Survey of image recognition technology based on convolution neural network

Zhichao Xian

Century College, Beijing University of Posts and Telecommunications, China

Keywords: Convolutional Neural Network (CNN); image recognition; deep learning.

Abstract: In recent years, Convolutional Neural Network (CNN) has made a series of breakthrough research results in the fields of image recognition and so on. The powerful ability of CNN for image recognition attracts wide attention. However, due to the widespread application of CNN's deep neural network model in the field of image processing, many of these problems have been solved and even the recognition accuracy of some algorithms has exceeded the human eye recognition accuracy. Therefore, CNN has very important research value in image recognition. This article discusses the current status and prospects of its technology development.

1. Introduction

CNN is currently the most widely used deep neural network model in the image processing field. It has a wide range of applications, such as image processing, audio processing, natural language processing. Image recognition based on CNN, features are automatically learned from various data, but features in traditional image recognition algorithms are labeled manually. Good features can improve the performance of the recognition systems. Besides, CNN does not require human intervention to extract features. Instead, the CNN automatically extracts image features and classification, which greatly reduces the calculation cost.

2. Principle and composition of CNN

2.1 Composition of CNN

CNN is composed of input and output layers and multiple hidden layers. The hidden layer can be divided into convolution layer, pooling layer and fully connected layer. CNN's input is usually a two-dimensional vector which can have height, for example, RGB image. Convolution layer is the core of CNN, and its parameters are composed of a group of learnable filters or kernels, which have a small receptive field and extend to the whole depth of input volume. During feedforward, each filter convolutes the input, calculates the dot product between the filter and the input, and generates a two-dimensional activation graph of the filter (input general two-dimensional vector, but may have a height (i.e. RGB)). Generally, the pooling layer changes the maximum value of four pixels per neighborhood into one pixel. The all connected layer is a conventional neural network, which is used to fully connect the advanced features obtained from multiple convolution layers and multiple pooling layers.
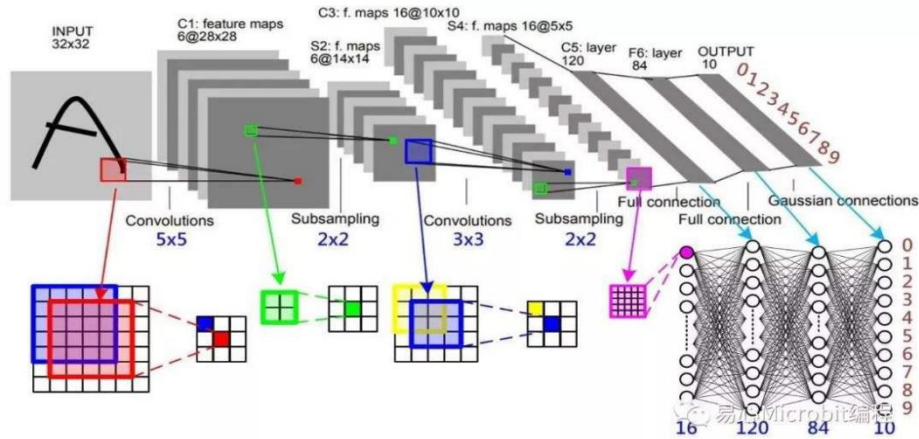
Figure 1 Convolutional Neural Network (CNN) lenet-5

2.2 Principle of CNN

Convolution is the main function of extracted features using the network, has a certain shift invariance and also has the effect of a certain dimension reduction. Convolution layer is composed of different convolution kernels, each convolution kernel used to calculate different feature maps. A number of different kernels describe the characteristics of the complete map, the parameters of the convolution operation sharing mechanism to reduce the complexity of the model and make the neural network training more easily Pooling operation will be referred to as sampling. Pooling window is set according to the set of rules, such as the maximum or average, moving in sequence in the input feature map, an element of the output characteristic figure, mainly is a function of dimension reduction. At the same time the neural network has better performance for small displacement or change the connection to the former layer in all the characteristics of the figure is connected to the current layer connection diagram Which can form a kind of global semantic information. It is essentially a perceptron because of the nonlinear nature, the network has the ability of approximation. The activation function can refer to the non-linear nature of the neural network. The excitation function Relu avoids the problem of gradient disappearance. It has faster convergence and easier learning and optimization, makes the network sparse and can also alleviate the over fitting problem.

