Aquaculture 507 (2019) 457–465

Contents lists available at ScienceDirect

Aquaculture

journal homepage: www.elsevier.com/locate/aquaculture

Evaluation of fish feeding intensity in aquaculture using a convolutional neural network and machine vision

Chao Zhou[a,c,d], Daming Xu[a,c,d], Lan Chen[a,c,d], Song Zhang[a,c,d], Chuanheng Sun[a,c,d], Xinting Yang[a,c,d], Yanbo Wang[b,*]

[a] Beijing Research Center for Information Technology in Agriculture, Beijing 100097, China
[b] School of Food Science and Biotechnology, Zhejiang Institute of Food Quality and Safety Engineering, Zhejiang Gongshang University, Hangzhou 310018, China
[c] National Engineering Research Center for Information Technology in Agriculture, Beijing 100097, China
[d] National Engineering Laboratory for Agri-Product Quality Traceability, Beijing 100097, China

ARTICLE INFO

Keywords:
Aquaculture
Machine vision
Feeding intensity grading
Convolutional neural network
Data augmentation

ABSTRACT

In aquaculture, information on fish appetite is of great significance for guiding feeding and production practices. However, most fish appetite assessment methods are inefficient and subjective. To solve these problems, in this study, an automatic method for grading fish feeding intensity based on a convolutional neural network (CNN) and machine vision is proposed to evaluate fish appetite. The specific implementation process was as follows. First, images were collected during the feeding process, and a dataset was constructed and extended using rotation, scale, and translation (RST) augmentation techniques and noise-invariant data expansion. Then, a CNN was trained on the training dataset, and the fish appetite levels were graded using the trained CNN model. Finally, the performance of the method was evaluated and compared with other quantitative and qualitative feeding intensity assessment methods. The results show that the grading accuracy reached 90%; thus, the model can be used to detect and evaluate fish appetite to guide production practices.

1. Introduction

In aquaculture, assessments of fish appetite can be used to guide feeding and production practices. Such assessments are useful for promoting production efficiency and reducing the unnecessary feedings (Zhao et al., 2019; Zhou et al., 2018a). Fish appetite reflects feeding-related conditions and is a highly useful input when developing a precise feeding or production management system (Saberioon et al., 2017; Sun et al., 2016).

Fish feeding intensity reflects the intensity and amplitude change of action and behavior during fish feeding, which can directly reflect fish appetite. Many attempts have been made to monitor and analyze the hunger status of farmed fish. For example, fish appetite can be rated according to a 4-step scale (Overli et al., 2006), but such ratings are often influenced by factors such as the individual observer's experience. Moreover, in commercial-scale fisheries, manual observations increase labor and time costs. Acoustic sensors have also been used to obtain and quantify fish behavior and feeding intensity (Kolarevic et al., 2016; Rakowitz et al., 2012) and monitor feed consumption (Juell and Westerberg, 1993), and this approach functions under low- and uneven-illumination conditions. However, acoustic technology is difficult to use in real production practice due to high costs and development difficulties of it (Zion and Barki, 2012). Mechanical switches and infrared photoelectric sensors have been used to detect fish feeding behavior (Rubio et al., 2004; Yukinori et al., 2016), but these approaches require the fish to be trained. Therefore, there is an urgent need for an automated and objective assessment of fish feeding intensity (Liu et al., 2014; Mallekh et al., 2003).

Machine vision is an ideal automatic, noninvasive, economical and effective method for assessing and detecting fish appetite (Wishkerman et al., 2016). The visual properties of fish change during the feeding process, for example, texture and optical flow (Pautsina et al., 2015; Zhao et al., 2017; Zhou et al., 2018b). Previous studies have commonly used feeding intensity metrics to assess appetite. Appetite is assessed by detecting fish feeding activity and feed consumption. Then, this information is used to automatically control the feeding process (Atoum et al., 2015; Sadoul et al., 2014). By using machine vision, the differences between fish feeding images can be assessed using the interframe difference method. Then, quantitative indicators of feeding behavior were proposed based on the differences (Duarte et al., 2009). Liu et al. (2014) proposed a computer vision-based feeding activity index for Atlantic salmon. The linear correlation coefficient between this method

* Corresponding author.
E-mail addresses: wangyb@mail.zjgsu.edu.cn, wyb1225@163.com (Y. Wang).

https://doi.org/10.1016/j.aquaculture.2019.04.056
Received 17 August 2018; Received in revised form 21 April 2019; Accepted 21 April 2019
Available online 23 April 2019

and manual observations of the feeding activity index reached 0.92. Based on these findings, Ye et al. (2016) used optical flow to extract the behavioral characteristics (speed and rotation angle) of a population, and the gastrointestinal digestion index and combined entropy were used to evaluate the feeding intensity of a fish population. In addition, this laboratory also proposed a method to measure the feeding intensity of a fish population to assess the real-time appetite of the fish swimming in a recirculating aquaculture system (Zhao et al., 2017), in which the assessment accuracy of the appetite of the school was maintained at a low nonmatch rate ((2.19 ± 0.81)% best) under ten different sampling durations. The studies mentioned above demonstrate that with respect to the feeding process, using machine vision to study feeding behavior and evaluate the feeding process has become a trend in research and development (Zhou et al., 2018b). Thus, the feasibility of using machine vision to assess the feeding intensity of fish has already been demonstrated.

The abovementioned research essentially utilized manually predesigned features for identification. The applicability of using such features to represent fish feeding intensity evaluations is often low, and it is difficult to determine the optimal solution. Simple histograms or feature fusion often ignore the relative spatial information of the image itself; thus, such approaches have many limitations. Compared with traditional methods, deep learning models use data to conduct self-learning of characteristics and expression relations directly to extract representative features (LeCun et al., 2015).

In aquaculture, deep learning approaches have already been employed in fields such as fish species classification (Ahmad et al., 2016; Siddiqui et al., 2018), behavior analysis and trajectory tracking (Beyan et al., 2018; Wang et al., 2017; Xu and Cheng, 2017; Zhao et al., 2018b), live fish recognition (Meng et al., 2018; Qin et al., 2016; Zhao et al., 2018a), and water quality prediction (Ta and Wei, 2018). Among deep learning methods, convolutional neural network (CNN) models, which have gained popularity in recent years, have been widely used in tasks involving image recognition, such as fish species classification, counting fruit, object detection (Ahmad et al., 2016; Moniruzzaman et al., 2017; Sa et al., 2016) and many others. The main objective of a CNN is to autonomously learn the implicit relationships between image pixel features, underlying features, high-level abstract features and predict final categories through layer-by-layer mapping of deep neural networks. This method is conducive to capturing the rich connotations of the data itself while avoiding the complicated manual feature design process. Successful CNN models include LeNet5 (Lecun et al., 1998), AlexNet (Krizhevsky et al., 2012), GoogLeNet (Szegedy et al., 2015) and ResNet (He et al., 2016), all of which have been used for image recognition tasks by an increasing number of scholars. Thus, CNNs are feasible for classifying fish hunger levels.

Based on a simulation of a commercial aquaculture site, a grading method for fish feeding intensity based on machine vision and a CNN is proposed in this study. The goal was to achieve automatic grading of feeding intensity. The performance of the proposed method is evaluated, providing valuable information for real-time feedback and automatic control of fish feeding.

