Research on Small Sample Image Recognition Based on Transfer Learning

Guhai Yang\textsuperscript{1,3}, Baihui Tang\textsuperscript{2,4,6} and Sanxing Cao\textsuperscript{2,5,7}

\textsuperscript{1}Department of Integrated Circuit Engineering, Communication University of China, Beijing, China.
\textsuperscript{2}Key Laboratory of Intelligent Fusion Media Ministry of Education, Communication University of China, Beijing, China.
\textsuperscript{3}Email: 15801069668@163.com
\textsuperscript{4}Email:tangbaihuiyt@163.com
\textsuperscript{5}Email:caosx_cuc@126.com
\textsuperscript{6}Second author
\textsuperscript{7}Corresponding author

Abstract. With the improvement of electronic hardware configuration performance, high-speed computing of large and complex data can be realized, and artificial intelligence, which requires a large amount of data calculation has been rapidly developed. Nowadays, artificial intelligence has developed rapidly in image recognition, natural language processing and speech recognition, and also has achieved good results in these areas. Among them, the convolutional neural networks plays an important role in image recognition. Some excellent convolutional neural networks such as VggNet, GoogleNet and ResNet have reached more than 90% in image recognition accuracy. But these convolutional neural network models are based on big data, consume a lot of computing time, which is very costly. Therefore, when data is limited, hardware configuration is low, and time is tight, a research on small sample image recognition based on transfer learning becomes very necessary. In this paper, the inception-V3 model is used to transfer learning. The training nodes of the model are skillfully used to train the model on the data set of ten kinds of fruits. The model is tested by the training model and good accuracy is obtained. The advantage of the experiment is that the training time is short, the required data set is small, and the image recognition accuracy is high.

1. Neural Network and Transfer Learning

This chapter will introduce the current mainstream concepts of convolutional neural networks and transfer learning detailedly.

1.1. Convolutional Neural Network

With the development of convolutional neural network, it has become a research hotspot in image processing, classification and recognition. At present, the mainstream convolution neural network is mainly composed of convolution layer, pooling layer and full connection layer. By orderly linking up these layers, a complete convolutional neural network can be built. Convolutional neural networks use local connection and parameter sharing to reduce the number of training parameters in the network, so as to reduce training time and improve the efficiency. The following are some commonly used convolutional neural networks.
1.1.1 VggNet. In ILSVRC 2014, Karen Simonyan and Andrew Zisserman proposed VggNet. Compared to earlier neural networks, the model improves the processing power of images by increasing the number of layers in the network.

1.1.2 GoogleNet. GoogleNet was proposed by Christian Szegedy in 2014. It is a new deep learning structure. Unlike AlexNet and VGG, the inception model can extract more features under the same computing power by using computing resources efficiently, and improve the training effect. Figure 1 shows the inception model structure of GoogleNet.

1.1.3 ResNet. Kaiming He, Xiangyu Zhang, Shaoqing Ren and Jian Sun proposed ResNet. The ResNet model is made up of residual blocks like Figure 2(a). The residual block is shown in Figure 2(b). Each residual block of ResNet consists of one 3 × 3 convolution kernel and two 1 × 1 convolutions. The purpose is to reduce the number of parameters. The first 1×1 convolution reduces the number of 256-dimensional channels to 64 dimensions, and then recovers with a 1×1 convolution in the end. As the number of layers increases, the amount of computation is much less than other networks. In the general network, there is only the part that don’t have curve in Figure 2(a), and the core of the residual network lies in this curve. The final result is \( H(x) \), and it can be seen from Figure 2(a) that \( H(x) = F(x) + x \), which guarantees that the gradient disappears during the gradient propagation. Then \( F(x) = H(x) - x \) is a residual value, so the network is called residual network.

1.2. Transfer Learning

The essence of transfer learning is to apply the model or method of solving a problem to a new problem. In the neural network, transfer learning is to apply a well-trained network node to a new network structure through transfer, thus completing a new task, which greatly reduces the number of network layers that need training, and also the training cost and time are greatly reduced. The value of transfer learning is that it can use existing knowledge achievements to complete a great deal of work, and it does not need to spend a lot of manpower and material resources to obtain huge new data sets. The most important thing is that it has timeliness, and can use the original trained model to efficiently and quickly train the new model to solve new problems. According to the method of transfer learning, it can be roughly divided into four categories. The first is sample transferring, which centralized...
search and target similar data in the original data set, enlarges the similar data and matches the target data, and gives more weight to the important data. In this way, when predicting the target object, the data similar to the target has a greater probability to become the prediction result, which enhances the accuracy of the prediction. Sample transferring is characterized by the need to weigh different examples and to train data. The second is feature transferring. It finds the representative features of the target data, makes its distribution consistent with the distribution of the data with the same characteristics in the original data through mapping transformation, and then for machine learning. The third is model transferring, which reduces the training steps by sharing the node parameters of the trained network to the target network. The fourth category is relationship transferring, which transfers similar relationships, such as the relationship between teachers and students to the relationship between superiors and subordinates.

