Comparison of Convolutional Neural Networks in Real-Time Monitoring of Aquaculture Water State

Hanxiang Qin\textsuperscript{1,2}, Siyuan Mei\textsuperscript{1,2}, Huihui Yu\textsuperscript{3}, Yeqi Liu\textsuperscript{1,2}, Ling Yang\textsuperscript{1,2}, Yupeng Mei\textsuperscript{1,2} and Yingyi Chen\textsuperscript{1,2}

\textsuperscript{1} College of 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} School of Information Science and Technology, Beijing Forestry University, Beijing, 100083, China
Email: chenyangyi@cau.edu.cn

Abstract. Real-time monitoring of the aquaculture water state is essential for shrimp aquaculture, and it plays a vital role in the growth of shrimp. The main purpose of this research is to find an effective method to achieve rapid detection and classification with water state images. We obtained a water state image dataset containing four categories in the laboratory, and gained the corresponding Hough datasets through Hough transform, including blue-green and turbid water, reddish brown and turbid water, light green and clear or tawny and clear water, colorless and transparent water. We compare the performance and spatio-temporal complexity of 12 widely used convolutional neural network models in water state image classification tasks, and then optimize the best performing model on the optimizer and hyperparameters. The experimental results show that the InceptionResNet model has the best effect and the InceptionV3 model has the smallest spatio-temporal complexity, while both models realize the task of water state image classification excellently.

Keywords. Deep learning; convolutional neural network; computer vision; aquaculture.

1. Introduction

Shrimp is one of the most important species in aquaculture industry. It is not only delicious in meat, but also rich in trace elements that are beneficial to the human. It has high nutritional value so that it is welcomed by many consumers. In the process of breeding, it is the consensus of the shrimp farmers to “Raise water before raising shrimp”. The state of aquaculture water directly affects the success of shrimp breeding and it is a crucial indicator for evaluating the ecological balance of aquaculture water \cite{1,2}.

Generally, farmers can judge the state of aquaculture water by observing the water color and water transparency. The type and quantity of plankton in the water is the main reason for the reflection of the water state. When the water is blue-green and turbid, the concentration of Cyanobacteria in the water increases, resulting in turbid water and low transparency of the water \cite{3}, which may easily lead to a large number of deaths of shrimp; when the water is reddish brown and turbid, red tide organisms like Dinophyta rapidly breed and lots of dead algae appear on the water surface, causing serious water pollution to the aquaculture water \cite{4}; when the water is colorless and transparent, the organisms in the aquaculture water become low and the amount of fertilizer applied to the water is insufficient,
which is not conducive to the growth of shrimp. Oppositely, when the water is brown or light green and clear, the beneficial algae like Bacillariophyta or Chlorophyta grow stably and absorb excess nitrogen fertilizer in the aquaculture water, the aquaculture water is in an ideal state as it has high biodiversity to maintain better ecosystem functions, improve productivity and nutrient retention [5, 6], which make the shrimp grow fast and dynamic.

Nowadays, computer vision (CV) technology has developed rapidly. It became one of the most successful application territories of artificial intelligence (AI) and has been used in water state detection [7]. Mullins et al. [8] proposed a turbidity detection method based on image processing. The average light intensity in the image was selected as the feature and the model relationship with the camera simulated turbidity was established to realize the turbidity of domestic and industrial wastewater. Hamidi [9], Karnawat [10], Li’ang [11] converted the collected RGB images into grayscale images for the color feature extraction, using grayscale histograms, gray-level co-occurrence matrix and other methods, and established relationship models to realize the turbidity detection of different water state at low cost. Chai et al. [12] proposed a method to detect the number of circular rings in different water state based on the image shape, and used the number of circular rings as features to establish a relationship model with turbidity to achieve turbidity detect.

Although these researches have achieved results, there are still some limitations in these detection methods [13]. The traditional machine learning (ML) method is adopted in the water state image processing and the feature extraction is performed manually. These methods depend on the extracted feature extremely and have many shortcomings such as complex process and poor performance, which greatly limits the accuracy and generalization of models.

In recent years, with the popularization of high-performance hardware and the production of big data, deep learning (DL) technique has achieved rapid development in theory and application, especially in water state prediction tasks [14], which can solve more breeding problems in real scenarios [15, 16]. However, these researches mainly focus on the analysis of sensor data such as dissolved oxygen and water temperature. There is still a lack of a method based on the combination of CV and DL for efficient and effective detection of water state.