2. Materials and methods

In this paper, a CNN was used to automatically extract the characteristics of fish feeding images to classify feeding intensity. Four levels of fish feeding intensity were evaluated: "none", "weak", "medium" and "strong". The method is outlined in Fig. 1:

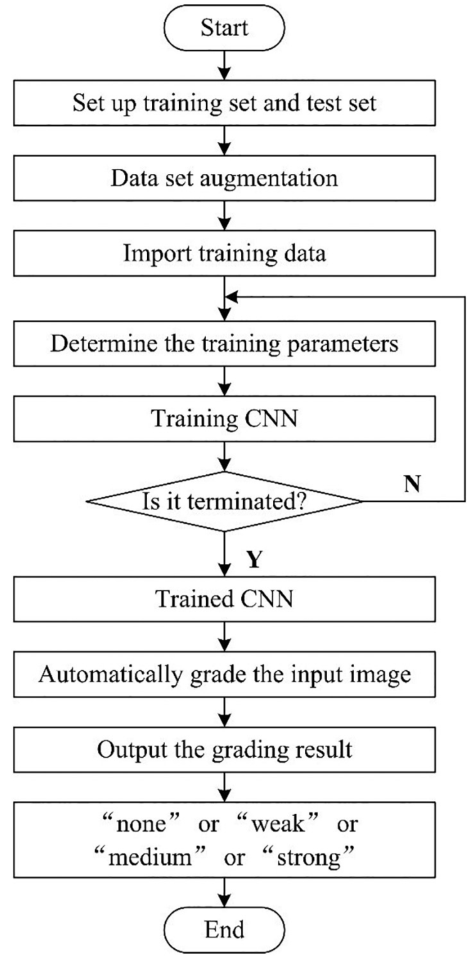


Fig. 1. The flowchart of the grading method for fish feeding intensity.

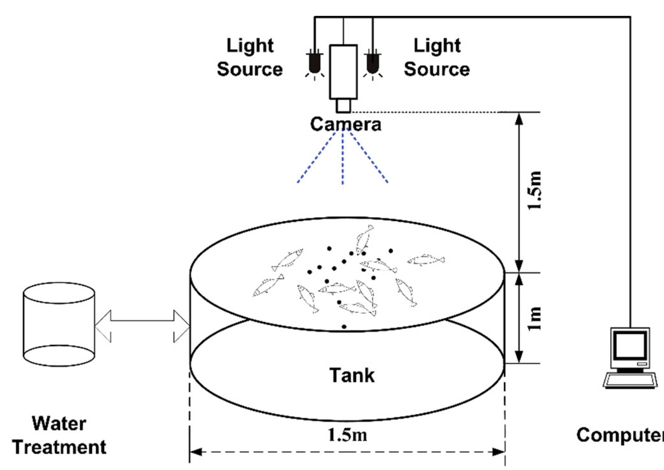


Fig. 2. The structure of the experimental system.

2.1. Experimental system and materials

Experiments were carried out in the fish rearing laboratory at the Xiaotangshan National Experiment Station for Precision Agriculture (Changping, Beijing, China). The experimental system consisted of six tanks, each with a diameter of 1.5 m and a water depth of 1 m (Fig. 2) (Zhou et al., 2017; Zhou et al., 2018c). Image acquisition and processing were conducted using a near-infrared industrial camera (Mako NIR G-223B), a near-infrared light source, and a computer. The camera was fixed to the top of the tank. Aeration was performed using a biofilter to prevent bubbles from influencing the image processing results. At the same time, to avoid water surface fluctuations as much as possible, the water inlet was submerged below the water surface, and the water flow entered from a tangential direction. The computer was placed in a control room next to the tank and connected to the camera via a 30-m gigabit twisted cable to reduce fish behavioral abnormalities caused by human activities. Image acquisition and analysis were performed using self-developed software. The software development toolkit was provided by AVT and used to convert the raw image stream data into other formats, such as bmp. In addition, the image acquisition frame rate, exposure time and resolution were configurable. The data transport layers supported Allied Vision cameras with GigE Vision, USB3 Vision, IEEE 1394, and Camera Link interfaces. Image acquisition began ten minutes before each feeding at a frame rate of 1 frame/s. Thus, images were collected of all three stages (before, during and after feeding). Ten minutes after feeding, image acquisition was terminated.

In these experiments, tilapia was used as the experimental object. The tilapia was purchased from Xiao Tangshan Aquaculture Breeding Development Center (Changping, Beijing, China). Thirty fish (weight: 223.02 ± 5.5 g, size: 19.7 ± 2.8 cm) were cultured in each tank and fed with floating pellets. The oxygen level was maintained within a range of 5.8–8.2 mg/L, and the water temperature was maintained between 20–25 °C. Before the experiment, the fish were cultured in the experimental system for 4 weeks; thus, they were considered to have adapted to the rearing environment.

2.2. CNN framework

A CNN is an artificial neural network based on deep learning theory that uses images directly as input, thus avoiding complex operations such as feature extraction. A CNN can be used to classify feeding intensity and is highly suitable for high-density aquaculture fields (Ding and Taylor, 2016). Among the different types of CNNs that have been developed, the LeNet5 framework-based CNN has a strong processing capability for training on datasets with a relatively small volume (Ding and Taylor, 2016). This method is easy to implement and can achieve a high recognition rate (LeCun et al., 2015).

Based on the LeNet5 CNN, combined with fishery image data and the characteristics of the samples to be tested, a CNN model was constructed; the basic parameters and structure are shown in Fig. 3. The model includes 7 layers (not counting the input): C1 (convolutional layer), S2 (subsampling layer), C3 (convolutional layer), S4 (subsampling layer), C5 (convolutional layer or first fully connected layer) and F6 (fully connected layer). The input is a 50 × 80-pixel image. C1 contains 5 46 × 76 feature maps; each unit in each feature map is connected to a 5 × 5 neighborhood of the input. C1 contains 130 trainable parameters and 454,480 connections. S2 is a subsampling layer with 5 23 × 38 feature maps, and each unit in each feature map is connected to a 2 × 2 neighborhood of the corresponding feature map in C1. Layer S2 has 10 trainable parameters and 21,850 connections. Layer C3 is a convolutional layer with 10 feature maps, and each unit in each feature map is connected to several 5 × 5 neighborhoods at identical locations in a subset of the feature maps in S2. C3 contains 1260 trainable parameters and 813,960 connections. Layer S4 is a subsampling layer with 10 5 × 5 feature maps, 20 trainable parameters and 7650 connections. Layer C5 is a convolutional layer with 120 feature maps. Layer S4 has 30,120 trainable parameters and 1,957,800 connections. Layer F6 contains 84 units, and layer S4 has 10,164 trainable parameters and 660,660 connections. The output layer contains 4 neurons that represent the four hunger status levels in the fish feeding process. The output layer classification function uses the “softmax” function.

Table 1
CNN training parameters.

