Research on Similar Animal Classification Based on CNN Algorithm

Peiyi Zeng1*

1 University College London, London, United Kingdom, WC1E 6BT
*Corresponding author’s e-mail: zcempze@ucl.ac.uk.

Abstract. Animal image classification with CNN (convolutional neural network) is commonly investigated in aera of image recognition and classification, but major studies focus on species pictures classification with obvious distinctions. For example, CNN is usually employed to distinguish images between dogs and cats. This article puts the effort on similar animal images classification by applying simple 2D CNN via python. It focus on the binary classification for snub-nosed monkeys and normal monkeys. This distinguished is hard to be done manually in a short time. For constructing complete convolutional neural network, some preparations are done in advance, such as the database construction and preproces. The database is constructed by python crawler (downloading from google images), with 800 and 200 images for each class respectively as train data and test data. The pre-work includes image resizing, decoding and standardization. After that, the model is trained and then tested for verifying the model reliability. The training accuracy is 96.67% without any abnormality. On the basis of successful training, the test accuracy almost coincides with train accuracy in each 50 generations and plots in a graph. It indicates similar trends and results for them in the whole process. Because of this, CNN model in the study can help people identify rare animals in time and then people can effectively protect them. Therefore, CNN will be helpful in field of animal conservation, especially for rare species.

1. Introduction

CNN (convolutional neural network) is one category of deep learning neural networks. It is a kind of multilayer neural network with artificial neurons. The multilayer mainly consists of three basic types: input layer, hidden layer and output layer. The kernel existed in first hidden layer can gain the convolved feature from input images and then the convolved feature can be transmitted to the next layer. During the process, the feature obtained becomes more complex as layer is deeper. With different image classes, CNN can extract picture features respectively and proceed identification. Because of its working principle, it has a wide range of applications in image classification. For example, it can be found in Facebook’s photo tagging and in self-driving areas as the core technology [1]. Although some portion of CNN image recognition technologies are established well in certain areas, the higher level of CNN image recognition is still in infancy, such as the identification for road obstacles which is challenging under the complicated lane conditions. The main drawback of CNN is the requirement for large amount of data (e.g. same image with different angles and different lighting conditions) and failed to understand image contents.

In this paper, CNN model are applied for similar rare creature image classification. High similarities exist in various species of the same class. Hence, it is hard to distinguish them by eyes in real time. This model can help develop animal conservation because it can fast recognize endangered animals remotely. Then, taking further actions to protect them from catch and hunt. The model is trained to classify images
for snub-nosed monkeys, a kind of rare specials in monkey group, and normal monkeys which have high similarity to snub-nosed monkey. The CNN model extracts image features with input data to identify other similar images [2]. Training process enables CNN to identify differences among images. In test stage, it is able to classify images into different groups. Database construction and data preprocess are also contained in this article. Additionally, accuracy obtained in test data can be used to evaluate the model reliability. However, the model is a simple 2D convolution neural network, which means limitation applications existed in this model. Although some improvements are feasible on model reliability, it only can process 2D problems.

This paper aims to investigate the similar species discrimination according to fine distinctions in animal groups. Taking monkey group as an example, the neural network can distinguish between the snub-nosed monkey and normal monkey. This study based on CNN algorithms is quite constructive in rare animal conservation aera, which can be further utilized in other rare animal groups.

2. Method

There is binary classification in CNN model, one class is for snub-nosed monkey images, and the other is for normal monkey images. The database construction is through python crawler to download snub-nosed monkey images and normal monkey images from google respectively. For preventing some other types of images from already saving in the same folder before downloading, some coding are done for cleaning other image formats like “PNG” “JPEG”. However, the images downloaded are not always exactly the real monkey. Some TV characters, animation figures and drawing related to monkeys are all included. It is necessary to check and delete these inaccurate and unclear images at first before use. Thus, there are 800 images for each class as train datasets, and 200 images for each class as test datasets.

2.1. Preprocess for input image

Before setting up the CNN model, it is unavoidable to preprocess these images, because standardization is required for different pictures in Neural network. Every image for snub-nosed monkeys and normal monkeys must be classified and marked. The training datasets and test datasets have been separated in previous step. Thus, every image has corresponding unique label for neural work to learn. Consequently, naming the pictures are also made in standard form to ensure automatic reading by the algorithm. Besides, all images are set to 28×28 pixels, 3 RGB channels (black and white pictures are deleted). Images and labels are correspondingly made to lists, which are shuffled to create randomness and then eliminate the sequential influence from the crawler. The sequential influence is from the searching engine google (most related, time-based search, etc. options). The input data should be in batch process, because it is convenient for neural network training. As a result of the preprocessing, the image batch and label batch are generated as the input data for neural network.