In order to overcome these challenges in the detection of water state, we compare 12 widely used CNN models in the water state image. Firstly, the water state image dataset called original dataset is acquired in the laboratory conditions, and then we build the Hough dataset as comparison using Hough transform. Then, we utilize a series of CNN models and compare the performance and spatio-temporal complexity in the original dataset and Hough dataset respectively. Finally, we explore the optimizing strategies of the best performing model, including optimizers and hyperparameters. We summarized the performance of different models on water state image classification task in both image dataset. The experimental results show that the method based on DL and CV can solve the practical problems for rapid detection and real-time monitoring of water state.

Conclusively, the main contributions of this paper are described as follows:

1. We create two datasets for the monitoring of water state, one is original dataset and the other is Hough dataset, each of them contains 4 categories.
2. We compare the performance and spatio-temporal complexity of 12 classical CNN models on the image classification task in two datasets.
3. We analyze the advantages and disadvantages of these models, and then optimize for the best performing model. Through experiments, we have proved that CNN models can perform outstanding tasks in real-time monitoring of water state.

The rest of this paper is organized as follows: Section 2 describes the datasets and models. In Section 3, we compare the performance of many methods in original dataset and Hough dataset, we also analyze the effectiveness of hyperparameter settings on the experimental results. Finally, we give the conclusion of this paper in Section 4.
2. Materials and Methods

2.1. Aquaculture Water State Image Datasets
In this study, we obtain four different types of water state images in the laboratory: blue-green and turbid, reddish brown and turbid, light green and clear or tawny and clear, colorless and transparent, and number them respectively as I, II, III and IV, which represent the four types of water state respectively: the concentration of Cyanobacteria is too large, the concentration of Dinophyta is too large, the nutrients in the water are suitable and the nutrients in the water are few.

2.1.1. Image Acquisition. In order to obtain images of different water state, we used equipment such as beaker, camera and Secchi disk to build a simple device for image acquisition, as shown in figure 1. When acquiring images, we install the camera and Secchi disk at the fixed position in the device, pour the same volume of water samples into the beaker, and then use the camera to collect images. The image resolution is 640×480 pixels and the image format is JPG. The dataset collected in the experiment includes 410 pictures in 4 categories. Details are described in table 1 and examples are shown in figure 2.

![Figure 1. Datasets are acquired using the image acquisition device in the laboratory conditions.](image)

2.1.2. Data Preprocessing. Since the region of interest (ROI) in the image is a circular, we adopt an improved circular gradient symmetrical Hough transform algorithm [17] to realize the detection and segmentation of ROI, and creates the Hough dataset used for comparison with the original dataset. Then we use a variety of strategies to expand the original dataset and Hough dataset, such as horizontal flipping, vertical flipping, random rotation and so on. We also uniformly transform the image, including centering the image, adjusting the image to 224×224 pixels, standardizing the image and so forth. We train a large amount of classification models for the two datasets.

| Label | Category | Image Size | Number of datasets | Description |
|-------|----------|------------|--------------------|-------------|
| I     | blue-green and turbid | 224×224 | 110 | The concentration of Cyanobacteria in the water is excessive. |
| II    | reddish brown and turbid | 224×224 | 103 | The concentration of Dinophyta in the water is excessive. |
| III   | light green and clear or tawny and clear | 224×224 | 101 | The concentration of Bacillariophyta or Chlorophyta in the water is suitable |
| IV    | colorless and transparent | 224×224 | 96 | The concentration of microorganism in the water is insufficient |
2.2. Models
CNN is a groundbreaking neural network, which has achieved great success in the field of CV. Since AlexNet proposed by Alex et al. [18] achieved amazing results in the LSVRC-2012 ImageNet competition, CNN has been widely used in various image-related tasks. In this article, we compare several classic model structures such as VGGNet [19], ResNet [20], Inception [21-24] and ResNeXt [25] in the application of water state image classification and figure 3 shows the network structure of several methods.

2.3. Hardware and Software Platform
The hardware platform we use is the server including i7-8700K CPU, a single NVIDIA GeForce GTX 1080Ti GPU and 8G of memory. All experiments are run in the environment of Python3.7 and Pytorch1.2.

3. Experiment
First, we compare several representative deep learning models in order to prove the effectiveness of CNN in the classification task of aquaculture water quality. Secondly, we evaluate the different training strategies for the single network so that we can get a more suitable model. Finally, we adjust the hyper-parameters of the model to obtain the best performing deep convolution model and its configuration. In the experiment, we divided the dataset into training set and test set, the proportions were 80% and 20% respectively. All models use the same parameter settings to ensure that the experiment is fair. We uniformly use batch normalization (BN) layer and ReLU activation functions after each convolution operation.

