Convolutional Neural Networks

Quan Zhang
School Of Electronic & Information Engineering, Suzhou University of Science and Technology, Suzhou 215009, China
woquan0412@163.com

Keywords: Convolutional Neural Network, feature extraction, deep learning.

Abstract: The earliest Convolution Neural Network (CNN) model is leNet-5 model proposed by LeCun in 1998. However, in the next few years, the development of CNN had been almost stopped until the article 'Reducing the dimensionality of data with neural networks' presented by Hinton in 2006. CNN started entering a period of rapid development. AlexNet won the championship in the image classification contest of ImageNet with the huge superiority of 11% beyond the second place in 2012, and the proposal of DeepFace and DeepID, as two relatively successful models for high-performance face recognition and authentication in 2014, marking the important position of CNN. Convolution Neural Network (CNN) is an efficient recognition algorithm widely used in image recognition and other fields in recent years. That the core features of CNN include local field, shared weights and pooling greatly reducing the parameters, as well as simple structure, make CNN become an academic focus. In this paper, the Convolution Neural Network's history and structure are summarized. And then several areas of Convolutional Neural Network applications are enumerated. At last, some new insights for the future research of CNN are presented.

1 INTRODUCTION

Convolutional Neural Network is a category of neural networks that have proved very effective in areas such as image recognition and classification (Krizhevsky, 2012). Convolutional Neural Network has made a series of breakthrough research results in the fields of face recognition (Lawrence, 1997), sentence classification (Kim, 2014), semantic segmentation and so on. So, it is of great value to review the works in this research field.

Convolutional Neural Network is inspired by the biology of the visual cortex. A small percentage of cells in the visual cortex are sensitive to the visual area of a particular part. An interesting experiment by Hubel and Wiesel (Kandel, 2009) in 1962 illustrates this concept in detail. They demonstrated that some individual neurons in the brain respond only when they are in the edges of a particular direction. For example, some neurons are only excited about vertical edges, while others are excited about horizontal or diagonal edges. Hubel and Wiesel found that all of these neurons are arranged in the form of columnar structures and work together to produce visual perception. The idea that a particular component of such a system has a particular task (neuronal cells in the visual cortex looking for specific features) is equally applicable in machines, which is the basis of the CNN.

Convolution has important points that local connection and weight sharing. Local connection and weight sharing reduce the amount of parameters, greatly reducing training complexity and alleviating over-fitting. At the same time, weight sharing also gives the convolutional network tolerance for translation.

After years of development, the Convolution Neural Network has been gradually extended to more complicated fields such as pedestrian detection, behavior recognition and posture recognition from the initial simpler handwritten character recognition, etc. Recently, the application of Convolutional Neural Networks further develops to a deeper level of artificial intelligence, such as natural language processing, speech recognition, and others. Recently, Alpha go, an artificial intelligence go program developed by Google, has successfully used convolutional neural network to analyze the information of Go board and win the European Champion and Go champion successively in the Challenge, attracting wide attention. From the current trend of research, since the application prospect of convolutional neural network is full of possibilities, it also faces some research problems,

Zhang, Q.
Convolutional Neural Networks.
In *3rd International Conference on Electromechanical Control Technology and Transportation (ICECTT 2018)*, pages 434-439
ISBN: 978-989-758-312-4

such as how to improve the structure of Convolution Neural Network to increase the learning ability, and how to convert Convolution Neural Network from a reasonable form into the new application model.

The outline of the paper is as follows. Section two describes the background to the CNN including the theoretical proposition stage, the model realization stage, and the extensive research stage. Section three defines the structure of CNN that mainly composed of input layer, convolution layer, subsampling layer (pooling layer), fully connected layer and output layer. Section four gives two examples of Convolution Neural Network applications and discusses the future of other applications.

2 BACKGROUND

The history of Convolutional Neural Networks can be roughly divided into three stages: the theoretical proposition stage (Yandong, 2016), the model realization stage (Fukushima, 1988), and the extensive research stage.

2.1 The theoretical proposition stage

In the 1960s, Hubel showed that the biological information from the retina to the brain is stimulated by multiple levels of receptive field. In 1980, for the first time, Fukushima proposed Neocognitron (Ballester, 2016) based on the theory of receptive fields. Neocognitron is a self-organizing multi-layer neural network model. The response of each layer is obtained by the local sensory field of the upper layer. The recognition of the model is not affected by position, small shape changes and the size of the scale. Unsupervised learning by Neocognitron is also the dominant learning method in early studies of convolutional neural networks

2.2 The model realization stage

In 1998, Lecun proposed LeNet-5 to use a gradient-based backpropagation algorithm to supervise the network. The trained network converts the original image into a series of feature maps through the convolutional layer and the down-sampling layer alternately connected. Finally, the feature representation of the images is classified by the fully connected neural network. Convolutional kernel completed the receptive field function, and it can lower the local area information through the convolution kernel excitation to a higher level. The successful application of LeNet-5 in the field of handwritten character recognition has drawn the attention of academia to the convolutional neural network. In the same period, research on Convolutional Neural Networks in speech recognition, object detection, face recognition and so on has been gradually carried out.

