Understanding of the Convolutional Neural Networks with Relative Learning Algorithms

Jieluo Peng
School of Automation Science and Electrical Engineering, Beihang University. No. 37 Xueyuan Road, Haidian District, Beijing, China
{707031210}@qq.com

Keywords: Convolutional Neural Networks, back-propagation, applications.

Abstract: With the development of calculating ability, image detection has become one of the most popular research fields recently. Convolutional Neural Network is a kind of depth feed-forward network, which has been successfully applied in image recognition. Its hierarchical structure provides the power of weight-sharing and down-sampling. The Convolutional Neural Network effectively combines the two stages of feature extraction and classification in the traditional pattern recognition, and applies the gradient descent algorithm and the back-propagation algorithm to realize the network training. This article will explore the structure and function of Convolutional Neural Networks, with the introduction of the back-propagation algorithm. Then it will introduce how to apply Convolutional Neural Networks in the application of face recognition. The advantages of applying Convolutional Neural Network to face recognition are analyzed. This article also introduces the application of Convolutional Neural Network in other aspects as well.

1 INTRODUCTION

In the current era of rapid development of information technology, how to determine a person's identity becomes particularly important. The traditional authentication technology such as passwords and documents became difficult to meet the needs of society, because they are easily falsified and lost. And the use of biometric identification technology (Matey J. R., 2010; Shan S. G., 2004; Zhao, 2011) to test the identity attract more and more people's attention. The identification of biological features such as human face, fingerprint, retina, iris, etc. which has human unique identification has great research significance and value. And face recognition has non-mandatory, non-contact, intuitive, and simplicity and other characteristics (Chugh T, 2017; Yang, 2013; Yan, 2013)，thus it has been becoming the most important way to identify each other. Especially in the access control system, criminal investigation, video surveillance, network applications and Human-computer interaction (Tang, 2013; Wang, 2007)，it has a wide range of applications. That is why face recognition is a very popular research direction at home and abroad.

However the academic research of face recognition has a history of half a century, and scholars also put forward many efficient and practical methods, the face recognition technology still faces enormous challenges.

Convolutional Neural Networks (CNNs) have developed rapidly in recent years, and widely used in the field of pattern recognition and image processing. CNNs have great advantages in face recognition because of its excellent ability of image recognition Therefore, there is a great significance to study how to apply Convolutional Neural Network to face recognition and solve the problems encountered in its application.

2 BACKGROUND

Face recognition research can date back to the 1888 Galton’s paper published in Nature. He used a set of numbers to represent different facial features and analyzes his own face recognition ability from a psychological point of view. In 1910, Galton proposed the use of the key points of the face and the distance between the key points to form the eigenvector representing the face, and used this

Peng, J.
Understanding of the Convolutional Neural Networks with Relative Learning Algorithms.
In *3rd International Conference on Electromechanical Control Technology and Transportation (ICECTT 2018)*, pages 657-661
ISBN: 978-989-758-312-4

vector for identity recognition, which was the prototype of the face recognition method based on geometric features, yet not really automatic face recognition.

The earliest research on automatic face recognition originated in the 1960s. The representative result was published by Chan at Panoramic Research Incorporated in 1965(Bledsoe W. W., 1965). Domestic face recognition research started late, 1979 Journal of Automation published a "review of artificial intelligence at home and abroad" (Li, 1979), which is the first time to retrieve the domestic journals "face recognition" concept. In 1992, Hong published the "Image Algebra Feature Extraction for Image Recognition" in Journal of Automation (Zi-Quan Hong, 1992) and Zheng Jianping "Standard Frontal Face Recognition in Computer Engineering, Is the earliest academic papers retrieved in the field of face recognition research (Zheng J, 1992).

In the past decades, more and more face recognition technology has attracted the attention of domestic and foreign researchers. Especially in the 21st century, with the rapid development of artificial intelligence, the use of advanced algorithms for face recognition has been pushed to the peak of research. However, face recognition technology has received extensive attention and research. It is still a challenging task because of changes in light, gesture changes, facial expressions and occlusion and other factors.

