

## Research Article

# Convolutional Neural Network-Based Fish Posture Classification

Xin Li,<sup>1</sup> Anzi Ding,<sup>1</sup> Shaojie Mei,<sup>2</sup> Wenjin Wu ,<sup>1</sup> and Wenguang Hou <sup>2</sup>

<sup>1</sup>*Institute of Agricultural Products Processing and Nuclear Agricultural Technology, Hubei Academy of Agricultural Sciences, Wuhan 430064, Hubei, China*

<sup>2</sup>*College of Life Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, Hubei, China*

Correspondence should be addressed to Wenjin Wu; 272081603@qq.com and Wenguang Hou; houwg@hust.edu.cn

Received 22 March 2021; Revised 17 April 2021; Accepted 11 June 2021; Published 21 June 2021

Academic Editor: Danilo Comminiello

Copyright © 2021 Xin Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Fish killing machines can effectively relieve the workers from the backbreaking labour. Generally, it is necessary to ensure the fish to be in unified posture before being input into the automatic fish killing machine. As such, how to detect the actual posture of fish in real time is a new and meaningful issue. Considering that in the actual situation, we only need to determine the four postures which are related to the head, tail, back, and belly of the fish, and we transfer this task into a four-kind classification problem. As such, the convolutional neural network (CNN) is introduced here to do classification and then to detect the fish's posture. Before training the network, all sample images are preprocessed to make the fish be horizontal on the image according to the principal component analysis. Meanwhile, the histogram equalization is used to make the grey distribution of different images be close. After that, two kinds of strategies are taken to do classification. The first is a paired binary classification CNN and the second is a four-category CNN. In addition, three kinds of CNN are adopted. By comparison, the four-kind classification can obtain better results with error less than 1/1000.

## 1. Introduction

Fish is one of the most important foods for human beings. At present, lot of manual dissection is needed in making fish production. The production quality and processing speed depend on the worker's proficiency and efficiency. Moreover, the fish processing workshops are relatively humid and workers must touch the water and fish by hand for a long time. It greatly threatens the health of workers. In view of this situation, researchers try to build an automatic fish processing equipment to complete the automatic killing, descaling, and cutting. Generally, the fish should be orderly placed on the rolling shaft of the machine in a uniform orientation. On the conveyor belt, the mackerel scale is removed by the rolling blade. A knife is then used to cut the fish belly and remove internal organs. Finally, the processed fish is sent to the output of the machine [1, 2], that is, the machine mainly carries out descaling and laparotomy. However, manual operation is still needed to put the fish into inlet of the machine. Our team is trying to make an automatic fish input device, which is used to adjust the fish

posture to make fish be in uniform orientation before inputting into the machine [3]. Therefore, it is necessary to detect the actual orientation of the fish on the conveyor belt in real time and adjust it accordingly.

In another side, we can keep the fish horizontal relying on basis of the specially designed mechanical equipment [2, 3]. The posture of fish includes four cases: head forward and back up, head forward and back down, head back and back up, and head back and back down [1], i.e., we only need judge the four postures of fish from the image captured from the processing line. As such, the motivation of this study is to establish a method to judge fish's four kinds of posture. It should be mentioned that the experimental data in this study are not collected on the real production line. They are captured by camera or mobile phone artificially, i.e., the fish is not horizontally located on the image. In order to simulate the real scene, we first use the image processing algorithm to rotate the fish image to the horizontal; second, the image is preprocessed through histogram equalization and the blank cutting. These image operations will lessen the difficulty in later image classification. After preprocessing, the images

will be taken as training examples. Meanwhile, a convolutional neural network is designed to classify the pose of the image. According to the above discussion, this study intends to solve a four-kind classification problem. It can also be seen as two binary classification problems: the first is to judge the head direction and second judges the orientation of belly. Then, the fish posture is classified into four kinds on basis of the two dichotomies results. Generally, the traditional feature-based classification methods cannot be comparable to the deep learning-based method in the classification tasks which can be found in many related fields. But, the deep learning-based method relies on a certain number of samples. In our applications, however, the samples can be easily obtained. As such, the deep learning-based method is introduced here. The flowchart of the proposed method is shown in Figure 1. Generally speaking, there are a small amount of related works about the fish posture classification. Their application background is different from us [4–6]. Yet, we all want to introduce the image processing method to deal with the fish processing related issue. This study is organized as follows. Section 2 simply introduces the image preprocessing method. Section 3 reports the case on basis of binary classification, and Section 4 details the network structures and the related experiment in case of four-kind classification. At last, the conclusion, limitation, and possible further research studies are discussed. The operating system is Linux, while the CPU is Inter(R) Core(TM) i7-6700k and GPU is a GeForce GTX TITAN X, equipped with 3072 processor cores and 12 GB memory.

