A detection method of the turned white belly fish based on improved SSD

TIAN Ganglu\textsuperscript{1,2,3}, LI Daoliang\textsuperscript{4,2,3}, LI Wensheng\textsuperscript{4}, ZHANG Lu\textsuperscript{1,2,3}, ZHANG Hongxu\textsuperscript{1,2,3} and Qingling Duan\textsuperscript{1,2,3,*}

\textsuperscript{1}Information and Electrical Engineering, China Agricultural University, Beijing 100083, China
\textsuperscript{2}National Innovation Center for Digital Fishery, China Agricultural University, Beijing 100083, China
\textsuperscript{3}Agricultural Internet of Things Engineering Technology Research Center, China Agricultural University, Beijing 100083, China
\textsuperscript{4}Laizhou Mingbo Aquatic Products Co., Ltd., Laizhou, 261400, China
*Corresponding author’s e-mail: dqling@cau.edu.cn

Abstract. Aquaculture water pollution and fish disease caused the phenomenon of fish turning white belly. The accurate detection of the phenomenon of fish turning white belly was of great significance for emergency warning, reducing loss and improving the quality of aquatic products. There were few studies on the detection of the turned white belly fish in the fish swarm. The current research was mainly to identify single the turned white belly fish under a simple background, which was not applicable in the real aquaculture environment. In order to solve the above problems, a detection method of the white belly fish based on improved SSD was proposed in a complex environment. Crucian carp was taken as the experimental object and the proposed detection method of the turned white belly fish had been tested on the real dataset. The results showed that the average detection accuracy of white-bellied fish was 99.8\%, and the average detection accuracy of normal fish was 98.8\%, which was suitable for the complex environment with high breeding density and many interferences.

1. Introduction
Fish is an important part of food supply, providing humans with a large amount of high-quality protein, unsaturated fatty acids and various trace elements[1-2]. Fish often turn white belly due to their own diseases, lack of oxygen in the aquaculture waters, and water temperature discomfort. Once the phenomenon of fish turning white belly is not found in time and measures are taken, it is easy to affect other fish and cause large-scale death of fish. In fish farming at this stage, artificial pond patrols are often used to detect the fishes, which requires a lot of labor and cannot guarantee the timeliness of monitoring. In recent years, machine vision technology has been welcomed by domestic and foreign scholars in the field of aquatic product detection with its advantages of fast, objective, and non-contact[3]. Researchers at home and abroad conducted some studies on the fish target detection model based on machine vision, which were divided into traditional target detection methods based on manual features and target detection methods based on deep learning[4-6].
Shao Tengfei[7] took the fish target as the input of the neural network, and used the error back propagation algorithm to train the network to determine whether the fish target was in the active state. However, the average accuracy of multiple fish activity detection was low, only reaching more than 80%. Liu Xingqiao[8] binarized the original image, separated the white abdomen of the fish body from the background, and then counted the number and size of the white area to determine the white belly. The research mainly used image segmentation to recognize the turned white belly fish under simple background. Traditional target detection methods based on manual features usually had low detection efficiency and poor accuracy. Yuan Hongchun[9] proposed a multi-scale retinal enhancement algorithm (msrcr) based on Faster R-CNN secondary transfer learning and color recovery. The clever use of two migration learning to achieve accurate detection of small-scale underwater fish data sets. In the research of binocular fish image detection, Shen Junyu[10] proposed a splicing method based on an improved SSD network model, and the accuracy of small target fish detection reached more than 90%. High and low feature fusion improves the detection accuracy of small target fish. The target detection algorithm of deep learning is mostly used for the detection of fish target, but there are few researches on the detection of fish turning white belly.

At present, the research of fish turning white belly detection mainly uses the traditional target detection method to recognize a single turned white belly fish in a simple background. In response to the above problems, the detection method of fish turning white belly in complex environment was studied in the paper. The main contributions of this paper: firstly, the deep learning and automatic color scale algorithm are combined to detect the white belly of fish, which solves the problem of low detection performance caused by the complex underwater environment; secondly, K-means and SSD target detection framework were fused to find the more suitable size and aspect ratio anchor box for fish, thereby improving the accuracy of the turned white belly fish detection.

2. Materials and methods

2.1. Fish image collection

In the paper, the fish video were collected based on the fish turning white belly detection experimental data collection platform at the National Digital Fisheries Innovation Center. The experimental platform is shown in Figure 1.