2. Design and Analysis of Experiments
This chapter mainly introduces the experimental environment, then analyses and sets the appropriate parameters, then trains the model, records the experimental results and analyses them, and draws the experimental conclusions based on the analysis.

2.1. Experiment Setup

2.1.1 Experimental environment. The experiment uses Windows 10 system, Python 3.7 scripting language, AMD2600X processor, Intel 760P256G solid-state hard disk and 3200Hz 8G memory.

2.1.2 Experimental principle. The trained inception—V3 model is used in the experiment. The principle of the inception—V3 model is to take the output of the image in the bottleneck layer as the feature of the image, and then pass these features through a single-layer full-connected layer neural network, so that 1000 kinds of images can be distinguished well. Therefore, the output vector of any image in the bottleneck layer can be used as the feature vector of the image. In this experiment, the trained image is processed by the trained inception—V3 model, and the output vector of the image in the bottleneck layer is used as the feature vector of the image. Then the extracted feature vector is trained by a single-layer fully connected neural network and the image is classified. Generally, the effect of this transfer learning is slightly worse than the result of complete retraining, but it needs less data, less training steps, lower hardware configuration requirements but the image recognition accuracy can reach more than 90%.

2.1.3 Experimental data. The experimental data were collected from the Internet by web crawlers according to keywords and screened more than 5000 pictures of 10 kinds of fruits. The format of the pictures was jpg. The similarity of their characteristics was low. They were apple, banana, blueberry, pitaya, grape, litchi, mango, orange, strawberry and kiwifruit. The experimental data were divided into six groups according to the number of samples in the data set. The six groups of data contained 50, 100, 200, 300, 400 and 500 samples of each fruit, respectively. The experiment divides the data set into training set, validation set and test set. The training set is a data sample for model fitting. The validation set is a separate sample set in the process of model training. It can be used to adjust the model's super parameters and to evaluate the model's ability preliminarily. The test set is used to evaluate the generalization ability of the final model, but it cannot be used as a call of the selection of features and other related selection criteria. In the experiment, the training set accounted for 80%, and the verification set and the test set each accounted for 10%. In addition, the experiment adds 1000 sets of application set images, each of them contains 100 sheets, which is used to test the accuracy of the trained model in practical applications.

2.2. Model Training

2.2.1 Training parameter settings. Batch_size settings cannot be too large or too small, too large will increase memory occupancy, run a complete data set of iterations to reduce the number of times
needed to achieve the same accuracy increased, too small will make the network training time increased and unsuitable for memory. After many experiments, the batch_size was set to 100. The appropriate batch_size needs relatively small memory to train, and the training network is fast.

2.2.2 The setting of learning rate. The learning rate set in the experiment does not adopt the traditional fixed and invariable learning rate, but adopts the learning rate with better effect of exponential attenuation. The significance of the exponential attenuation of learning rate is that it can dynamically adjust the learning rate. The larger learning rate at the initial stage of training can accelerate the convergence time of the model. When the convergence of the model reaches a certain degree, the larger learning rate can be achieved. The learning rate no longer improves the accuracy of the model, which may cause the parameters to move back and forth on both sides of the optimal value, but reduces the optimization speed. Therefore, reducing the learning rate at this time can make the model more stable and improve the efficiency. The exponential attenuation formula is shown in Formula (1), where M is the learning rate used in each round of optimization, N is the initial learning rate, A is the attenuation coefficient, B is the current training round, and C is the attenuation speed.

\[ M = N \times A^\frac{B}{C} \]  

2.2.3 Training steps. After many experiments, the training accuracy of 500, 500, 1000, 1000, 1000, 1000, 1000 and 2000 is the best when the sample number was 50, 100, 200, 300, 400 and 500. Figure 3 shows a broken line chart of training accuracy and loss function varying with step number when the sample number was 50, 200 and 500. The smoothing value was set to 0.6, which makes the observation better. From Figure 3, it can be seen that the model fitting near the step number set by the experiment achieves the best, the highest precision and the lowest loss function value.

![Figure 3. Training accuracy and loss function breakdown diagram](image)

2.2.4 Experimental steps. The experiment is divided into training stage and testing stage. Training stage: According to the set parameters, the code of model training was compiled, and then the model was trained for different sample data sets, and the trained model and its corresponding labels were saved. Application testing stage: Import the trained model, load the image of application set, use the trained model to get the feature vector of the image, judge the label according to the feature vector, record the correct or wrong judgment according to the output label, and record the recognition accuracy after all the test set image recognition. The accuracy of the trained model in the verification set and the test set was obtained in the training step, while the accuracy of the application test set was obtained in the application test stage.
3. Experimental Results and Analysis

According to the experimental results, the accuracy of the training model on the verification set, test set and application set were recorded, and the broken-line graph as shown in the Figure 4 was obtained.

