Efficient Data Acquisition and CNN Design for Fish Species Classification in Inland Waters

Jin-Hyun Park[1] and **Young-Kiu Choi**[2*], *Member*, *KIICE*

[1]Department of Mechatronics Engineering, Gyeongnam National Univ. of Science and Technology, Jinju 52725, Republic of Korea
[2]Department of Electrical Engineering, Pusan National University, Pusan 46241, Republic of Korea

Abstract

We propose appropriate criteria for obtaining fish species data and number of learning data, as well as for selecting the most appropriate convolutional neural network (CNN) to efficiently classify exotic invasive fish species for their extermination. The acquisition of large amounts of fish species data for CNN learning is subject to several constraints. To solve these problems, we acquired a large number of fish images for various fish species in a laboratory environment, rather than a natural environment. We then converted the obtained fish images into fish images acquired in different natural environments through simple image synthesis to obtain the image data of the fish species. We used the images of largemouth bass and bluegill captured at a pond as test data to confirm the effectiveness of the proposed method. In addition, to classify the exotic invasive fish species accurately, we evaluated the trained CNNs in terms of classification performance, processing time, and the number of data; consequently, we proposed a method to select the most effective CNN.

Index Terms: Convolutional neural network, Data acquisition, Efficient classification, Exotic invasive fish species

I. INTRODUCTION

The Convention on Biological Diversity (CBD), which came into force in 1993, aims at a broad international agreement to address the conservation of biodiversity on the planet, the sustainable use of its components, and the fair and equitable sharing of benefits arising out of the utilization of genetic resources [1]. Several species are being monitored in every country for fulfilling this purpose. In particular, the 10th Conference of the Parties to the Convention on Biological Diversity announced the prioritization of the "elimination of exotic invasive species" worldwide [1, 2]. In South Korea, the largemouth bass and bluegill species cause ecosystem disturbance in domestic waters. They are known to be the most important factors in reducing the population of native species in domestic freshwater. Therefore, an efficient and

reliable system needs to be developed for the elimination of these exotic invasive species.

Studies conducted on fish recognition so far have mainly been limited to the ground or restricted environments. However, despite the high demand for underwater object recognition, studies on data acquisition and automatic data processing have been lacking, due to the limitations of underwater environments [3]. Lee et al. [4] used fish contour matching to recognize fish in a fish tank, whereas Strachan et al. [5-7] used color and shape to recognize fish on a conveyor belt with 92% accuracy [7]. Recently, French et al. [8] have studied a system that simply recognizes and counts fish using a convolutional neural network (CNN). Qin et al. [3] used a CNN to classify fish species in the ocean. The fish images in the water have been designed with low resolution, due to the large amount of data obtained using a video system, and the

quality of the fish images is low, due to the extreme conditions in the water [9]. Therefore, it is difficult to obtain constant proportional data of several fish species from video images captured underwater, and the images require preprocessing to remove the background to improve the accuracy of the CNN. Thus, this study aims to design a system that is robust to the surrounding environment for devices classifying fish species in natural ecosystems, such as inland water and rivers. Accordingly, we consider the following problems. The first problem is the acquisition of the fish species data. Several studies performed performance evaluation using small datasets under limited constraints. This is because there are several limitations to obtaining the images of multiple species in a wide range of natural environments. Therefore, a system constructed under limited conditions cannot be used for the performance evaluation of fish images obtained in a wide range of natural environments. Second, the captured fish images differ owing to the visual effects of the depth and turbidity of the water [3]. In a natural environment, the light intensity changes frequently due to water and sunlight, and the field of view is limited according to the depth and turbidity of the water. Therefore, obtaining a large number of fish images in a natural environment is difficult. Third, no studies or data exist on the number of fish images and image clarity, to design a fish classification system. To solve these problems, we use our previously proposed method to construct a large number of fish images in different natural environments [10], and we verify the performance against numerous experimental fish images.

A learning system accumulates data from experience and learns by itself to improve its performance. Machine learning and deep learning have attracted increasing interest for several years, and they are being applied in various fields. Google Assistant, Apple Siri, and IBM Watson use machine learning to acquire knowledge from data, and to solve real problems [11]. In particular, deep learning is a method of machine learning in which images can be classified and learned directly into a network model, where the computer learns directly from the data [12-14]. Among them, CNN, which was proposed by LeCun [12] in the late 1990s and which mimics the visual processing process of humans and animals, has been successfully applied to the field of image

recognition [12-19].

In this paper, we present appropriate criteria for acquiring the fish species data, the classification performance by turbidity, and the number of learning data, and for selecting the most efficient CNN for the accurate classification of exotic invasive species for their elimination. The raw data of the fish images captured in a laboratory were composed of 5,000 [images/fish species] to overcome the difficulty of obtaining a large number of actual fish species images in a natural environment. These image data were converted into an image dataset similar to the images captured in a natural environment, which was then used in experiments. We also confirmed the effectiveness of the proposed method by using 1,000 [images/fish species] of largemouth bass and bluegill captured at a pond for several months. The CNNs evaluated in this study were AlexNet, the recently used VGGNet (VGG16, VGG19), and GoogLeNet [16-19]. We present a method to choose the most efficient CNN for the classification of exotic invasive species.