## 2. Image Preprocessing

The image of mackerel is considered in this article. Figure 2(a) shows its original image which is captured by a hand-held camera in a fish processing workshop. The processing workshop is drippy and illumination is uneven. The grey distribution is diverse in different images. Then, the histogram equalization is conducted for all original images to make the grayscale distribution of the images collected by different cameras under different illuminations be close as possible. The positions of the fish in our captured images are diverse. In actual application scene, however, the fish should be horizontally located on the image. Therefore, the first step of image preprocessing is to rotate the image, so that the fish is basically in a horizontal status in the rotated image. It is to generate the similar samples as the actual case. The basic principle is to perform principal component analyze (PCA) to the original image [7]. In the original image, the main significant target is the fish, which contains a small amount of background color, that is, the main axis direction of the original image is generally the direction of the fish body. PCA transformation is taken to compute the main axis of the image. The main calculation process is as follows: (a) binarize the original image on basis of the Otsu algorithm [8]; roughly divide the image into two kinds: target and background; (b) set the fish as the high grey (target) and the background as the low grey level; (c) the coordinates of all the target after binarization are formed into a matrix:

$$X = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \\ y_1 & y_2 & \cdots & y_n \end{bmatrix}, \quad (1)$$

where  $N$  denotes there are  $N$  highlight points in the image. Calculate the average as  $\bar{X} = [(1/n) \sum_{i=1}^n x_i, (1/n) \sum_{i=1}^n y_i]$ ,  $i = 1, 2, \dots, n$ . Then, we can get the variance and covariance matrix:

$$C = (X - \bar{X})(X - \bar{X})^T. \quad (2)$$

The size of the covariance matrix is  $2 * 2$ , and formula  $C = U^T \Lambda U$  is obtained after orthogonal decomposition, where  $\Lambda$  is the eigenvalue vector, and the eigenvalues are arranged from large to small.  $U$  is a unit orthogonal matrix with a size of  $2 * 2$ . Its first column  $U_1$  represents the main direction in the image, which is the direction of the fish body. The second column of  $U$  indicates the direction perpendicular to the main axis. It is easy to calculate the angle  $\theta$  between the vector  $U_1$  and  $X$  coordinate axis. The original image is rotated by angle  $\theta$ . The relationship between original pixel and its location on the rotation image is shown as equation (3), in which,  $X_i$  and  $Y_i$  denote the position of each pixel. After rotation, the fish will be in a horizontal position in the rotated image. After main axis transformation and histogram equalization [9], the image is obtained as shown in Figure 2(b). It can be seen that the fish is horizontally located in the image, and the grey distribution is relatively uniform, i.e., the samples for CNN training is obtained.

$$Y_i = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} X_i. \quad (3)$$

## 3. Paired Binary Classification

After image preprocessing, the paired binary classification CNN is used to compute the posture of fish. One is to judge the direction (left and right) of fish head or tail. The second one estimates the orientation (up and down) of the fish belly or back. After the paired binary classification, the posture of the fish can be obtained by finding their intersection.

*3.1. CNN Structure.* The designed binary classification convolution neural network is based on the VGGNet model [10]. Compared with the original structure, its network structure is greatly simplified. It includes eight layers: six convolutional layers and two fully connected layers, as shown in Figure 3. The output is the probability that the image belongs to two classes, represented by a one-dimensional vector including two elements. The Softmax function is used to normalize the output probability value to 0-1. Furthermore, the max pooling followed the Conv1, Conv3, and Conv6 layers. After the first fully connected layer, the dropout sublayer is introduced to prevent network overfitting and improve the generalization ability of the model. Moreover, batch normalization (BN) after the activation function is utilized in each layer to significantly accelerate the training speed and prevent the gradient from

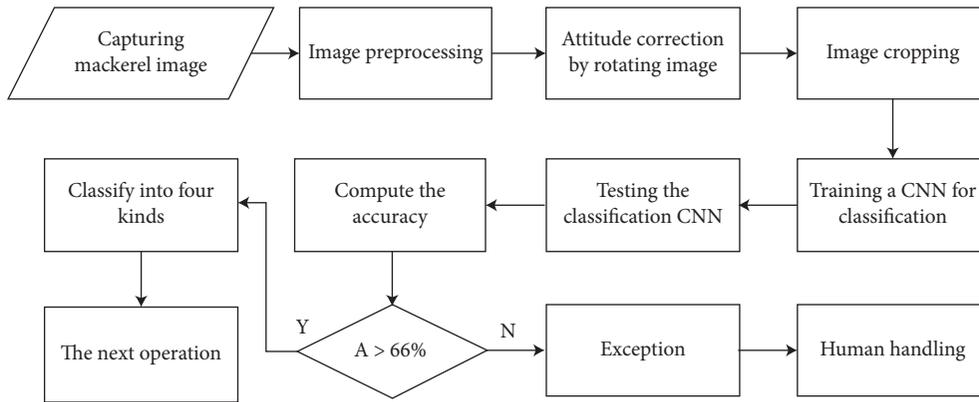


FIGURE 1: The flowchart of the proposed method.