2.2. The model design and construction

The framework of hidden layers in CNN consists of these mainly categories of layers, convolutional layer (Convo layer), pooling layer, normalization layer and fully connected layer [3]. All of these layer categories are involved in this model.

The initial model is mainly composed of seven convolution layers and four fully connected layers. Every convolution layer is followed by a pooling layer and a normalization layer. The first convolution layer begins with 1024 convolution kernels, which has two-dimensional size. However, with this model design the training result is easily to show 100% accuracy, which is as known as overfitting and is willing to have negative impact on test result. Through many adjustments, for example, some changes on the number of layers, number of convolutional kernels, size of convolutional kernels, etc, the training result can reach a high accuracy without overfitting in the whole training process.

As indicated in figure 1, it is the showcase for the final model structure. There are only three convolution layers followed by two fully connected layers. Moreover, every convolution layer has its own max pooling layer and normalization layer, which has the uniform size of 2 × 2 and strides of 2. In addition, all the convolution layers apply “relu” as linear activation function, and the same size of
convolution kernel which is 3×3. The first convolution layer has 128 convolution kernels, and 64, 32 for the second and third layer respectively. In addition, the bias variable for each convolution layer corresponds to the number of convolution kernels. There are 256 neurons in fully connected layers. At the same time, it applies linear activation function “relu” as same in convolution layer. For preventing overfitting from happening in the training result, the dropout layer is followed by the second fully connected layer with dropout rate 0.6. The possibility that each image accurately identification for each sample is calculated by the “softmax” classifier, in which type of classifier can make the sum of output possibility from the neural network is equal to 1. At the end of model design, there is the loss value optimization and output of accuracy calculation.

2.3. Model training process with training datasets

There are two sample classes in the training process. With the initial model design, it defines 64 as batch size, 200 as capacity, 800 as maximum step, 0.0001 as learning rate. At the same time, the epoch is set as 100, which means the training result would be shown in every 100 steps. In this case, overfitting always occurs around 500, 600 and 700 steps during training process as shown in figure 2 and figure 3. For solving this problem, data augmentation and dropout can be applied [4]. Increasing the data size through data augmentation can make the model more robust in general. Thus, the initial number of 500 images for each class as train datasets is increased to 800 images. In model improvement, the last fully connected layer is closely followed by a dropout layer, a regularization method proposed by Srivastava et al, aiming to reduce the overfitting problem in convolutional neural networks [5]. Based on these changes, overfitting problem still occurs at the end of training step. Afterwards, taking some other simple optimization methods for improving this result, such as hyper-parameter adjustments and layers advancements. Hyperparameters are variated at first. The batch size is changed into 32. However, this change does not have significant influence on the result. When considering the number of training datasets, 800 images in each category, the model with high complexity is easier to have overfitting problem. Thus, the structure of model is simplified based on this consideration. The number of convolution layers and fully connected layers are attempted to be reduced, the maximum step is changed into 1000 and the epoch is 50. By observing the training result, from 900 to 950 steps, the researcher finds that overfitting is likely to occur. With simplifying the model structure in further and observing the training result, the final model structure is determined which has three convolution layers and two fully connected layers as indicated in figure 1. With this fixed model design, one of the methods for protecting the training result from overfitting is to reduce the maximum steps. Based on previous observation, the maximum step is decreased to 900. Training the model again, the accuracy is up to 96.88% as shown in figure 4. The accuracy is very close to 100% but it is not overfitting, hence, it is feasible to move to the next step with this successful training result.
3. Testing the model with test datasets
In previous training process, the data accuracy has achieved a satisfactory level and no abnormal occurrence in this case. On the basis of approving result, the model can be in the test process. Through aforementioned python crawler at the beginning, 200 images for golden monkeys and 200 images for normal monkeys as test datasets are obtained. As shown in figure 5, there is a direct comparison between the train accuracy and test accuracy. It indicates that there is no distinctive differences between them in the whole process variation. The more generations (number of training), the higher test accuracy. As we can see, there is a significantly jump up for test accuracy and train accuracy during the first 200 generations. At the end of the generation (around 900), the two accuracies are stabilized at approximately 96%, in which result is similar to the training accuracy at 900 steps as shown before. According to this plot, it clearly shows the high similarity between train results and test results. Thus, the CNN model is equipped with high reliability and is able to be applied in animal conservation field.