Parameter	Value
learning_rate	0.05
n_epochs	200
nkerns	[5, 10]
batch_size	40

The choice of parameter values is a key step in this method and is essential for achieving high precision during training. After many trials, the initial parameters of CNN were set to the values shown in Table 1.

2.3. Training set production

Insufficient sample size is a frequent problem in deep learning, especially when only small sample training datasets are available. Expanding the sample size is the most direct way to solve this problem; such expansion not only improves the classification effect but also the generalization ability of the model. During the fish feeding process, due to the influence of feeding activities, the region of interest is often not in the center of an image and is occasionally accompanied by image

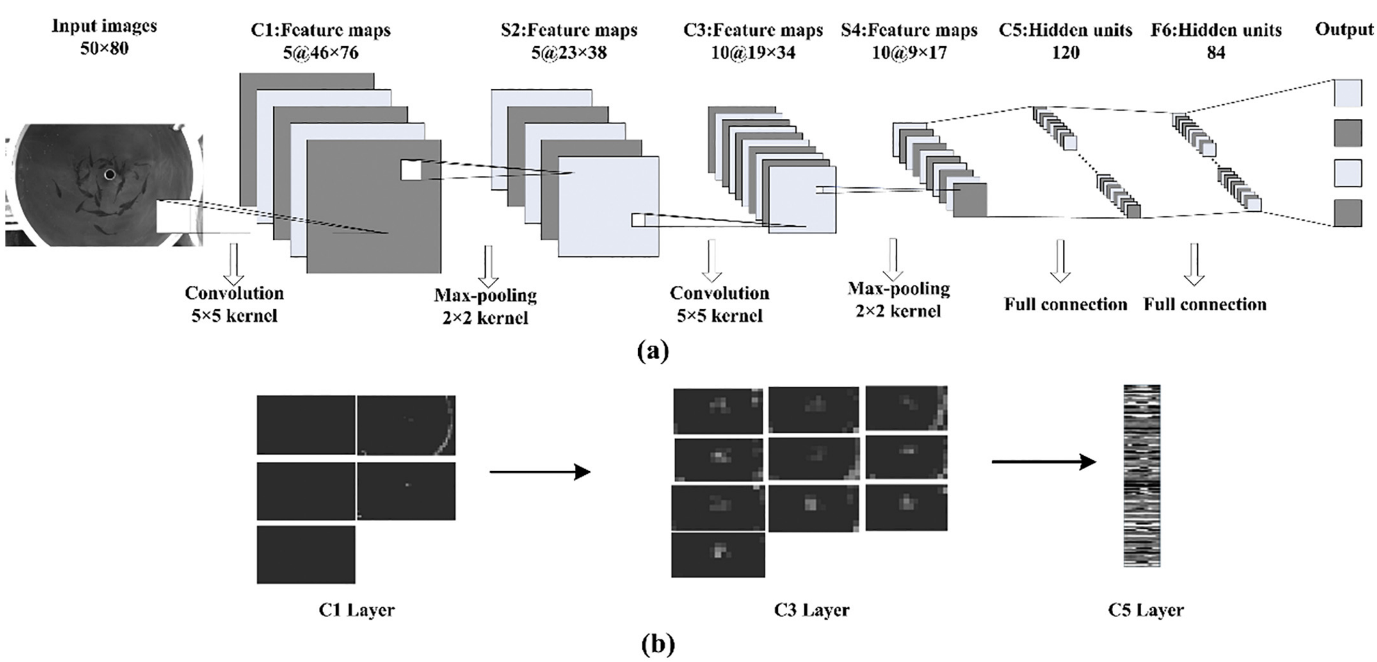


Fig. 3. CNN architecture and stages: (a) LeNet structure; (b) example images from each stage.

rotation. In addition, most environmental conditions are uncontrollable: images may be affected by illumination and other conditions. Image preprocessing cannot completely remove these noises. Therefore, based on the characteristics of the fish feeding images, the dataset for this study was augmented using rotation, scale, and translation (RST) techniques and noise-invariant data expansion. The expanded mathematical form is as follows, where g is the augmented image and f is the original image.

$$g(x,y) = f(T(x,y)) + n(x,y)$$
$$T(x,y) = \begin{pmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} s_x & & \\ & s_y & \\ & & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \tag{1}$$

where t_x and t_y are the translational quantities, s_x and s_y are the grading scales, and $n(x,y)$ is the distance by which the image pixels are moved. Specifically, the augmentation approach was conducted as follows:

(1) The original image was flipped horizontally to obtain its vertically symmetrical image. Mathematically, this transformation is expressed as follows:

$$g_1(x,y) = f(-x,y) \tag{2}$$

The original image and its horizontally flipped image were then shifted by 5 pixels in the upward, downward, left and right directions. Finally, 8 translational augmented images were obtained. These translations are expressed mathematically as follows:

$$\begin{aligned} &g_2(x,y) = f(x+5,y),\ g_3(x,y) = f(x-5,y) \\ &g_4(x,y) = f(x,y+5),\ g_5(x,y) = f(x,y-5) \\ &g_6(x,y) = g_1(x+5,y),\ g_7(x,y) = g_1(x-5,y) \\ &g_8(x,y) = g_1(x,y+5),\ g_9(x,y) = g_1(x,y-5) \end{aligned} \tag{3}$$

(2) The original image and its horizontally flipped image were each rotated by 5 degrees to the left and right to obtain 4 rotated images. These transformations are expressed mathematically as follows:

$$\begin{aligned} &g_{10}(x,y) = f(x\cos(5°) + \text{ysin}(5°), y\cos(5°)\text{-xsin}(5°)) \\ &g_{11}(x,y) = f(x\cos(-5°) + \text{ysin}(-5°), y\cos(-5°)\text{-xsin}(-5°)) \\ &g_{12}(x,y) = g_1(x\cos(5°) + \text{ysin}(5°), y\cos(5°)\text{-xsin}(5°)) \\ &g_{13}(x,y) = g_2(x\cos(5°) + \text{ysin}(-5°), y\cos(-5°)\text{-xsin}(-5°)) \end{aligned} \tag{4}$$

(3) Two groups of noise (salt-and-pepper noise and Gaussian noise) were added to the original image and its horizontally flipped image. The densities of the salt-and-pepper noises were 0.01 and 0.04, and the variances of the Gaussian noises were $10^2/255^2$ and 0.01. Finally, 8 noise images were obtained.

All the actions mentioned above were applied to the 80 × 50-pixel images. Examples of the expanded images are shown in Fig. 4.

In this study, feeding intensity was divided into four levels following the studies of Eriksen et al. (2011) and Ye et al. (2016). The grading standards are listed in Table 2. A total of 400 images at these 4 levels were manually selected as the dataset. After data augmentation, each original image was expanded to 22 images, which increased the total number of samples in the training set to 8800. Of those images, 80% were used as the training set, 10% were used as the validation set, and 10% were used as the test set.

After the CNN was trained using the selected parameter values, the image features at each level were automatically extracted, and the mapping correspondence between the features and the data tags was established. Finally, the trained CNN model was used to classify the feeding images to obtain the four fish feeding intensity levels: "none", "weak", "medium" and "strong".