2.3 The extensive research stage

In 2012, AlexNet proposed by Krizhevsky won the championship in the image classification contest of ImageNet, a large image database, with the huge superiority of 11% beyond the second place, making the Convolutional Neural Network become an academic focus. After AlexNet, new models of convolutional neural networks have been proposed, such as Visual Geometry Group (VGG), Google's GoogLeNet, Microsoft's ResNet, etc. These networks refresh AlexNet Record created on ImageNet. Furthermore, convolutional neural network is continuously merged with some traditional algorithms, and with the introduction of migration learning method, the application of Convolutional Neural Networks has been rapidly expanded. Some typical applications include: Convolutional neural network combined with Recurrent Neural Network (RNN) for image summarization and image content quiz; Convolution Neural Networks have achieved significant accuracy gains in small sample image recognition databases; and video-oriented behavioral recognition models - 3D Convolutional Neural Networks.

3 STRUCTURE

As shown in the figure 1, the typical Convolutional Neural Network is mainly composed of input layer, convolution layer, subsampling layer (pooling layer), fully connected layer and output layer.

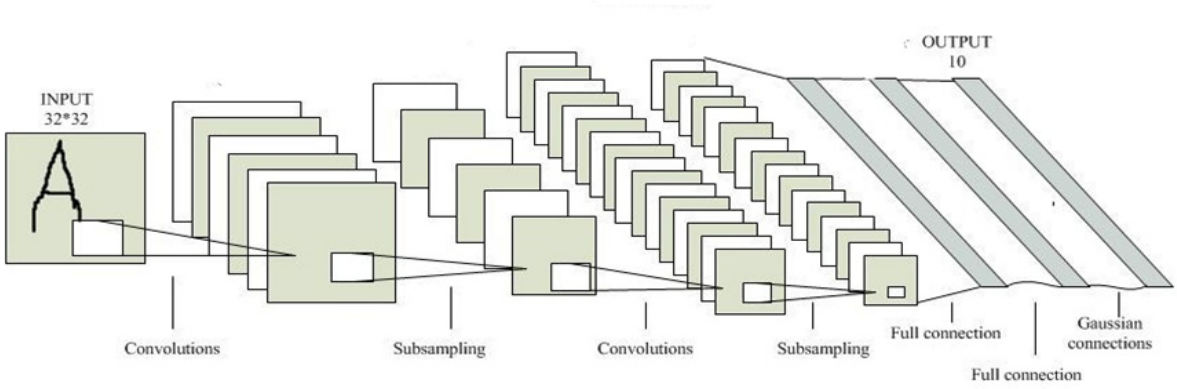

Figure 1: The structure of CNN [12]

3.1 Input layer

When we enter an image, what the computer sees is a matrix of pixel values. Depending on the resolution and size of the image, the computer will see different matrices, such as a 32 x 32 x 3 matrix (3 refers to RGB values). Each digit in the matrix has a value from 0 to 255, which describes the pixel's gray level at that point.

In essence, each picture can be expressed as a matrix of pixel values.

3.2 Convolution layer

The fundamental purpose of convolution is to extract features from the input picture. Convolution uses a small square matrix, which preserves the spatial relationships among pixels, to learn image features.

The convolution operation is shown in the figure 2. The matrix that slides the filter on the original image and performs the convolution operation is called the feature map.

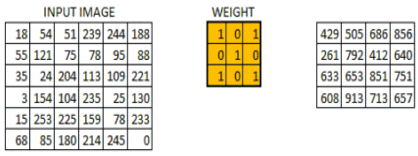

Figure 2: Convolution operation (slide=1) [13]

For every feature map, all neurons share the same weight parameter that is known as filter or kernel (yellow part in the figure 2). The filter is a feature detector for the original input picture. Different filters will produce different feature maps for the same picture. By simply adjusting the filter values, we can perform effects such as edge detection, sharpening, blurring, etc. that mean different filters detect different features, such as edges, curves, etc. from the picture.

The image specifications continue to decrease when the step size is increased. Filling the '0' borders around the input image can solve this problem.

