Advances in Intelligent Systems Research, volume 147
International Conference on Network, Communication, Computer Engineering (NCCE 2018)

Survey of Convolutional Neural Network

Xv Zhang [1, a)], Chenxi Xv [1], Ming Shen [1], Xin He [1], Wei Du [2, b)]

[1] *School of Software Engineering, JiLin University, Changchun JiLin 130000, China*
[2] *School of Computer Science and Technology, JiLin University, Changchun JiLin 130000, China*

[a)] The first author: 1075910752@qq.com
[b)] Corresponding author: weidu@jlu.edu.cn

Abstract. In recent years, the breakthrough of deep learning in the field of artificial intelligence algorithms has triggered an academic upsurge which attracted more and more researchers. As a multi-layer perceptron, the key to its success lies in the local link and weight-sharing method. On the one hand, it reduces the quantity of weights and makes the network easier to optimize. On the other hand, it reduces the risk of over-fitting. A weight-sharing network's structure of the convolutional neural network makes it more similar to a biological neural network, which reduces the complexity of the network model and quantity of weights. In the processing of image problems, especially recognizing displacement, scaling, and other forms of distortion invariant applications, it has better robustness and operation efficiency. First of all, this paper reviews the development history of convolutional neural network. Secondly, it introduces the basic structure of convolutional neural network, and elaborates its differences from ordinary artificial neural networks in terms of operating principles. It also analyzes the details of convolutional neural network's structural framework which includes convolutional layers, subsampling layers, and fully connected layers. Finally, the advantages of convolutional neural network in image processing, speech analysis, and other fields are given at last.

Key words: Convolutional Neural Network; deep learning; image processing; computer vision.

INTRODUCTION

Convolutional Neural Network Overview

The convolutional neural network, also known as CNN, is one of the artificial neural networks. It is a special way of image recognition, and a very effective network with forward feedback [1]. The main goal of the CNN is to identify two-dimensional graphics. Its network structure is highly invariant to translation, scaling, slanting or other forms of deformation. The reason why CNN has these characteristics is that CNN focuses on different kinds of features at each level [2]. At first layer which is close to the original image, focused is on the pixel level. And after multiple feature extractions, Features such as relational, sequential, or structured types (which is used to be called topology) are extracted, and the consistency is close to the object itself.

The application range of convolutional neural networks is not only limited to the field of image recognition but can also be applied in the face recognition [3], text recognition [4] and other directions.

The Development of Neural Networks

In 1958, David Hubel [5] and Torsten Weiesel studied the corresponding relationship between the pupil area and the cerebral cortex in Johns Hopkins University. They opened a 3 mm hole on the cat's hindbrain bone and inserted some electrodes in it to test the activity of the neurons. They showed various shapes and brightness objects in front of the eyes of cats at the same time. When showing each object, they also changed the position and angle of it.Six days later, the researchers discovered a unique type of neuron and name direction-selective cell. When the pupil

captures the edge of the object in the eyes and this edge points in a certain direction, the neuron cells will be active. Then the concept of receptive field [6] is put forward.

In 1984, Japanese scholar Fukushima [7] proposed a neurocognitive machine based on the concept of receptive field. The neurocognitive machine can be regarded as the first realization network of the convolutional neural network, and it is also the application of the concept of receptive field in the artificial neural network. The neurocognitive machine decomposes a visual mode into a few sub features, and then processes it into the hierarchical plane of the hierarchical necklace. It tries to model the visual system, enabling it to recognize even when the object is moving or slightly deforming. Usually, the neurocognitive machine contains two neurons, that is, the S-element that bear the feature extraction and the anti-deformation C-element. There are two important parameters in the S-element, namely receptive field and threshold parameter, the receptive field determines the number of input connections, and the threshold controls the reaction degree of the sub features. Each visual fuzzy quantity brought by C-element in the receptive field of S-element obviousness normal distribution. Corresponding to the actual physiological phenomenon, if the eyes feel the object is moving, that is already feeling single blur and ghost image, S-element feelings will adjust the recognition mode, it will not extract all the features to the brain completely and will only get some important features to the brain and shield other visual interference. The final conclusion is that when the eye sees the moving object, the C-element determines the overall characteristic sensory control first, and then extracts the corresponding features from the S-element sensory region. And the discovery of neurocognitive science is a reference to the mathematical simulation of the convolutional neural network.