3. Research progress of image recognition based on convolutional neural network

Research progress on convolutional neural network deeply affects the technology of image recognition. People found that when increasing the depth of the neural network, there will be a problem, that is, the accuracy will rise first, then reach saturation, and then continue to increase the depth will lead to the decline of accuracy. This is not an over fitting problem, because not only in the test set, but also in the training set itself. In the traditional convolution layer or full connection layer, there are some problems such as information loss. Therefor, RESNET was proposed by Kaiming he and the other four Chinese in Microsoft Research Institute.[2] Through the successful training of 152 layer deep neural network by using residual units, RESNET won the championship in ilsvrc 2015 competition, with the error rate as low as 3.57%, and the parameter amount is lower than vggnet, with good results. It uses a connection method called "shortcut connection". It makes a reference (x) for the input of each layer, and learns to form the residual function instead of learning some functions without reference (x). This kind of residual function is easier to optimize and can make the number of network layers deepen greatly.
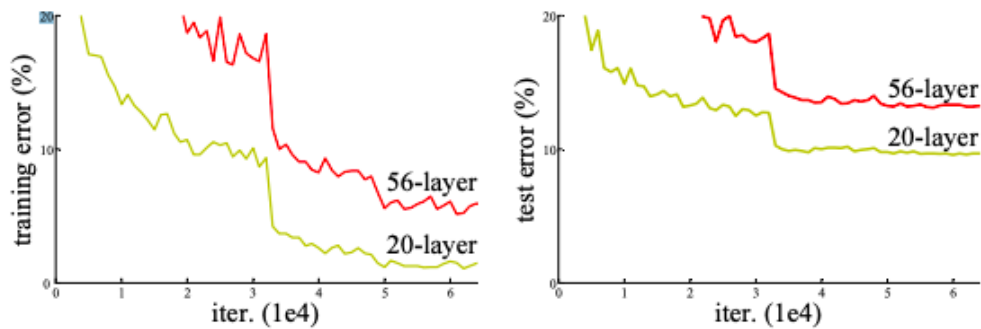
Figure 2. Training error (left) and test error (right) on CIFAR-10 with 20-layer and 56-layer "plain" networks. The deeper network has higher training error, and thus test error. Similar phenomena on ImageNet are presented in Fig. 2.

RESNET solves this problem in a certain sense by bypassing the input information directly to the output to protect the integrity of the information. The whole network only needs to learn the difference between the input and output to simplify the learning objectives and difficulties.

4. Application analysis of image recognition based on convolutional neural network

As an important technical means in the field of image recognition, CNN has broad application prospects. The existing research is worth discussing such as face recognition. Face recognition technology is the most commonly used technical means in daily life and is widely used in video surveillance, access control systems, e-commerce, etc. Face recognition comparison is one of the most core technologies in face recognition systems. The accuracy of face recognition comparison directly affects the accuracy of the entire system.

Fu et al. [3] proposed a new CNN architecture-Guided Convolutional Neural Network (Guided-CNN), which solves the problem of low matching of face images at different resolutions. In different test methods, All achieved a good accuracy rate, and the highest matching accuracy rate reached 97.4%. Deng et al. [4] proposed an improved CNN to solve the problem of low recognition rate of densely packed objects. The method is based on the latest residual network (ResNets) and consists of two subnetworks: an object recommendation network and Object detection network.

5. Defects and development of image recognition based on convolutional neural network

5.1 Defects of image recognition based on convolutional neural network

Based on an area of empirical research, CNN research is still relatively backward so the research on CNN is still relatively important. There is still a large space for research on the structure of convolutional neural network. Over fitting and network degradation require more reasonable network design. The parameters of volume and network are very large. The quantitative analysis and research of parameters are still a problem to be solved. CNN's structural model improvement can't meet the current needs of the old data sets. The data sets have some important significance for structural research, migration learning and training, with more numbers and categories and more complex data situation is the development trend of the current data set

5.2 Development of image recognition based on convolutional neural network

Multi layer architecture has a great contribution to the field of image recognition,The most prominent multi-layer architecture used in the field of computer vision. Before the recent success of deep learning based network, the most advanced computer vision system for recognition relies on two separate but complementary steps. The first step is to transform the input data into a proper form through a set of manually designed operations . Secondly, the transformed data is usually used to train

some types of classifiers to identify the content of the input signal. Generally speaking, the performance of any classifier will be seriously affected by the transformation method used.

