Facial Recognition Research of Convolutional Neural Network (CNN)

He Yuanzi

Nanchang Institute of Science and Technology, Nanchang, China

Keywords: convolutional neural network; Face recognition; Typical applications

Abstract: in recent years, with the continuous development of deep learning technology, convolutional neural network has been widely used in the field of pattern recognition research with its excellent feature extraction ability. Compared with traditional face recognition, convolutional neural network has many advantages. First of all, traditional face recognition is to extract features by artificial means based on different methods, which is not only slow, but also dependent on human subjectivity. Secondly, convolutional neural network can extract deeper facial features, which are more expressive.

1. Overview of convolutional neural network

Convolutional neural network is a model structure of deep learning. It is a feedforward neural network composed of one or more convolution layers, pooling layers and full connection layers, as well as associated weight matrix. Different from other deep learning methods, convolutional neural network has good effects in image recognition and speech recognition. Compared with other feedforward neural network, the convolutional neural network in large scale image processing and speech recognition, etc, and has the right to share values and local connection features are concerned, it largely reduce the need to learn in the process of model training and calculation parameters, effectively reduced the complexity of the network model, making it very promising a deep learning structure. The application of convolutional neural network in computer vision and speech recognition has become the research focus of scholars at home and abroad. In the experimental simulation and practical application, the convolutional neural network can use pictures as input to extract convolution features automatically, and abstract the learned features layer by layer to extract effective information, so as to effectively solve the complex and time-consuming feature extraction problems in traditional methods. In the case of network model convergence, the semantic features automatically learned by convolutional neural network during training are proved to be highly invariable to the translation, occlusion and perspective change of images.

2. Convolutional neural network characteristics

Traditional convolutional neural network model architecture generally includes convolution layer for filtering operation, pooling layer for reducing dimension and extracting significance features, loss function layer for measuring the quality of model and full connection layer for classifier. Pooling layer - usually appears alternately after the convolution layer to connect with the full connection layer. CNN adopts local connection, weight sharing and pooling methods, so as to obtain good translation, scaling and distortion invariance and make the extracted features more distinguishable

2.1 Local connection

Local connection is an important feature in CNN. Due to the statistical characteristics of the image itself: the statistical characteristics of one region of the image are the same as those of other regions, so it is possible to carry out convolution operation in each region of the same check image. Local connections are helpful to extract local features such as edge information, reduce the number of connections between adjacent layers by 0, and reduce the computational complexity of the model.

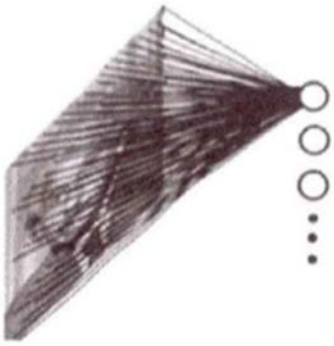

Figure 1 full network connection mode

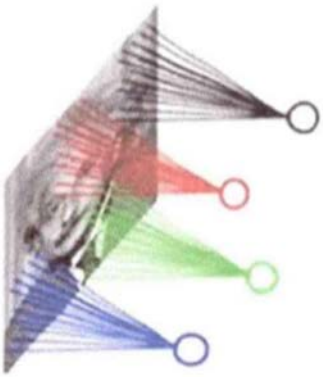

Figure 2 local connection mode

If the image size is 400x400 pixels and there are 100 hidden neurons in the network, then the fully connected network shown in figure 1 requires 400x400X100 = 1.6x107 connections. If the local area used for local connections is 10X10 pixels, then the locally connected network shown in figure 2 only needs 10X10X100 = 104 connections. Therefore, local connections are used to reduce the number of connections in the network structure, which is helpful to improve the computing speed. If the image size is large, the full connection method requires more hardware, while the local connection method can not only effectively reduce the computational complexity, but also reduce the hardware requirements.

2.2 Weight sharing

In order to reduce the number of trainable parameters, the convolutional neural network adopts the mechanism of weight sharing, and the same feature graph USES the same filter to calculate the previous layer. In the design of network architecture, multi-size filters are commonly used to extract multi-level features of the image and enhance the robustness of the model. Similarly, it is assumed that the size of the image is still 400X400 pixels, there are 100 hidden units in the network, and the size of the filter is still 10X10. 400X400X100 = 1.6x107 parameters need to be learned for full connection, while only 100 parameters need to be learned if weight sharing is adopted. Therefore, the weight sharing mechanism greatly reduces the number of trainable parameters in the network structure and reduces the learning difficulty of the network

2.3 Spatial correlation pooling

Principles based on image spatial correlation, convolution neural network joined the pooling mechanism, pooling generally followed after the convolution layer, while the convolution layer reduced neuron structure by means of partial connection number of connections, but due to the increase in figure number, makes the feature dimension is too large, the network has more training, and prone to fitting, by pooling operation such as maximum pool, pool draw can effectively reduce the dimension of the feature and retain image information effectively, to speed up the training of the network, to a certain extent, to avoid the over fitting phenomenon.