In recent years, the research of convolutional neural network has made great progress in the field of real time application.Gishick et al[8,10]. And Ren et al [9]. have made deep research in the area of object detection based on convolutional neural network. R-CNN [8], Fast R-CNN [10] and R-CNN models [9] have been put forward, breaking through the bottleneck of real-time application of convolutional neural network.

THE STRUCTURE OF THE CONVOLUTIONAL NEURAL NETWORK

The Basic Structure of Convolutional Neural Network

The early structure of convolutional neural network is relatively simple, such as the classical LeNet-5 model [11], which is mainly applied in handwritten character recognition, image classification and other relatively simple computer vision applications.

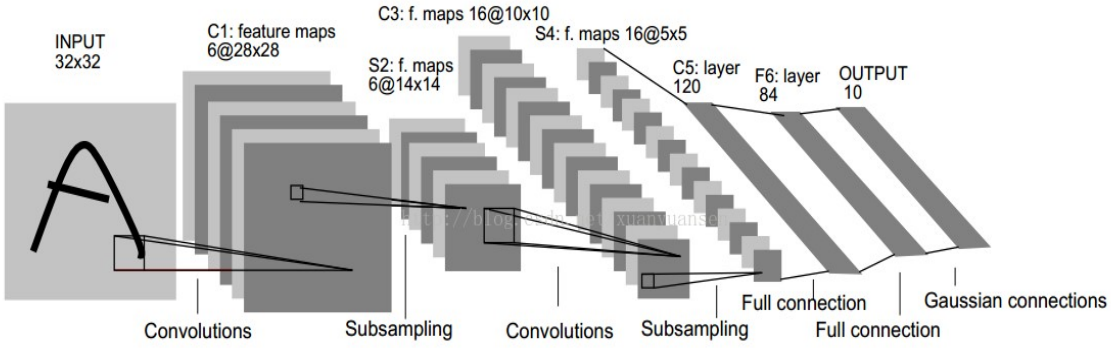

FIGURE 1. Basic structure of LeNet-5

The important components of the convolutional neural network include:

Convolutional Layer: The most important part of the structure of convolutional layer neural network is called the filter or the kernel. Which can convert a sub node matrix on the current layer of neural networks into a unit node matrix on the next layer of neural networks. The unit node matrix refers to a node matrix with a length and a width of 1, but the depth is not limited.

$$Depth_{\text{filter}} = \text{D}epth_{node} \tag{1}$$

The forward propagation process of the filter is the process of calculating the nodes in the right unit matrix by the nodes in the left small matrix. Suppose that the $w^i_{x,y,z}$ is used to represent the node i in the output unit node matrix. The filter input the weight of the node (x, y, z) ,and using b^i to represent the offset term parameters for the output node i .Then the value of the node i in the unit matrix is:

$$g(i) = f(\sum_{x=1}^{2}\sum_{y=1}^{2}\sum_{z=1}^{2} a_{x,y,z} * w^i_{x,y,z} + b^i) \tag{2}$$

Where $a_{x,y,z}$ is the value of node (x, y, z) in the filter, and f is the activation function [17].

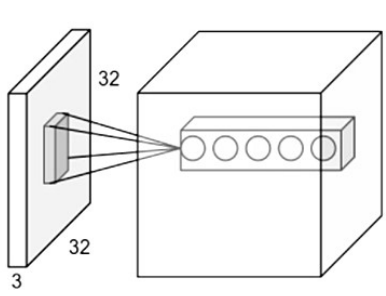

FIGURE 2. Structure diagram of the convolutional layer's filter

Several filters are applied to the input data, one input parameter is used to do many kinds of feature extraction, and the result of applying a filter to the image is called Feature Map [12] whose quantity is equal to the number of filters. If the previous input layers are convolutional layers, FM applies the filter and the filter outputs another FM, which means when distributing the eigenvalue of filters to the whole image, the feature will be location-independent. Meanwhile, multiple filters could detect different features at the same time.

