Journal of information and communication convergence engineering
http://jicce.org

J. Inf. Commun. Converg. Eng. 19(2): 79-83, Jun. 2021 Regular paper

Novel Image Classification Method Based on Few-Shot Learning in Monkey Species

Guangxing Wang[1], Kwang-Chan Lee[2], and Seong-Yoon Shin[2*], *Member*, *KIICE*

[1]Information Technology Center, Jiujiang University, Jiujiang, 332005 China
[2]School of Computer Information & Communication Engineering, Kunsan National University, Kunsan, 54150 Korea

Abstract

This paper proposes a novel image classification method based on few-shot learning, which is mainly used to solve model overfitting and non-convergence in image classification tasks of small datasets and improve the accuracy of classification. This method uses model structure optimization to extend the basic convolutional neural network (CNN) model and extracts more image features by adding convolutional layers, thereby improving the classification accuracy. We incorporated certain measures to improve the performance of the model. First, we used general methods such as setting a lower learning rate and shuffling to promote the rapid convergence of the model. Second, we used the data expansion technology to preprocess small datasets to increase the number of training data sets and suppress over-fitting. We applied the model to 10 monkey species and achieved outstanding performances. Experiments indicated that our proposed method achieved an accuracy of 87.92%, which is 26.1% higher than that of the traditional CNN method and 1.1% higher than that of the deep convolutional neural network ResNet50.

Index Terms: Deep learning, Feature extraction, Few-shot learning, Image classification

I. INTRODUCTION

In recent years, deep learning models have been applied to computer vision tasks with great success, such as face recognition, object recognition, image classification, and semantic segmentation etc. [1-3]. Furthermore, several outstanding deep learning models have emerged, such as LeNet, AlexNet, GoogLeNet, VGG(Visual Geometry Group), and ResNet [4-8]. These deep learning models based on the convolutional neural network (CNN) model have different characteristics and can achieve satisfactory results for different tasks. Deep learning models can automatically learn features from data, which generally requires a large amount of available training data, particularly for very high-dimensional input samples such as image and video processing. If the number of samples is small, the deep learning model can extract minimal features, and the expected results produced by the model are unsatisfactory.

CNNs are also called shift-invariant artificial neural networks [9] because they can express learning and can perform shift-invariant classification of input information according to their hierarchical structure. As the deep learning model mainly relies on the sample data features extracted by the convolutional layer to realize object recognition and image classification, the number of samples determine the accuracy of the model prediction to a certain extent. The aforementioned classic CNNs are all models trained on large image datasets such as ImageNet [10, 11], which can achieve better results in computer vision tasks. However, in realistic scenes, minimal features are extracted by the model owing to the small amount of sample data, resulting in non-convergence and over-fitting of the deep learning model during training

Received 02 March 2021, **Revised** 13 April 2021, **Accepted** 15 April 2021
***Corresponding Author** Seong-Yoon Shin (E-mail: s3397220@kunsan.ac.kr, Tel: +82-63-469-4860)
School of Computer Information & Communication Engineering, Kunsan, 54150 Korea.

Open Access https://doi.org/10.6109/jicce.2021.19.2.79 print ISSN: 2234-8255 online ISSN: 2234-8883

that directly affects the prediction results of the model.

This paper proposes an optimized image classification method that can solve the problem of image classification in small datasets. In these methods, we use a deepened model depth to realize the feature extraction of small dataset images. On the one hand, we use the data augmentation technology to increase the number of training samples and the regularization technology to suppress model overfitting. On the other hand, by setting a lower learning rate, shuffle, and other model optimization methods, model convergence is accelerated. The use of these model optimization methods can significantly improve the prediction accuracy.

II. BACKGROUND

A. Few-Shot Learning

Humans can recognize new objects using minimal samples. For example, children can recognize simple objects, such as apples and strawberries, with only individual pictures in the book. Cognitive ability refers to the perception of objects. Nerve cells play an essential role in cognition. Researchers hope that a machine learning model can quickly achieve few-shot learning [12] by training a large amount of data in a particular category and then training only a small number of samples for a new category arising from a downstream task. Traditional few-shot learning considers that both the training data and test data come from the same domain. If the unknown domain is included in the downstream task, the traditional few-shot learning method is unsatisfactory.