As shown in the figure 3, the original shape of the image is preserved after adding '0' to the image. This is called same padding because the output image is as the same size as the input image.

The size of the image has also been preserved due to the '0' borders.

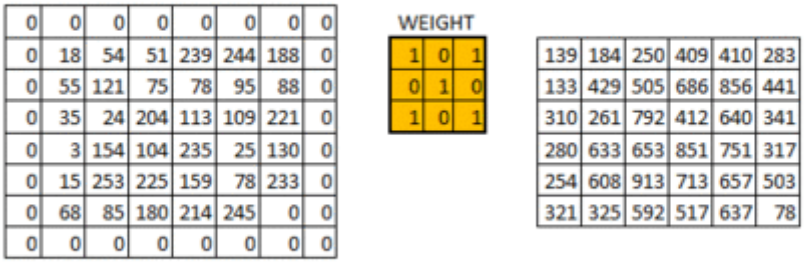

Figure 3: Same padding (slide=2) [13]

3.3 Subsampling layer

The subsampling layer is also called the pooling layer.

Sometimes the image is too large and we need to reduce the number of training parameters .It is required to periodically introduce a pooling layer between subsequent convolution layers. The only purpose of pooling is to reduce the size of the image space.

Pooling can take many forms: Max Pooling, Average Pooling, Sum Pooling, and so on. The most common form of pooling is Max Pooling that is shown in the figure 4.

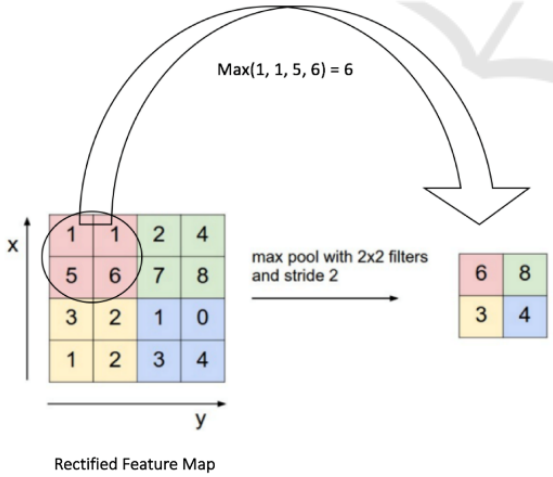

Figure 4: Max pooling [14]

Pooling operations are applied to each feature map separately

After convoluting and pooling, the image still retains most of the information as well as the size of the image has been reduced. The main role of the Pooling layer is subsampling, which further reduces the number of parameters by removing unimportant samples in the Feature Map.

3.4 Fully connected layer

"Fully connected" means that every neuron in the upper level is interconnected with every neuron in the next level.

Convolution and pooling layer is a feature extractor, the fully connected layer is a classifier.

The fully connected layer receives the output of the upper layer (which represents the feature map of the higher-level features) and determines which category these features best fit.

For example, if the program determines the content of a picture as a dog, the feature map with higher values represent some of the more advanced features such as claws or four legs. Similarly, if the program determines the content of a picture as a bird, the feature map with higher values represent some of the more advanced features such as wings or beaks.

The fully connected layer is the Multi-Layer Perceptron that uses the Softmax excitation function as the output layer. The Softmax function converts a vector of any real value into a vector of elements 0-1 and 1.

In general, a fully connected layer observes which category the advanced features most closely matches and what weight they have. When calculating the weight and the dot product between previous layers, we can get the correct probability for the different categories.

This layer process input, and then output an N-dimensional vector (N represents the number of categories). Each number of N-dimensional vector represents the probability of a particular category.

3.5 Output layer

The fully connected layer is the Multi-Layer Perceptron that uses the Softmax excitation function as the output layer.

4 APPLICATION

CNN plays a very important role in different fields, such as image processing, natural language processing, etc. The next few cases are important applications based on CNN.

4.1 Instance

4.1.1 Traffic Statistics

The traffic is modeled as images, and traffic classification tasks become image classification.

Traffic statistics takes pedestrian counting for an example. The task of pedestrian counting uses video image analysis techniques to automatically tally the number of pedestrians passing a scene over a certain period of time.

Firstly, rough human head detection is carried out based on the Adaboost human head detector.

Secondly, a large number of false targets are eliminated by using the CNN-SVM human head classifier through migration learning technology to ensure a high detection accuracy rate. CNN, as a feature extractor, is used to train linear SVM classifier.

Finally, according to the head information, the target of the head of the related person is found accurately by the method of region restriction and feature matching, which greatly improves the counting accuracy.

Traffic statistics can be used in a wide range of areas such as commerce, finance, hospitality and transportation because traffic statistics is an important piece of information in many industries.