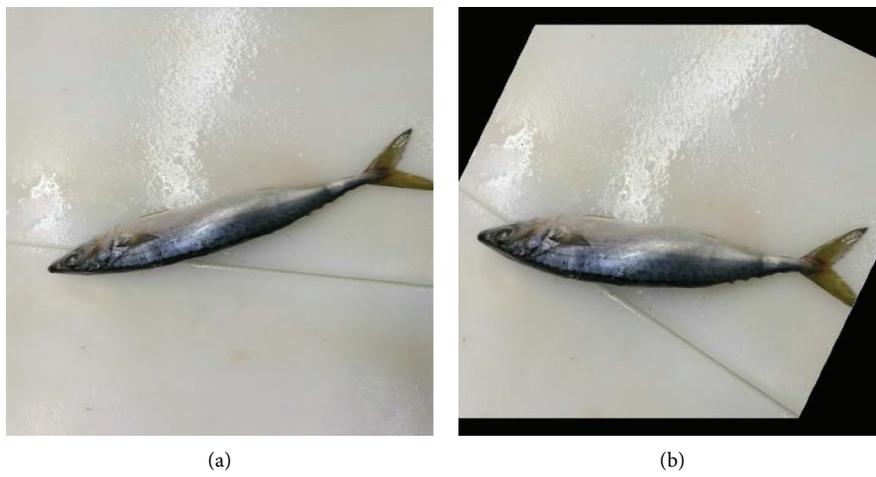


FIGURE 2: Comparison between (a) preprocessing and (b) postprocessing images.

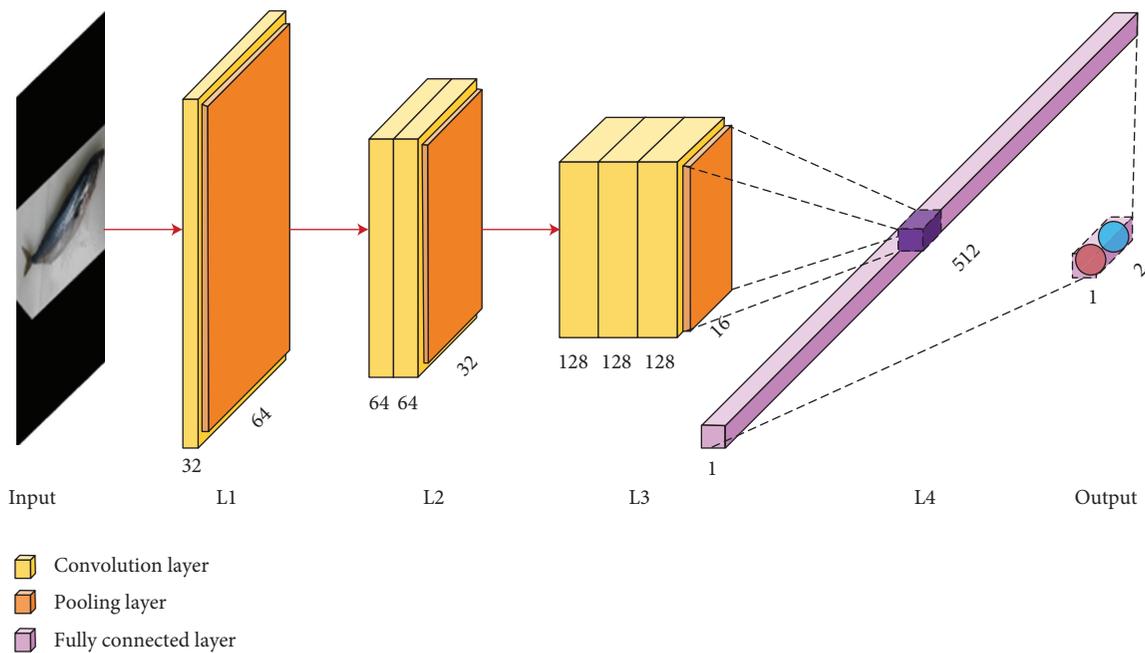


FIGURE 3: The structure of the CNN for binary classification.

disappearing. The architecture of the binary classification network is presented in Table 1.

**3.2. Dataset.** After the original mackerel images' pre-processing, we check them one by one to delete the invalid data. As shown in Figure 4, some fishes are curved, and their head or tails may be lost. The same character of those abandoned images is that their orientation cannot be judged and should be individually processed. After eliminating the invalid images, a total of 4440 images are remained. The number of samples in different classes is given in Table 2. The number of fishes whose faces are in left is 2240, and the fish's number in reverse direction is 2200. Correspondingly, the number of fishes whose belly is up is 2000, and the fish's number in reverse direction is 2440. Then, two classification CNNs are trained supervised by taking the image as input and their label as output. Figure 5 shows the examples of images in the binary classification task. For the classification tasks in judging fish head direction, we set the labels of the sample to be 0 and the reverse is 1. Correspondingly, the fish belly orientation computation, the sample whose belly is up is set to be 0 and the reverse is set to be 1.

**3.3. Network Training.** In the process of training, Adam is taken to optimize the network. The batch size is set to 4, and the loss function is binary cross-entropy. The network is trained 30 epochs, the initial learning rate is set to 0.01, and the learning rate is adjusted adaptively. The parameter number is about 3.8 M. Each epoch costs less than 10 seconds. Then, the total computation time is about 300 seconds. The network structure is established on basis of our experience, and the initial value of each parameter is randomly generated.

**3.4. Result.** In training, the cross-entropy loss is utilized as the loss function, and the stochastic gradient descent (SGD) is adopted to optimize the network, initialized by the truncated normal distribution with a batch size of 32. Furthermore, the learning rate is set to be 0.01, and a momentum rate is set to be 0. Moreover, the network is trained for 30 epochs. Meanwhile, an 80% cross-validation method is used to compute the classification accuracy in the training process to then further determine the classification threshold accordingly. In addition, we set the threshold of classification probability to be 66%, that is, the samples whose probability to each kind is less than 66% are considered as indistinguishable samples. The test results of two two-kind networks are, respectively, given in Tables 3 and 4.

