ISSN: 2414-1895 DOI: 10.6919/ICJE.202108_7(8).0019

Animal Image Classification Recognition based on Transfer Learning

Haomin Liu

Key Laboratory of Electronic and Information Engineering, University Southwest Minzu University, Chengdu, China.

Abstract

When the deep learning model is applied to small sample image classification, there are problems of long training time and over-fitting. For this reason, a small sample image classification method based on transfer learning is proposed. First of all, for the problem of unbalanced data set, we use data augmentation technology in the training set. Secondly, the MobileNet-V2 lightweight deep convolutional neural network is placed in a large data set for pre-training, and then the basic parameters in the source network pre-training process are retained and frozen, and then feature extraction training is performed on the data set in the training set. Finally, fine-tune the pre-trained network model, and unfreeze part of the hierarchy, which is used to adjust the network weight and train the target data set again. After 50 times of training, the migration learning method of MobileNet-V2 was used to obtain an accuracy of 95.16% on the Animals-10 data set.

Keywords

Transfer Learning; MobileNet V2; Image Classification.

1. Introduction

With the continuous development of deep learning methods, animal facial recognition as a kind of modern biological information recognition, it has received extensive attention in animal monitoring and protection, as well as in animal breeding and management. The use of deep neural networks for transfer learning is widely respected. Compared with traditional methods, the use of deep neural networks can directly extract more expressive features, and deep neural networks can meet the end-to-end needs of migration learning in real-world applications. At the level of deep learning, adaptation in transfer learning mainly completes two parts of work. One is to use different layers of the network to determine the degree of learning of image features by the deep convolutional neural network, and the other is to use different measurement criteria to determine the degree of learning of image features. Observe the generalization ability of the network [1-2]. Krizhevsky et al [3] won the ImageNet large-scale visual recognition challenge image classification and target positioning task champion in 2012 with AlexNet, demonstrating the huge potential of convolutional neural networks in the image field. After that, the proposal and development of vgg-16 [4], GoogleLeNet [5], ResNet [6] and other networks [7-8] greatly improved the classification accuracy of the classification task, while the application fields have gradually become diversified [9-10].However, the performance of these networks has been improved, and the depth of the model and the number of parameters have increased rapidly, which will lead to difficulties in model training, large storage space, and long training and prediction time. This is especially true for the method of integrating convolutional networks. In recent years, more and more researches have focused on the efficiency of network models. SqueezeNet using fire module published on ICLR in 2017, MobileNet using deep separable convolution published on CVPR in the same year, and ShuffleNet [11], Xception [12] and MobileNet V2 [13] published later,

ISSN: 2414-1895 DOI: 10.6919/ICJE.202108_7(8).0019

These networks greatly reduce the number of network parameters while maintaining similar accuracy, and solve the efficiency problem of the network by reducing the network parameters and the amount of calculation.

This article is based on the classification of animal images based on the Mobile Net V2 network. As a new broad market, the animal application field provides a new direction for the development of facial recognition technology. In the future, people can use the faster commercialization of animal face recognition to reverse the better development of face recognition. Animal facial recognition is a good new track and new opportunity.