Recently, along with the rapid development of machine learning, the development of few-shot learning in the image processing field has surpassed that in the natural language processing field, and excellent few-shot learning has been focused on [13, 14]. Few-shot learning is an application of meta-learning in the field of supervised learning. Meta-learning, also known as learning-of-learning, decomposes a dataset into multiple meta-tasks in the meta-learning phase to learn the model generalization ability when categories change. There is no need to face brand new categories in the meta-testing stage. The classification can be completed by changing the existing model. The definition of small-sample learning is as follows: a few-shot learning set contains multiple categories, each with multiple samples. C categories are randomly chosen from the training set of the training phase, along with K samples for each category (total CK data), and the meta-task is composed of inputs to the model support set. Extract the batches of samples from the remaining data as model prediction objects (batch sets). The model must learn to distinguish these C categories from C*K data, and these tasks are called C-way K-shot problems.

In a learning process with only a few-shot learning session, different meta-tasks are sampled for each training (episode); therefore, the training as a whole involves different combinations of categories. This mechanism allows the model to learn the common parts of various meta-tasks, such as extracting important features and comparing sample similarities, but removes the relational task-specific parts from the meta-tasks. Models trained using this learning mechanism can classify well even when faced with new and unseen meta-tasks.

Few-shot learning models can be broadly divided into model-based, metric-based, and optimization-based models. The mathematical definition of the model-based model is shown in (1).

$$P_\theta(y|x,S) = f_\theta(x,S), \tag{1}$$

where P represents probability, S represents the sample, x denotes the input sample, y denotes the output sample, and θ denotes the weight. f represents the entropy function, also known as full conditional embedding, and can be understood as a model predicted by the model. The mathematical definition of a metric-based model is shown in (2).

$$P_\theta(y|x,S) = \sum_{(x_i,y_i)\in S} k_\theta(x,x_i,S)y_i\,. \tag{2}$$

where k represents the number of samples. The mathematical definition of optimization-based model is shown in (3).

$$P_\theta(y|x,S) = f_\theta(S)(x). \tag{3}$$

B. Dataset

The study of monkey species is conducive to researchers on the habits, population classification, and genetic characteristics of monkeys and is of great significance for studying human evolutionary history. This study established a small deep learning model to classify approximately 1,400 monkey group images of 10 species. This dataset was taken from Wikipedia's monkey cladogram [15], named 10 monkey species. The training dataset had 1098 images, and the test dataset had 272 images. The number of images in each category was not uniform. Compared to the Dogs vs. Cats dataset [16] (approximately 25,000 images), the number of samples in this dataset was minimal. Therefore, it was challenging to establish a deep model to achieve monkey species classification and improve the classification accuracy.

III. PROPOSED METHOD

A. Model Architecture

When there are enough data samples, the CNN model can be competent for most image recognition and classification

tasks owing to its simple structure, small number of parameters, fast data feature extraction, and high prediction accuracy. However, when dealing with image classification tasks of small datasets, although pre-training models (such as VGG and ResNet) can extract more data features, the result of the pre-trained model will increase the number of model parameters, resulting in a deeper model hierarchy. The model training time is extremely long. Therefore, we reformed and optimized the CNN model structure without significantly increasing the model parameters, mainly including the following aspects.

First, we extracted more data features by adding a small number of convolutional and pooling layers. Here, we used a 3 × 3 convolution kernel, a convolution stride size of 1 × 1, and set the convolution layer activation function to ReLU. We set the maximum pooling size to 2 × 2 and the pooling stride size to 2 × 2. Moreover, we set up a combination module of two layers of convolution and a maximum pooling layer, and added two such combination modules before the fully connected layer.

Second, in the fully connected layer, we used regularization techniques such as dropout to reduce unnecessary neurons, provide essential data features for the classifier, and improve classification accuracy. In the fully connected layer, we set the two layers dropout values to 0.5 and 0.25.

Third, we set softmax as a classification function to realize the classification output of input image feature neurons and output the classification results of 10 categories.

Finally, in the model training process, we used the categorical cross-entropy verification function as the loss function, and we used Adam [17] as the optimizer of stochastic gradient descent to optimize the training process of the model and accelerate the model convergence. The structure of the proposed model is illustrated in Fig. 1.

B. Data Argumentation

Data enhancement, also known as data expansion, refers to the value of limited data corresponding to more data without significantly increasing the data. Data augmentation is very effective for small sample datasets. Data enhancement can be divided into two methods: supervised and unsupervised enhancement. Supervised data enhancement can be divided into single-sample data enhancement and multi-sample data enhancement methods. Unsupervised data enhancement can be divided into two directions: generating new data and learning enhancement strategies. In this study, we mainly adopted the data enhancement method of geometric transformation, as presented in Table 1. Data enhancement by such geometric transformation can increase the number of training samples to improve the model generalization ability, effectively suppress model overfitting, and improve the accuracy of model classification.

