Research on Image Classification Algorithm Based on Convolutional Neural Network-TensorFlow

Jiantao Zhao¹, Lili Li¹*

¹School of Control and Computer Engineering, North China Electric Power University, Beijing 102206, China

*lilili@eplian.com

Abstract. Image Classification is one of the important processing methods in the field of image processing. Traditional Classification algorithms, such as KNN (K-NearestNeighbor), have been unable to meet the accuracy standards of image Classification. And convolutional neural networks (CNN) are increasingly used to solve the problem of image classification. This paper briefly ces the convolutional neural network, which is constructed using the TensorFlow deep learning framework, and conducts experiments on 2000 pieces of custom data set of 10 classification obtained by crawler technology. At the same time, a comparative experiment is conducted on the traditional image classification KNN algorithm. Experimental results show that the application of convolutional neural networks in image classification has relatively large advantages, and the accuracy has been greatly improved.

1. Introduction
Image classification technology as the focus of the field of vision technology exploration and research, scholars put their energy into the study of these image classification algorithms. In recent years, the convolutional neural network has attracted more and more attention due to its outstanding feature extraction ability and outstanding effect in the application of image classification. Compared with other traditional methods, Convolutional Neural Network has a higher ability of feature extraction and representation in the database. Using the designed network structure, the neural network can not only learn the underlying features of the image, but also easily learn the high-level semantic abstract features. At the same time, its learning is end-to-end, that is, the original data without preprocessing can be input into the convolutional neural network to get the final result, and the efficiency can be greatly improved. Compared with traditional machine learning, the advantages of image classification algorithm implemented by convolutional neural network are very prominent.

2. Research theory and method
For image classification problem, first of all, there should be a fixed set of classification tags containing K classes and a set of n images, and the label of each picture in the set belongs to one of the known classification tags. Then, the image set should be input to the classifier for learning. After learning, input an image without classification label to the classifier, and ask it to find a certain classification label from the known collection of classification labels and assign it to the input image. The category label predicted by the image classifier should be consistent with the real category label of the image as much as possible, which also represents the better training of the classifier.
2.1 The traditional image classification algorithm KNN
Common classification algorithms include KNN, SVM classification algorithm, etc. The basic idea of KNN is to first input a picture whose category is unknown, first find the k neighbors closest to the picture, and then select the category that has the most frequency of occurrence among the first k closest neighbors, and take the result as a prediction category of the input picture[1].

2.2 Overview of Convolutional Neural Network
Convolutional neural network has a powerful ability to extract image features, and researchers use convolutional neural network to take the original image as input to carry out image classification.

The network model is mainly composed of input layer, convolution layer, pooling layer, full connection layer and output layer[2]. The convolutional layer and the pooling layer are connected as sub-modules. The network connects several sub-modules together to form an important core module, while the full-connection layer is the high-level part of the network.

2.2.1 Convolutional Layer. The convolution layer is used to extract the characteristic information of the input image. Each convolution unit includes forward propagation and backward feedback: signal is transmitted forward through the forward propagation process, and weight matrix is updated through learning parameters in the backward feedback process to optimize the model. Feature information is extracted by moving the input image regularly through the receptive field and convolution operation with the corresponding block. Low-level convolution can only extract low-level features such as color and contour, while higher-level convolution can extract deeper semantic information about images.

In this paper, the ReLU function is used to perform nonlinear operation on the image. The analytic expression of ReLU function is as follows:

\[
\text{ReLU} = \max(0, x)
\]

The ReLU function is really just a function that takes the maximum value, as shown in Figure 1.

![ReLU function](image1.png)

ReLU function can perfectly deal with the problem of the disappearance of a gradient in the process of calculation of convolution. Compared with sigmoid and tanh function, its calculation speed and its convergence rate are far higher than those of the first two.

2.2.2 Pooling layers. The pooling layer is fixed behind the activated convolutional layer and between successive convolutional layers. Its purpose is to reduce the size and dimension of the feature graph, simplify the complexity of network computation, and thus improve the robustness of the feature and the computing power of the network. Not only that, but pooling layers can also reduce overfitting. Max pooling and average pooling are two common pooling functions[3], as shown in Figure 2.
Fig. 2 Diagram of pooling operation

In Figure 2, \(2 \times 2\) convolution kernel is used, the step size is set to 2, and no padding is performed. As can be seen from the figure, maximum pooling is to find the maximum value in each convolution process. The mean pooling is to calculate the average value in each convolution region to obtain the main features. In the pooling operation, the most commonly used convolution kernel size is 2 and the step size is 2, thus reducing the feature graph.

2.2.3 Fully connected layers. The full connection layer is located at the end of the convolutional neural network, which is, the last layer, which outputs the final classification result, which is equivalent to a classifier. In the full connection layer, the feature map will lose its original spatial structure, and the feature map will be expanded into feature vectors, and the features calculated at the previous layer will be output through nonlinear combination. The calculation formula is formula (2) as follows.

\[
f(x) = W \times x + b
\]

In formula (2), \(W\) is the weight coefficient, \(x\) represents the input of the full connection layer, and \(b\) is the bias. The full connection layer connects all the features calculated by the previous layers and carries out nonlinear combination to get the output results, which are input to the output layer. For different problems, the output layer has different meanings. Taking the image classification problem as an example, the output layer uses normalized exponential function (softmax) or logical to output the results, and the attribute of the output is category label.

2.3 TensorFlow framework

TensorFlow framework is one of the widely used deep learning frameworks, which can simplify the process of implementing deep learning models and help solve the technical problem that deep learning is too large to operate. TensorFlow, an interface and framework for deep learning algorithms, can simplify the development process and reduce the difficulty of implementation[4]. With it, users can easily and quickly design neural network models and train on data sets with large amounts of data.