There are two types of classification errors. First, the predicted probability of the input image belonging to each category is less than 66%. It will be regarded as uncertainty sample, which means the network cannot accurately classify this image. In real situation, these fish will be handled exceptionally. It can be drawn from the tables that the proportions of uncertain situations in two binary classification tasks are, respectively, 2% and 4%. Both are less than 5%. It indicates that very few cases require manual intervention for

exception handling in practice. The second error is that the prediction accuracy of the network is higher than 66%, but the classification result is wrong. The error rates of the two binary classification tasks about head direction and belly orientation are both equal to 4%, not exceeding 5%. The results demonstrate that the network achieves high accuracy on both classification tasks. In contrast, the error rate and uncertain probability of the head direction classification task is smaller than those of the belly orientation classification task. The main reason is that there are some special samples, which should be rare in practice. The confusion matrix of two binary classification tasks and the confusion matrix obtained by superimposing two binary classification results are shown in Figure 6. The error is not serious.

It is clear that the proposed classification method is effective, and the robustness is ensured due to that a threshold is introduced. Higher threshold will result in more classification accuracy but lead more samples to be unprocessed. According to our experiments, 0.66 should be a balance. Although high accuracy can be obtained, there still exist some error cases which can be easily judged by ordinary people. After carefully checking these samples, we found that the fish's scale is relatively big in these images. As such, we advise the fish's scale should be moderate and the enough margins should be retained in the captured image.

## 4. Four-Type Classification

In this section, we directly classify the image into four types: fish head forward, fish belly up; fish head forward, fish belly down; fish head back, fish belly up, and fish head back, fish belly down. Three networks are taken to do this task.

**4.1. CNN Structure.** Since the four-type classification task is more complicated than the two-type classification task, we conduct training and testing on three different networks. As shown in Figure 7, the structure of the first convolutional neural network we designed is basically same as the abovementioned binary network. The only difference lies in that is the output is a quaternion vector which corresponds to the probability of belonging to a certain type. Softmax function is used to normalize the output probability value to 0-1. The specific structure of each layer within the network is also the same as in Table 1.

As shown in Figure 8, we have made some modifications on the network parameters and the fully connected layer on basis of the VGG16 network [11, 12]. The network mainly includes 15 layers: 13 convolutional layers and 2 fully connected layers. The Softmax function is used in the output layer to do normalization. In addition, maximum pooling is used in the Conv2, Conv4, Conv7, Conv10, and Conv13 layers to reduce dimension of detected features. The dropout layer is used after the first fully connected layer to prevent overfitting and improve the generalization ability of the model. At the same time, batch normalization processing is used in each layer [13].

Figure 9 shows the complete structure of the ResNet-18 network which is mainly composed by 17 convolutional

TABLE 1: The parameters of convolutional neural networks.

Layer	Conv1	Conv2	Conv3	Conv4	Conv5	Conv6	FC1	FC2
Input size	64 * 64	32 * 32	32 * 32	16 * 16	16 * 16	16 * 16	8 * 8 * 128	512
Convolution kernel	3 * 3	3 * 3	3 * 3	3 * 3	3 * 3	3 * 3		
Output	32	64	64	128	128	128		
Boundary filling	1	1	1	1	1	1		
Step	1	1	1	1	1	1		
Activation function	ReLU	ReLU						
Pooling	2 * 2		2 * 2			2 * 2		
Dropout							0.6	

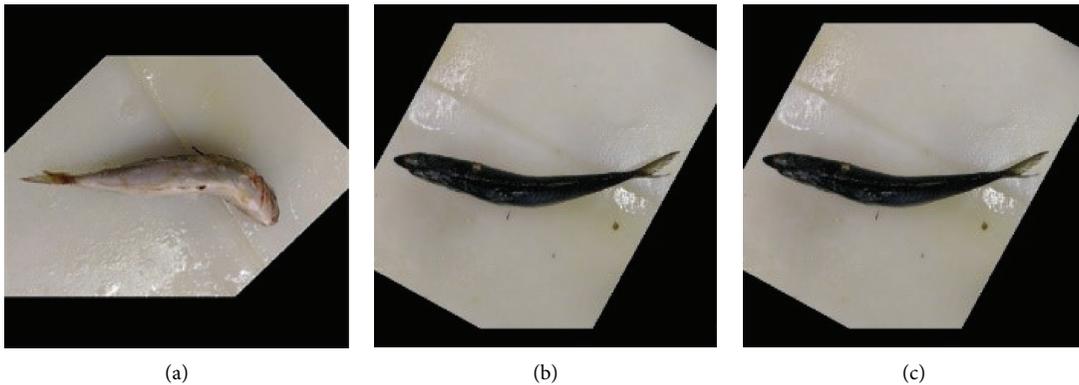


FIGURE 4: Invalide images. They will be removed from the training set.

TABLE 2: The number of samples.