Subsample layer, which is also called pooling layer [13], is mainly used to reduce the size of the input data. There are many methods to realize sub-sampling which can reduce effectively the computational complexity of the computer. Moreover, the most common methods are maximum merging, average merging and random merging. Pooling layer, which often appears behind the convolutional layer, can reduce the size of the matrix very effectively. It can not only accelerate the calculation speed, but also prevent the over-fitting problem. The process propagating to before pooling is also accomplished by moving a structure similar to the filter. However, the calculation in pooling layer filter is not a weighted sum of nodes, but a simpler maximum or average operation. In addition to moving in the length and width dimensions, the filter for pooling layer also needs to move in depth dimension. The final sub-sampling layer is usually connected to one or more fully connection layers, whose output is the final output. In addition, each neuron in the full connection layer will be connected to each neuron in the previous layer.

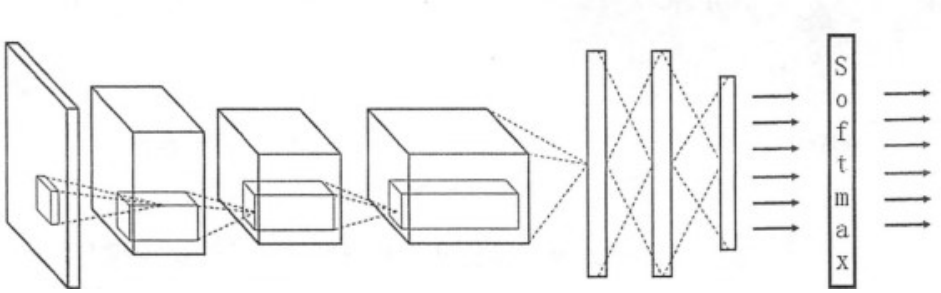

FIGURE 3. The convolutional neural network architecture diagram of classification problem

The convolutional neural network is a multi-layer neural network, each layer is composed of multiple two-dimensional planes, and each plane is composed of multiple independent neurons. Figure 4 is a conceptual demonstration of the convolutional neural network. The input image is convoluted with three tradable filters and addressable biases, and three feature maps are generated in the C1 layer after convolution. After the four pixels of each group of the feature map are summed, weighted and biased, we can get three feature maps of S2 layer by a sigmoid function [14]. These maps are then filtered to get the C3 layer. This hierarchy then produces S4 in the same way of S2. Finally, these pixel values are processed regularly. Then connect them into a vector and input the vector to the traditional neural network to get the output. That is to say, the complex data is removed and the simple data that can be extracted is left behind. The middle part of the convolutional neural network is the part that is really completing the convolution work. This part is composed of two parts, one is the feature-extracting [15] layer, and the other is the feature-mapping [16] layer. Layer C is the feature-extracting layer, the input of each neuron is connected with the local sensing region of the previous layer, then extract the local characteristics. The layer S is the feature-mapping layer. Each computing layer of the network is composed of multiple feature maps. Each feature is mapped into a plane on which all neurons have equal weights. The feature mapping structure uses the influence function and the small sigmoid function as the activation function of the convolutional neural network, which makes the feature mapping have unique invariance. The number of network free parameters is reduced and the complexity of network parameter selection is reduced because of the neuron sharing weights on the same surface. A computational layer called S-layer which is used to calculate local average and secondary extraction, following each feature extraction layer in the convolutional neural network. This unique two-step featured extraction structured to enable the network to identify the input samples with higher tolerance to distortion. This ability is one of the reasons why multi-layer convolutional neural network is easy to use.

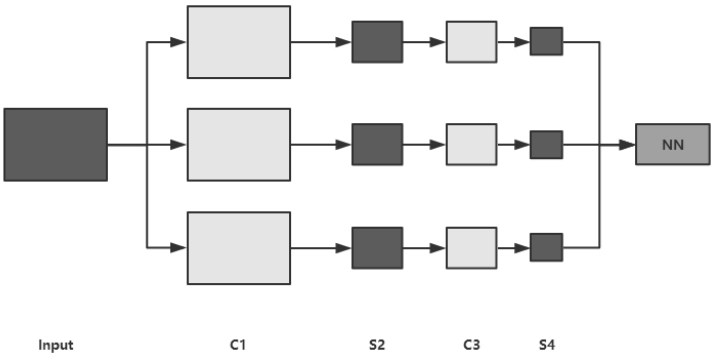

FIGURE 4. A simplified model of the convolutional neural network

The Advantages of Convolutional Neural Network

Conventional neural networks do not adapt well to all images.