II. CONVOLUTIONAL NEURAL NETWORK

CNN, proposed by LeCun [12], is a deep learning algorithm that exhibits the best performance in the field of image recognition. It can perform recognition even when the size and position of objects are changed, by mimicking the visual processing process of an animal [12-14]. CNN is capable of re-learning the network to be applied to new recognition work through transfer learning, thereby enabling high accuracy and quick learning [15-19]. Fig. 1 shows the general hierarchical structure of CNN, which can be divided into a feature extraction layer and classification layer [12-14]. The feature extraction layer can be divided into a convolution layer, rectified linear unit (ReLU) layer, and pooling layer. The convolution layer activates image features by using a convolution filter on the input image. The ReLU layer eliminates the vanishing gradient of the network, which allows the network to learn quickly. The pooling layer performs a nonlinear downsampling of the image to simplify the output, and to reduce the number of network parameters. In addition, the classification layer consists of nodes connected
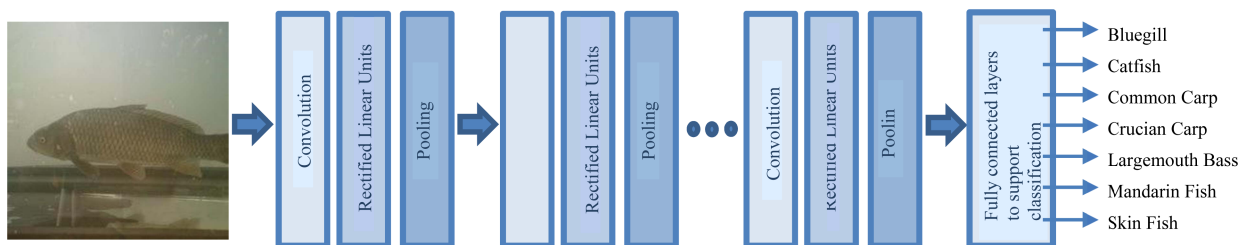

Fig. 1. CNN system

after passing through the feature detection layer. It is generally composed of a fully connected layer, and it provides the number of classifications that can be predicted by the network as a K-dimensional vector [12-14]. In this study, we evaluated the performance after transfer learning using Alex-Net, VGGNet (VGG16, VGG19), and GoogLeNet.

III. DATASET CONSTRUCTION

The datasets are composed of the basic dataset, data group I, data group II, and dataset Z. The basic dataset consists of the raw images captured in the laboratory. Data group I and Data group II consist of the learning data and the test data for the CNNs, respectively. Dataset Z is composed only of the test data for the performance evaluation of the CNNs.

A. Basic Dataset

The learning, validation, and test data of the CNNs for the classification of fish species are difficult to secure, because acquiring data using Internet images or video images requires a considerable amount of time and numerous copyright permissions [15]. This is because there are several limitations to obtaining a large number of fish images for different species in a wide range of natural environments. In a natural environment, the light intensity changes frequently due to water and sunlight, and the field of view is limited according to the depth and turbidity of the water. Therefore, obtaining a large number of fish images in a natural environment is difficult. Fig. 2(a) is an example of largemouth bass

images captured at the pond for several months. Distinguishing individual fish species by the influence of sunlight, illuminance, green algae, and moss is difficult. Therefore, an effective method is required to acquire a large number of fish species images in different natural environments, to solve these problems.

To acquire a large number of fish images for the same fish species, we recorded the videos of fish of specific species in the laboratory environment and used the captured video images as the basic data [15]. Seven individual fish species were tested for the fish classification. Two species were largemouth bass and bluegill, which are exotic invasive species, whereas the other five species were common carp, crucian carp, catfish, mandarin fish, and skin carp, which reside in rivers and inland waters, and are similar in size or shape to the exotic invasive species. We used 5,000 [images/fish species] data from the videos of the seven individual fish species as the basic data. Fig. 2(b) shows an example of the captured fish images.

B. Data Group I

The image of a largemouth bass shown in Fig. 2(a) captured at the pond is different from the image shown in Fig. 2(b) captured in the laboratory, due to the effects of sunlight, illumination, green algae, and moss. Eight underwater images were selected for realizing fish images close to the natural environment. The eight underwater images consist of deep underwater images and images of green algae and muddy water often observed in streams and inland water [10]. We created Data group I through a random synthesis of

(a) Largemouth bass images in a natural environment.