![Fish image collection platform diagram](image)

1. Camera (Hikvision 3T86FWDV2-I3S, 8 million pixels, 4mm focal length)  
2. Fish tank (1.5m*5m)  
3. Aeration type aerator  
4. Desktop computer  
5. Router

Figure 1. Diagram of the fish-image collection platform.

The fish video data used in the paper were collected by the above-mentioned experimental platform. One frame was intercepted every 4 seconds. The resolution of the image set was 3840x2160, the format was JPG. The expanded image data was a total of 9053. The number of normal fish targets was 11943, and the number of white bellied fish targets was 2705.

2.2. Fish image preprocessing

(1) Fish image scaling and region of interest extraction.

The data set of the paper was the fish image obtained from the fish video frame. The file format was large, which was not conducive to subsequent image batch processing. For the convenience of
operation, the image resolution was processed to 300x300 through the image scaling function and the region of interest extraction function of OpenCV.

(2) Fish image enhancement

The automatic color scale algorithm pair was used to defog the fish image and enhance the image contrast. First, the original fish image was used as the input of the automatic color scale algorithm, and the histogram of the fish image was counted. According to the histogram algorithm, the upper and lower thresholds of the fish image were automatically calculated. If the pixel value of the fish image was less than the lower threshold, the pixel value was assigned a value of 0. If the pixel value was greater than the upper threshold, the pixel value was assigned a value of 255. If the image value was between the upper and lower thresholds, the threshold difference was calculated, and the pixel value--the lower threshold was multiplied by 255 to divide the threshold difference[11-12]. The threshold difference is the upper threshold--the lower threshold, and the calculation formula is as shown in equation (1):

$$D = \text{Maxlevel} - \text{Minlevel}$$

Finally, the images processed by these three parts were integrated to obtain the fish image after defogging.

(3) Dataset expansion

The paper used rotation, translation, mirror image, brightness adjustment and noise to enhance the data set, so as to improve the rotation invariance and scale invariance of the algorithm and reduce the risk of over fitting. The data set was expanded to 10 times of the original, and Labellmg was used to mark the position coordinates of fish.

2.3. The detection of fish turning white belly based on SSD

The SSD algorithm was used to detect the turned white belly fish. The fish image of 300x300x3 was used as the input layer of SSD model, the SSD model used VGG-16 as the backbone network. The fish image was feature extracted through the improved VGG-16 and multiple convolutional layers, and 6 effective fish body feature layers were extracted. The six effective feature layers obtained were divided into six different size grids (38x38, 19x19, 10x10, 5x5, 3x3, 1x1) for multi-size fish target detection. SSD added a default box mechanism to construct a priori box of different scales for each point on the acquired effective feature layer. SSD made predictions on different feature maps, that was, the priori box was adjusted to predict the position offset and confidence of the fish target. Finally, the Non-Maximum Suppression algorithm[13] (NMS) was used to screen the fish target prediction results with the highest confidence. The structure diagram is shown in Figure2.
2.4. Priori box selection based on K-means clustering
Generally, the priori box was set by the original author of the model according to the target size in the VOC data set and personal experience, and the categories of these data sets were not suitable for the image data set of fish in the paper. In order to find a more suitable priori box for the target size of fish, K-means was used to cluster the labeled data set in the paper [14-15]. According to the size obtained by clustering, the priori box size of regression prediction network was modified so as to improve the accuracy of target detection of fish.

First extract the coordinates of the fish bounding box manually labeled in the data set; then convert the coordinate data into the width and height of the box; randomly select k bounding boxes as the initial value of the anchor boxes; calculate the IOU between each bounding box and each anchor box, and the function is as shown in equation (2):

$$
\text{d}(\text{box, centroid}) = 1 - \text{IOU}(\text{box, centroid})
$$

(2)

The bounding box is assigned to the anchor box with the largest IOU value. Calculate the median width and height of the bounding box of each anchor box class, and use the median as the latest anchor box size. Repeat four to six steps, so that the optimal size of the fish target is obtained through k-means clustering.

3. Experiments and results

3.1. Experimental design
Crucian carp was taken as the experimental object. There were 9053 experimental data sets, including 5,069 training sets, 2716 test sets, and 1268 verification sets. The Keras deep learning framework was used in the experiment, training was carried out on 64-bit Windows7 system and i7-8700CPU. The stochastic gradient descent algorithm was used to optimize the training stage. The batch size was set to 4, and the initial learning rate was set to 0.00001.