Users operate on branches of TensorFlow program (loop operation or conditional control) through data operations in programming languages such as Python. A variable is a class in TensorFlow that can be used to store parameters in a machine learning model and to update the parameters. In the neural network, variables are represented as weight and bias, which need to be initialized before use. The trained model is stored on the disk, and the following model analysis can directly load it for use.

3. Neural network construction

3.1 Data overview and presentation

The data set used in this paper is self-collected data of custom classification, which is obtained through crawler technology. The dataset includes 2000 images of 10 categories. The correspondence between
the image label and the actual category is shown in Table 1. Each category corresponds to a folder named with the tag name, and each folder contains 200 pictures respectively.

| label | type  | label | type  |
|-------|-------|-------|-------|
| 0     | apple | 5     | banana|
| 1     | dog   | 6     | flower|
| 2     | cat   | 7     | panda |
| 3     | truck | 8     | people|
| 4     | orange| 9     | elephant|

3.2 Build neural network model

Before the training images are input into the model, the images need to be preprocessed the size of the 32×32×3. The neural network model is built using the TensorFlow. The convolution block can be divided into two Conv0 and Conv1 small convolution blocks, each of which contains a convolution operation, a batch normalized BN and an activation operation. The activation operation uses ReLU function. The pooling functions used are maximum pooling, which are connected after each convolution block of the image so as to retain the most significant features. The step size is set to 2, thus reducing the size of the image. The full connection layer connects the data into a feature vector with a length of 400 after the last maximum pooling layer. More specifically, the Flatten function is used before the full connection layer, in order to flatten the input data, while the ReLU activation function is still selected for nonlinearization by the full connection layer. In order to reduce the over-fitting problem of the model, the Dropout layer is used after the full connection layer, and the set is 0.25. Logits is the output layer, including ten numbers from 0 to 9, and the final output is the number of categories of the image. At the same time, Adam was used as the optimizer, cross-entropy was used to define the loss, accuracy was used as the indicator, and learning efficiency was set at 1e-4. Specific parameters of the model are shown in Table 2.

| network structure | type                  | Size of convolution kernel | output size   |
|-------------------|-----------------------|----------------------------|---------------|
| Input layer       | The data layer        | 32×32×3                   |               |
| Conv0             | Convolution layer     | 5×5                       | 28×28×20      |
| pool0             | Maximum pooled layer  | 2×2                       | 14×14×20      |
| Conv1             | Convolution layer     | 4×4                       | 11×11×40      |
| pool1             | Maximum pooled layer  | 2×2                       | 5×5×40        |
| FC                | Connection layer      | 1×1×400                   |               |
| Output layer      |                       | 1×1×10                    |               |

4. Experimental results and analysis

4.1 Traditional image classification algorithm based on KNN

The traditional image classification algorithm KNN is selected to carry out image classification experiments on the data set. KNN algorithm is powerful and easy to implement, and we mainly use Sklearn to implement it. Firstly, 2000 images are randomly divided according to the proportion of 80% in the training set and 20% in the test set. Finally, the algorithm evaluation is shown in Figure 3, including Precision, recall rate and F1-score. It can be seen that the average accuracy of KNN algorithm is only 0.46, the average recall rate is 0.46, and the average F value is 0.41. Clearly, the results are not good.
4.2 image Classification Algorithm Based on Convolutional Neural Network

The data set contains a total of 2000 pictures of 10 categories. 80% of them are randomly selected as the training set and the remaining 20% as the test set, so as to carry out the final result test. After the experiment, the loss change curve of the model in Figure 4 on the training set and test set can be obtained. It can be seen from the curve that the loss rate gradually decreases to a flat level when the training times increase. After the 160th training session, the loss rate on the training set can reach 16%, while the loss rate on the test set can reach around 36.

After the training, 400 pictures were tested. The final prediction was correct for 356 pictures, with an accuracy of 89%. The test accuracy of each category is shown in Figure 5. The accuracy rate for dogs was 95.6 percent, while the accuracy rate for trucks was 83.8 percent. Compared with the previous use of the traditional image classification algorithm KNN accuracy of 46% is still a very large improvement.
5. Conclusion

Image classification algorithms play an important role in various computer vision problems.

Traditional image classification algorithms based on machine learning have been unable to meet people's accuracy requirements. The emergence of convolutional neural network shows it has the ability to learn more discriminative and representative features.

In this paper, the convolutional neural network is used to study the image classification algorithm, and the Keras API of TensorFlow is used to build the convolutional neural network model to realize the image classification and recognition of 2000 images in the custom 10 types. In comparison with the traditional image classification KNN, it can be see, the accuracy of image classification algorithm based on neural network has been greatly improved.

Acknowledgments

This work was supported by the Institute of Network and Information Systems of North China Electric Power University.

References

[1] Shokrzade Amin, Ramezani Mohsen, Akhlaghian Tab Fardin, Abdulla Mohammad Mahmoud. A novel extreme learning machine based kNN classification method for dealing with big data[J]. Expert Systems With Applications, 2021, 183.

[2] Li Shaojuan, Wang Lizhi, Li Jia, Yao Yuan. image Classification Algorithm Based on Improved AlexNet[J]. Journal of Physics: Conference Series, 2021, 1813(1).

[3] Huang Zhiwen, Zhou Quan, Zhu Xingxing, Zhang Xuming. Batch Similarity Based Triplet Loss Assembled into Light-Weighted Convolutional Neural Networks for Medical image Classification[J]. Sensors, 2021, 21(3).

[4] Osah Samuel, Acheampong Akwasi A, Fosu Collins, Dadzie Isaac. Deep learning model for predicting daily IGS zenith tropospheric delays in West Africa using TensorFlow and Keras[J]. Advances in Space Research, 2021, 68(3).