2.4. Performance evaluation

In this study, accuracy, precision, recall and specificity were used to evaluate the performance of the proposed method. Accuracy is the ability to correctly identify the various feeding intensities; it is the ratio of the number of correctly graded samples to the total number of samples. At a specific level of feeding intensity, precision is the proportion of correctly graded images among all the images predicted to belong to a specific class. Recall is the proportion of correctly classified items among all items to be classified. These three indicators can be defined as follows:

$$Accuracy = \frac{\text{TP} + TN}{TP + FN + FP + TN} \times 100\% \tag{5}$$

$$\text{Precision} = \frac{\text{TP}}{TP + FP} \times 100\% \tag{6}$$

$$\text{Recall} = \frac{\text{TP}}{TP + FN} \times 100\% \tag{7}$$

$$\text{Specificity} = \frac{\text{TN}}{TN + FP} \times 100\% \tag{8}$$

where true positives (*TP*), false positives (*FP*), false negatives (*FN*) and true negatives (*TN*) respectively indicate that a positive class was judged correctly as a positive class, a negative class was judged to be a positive class, a positive class was judged to be a negative class and a negative class was judged to be a negative class.

3. Results and discussion

3.1. Feeding intensity grading results of CNN

The CNN model was trained and tested. Fig. 5 indicates that after 80 iterations, the verification error of the original samples eventually stabilized at approximately 40%. After augmenting the dataset, the verification error of the model decreased, reaching 10%, where it remained stable. The convergence speed also increased after augmentation, which reduced the training time. Therefore, the robustness of the model was improved by including a variety of interference and noise in the training set.

To further verify the performance of the method, 80 photographs were selected and graded using the trained CNN. The results are shown in Table 3. Seventy-two frames were correctly identified, and the final accuracy reached 90%.

Fig. 6 shows the sample image results. Panels (a)–(d) are examples of the four levels of feeding intensity. The results show that the grading results are essentially consistent with human intuition. The algorithm achieved the best grading effect when the feeding intensity was either "none" or "strong". Therefore, the CNN can detect feeding intensity and offers a practical method for assessing fish appetite.

3.2. Method performance

3.2.1. Comparison with qualitative feeding intensity assessment methods

To assess the performance of the method for each feeding intensity grade, 80 samples containing individual feeding intensities were randomly selected and their feeding intensity levels were graded using a support vector machine (SVM), back propagation neural network (BPNN) and the proposed CNN. Feature extractions for the SVM and BPNN were accomplished using the gray level cooccurrence matrix (Haralick, 1973; Zhou et al., 2018b), and five features were extracted: contrast, energy, correlation, inverse gap and entropy. Four labels ("10", "20", "30" and "40") were used to represent the four levels of feeding intensity ("none", "weak", "medium" and "strong", respectively). The SVM used the radial basis function kernel function during the grading process. After the penalty coefficient c and the kernel parameter g of the SVM were selected and trained, 21 of the 80 samples were misclassified; the arrows in Fig. 7 indicate examples of SVM classification errors. The sample grading accuracy reached 73.75%. Fig. 8 shows the classification results of the BPNN. The initial

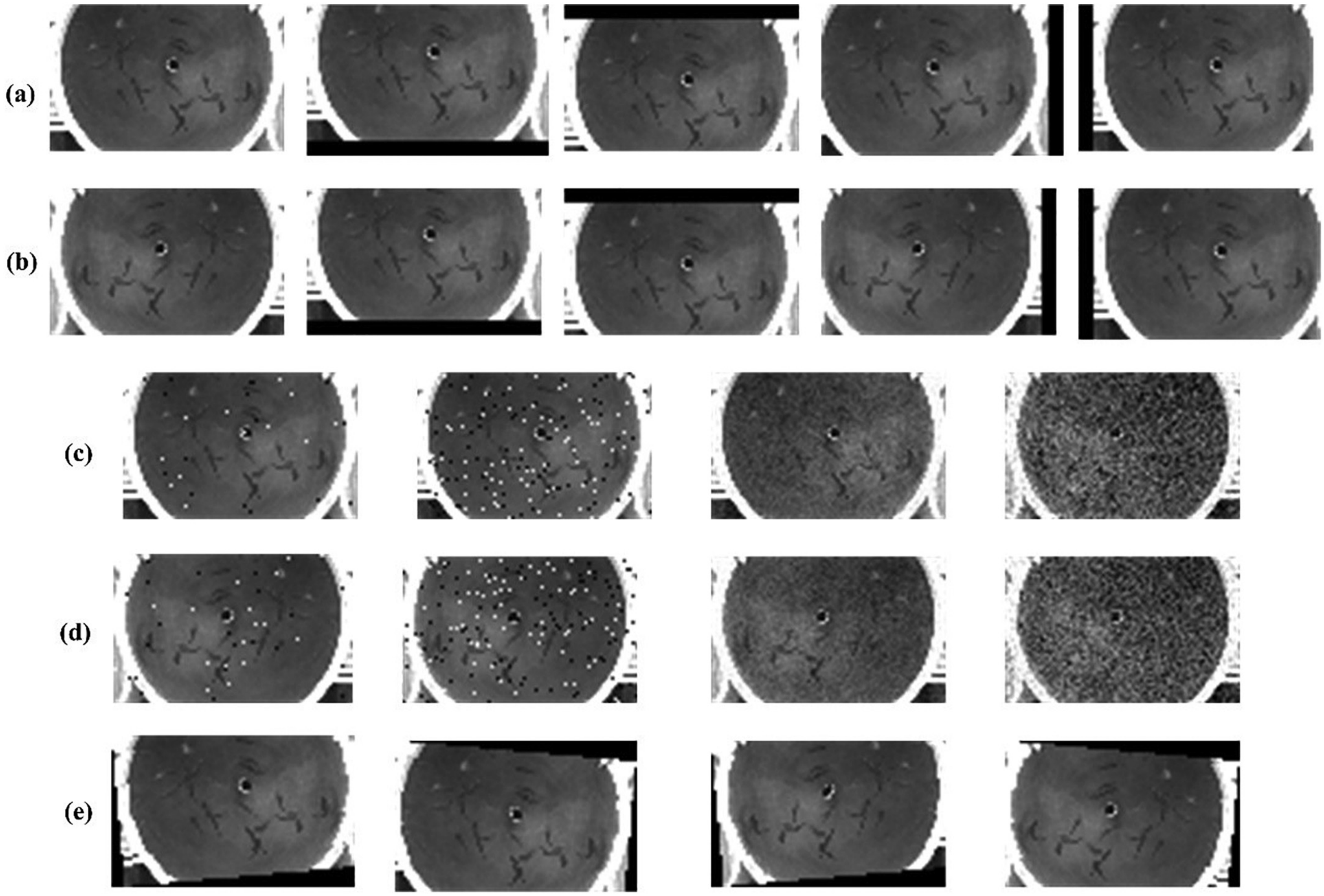


Fig. 4. Data augmentation: From left to right, the rows show (a) the original image and the original image shifted up, down, left and right by 5 pixels; (b) the horizontally flipped original image and the horizontally flipped image shifted up, down, left and right by 5 pixels; (c) the original picture with added salt-and-pepper noise (density: 0.01; density: 0.04) and Gaussian noise (variance: $10^2/255^2$; variance: 0.01); (d) the horizontally flipped image with added salt-and-pepper noise (density: 0.01; density: 0.04) and Gaussian noise (variance: $10^2/255^2$; variance: 0.01); and (e) the original image rotated 5 degrees to the left and right and the horizontally flipped image rotated 5 degrees to the left and right.