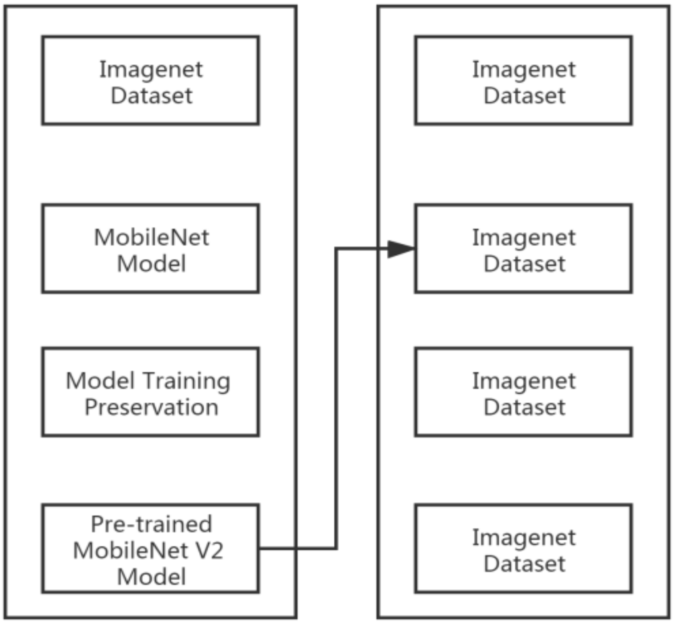

Fig. 1 Flow chart of transfer learning

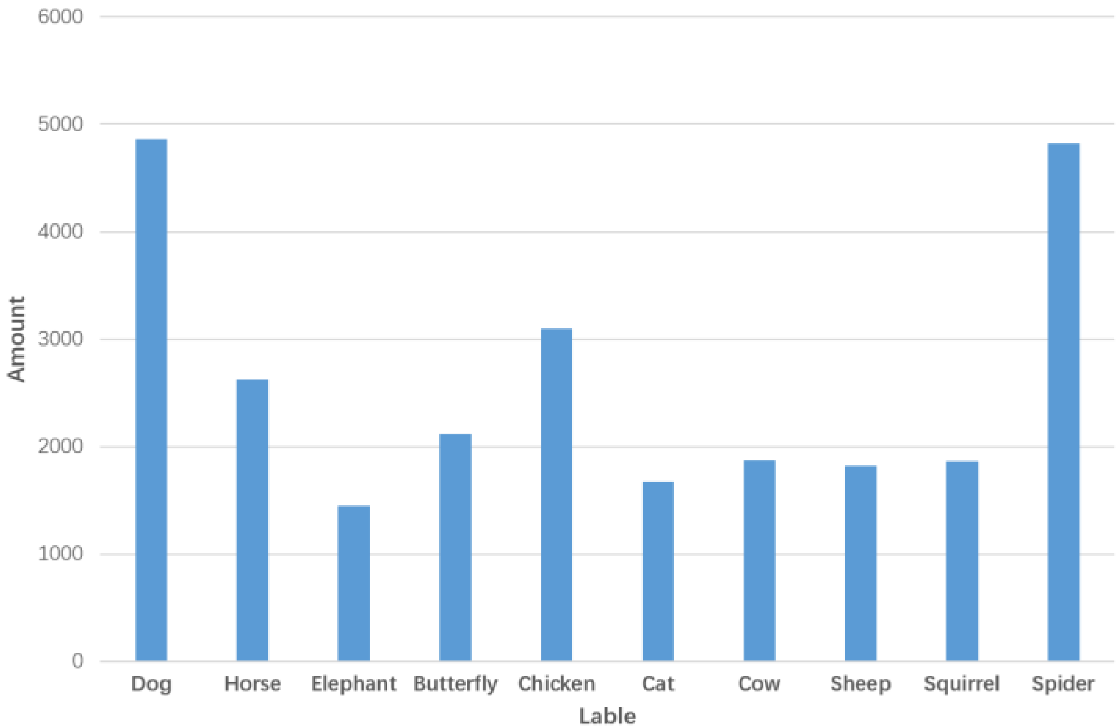

Fig. 2 Distribution of Animal-10 dataset categories

ISSN: 2414-1895 DOI: 10.6919/ICJE.202108_7(8).0019

2. Deep learning model

In the experiment, the backbone network used the MobileNet V2 pre-training model based on migration learning for feature extraction and fine-tuned the pre-training network. The transfer learning process used in the experiment is shown in Figure 1.

2.1 Data enhancement

The dataset used in the experiment is the Animals-10 dataset provided by kaggle, which consists of 28,000 images of medium quality. The Animals-10 dataset has 10 categories, namely dog, cat, horse, spider, butterfly, chicken, sheep, cow, squirrel, elephant. The data distribution of each category is shown in Figure 2. You can see that the amount of data in each category is different, with Dog having the most data and Elephant having the least data. The imbalance of data categories has an important impact on the recognition effect of the model. For this reason, we use data enhancement technology to solve such problems. Data enhancement can solve the problem of too few samples in the data or unbalanced sample categories. For the Animals-10 dataset, in the training set, this article mainly uses horizontal flip, translation, random rotation, scaling, and cropping of the angle of the image proportionally, as well as the proportional adjustment of the height and width of the image [14].

2.2 Transfer learning

The idea of transfer learning is to apply a model trained on one problem to another similar new problem, that is, to extract information from one or more source tasks, and then apply it to the target task [15]. The emergence of transfer learning enables the new generation of neural networks to be more advanced on the original basis without having to restructure the network.

Whether in unsupervised learning, supervised learning or reinforcement learning, transfer learning has a wide range of applications. With the rapid development of deep learning technology based on basic neural networks, transfer learning is gradually being applied to deep learning models. As we all know, in order to fully capture the connection and detailed information of different parts of the data, the complexity of the construction of the convolutional neural network model is getting higher and higher, and its requirements for the number of labeled datasets are getting higher and higher. At the same time, the model The amount of parameters has also been greatly increased. In transfer learning, the model used is a pre-trained model, which also shows good generalization performance for pictures outside the dry training dataset, so only the pre-trained model needs to be placed on the new dataset. By fine-tuning the parameters, a better model can be obtained while reducing the amount of calculation.