4.1.2 Medical image segmentation

In the traditional medical image processing techniques and machine learning algorithms assisting the segmentation process, the need for human participation is a vital thing in the selection of features. The medical images have complex information including uneven gray distribution, large noise and easy deformation of the organ tissue, which increase the difficulty of feature selection and image segmentation.

Convolutional Neural Network, as one of the most important models of deep learning, has produced amazing segmentation results in medical image segmentation.

The traditional CNN is two-dimensional network. If it is directly extended to three-dimensional, more parameters and calculating processes are needed. The three-dimensional network is more complicated, which requires longer training time and more training data. The simple use of two-dimensional data does not make use of three-dimensional features, may result in decreased accuracy. Adhish has adopted a compromise solution for this purpose: use three 2D CNNs, which are shown in the figure 5.

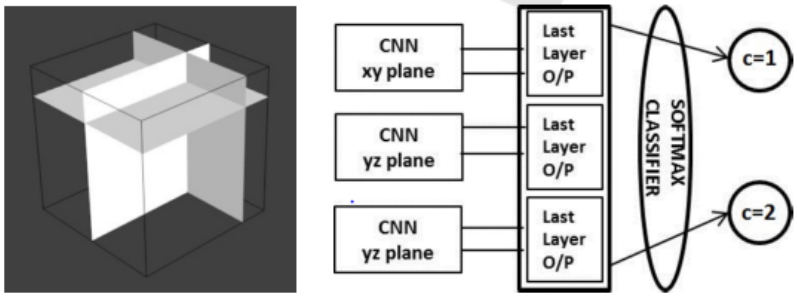

Figure 5 : An example for three-dimensional CNN [18]

Three 2D CNNs are responsible for the processing of the xy, yz, and xz plane, and their outputs are connected together through a Softmax layer to produce the final output.

4.2 Discussion

The application of CNN is not limited to the above cases, and it also includes object detection of self-

driving, handwritten character recognition, etc.

CNN's vision is also very abundant, for example, we can monitor the patients' body image to analyze the patients' conditions through the camera instead of wearing a certain sensor.

From a variety of applications and ideas, CNN's applications also need more in-depth studies.

5 CONCLUSION

In this paper, the Convolution Neural Network's history and structure are summarized. And then several areas of Convolutional Neural Network applications are enumerated. At last, some new insights for the future research of CNN are presented.

According to the above, on one hand, how to improve the performance of the network deserves more in-depth studying. On the other hand, CNN should not be confined to the traditional field of computer vision.

So, how to develop Convolution Neural Network into the complicated, intelligent and real-time application such as self-driving and traffic statistics is also a problem worth noting.

REFERENCES

Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks [J]. Communications of the Acm, 2012, 60(2):2012.

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.

Kim Y. Convolutional Neural Networks for Sentence Classification [J]. Eprint Arxiv, 2014.

Kandel E R. An introduction to the work of David Hubel and Torsten Wiesel [J]. Journal of Physiology, 2009, 587(12):2733.

Yandong L I, Hao Z, Lei H. Survey of convolutional neural network [J]. Journal of Computer Applications, 2016.

Fukushima K. Neocognitron: A hierarchical neural network capable of visual pattern recognition [J]. Neural Networks, 1988, 1(2):119-130.

Lecun Y. LeNet-5, convolutional neural networks [J].

Ballester P, Araujo R M. On the performance of GoogLeNet and AlexNet applied to sketches[C]// Thirtieth AAAI Conference on Artificial Intelligence. AAAI Press, 2016:1124-1128.

http://subscribe.mail.10086.cn/subscribe/readAll.do?columnId=563&itemId=4111011

https://www.analyticsvidhya.com/blog/2017/06/architecture-of-convolutional-neural-networks-simplified-demystified/

https://ujjwalkarn.me/2016/08/11/intuitive-explanation-convnets/

ZHANG Yajun, GAO Chenqiang, LI Pei, LIU Jiang, CHENG Hua. Pedestrian counting based on convolutional neural network [J]. *Journal of Chongqing University of Posts and Telecommunications:Natural Science Edition*,2017, 29(2): 265-271.

Li X, Wang L, Sung E. AdaBoost with SVM-based component classifiers [J]. Engineering Applications of Artificial Intelligence, 2008, 21(5):785-795.

Prasoon A, Petersen K, Igel C, et al. Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network[C]// Medical Image Computing & Computer-assisted Intervention: Miccai International Conference on Medical Image Computing & Computer-assisted Intervention. Med Image Comput Comput Assist Interv, 2013:246.

http:// blog.csdn.net/taigw/article/details/50534376