2.4 Classification function

The key point of CNN's identification module is classification function, which is connected to the network optimization module successively. As the last link of network structure, it determines the final identification result.

The classification and recognition of face data is a kind of nonlinear mapping. Logistic, Softmax and SVM are common classifiers. From the objective function, the purpose of these three loss functions is to increase the weight of data points that have a great impact on classification, and to reduce the weight of data points that have a small relationship with classification. Among them, SVM is a support vector basis and only considers a few points most relevant to classification to learn classifier. Logistic and Softmax Logistic regression reduce the weight of the points far from the classification plane through nonlinear mapping, so as to improve the weight of the data points

most relevant to the classification. Logistic regression is relatively simple to calculate, effective and convenient to apply, especially in large-scale linear classification. Logistic regression is only suitable for binary classification problems. Softmax regression algorithm is mainly applied to more than two multi-classification problems. Softmax classification function is used to design the classification function layer for face recognition.

CNN's local connection, weights of sharing and pooling operation and strong characteristics of the automatic extraction of features to make it more close to the real world of neural networks, can effectively reduce the complexity of the network structure, and obtain the image scaling, translation, characteristics of distortion invariance has good robustness, and is widely used in target detection, image classification, target tracking, etc. CNN adopts local connection, weight sharing and pooling methods, so as to obtain good translation, scaling and distortion invariance and make the extracted features more distinguishable.

3. Application of convolutional neural network in Le net-5 handwritten numeral recognition

The convolutional neural network has invariance of rotation, translation and scaling of spatial position on the image. Even if there is some change in the shape and posture of the target recognition object, it still does not affect the recognition effect. Layers through the network to identify the image of the topological structure of the data in the abstract, you can get to the image of the implied characteristics of image information, which does not need before the statistical characteristics of the classification of some of the characteristics of image manually extract (manually extract characteristics, often is not always available), the advantage of feature extraction, and because the convolutional neural network makes the network can be directly for training and don't need too much pre-processing operation. The weight distribution in the convolutional neural network is modeled according to the principle of real biological neural network system. The weights of neurons in this network are Shared, which greatly reduces the complexity of the computation of the convolutional neural network. In view of the complex influencing factors in face recognition, these advantages of this network can effectively improve the efficiency of face recognition. In the application of convolutional neural network, it is necessary to set the class label on the input data. The data without class label in common data sets accounts for most of the data, so it is necessary to manually set the class label.

Le net-5 is a network model constructed by Lecun and other researchers, which marks the completion of the prototype of neural network. Its structure is shown in figure 3. Le net-5 is so good at recognizing handwritten Numbers that it has been used to read more than 10 percent of checks in the United States. Le net-5 USES a 32X32 pixel grayscale image as its input data, and the network structure has seven layers. Its first layer is a convolution layer C1, which has 6 Feature maps with convolution kernel of 5X5 size, and each Feature Map has (32-5+1) X (32-5+1), that is, 28X28 neurons, each of which is connected to a region of 5X5 size in the input layer. Therefore, there are a total of (5X5+1) X 6=156 training parameters in C1 layer. The number of connections between the two layers is 156 X (28 X 28) =122304. The second layer is a lower sampling layer S2, which has 6 feature maps of size 14 X 14, and each neuron in each feature map is connected to a region of size 2 X 2 in the corresponding feature map of C1 layer. Each characteristic diagram of S2 has 14 X 14 neurons, 2 X 6=12 parameters, and (4+1) X (14 X 14) X 6=5880 connections. The purpose of the lower sampling layer is to reduce the over-fitting degree of network training parameters and models. The third layer is also a sampling layer C3, and the convolution kernel of 5x5 is used to deal with layer S2. The number of neurons to calculate the characteristic diagram of C3 is (14-5+1) X 14-5+1), that is, 10 X 10. C3 has 16 feature graphs, each of which is derived from different combinations of the previous layer's feature graphs. Figure 4 summarizes the combination of feature graphs of C3 layer. The 0th feature graph is obtained by combining the 0th, 1st and 2nd feature graphs of S2 layer. The 1st feature graph is a combination of the 1st, 2nd and 3rd feature graphs of S2 layer, and so on. Convolution layer C3 does not carry out convolution operation with full connection for sampling layer S2. The advantage of this method is that it can greatly reduce the parameters that need to be learned in the training process and effectively prevent the occurrence of