3.1. Accuracy Comparison of Models on Different Data Sets

As can be seen from Figure 4, under the same number of experimental samples, the model has the highest accuracy on verification set, the lowest accuracy on Application set, and the accuracy on test set lies between them. This is because the model’s hyperparameters are adjusted according to its accuracy on verification set, so the model has the highest accuracy on verification set, and the test set acts as generalization energy on model. In force evaluation, the accuracy of the model is generally less than that of the verification set. The application set is used to detect the accuracy of the model in the actual application process. It is used to evaluate the practical application performance of the model. The accuracy of the model is generally less than that of the former two because of the quality and quantity of the image in the test set. When the parameters of the model are set properly, the accuracy of the trained model on the three data sets is not much different.

3.2. The Accuracy of the Model on the Same Data Set Varies with the Number of Training Steps

The green polyline in Figure 4 shows that the accuracy of the model on the verification set varies with the number of samples, and the values are between 0.98 and 1. It shows that the accuracy of the model on the verification set is less affected by the number of experimental samples. Blue line chart shows that the accuracy of the model in the experimental equipment with the change number of the samples changed. The values fluctuate between 0.97 and 0.99 and the fluctuation range is small. It shows that the number of samples has little influence on the accuracy of the model on the test set. The yellow polygraph is the situation that the accuracy of the model varies with the number of samples, and the yellow virtual line is its overall trend. Combining the two, we can see that with the increase of the number of samples, the accuracy of the model in practical application will be improved, but the increase is not large. When the training sample is small, the accuracy of the training model will only slightly decrease. It shows that the model obtained from small sample training in transfer learning still has good performance in practical application. The experimental results show that small samples can still obtain better training results and test results in the transfer learning, while achieving good recognition rate, saving training costs and timeliness.

Figure 4. Experimental results
4. Summary
With the application of convolutional neural network in the field of image recognition, the accuracy of image recognition is getting higher and higher. Most of the existing convolutional neural network models are based on a large number of data and training time, and are not suitable for image recognition in specific areas. This paper studies a small sample recognition model based on transfer learning. The experiment migrates the inception—V3 model and trains the experimental data of fruit images with 50, 100, 200, 300, 400 and 500 samples respectively. The accuracy of the model in the test set and application set is above 94%, and with the decrease of the number of samples, the accuracy does not decline significantly. This kind of transfer learning, which only needs less samples and training time and does not cost a lot of accuracy loss, is very suitable for image recognition in specific areas, and can obtain better accuracy.

5. Acknowledgments
This research project is supported by the Interdisciplinary Team Project of the "Double First-Class" Construction Project of the Communication University of China, which is named "Virtual Reality of New Media Independent Innovation Capability Improvement and Application Promotion" (No. YLTS180522).

6. References
[1] Wang Shuang. Research and Implementation of Flower Recognition Algorithms Based on Machine Learning [D]. University of Electronic Science and Technology, 2018.
[2] Zhang Cedar, solemn, Yan Fei, Wang Wei. Research and progress of classification-level object recognition and detection based on transfer learning [J/OL]. Journal of Automation: 1-20 [2019-05-06], https://doi.org/10.16383/j.aas.c180093.
[3] Wang Keli, Yuan Hongchun. Aquatic Animal Image Recognition Method Based on Transfer Learning [J]. Computer Applications, 2018, 38 (05): 1304-1308+1326.
[4] Yan Bin, Zhou Peng, Yan Li. Recognition of crop diseases with small samples based on transfer learning [J]. Modern Agricultural Science and Technology, 2019 (06): 87-89.
[5] Li Quanjie. Research on Vehicle Logo Recognition Method Based on Transfer Learning and Deep Convolution Neural Network [D]. Guangdong University of Technology, 2018.
[6] Wu Di, Liu Xiuli, Hou Lingyan, Liu Xuhong and Li Hongchen. Flower image classification based on saliency detection and transfer learning [J]. Journal of Beijing University of Information Science and Technology (Natural Science Edition), 2019, 34 (01): 55-63.
[7] Li Nianhua. Research on Key Problems of Transfer Learning of Deep Neural Networks [D]. University of Electronic Science and Technology, 2018.
[8] Jinzhao. Contrast and Analysis of Different Deep Convolutional Neural Networks Based on TensorFlow [J]. Electronic World, 2018 (06): 25-26.
[9] Majie, Zhang Xiudan, Yang Nan, Tian Yalei. Gesture Recognition Method Combining Dense Convolution and Space Conversion Network [J]. Journal of Electronics and Information, 2018, 40(04): 951-956.
[10] Liu Yicheng. Food Image Classification Based on Transfer Learning [J]. Information Communication, 2018(12):103-104.