Animal Image Classifier Based on Convolutional Neural Network

Qinhan Yu

Statistics, Tianjin Nankai University
**Corresponding author. Email: 1943547556@qq.com*

ABSTRACT
In modern society, there are dogs and cats around people, as well as rare wild animals living in nature. The relationship between human beings and animals is getting closer and closer. The rapid development of machine learning and deep learning technology has been widely used in the academic field. Aiming at the problem of animal image classification, this paper uses Pytorch to learn about 10,000 pictures containing cats, dogs, and wild animals (tiger, lion, etc.) based on the research algorithm of convolutional neural network in the field of image classification. And a convolutional neural network model that can realize the animal image classifier is established and optimized, so that the model can efficiently classify cats, dogs and wildlife pictures. The results show that the accuracy of the two models is above 90%, and the model loss ranges from 0.706 to 0.061, and 0.807 to 0.051, respectively, showing the characteristics of good model fitting effect and strong optimization ability. Meanwhile, The accuracy of the model can be increased by properly increasing the number of full connection layers. Therefore, by constructing the convolutional neural network, the accurate detection of national ecological protection animal images can be realized.

Keywords: *Image classification, CNN, Pytorch, SGD*

1. INTRODUCTION

In the information age, image has become a medium and carrier to convey information, and is widely used in various fields [1][2]. Animal image classification is used in forests to classify animals in real time with profound research significance. In the past, many computer vision technologies were introduced, but due to the lack of accuracy, these technologies failed to meet the requirements [3]. Convolutional neural network (CNN) is a typical algorithm for realizing deep learning with its deep structure and good learning representation ability, which has been widely used in computer vision and other fields [4].

Many researchers have studied on the image classification. The purpose of the study is to establish a model that can realize the animal image classifing by using CNN. In this paper, first of all, a suitable data set is collected for the research. Secondly, the parameters and layers of the neural network are chosen. Finally, a suitable criterion is established for evaluating the model so as to find the most suitable model to solve the research problem. Through this study, a better neural network model can be obtained to realize the classification of animal images, and the further research of neural network is realized by combining the machine learning and deep learning technology with the ecological protection.

2. THEORETICAL BACKGROUND

2.1 Convolutional neural network

Convolutional neural network (CNN) is a kind of feed-forward neural network containing convolutional computing, and having a deep structure. It is one of the most representative algorithms of deep learning [5][6]. For general large-scale image classification problems, CNN can be used to build the class classifier [7]. A CNN mainly consists of the following three structures: the input layer, hidden layer and output layer.

The hidden layer includes three common constructions: the convolution layer, pooling layer and fully connected layer. The input layer is the input of the whole neural network. In the CNN for image processing, it generally represents the pixel matrix of a image. The convolution layer is the most important part of a CNN. It tries to analyze every small block in the neural network more deeply, so as to get more abstract features. The convolution layer has built-in parameters, mainly

including the convolution kernel size, step size and padding, which together determine the size of the output feature map of the convolution layer, and are super parameters of CNN. At the same time, it contains excitation functions to help express complex features; the pooling layer can transform an image with higher resolution into an image with lower resolution, and by further reducing the number of nodes in the last fully connected layer, the purpose of reducing the parameters in the whole neural network can be achieved. While the fully connected layer is used to complete the classification task, the feature map will lose its spatial topology in the fully connected layer, and be expanded into a vector. Finally, through the output layer, different kinds of probability distributions can be obtained.

2.2 Stochastic gradient descent

Stochastic gradient descent (SGD) is a simple but extremety effective method, which is often used to learn linear classifiers under convex loss functions such as support vector machines and logistic regression. In the establishment of CNN, SGD can calculate the model parameters, and minimize the loss function, so as to learn and optimize the model.

The direction of the maximum value of the derivative in the upper direction of the surface represents the direction of the gradient. Therefore, when the gradient descends, the weight should be updated along the opposite direction of the gradient, which can effectively find the global optimal solution. The working principle of SGD is as follows.

Suppose a loss function is

$J(\theta) = \frac{1}{2}\sum_{i=1}^{m}(h_\theta(x) - y)^2$,

and $h_\theta(x) = \theta_0 + \theta_1 x_1 \ldots + \theta_2 x_2 + \theta_n x_n$.