Table 2
The grading standards for fish feeding intensity.

Grading	Behavior
None	Fish do not respond to food
Weak	Fish eat only pellets that fall directly in front of them but do not move to take food
Medium	Fish move to take food, but return to their original positions
Strong	Fish move freely between food items and consume all the available food

Table 3
Grading result for each sample type.

Feeding intensity	Grading result				Precision/%	Recall/%	Accuracy/%
	no	weak	medium	strong			
none	19	0	1	0	95.00	95.00	90.00
weak	1	17	2	0	89.47	85.00	
medium	0	2	17	1	90.48	85.00	
strong	0	0	1	19	95.00	95.00	

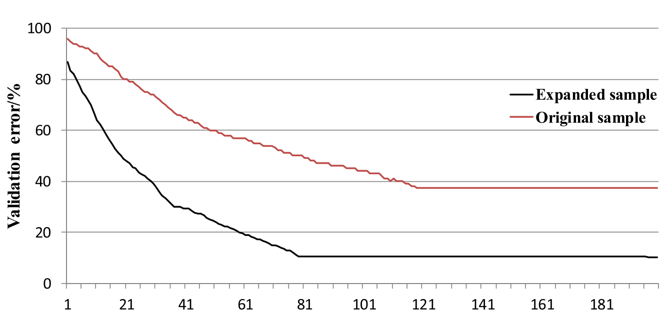


Fig. 5. Training results of the convolutional neural network.

parameters of the BPNN included 12 hidden layer nodes, a learning rate of 0.02, a mean squared error (MSE) performance function, and a minimum training error of e^{-6}; the training algorithm was the Levenberg-Marquart optimization algorithm. After training, the BPNN misclassified 15 of the 80 samples; the circles in Fig. 8 indicate examples of BPNN grading errors. The sample grading accuracy reached 81.25%. The results are shown in Table 4.

While the grading accuracies of all three abovementioned methods exceeded 70%, the CNN performed best. In general, the precision, recall and specificity values reflect the classification accuracy level, as shown in Table 4. In Fig. 9, the precision values for several grades were very high, indicating that the number of false positives was very low. Among these grades, those with less precision, such as “weak” and “medium”, indicate higher levels of false positives. These results occur because the algorithm incorrectly marks a large number of images as belonging to

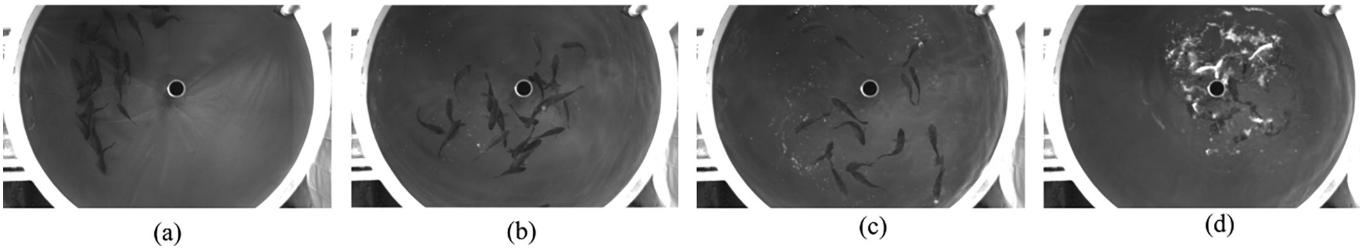


Fig. 6. Result images: (a) none; (b) weak; (c) medium; and (d) strong.

these classes. The unclear boundaries between adjacent feeding intensities leads to ambiguity when associating particular image categories with feeding intensity levels. As shown in Fig. 10, for nearly every model and each level, instances occurred in which images were incorrectly assigned to another level. In other words, nearly every model failed to classify a small portion of the feeding intensity images into the correct class. For example, the CNN model's grading results are shown in Table 3. Most of the false positive and false negative rates are similar between adjacent levels. Due to the commonalities between specific examples, the similarities between classes led to false positives and false negatives between the "weak" and "medium" feeding intensity levels, resulting in low average grading accuracies for these levels. This result is consistent with the abovementioned results: when the feeding intensity was "strong" or "none", the feature gap was obvious, and the classification accuracy was high.

In summary, the most important result of this study is that the grading accuracy of the proposed CNN reached 90%, exceeding the accuracy obtainable using other classic methods such as an SVM or BPNN (Fig. 9, Fig. 10 and Fig. 11). Table 4 shows the performances of the SVM and BPNN compared to that of our proposed method. The accuracies of the other two classifiers were significantly lower than that of our proposed method. This result occurred because when using the abovementioned methods to automatically identify the region of interest from an image, it is necessary to overcome two great environmental change challenges: visual distortion and image noise. When variability and distortion are present, the objects of interest in the image are expressed in a highly nonlinear input space, making it difficult to identify the objects of interest in the feature space efficiently (Larochelle et al., 2009). Because the SVM and BPNN use features extracted manually before grading, their classification accuracy is affected by the feature extraction. This effect could be corrected by using a nonlinear automatic learning system, such as a CNN multilayer neural network. The supervised training in a CNN ensures automatic learning of complex, nonlinear and discriminatory features (Siddiqui et al., 2018). In addition, because a CNN extracts the features of an image directly, without complex steps, it has strong adaptability and high accuracy. The experimental results also show that the CNN results have a high correlation with the 4-step scale method.

3.2.2. Comparison with quantitative feeding intensity assessment methods

To further verify the performance of the proposed method, we compared it with the method proposed by Zhou et al. (2018b), which achieved good results for making feeding decisions, resulting in a feed conversion rate improvement of 10.77% and reduced water pollution. This method used Delaunay triangulation and the gray level cooccurrence matrix to extract image features to represent fish feeding intensity and ultimately resulted in two fish feeding intensity indexes: FIFFB (Flocking index of fish feeding behavior) and SIFFB (Snatch intensity of fish feeding behavior). As a quantitative fish appetite evaluation method, the exact quantitative indicators can only reflect relative trends of fish feeding intensity. To classify fish appetite, appropriate thresholds must be selected. Therefore, the threshold selections affect the method's accuracy. In this study, by analyzing images of the fish feeding process, the thresholds for the four feeding intensity levels of FIFFB and SIFFB were set to 0.85, 0.57, 0.24 and 0.41, 0.19, and 0.09, respectively (Fig. 12).