3.2. Evaluation metrics
The results of our experiments were evaluated by Precision, Recall, F1-Score, AP and mAP. They are calculated in Equations (3) ~ (7).
\[
\begin{align*}
\text{Precision} &= \frac{TP}{TP + FP} \\
\text{Recall} &= \frac{TP}{TP + FN} \\
F1\text{- score} &= \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \\
AP &= \int_0^1 p_{\text{smooth}}(r) dr \\
\text{mAP} &= \frac{\sum AP_i}{c}
\end{align*}
\]

Where \( TP \) is the number of samples that are positive in reality and positive in prediction, \( FP \) is the number of samples that are actually negative but predicted to be positive, \( TN \) is the number of samples that are negative in reality and negative in prediction, \( FN \) is the number of samples that are actually positive but predicted to be negative, \( c \) is the number of categories detected, \( p_{\text{smooth}}(r)dr \) is the curve obtained by smoothing \( p(r) \) curve.

### 3.3. Results and discussion

#### 3.3.1. Fish image preprocessing results

The (a) to (d) in Figure 3 are the original fish image and the preprocessed fish images. Specifically, the preprocessed fish images include the segmented fish image, the enhanced fish image, and the annotated fish image.

(a)Original fish image (b) Segmented fish image (c)Enhanced fish image (d)Annotated fish image

Figure 3. Results of fish image preprocessing.

From Figure 3, it can be seen that the original fish image contained a large number of blank areas and there were problems such as plane reflections around the fish tank, so the image (b) was obtained by selecting the region of interest. From the comparison of image (b) and image (c), it can be seen that the automatic color scale algorithm used in this article can effectively defog. The contrast and clarity between the fish target and the background were effectively enhanced.

#### 3.3.2. Results of the priori box selection

In the paper, \( k = (1-12) \) was used to cluster the data set[16]. The clustering results are shown in Figure 4.
According to the size of fish, the priori box size and aspect ratio of the feature layer of 19*19 and 10*10 were mainly set. From the feature layer structure of the SSD, 6 prior boxes were drawn for each grid of 19*19 and 10*10 characteristic layers, so three different aspect ratios were required. The clustering results can be seen from Figure 8: when K<3, Avg_IOU grew faster, when K>3, Avg_IOU grew slowly, so K=3 was selected for clustering. The best aspect ratio was (82.07,54.08),(91.07,81.12), (53.96,74.88), the aspect ratio was 2/3, 1, 3/4.

3.3.3. The detection results of the turned white belly fish

In the paper, five weight models were selected for performance comparison using evaluation indicators. The comparison was as follows.

Table 1. Comparison of different models.

| Model number | White belly fish | Normal fish | mAP |
|--------------|------------------|-------------|-----|
|              | Precision Recall F1-Score | AP | Precision Recall F1-Score | AP |  |
| 0            | 92.5% 100.0% 96.1% 99.7% | 80.9% 98.9% | 89.0% 98.6% | 99.2% |
| 1            | 92.2% 100.0% 95.9% 99.7% | 80.8% 98.7% | 88.6% 98.5% | 99.1% |
| 2            | 92.0% 99.9% 95.8% 99.7% | 85.5% 99.0% | 91.8% 98.9% | 99.3% |
| 3            | 93.7% 100.0% 96.7% 99.8% | 86.1% 99.0% | 92.1% 98.8% | 99.3% |
| 4            | 95.2% 100.0% 97.5% 99.8% | 86.6% 99.0% | 92.4% 98.8% | 99.3% |

From table 1, the average accuracy rate mAP (0.5) of the five models obtained by training for fish swarm detection reached more than 99.1%, the average accuracy of the detection of the turned white belly fish was more than 99.7%, and the average detection accuracy for normal fish was more than 98.5%. The training model in the paper can effectively detect both normal fish and the turned white bellied fish in fish swarm. Overall, model 4 was the best, so model 4 was selected as the fish detection model.

3.3.4. Results and discussion of comparison methods

Figure 5 and Figure 6 were from real data collected by the experimental platform. The detection of the turned white belly fish was marked with a red box, and the detection of normal fish was marked with a green box. It can be seen from the figure that Faster R-CNN had a false detection for background interference in aerators. The SSD algorithm's detection effect on the image after defogging was significantly better than that of undefogged. The method based on improved SSD had made corresponding improvements on the original SSD method. Compared with SSD method, the detection accuracy of fish was improved, and the image clarity was more inclusive.
We compared the improved SSD method with several turned white belly fish detection methods. The comparison results are shown in Table 2.