The image dataset used for training in this experiment is similar to the animal images in the ImageNet dataset used by Google to train the model, Through the migration learning model that has been trained in Google, the final fully connected layer is trained. The training task can be completed only by using the experimentally constructed image dataset and ordinary CPU, and the accuracy of the model is high.

2.3 MobileNet V2

The high computing resources required by modern advanced networks far exceed the capabilities of mobile and embedded devices. The Mobilenetv2[15] network is a new type of network structure designed for this limitation. The network can effectively reduce the amount of parameters and calculations in the network while maintaining similar accuracy. The main contribution of MobileNetv2 comes from Linear Bottlenecks and Inverted Residual block. Linear Bottlenecks is the activation function Relu after removing the layer with the smaller output dimension in the network and changing it to linear activation. This improvement reduces the loss of information caused by using the Relu function. The design of the Inverted Residual block adopts the structure of first dimension reduction and then dimension reduction, which is contrary to the traditional Residual block structure of first dimension reduction and then dimension reduction, which reduces the loss of information. At the same time, the network is designed with an expansion coefficient t to control the size of the network.

3. Experiments and results

3.1 Experimental environment

The experiment was carried out on the Ubuntu system with the system version 18.04, and the graphics card was GeForce RTX 2080 Ti. The deep learning framework used in the experiment is Tensorflow, and the language is Python.

3.2 Experiment analysis

This article uses the MobileNet V2 network to implement the classification task, which is improved on the mobileNet V1 network. The MobileNet v1 network uses a depthwise separable convolution structure, which can be divided into depthwise convolution and pointwise convolution, which can reduce the number of network parameters and the amount of calculation corresponding to model training.

The main innovations of MobileNetV2 based on V1 are as follows:

(1) Introduce the residual structure, first increase the dimension and then reduce the dimension, enhance the propagation of the gradient, and significantly reduce the memory footprint required during inference (Inverted Residuals).

(2) Remove the ReLU after the Narrow layer (low dimension or depth), retain the feature diversity, and enhance the expressive ability of the network (Linear Bottlenecks).

(3) The network is fully convolutional, so that the model can adapt to images of different sizes; the use of RELU6 (the highest output is 6) activation function makes the model more robust under low-precision calculations.

It added a new PW convolution before DW convolution to achieve dimensionality improvement. And define the coefficient of Shengwei t=6, so no matter what the number of input channels C_{in} is, after the first PW realizes the dimension upgrade, DW can achieve feature extraction in a relatively higher dimension. Because the author believes that the activation function can effectively increase the nonlinearity in the high-dimensional space, but in the low-dimensional space, it will destroy its characteristics to a certain extent, which is less effective than the linear structure. Since the second PW is mainly to achieve dimensionality reduction, the author removed its activation function.