Convolutional Neural Networks are inspired by the structure of biological neural networks and visual systems. 1962 Hubel and Wiesel through the cat's visual cortical cell research, put forward the experience of receptive field concept (Hubel D H, 1962). In 1980, Fukushima first proposed a theoretical model based on the receptive field Neocognitron (Fukushima K, 1987). Neocognitron was a self-organized multi-layer neural network model. In 1998, Yan LeCun used the gradient descent optimization algorithm and the back-propagation error algorithm to train the convolution neural network on the handwriting, and achieved the best effect in the world at that time (Krizhevsky A, 2012). 2012 Geoffrey Hinton and others in the very well-known ImageNet on the Convolution Neural Network model to obtain the best results of the world. The results was far more than the second, which made the CNN attracting higher attentions.

3 THE STRUCTURE OF CNN

CNN is a specially artificial neural network designed to process two-dimensional input data, and each layer in the network consists of multiple planes. Each plane consists of multiple independent neurons. CNN was inspired by the early Time-Delay Neural Network (TDNN) (Waibel A, 1990). TDNN reduces the computational complexity of network training by sharing weights in the time dimension. It is suitable for processing speech and time-series signals. CNN adopts the weight-sharing network structure to make it more similar to the biological neural network. Compared with the fully connected layer network in each layer, CNN can effectively reduce the learning complexity of the network model, with fewer network connection layers and weight parameters, and thus easier to train.

The basic structure of CNN consists of input layer, convolution layer, pooling layer, fully connected layer and output layer. That is, a convolution layer connected to a pool layer, the pool layer and then connect a convolution layer, and so on. Since each neuron in the output feature of the convolutional layer is locally connected to its input, the corresponding connection uses the weights and local input weighted sum, plus offset value to get this neuron input value. The process is equivalent to the convolution process, and this is why CNN is called (Lecun Y, 1998).

In the convolution layer of CNN, each neuron of the feature map is connected with the local receptivity field of the previous layer. The local features are extracted through the convolution operation. In the convolutional layer, there are many feature maps. Each feature map extracts one feature. When extracting features, neurons in the same feature map share a set of weight convolution kernels. Different feature maps have different weights. And weight parameters are constantly adjusted during the training so that feature extraction is performed in a favorable direction.

There will be a pooling layer after the convolutional layer. Because the previous layer has a large amount of overlap when window sliding convolution is done. There is redundancy in the convolution value. A pooling layer is needed to simplify the output of the convolution layer. Pooling layer will retain the main information convolutional layer, while reducing the parameters and calculation, to prevent over-fitting. The most common pooling is max-pooling, which takes the largest feature points in the field. Max-pooling transmits only the parameters with largest value and takes others away.

When there is backward to put the maximum position, the other positions can be filled with zero. There are also mean-pooling (Boureau Y L, 2011) and average pooling in the pooling layers.

After multiple convolutional layers and pooling layers, there will be one or more fully connected layers. Each neuron in the fully connected layer is fully connected to all the neurons in its previous layer. The fully connected layer can integrate regional information in the convolutional layer or the pooling layer with category classification (Sainath T N, 2013).To improve the performance of CNN, the ReLU function is generally used for the excitation function of each neuron in the fully connected layer (O'Shea K, 2015).

4 BACKPROPAGATION ALGORITHM

For a fixed sample set containing m samples $\{(x^{(1)},y^{(1)}),\ldots,(x^{(m)},y^{(m)})\}$, we can use the gradient descent algorithm to solve the neural network. Specifically, for a single sample (x,y), its cost function is:

$$J(W,\mathrm{b};x,y)=\frac{1}{2}\left\|h_{W,b}(x)-y\right\|^2$$

And for a data set containing m samples, its overall cost function is:

$$J(W,b)=\left[\frac{1}{m}\sum_{i=1}^{m}J(W,b;x^{(i)},y^{(i)})\right]+\frac{\lambda}{2}\sum_{l=1}^{n_l-1}\sum_{i=1}^{s_l}\sum_{j=1}^{s_{l+1}}\left(W_{ji}^{(l)}\right)^2$$
$$=\left[\frac{1}{m}\sum_{i=1}^{m}\left(\frac{1}{2}\left\|h_{W,b}(x^{(i)})-y^{(i)}\right\|^2\right)\right]+\frac{\lambda}{2}\sum_{l=1}^{n_l-1}\sum_{i=1}^{s_l}\sum_{j=1}^{s_{l+1}}\left(W_{ji}^{(l)}\right)$$

The first term in the equation is a mean square error term. The second term is a regularization term, which is also called weight attenuation term, whose purpose is to reduce the magnitude of the weight and prevent over-fitting. Weight decay parameters λ are used to control the relative importance of two terms in a formula.

