Research on Principle and Application of Convolutional Neural Networks

Hang Zhou¹, Qichen Sun²

¹College of Science, Zhejiang University of Technology, Hangzhou, Zhejiang, 310023
²College of Computer Science & Technology, Zhejiang University of Technology, Hangzhou, Zhejiang, 310023

Abstract. At present, in the field of image recognition, convolutional neural networks have become a research hotspot. Based on the neural cognitive machine, it further reduces the complexity of network models. Convolutional neural network (CNN) has characteristics of simple structure, less training parameters and strong adaptability because it reduces the weight of nodes. In this paper, the network structure, neuron model and training algorithm of convolutional neural network are studied. On the basis of this, the application of CNN in shape recognition is taken as an example, and the related network structure and experimental results are given.

1. Introduction
In 1984, Japanese scholar Fukushima proposed a neurocognitive machine based on the concept of receptive field, which is considered to be the first CNN to be realized, and the first application of the concept of the receiving field in the field of artificial neural networks. A neural cognitive machine based on the concept of receptive field decomposes the visual pattern into multiple sub-patterns and then performs hierarchically-scaled feature plane processing, which attempts to model the visual system so that in the case of object displacement or slight deformation, achieve target recognition. Neurocognitive machines can learn from excitation patterns using constant displacement capabilities and can recognize changes in these patterns. In subsequent applied research, Fukushima used a neurocognitive machine to identify handwritten numbers. Subsequently, researchers at home and abroad proposed various convolutional neural network forms, which have been widely used in postal code recognition and facial recognition.

2. CNN Principle
Deep learning is a special machine learning that uses the nested concept hierarchy to represent and implement great functionality and flexibility. Each concept is defined as associated with a simple concept, and a more abstract representation. It is calculated in a less abstract way. Convolutional neural network is a feedforward neural network inspired by the biological natural visual cognition mechanism. Now, CNN can be applied to image classification, target recognition, target detection, semantic segmentation, etc. This paper mainly introduces the volume that can be used for image classification. The basic structure of the neural network.
3. CNN Training Process

Convolutional neural network (CNN) is an improvement on traditional neural networks. Feedforward neural network. Its biggest feature is local perception and parameter sharing, that is, shared convolution kernel, which has no pressure on high-dimensional data processing. In the convolutional neural network, the neurons are not fully connected, and the connection weights of some neurons in the same layer are also the same, which greatly reduces the number of parameters and weights, and does not need to manually select features, i.e. feature classification. The effect is better.

A typical CNN generally consists of an input layer, a convolution layer, a pooling layer, a softmax layer, and an output layer. The convolution layer, the pooling layer, and the softmax layer can be configured in multiple in a CNN structure. Except that the pooling layer does not have a weight matrix, the other layers have their own weight matrix.
As shown in the figure, the general input data is an image, and the input image is convolved by three trainable filters and an addable bias.

(1) Input layer
The processing to be done by this layer is mainly to preprocess the original data, including:
• De-average: centralizes each dimension of the input data to 0, the purpose of which is to pull the center of the sample back to the origin of the coordinate system;
• Normalization: The amplitude is normalized to the same range, i.e., the interference caused by the difference in the range of values of the data in each dimension;
• PCA/whitening: Dimensionality with PCA, which is normalized to the amplitude of each characteristic axis of the data.

(2) Convolutional layer
There are two key operations in this layer: first, local association, that is, each neuron is treated as a filter, followed by a receptive field sliding, that is, calculation of local data. Suppose there are matrices \( A_{3\times3} \) and \( B_{2\times2} \):

\[
A = \begin{bmatrix}
1 & 2 & 3 \\
4 & 5 & 6 \\
7 & 8 & 9 
\end{bmatrix} \quad B = \begin{bmatrix}
1 & 2 \\
3 & 4 
\end{bmatrix}
\]

Then the result of \( b \) convolution \( a \) is to let \( b \) slide on matrix \( a \). In other words, all \( 2 \times 2 \) consecutive sub-matrices of \( b \) and \( a \) do the "sum of corresponding element product" operation, so the result \( c \) at this time should be:

\[
C = \begin{bmatrix}
37 & 47 \\
67 & 77 
\end{bmatrix}
\]

Therefore, assuming that the size of \( A \) is \( h_a \times w_a \) and the size of \( B \) is \( h_b \times w_b \), the size of \( C \) is \( (h_a-h_b+1) \times (w_a-w_b+1) \), matrix \( B \) is called a convolution kernel or filter, and matrix \( C \) is called a feature map. The above operation is called narrow convolution. If matrix \( A \) is added to \( h_b-1 \) row zero vector in advance, \( w_b-1 \) column zero vector is added to the left and right, and then convolved with \( B \), it is called wide convolution. Narrow convolutions and wide convolutions are represented by \( \text{conv2} \) (\( A, B, \text{'valid'} \)) and \( \text{conv2} \) (\( A, B, \text{'full'} \)) respectively.