Classification task	Image size	Image number	Direction	Number
Face head	224 * 224	4440	Left	2240
			Right	2200
Fish belly	224 * 224	4440	Up	2000
			Down	2440

layers and 1 fully connected layer [14]. Among them, we set the step size of the Conv6, Conv10, and Conv14 layers to be 2. Pooling operation downsamples the detected feature. The solid line in the short-circuit connection in the figure indicates that the input and output are directly added, and the dashed line indicates that the input  $x$  is first upsampled and downsampled until it has the same size as the output; due to that, the input and output dimensions may be different. Similarly, batch normalization is used in each layer to speed up CNN convergence and prevent overfitting.

**4.2. The Data.** As mentioned above, we have obtained a total of 4440 mackerel pictures with different poses whose size is 224 \* 224. The four different postures are fish head forward, fish belly up; fish head forward, fish belly down; fish head back, fish belly up, and fish head back, fish belly down. The number of pictures for each posture is 1000, 1240, 1000, and 1200, respectively. Figure 10 shows the samples in the four-type classification task.

**4.3. Network Training.** In the process of training, Adam is taken to optimize the network. The batch size is set to 4, and the loss function is binary cross-entropy. The network is trained 40 epochs, the initial learning rate is set to 0.01, and the learning rate is adjusted adaptively. The parameter numbers of VGGNet, VGG16, and ResNet-18 are 4.6 M, 4.2 M, and 11 M. Each epoch costs about 10 seconds. Then, the training times of three networks are about 265, 341, and 432 seconds. The network is not complex, and the parameter of network is experienced, established by borrowing the basic network structure from the related fields.

**4.4. Result.** The cross-entropy loss function is used for training, and the networks are still optimized by the stochastic gradient descent method. The learning rate is 0.01, the momentum of the simple VGGNet and VGG16 is set to 0, the momentum of ResNet-18 is set to 0.9, batch size is set to 32, the iteration is 30 times, and the classification threshold is 66%. Table 5 provides the test statistics of three different networks.

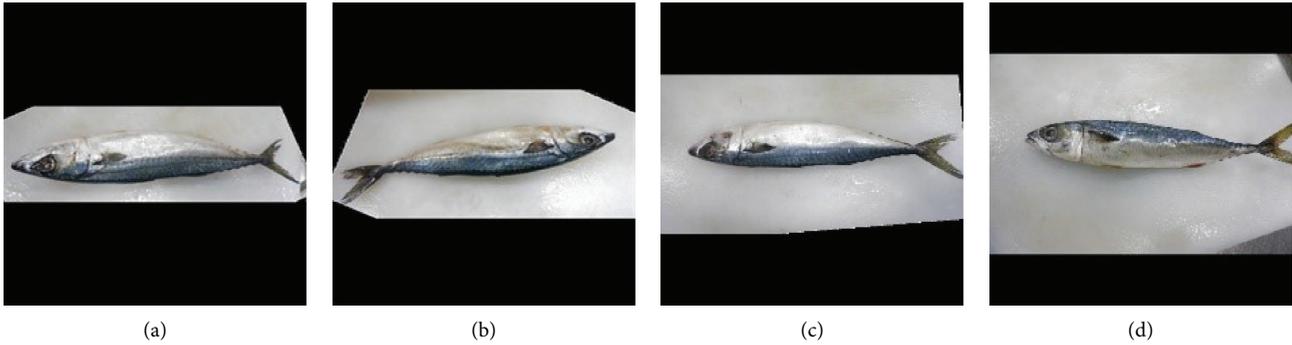


FIGURE 5: Two kinds. (a)-(b) Head is on left or right. (c)-(d) Belly is up or down.

TABLE 3: Test result about fish head direction.

Correct	Rate > 66	Wrong Rate < 66	Total	Error ratio (%)	Uncertainty (%)
4413	16	11	27	4	2

TABLE 4: Test result about fish belly direction.

Correct	Rate > 66	Wrong Rate < 66	Total	Error ratio (%)	Uncertainty (%)
4405	18	17	35	4	2

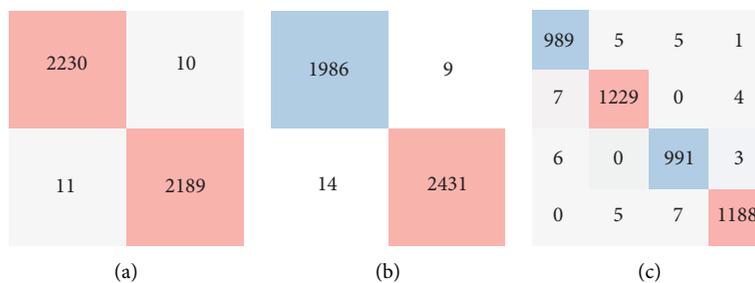


FIGURE 6: Two classification confusion matrix. (a) The result of head direction, (b) the face back direction, and (c) their superposition.

It can be concluded from the above table that for the first type of uncertainty rate, Simple VGGNet has the largest uncertainty probability, while the uncertainty probability of VGG16 and ResNet-18 is very small, only 2%. It shows that the deepening of the network is beneficial to eliminate the uncertainty of classification. For the four-type classification task, the uncertainty probability of the three networks is very small. For the second type of classification error, ResNet-18 performs best. Simple VGGNet and VGG16 have similar error rates, but the classification error rates of the three are very low, indicating that the probability of actual errors in the four classifications is very small, and the accuracy rate is very high. Figure 11 is the confusion matrix of the three-network cross-validation. According to experiments, the two classification methods have achieved better classification

results. To ensure the fairness of the comparison, the training and testing samples are same for all networks and the sizes of the networks are close.