Table 1. Data argumentation

Parameters	Values
Rescale	1.0/255
Rotation (º)	40
Width shift	0.2
Height shift	0.2
Zoom range	0.2
Shear range	0.2
Horizontal flip	TRUE
Fill mode	NEAREST

IV. RESULTS

To test the classification effectiveness and performance of the proposed model, we used the Python computer programming language to construct a deep learning model and deployed the model on a graphics workstation equipped with an Intel Core i7-4790 chip, 16 GB memory, 2T hard disk, and a GTX960 graphics display card. The experimental procedure was as follows.

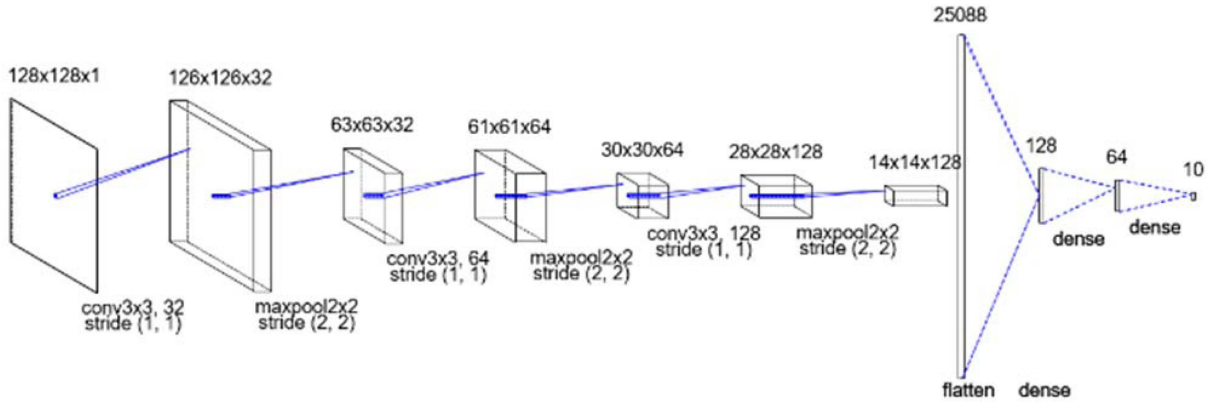
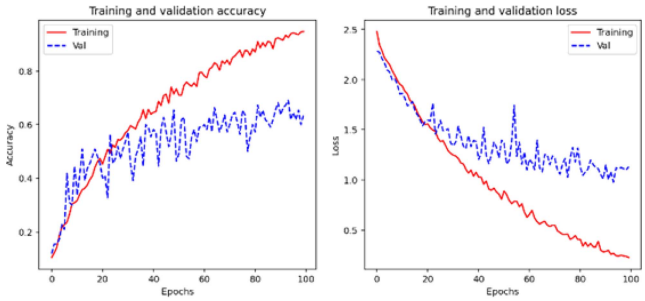
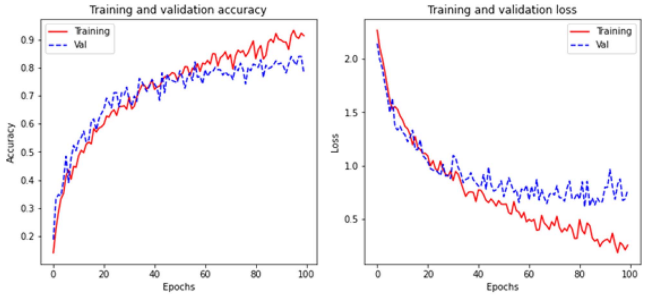


Fig. 1. Proposed model structure.



(a) (b)

Fig. 2. Accuracy and loss rate curves of the CNN model.



(a) (b)

Fig. 3. Accuracy and loss rate curves of the proposed model.

Table 2. Model performance

Model	Accuracy	
	Training	Test
CNN	0.8876	0.6974
VGG16	0.9191	0.8213
ResNet50	0.9268	0.8696
Ours	0.9358	0.8792

Classification experiments were performed on the preprocessed dataset in the CNN model and the proposed model; the number of batches of training data was set to 64, and the number of training rounds was set to 100. Fig. 2 depicts the accuracy and loss rate curves of the CNN model, and Fig. 3 shows the accuracy and loss rate curves of the proposed model during training. The average accuracy of the CNN model in the prediction dataset was 69.74%, and the accuracy of the proposed model was 87.92%. In addition, to comprehensively evaluate the performance of the proposed model, we trained and tested the preprocessed data in deep learning models such as VGG16 and ResNet50. The test results are presented in Table 2.