overfitting. Moreover, this sparse connection mode can further improve the generalization ability of the network, because it produces asymmetric structure in the original network, so the learned features are more essential and valuable. The number of training parameters in C3 layer can be calculated as (5 X 5 X 3+1) X 6+ (5 X 5 X 4+1) X 9+ (5 X 5 X 6+1) X 1=1516. The fourth layer is a lower sampling layer S4, which is composed of 16 feature maps of 5 X 5 size. Each neuron is connected to a region of 2 X 2 size in the corresponding feature map C3. Similarly, 2 X 16=32 parameters and 2000 connections are calculated. The fifth layer is another convolution layer C5, and the convolution kernel of 5 X 5 is also used. Each feature graph has (5-5+1) X (5-5+1), that is, 1 X 1 neuron. Each neuron is fully connected to the 5x5 region of all 16 feature graphs in S4 layer. The sixth layer is the full connection layer F6, with 84 feature diagrams, and each feature diagram has only one neuron fully connected to the C5 layer, so (1 X 1 X 120+1) X 84=10164 parameters and connections. The seventh layer is a full connection layer, which is used to classify the 10 handwritten digits. There are 10 nodes in total, and each node represents an output result. That is, if the value of node I is 0, the result of network recognition is I.

$$y_i = \sum_j (x_j - w_{ij})^2$$

The output value y is determined by the bitmap encoding of I. The closer the bitmap encoding is to I, the closer the value is to 0. This layer has 84X 10=840 set parameters and connections:

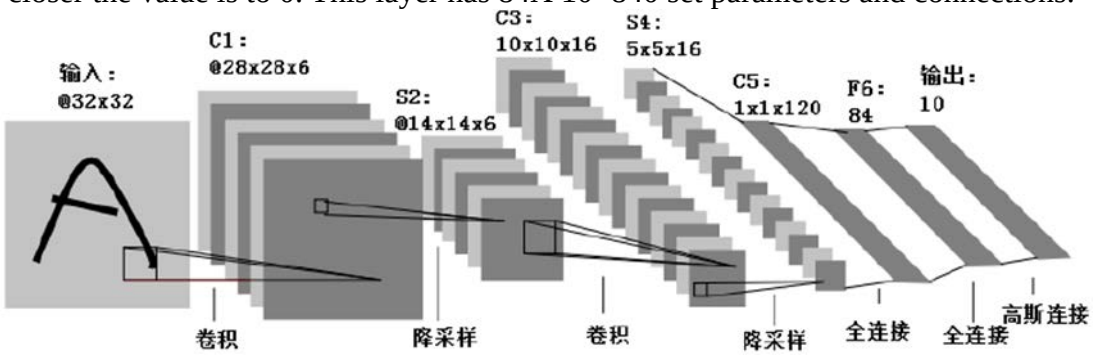

Figure 3 network structure diagram of Le net-5

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	X				X	X	X			X	X	X	X		X	X
1	X	X				X	X	X			X	X	X	X		X
2	X	X	X				X	X	X			X		X	X	X
3		X	X	X			X	X	X	X			X		X	X
4			X	X	X			X	X	X	X		X	X		X
5				X	X	X			X	X	X	X		X	X	X

Figure 4 incomplete connection between C3 layer and S2 layer in Le net-5

4. Summary and prospect

After decades of exploration and development, face recognition technology has become increasingly mature. Convolutional neural network has been successfully applied in many fields, especially in the field of pattern recognition. Some applications developed on the basis of convolutional neural network technology, such as handwriting font recognition, have been successfully applied in handwriting functions such as mobile phones and computers. But there are still not good problem solving, such as how to change in the objective environment of complex situations also can undertake good face recognition, and how to guarantee the accuracy on the basis of improving training speed and recognition rate of face recognition algorithm, these are all we need to further thinking to solve problems, think the convolution neural network can also in-depth

research in the following aspects: (1) in the convolutional neural network into space pyramid pooling method can adapt to any size of the input image, using a greater depth of network try a bigger size of the image face recognition. (2) by training in a certain way (with or without supervision), a better initial value can be provided for the network, so as to improve the overall convergence speed of training using back propagation algorithm.

Acknowledgement

This research project is funded by the science and technology research project of Jiangxi provincial department of education in 2018 (GJJ181055).

References

[1] Yang Ziwen, Facial Recognition Research Based on Deep Convolutional Neural Network [D] Guangxi Normal University, 2017

[2] Hu Chen, Facial Recognition Research Based on Convolutional Neural Network [D], Wuhan University of Technology, 2017

[3] Wang Fei, Facial Recognition Research Based on Convolutional Neural Network [D], Xiangtan University, 2018

[4] Yan Kevin, Facial Recognition Research Based on convolutional neural network [D], Hangzhou Electronics University of Science and Technology, 2018

[5] Wei Pengfei, Research on Face Detection and Recognition Technology Based on Improved CNN [D], Southwest Jiaotong University, 2014

[6] Wu Yao, Qiu Weigen. Face Recognition Based on Improved Deep Convolutional Neural network [J]. Computer Engineering and Design, 2017, 38 (08)