The updating formula of SGD can be defined as follows:

Loop{For i=1 to m, { $\theta_j := \theta_j + \alpha(y^{(i)} - h_\theta(x^{(i)}))x_j^{(i)}$ (for every j).}}

3. METHODOLOGY

To build an animal image classifier, the author mainly establishes a CNN, which belongs to the field of machine learning and deep learning. Pytorch is the programming language.

3.1 Data collection

In order to better analyze the data, a suitable data set should be chosen for building the model. The data set used is from Kaggle, which contains a training set and a test set. The training set includes 5,153 images of cats, 4,739 images of dogs, and 4,738 images of wildlife. The test set includes 500 images of cats, 500 images of dogs, and 500 images of wild animals.

3.2 Data processing

Since the datasets are not sufficient for building a good model, the data progressing needs to be performed. Through rotating, cropping and loading, the data image can be suitable for subsequent operations.

3.3 CNN model

In this paper, the author mainly builds and optimizes the full connection layers and parameters of the convolution layer, where relu is the nonlinear activation function in the model. Two models with different full connection layers are built and compared to choose the better one.

In the model evaluation, the accuracy and loss are used as criteria to evaluate the established neural network, and the algorithm of Cross-Entropy loss and SGD are used to optimize it.

3.3.1 Load and normalize datasets

First, loading and normalizing the training and test datasets by using torchvision. Using torchvision.datasets.ImageFolder to load the data, and randomly output four images and their corresponding labels in the training set as shown in Figure 1.

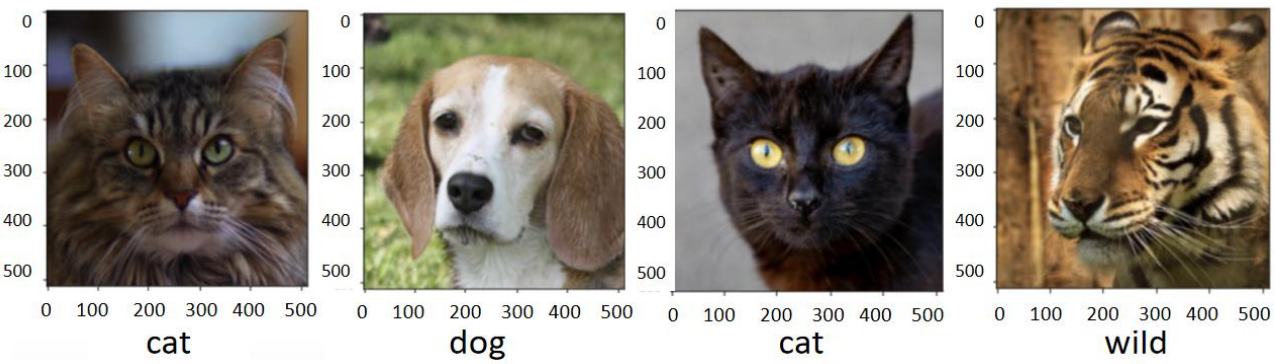

Figure 1 Picture preview.

3.3.2 CNN1 model building

The torch.nn package is used to build a CNN model with two convolution layers and two fully connected layers, setting parameters and weights. The model has 3 input image channels, 6 output channels, and 5x5 square convolution. Also, the relu is regarded as a non-linear activation function to obtain the final model. The model of two fully connected layers is recorded as CNN1.

3.3.3 CNN2 model building

Then, the same parameters are used to construct the second neural network, still with 3 input image channels, 6 output channels and 5x5 square convolution. But the difference is that the number the full connection layer is changed to three, and the neural networks of the three full connection layers is named as CNN2.

3.3.4 Loss function and optimizer

The application of the loss function is also important in the process of evaluating and optimizing the neural network. A loss function takes the pair of inputs, and computes a value that estimates how far away the output is from the target. The author uses a Classification Cross-Entropy loss and SGD with momentum as 0.9 and learning rate as 0.001.

3.3.5 Train the CNN

The test set is trained by applying the above established model and function into the training set for training, and cycling the dataset 10 times to finally get each loss, which is set every 2000 mini-batches output. The results of the two models are presented as follows.

Table 1 Loss of CNN1.