After the thresholds were selected, their performance was tested using the 80 samples mentioned above. The results are shown in

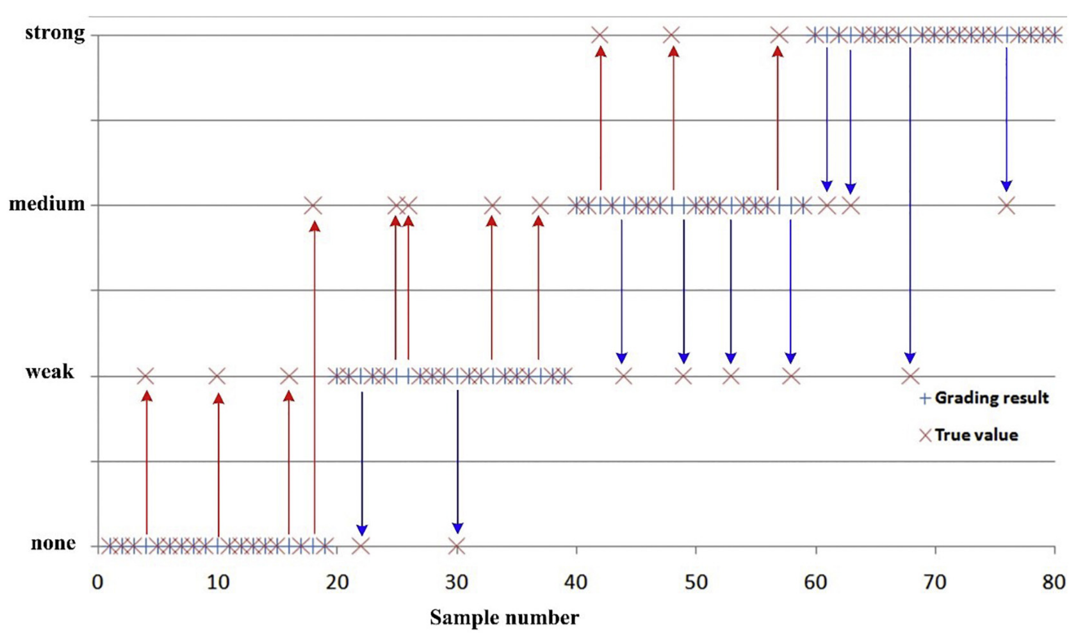


Fig. 7. Grading results of SVM.

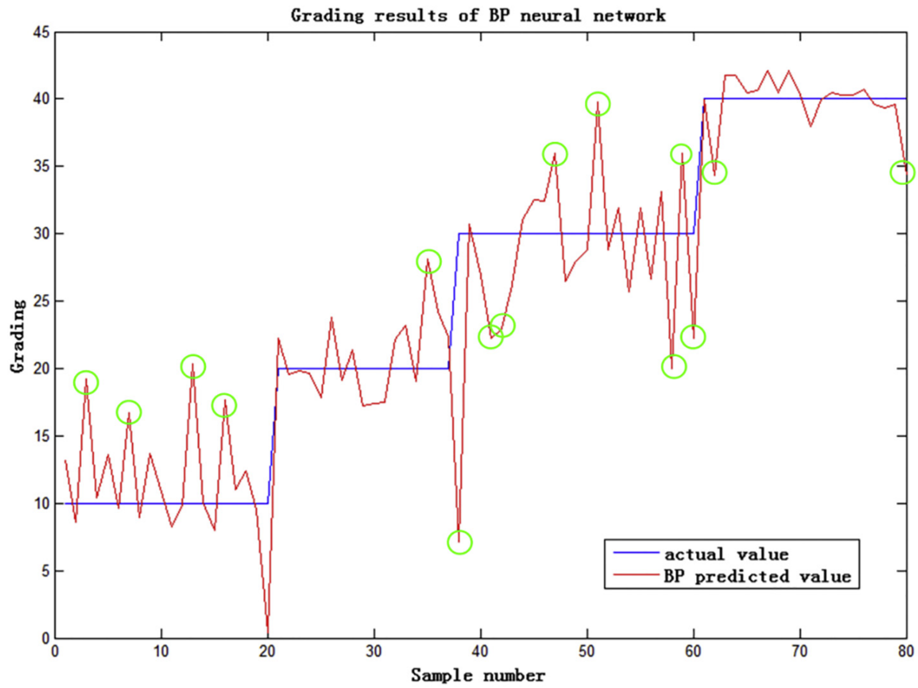


Fig. 8. Grading results of the BP neural network.

Table 4
Feeding intensity grading results of test samples.

Method	Total number	Identification number	Accuracy/%
SVM	80	59	73.75
BPNN	80	65	81.25
CNN	80	72	90.00

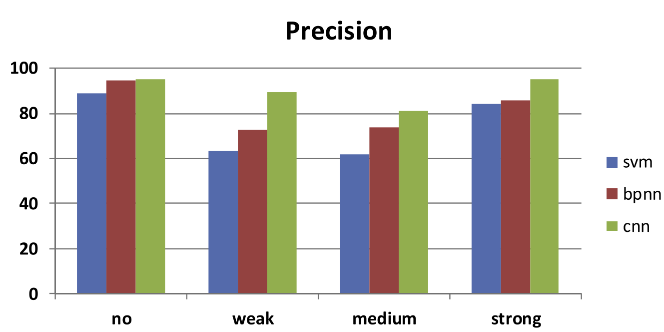


Fig. 9. Performance comparison in terms of precision between the different models for each feeding intensity.

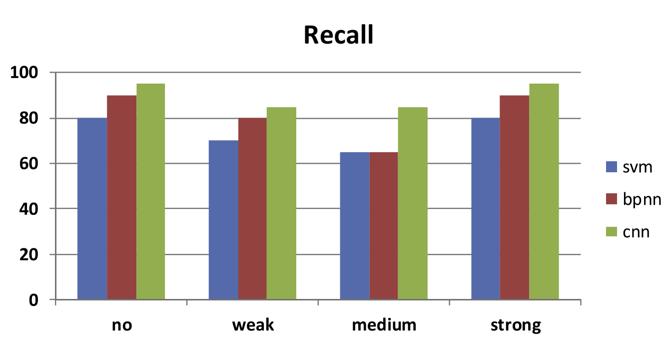


Fig. 10. Performance comparison in terms of recall between the different models for each feeding intensity.

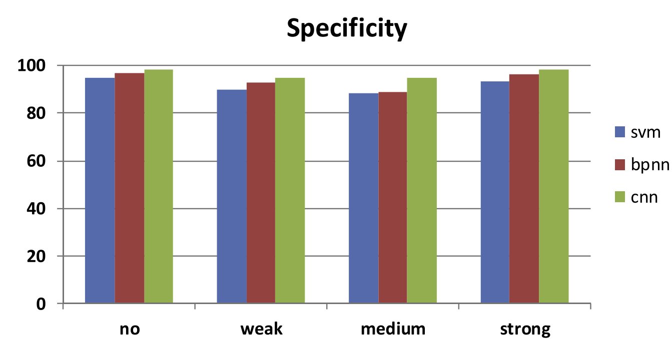
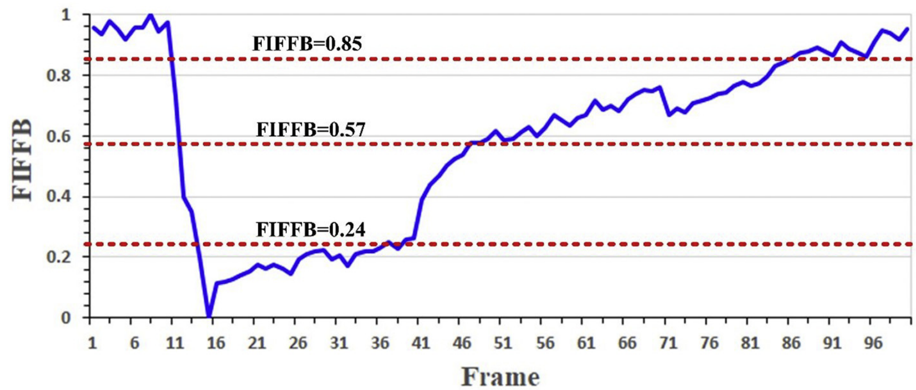


Fig. 11. Performance comparison in terms of specificity between the different models for each feeding intensity.

