of our method for measuring BL was higher than for BW. This is mostly because, as was mentioned in Section 2.2, it was difficult to differentiate the boundary of the pectoral fin and the fish body. Thus, the IoU of the pectoral fin was low, which caused BW measurement accuracy to be much worse. Meanwhile, different turbot placement angles and bad results of semantic segmentation may be the reason for errors. Owing to the line connecting the center of gravity of the fish body and the center of the caudal fin deviating from the center axis, the estimation of body length and width slightly differed from the actual value. In contrast to other research, Huang et al. [34] proposed an algorithm based on midline estimation to measure the fish length, with a mean absolute error of 1.5%, respectively. Lee et al. [35] measured the length and width of grass carp in a bent state using an object detection network, and its mean relative error was 2.1%. Consequently, our method could accurately estimate the BL and BW of the turbot, which contributed to the injection position estimation of the turbot.

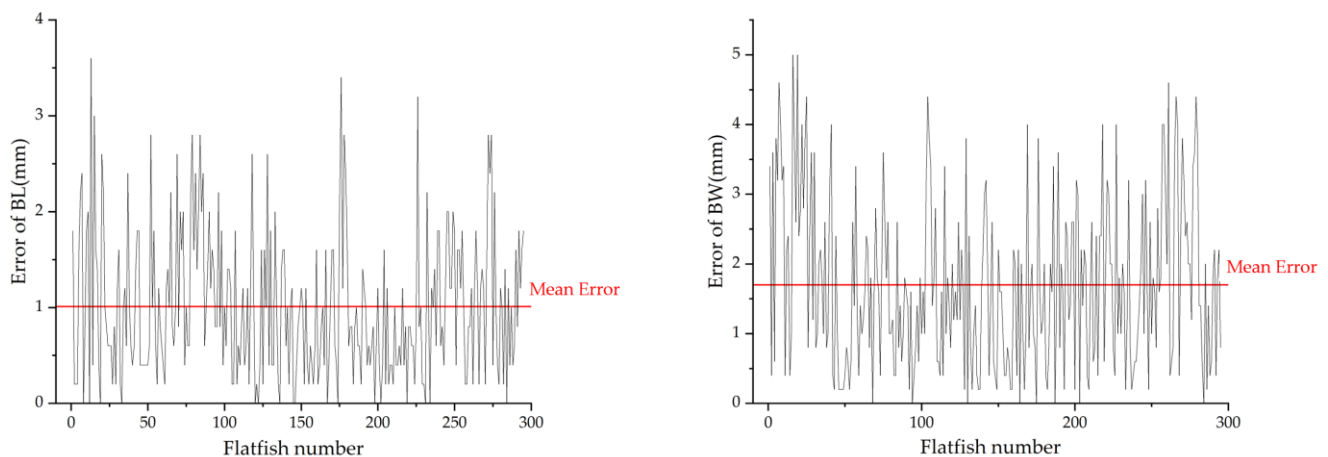


Figure 8. Absolute error of BL and BW.

3.3. The Performance of the Injection Position Estimation Model

Dataset 2 was used to evaluate the performance of the turbot injection position evaluation algorithm, where the maximum allowable error for each turbot injection position was the radius of the injectable area, which was annotated in Section 2.1, and the estimation error of the injection position was calculated using Equation (11).

As shown in Figure 9, the proportions of turbot for which injection position was correctly estimated the injection position were 94, 90, 89, and 84 percent for Atten-Deeplabv3+, Deeplabv3+, PSPnet, and Unet, respectively. The error of injection position for Atten-Deeplabv3+ ranged from 0.3 mm to 4.2 mm, and the maximum allowable error was $3.4 \text{ mm} \pm 0.3 \text{ mm}$. Some researchers have performed studies on the estimation of injection sites in olive flounder. Lee et al. [36] designed an automatic device for olive flounder based on template matching, which had a position recognition error of 0 to 0.6 mm. However, this method can only identify fish bodies with similar sizes to the template, and the recognition effect is poor for fish bodies with larger sizes than the template. In addition, Lee et al. [17] also designed an olive flounder injection device based on morphological recognition in a dark environment, with an error of 0 to 1.9 mm. However, PV500, a visual recognition system that they used, was much more expensive than the camera we used. In general, the vast majority of turbot in our research were correctly identified to the injection position, and only a few turbot injection positions exceeded the maximum allowable range. Overall, the injection position for most of the turbot in our study was correctly estimated, and only a few were injected beyond the maximum allowable range. Therefore, though estimation errors occurred in body length and body width, the error of the injection position was still acceptable, which means that this algorithm can be used as a turbot injection position estimation algorithm and would achieve successful estimation results. Although there were some errors in the evaluation process of injection position, the immune success rate and survival rate after the injection of turbot near the maximum allowable range need to be further verified in a vaccination injection experiment.