3.1. Comparison of Model Performance
We compare the performance of 12 CNNs on the original dataset and the Hough dataset, including VGG16, InceptionV1, InceptionV3, InceptionV4, InceptionResNet, ResNet18, ResNet50, ResNet101, ResNeXt34, ResNeXt50, ResNeXt101 and ResNeXt152. Based on the classic model, we modified the input of all models to 224×224 pixels, and modified the output of the corresponding part of the model. We use the standard stochastic gradient descent (SGD) algorithm as the optimizer and cross-entropy loss as the loss function to train the model. We set the Batchsize to 8, the learning rate to 10-4, and perform 100 rounds of model training. The accuracy of the model on different dataset is shown is figure 4. From figure 4a, we can observe that the classification accuracy on the original dataset is 97.56% for ResNet101, 98.78% for InceptionV4, and 100% for other models. From figure 4b, we can observe likewise that the classification accuracy on Hough dataset is 97.56% for ResNet101 and 100% for other models. The classification loss of the model on different dataset is shown in figure 5. Through figures 4 and 5, we observe that the InceptionResNet model has better performance in the
original dataset and Hough dataset. Compared with other models, it has a faster convergence speed and strong stability, and has a better effect on the quality classification task of aquaculture water.

Figure 3. Illustration of images classification with different networks: (a) InceptionV3 network (b) ResNet50 network (c) InceptionResNet network.

Figure 4. Classification accuracy of different models: (a) represents the accuracy of the model on the original dataset, (b) represents the accuracy of the model on the Hough dataset.

Figure 5. Classification loss of different models: (a) represents the loss of the model on the original dataset, (b) represents the loss of the model on the Hough dataset.
In order to show the role of models in the training process, we use class activation map (CAM) to visualize the image. CAM is a two-dimensional score grid related to the output category. For the output feature of a convolutional layer, each channel in the feature map is weighted with a gradient, which indicates the importance of each position to the category. The highlighted area shows that the network is looking at the correct location when making classification judgment. In this article, we use the InceptionResNet model to generate the CAM of the original dataset and the Hough dataset. From figure 6, we can see that for the images of the two datasets, the CAM highlights are mainly concentrated in the liquid circle area. Furthermore, the highlight area of the Hough dataset image is more concentrated than the original dataset image, and the circular area is better used as the classification standard, so its classification performance is better.

![Figure 6. The visualized images with CAM. The first line represents the visualization of the original dataset and the second line represents the Hough dataset. (a) represents the input initial image, (b) represents the CAM, and (c) represents the superimposition of the CAM with initial image.](image)

3.2. Comparison of Spatio-Temporal Complexity

We compared the spatio-temporal complexity of the model, including the parameters, memory, MAdd and Flops. The Parameters is the sum of the number of parameters in each layer of the model, which represents the space complexity of the model, model memory represents the memory without any optimizer. MAdd represent the number of multiply-adds and Flops represents floating point Operations. MAdd and Flops can represent the amount of model calculations, and they are often used to evaluate the temporal complexity of the model.

Table 2 lists the spatio-temporal complexity of 12 models and we visualize these data in order to facilitate observation and analysis. As shown in figure 7, we use the parameters as the X axis and Flops as the Y axis to compare the spatio-temporal complexity of 11 models of 4 types (the VGG16 model is not shown in figure 7 as its parameters and Flops are much larger than other models). We can see that the series of ResNet is better than other three series. However, the spatio-temporal complexity of InceptionV1 and InceptionV3 is less than any model in the series of ResNet, which have better computing efficiency under the condition of limited resources. It is worth mentioning that InceptionResNet, which has the best model performance in the previous section, also has satisfactory spatio-temporal complexity.

We visualized the memory and MAdd in figure 8. We can easily observe that the spatial complexity of VGG16 model is much bigger than others due to the extensive use of fully connected
layers, while the memory and MAdd of the models will continue to increase as the networks deepen. Comparing table 2, figures 7 and 8, we can find that InceptionV1 and InceptionV3 have low spatio-temporal complexity so that they are suitable for deployment in lightweight devices such as mobile terminals. Further, according to comparison of their performance in figure 4, we infer that the InceptionV3 is more befitting, although the spatio-temporal complexity of InceptionV1 is slightly less than that of InceptionV3. The model performance of InceptionV3 has stronger stability and it is completely acceptable to sacrifice a little spatio-temporal complexity for more stable performance. Considering all factors, choosing the InceptionV3 model can get the best results in the case of limited hardware conditions.