Cycle number	Loss
1	0.706
2	0.286
3	0.200
4	0.165
5	0.141
6	0.109
7	0.094
8	0.081
9	0.064
10	0.061

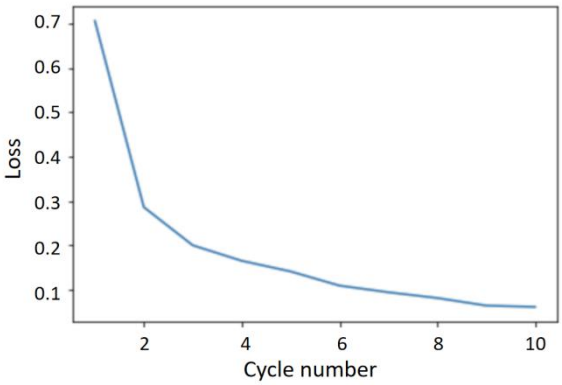

Figure 2 Loss of CNN1.

Table 2 Loss of CNN2.

Cycle number	Loss
1	0.807
2	0.325
3	0.227
4	0.171
5	0.143
6	0.108
7	0.092
8	0.069
9	0.066
10	0.051

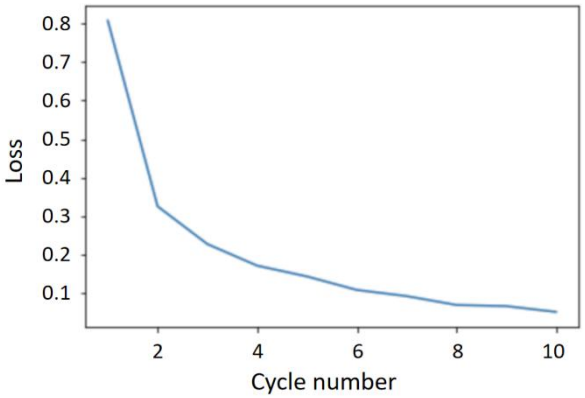

Figure 3 Loss of CNN2.

3.3.6 Test the CNN

In the final step, the models are further tested on the test set. As shown in figure 4, four images and labels of the test sets are randomly selected first.

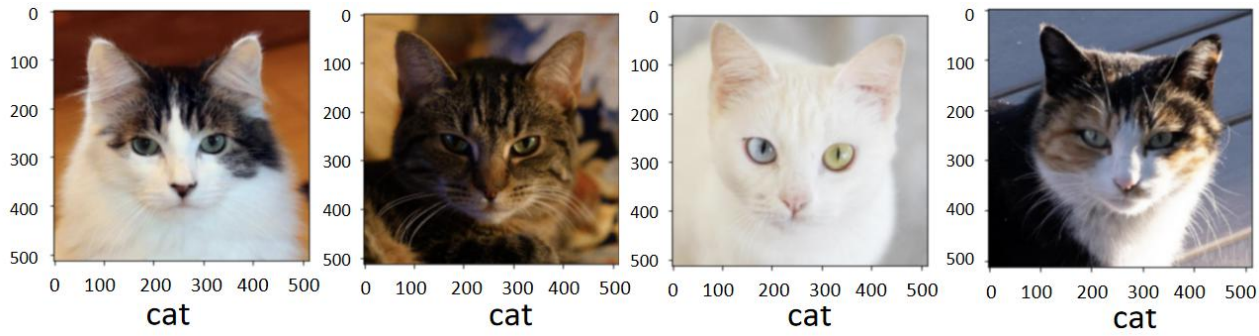

Figure 4 Images of the test.

It can be seen from figure 4 that the tested images are four cats. Then, the images are put into the CNN1 and CNN2, and the prediction results are all cats,.which means that the prediction results of both models are correct.

4. RESULTS AND FINDINGS

Therefore, it can be preliminarily judged that the training effects of both CNN1 and CNN2 are excellent. Additionally, all 1500 test sets are brought into the neural network, yielding the model accuracy to 93% and 94%, respectively. At the same time, in order to better understand the classification of each model, each image is analyzed accurately, obtaining the results as shown in Table 3

Table 3 Accuracy of each class of pictures.

	Cat	Dog	Wild
Accuracy rate of CNN1	93%	96%	89%
Accuracy rate of CNN2	96%	96%	90%

It can be seen from Table 3 that the prediction results of the two models for each category also show a good trend.