(b) Fish images of basic data

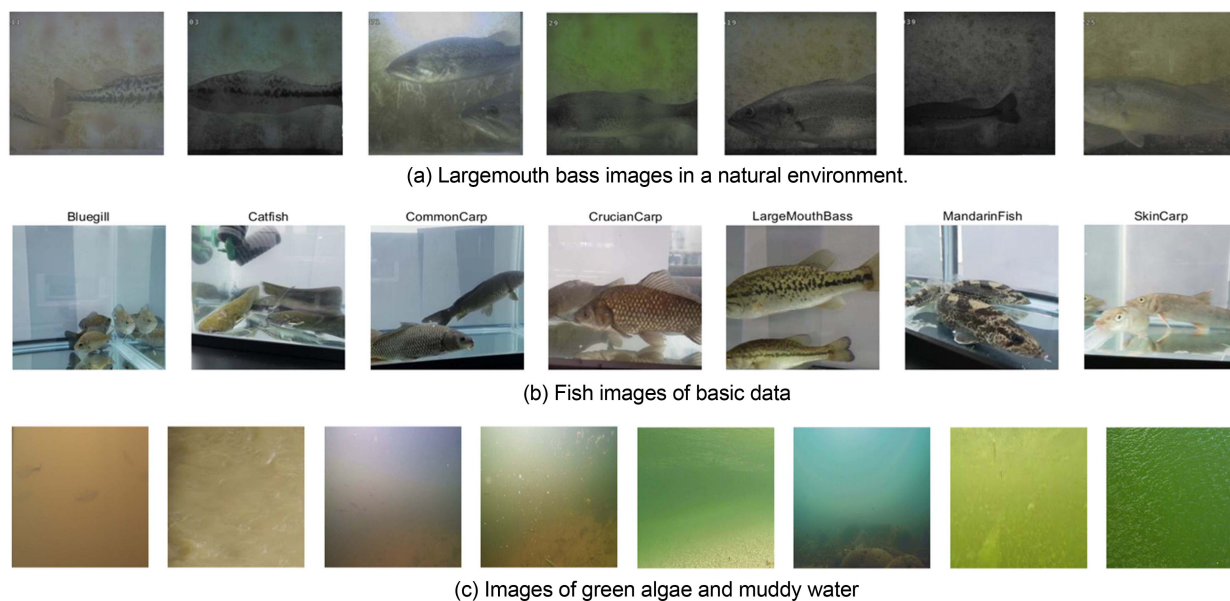(c) Images of green algae and muddy water

Fig. 2. Images of largemouth bass captured in a natural environment, fish images of basic data captured in a laboratory, and underwater images

the selected eight underwater images and 5,000 [images/fish species] fish images. Data group I consists of eight datasets named Datasets A–H.

To realize an image in a natural environment, which changes according to the change in luminous intensity due to sunlight, water depth, and turbidity, we created Datasets A–H, by mixing randomly selected underwater images and fish images in different ratios.

The image composition can be expressed simply as a linear combination of two input images pixels as in (1):

$$Image_{Comp} = (1 - x) \times Image_1 + x \times Image_2, \qquad (1)$$

where x s.t. $0 \leq x \leq 1$ is the image composition ratio, $Image_1$ and $Image_2$ are the two image pixel values to be synthesized, and $Image_{Comp}$ is the synthesized image pixel value.

Fig. 2(c) shows the eight underwater images used in this study. Table 1 shows the eight Datasets A–H and the proportions of the synthesized images.

Fig. 3 shows an example of images from Datasets A–H obtained by synthesizing the muddy water image of Fig. 2(c) for a bluegill. As evident from Fig. 3, it is possible to realize a fish image similar to that in a natural environment, which changes according to the change in luminous intensity due to

sunlight, and the depth and turbidity of the water. In particular, Datasets E–H are highly turbid, so that it is difficult for laypersons or even fish specialists to distinguish fish species.

C. Data Group II

Existing CNN studies have not considered the amount of data required for learning, validation, and testing. In general, the learning of a CNN is greatly affected by the number of data depending on the number of classification categories, the similarity of classification images, the size of images, and the sharpness of images. Therefore, the classification performance is known to improve as the number of images increases, but there are no studies on the appropriate number of data for learning.

The ImageNet Large Scale Visual Recognition Challenge (ILSVRC), an image contest held from 2010 to 2017, used 10,000 [image/class] learning data for 1,000 classifications [20]. The training data of the Modified National Institute of Standards and Technology (MNIST) database employed as a benchmark test for machine learning and deep learning used 6,000 [images/classes] [21], whereas the image training data of CIFAR-10 and CIFAR-100 built by the Canadian Institute for Advanced Research (CIFAR) [22] used 5,000 [images/classes] and 500 [images/classes], respectively. Therefore, we attempt to evaluate experimentally the number of fish images required to classify fish species in water.

Data group II consists of Datasets L–P constructed with learning, validation, and test data in a certain ratio from Datasets A–H. However, Datasets D–H randomly contain fish images that are difficult for humans to recognize. Assuming that the fish images to be classified are generally in their natural environment, only Datasets A–C are sufficient. Therefore, it is practical to construct datasets of fish images newly from the Datasets A–C at random. Table 2 shows the number of data in the newly constructed Datasets