(3) Pooling layer
The pooled layer main functions are as follows:
• Feature invariance: that is, the scale invariance in image processing. The pooling operation is image compression, and the information left is the feature with scale invariance, which is the feature that can best express the image.
• Feature Dimension Reduction: Extracting the most important features is also the primary role of pooling operations.
• Prevent over-fitting to a certain extent and make it easier to optimize.

(4) Softmax layer
The role of this layer is mainly to make the non-linear mapping of the convolutional layer output. The excitation function used by CNN is generally ReLU, and the image is as follows.
Figure. 3 Excitation function image

In summary, the convolutional network acts as a multi-layer perceptron whose network structure is highly invariant to translation, scaling, tilting or common forms of deformation. Convolution of each neuron affects only a subset of neurons in the adjacent layer, with local receptive fields. Therefore, the network has a strong ability to capture local features; on the other hand, it is significantly reduced by weight sharing and pooling. The computational complexity of the network makes CNN widely used.

4. Shape Recognition Based on CNN

CNN is a target recognition based on the shape of the object boundary, which can recognize a closed shape and has a high recognition rate for an unclosed shape. The closing value of the S-ary in the network is a preset value, and the update weight during training is based on the enhanced learning rules proposed by Fukushima. The network training method uses an unsupervised learning method. Figure 1 below is a partial experimental sample diagram.

Figure. 4 Part of the training sample plan

The experimental samples were divided into four types: triangular, quadrilateral, octagonal and circular. Each sample had 10 samples and 80 samples. The experimental training used 40 samples and the remaining 40 samples were used for testing. The final recognition results are shown in Table 1.

Table. 1 Identification results

| The shape to be recognized | Recognition result |
|---------------------------|--------------------|
| Triangle                  | 10 pairs           |
| Quadrilateral             | 10 pairs           |
| Octagon                   | 10 pairs           |
| Round                     | 9 pairs, the other is misunderstood as a quadrilateral |

An example of identifying errors is shown in Figure 2:
The reason for the recognition error is that the training uses fewer sampling modes and does not cover all loop modes, so when the test mode is input to the network, a similar quad mode wins and eventually gets the wrong output. Shape recognition using convolutional neural networks is primarily to verify the pattern recognition capabilities of convolutional neural networks. Therefore, although the sample image used is small, it has been explained that the convolutional network has a higher recognition rate in shape recognition. Due to the limitations of preprocessing (positioning, segmentation) capabilities, the distortion-resistant and pre-identification license plate characters have some noise and distortion, so the convolutional neural network can be applied to the license plate recognition system.

5. Conclusion

Based on the neural cognitive machine, this paper proposes a convolutional neural network to deal with the problem of pattern recognition. The network is a layered neural network with neurons of the same type in each layer. The basic convolutional neural network structure and its neuron model are also introduced. Then the training process of convolutional neural networks is discussed. When the desired features are predetermined, a supervised algorithm is used, and the network learns layer by layer, and vice versa.

References

[1] Gu Jialing, Peng Hongjing. Growing convolutional neural network and its application in face detection [J]. Journal of System Simulation, 2009, 21(8): 2441-2445.

[2] Zhao Zhihong, Yang Shaopu, Ma Jianqiang. Research on license plate character recognition based on convolutional neural network LeNet-5 [J]. Journal of System Simulation, 2010, 22(3): 638-641.

[3] T. H. Hildebrandt, Optimal Training of Thresholded Linear Correlation Classifiers, IEEE Transactions on Neural Networks Z(6), PP.577-588, Nov. (1991).

[4] K. Fukushima, “Neocognitron: A self-organizing neural-network model for a Mechanism of Pattern recognition unaffected by shift in position, Biol. Cybern., vol.36, pp.193-202, 1980.

[5] C. Neubauer, Shape, position and size invariant visual pattern recognition based on principles of neocognitron and perception in Artificial Neural Networks, 1. Alexander and J. Taylor, Eds. Amsterdam the Netherlands: North-Holland, vol.2, 1992, PP.833-837.