4. Conclusion
In conclusion, this model has accuracy around 96% after many attempts of optimization in terms of hyperparameters and model structures. With this model, it can distinguish normal animals and rare animals in a very short time. Despite striking similarities existed in biology, groups of animals do differ in some ways [6]. Thus, the model can be trained to help humans effectively protect rare specials of animals from other common creatures. On the other hand, the model is possible to find out new species with anomalous behavior, like much lower test accuracy.

Nevertheless, the model has some limitations. The most obvious one is the dataset, which has only two categories. With this model, only the snub-nosed monkey can be identified in the group of monkeys. It does not work in other kinds of rare monkeys, including sapajou, Hainan gibbons, etc. [7]. Besides,
the model cannot be used to identify a particular creature in multitype animals because of this limitation. One of the solutions is to enhance image numbers and image categories. It would be another difficulty for constructing this model because a large amount of training data is required. In addition, this binary classification in this CNN model can be improved to multiclass image classification. Thus, the model can identify more than one types of rare animals at the same time. Moreover, there are 2D convolutional kernels used in this model, which can utilize context across the width and height of single slice to make predictions [8]. However, compared with applying 3D convolutional kernels which can capture voxel information from adjacent slices (whole volume as inputs), 2D CNN (single slice as inputs) has lower accuracy, and it can only process 2D images.

Although some limitations existed in this CNN model, there are many spaces for improvements and more applications in a wider range of area. The development of CNN image recognition technology is still on the road. With more complex convolution neural network, it has high computational cost and high configuration requirement [9]. Moreover, one of the core difficulties for complex neural network is associated with the training dataset as it is the foundation for neural network learning. Along with accomplishing these challenges, the accuracy for image classification will be improved further and the network will be more concise in the future [9].

Acknowledgements
Gratefully acknowledges assisting by Carnegie Mellon University professor Ta'asan, who provides the knowledge foundation in this study.

References
[1] W. Rawat and Z. Wang, Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review, in Neural Computation, vol. 29, no. 9, Sept. 2017, pp. 2352-2449.
[2] Kavish Sanghvi, Adwait Aralkar, Saurabh Sanghvi and Ishani Saha, Fauna Image Classification using Convolutional Neural Network, International Journal of Future Generation Communication and Networking Vol. 13, No. 1s, (2020), pp. 08-16.
[3] Gousia Habib, Shaima Qureshi, Optimization and acceleration of convolutional neural networks: A survey, Journal of King Saud University - Computer and Information Sciences, October 20, 2020.
[4] T. Sinha, B. Verma and A. Haidar, “Optimization of convolutional neural network parameters for image classification”, 2017 IEEE Symposium Series on Computational Intelligence (SSCI), 2017, pp. 1-7.
[5] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever and R. Salakhutdinov, “Dropout: A simple way to prevent neural networks from overfitting”, The Journal of Machine Learning Research, vol. 15, no. 2014, pp. 1929-1958.
[6] S. Y. Chaganti, I. Nanda, K. R. Pandi, T. G. N. R. S. N. Prudhivith and N. Kumar, Image Classification using SVM and CNN, 2020 International Conference on Computer Science, Engineering and Applications (ICCSEA), 2020, pp. 1-5.
[7] Yang Y, Tian Y, He, C, Huang Z, Dong S, Wang B, Xiao W. The Critically Endangered Myanmar snub-nosed monkey Rhinopithecus strykeri found in the Salween River Basin, China. Oryx, 2018, 52(1), 134-136.
[8] M. Zreik, T. Leiner, B. D. de Vos, R. W. van Hamersvelt, M. A. Viergever and I. Išgum, Automatic segmentation of the left ventricle in cardiac CT angiography using convolutional neural networks, 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI), 2016, pp. 3-4.
[9] Jiaohua Qin, Wenyuan Pan, Xuyu Xiang, Yun Tan, Guimin Hou, A biological image classification method based on improved CNN, Ecological Informatics, vol.58, July 2020, pp. 1-2.