Table 1. Data group I

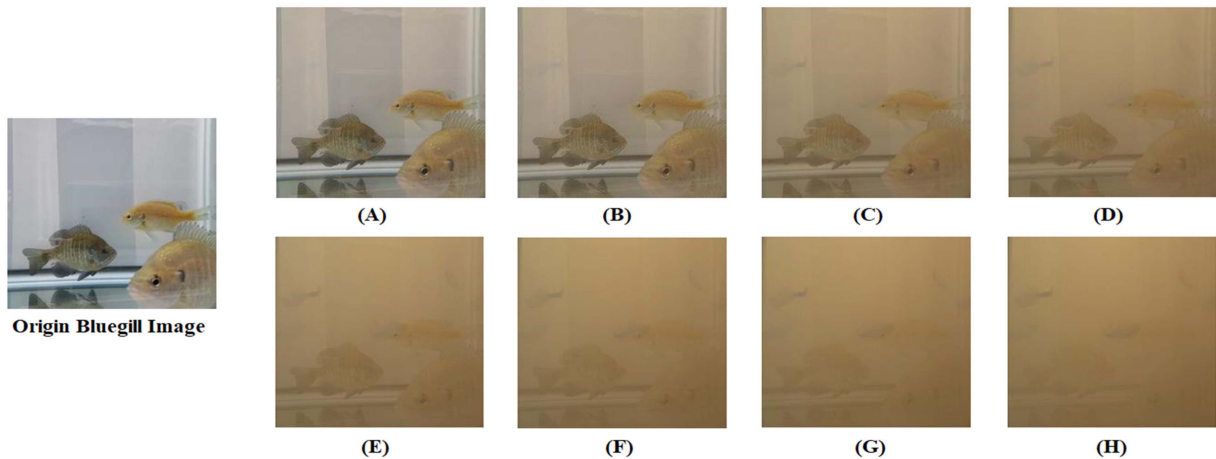| Dataset	Random underwater images (%)	Fish images (%)
Dataset A	30	70
Dataset B	50	50
Dataset C	70	30
Dataset D	80	20
Dataset E	85	15
Dataset F	90	10
Dataset G	92.5	7.5
Dataset H	95	5

Origin Bluegill Image

Fig. 3. Synthesis of muddy water image. (A)–(H) Datasets A–H.

Table 2. Data group II

Datasets	No. of training data	No. of validation data	No. of test data
Datasets L, Q	2,800	1,200	1,000
Datasets M, R	2,100	900	1,000
Datasets N, S	1,400	600	1,000
Datasets O, T	700	300	1,000
Datasets P, U	350	150	1,000

L–P and Q–U. Table 2 shows that the numbers of learning and validation data were selected differently, and the number of test data was maintained the same to ensure the reliability of the experiment, and to evaluate the classification performance. Data group II was created to evaluate the number of fish images and the effect of the turbidity of the synthesized images.

D. Dataset Z

The synthesized fish images are images that are very similar to, or that have been severely synthesized to be similar to those captured in their natural environment. We believe that these images could be replaced by fish images captured in a wide range of natural environments, as in Fig. 2(a). Therefore, the performances of the CNNs learned with Datasets L–P and Q–U were evaluated using Dataset Z as the test data.

Dataset Z was created by selecting 1,000 [images/fish species] of largemouth bass and bluegill captured at the pond for several months. The purpose of this study is to classify the exotic invasive species living in rivers and inland waters, and hence, Dataset Z is used as the test dataset of the exotic invasive species in the natural environment. Therefore, this dataset is used for verifying the acquisition of the data of several species and the number of learning data, and the selection of the most efficient CNN.

IV. DISCUSSION AND CONCLUSIONS

The images of the fish captured in the laboratory environment were synthesized to approximate the images captured in the natural environment, and were then used to construct Data group I. Furthermore, Data group II was constructed to select proper data for the training of the CNNs. We experimentally evaluated the performances of various CNNS such as AlexNet, which has a simple structure and high performance, VGGNet (VGG16, VGG19), and GoogLeNet. Each CNN attempts to reduce the learning times using the transfer learning provided by MATLAB (MathWorks). It is also important to consider the learning and computation times when CNNs are applied to the detection systems for exotic

invasive species in rivers or inland water.

The optimization method used by each CNN was Adam [23], with the initial learning rate set to $1.0 \times e^{-4}$. The size of the mini-batch was adjusted according to the performance of each CNN. Intel i9-7900 3.30 GHz CPU and four NVIDIA GeForce GTX1080Ti were the hardware devices used.

A. Experimental Results and Analysis by Data Group I

Data group I for learning, validation, and testing consisted of 5,000 [images/fish species] images. First, 20% of the total data were selected as the test data, and 70% and 30% of the remaining data were selected as the learning and validation data, respectively.

For training AlexNet using Dataset D from Data group I, the mini-batch size was 1,024 and the total number of learning data was 450; hence, 23.56 epochs were performed. After learning, the performance of AlexNet showed a high accuracy of 98.33% for 7,000 (1,000 [images/fish species]) test images.

Table 3 shows the mini-batch size, the processing time for one image, and the classification accuracy for the four CNNs with Data group I. The processing time for one image and the accuracy of classification for developing a classification system are equally important factors for the elimination of exotic invasive species. If the fish species classification system is installed in a natural environment, such as a river or inland water, rather than a laboratory environment, a short processing time is necessary. The processing time of AlexNet is approximately 4.3 times, 11.1 times, and 11.6 times shorter than that of GoogLeNet, VGG16, and VGG19, respectively. AlexNet is not deeper than the other CNNs, which indicates that the processing time is shorter. This is an advantage of the application of AlexNet to a fish species classification system.

