ISSN: 2664-9640

Summary of Convolutional Neural Network Research

Qi Tang, Zhihe Li, Hang Yan

Chengdu University of Technology, Chengdu, China

Abstract

After several years of development, deep learning has attracted the attention of researchers. As an important structure of deep learning model, convolutional neural network is widely used in various fields. This article mainly introduces the structure model of the convolutional neural network and the technical point of feature extraction through operations such as convolution pooling. And summarized some of the existing deficiencies of CNN.

Keywords

Convolutional Neural Network; Deep Learning; Deep Learning Network.

1. Introduction

Convolutional neural network is an important branch of deep learning model structure. In 2006, Geoffrey Hinton and his student Ruslan Salakhutdinov published an article in the international journal "Science" [1], which first proposed the idea of deep learning. Krizhevsky et al. [2] used CNN in the LVSRC-12 competition for the first time. By deepening the depth of the CNN model and using ReLU+dropout technology, they achieved the best classification results at the time. CNN improves some of the shortcomings of ANN through several important ideas such as sparse connection, parameter sharing, and equivariant representation. The space complexity of the network structure is reduced, and the generalization ability of the network is improved. With the development of model of deep learning, convolutional neural networks have been widely used in many fields.

2. Theory

2.1. Artificial Neural Network

Artificial neural network is a simulated biological process based on modern neurobiology research. It reflects the computational structure of certain characteristics of the human brain. In artificial neural networks, artificial neurons are called "processing units", also called "nodes." An Artificial neuron is a formal description of biological neurons. It uses mathematical language to abstract and describe the process of information processing of biological neurons. And use mimetic diagrams to simulate and express the structure and function of biological neurons.

2.1.1. Neuron modeling

The neuron model is the most basic part of the neural network. The neuron receives input signals from n other neurons. These input signals are passed through weight to connect.

The summary of input value received by the neuron will be compared with the threshold of the neuron, and then "the activation function "To be processed to produce output.

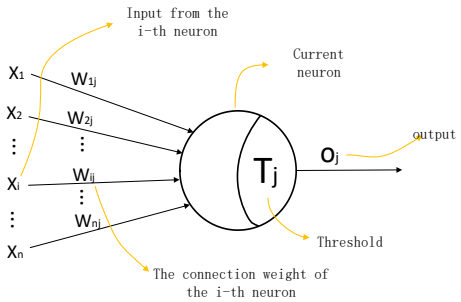

Figure 1-1 Schematic diagram of artificial neuron input-output model

2.2. Convolutional Neural Network

The biological basis of convolutional neural networks is Receptive Field, which was discovered by Hubel and Wiesel in 1962 through the study of cat visual cortex cells. The stimulus area that a neuron responds to or dominates is the receptive field of the neuron, and different neurons have their own different receptive fields.

On this basis, Fukushima proposed a neurocognitive machine model in 1984. The neuro cognitive machine model is the first realization of the convolutional neural network. It is also the first application of the concept of the receptive field in the field of artificial nerves.

2.2.1. Concepts and principles of convolutional neural networks

Convolution operation helps to improve the machine learning system through three important ideas: sparse interaction, parameter sharing, and equivariant representations. In addition, convolution provides a way to handle variable-size input.

2.2.1.1 sparse connectivity

Traditional neural networks use matrix multiplication to establish the connection between input and output. Among them, each independent parameter of the parameter matrix describes the interaction between each input unit and each output unit. This means that every output unit interacts with every input unit. However, convolutional neural networks have the characteristics of sparse interactions (also called sparse connectivity or sparse weights). This is achieved by using cores that are much smaller than the input size. Storing fewer parameters not only reduces the storage requirements of the model, but also improves the statistical efficiency of the model. At the same time, it also requires less calculation to obtain the output, which significantly improves the efficiency.

2.2.1.2 Parameter sharing

Parameter sharing refers to using the same parameters in multiple functions of a model. In a traditional neural network, when calculating the output of a certain layer, each element of the weight matrix is used only once. When it is multiplied by an element of the input, it will not be used again. Parameter sharing can also be said that a network contains bound weights, because the weights used for one input will also be bound to other weights. In a convolutional neural network, every element of the kernel acts on every position of the input (except for some boundary pixels that depend on the decision design). Unlike each position that needs to learn a separate parameter set, the parameter sharing in the convolution operation ensures that the network only needs to learn one parameter set. Although this will not change the forward propagation time (still $O(k \times n)$), it significantly reduces the storage requirements of the model, reducing the number of parameters m to k, and k is usually much smaller than m Magnitude. Because the scales of m and n are usually very close, in practice, the scale k is small relative to m×n. Therefore, convolution is greatly superior to dense matrix multiplication in terms of storage requirements and statistical efficiency.

ISSN: 2664-9640

2.2.2. Convolutional Neural Network Model

Convolutional neural networks mainly include four layers: the input layer, the convolution layer, the pooling layers, and the output layer. The input layer has only one layer, which directly accepts two-dimensional vision; the convolutional layers and the pooling layers are generally set with multiple layers; the output layer is generally a one-dimensional linear array for classification. The convolutional neural network model is shown in Figure 1-6.

