Research on Image Recognition Technology Based on Convolution Neural Network

Wang Jinghe

Hong Kong University of Science and Technology, Hong Kong, China

Keywords: Image recognition, Accuracy rate, Convolution neural network

Abstract: In this paper, the convolution neural network structure is used for image recognition. The initialized image is convoluted with the convolution core in the convolution layer to extract the features of the image. The extracted image features are compressed by the pooling layer by the full connection layer. The convolution neural network iterates the image for several times and compresses the image layer by layer to recognize the image and output the recognized image. Experiments show that compared with BP neural network, the network structure can significantly improve the accuracy of image recognition.

1. Introduction

In the fields of pattern recognition, computer vision, artificial intelligence, machine learning and in-depth learning, image processing and recognition technology has been the first problem that researchers began to try and explore [1]. In the traditional image processing technology, it is necessary to pre-process the image first, and then complete the information recognition through a series of manual operations. The whole process is very difficult, and it is difficult to meet the needs of the future society. Therefore, we must find an image recognition method that can adapt to the current work to better meet the needs of society. Convolutional neural network is an efficient and accurate classification technology, which is a modern deep learning method developed on the basis of multi-layer neural network. Convolutional neural network is proposed based on human visual feature mechanism. Its main feature is that it has local weight sharing, which can significantly reduce the difficulty of information processing. The most important part of convolution layer of convolution neural network is convolution core. The convolution nucleus can not only reduce the number of connections between neurons, but also reduce the over-fitting error. The sub-sampling process is the pooling process. The convolution process is to convolute the convolution core and the pre-test image. Subsampling can simplify the network model, reduce the complexity of the network model, and thus reduce the parameters. In image recognition, the input image needs to be initialized first, then the initialized image is convoluted and sampled, and fed back to the full connection layer. The output layer can be calculated by transformation. Finally, the linear regression between feature enhancement effect and logic can be used to judge whether the desired effect of image recognition meets the expectation of the image recognition, and the cycle is reciprocating. The algorithm iterates one after another to recognize the image [2].

2. Structure of Convolution Neural Network

2.1 Convolution Layer.

In convolution layer, the feature map connected between layers is extracted by convolution operation from one or more convolution cores. The result of convolution operation is transformed by activation function mapping to form the feature mapping relationship from input to output. Each convolution core traverses the whole feature map using a mechanism similar to sliding window. The convolution core aggregates and fuses the feature information of each small area to complete the characterization of a small local area in the image. The biggest characteristics of convolution layer are local perception and weight sharing. Traditional neural networks have strong expressive ability.

  DOI: 10.25236/iwmecs.2019.028

Two-layer neural networks can approximate all functions, but it is difficult to train. For example, in the field of image, each pixel of image is connected with each neuron, and the number of weight parameters to be learned is very large, so it is difficult to train. Convolution layer is proposed to extract the same features in each region of the image using convolution kernel parameters, which only include the number of parameters of the convolution core. To extract multiple features, we use multi-convolution kernels. The same parameters are used to extract features in different regions of the image. These features are some local features. In addition, the same features can be extracted from the same or similar regions in different locations of the image to ensure the robustness of the features. Generally speaking, in the case of 2-D images, the convolution layer is as follows:

$$z(x, y) = f(x, y) * g(x, y) = m \sum_{m} f(x - m, y - n) g(m, n)$$

Among them, f stands for input, usually 2-D image and g stands for convolution core. M and n are the dimensions of convolution core respectively.

2.2 Activation Layer.

Beginning with the early artificial neural network, the mapping relationship between interconnected ganglion points is established from input to output by activation function. Activation layer mainly sets up activation function in convolution network. Its essence is a kind of function mapping. It maps and transforms input data to provide the ability of non-linear modeling of network [3]. In the process of calculation, the size of the original data is not changed by element calculation, that is, the size of the input and output data is equal. In real environment, deep learning model is used to fit real data, and the data itself often shows the characteristics of non-linear distribution. These non-linear activation functions enable deep neural network to have the ability of layered non-linear mapping learning. Therefore, the activation function in the network is an important part of the deep neural network. Activation function is used to add non-linear factors, because the expression of linear model is not enough. We know that in the neural network, we mainly use convolution to process images, that is, to assign a weight to each pixel, this operation is obviously linear. But for our samples, it is not necessarily linear separable. To solve this problem, we can make linear changes, or we can introduce non-linear factors to solve problems that linear models cannot solve. Logistic-Sigmoid function and Tanh-Sigmoid hyperbolic tangent function are the most commonly used activation functions in traditional neural networks. Their function forms are as follows:

$$f(x) = \frac{1}{1 + e^{-x}}$$

$$\tanh(x) = \frac{1 + e^{-2x}}{1 + e^{2x}}$$

2.3 Pooling Layer.

Generally, convolution neural networks often periodically insert pooling layers between successive convolution layers. Based on the recognition that image features useful to one region are likely to be equally applicable to another region, the pooling layer combines semantically similar features, reduces the feature vectors of convolution layer output by pooling operation, and prevents over-fitting. The pooling unit calculates the value of a local block in the feature graph, and the adjacent pooling unit reads data from a small area by moving a row or column, which reduces the dimension of data expression, ensures the translation invariance of data, and greatly reduces the number of parameters and the amount of computation in the network. The maximum value of the image area is calculated by maximizing the pooling operation as the result of the pooling of the area, while the average value of the image area is calculated as the result of the pooling of the area, and the response value of the random pooling method is selected randomly according to the size of the probability matrix. In convolution neural networks, we often encounter pooling operations, and the pooling layer is behind the convolution layer, which reduces the output eigenvectors of the convolution layer and improves the results. Because the image has a static property, it means that useful features in one image area are likely to be applied in another area as well. Therefore, in order to describe large images, it is a natural idea to aggregate statistics of features in different locations [4].

2.4 Fully Connection Layer.

Full connection layer mainly collates and merges information. The input of full connection layer is the output of convolution layer and pooling layer. Among the visual features, the features such as the color of the nearest point are the most similar, and the pixels are the same. Full connection layer appears at the end of the network structure, which is a traditional multi-layer perceptron network. Each neuron in the full connective layer is fully connected to each neuron in the previous layer. This is also the source of full connection layer names. The final output of the network is the high-level features of the input image, and then the probability of the corresponding class labels of the input image is calculated by classifier statistics. Regression classification model is often used as the last layer of the full connection layer, and the output value is the probability of each category between 0 and 1. In a fully connected network, the feature maps of all two-dimensional images are spliced into one-dimensional features as input of the fully connected network. The output of the fully connected layer can be obtained by summing the weighted inputs and by responding to the activation function.

$$x_j^l = f(u_j^l)$$

$$u^l = w^l x^{l-1} + b^l$$

Among them, $u^l$is called net activation of full connection layer l, which is obtained by weighting and biasing the output characteristic graph x^{l-1} of the previous layer. w^l is the weight coefficient of the fully connected network and $b^l$is the bias term of the full connection layer l.

3. Image Recognition Technology Based on Convolution Neural Network

3.1 Problem Description.

License plate recognition is an important part of intelligent transportation and even intelligent city construction. General license plate recognition technology mainly includes vehicle detection, license plate detection, license plate character recognition and other important links. As the last step of the whole license plate recognition task, the recognition accuracy directly affects the performance of the whole license plate recognition system. In the real license plate recognition system, due to the influence of practical application environment, compared with other optical character recognition tasks, license plate character recognition in natural scenes has its difficulties. Character datasets are small in size and size. In our country, license plate characters mainly include Chinese characters, English characters and Arabic numeric characters. The collected license plate characters have lower pixels, less effective feature information and smaller data scale to participate in the practical application of the algorithm. Character image quality is greatly affected by the environment. In natural environment, uneven illumination, weather change, car body movement, rivets, occlusion, license plate stain and low resolution caused by camera shooting angle will lead to character blurring, distortion and even fracture in the extracted license plate character image, which will affect the recognition rate of the whole license plate. Misunderstanding of structurally similar characters. Character extraction with similar structure has high similarity, and it is easy to cause misunderstanding in recognition stage. The characteristic extraction ability of convolutional neural network, such as local receptive field, weight sharing can learn more useful information from the original image. Continuous hierarchical structure can gradually upgrade the bottom features of the original image to more Abstract high-level features, so that the network has good feature learning ability, and can satisfy the image rotation while reducing the computational complexity. Multi-scene requirements such as rotation, translation or tailoring. Therefore, this paper designs and constructs a convolutional neural network model for license plate character recognition in natural scenes.