Fig. 4 plots the classification performances. For Datasets A–F, the CNNs showed a classification accuracy of more than 95%, and Datasets G–H showed a classification accuracy of more than 85%. CNNs are very robust to noise. In

Table 3. Classification performance of CNNs using test data in Data group I

CNN	AlexNet	GoogLeNet	VGG16	VGG19
Mini batch	1,024	512	256	200
Processing time for one image	$1.234e^{-2}$	$5.347e^{-2}$	$1.375e^{-2}$	$1.430e^{-2}$
Accuracy of Dataset A (%)	99.63	99.93	99.90	99.64
Accuracy of Dataset B (%)	99.46	99.90	99.79	99.74
Accuracy of Dataset C (%)	99.30	99.61	99.31	99.31
Accuracy of Dataset D (%)	98.33	98.36	99.16	98.70
Accuracy of Dataset E (%)	97.63	97.74	98.70	98.50
Accuracy of Dataset F (%)	95.70	95.01	98.13	97.41
Accuracy of Dataset G (%)	90.90	91.29	95.73	94.33
Accuracy of Dataset H (%)	88.30	85.56	92.70	87.31

Fig. 4. Classification accuracy of CNNs using Data group I

Table 4. Classification performance of CNNs using test data in Data group II

CNN	AlexNet	GoogLeNet	VGG16	VGG19
Accuracy of Dataset L (%)	93.10	93.03	96.11	96.11
Accuracy of Dataset M (%)	90.17	91.49	95.76	95.30
Accuracy of Dataset N (%)	86.76	87.83	92.19	91.23
Accuracy of Dataset O (%)	77.50	80.00	77.36	76.87
Accuracy of Dataset P (%)	71.01	75.84	72.93	68.53
Accuracy of Dataset Q (%)	99.47	99.66	99.76	99.57
Accuracy of Dataset R (%)	98.98	98.07	98.92	98.49
Accuracy of Dataset S (%)	95.90	97.64	98.03	97.56
Accuracy of Dataset T (%)	91.94	92.31	89.59	87.47
Accuracy of Dataset U (%)	85.69	86.61	83.69	76.10

particular, Datasets D–H are very likely to be misrecognized by humans, because the turbidity of images is very high. Overall, the recognition rate achieved by CNNs was better than that achieved by humans. Even though AlexNet is not as deep as the other CNNs, its performance is excellent. This indicates that a small network is more useful, because the number of species to be classified is small.

Through the performance evaluation for Data group I, we have confirmed that CNNs show excellent performance in learning and classifying, even if the images change depending on the brightness and the depth and turbidity of the water, as in the natural environment.

B. Experimental Results and Analysis by Data Group II

The previously designed CNNs showed excellent classification performances of the fish species for Data group I, including images with a wide range of natural environments. However, the number of data required for CNNs to classify fish species is unknown. Considerable amounts of time and effort are required to acquire images of fish in diverse natural environments. In this study, we attempt to evaluate the proper number of image data required for designing a classification system for fish species using CNNs. Table 4 shows the classification performance results of Data group II, and Fig. 5 graphically shows the classification performances.

The results of the performance evaluation of Data group II confirmed that the greater the number of learning data, the higher is the classification performance. VGG16 and VGG19 showed better performance than the other CNNs when the number of learning data was large; however, as the number of learning data decreased, their performance was extremely degraded. This is because VGG16 and VGG19, which have more parameters than those of AlexNet and GoogLeNet, are difficult to train. Therefore, one of the reasons for their performance degradation is the lack of learning data.
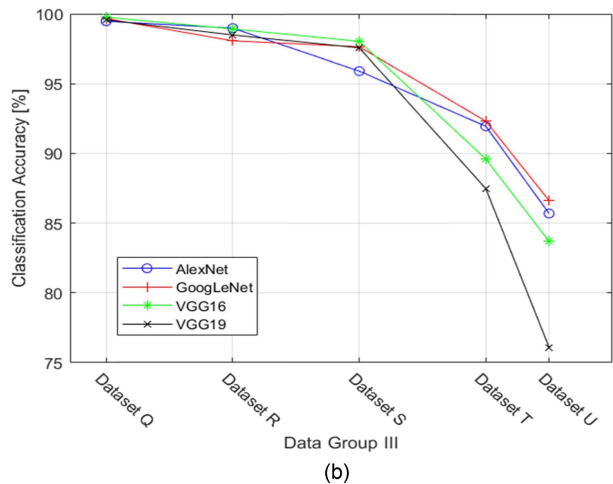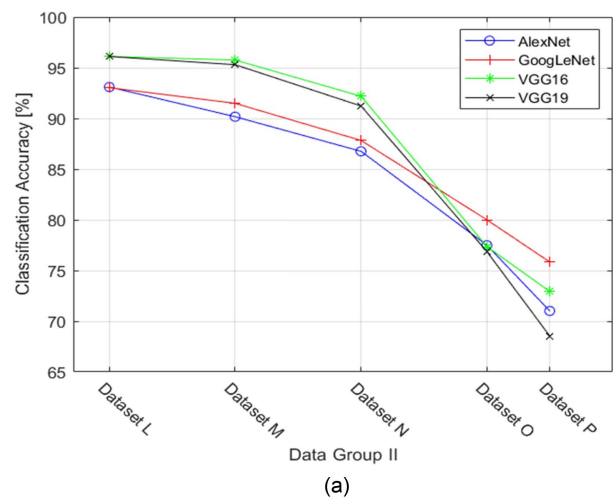

(a)

(b)

Fig. 5. Classification accuracy of CNNs using Data group II (a) Datasets L–P and (b) Datasets Q–U.