Although the classification accuracy of a simple shallow network should also be able to meet the needs of practical applications, the ResNet-based four-classification network directly has higher robustness. It is due to that the ResNet can make the loss error be inversely delivered to all parameters leading to finely optimize the parameters. In another sides, the differences among three networks are not obvious. Compared with the paired binary classification, the four-type classification obtains better accuracy and due to that, the direction of head and orientation of fish back are considered at the same time. These two features mutually verified leading to finer classification results. As such, it is

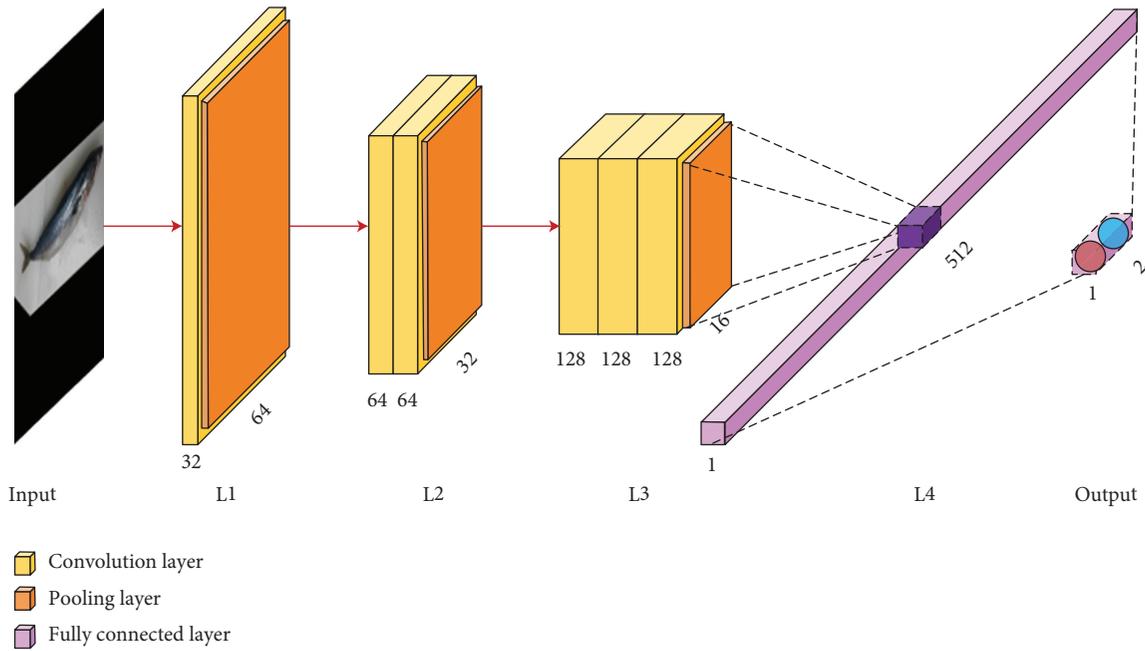


FIGURE 7: A designed simple VGGNet.

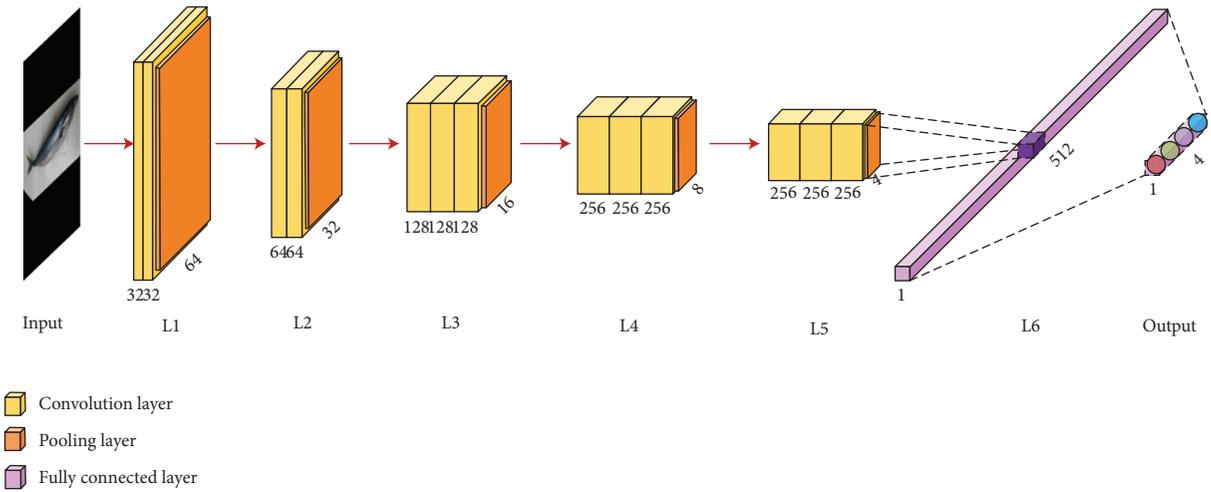


FIGURE 8: The structure of VGG16 CNN.