3.2 Experiment Process.

The license plate character data set in this paper comes from the complete license plate images collected by our laboratory under different natural scenes of a tunnel entrance, totaling 5261. These complete license plate images are segmented into 30,000 character-images. The experimental environment is MATLAB 2015a, Windows 7 and WLAN. Error recognition rate is to describe the probability of all image recognition errors in the whole sample in the process of image recognition. The error rate of the network model algorithm is compared with that of the original algorithm. In the training process of convolution neural network model, the random gradient descent algorithm with 64 batch-size, 0.9 momentum parameter and 0.01 learning rate is used to adjust the model parameters. The regularization constraints and penalties are added to the model training to restrict the change of network weights in training and optimize the whole convolution network; data enhancement technology is introduced into the pretreatment of experimental data. The scaling factor, translation factor, flip factor, scaling factor and clipping factor are 0.1.

3.3 Experiment Result.

In the process of this experiment, the whole experiment is completed under the standard convolution core experimental environment, and the final data is obtained by running convolution neural network. The experiment makes use of the ideal computing environment, and according to the corresponding graphical characteristics to analyze the changes of parameters, to make clear the results of the whole visual data analysis. In this experimental analysis, the parameters of convolution neural network are set according to the conditions. In order to obtain the standard image recognition information results, the collected image data are processed and the key information of the image is counted. The network situation is tested on the test set, and the corresponding accuracy and output value are generated. In the process of data analysis, in order to obtain the most accurate data, after each image information analysis, it is necessary to obtain a snapshot of the current state. All the results were generated after 100 iterations. The accuracy rate of convolutional neural network (%) was 98.2%, while accuracy rate of bp neural network (%) was only 95.1%. As the number of iterations increases in the experimental process, the image recognition results based on convolution neural network will change correspondingly, and the image recognition rate will be further improved. This result suggests that when using convolutional neural network to identify image content, increasing the number of iterations appropriately helps to ensure the results of data analysis. But in fact, the recognition rate oscillation will also affect the image analysis results. Therefore, in the experimental process, we must actively control the image recognition oscillation rate. When the data set is relatively small, we can determine the descent direction by increasing the size to ensure the image data analysis results.

Table 1. Comparison of Accuracy Rates of BP Neural Network and Convolutional Neural Network

Number of iterations	Accuracy Rate of BP Neural Network (%)	Accuracy Rate of Convolutional Neural Network (%)
10	89.3	89.9
20	90.1	91.4
50	93.6	95.7
100	95.1	98.2

4. Conclusion

Convolutional neural network plays an important role in image recognition. Therefore, for the relevant personnel, it is necessary to further clarify the implementation path of convolutional neural network method and master the basic structure of convolutional neural network algorithm. We should extract the feature information of the image, further optimize the strategy of image recognition, and improve the image recognition process from the whole, so as to lay a foundation for further improving the ability of image analysis and processing.

References

[1] Xie Huifang, Liu Yihang, Wang Zi, et al. Research on image recognition based on convolution neural network [J]. Wireless Internet Technology, 2018(14): 23-24.

[2] Zhang Ronglei, Tian Aikui, Tan Hao, et al. Research on image recognition algorithm based on convolution neural networks [J]. Journal of Shandong University of Technology (Natural Science Edition), 2018, 32(1): 48-50.

[3] Yang Wenzhu, Liu Qing, Wang Sile, et al. Down Image Recognition Based on Deep Convolution Neural Networks [J]. Journal of Zhengzhou University (Engineering Science), 2018, 39(2): 11-17.

[4] Cai Chunhua, Wang Feng. Research on Image Recognition Based on Deep Learning [J]. Electronics Quality, 2018(9): 7-9+12.