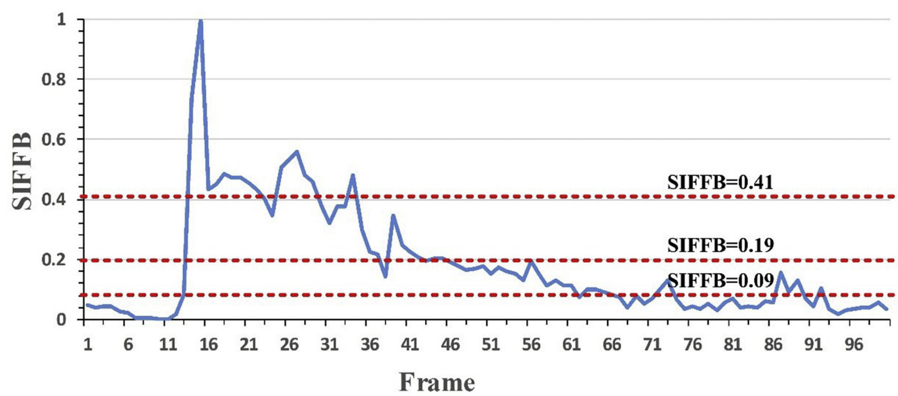
Table 5. Compared with FIFFB and SIFFB, the CNN still achieves the best accuracy.

In addition, extraction FIFFB and SIFFB requires image processing operations such as background generation and target extraction. The increased number of image processing steps not only leads to longer computational times but also increases the risk that errors will be introduced. However, the advantage of this method is that it accurately and objectively describes every moment of the entire feeding process and can comprehensively reflect fish feeding behavior.

Although the proposed method does not accurately quantify fish feeding intensity, it is both practical and easy to apply to production in practice compared with other methods that more accurately quantify the intensity of the feeding process (Liu et al., 2014; Zhao et al., 2017). The implementation process for the proposed method is simple; after the model was trained, the degree of fish hunger could be quickly ascertained. The proposed model also avoids having to manually extract complex features. Moreover, during the fish feeding process, some studies treated fish appetite as a fuzzy variable (Soto-Zarazúa et al., 2010; Zhao et al., 2019; Zhou et al., 2018b). In contrast, the proposed method is essentially intended to classify fish feeding intensity or hunger. In previous studies, most feeding control methods have been implemented by setting a specific threshold for the quantified feeding intensity, allowing judgments concerning the start and end of feeding.



(a) FIFFB



(b) SIFFB

Fig. 12. Threshold selection of FIFFB and SIFFB.

Table 5
Feeding intensity grading results of test samples using FIFFB and SIFFB.

Method	Total number	Identification number	Accuracy/%
FIFFB	80	69	86.25
SIFFB	80	67	83.75
CNN	80	72	90.00

Therefore, in some cases, there was essentially no significant difference between the precise quantification and grading of fish feeding intensity.

4. Conclusions

To solve the problems of low efficiency and low objectivity exhibited by methods that assess fish appetite, this study proposed a method for grading the degree of fish hunger to detect fish feeding intensity. In this study, a CNN used for deep learning proved to be an effective way to learn fish feeding intensity characteristics and apply them to new images. The most important result of this study is that the grading accuracy reached 90%. Moreover, because the method can extract features directly and does not require complicated steps such as target extraction, the adaptability of this method is strong, and its operation is simple.

After grading feeding intensity, by incorporating intelligent algorithms such as fuzzy control, it is possible to make decisions concerning when to stop feeding. The input of fuzzy information does not require precise variables; only fuzzy language variables are required. Therefore, more precise feeding control can be achieved. In addition, implementing this method in production practices could reduce labor costs and lay a theoretical foundation for guiding production practice. However, despite the dataset augmentation applied in this study, the training set was still limited due to various conditions—for example, the quantity and consistency of labeled samples. In future work, we plan to address this problem by designing training algorithms that are more suitable for small sample datasets.

Acknowledgments

The work was supported by the National Key Research and Development Program of China (2017YFD0701700) and the Beijing Excellent Talents Development Project (2017000057592G125).

References

Ahmad, S., Ahsan, J., Faisal, S., Ajmal, M., Mark, S., James, S., Euan, H., 2016. Fish species classification in unconstrained underwater environments based on deep learning. Limnol. Oceanogr. Methods 14, 570–585.
Atoum, Y., Srivastava, S., Liu, X., 2015. Automatic feeding control for dense aquaculture fish tanks. IEEE Signal Process. Lett. 22, 1089–1093.
Beyan, C., Katsageorgiou, V.M., Fisher, R.B., 2018. Extracting statistically significant behaviour from fish tracking data with and without large dataset cleaning. IET Comput. Vis. 12, 162–170.
Ding, W., Taylor, G., 2016. Automatic moth detection from trap images for pest management. Comput. Electron. Agric. 123, 17–28.
Duarte, S., Reig, L., Oca, J., 2009. Measurement of sole activity by digital image analysis. Aquac. Eng. 41, 22–27.
Eriksen, M.S., Færevik, G., Kittilsen, S., McCormick, M.I., Damsgård, B., Braithwaite, V.A., Braastad, B.O., Bakken, M., 2011. Stressed mothers – troubled offspring: a study of behavioural maternal effects in farmed Salmo salar. J. Fish Biol. 79, 575–586.
Haralick, R.M., 1973. Texture features for image classification. IEEE Trans. Syst. Man Cybern. 3, 610–621.
He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. 2016 IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR) 770–778.
Juell, J.-E., Westerberg, H., 1993. An ultrasonic telemetric system for automatic positioning of individual fish used to track Atlantic salmon (Salmo salar L.) in a sea cage.