| No | Model          | Parameters (M) | Memory (M) | MAdd (G) | Flops (G) |
|----|----------------|----------------|------------|----------|-----------|
| 1  | VGG16          | 131.95         | 109.3      | 30.96    | 15.5      |
| 2  | ResNet18       | 10.66          | 30.72      | 3.64     | 1.82      |
| 3  | ResNet50       | 22.43          | 137.53     | 8.22     | 4.12      |
| 4  | ResNet101      | 40.54          | 209.12     | 15.66    | 7.85      |
| 5  | InceptionV1    | 5.52           | 35.91      | 3.08     | 1.54      |
| 6  | InceptionV3    | 7.4            | 48.57      | 3.54     | 1.78      |
| 7  | InceptionV4    | 39.27          | 122.43     | 16.36    | 8.18      |
| 8  | InceptionResNet| 20.05          | 97.25      | 8.2      | 4.11      |
| 9  | ResNeXt34      | 39.27          | 93.79      | 18.06    | 9.04      |
| 10 | ResNeXt50      | 21.96          | 165.57     | 8.53     | 4.29      |
| 11 | ResNeXt101     | 40.25          | 243.67     | 16.04    | 8.05      |
| 12 | ResNeXt152     | 55.32          | 340.14     | 23.56    | 11.83     |

3.3. Optimization of Model

We explored the effects of different optimizers on the accuracy and robustness of the model on the original dataset and the Hough dataset. The optimizer can perform model optimization and accelerate model convergence. Common optimizers include SGD, ASGD, Rprop, RMSprop, Adam and Adagrad. We used the well-performing InceptionResNet in the previous section as the model for 100 rounds of training. As in the previous section, we still use the cross-entropy loss as the loss function and set the Batchsize to 8 and learning rate to 10-4. The performance of different optimizers is shown in figure 9. Due to better stability for the two datasets, the InceptionResNet model with Adagrad optimizer has the best performance.
Figure 9. The classification accuracy of the InceptionResNet model with different optimizers: (a) represents the accuracy of different optimizers in the original dataset, (b) represents the accuracy of different optimizers in the Hough dataset.

We also tried to train the InceptionResNet model with Adagrad using different hyper-parameter. We respectively set the Batchsize to 4, 8, 16 and set the learning rate to 10^-4, 5×10^-5, 10^-5. The test set curve is shown as figure 10. We use different colors to represent different Batchsizes and different curve forms to represent different learning rates. For example, we use red line when Batchsize is 4, use blue line when Batchsize is 8 and use green line when Batchsize is 16. Meanwhile, the line is solid and point is dot when the learning rate is 10^-4, the line is dashed and the point is boxes when the learning rate is 5×10^-5, the line is dots and the point is triangle when the learning rate is 10^-5. When the Batchsize is the same, the model chooses the learning rate of 10^-4 to have the best performance, and the model chooses 5×10^-5 has greater fluctuations. Similarly, when the learning rate is the same, the model chooses the Batchsize of 4 or 8 performs better. Further, the Batchsize of 8 is better than the Batchsize of 4, and the Batchsize of 16 has a larger curve fluctuation, and the performance of the model is significantly lower than other models.

In summary, through multiple sets of comparative experiments, we can find that when using InceptionResNet as a CNN structure, choosing Adagrad as the optimizer, setting the Batchsize to 8 and learning rate of 10^-4 can make the model have better performance and strong robustness.

Figure 10. The classification accuracy of the InceptionResNet model with different hyper-parameters: (a) represents the accuracy of different hyper-parameters in the original dataset, (b) represents the accuracy of different hyper-parameters in the Hough dataset.

4. Conclusions

Although deep learning algorithms have been widely used in water state prediction, the use of computer vision technology for real-time monitoring is the development trend of evaluation the state of aquaculture water. Deep learning methods can automatically extract the characteristics of aquaculture water state image and it is a feasible method to recognize and classify the water state images based on CNN. This paper compares the performance of widely used CNN on the water state image dataset, and gives the optimal network structure and parameters for the collected image dataset.
In our experiment, we tested 12 commonly used CNN models. In terms of the model complexity, InceptionV3 has a small number of parameters and calculations and can be used in lightweight devices. In terms of the model performance and stability, InceptionResNet is more suitable. Therefore, we choose InceptionResNet as the network structure, and further studied the influence of different training strategies and hyper-parameters on the performance of InceptionResNet in the task. The results shows that when InceptionResNet uses Adagrad as the optimizer and sets the Batchsize to 8 and the learning rate to 0.0001, it has the advantages of fast convergence and stable performance compared with other training strategies and hyper-parameters. With the continuous development of agricultural Internet of Things (IoI) technology, the classification model can be embedded in the IoI platform in the future to realize real-time monitoring and rapid classification of the aquaculture water quality.