V. CONCLUSION

This paper proposes a novel image classification method based on few-shot learning in monkey species. This method increased the number of convolutional layers to achieve the rapid extraction of sample data features from a small dataset based on a CNN. Then, by fine-tuning the fully connected layer and adopting the dropout mechanism, the most extensive feature data was retained for the classification function to achieve a fast and accurate classification, thereby improving the classification accuracy.

We trained and tested the proposed model on a dataset of 10 monkey species and obtained a test accuracy of 87.92%. Compared to the CNN model, the accuracy rate increased by 26.1%. Compared to the VGG16 and ResNet50 deep learning models, the accuracy of the proposed model increased by 7% and 1.1%, respectively. The experimental results indicated that the proposed model exhibited an outstanding performance in image classification tasks on a small dataset.

ACKNOWLEDGEMENTS

"This research is partially supported by Institute of Information and Telecommunication Technology of KNU."

REFERENCES

[1] L. Jiao, F. Zhang, F. Liu, S. Yang, L. Li, Z. Feng, and R. Qu, "A survey of deep learning-based object detection," *IEEE Access*, vol. 7, pp. 128837-128868, 2019. DOI: 10.1109/ACCESS.2019.2939201.

[2] H. Laga, L. V. Jospin, F. Boussaid, and M. Bennamoun, "A survey on deep learning techniques for stereo-based depth estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1-27, 2020. DOI: 10.1109/TPAMI.2020.3032602.

[3] G. X. Wang and S. Y. Shin, "An improved text classification method for sentiment classification," *Journal of Information and Communication Convergence Engineering*, vol. 17, no. 1, pp. 41-48, 2019. DOI: 10.6109/jicce.2019.17.1.41.

[4] 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, Nov. 1998. DOI: 10.1109/5.726791.

[5] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications on ACM,* vol. 60, no. 6, pp. 84-90, 2017. DOI: 10.1145/3065386.

[6] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, pp. 1-9, 2015.

[7] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proceedings of International Conference on Learning Representations,* San Diego, CA, USA, 2014. DOI: arxiv.org/abs/1409.1556.

[8] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of Computer Vision & Pattern Recognition 2016,* Las Vegas, NV, USA, pp. 770-778, 2016.

[9] W. Zhang, "Shift-invariant pattern recognition neural network and its

optical architecture," in *Proceedings of Annual Conference of The Japan Society of Applied Physics*, Montreal, CA, 1988.

[10] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, Miami: FL, pp. 248-255, 2009. DOI: 10.1109/CVPR.2009.5206848.

[11] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet large scale visual recognition challenge," *International Journal of Computer Vision*, vol. 115, pp. 211-252, 2015. DOI: 10.1007/s11263-015-0816-y.

[12] N. Bendre, H. T. Marín, and P. Najafirad, "Learning from few samples: A survey," *Computer Vision and Pattern Recognition*, pp. 1-17, 2007.

[13] J. Lu, P. Gong, J. Ye, and C. Zhang, "Learning from very few samples: A survey," *Computer Vision and Pattern Recognition*, pp. 1-30, 2020.

[14] Y. Wang, Q. Yao, J. T. Kwok, and L. M. Ni, "Generalizing from a few examples: A survey on few-shot learning," *ACM Computing Survey (CSUR)*, vol. 53, no. 3, pp. 1-34, 2020.

[15] Kaggle, 10 monkey species [Internet], Available: https://www.kaggle.com/slothkong/10-monkey-species.

[16] Kaggle, Dogs vs. cats [internet], Available: https://www.kaggle.com/salader/dogsvscats.

[17] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," in *The 3rd International Conference for Learning Representations*, San Diego, CA, USA, pp. 1-15, 2015.

Guangxing Wang

He received his M.S. degree in Computer Application Technology from Huazhong University of Science and Technology, Wuhan, China in 2009. He received Ph.D. degree from the School of Computer Information Engineering of Kunsan National University Korea in 2020. From 2016 to the present, he has been an associate professor in the Information Technology Center of Jiujiang University in China. His research interests include data science, information system, and artificial intelligence.

Kwang-Chan Lee

He received his M.S. degree from the Dept. pf Management Information of Hankuk University of Foreign Studies, Seoul, Korea, in 2001. He a doctoral student from the Dept. of Computer Information Engineering of Kunsan National University, Gunsan, Korea, in from 2018 present. From 2006 to the present, he has been a professor in the same department. His research interests include image processing, big dara, and ERP.

Seong-Yoon Shin

He received his M.S. and Ph.D. degrees from the Dept. of Computer Information Engineering of Kunsan National University, Gunsan, Korea, in 1997 and 2003, respectively. From 2006 to the present, he has been a professor in the same department. His research interests include image processing, computer vision, and virtual reality.