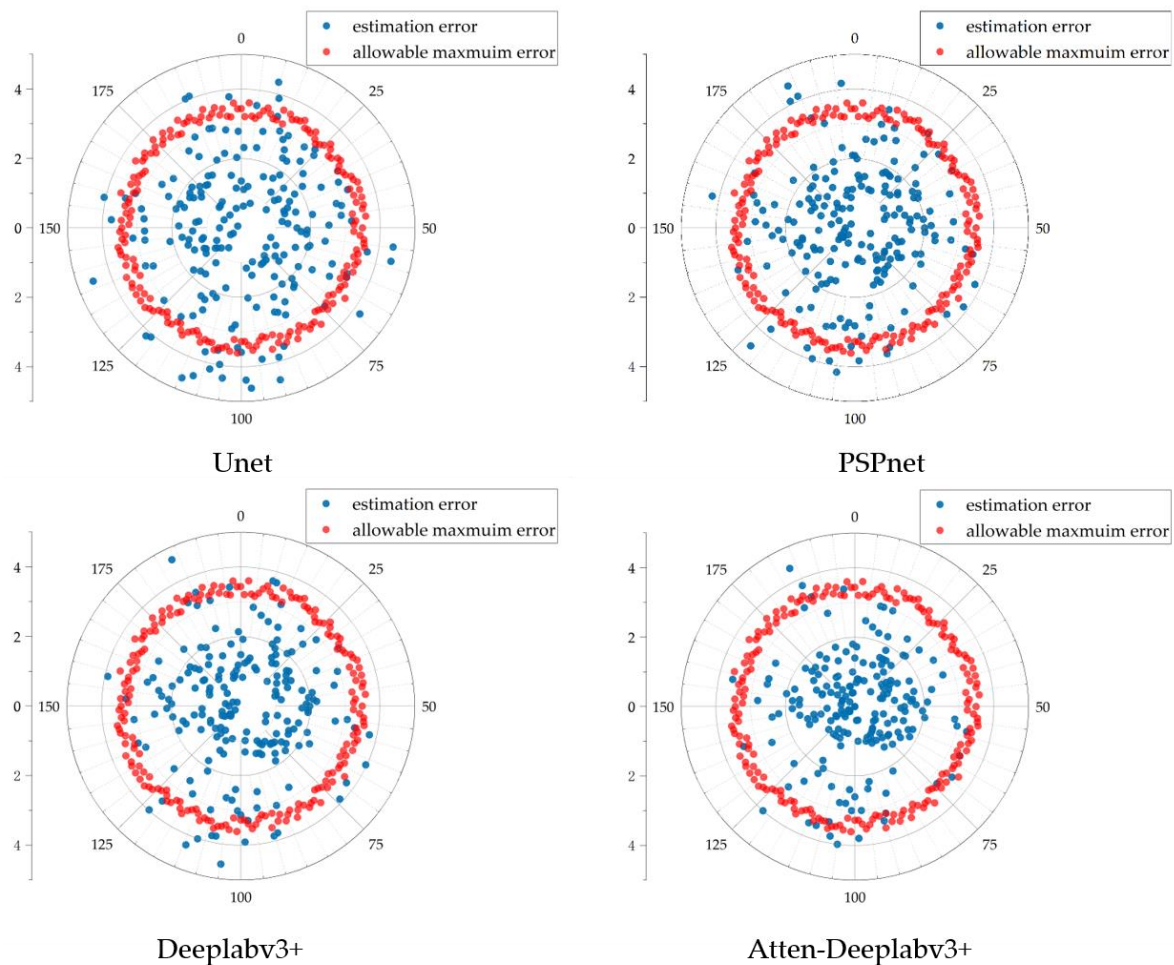


Figure 9. The error of injection position for each network.