Acknowledgments
This work was supported by Science and Technology Planning Project of Guangdong Province of China No.2017B010126001 and The National Key Research and Development Program of China No.2018YFD0700904-2.

References
[1] Correia E, Pereira J, Apolinárió M, Horowitz A and Horowitz S 2002 Effect of pond aging on natural food availability and growth of the freshwater prawn Macrobrachium Rosenbergii Aquac Eng 26 61-9.
[2] Brito L, Arantes R, Magnotti C, Derner R, Pchara F, Olivera A and Vinatae L 2014 Water quality and growth of Pacific white shrimp Litopenaeus vannamei (Boone) in co-culture with green seaweed Ulva lactuca (Linaeus) in intensive system Aquacult Int 22 497-508.
[3] Visser P, Verspagen J, Sandrini G, Stal L, Matthijs H, Davis T, Paerl H and Huisman J 2016 How rising CO2 and global warming may stimulate harmful cyanobacterial blooms Harmful Algae 54 145-59
[4] Amorim C and Moura A 2020 Ecological impacts of freshwater algal blooms on water quality, plankton biodiversity, structure, and ecosystem functioning Sci. Total. Environ 758 143605.
[5] Tilman D, Knops J, Wedin D, Reich P, Ritchie M and Siemann E 1997 The influence of functional diversity and composition on ecosystem processes Science 277 1300-02.
[6] Ptacnik R, Solimini A, Andersen T, Tamminen T, Brettum P, Lepisto L, Wille’n E and Rekolainen S 2008 Diversity predicts stability and resource use efficiency in natural phytoplankton communities Proc. Natl. Acad. Sci. U.S.A 105 5134-8.
[7] Duan Y, Li D, Li Z and Fu Z 2015 Review on visual characteristic measurement research of aquatic animals based on computer vision Transactions of the CSAE 31 1-11. (in Chinese)
[8] Mullins D, Coburn D, Hannon L, Jones E, Clifford E and Glavin M 2018 A novel image processing-based system for turbidity measurement in domestic and industrial waste-water Water, Sci. Technol. 77 1469-82.
[9] Hamidi F, Zainuddin M, Abbas Z and Ahmad A 2017 Low cost and simple procedure to determine water turbidity with image processing ICISPC 30-34.
[10] Karnawat V and Patil S 2016 Turbidity detection using image processing ICCCA.
[11] Li A 2012 Design and Implement of Turbidity Detection System Based on Image Recognition (Harbin: Harbin Institute of Technology Press) p 27.
[12] Chai M, Ng S and Chua H 2016 An alternative cost-effective image processing based sensor for continuous turbidity monitoring AIP Conference Proceedings.
[13] Liu Y, Chen Y and Fang X 2018 A review of turbidity detection based on computer vision IEEE Access 6 60586-604.
[14] Chen Y, Cheng Q, Cheng Y, Yang H and Yu H 2018 Applications of recurrent neural networks in environmental factor forecasting: A review Neural. Comput. 30 2855-81.
[15] Liu Y, Zhang Q, Song L and Chen Y 2019 Attention-based recurrent neural networks for accurate short-term and long-term dissolved oxygen prediction Comput. Electron. Agric 165 104964.

[16] Cao X, Ren N, Tian G, Fan Y and Duan Q 2021 A three-dimensional prediction method of dissolved oxygen in pond culture based on Attention-GRU-GBRT Comput. Electron. Agric. 181 105955.

[17] Cai J 2005 Research and Application of Relative Algorithms in Real-time Automatic Vision Inspection System (Hangzhou: Zhejiang University Press) p 72.

[18] Krizhevsky A, Sutskever I and Hinton G 2012 ImageNet classification with deep convolutional neural networks, Proceedings of the 25th ICNIPS pp 1097-105.

[19] Simonyan K and Zisserman A 2015 Convolutional networks for large-scale image recognition ICLR.

[20] He K, Zhang X, Ren S and Sun J 2016 Deep residual learning for image recognition CVPR.

[21] Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V and Rabinovich A 2014 Going deeper with convolutions CVPR.

[22] Ioffe S and Szegedy C 2015 Batch normalization: Accelerating deep network training by reducing internal covariate shift, Proceedings of the 32nd ICICML 448-56.

[23] Szegedy C, Vanhoucke V, Ioffe S and Shlens J 2016 Rethinking the inception architecture for computer vision CVPR 2818-28.

[24] Szegedy C, Ioffe S and Shlens J 2017 Inception-v4, Inception-ResNet and the impact of residual connections on learning AAAI 4278-84.

[25] Xie S, Girshick R, Dollar P, Tu Z and He K 2017 Aggregated residual transformations for deep neural networks CVPR pp 5987-95.