Research for improving identification accuracy of specific fish species with CNN

Yuta Shoji^{a,*}, Lifeng Zhang^a

^aFaculty of Engineering, Kyushu Institute of Technology,
1-1 Sensui-cho, Tobata-ku, Kitakyushu, Fukuoka, 804-8550, JAPAN

* Corresponding Author: p108053y@mail.kyutech.jp

Abstract

In recent years, the ecosystem collapses happened around in Japan due to the invasion of alien species and the increasing of the amount has become a serious environmental problem. Especially to the aquatic alien species, different with the terrestrial alien species, is difficult to survey because human have not an efficient visual capability to see through a long distance in water. Therefore, in order to easily conduct underwater surveys, this research aimed to develop a system that can automatically find specific fish using a camera. Although an existing convolutional neural network has performed fish identification, it is desirable that machines used underwater be power saving and low load. Therefore, this study aims at the accuracy improvement of the low power consumption, low load identification system.

keywords: CNN, Deep Learning, classification system

1. Introduction

The number of foreign-originated creatures living in the field in Japan is only about 2000. As a problem of alien species, ecosystems are established with an exquisite balance, but when organisms invade from here, they can have a wide range of adverse effects. It is difficult to survey because human have not an efficient visual capability to see through a long distance in water. Therefore, in order to easily conduct underwater surveys, this research aimed to develop a system that can automatically find specific fish using a camera. Although an existing convolutional neural network has performed fish identification, it is desirable that machines used underwater be power saving and low load. Therefore, this study aims at the accuracy improvement of the low power consumption, low load identification system. This is way-changed the number of nodes and the condition of classifica-

tion as a method. The rest of paper is organized as follows. Section 2 shows the principle of machine learning. Section 3 depicts and explains a flow of the processing to make an identification system. Section 4 shows the experiments, and their results discuss them. Finally, Section 5 concludes this paper and indicates the future work.

2. Principle

2.1 Convolutional Neural Network

When humans recognize images, the human visual cortex can also capture spatial features firmly. Furthermore, in human visual cortex, experiments have shown that multiple cortices perform image recognition by capturing structures with different levels of abstraction in each hierarchy. The first step is to gradually capture more complex higher-order information from unit elements in pixel units, and from there contours, colors, faces, and bodies. The structure of CNN is shown in Fig.1. Convolutional neural networks (CNNs) were invented with the help of human visual cortex. CNN is an innovative technology that has sparked the recent deep learning boom. It was originally invented to perform image recognition, but its effect is known to be effective for dealing with series data such as text processing, video, audio, etc. beyond genres.

2.1.1 Convolutional layer

The structure of convolutional layer is shown in Fig.2. CNN is divided as an intermediate layer into a feature extraction part of the convolutional layer and the pooling layer and a classification part by the total joint layer. In convolutional layer, you can capture spatial features by preparing a filter called kernel and sliding it from left to right. Performs a product-sum operation on kernel and input data to generate a

Fig. 1. CNN structure

Fig. 2. Convolutional layer

Fig. 3. Pooling layer

feature map.

2.1.2 Pooling layer

The structure of pooling layer is shown in Fig.3. The pooling layer shrinks the data by dividing the input data into smaller areas and performing Max pooling to extract the maximum value of each area or Average pooling to extract the average value. In addition to reducing computational costs to reduce data, you can build robust models against small position changes to ignore differences within each region.

2.1.3 Dropout

In this CNN, Dropout is applied. Dropout is a method for optimizing neural networks with deep hierarchy with high accuracy. When learning neural network, Dropout performs learning by invalidating some of the nodes in the layer at a certain update. In the next update, another node is invalidated and learning is repeated. This makes it possible to forcibly reduce the degree of freedom of the network at the time of learning to improve generalization performance and to avoid over-learning. In the layer, it is generally said that it is better to invalidate about 50

Fig. 4. Study flowchart

3. Structure

Fig.4 shows the flow of deep learning used in the research. In the learning stage, the prepared data set is preprocessed data augmentation and resize image. Next, deep learning using CNN is performed on the prepared data set, and the model and weight of the learned data are acquired. Give test data to the learned model and evaluate the identification accuracy. In this study, we examine differences in identification accuracy and power consumption by changing the input data, image size, or CNN model configuration.

Fig. 5. fish example

Fig. 6. augmentation example

4. Experiment

In this study, we conducted experiments by changing the configuration of three parameters of CNNs, and investigated the optimal condition that obtains a best performance tradeoff between high accuracy and less calculation time. Two hundred fish of each of four types were prepared from ImageNet as identification targets, and they were refilled to data augmentation of 1400 each. We performed deep learning with this model and evaluated the prediction system of test data. Three experiments were conducted this time. Fig.5 shows an example of classified fish and data augmentation image example is shown in Fig.6.

4.1 Experiment 1

In Experiment 1, the size of the input image is changed as 32×32 , 64×64 and 96×96 pixels. Experiment 1 example is shown in Fig.7. The size of the input image increases

Fig. 7. Experiment 1

the number of nodes in the input layer and increases the computation time of the neural network. Therefore, the size of the input image suitable for identification was verified.

4.2 Experiment 2

In Experiment 2, the number of nodes in all coupled layers was changed by 128, 256 and 512. In the all-connection layer, the data that has been convoluted and pooled in the hidden layer is classified according to the number of nodes in the all-connection layer, so the identification ability improves as the total connection layer increases, but the calculation time increases. Therefore, we verified the optimal number of nodes to identify.

4.3 Experiment 3

In the experiment 3, the type of identification object was changed in the model size optimized by the experiments 1 and 2. As the type of classification changes, the amount of features provided changes, so we investigated how it affects the accuracy.

5. Conclusion

In this study, we conducted experiments by changing the configuration of three parameters of CNNs, and investigated the optimal condition that obtains a best performance tradeoff between high accuracy and less calculation time. Simulation result shows, with the reduced number of nodes to 128×2 and keep the training image size of 64×64 , 91.77% identification accuracy could be obtained. It improves the identification accuracy of specific fish species and expect reducing the load

and power consumption to less 1/4 of the original one estimatedly. Such a result is good for underwater camera implementation. As future tasks, it is necessary to consider the implementation of a system that can identify the fish species of multiple fishes by combining accuracy improvement by aligning learning images and position detection.

References

- (1) Alex Krizhevsky, Ilya Sutskever and Geoffrey E. Hinton, *ImageNet classification with deep convolutional neural networks*, NIPS Proceedings, 2012
- (2) Google Researchers, *TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems*, 2016
- (3) Diederik P. Kingma, Jimmy Ba, *Adam: A Method for Stochastic Optimization*, 2014
- (4) Yann LeCun, Leon Bottou, Genevieve B. Orr, and Klaus-Robert Muller *Efficient BackProp*, 1998
- (5) Pierre Baldi and Peter J. Sadowski, *Understanding Dropout*, NIPS Proceedings, 2013
- (6) Ministry of the Environment "Countermeasures against alien species in Japan", <https://www.env.go.jp/nature/intro/2outline/list/Distribution/sa/PDF/L-sa-05.pdf>, Access Date: 2019/1/29
- (7) Handwritten Character Recognition Using Neural Network, http://nnadl-ja.github.io/nnadl_site_ja/chap1.html, Access Date: 2019/1/29