Our goal is to minimize $J(W,b)$ as a function of W and b . To train our neural network, we will initialize each parameter $W_{IJ}^{(l)}$ and each $b_i^{(l)}$ to a small random value near zero, and then apply an optimization algorithm such as batch gradient descent. Gradient descent method in each iteration according to the following formula to update the parameters W and b :

$$W_{ij}^{(l)}=W_{\mathrm{ij}}^{(l)}-\alpha\frac{\partial}{\partial W_{ij}^{(l)}}J(W,b)$$

$$b_i^{(l)}=b_i^{(l)}-\alpha\frac{\partial}{\partial b_i^{(l)}}J(W,b)$$

And α is the learning rate. The most important step is to calculate the partial derivative. Back-propagation algorithm is a very effective way to calculate partial derivatives.

$$\frac{\partial}{\partial W_{ij}^{(l)}}J(W,b)=\left[\frac{1}{m}\sum_{i=1}^{m}\frac{\partial}{\partial W_{ij}^{(l)}}J(W,b;x^{(i)},y^{(i)})\right]+\lambda W_{ij}^{(l)}$$
$$\frac{\partial}{\partial b_i^{(l)}}J(W,b)=\frac{1}{m}\sum_{i=1}^{m}\frac{\partial}{\partial b_i^{(l)}}J(W,b;x^{(i)},y^{(i)})$$

These two equations are the partial derivatives of the cost function of a single sample by using a backpropagation algorithm.

We must first conduct "forward conduction" operation. The purpose is to calculate all the network activation values. By using the forward conduction formula, we can compute the activations for layers L_2 , L_3, and so on up to the output layer L_{n1}. Then, for each node i in layer l, we would like to compute an "error term" $\delta_i^{(l)}$ that measures how much that node was "responsible" for any errors in our output. For an output node, we can directly measure the difference between the network's activation and the true target value, and use that to define $\delta_i^{(n_l)}$ (where layer n_1 is the output layer). How about hidden units? For those, we will compute $\delta_i^{(l)}$ based on a weighted average of the error terms of the nodes that uses $\alpha_i^{(l)}$ as an input. Afterwards, for each node i in the first layer, we calculate its "residual" $\delta_i^{(l)}$, which shows how much the node affected the residual of the final output value. For the final output node, we can directly calculate the difference between the activation value generated by the network and the actual value. We define this gap as $\delta_i^{(n_l)}$ For each output unit i in layer n_1 the output layer set:

$$\delta_i^{(n_l)}=\frac{\partial}{\partial z_i^{(n_l)}}\;\frac{1}{2}\left\|y-h_{W,b}(x)\right\|^2=-(y_i-a_i^{(n_l)})\cdot f'(z_i^{(n_l)})$$

And the residual of the i node in layer l is calculated as follows:

$$\delta_i^{(l)}=\left(\sum_{j=1}^{s_{l+1}}W_{ji}^{(l)}\delta_j^{(l+1)}\right)f'(z_i^{(l)})$$

After computing the desired partial derivatives, we can bring it into the gradient descent algorithm to update the parameters. Then we repeat the iterative step of the gradient descent method to reduce the cost function value and solve our neural network.

5 THE APPLICATIONS OF CNN

5.1 Image classification

Image classification is through the analysis of the image. The image is divided into a certain category. The main emphasis is on the image to determine the overall characteristics. In the field of image classification, the ImageNet Larget Scale Visual Recorder-nition Challenge (ILSVRC) is one of the most important events in evaluating image classification algorithms. ILSVRC2012 was a turning point. AlexNet, for the first time to apply deep learning to large-scale image classification and achieved good results. Since then, deep learning-based convolutional neural networks have begun to occupy the dominant position of ILSVRC. The new CNN model has been put forward constantly. When the game record is refreshed, the ability of CNN model to extract image features is also continuously improved. At the same time, the emergence of large-scale data sets such as ImageNet also aided the training of CNN, which also promoted CNN's classification learning.

5.2 Object detection

In the field of computer vision, object detection is a more complicated problem than image classification. There are many objects in a picture, and they belong to different categories. We need to classify each of them into different types. Therefore, the CNN model used in object detection will be more complicated. At present, CNN-based object detection model mostly attributes the object detection problem to two sub-problems of how to propose candidate regions and how to classify the candidate regions. In the development of CNN-based object detection, many models focus on the optimization of training methods and procedures. It makes the CNN model improved in accuracy.