The three-dimensional structure of a network capacity: Convolutional neural networks have great advantages. For inputs which contain a large number of pictures, it limits the structure in a more rational way. Each layer of the convolutional neural network converts the three-dimensional input into a three-dimensional output value. When the image is recognized, the input layer retains all the details of the image, two dimensions of the three-dimensional vector diagram represent the original width and height of the image, and depth represents the color of image.

CNN is mainly used to identify displacement, scaling and other forms of distortion invariant [18] two-dimensional graphics. Since the feature detection layer of CNN learns through training data, explicit feature extraction is avoided when using it, and learning is implicitly performed from the training data.Secondly,on account of the same neuron weights on the same feature map, the network can learn in parallel. Because of the special structure of local weight sharing, convolutional neural network has unique advantages in speech recognition and image processing. Its layout is closer to the actual biological neural network [19]. Weight sharing reduces the complexity of network and the complexity of data reconstruction in feature extraction and classification is avoided.

Convolutional neural networks have the following advantages over general neural networks in image processing:

1)The input image and topology of the network can be better matched.

2)Feature extraction and pattern classification are performed at the same time and simultaneously in training.

3)Weight sharing can reduce the training parameters of the network, making the neural network structure simpler and more adaptable.

SUMMARY

This article briefly introduces the development history of convolutional neural networks and the advantages of conventional neural networks. The structure of the convolutional neural network is highlighted. Convolutional neural network has become a research hotspot in the field of speech analysis and image recognition. Its weight-sharing network structure makes itself more similar to biological neural networks, which reduces the complexity of the network model and reducing the number of weights. This advantage of the convolutional neural network is more pronounced when the input of the network is a multi-dimensional image, so that the image can be directly used as the input of the network, avoiding the complicated feature extraction and data reconstruction process in the traditional recognition algorithm.

REFERENCES

1. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(4):640-651.
2. Uijlings J R R, Sande K E A V D, Gevers T, et al. Selective Search for Object Recognition[J]. International Journal of Computer Vision, 2013, 104(2):154-171.
3. Lawrence S, Giles C L, Tsoi A C, et al. Face recognition: a convolutional neural-network approach[J]. IEEE Transactions on Neural Networks, 1997, 8(1):98-113.
4. Vaillant R, Monrocq C, Cun Y L. An original approach for the localization of objects in images[C]// International Conference on Artificial Neural Networks. IET, 1993:26-30.
5. Hubel D H, Wiesel T N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. [J]. J Physiol, 1962, 160(1):106-154.
6. bz(kl). A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position[J].
7. Fukushima K. Noncognition: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position[J]. Biological Cybernetics, 1980, 36(4):193-202.
8. GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C] // Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2014:580-587.
9. Ren S, He K, Girshick R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6):1137-1149.
10. Girshick R. Fast R-CNN[J]. Computer Science, 2015.
11. Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[C]// International Conference on Neural Information Processing Systems. Curran Associates Inc. 2012:1097-1105.
12. Koh J, Suk M, Bhandarkar S M. A multilayer self-organizing feature map for range image segmentation[J]. Neural Networks, 1995, 8(1):67-86.
13. Liu L, Shen C, Hengel A V D. Cross-Convolutional-Layer Pooling for Image Recognition[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015, 39(11):2305.
14. Yonaba H, Anctil F, Fortin V. Comparing Sigmoid Transfer Functions for Neural Network Multistep Ahead Streamflow Forecasting[J]. Journal of Hydrologic Engineering, 2010, 15(4):275-283.
15. Cayiroglu I. A new method for machining feature extracting of objects using 2D technical drawings[J]. Computer-Aided Design, 2009, 41(12):1008-1019.
16. Pajares G, Cruz J M, Aranda J. Stereo matching based on the self-organizing feature-mapping algorithm[J]. Pattern Recognition Letters, 1998, 19(3–4):319-330.
17. Solazzi M, Uncini A. Regularising neural networks using flexible multivariate activation function[J]. Neural Networks the Official Journal of the International Neural Network Society, 2004, 17(2):247.
18. Kanaoka T, Chellappa R, Yoshitaka M, et al. A higher-order neural network for distortion invariant pattern recognition[J]. Pattern Recognition Letters, 1992, 23(8):977-984.
19. Vaidyanathan S. 3-Cells Cellular Neural Network (CNN) Attractor and its Adaptive Biological Control[J]. International Journal of Pharmtech Research, 2015, 8(4):632-640.