Although various CNN models continue to promote the current best performance in various computer vision applications, there is still limited progress in understanding how these systems work and why they are so effective. This problem has aroused the interest of many researchers, so there are many ways to understand CNN. Generally speaking, these methods can be divided into two directions: visualizing the learned filter and the extracted feature map, expanding study inspired by the biological method of understanding the visual cortex.[5]

5.2.1 Based on the above discussion, there are the following key research directions of visualization based methods

First of all, it is very important to develop a more objective method for visual evaluation, which can be achieved by introducing indicators of the quality and meaning of the visual image generated by the evaluation.

In addition, although it seems that network centric visualization methods are more promising, it seems necessary to standardize their evaluation process. One possible solution is to use a benchmark to generate visual results for networks trained under the same conditions. Such a standardized approach, in turn, enables indicators based on assessment rather than current interpretative analysis. [6]

5.2.2 The following are the potential research directions of the method based on the introduction study

Data sets are organized in a common and systematic way, with different challenges common in the field of computer vision , and there must be more complex categories. In fact, such datasets have emerged recently. Using annotation study on such datasets and analyzing the confusion matrix, we can determine the pattern of CNN architecture errors and achieve a better understanding. In addition, the systematic study on the influence of multiple collaborative expansions on model performance is of great concern. Such a study should extend our understanding of how independent units work.

Finally, these controlled methods are promising research directions in the future, because they can give us a deeper understanding of the operation and representation of these systems compared with the completely learning based methods.

6. Conclusion

In this paper, the principle and composition of convolution neural network are introduced, the achievements of the full convolution neural network in the field of image processing are systematically combed and analyzed, focusing on the analysis of some basic concepts of the research direction of image processing, as well as the advantages of the current research such as multi-layer architecture and the areas that can be improved. At the same time, some of the latest methods are analyzed and summarized. Convolutional nerve is a research field with high heat, and has a broad research prospect.

References

[1] Saad Albawi, Tareq Abed Mohammed. Understanding of a Convolutional Neural Network, ICET2017, Antalya, Turkey,Aug.

[2] Kaiming He, Xiangyu Zhang, Shaoqing Ren et al, Deep Residual Learning for Image Recognition, arXiv:1512.03385v1 [cs.CV] 10 Dec 2015.

[3] Fu T,Chiu W,Wang F.Learning guided convolutional neural networks for cross-resolution face recognition[C]// Proceedings of 27th International Workshop on Machine Learning for Signal Processing(MLSP).Tokyo, p.2017:1-5. Information on: http://neuralnetworksanddeeplearning.com/

[4] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks.In NIPS, 2012,Volume 1,December.p.1097-1105.

[5] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri. Learning spatiotemporal features with 3D convolutional networks. In ICCV, 2015,p. 4489-4497.

[6] Geoffrey Hinton,Vinod Nair Rectified Linear Units Improve Restricted Boltzmann Machines.ICML, 2010. P. 807-814.

[7] Deep Learning, book by Ian Goodfellow, Yoshua Bengio, and Aaron Courville.The MIT Press.2016.Cambridge.

[8] Information on: http://cs231n.stanford.edu/

[9] Kalchbrenner N, Grefenstette E, Blunsom P. A Convolutional Neural Network for Modelling Sentences[J]. Eprint Arxiv, 2014, 1.

[10] Recurrent Convolutional Neural Network for Text Classification, Siwei Lai etc.AAAI'15: Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence. 2015,January. p.2267-2273

[11] Yoon Kim. Convolutional Neural Networks for Sentence Classification.Empirical Methods In Natural Language Processing,2014.p.1746-1751

[12] Zhang Ye, et al. A Sensitivity Analysis of (and Practitioners' Guide to) Convolutional Neural Networks for Sentence Classification.International Joint Conference On Natural Language Processing,2017.Volume 1,p.253-263