5.3 Gesture estimation

With the development of various online games and the popularization of animation video, it has become a very hot topic to recognize and understand the human gesture in the image. Attitude estimation is one of the most important computer vision challenges nowadays, because it can be quickly applied to character tracking, motion recognition and video-related video analysis. The traditional approach is local modeling, yet the ability to express is limited. While the CNN has the ability to deal with the entire picture so that the use of attitude estimation has achieved good results.

5.4 Image segmentation

CNN has achieved great success in image classification, target detection and pose estimation. The further development is the prediction of every pixel in the image. This task is image segmentation. Image segmentation is such an issue: for a graph, it may have multiple objects, multiple people or even multi-layer background. The image segment is use to predicted or classify each pixel on the original graph, to the part it belongs.

6 SUMMARY

The upsurge of deep learning makes artificial neural network once again become the hot spot of research. Convolutional Neural Network, as a kind of deep learning, integrates three core ideas of local receptive field, weight sharing and down-sampling structure effectively. It effectively combines traditional mode recognition, feature extraction and classification. The application of gradient descent algorithm and back propagation algorithm works for network training.

This article analyzes the structure of CNN in detail. The structures and functions of convolutional layer, pooling layer and full connection layer in CNN are respectively introduced, as well as the back-propagation algorithm.

This article describes how to make use of CNN for face recognition, which is the hot topic at present. It summarizes and discusses previous studies and analyzes the advantages of using CNN to face recognition, as well as the existing shortcomings and the future development prospects.

REFERENCES

Matey J. R. & Bergen J. R. Methods and systems for biometric identifications: U.S. Patent 7,751,598[p].2010-7-6.

Shan S. G. Research on the several issues in face recognition [D]. Institute of Computing

Technology, Chinese Academy of Sciences，2004

Zhao X. P. Summary of the development of biometric features recognition [J]. Forensic science and technology，2011,06:46-50

Chugh T, Singh M, Nagpal S, et al. Transfer Learning Based Evolutionary Algorithm for Composite Face Sketch Recognition[C]// Computer Vision and Pattern Recognition Workshops. IEEE, 2017:619-627.

Yang C, Zhang L, Lu H, et al. Saliency Detection via Graph-Based Manifold Ranking[J]. 2013:3166-3173.

Yan Q, Xu L, Shi J, et al. Hierarchical Saliency Detection[C]// Computer Vision and Pattern Recognition. IEEE, 2013:1155-1162.

Tang J D. Research on the technology of picture feature extracting and matching in face recognition [D]. Dalian Maritime University, 2013

Wang X J. Research on the automatic face recognition based on statistic learning [D]. University of Science and Technology of China, 2007

Galton F. Personal Identification and Description[J]. Journal of the Anthropological Institute of Great Britain & Ireland, 18(973):177-191.

Galton F, Galton F. Numeralised Profiles for Classification and Recognition[J]. Nature, 1910, 83(1):127-130.

Bledsoe W. W., Chan H. A Man-Machine Facial Recognition System; Some Preliminary Results. Technial Report,PRI 19A.Palo Alto,USA:Panoramic Research Incorporated,1965.

Li C. C., Wang Y. J., Tu X. Y., et al. Summary of domestic and foreign researches on artificial intelligence. Acta Automatica Sinica,1979,5(1):74-87.

Zi-Quan Hong. Algebraic feature extraction of image for recognition[J]. Acta Automatica Sinica, 1992, 24(3):211-219.

Zheng J. RECOGNIZING A TYPICAL HUMAN FACE FROM THE FRONT SIDE[J]. Computer Engineering, 1992.

Hubel D H, Wiesel T N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex[J]. Journal of Physiology, 1962, 160(1):106.

Fukushima K. Neural network model for selective attention in visual pattern recognition and associative recall[J]. Applied Optics, 1987, 26(23):4985-92.].

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.

Waibel A, Hanazawa T, Hinton G, et al. Phoneme recognition using time-delay neural networks[J]. Readings in Speech Recognition, 1990, 1(2):393-404.

Lecun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11):2278-2324.

Boureau Y L, Roux N L, Bach F, et al. Ask the locals: Multi-way local pooling for image recognition[C]// IEEE International Conference on Computer Vision. IEEE, 2011:2651-2658.

Sainath T N, Mohamed A R, Kingsbury B, et al. Deep convolutional neural networks for LVCSR[C]// IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 2013:8614-8618.

O'Shea K, Nash R. An Introduction to Convolutional Neural Networks[J]. Computer Science, 2015.