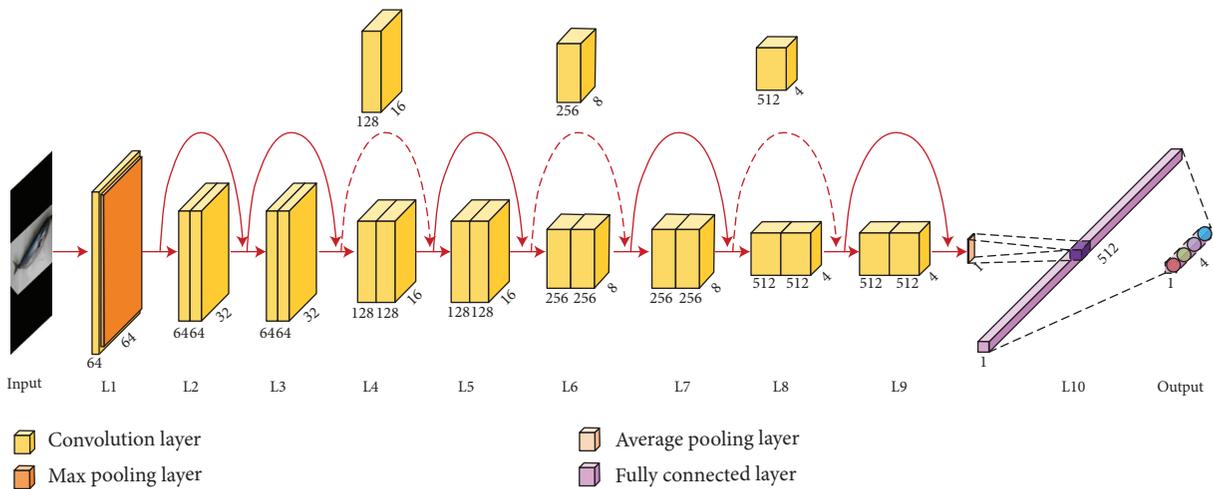


FIGURE 9: The structure of ResNet-18 CNN.

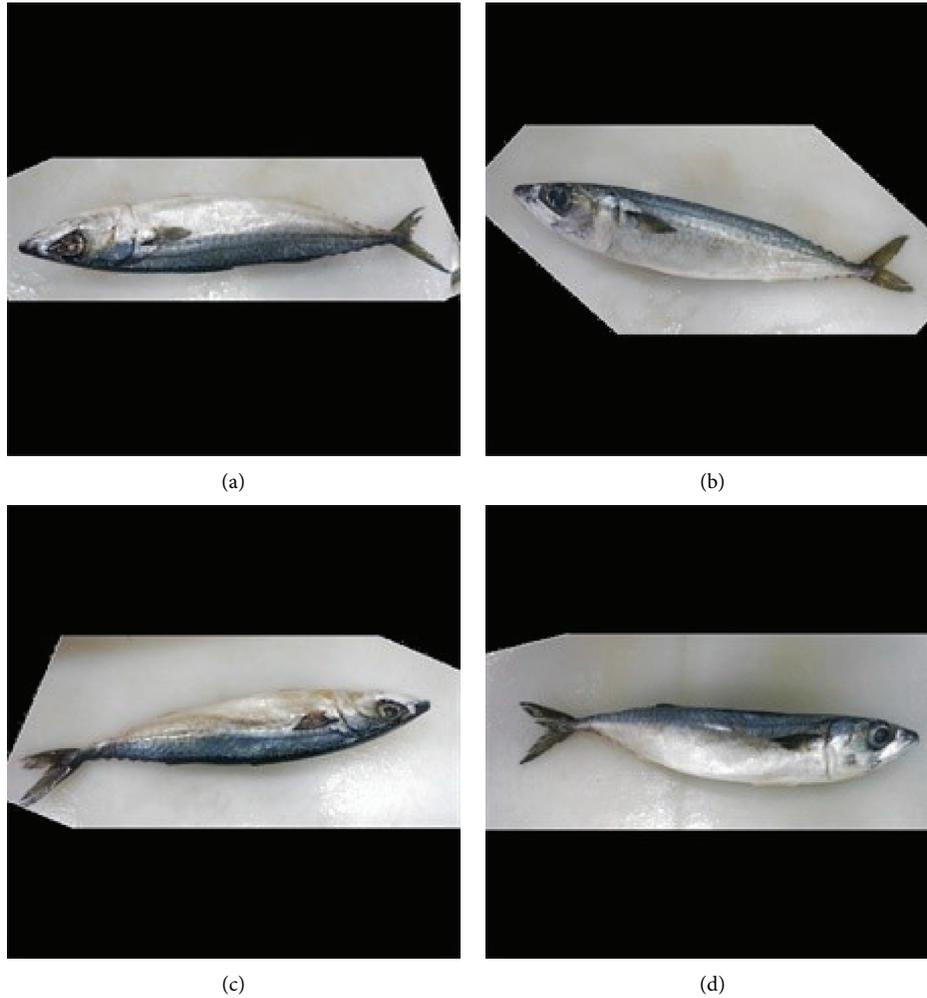


FIGURE 10: Four categories classification result. (a) Head forward, belly up. (b) Head forward, belly down. (c) Head back, belly up. (d) Head back, belly down.

TABLE 5: Testing result of three networks.