Fig. 5(a) shows that, if the number of fish species to be classified is small, and a classification accuracy of more than 90% is required, the number of learning data should be set to more than that of Dataset M among Datasets L–P. If the

required classification accuracy is 85% or more, the number of learning data can be set to approximately that of Dataset N. Thus, tens of thousands of image data are not required, because there are only a few species to be classified. If the number of fish species to be classified is large, the number of fish image data will increase as well.

Datasets L–P randomly contain fish images that are difficult for people to recognize. Therefore, we performed the same experiment using Datasets Q–U, which are assumed to be in a natural environment that humans can recognize with 100% accuracy. In Fig. 5(b), the overall data trend is similar to that of Fig. 5(a), but the classification performance is greatly improved. In particular, AlexNet and GoogLeNet showed a classification accuracy of over 90% for Dataset T. When the number of fish species to be classified is small, even with a small number of learning data, AlexNet and GoogLeNet, which have a shallow network or small parameters, are efficient.

The number of learning data can be expressed in an exponential function form using the classification accuracy of CNNs with Data group II as shown in Fig. 6. The number of learning data increases exponentially with the classification accuracy. Therefore, the number of learning data is affected by the number of fish species to be classified, classification accuracy, and data reliability. The fish species classification system requires the learning data of 1,500 [images/fish species] for Datasets L–P, and 700 [images/fish species] for Datasets Q–U. Therefore, if the data recognition reliability of Datasets Q–U is assumed to be 100% and the required classification accuracy is 90% or more, it is appropriate to set the number of learning data to approximately 700 [images/fish species], which is 100 times the number of fish species to be classified. To obtain a fish species classification accuracy of 95%, approximately 1,170 [images/fish species] are required, which is 167 times the number of fish species to be classified. Assuming that the Datasets D–H are difficult for humans to recognize, the data recognition reli-

ability of Datasets Q–U is 37.5% from the ratio of Datasets A–H. Therefore, to obtain a classification accuracy of more than 90% and 95% for Datasets L–P, the number of learning data can be estimated as 1,867 and 3,120 [images/fish species], respectively. These numbers are inversely proportional to the data recognition reliability. These results are similar to the results shown in Fig. 6.

C. Experimental Results and Analysis by Dataset Z

Dataset Z consists of 1,000 [images/fish species] of largemouth bass and bluegill captured in the pond over several months. This dataset was used as test data for the performance evaluation of the CNNs. The learning of CNNs requires a large number of fish images captured directly from a natural environment, but obtaining such images involves several difficulties. To solve these problems, we attempt to replace the fish images captured in a natural environment by CNN learning with the fish images synthesized in a laboratory, such as Datasets L–U. We will discuss the number of fish images required for learning.

Table 5 shows the classification performance of the CNNs for Dataset Z, and Fig. 7 graphically shows the classification performances. The results of the performance evaluation of the CNNs showed a similar trend to the experimental results for Data group II. This suggests that CNN learning can be achieved even with the synthesized fish images, rather than the images captured directly from the natural environment. The performances of AlexNet and GoogLeNet are generally better than those of VGG16 and VGG19. If the number of species to be classified is small, as in this study, smaller networks are more useful. The CNNs learned with Datasets O, P, T, and U show poorer performances compared with the CNNs learned with Datasets L–N and Q–S. This can lead to over-fitting when training the CNN with fewer learning data, which yields an incorrect result for the untrained image. In this study, the classification accuracy was over 80% when at least 2,000 [image/ fish species] (learning: 1,400, validation: 600) data were used for CNN learning. The CNNs learned
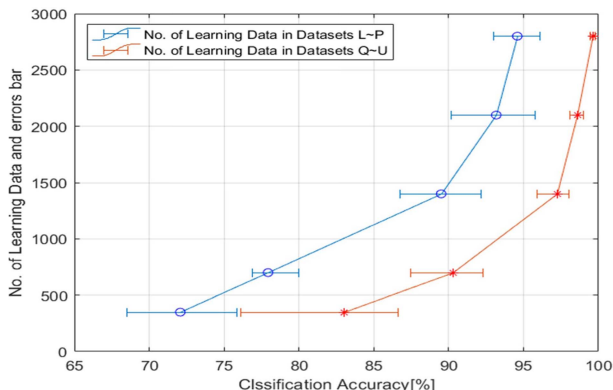

Fig. 6. Classification accuracy of CNNs using Data group II (a) Datasets L–P, and (b) Datasets Q–U.