Aquac. Eng. 12, 1–18.
Kolarevic, J., Aas-Hansen, Ø., Espmark, Å., Baeverfjord, G., Terjesen, B.F., Damsgård, B., 2016. The use of acoustic acceleration transmitter tags for monitoring of Atlantic salmon swimming activity in recirculating aquaculture systems (RAS). Aquac. Eng. 72-73, 30–39.
Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet classification with deep convolutional neural networks. In: Proceedings of the 25th International Conference on Neural Information Processing Systems. vol. 1. Curran Associates Inc., Lake Tahoe, Nevada, pp. 1097–1105.
Larochelle, H., Bengio, Y., Louradour, J., Lamblin, P., 2009. Exploring strategies for training deep neural networks. J. Mach. Learn. Res. 10, 1–40.
Lecun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998. Gradient-based learning applied to document recognition. Proc. IEEE 86, 2278–2324.
LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436.
Liu, Z., Li, X., Fan, L., Lu, H., Liu, L., Liu, Y., 2014. Measuring feeding activity of fish in RAS using computer vision. Aquac. Eng. 60, 20–27.
Mallekh, R., Lagardère, J.P., Eneau, J.P., Cloutour, C., 2003. An acoustic detector of turbot feeding activity. Aquaculture 221, 481–489.
Meng, L., Hirayama, T., Oyanagi, S., 2018. Underwater-drone with panoramic camera for automatic fish recognition based on deep learning. IEEE Access. 6, 17880–17886.
Moniruzzaman, M., Islam, S.M.S., Bennamoun, M., Lavery, P., 2017. Deep Learning on Underwater Marine Object Detection: A Survey. vol. 10617. pp. 150–160.
Overli, O., Sorensen, C., Nilsson, G.E., 2006. Behavioral indicators of stress-coping style in rainbow trout: do males and females react differently to novelty? Physiol. Behav. 87, 506–512.
Pautsina, A., Císař, P., Štys, D., Terjesen, B.F., Espmark, Å.M.O., 2015. Infrared reflection system for indoor 3D tracking of fish. Aquac. Eng. 69, 7–17.
Qin, H., Li, X., Liang, J., Peng, Y., Zhang, C., 2016. DeepFish: accurate underwater live fish recognition with a deep architecture. Neurocomputing 187, 49–58.
Rakowitz, G., Tušer, M., Říha, M., Jůza, T., Balk, H., Kubečka, J., 2012. Use of high-frequency imaging sonar (DIDSON) to observe fish behaviour towards a surface trawl. Fish. Res. 123-124, 37–48.
Rubio, V.C., Vivas, M., Sánchez-Mut, A., Sánchez-Vázquez, F.J., Covès, D., Dutto, G., Madrid, J.A., 2004. Self-feeding of European sea bass (Dicentrarchus labrax, L.) under laboratory and farming conditions using a string sensor. Aquaculture 233, 393–403.
Sa, I., Ge, Z., Dayoub, F., Upcroft, B., Perez, T., McCool, C., 2016. DeepFruits: A Fruit Detection System Using Deep Neural Networks. vol. 16. pp. 1222.
Saberioon, M., Gholizadeh, A., Cisar, P., Pautsina, A., Urban, J., 2017. Application of machine vision systems in aquaculture with emphasis on fish: state-of-the-art and key issues. Rev. Aquac. 9, 369–387.
Sadoul, B., Evouna Mengues, P., Friggens, N.C., Prunet, P., Colson, V., 2014. A new method for measuring group behaviours of fish shoals from recorded videos taken in near aquaculture conditions. Aquaculture 430, 179–187.
Siddiqui, S.A., Salman, A., Malik, M.I., Shafait, F., Mian, A., Shortis, M.R., Harvey, E.S., 2017. Automatic fish species classification in underwater videos: exploiting pre-trained deep neural network models to compensate for limited labelled data. ICES J. Mar. Sci. 75, 374–389.
Soto-Zarazúa, G.M., Rico-García, E., Ocampo, R., Guevara-González, R.G., Herrera-Ruiz, G., 2010. Fuzzy-logic-based feeder system for intensive tilapia production (Oreochromis niloticus). Aquacult. Int. 18, 379–391.
Sun, M., Hassan, S.G., Li, D., 2016. Models for estimating feed intake in aquaculture: a review. Comput. Electron. Agric. 127, 425–438.
Szegedy, C., Wei, L., Yangqing, J., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going deeper with convolutions. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1–9.
Ta, X., Wei, Y., 2018. Research on a dissolved oxygen prediction method for recirculating aquaculture systems based on a convolution neural network. Comput. Electron. Agric. 145, 302–310.
Wang, S.H., Zhao, J.W., Chen, Y.Q., 2017. Robust tracking of fish schools using CNN for head identification. Multimed. Tools Appl. 76, 23679–23697.
Wishkerman, A., Boglino, A., Darias, M.J., Andree, K.B., Estévez, A., Gisbert, E., 2016. Image analysis-based classification of pigmentation patterns in fish: a case study of pseudo-albinism in Senegalese sole. Aquaculture 464, 303–308.
Xu, Z., Cheng, X.E., 2017. Zebrafish tracking using convolutional neural networks. Sci. Rep. 7, 42815.
Ye, Z., Zhao, J., Han, Z., Zhu, S., Li, J., Lu, H., Ruan, Y., 2016. Behavioral characteristics and statistics-based imaging techniques in the assessment and optimization of Tilapia feeding in a recirculating aquaculture system. Trans. ASABE 59, 345–355.
Yukinori, M., Rosli, Muhammad Khairulanwar, Tan, N.H., Chung, F.L., 2016. Demand feeding system using an infrared light sensor for brown-marbled grouper juveniles, epinephelus fuscoguttatus. Sains Malays 45, 729–733.
Zhao, J., Bao, W.J., Zhang, F.D., Ye, Z.Y., Liu, Y., Shen, M.W., Zhu, S.M., 2017. Assessing appetite of the swimming fish based on spontaneous collective behaviors in a recirculating aquaculture system. Aquac. Eng. 78, 196–204.
Zhao, J., Li, Y., Zhang, F., Zhu, S., Liu, Y., Lu, H., Ye, Z., 2018a. Semi-supervised learning-based live fish inentification in aquaculture using modified deep convolutional generative adversarial networks. Trans. ASABE 61, 699–710.
Zhao, J., Bao, W., Zhang, F., Zhu, S., Liu, Y., Lu, H., Shen, M., Ye, Z., 2018b. Modified motion influence map and recurrent neural network-based monitoring of the local unusual behaviors for fish school in intensive aquaculture. Aquaculture 493, 165–175.
Zhao, S., Ding, W., Zhao, S., Gu, J., 2019. Adaptive neural fuzzy inference system for feeding decision-making of grass carp (Ctenopharyngodon idellus) in outdoor intensive culturing ponds. Aquaculture 498, 28–36.
Zhou, C., Zhang, B., Lin, K., Xu, D., Chen, C., Yang, X., Sun, C., 2017. Near-infrared imaging to quantify the feeding behavior of fish in aquaculture. Comput. Electron. Agric. 135, 233–241.
Zhou, C., Xu, D., Lin, K., Sun, C., Yang, X., 2018a. Intelligent feeding control methods in aquaculture with an emphasis on fish: a review. Rev. Aquac. 10, 975–993.
Zhou, C., Lin, K., Xu, D., Chen, L., Guo, Q., Sun, C., Yang, X., 2018b. Near infrared computer vision and neuro-fuzzy model-based feeding decision system for fish in aquaculture. Comput. Electron. Agric. 146, 114–124.
Zhou, C., Sun, C., Lin, K., Xu, D., Guo, Q., Chen, L., Yang, X., 2018c. Handling water reflections for computer vision in aquaculture. Trans. ASABE 61, 469–479.
Zion, B., Barki, A., 2012. Ranching fish using acoustic conditioning: has it reached a dead end? Aquaculture. 344-349, 3–11.