Networks	Correct	Wrong		Total	Error ratio (%)	Uncertainty (%)
		Rate > 66	Rate < 66			
Simple VGGNet	4375	32	33	65	7	7
VGG16	4402	29	9	38	7	2
ResNet-18	4408	23	9	42	5	2

986	8	6	0	992	4	4	0	994	3	3	0
8	1225	0	7	5	1229	0	6	5	1231	0	4
7	0	992	1	3	0	995	2	4	1	993	2
0	2	2	1196	0	3	4	1193	0	1	2	1197
	(a)				(b)				(c)		

FIGURE 11: Four-class cross-validation confusion matrix based on (a)-(c) VGGNet, VGG16, and ResNet-18.

better to do classification considering all indexes than to classify separately. In addition, the similarity threshold ensures the robustness and reduces the error.

## 5. Conclusion

To unify the posture of fish is the necessary step to do automatic processing, how to design an automatic fish posture adjustment equipment to replace the current manual operation has an important practical significance. Considering that in the actual situation, we only need to determine the orientation of the head and belly of the fish. As such, we transfer it into a four-kind postures classification problem. The convolution neural network is introduced here to do classification. Yet, the posture in the captured image is more complex which makes the classification error be high without any preprocessing operations. Then, the main axis transformation is taken to rotate the image to make the subsequent classification be same as the practical situation. Meanwhile, histogram equalization is used to make the grey distribution be similar in different images. They all help to lessen the difficulty in image classification. The experiments validate the effectiveness of the proposed method. To the best of knowledge, less report can be found about deep learning-based classification usage in applications of fish food. Yet, many spatial posture recognition problems can be transformed into classification question. As such, the proposed method can be extent to other applications such as production orientation detection and appearance quality monitoring.

Meanwhile, the fish in the tested image is rather distinct, and its background is plain which effectively reduce some error classification. Moreover, the image is captured by a high-resolution camera and mobile phone. It takes a certain amount of time to capture the demanded image. Yet, in actual production line, the image capturing time may be limited. It means that more image preprocessing operations should be done before inputting it into the classification network. Another limitation is that some special cases such as frizzy fish cannot be judged in this system. As such, the future work should try to improve the robustness and prediction effect and cover more situations into the scope of this solution.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This study was supported by the National Key Research and Development Program “Research and Development of aquatic raw material morphology accurate identification technology and intelligent pretreatment equipment” (2019YFD0901801).

## References

- [1] L. Feng, C. Zhu, and L. Xiaoguang, “Control system design of auto fish processing equipment based on plc,” *Mechanical Engineer*, vol. 9, pp. 33-34, 2015.
- [2] L. Zhao and C. Yang, “Design of a small automatic fish killing machine,” *Modern Manufacturing Technology and Equipment*, vol. 2, pp. 168-169, 2018.
- [3] Q. He, J. Ding, Y. Yang et al., “Design of small integrated scale machine,” *Technology Wind*, vol. 2, pp. 9-10, 2021.
- [4] Z. Hu, Y. Zhang, Y. Zhao, L. Cao, Y. Bai, and M. Huang, “Fish eye recognition based on weighted constraint adaboost and pupil diameter automatic measurement with improved hough circle transform,” *Nongye Gongcheng Xuebao/Transactions of the Chinese Society of Agricultural Engineering*, vol. 3323, pp. 226–232, 2017.
- [5] G. Ma, Y. Qi, and Y. Tian, “A preliminary study on automatic shape evaluation of blood parakeet fish based on image processing,” *Information Technology and Informatization*, vol. 9, pp. 27–30, 2019.
- [6] R. A. Tidd and J. Wilder, “Fish detection and classification system,” *Journal of Electronic Imaging*, vol. 10, no. 1, pp. 283–288, 2001.
- [7] K. Cheolmin and D. Klabjan, “A simple and fast algorithm for l1-norm kernel pca,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 428, pp. 1842–55, 2020.
- [8] L. Wang, Y. Zhao, P. Qin, and Y. Gao, “Harris corner detection algorithm optimization based on otsu,” *Recent Advances in Electrical and Electronic Engineering*, vol. 112, pp. 91–96, 2018.
- [9] C. Xu, Z. Peng, X. Hu, W. Zhang, L. Chen, and F. An, “FPGA-based low-visibility enhancement accelerator for video sequence by adaptive histogram equalization with dynamic clip-threshold,” *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 67, no. 11, pp. 3954–3964, 2020.
- [10] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” arXiv:1409.1556, 2015.
- [11] Y. Cao, S. Xing, and J. Pang, “Detection of pulmonary nodules based on improved vgg-16 convolution neural network,” *Chinese Journal of Medical Physics*, vol. 374, pp. 940–944, 2020.
- [12] G. S. B. Jahangeer and T. D. Rajkumar, “Early detection of breast cancer using hybrid of series network and VGG-16,” *Multimedia Tools and Applications*, vol. 80, no. 5, pp. 7853–7886, 2021.
- [13] K. He, X. Zhang, and S. Ren, “Deep residual learning for image recognition,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, IEEE, Las Vegas, NV, USA, June 2016.
- [14] A. Zanchetta and S. Zecchetto, “Wind direction retrieval from Sentinel-1 SAR images using ResNet,” *Remote Sensing of Environment*, vol. 253, Article ID 112178, 2021.