Table 5. Classification performance of CNNs using test data in Dataset Z

| CNN	AlexNet	GoogLeNet	VGG16	VGG19
Accuracy of Dataset L (%)	92.85	94.85	83.05	87.75
Accuracy of Dataset M (%)	92.35	83.65	86.80	84.50
Accuracy of Dataset N (%)	94.45	92.60	87.10	84.10
Accuracy of Dataset O (%)	72.65	84.10	52.35	70.05
Accuracy of Dataset P (%)	45.10	58.25	39.90	47.25
Accuracy of Dataset Q (%)	91.05	97.65	94.10	89.70
Accuracy of Dataset R (%)	87.65	98.35	87.05	85.60
Accuracy of Dataset S (%)	88.75	96.40	85.50	96.00
Accuracy of Dataset T (%)	83.50	94.20	62.95	68.10
Accuracy of Dataset U (%)	51.50	48.70	49.35	48.75
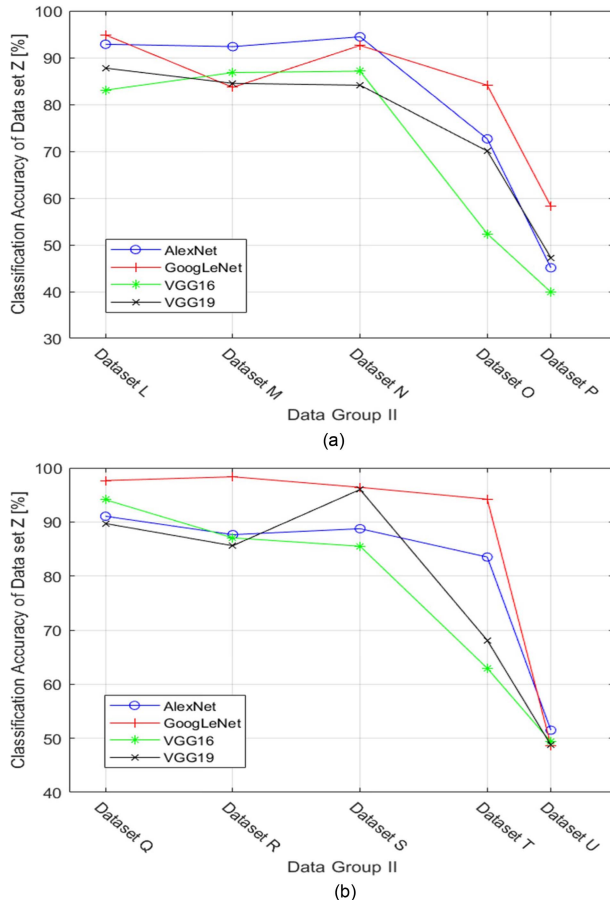
Fig. 7. Classification accuracy of CNNs using Dataset Z (a) Datasets L–P and (b) Datasets Q–U.

with Datasets Q–U exhibit a better overall performance than those learned with Datasets L–P. Dataset Z used as the test data consists of images captured in a natural environment that can be recognized by humans, as shown in Fig. 2(a), so that the CNNs trained with Datasets Q–U showed a better performance. Datasets L–P, which contain random images of fish that are difficult to recognize, rather adversely affected the classification performance of the CNNs.

Therefore, we have drawn the following conclusions from this study:

1) As the number of learning data for the CNNs increases, their classification accuracy increases.

2) The smaller the number of objects to be classified, the better is the performance of a small CNN in terms of learning time, processing time, and performance.

3) The learning data should be composed of data that are at least lower than the recognition reliability of the test data, and the CNN should be trained.

4) The proposed data acquisition method solves the difficulty of acquiring a large number of data in a natural environment, and its effectiveness is confirmed via

experiments.

V. CONCLUSIONS

In this paper, we propose appropriate criteria for obtaining fish species data and the number of learning data, and for selecting the most appropriate CNN for the efficient classification of exotic invasive fish species for their extermination. The acquisition of fish species data for learning CNNs has several limitations, due to the changes in luminous intensity according to sunlight, and the depth and turbidity of water. To solve these problems, a large number of fish images are acquired for various fish species in a laboratory environment, and the obtained fish images are converted into fish images obtained in different natural environments through simple image synthesis. Data group I consisted of images from a wide range of natural environments, whereas Data group II was constructed by changing the number of data in Data group I.

The experiments confirmed that the classification performances of the four CNNs were excellent in Data group I, which included various environmental conditions. In particular, AlexNet showed the shortest processing time, and it was confirmed to be suitable for developing systems with a small number of objects to be classified. In experiments for selecting the number of learning data using Data group II, if the number of fish species to be classified is small, AlexNet, which is a shallow network, and GoogLeNet are more effective. In addition, we confirmed that the number of learning data is affected by the number of classifications, classification accuracy, and data recognition reliability.

Therefore, the data acquisition method via image synthesis was presented to solve the difficulty of acquiring a large number of data in a natural environment, and its effectiveness was confirmed via experiments.

REFERENCES

[1] Convention on Biological Diversity, UN environment programme. [Internet], Available: https://www.cbd.int/abs/default.shtml.

[2] Korea Institute for International Economic Policy. 2010, 10th Session of the Conference of the Parties to the Convention on Biological Diversity: Nagoya Protocol, [Internet], Available: http://www.kiep.go.kr/sub/view.do?bbsId=global_econo&nttId=185515.