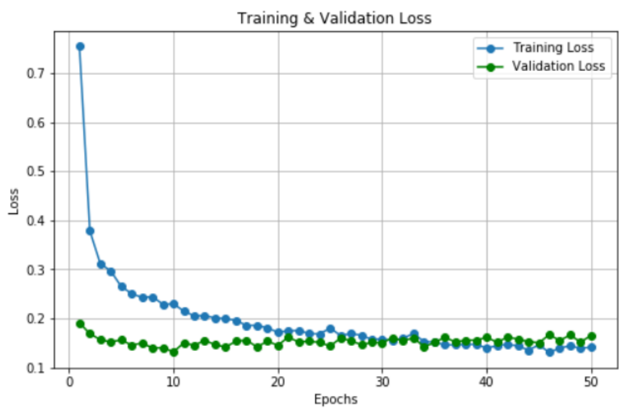

Fig. 3 Model loss

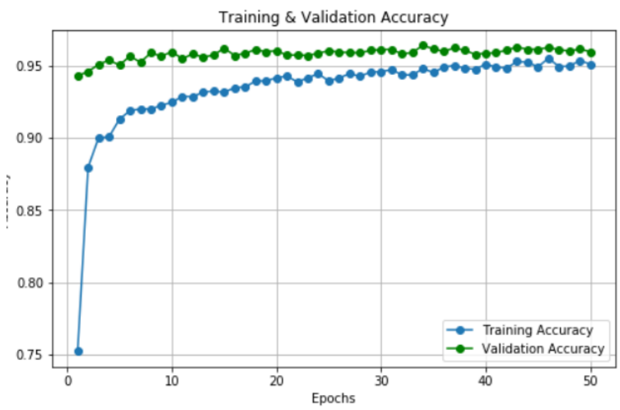

Fig. 4 Model accuracy

3.3 Experiment results

In the experiment, the MobileNet V2 pre-training model is applied to the target dataset for feature extraction, and the predicted category is changed to the category required in the experiment. Then use layer freezing, Dropout and other technologies to adjust the parameters of the migration training model. Finally, the network is fine-tuned to unfreeze some layers, and the extracted classification features are more suitable for the target dataset, which further improves the classification effect. In

ISSN: 2414-1895 DOI: 10.6919/ICJE.202108_7(8).0019

the experiment, the softmax function is used as the classification function, the loss function uses the cross-entropy loss function, and the optimizer uses Adam. The experimental results are shown in Figure 3 and 4. It can be seen that after 40 trainings, the accuracy of the model is basically balanced. After 50 iterations of training, the final test set accuracy rate is 95.16%.

It can be seen that the model in this paper has a good effect on animal classification tasks, and its accuracy rate has reached 95.16%. The classification effect diagram is shown in Figure 5:

Fig. 5 Experimental classification effect diagram

4. Conclusion

New learning for traditional construction of convolutional neural network models to train classification models from scratch requires large data sets, and the computational process requires a lot of computational resources and takes a lot of time. In this paper, we use transfer learning methods to train a model for animal image classification. The results show that transfer learning can take advantage of features learned from large-scale data, as opposed to brand-new learning. The convergence of the neural network is significantly accelerated, improving the classification performance and saving a lot of time. In this experiment, the data set comes from a competition on kaggle, because the images in the data set are relatively single, the accuracy of the experiment reached 95.16%, in later experiments, more data sets will be used to test the model, and will try to add the target detection function to the model, and further improve the recognition rate and practicality of the model.

References

[1] Xing Yan fang, Duan Hongxiu, He Guangwei. Application of TensorFlow in Image Recognition System [J]. Computer Technology and Development, 2019, 29(5): 192-196.

[2] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]/ /2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE, 2016: 770-778.

[3] KRIZHEVSKY A, SUTSKEVER I, HINTON G. ImageNet classification with deep convolutional neural networks[C]// International Conference on Neural Information Processing Systems, Dec 3-6, 2012, Lake Tahoc, Nevada. Red Hook: Curran Associates Inc,2012: 1097-1105.

[4] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv: 1409. 1556. 2014.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202108_7(8).0019

[5] SZEGEDY C,LIU W, JIA Y, et al. Going deeper with convolutions [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015 : 1-9.

[6] HEK, ZANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.

[7] HEK, ZHANG X, REN S, et al. Identity mappings in deep residual networks[C]//4th European Conference on Computer Vision, Oct 08-16, 2016, Amsterdam, Netherlands. Heidelberg: Springer, 2016: 630-645.

[8] VEIT A, WILBER M, BELONGIE S. Residual networks behave like ensembles of relatively shallow networks[C]//Advances in Neural Information Processing Systems, 2016:550-558.

[9] Huang Xueyu,Guo Qin. Game algorithm integrating environment model and deep reinforcement learning[J]. Journal of Jiangxi University of Science and Technology, 2018, 39 (3), 84-89.

[10] Wang Zheng, He Wen. The application of deep residual network in the diagnosis of pneumoconiosis[J]. Chinese Journal of Industrial Medicine, 2019,32 (1):31-33.

[11] ZHANG X, ZHOU X, LIN M, et al. ShuffleNet; an extremely efficient convolutional neural network for mobile devices [C]//Proceedings of the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE Press, 2018:6848-6856.

[12] FRANCOIS C. Xception: deep learning with depthwise separable convolutions[C]//CVPR 2017: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Hawaii, Jul 21-26, 2017. United States: IEEE, 2017: 1251-1258.

[13] SANDLER M, HOWARD A, ZHU M, et al. MobileNetV2: inverted residuals and linear bottlenecks[C] //The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018: 4510-4520.

[14] Zhong, Z, Zheng, L, Kang, G, Li, S, Yang, Y: Random erasing data augmentation. CoRR (2017)

[15] Cao Xiaojie, Mo Lao, Yan Yuling. Convolution using transfer learning Neural network flower image recognition[J] Computer Application and Software Pieces, 2020, 37(8): 142-148.