| Model          | White belly fish | Normal fish | Time/epoch |
|----------------|------------------|-------------|------------|
|                | Precision | Recall | F1-Score | Precision | Recall | F1-Score | (min) |
| Faster R-CNN   | 48.8%     | 98.0%  | 65.2%    | 66.0%     | 95.2%  | 78.0%    | 140.4 |
| SSD            | 91.2%     | 100.0% | 95.4%    | 82.9%     | 99.3%  | 90.4%    | 113.9 |
| improved SSD   | 95.2%     | 100.0% | 97.5%    | 86.6%     | 99.0%  | 92.4%    | 91.1  |

From Table 2, it can be seen that the F1 value of Faster R-CNN for the detection of the turned white belly fish was 65.2%, and the detection result was the worst. The detection accuracy of SSD for the turned white belly fish was 91.2%. Compared with SSD, the detection accuracy rate of the method based on improved SSD was increased by 4.0%, which showed that the priori box selection algorithm using K-means clustering to modify SSD effectively reduced the false detection rate of the turned white belly fish. From the perspective of model training time, the method based on improved SSD was also better than the comparison methods.

4. Conclusion
The innovation of this research was that the deep learning target detection method was applied to the detection of the turned white belly fish. The automatic color scale defogging method was used to enhance the contrast between the target and the background. The K-means and SSD algorithms were fused to improve the detection accuracy of fish target positions.

The experimental results showed that the average detection accuracy of the method based on improved SSD can reach 99.8% for the turned white belly fish, the average detection accuracy of normal fish can reach 98.8%, and the mAP of this model can reach 99.3%. Compared with SSD and Faster R-CNN, the method based on improved SSD had advantages in the detection accuracy of the turned white belly fish detection and model training speed. It can detect normal fish and the turned white belly fish in the fish swarm under complex background, and can be applied to actual production.

Acknowledgments
This study was supported by the Shandong Province Major Scientific and Technological Innovation Project under Grant 2019JZZY010703 and the Ningbo Public Welfare Science and Technology Project under Grant 202002N3034.
References

[1] Xu, X., Li, W., Duan, Q., 2021. Transfer learning and SE-ResNet152 networks-based for small-scale unbalanced fish species identification. Computers and Electronics in Agriculture 180.

[2] Gjedrem, T., Robinson, N., Rye, M., 2012. The importance of selective breeding in aquaculture to meet future demands for animal protein: A review. Aquaculture 350-353, 117-129.

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

[4] Fan, L., Zhao, H., Zhao, H., Hu, H., Wang, Z., 2020. A review of target detection based on deep convolutional neural networks. Optics and Precision Engineering. 28, 1152-1164.

[5] Guo, M., Zhao, Y., Xiang, J., Zhang, C., Chen, Z., 2014. Overview of target detection algorithms based on support vector machines. Control and Decision. 29, 193-200.

[6] Luo, H., Chen, H., 2020. A review of research on target detection based on deep learning. Chinese Journal of Electronics. 48, 1230-1239.

[7] Shao, T., 2017. Judgment of fish activity based on neural network and its application in water quality monitoring. Zhejiang University of technology. 29-40.

[8] Liu, X., Sun, Y., Zhao, D., Zhao, B., Cheng, L., Huang, Y., 2005. Monitoring and forecasting of fish health status using image processing technology. Transactions of the Chinese Society of Agricultural Engineering. 118-121.

[9] Yuan, H., Zhang, S., 2020. An underwater fish target detection method based on Faster R-CNN and image enhancement. Journal of Dalian Ocean University. 35, 612-619.

[10] Shen, J., Li, L., Dai, Y., Hu, F., 2020. Binocular fish school image detection based on improved SSD algorithm. Computer Engineering and Design. 41, 488-493.

[11] Huang, Y., Ding, W., Li, H., 2017. Unmanned aerial vehicle reconnaissance image defogging method based on image enhancement. Journal of Beijing University of Aeronautics and Astronautics. 43, 592-601.

[12] Wang, P., Xie, S., 2005. Photoshop CS Chinese Version Practical Course. Beijing: China Water Power Press.

[13] Sun, C., Vallotton, P., 2009. Fast linear feature detection using multiple directional non-maximum suppression. Journal of Microscopy. 234, 147-157.

[14] Zhao, L., Li, S., 2020. Object Detection Algorithm Based on Improved YOLOv3. Electronics 9.

[15] Yao, S., Chen, Y., Tian, X., Jiang, R., Ma, S., 2020. An Improved Algorithm for Detecting Pneumonia Based on YOLOv3. Applied Sciences 10.

[16] Bao, L., Wang, M., Liu, J., Peng, Z., Peng, S., 2020. Common tortoise identification method based on SSD target detection framework. Journal of Shenyang Agricultural University. 51, 238-244.