4. Conclusions

To develop an injection position and estimation algorithm for the turbot vaccination machine, a method based on injection position models and semantic segmentation was proposed. The Atten-Deeplabv3+ was used to recognize the fish parts, and the IoU values of the test set were 99.27, 96.47, 85.81, and 91.70 percent for the background, fish body, pectoral fin, and caudal fin, respectively. The error of injection position ranged from 0.3 mm to 4.2 mm, which is almost within the injectable range. In conclusion, the devised method was able to correctly differentiate the fish body from the background and fins in a pure background which was similar to the device we designed. Furthermore, the extracted area of fish could be successfully used to estimate the injection position, which would benefit the development of turbot vaccination machines. In the future, this method will be applied to vaccination machines, and the immune success rate and survival rate of turbot will be tested.

Author Contributions: Methodology, W.L. and J.L.; software, W.L. and C.L.; validation, W.L.; formal analysis, W.L. and C.L.; investigation, W.L.; data curation, W.L.; writing—original draft preparation, W.L.; writing—review and editing, K.W. and J.L.; visualization, W.L. and C.L.; supervision, J.L., Z.Y. and S.Z.; project administration, J.L.; funding acquisition, J.L. and Z.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key Research and Development Program of China (Grant No. 2019YFD0900103-05) and the National Modern Agriculture Industrial Technology System Special Project—the National Technology System for Conventional Freshwater Fish Industries (CARS-45-24).

Institutional Review Board Statement: The animal study protocol was approved by the guidelines of the Experimental Animal Welfare Ethics Committee of Zhejiang University (no. ZJU20190079).

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on demand from the first author at (weiluo0109@zju.edu.cn).

Acknowledgments: The authors are very grateful to the members of the research group for their help in the experiment. We also appreciate the work of the editors and the reviewers of the paper.

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

References

1. Sarvenaz, K.T.; Sabine, S. Nutritional Value of Fish: Lipids, Proteins, Vitamins, and Minerals. *Rev. Fish. Sci. Aquac.* **2018**, *26*, 243–253. [CrossRef]
2. Ninawe, A.S.; Dhanze, J.R.; Dhanze, R.; Indulkar, S.T. *Fish Nutrition and Its Relevance to Human Health*; CRC Press: London, UK, 2020; pp. 10–26.
3. FAO. *Fishery and Aquaculture Statistics*; Global Aquaculture Production 1950–2020 (FishstatJ); FAO Fisheries and Aquaculture Department: Rome, Italy, 2022; Available online: <http://www.fao.org/fishery/statistics/software/fish-stati/enStatistics> (accessed on 23 October 2022).
4. Statistics Bureau of the People's Republic of China. *China Fishery Statistical Yearbook*; China Statistics Press: Beijing, China, 2021.
5. Kalantzi, I.; Rico, A.; Mylona, K.; Pergantis, S.A.; Tsapakis, M. Fish farming, metals and antibiotics in the eastern Mediterranean Sea: Is there a threat to sediment wildlife? *Sci. Total Environ.* **2021**, *764*, 142843. [CrossRef] [PubMed]
6. Ma, J.; Bruce, T.J.; Jones, E.M.; Cain, K.D. A Review of Fish Vaccine Development Strategies: Conventional Methods and Modern Biotechnological Approaches. *Microorganisms* **2019**, *7*, 569. [CrossRef] [PubMed]
7. Plant, K.P.; LaPatra, S.E. Advances in fish vaccine delivery. *Dev. Comp. Immunol.* **2011**, *35*, 1256–1262. [CrossRef] [PubMed]
8. Santos, Y.; Garcia-Marquez, S.; Pereira, P.G.; Pazos, F.; Rianza, A.; Silva, R.; El Morabit, A.; Ubeira, F. Efficacy of furunculosis vaccines in turbot, *Scophthalmus maximus* (L.): Evaluation of immersion, oral and injection delivery. *J. Fish Dis.* **2005**, *28*, 165–172. [CrossRef] [PubMed]
9. Ellis, A. Health and disease in Atlantic salmon farming. In *Handbook of Salmon Farming*; Stead, S.M., Laird, L.M., Eds.; Springer Praxis: London, UK, 2002; Volume 10, pp. 373–401.
10. Zhu, Y.; Xu, H.; Jiang, T.; Hong, Y.; Zhang, X.; Xu, H.; Xing, J. Research on the automatic injection technology of Ctenopharynuodon idellus vaccine. *Fish. Mod.* **2020**, *47*, 12–19. [CrossRef]
11. Brudeseth, B.E.; Wiulsrød, R.; Fredriksen, B.N.; Lindmo, K.; Løklung, K.-E.; Bordevik, M.; Steine, N.; Klevan, A.; Gravningen, K. Status and future perspectives of vaccines for industrialised fin-fish farming. *Fish Shellfish Immunol.* **2013**, *35*, 1759–1768. [CrossRef] [PubMed]
12. Leira, H.L.; Baalsrud, K.J. Operator safety during injection vaccination of fish. *Dev. Biol. Stand.* **1997**, *90*, 383–387. [PubMed]
13. Understanding-Fish-Vaccination. Available online: <https://thefishsite.com/articles/understanding-fish-vaccination> (accessed on 23 October 2022).
14. Vaccination-Aquaculture-Nettsteder-Skala Maskon. Available online: <https://en.skalamaskon.no/aquaculture2/vaccination> (accessed on 23 October 2022).
15. Products-Lumic AS & Strømmeservice AS. Available online: <https://lumic.no/en/products/> (accessed on 23 October 2022).
16. The Fish Site. Evaluation of Fish Vaccination Machines in Norway. Available online: <https://thefishsite.com/articles/evaluation-of-fish-vaccination-machines-in-norway> (accessed on 23 October 2022).
17. Lee, D.-G.; Yang, Y.-S.; Kang, J.-G. Conveyor belt automatic vaccine injection system (AVIS) for flatfish based on computer vision analysis of fish shape. *Aquac. Eng.* **2013**, *57*, 54–62. [CrossRef]
18. Khankeshizadeh, E.; Mohammadzadeh, A.; Moghimi, A.; Mohsenifar, A. FCD-R2U-net: Forest change detection in bi-temporal satellite images using the recurrent residual-based U-net. *Earth Sci. Informatics* **2022**, *15*, 2335–2347. [CrossRef]
19. Fernandes, A.F.; Turra, E.; 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. *Comput. Electron. Agric.* **2020**, *170*, 105274. [CrossRef]
20. Liu, B.; Wang, K.; Li, X.; Hu, C. Motion posture parsing of *Chiloscyllium plagiosum* fish body based on semantic part segmentation. *Trans. CSAE* **2021**, *37*, 179–187. [CrossRef]
21. Li, Y.; Huang, K.; Xiang, J. Measurement of dynamic fish dimension based on stereoscopic vision. *Trans. CSAE* **2020**, *36*, 220–226. [CrossRef]
22. Russell, B.C.; Torralba, A.; Murphy, K.P.; Freeman, W.T. LabelMe: A Database and Web-Based Tool for Image Annotation. *Int. J. Comput. Vis.* **2008**, *77*, 157–173. [CrossRef]
23. Shorten, C.; Khoshgoftaar, T.M. A survey on Image Data Augmentation for Deep Learning. *J. Big Data* **2019**, *6*, 60. [CrossRef]
24. Chen, L.C.; Papandreou, G.; Schroff, S.; Adam, H. Rethinking atrous convolution for semantic image segmentation. In Proceedings of the Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017.

25. Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018.
26. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015.
27. 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.
28. Ba, J.; Mnih, V.; Kavukcuoglu, K. Multiple object recognition with visual attention. In Proceedings of the Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018.
29. Sandler, M.; Howard, A.; Zhu, M.L.; Zhmoginov, A.; Chen, L.C. MobileNetV2: Inverted residuals and linear bottlenecks. In Proceedings of the Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015.
30. Welcome to Python.org. Available online: <https://www.python.org/> (accessed on 23 October 2022).
31. Home—OpenCV. Available online: <http://www.opencv.org/> (accessed on 23 October 2022).
32. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid scene parsing network. In Proceedings of the Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017.
33. The Pascal Visual Object Classes Challenge 2012 (VOC2012) Results. Available online: <http://www.pascal-network.org/challenges/VOC/voc2012/workshop/index.html> (accessed on 23 October 2022).
34. Huang, T.-W.; Hwang, J.-N.; Rose, C.S. Chute based automated fish length measurement and water drop detection. In Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP), Shanghai, China, 20–25 March 2016.
35. Lee, C.; Li, J.; Zhu, S. Automated size measurement and weight estimation of body-curved grass carp based on computer vision. In Proceedings of the American Society of Agricultural and Biological Engineers Annual International Meeting, Virtual, 12–16 July 2021.
36. Lee, D.-G.; Cha, B.-J.; Park, S.-W.; Kwon, M.-G.; Xu, G.-C.; Kim, H.-J. Development of a vision-based automatic vaccine injection system for flatfish. *Aquac. Eng.* **2013**, *54*, 78–84. [[CrossRef](#)]