[3] H. W. Qin, X. Li, J. Liang, Y. G. Peng, and C. S. Zhang, "DeepFish: Accurate underwater live fish recognition with a deep architecture," *Neurocomputing*, vol. 187, no. 26, pp. 49-58, 2016. DOI: 10.1016/j.neucom.2015.10.122

[4] D. J. Lee, R. B. Schoenberger, D. Shiozawa, X. Xu, and P. Zhan, "Contour matching for a fish recognition and migration-monitoring system," in: *Two-and Three-Dimensional Vision Systems for Inspection, Control, and Metrology II*, International Society for Optics and Photonics, vol. 5606, pp. 37-48, 2004. DOI: 10.1117/12.571789

[5] N. J. C. Strachan, P. Nesvadba, and A. R. Allen, "Fish species recognition by shape-analysis of images," *Pattern Recognition*, vol. 23, no. 5, pp. 539-544, 1990.

[6] D. J. White, C. Svellingen, and N. J. C. Strachan, "Automated measurement of species and length of fish by computer vision," *Fisheries Research*, vol. 80, no. 2-3, pp. 203-210, 2006. DOI: 10.1016/j.fishres.2006.04.009

[7] N. J. C. Strachan, "Recognition of fish species by colour and shape," *Image Vision Computing*, vol. 11, no. 1, pp. 2-10, 1993. DOI: 10.1016/0262-8856(93)90027-E.

[8] G. French, M. Fisher, M. Mackiewicz, and C. Needle, "Convolutional neural networks for counting fish in fisheries surveillance video," in *Proceedings of the Machine Vision of Animals and their Behaviour* (MVAB), BMVA Press, pp. 7.1-7.10, 2015. DOI: 10.5244/C.29. MVAB.7

[9] A. Salman, A. Jalal, F. Shafait, A. Mian, M. Shortis, J. Seager, and E. Harvey, "Fish species classification in unconstrained underwater environments based on deep learning," *Limnology and Oceanography: Methods*, vol. 14, no. 9, pp. 570-585, 2016.

[10] J. H. Park, K. B. Hwang, H. M. Park, and Y. K. Choi, "Application of CNN for Fish Species Classification," *Journal of the Korea Institute of Information and Communication Engineering*, vol. 23, no. 1, pp. 39-46, 2019.

[11] G. López, L. Quesada, and L.A. Guerrero, "Alexa vs. Siri vs. Cortana vs. Google Assistant: a comparison of speech-based natural user interfaces," in *International Conference on Applied Human Factors and Ergonomics*, Springer, pp. 241-250, 2017.

[12] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," in *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278-2324, 1998.

[13] G. E. Hinton, S. Osindero, and Y. W. Teh, "A fast learning algorithm for deep belief nets," *Neural Computing*, vol. 18, no. 7, pp. 1527-1554, 2006.

[14] Y. Bengio, "Learning deep architectures for AI," *Foundations and trends in Machine Learning*, vol. 2, no. 1, pp. 1-127, 2009.

[15] J. H. Park, K. B. Hwang, and Y. K. Choi, "Design of CNN with MLP Layer," *Journal of Korean Society of Mechanical Technology*, vol. 20, no. 6, pp. 776-782, 2018.

[16] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *Advances in neural information processing systems 25*, pp. 1097-1105, 2012.

[17] E. Rezende, G. Ruppert, T. Carvalho, A. Theophilo, F. Ramos, and P. de Geus, "Malicious software classification using VGG16 deep neural network's bottleneck features," in *Information Technology-New Generations,* Springer, pp. 51-59, 2018.

[18] L. Chen, H. Zhang, J. Xiao, L. Nie, J. Shao, W. Liu, and T.-S. Chua, "Sca-cnn: Spatial and channel-wise attention in convolutional networks for image captioning," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 5659-5667, 2017.

[19] C. Liu, Y. Cao, Y. Luo, G. Chen, V. Vokkarane, and Y. Ma, "Deepfood: Deep learning-based food image recognition for computer-aided dietary assessment," in *International Conference on Smart Homes and Health Telematics*, Springer, pp. 37-48, 2016.

[20] Stanford Vision Lab, Stanford University, Princeton University, Large Scale Visual Recognition Challenge, [Internet], Available: https://www.image-net.org.

[21] The Courant Institute of Mathematical Sciences, New York University, THE MNIST DATABASE of handwritten digits, [Internet], Available: http://yann.lecun.com/ exdb/mnist/

[22] CIFAR, [Internet], Available: https://www.cifar.ca/

[23] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, 2014, [online] Available: https://arxiv.org/abs/1412.6980.

Jin-Hyun Park

received his B.S., M.S., and Ph.D. degrees in electrical engineering from Pusan National University, Pusan, Korea in 1992, 1994, and 1997, respectively. He is currently a professor in the Department of Mechatronics Engineering of Gyeongnam National University of Science and Technology. His research interests include robotics, intelligent control, and deep learning.

Young-Kiu Choi

received his B.S. degree in electrical engineering from Seoul National University, Seoul, Korea in 1980, and his M.S. and Ph.D. degrees in electrical and electronic engineering from KAIST, Seoul, Korea in 1982 and 1987, respectively. He is currently a professor in the Department of Electrical Engineering of Pusan National University. His research interests include robotics, intelligent control, and deep learning.