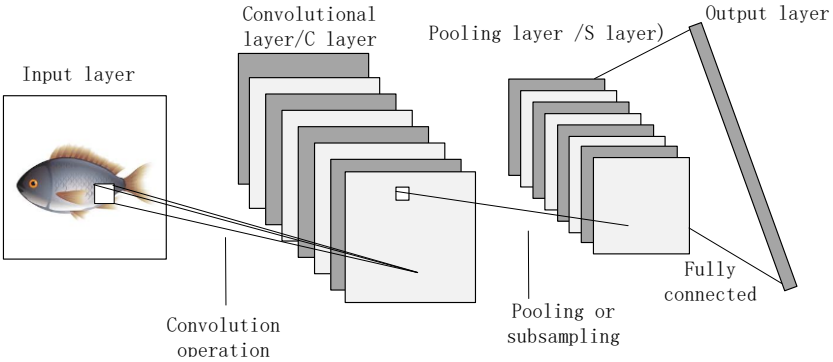

Figure 1-6 Schematic diagram of convolutional neural network (CNN) model

2.2.2.1 Convolution operation

Convolution operation is calculation method in analysis of mathematics. Assuming that the input data is the 5×5 matrix in Fig. 1-7a), the corresponding convolution kernel (ie, the convolution parameter) is a 3×3 matrix. In addition, when performing a convolution operation, each time a convolution operation is performed, the convolution kernel is set to move one pixel position, that is, the convolution step length is 1.

1	0	1
0	1	0
1	0	1

Convolution kernel

1	2	3	4	5
6	7	8	1	0
9	8	7	6	5
4	3	2	1	0
1	2	3	4	5

Input data

a) b)

Figure 1-7 Two-dimensional convolution kernel and input data

The first convolution operation starts from (0, 0) pixels, and the parameters in the convolution kernel are multiplied bit by bit with the image pixels at their corresponding positions, and then accumulated as the result of a convolution operation, which is 27, as shown in Figure 1-8. Show. The convolution kernel performs convolution operations on the input image in the order of top-down and left-to-right first according to the step size, and the final output is a 3×3 size convolution feature, and this result is used as the next Input for layer operation.

2.2.2.2 Pooling

The convolutional layer of a convolutional neural network usually contains three levels. In the first stage, multiple convolution operations are performed in parallel through the convolution layer to generate a set of linear activation functions. In the second stage, a non-linear activation function is applied to each linear output in the first stage. This level is sometimes called the detector stage. In the third stage, the pooling function is used to further adjust the output of the convolutional layer.

The pooling function takes the place of the overall statistical characteristics of the neighboring output of a certain location with the output of the network at that location, and integrates all

the feedback of the surrounding neighborhood. By integrating the statistical characteristics of k pixels in the pooling area, it is possible to make the pooling unit less than the detection unit.

2.2.2.3 Activation function

"Activation function", also known as "non-linear mapping function", is an indispensable key module in deep convolutional neural networks. In other words, the nonlinearity of the activation function brings most of the powerful representation capabilities of the deep network model. The activation function simulates the characteristics of biological neurons, generates output after receiving a set of signals, and simulates the activation or excitement of neurons through a threshold. Currently commonly used activation functions are: Sigmoid type function, tanh(x) type function, modified linear unit (ReLU) and its derived activation functions (Leaky ReLU, parameterized ReLU, randomized ReLU, exponential linear unit (ELU)) Wait.

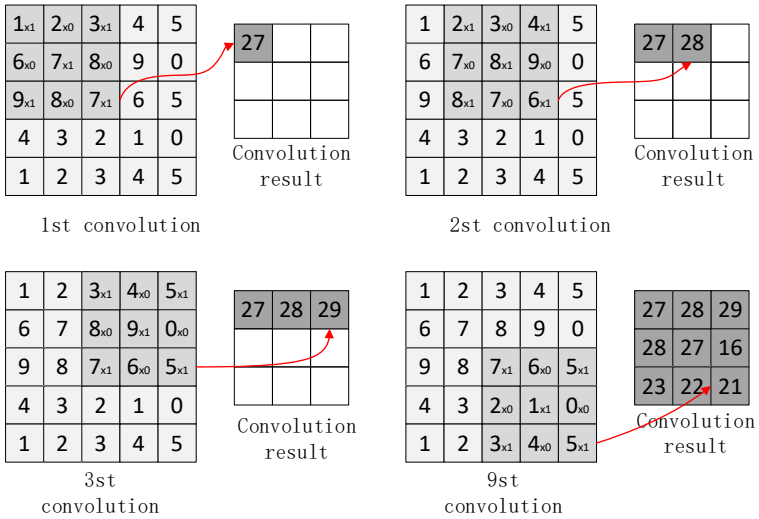

Figure 1-8 Convolution operation

2.2.3. Training method

As for the model of deep learning, the input data use the origin sample form which has not be processing. Then perform many operations on the data stacked on the input layer. These operations can be regarded as a complex function f(x) as a whole, and the data loss and the regularization loss of the model parameters together forms the final loss function. Driven by the final loss, the depth model updates the parameters of the model during training and back-propagates the error to all layers of the network. The training of the convolutional neural network model can be simply abstracted as the direct "fitting" from the original data to the final target. These components in the middle are playing the role of mapping the original data into features (ie feature learning) and then mapping them into sample labels (That is, the role of the target task.

3. Conclusion

CNN reduces the spatial complexity of the neural network structure through weight sharing and sparse connections, and improves the generalization performance of the network. However, although convolutional neural networks have achieved good results in some aspects, there are also deficiencies. For example, the determination of the activation function requires multiple training for comparison, which makes the operation process not easy. The number of layers of the convolutional neural network and the number of neurons in each layer are difficult to determine, and the error is large for test data that differs greatly from the training set.

ISSN: 2664-9640

References

[1]LoweDG.DistinctivemagagefeaturesforminvariantkeypointsJ].InternationalJournalofComputerVision, 2004, 60(2): 91-1110.

[2] KrizhevskyA,Sutskever Ⅱ ,Hinton G.Imagenet classification with deep convolutional neural networks// Proceedingsofthe Advancesin Neural Information Processing Systems.Lake Tahoe, USA, 2012: 1097-1105.