5. DISCUSSION

From the above analysis, it can be seen that the results of the two CNN models are both excellent, whether it is two fully connected layers or three fully connected layers. Through visually comparing the loss and accuracy of the two models, the specific contents can be found as follows:

Table 4 Comparison of models.

	The final loss	Accuracy	Accuracy of Cat	Accuracy of Dog	Accuracy of Wild
CNN1	0.061	93%	93%	96%	89%
CNN2	0.051	94%	96%	96%	90%

From the above table, it can be seen that the loss of CNN1 and CNN2 are small, which confirms the correctness of the model. The accuracy rate is basically above 90%. Especially in each classification, the dog set can be as high as above 95%. Therefore, a conclusion can be drawn that both neural networks can be used as a good animal image classifier.

However, by comprehensive comparison, the three-layer fully connected neural network shows better accuracy, so it can be further concluded that CNN2 is better to be used as an animal image classifier. In other

words, better training results can be achieved if the number of fully connected layers can be appropriately increased to make the network more complex without causing over-fitting.

6. CONCLUSION

In this paper, the animal image classifier is realized by establishing the convolutional neural network, which combines the machine learning and deep learning technology with ecological protection technology. The research results can get the following conclusion: the accuracy of the two models are both above 90%, with a loss range from 0.706 to 0.061, and 0.807 to 0.051 respectively, showing good model fitting effects and a strong ability for optimization. Therefore, a classifier for classifying three classifications of cats, dogs and wildlife can be obtained efficiently based on the models. Finally, to summarize, better training results could be obtained if we appropriately increase the number of fully connected layers to make the network more complex without causing over-fitting. Besides the advantages of simple with high efficiency and accuracy, the two models in this research also has their limitations. Both of the models can only classify three types of animals. In later studies, more kinds of animals can be tried to add to make the function of the animal image classification more powerful.

REFERENCES

[1] Chen Mian Shu, Yu Lu Lu, Su Yue, et al.Multi-label image classification based on convolutional neural networks [J].Journal of Jilin University (Engineering edition), 2020,50 (3): 1077-1084.

[2] Liu Qiong, Lee Zong Xian, Sun Fuchun, et al.Image recognition and classification based on deep-belief convolutional neural networks [J].Journal of Tsinghua University (Natural Science Edition), 2018,58 (9): 781-787.

[3] Likhith Prudhivi, M.Narayana, Ch Subrahmanyam, M.Gopi Krishna, Animal species image classification, Materials Today: Proceedings, 2021, ISSN 2214-7853,

[4] Zhang Xiaorui.Multi-label Image Classification and Recognition algorithm based on Convolutional Neural Network [J].Journal of Tonghua Normal College, 2022,43(02):75-82.DOI:10.13877/j.cnki.cn22-1284.2022.02.012.

[5] Goodfellow, I., Bengio, Y., Courville, A. . Deep learning (Vol.1) . Cambridge：MIT press，2016：326-366.

[6] Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, L., Wang, G.and Cai, J., 2015. Recent advances in convolutional neural networks.arXiv preprint arXiv:1512.07108.

[7] Srivastava, N.and Salakhutdinov, R.R., 2013. Discriminative transfer learning with tree-based priors.In Advances in Neural Information Processing Systems (pp.2094-2102).

[8] R.Karthik, R.Menaka, G.S.Kathiresan, M.Anirudh, M.Nagharjun, Gaussian Dropout Based Stacked Ensemble CNN for Classification of Breast Tumor in Ultrasound Images, IRBM, 2021, ISSN 1959-0318.

[9] Tahereh Hassanzadeh, Daryl Essam, Ruhul Sarker,EvoDCNN: An Evolutionary Deep Convolutional Neural Network for Image Classification,Neurocomputing,2022.

[10] Naftaly Wambugu, Yiping Chen, Zhenlong Xiao, Kun Tan, Mingqiang Wei, Xiaoxue Liu, Jonathan Li, Hyperspectral image classification on insufficient-sample and feature learning using deep neural networks: A review, International Journal of Applied Earth Observation and Geoinformation, Volume 105, 2021, 102603, ISSN 0303-2434.

[11] Additional Code link: https://github.com/qwershiqiqine/Code-of-image-classification.git