Evaluation Methods and Applications in Image Recognition based on Convolutional Neural Networks

Yiqi Lin
Department of Computer Science, Viterbi School of Engineering, University of Southern California, Los Angeles 90007, America
yiqil@usc.edu

Abstract. With the development of the technology in computer science and relative subjects, Machine Learning (ML) plays more and more important role in practical implementations. One of the most promising directions is the image recognition. This paper focuses on Convolutional Neural Networks (CNNs) which is the most common method in image recognition or pattern recognition issues. The basic understanding of CNNs' structure will be introduced, including the introduction of different layers. Also, two possible evaluation methods, based on NORB and MNIST respectively, will be displayed. The last part is some applications, followed by a summary.

1. Introduction
Human has a complex visual recognition system. We distinguish and classify objects independently, then save the information with memory for future uses. We can recall those fragments for some reasons, and the very first step is visual recognition. Actually, our brain is a deep and complex neural network. To process information more efficiently, researches worked on building proper structure to deal with massive data as well as provide ideal outputs. Many attempts have been made, and one potential soliton came out when researchers found some common points between computing architectures and neuroscience [1]. Artificial neural network is an artificial intelligence method which aims to imitate the brain or neural structure for computing. However, in the image processing or image cognition, even if it utilizes a serial of operations, there are still some noises or high error rate in practical uses. Thus, the development of relative methods or algorithms is a hot topic. In recent decades, Convolutional Neural Network (CNN) is one of the artificial neural networks, which has promising future in two-dimension image recognition. The original idea of CNNs is to simulate the mechanism of creature brains in scales, dealing with input and giving reliable outputs by computing [2]. The biggest feature of CNN is that it could extract visual features automatically and reduce calculation amount. Right now, it is still a popular research direction in ML and image processing.

CNNs have many application scenarios, including image recognition, pattern recognition, semantic recognition, video analysis, nature language processing and so on [3]. Different layers in CNNs keep the basic structure of neural networks as well as provide potential capability of auto feature extracting and multi-layer sensors. This characteristic relief plenty of computing pressure. There is no need of lots of pre-operations like noise reduction when applying CNNs. The related researchers put forward multilayer structure in CNNs, including image processing layer, convolution layer, whole connection layer, input layer and output layer, etc. Besides, to improve the effectiveness of image recognition, the rational allocation of network structure and parameters has significant impact [4].

This paper is going to talk about the layer structure of CNNs, the evaluation process and practical
implementations, including facial recognition, silence plate recognition and license plate location.

2. Convolutional Neural Networks

In the image processing, the very first step is to decompose the image information into pixel vectors, for example a picture with 1,000x1,000 pixels can be represented by 1,000,000 vectors. In this way, if the number of hidden layers equal to input layers, there will be 1,000,000x1,000,000 pieces of data, which is too huge to compute. Thus, the main task of image recognition is to reduce computing amount, to accelerate training speed [5].

CNNs are hierarchical neural networks combing different units and layers. To solve the issue above, CNNs could extract more than one features at one time and then use full-connected method to reduce computing times. Generally, the whole process contains three part. The first part is using convolutional layer to extract image features. The second part is to extract useful data from image information among layers to reduce computing quantity. And the third part is to output result and compare it with original image to complete training.

2.1. Convolutional layer

As shown in the name, in CNNs, the main function of convolutional layers is using convolution with kernels to extract features or useful data. More than one kernels could be used in one layer, which means convolutional layer could extract more than on features at one time. Besides, convolutional layers and pooling layers are appearing alternatively. The feature maps in the convolutional layers is associated with the local regions in previous layer. The features need to be calculated with convolutional sum, weighting, and non-linear functions with corresponding area in previous layer. In one feature map, only one kernel will be shared.

The principal advantages of this function are: 1) image information in adjacent area has high correlation and local characteristics can more easily be detected; and 2) local statistical characteristics of image and other signals remain unchanged. If one feature occurs in one part of the image, theoretically it is possible to be any part of the image. Therefore, CNNs use weights sharing to extract image features and reduce calculation amount. Specifically, CNNs use the convolutional areas with same weight to identify the different parts of the image. From mathematics perspective, the filtering process is the convolution operation process of discrete function. Thus, it is called Convolution Neural Network [6].

2.2. Sub-sampling/pooling layer

The basic purpose of sub-sampling/pooling layer is to offer translation invariance to extracted features to increase their robustness. There are two main ways: one is max-pooling and the other is average pooling. Both have the same process, pooling the inputs. However, max-pooling only chooses the max value in each pooling while average pooling chooses the average. No matter which pooling function is chosen, the extracted features will be more stable against noise and distortion.

2.3. Classification layer

Not all the layers in CNNs are fully connected. Full-connection will cause huge computing quantity. The image processing layers between input and output are usually called hidden layers. Among those hidden layers, some of them are partially connected based on regions or relative features, which could reduce the computing quantity. However, the output layer or topmost layer must be fully connected, with one-detention vectors. Thus, CNNs need classification layers to cluster feature maps, especially in describing large image with huge data [7]. Also, those classifiers need training by input features. One brief version of CNNs' whole process is shown in Figure 1.
3. Evaluation methods
This part will introduce two common datasets for image recognition evaluation, including basic introduction and implementations. Generally, in ML, a mature system should be trained by several steps. Before the training, the proper algorithms need to be set. And the testing datasets usually play as verification method to evaluate the error rate and stability of the algorithms.

3.1. NORB dataset
Caltech-10 is one of the most common dataset for textural recognition with few pose variations [9]. Another more object-based dataset is NORB [10]. The “small NORB” dataset (v1.0) consists images of 50 toys belonging to 5 generic categories: four-legged animals, human figures, airplanes, trucks, and cars. The objects were imaged by two cameras under 6 lighting conditions, 9 elevations (30 to 70 degrees every 5 degrees), and 18 azimuths (0 to 340 every 20 degrees). Several examples from “small NORB” dataset are displayed in Figure 2.

Figure 1. Architecture of a CNN [8]

Figure 2. Samples from small NORB dataset [6]
3.2. MNIST dataset

MNIST [11] is a database for the using of learning techniques and pattern recognition. It is a useful dataset of handwriting digits. Comparing to the shapes, handwritings are more complex and need different feature extracting strategy. MNIST consist of a set of training data with 60,000 examples, and a set of testing data with 10,000 examples. All of them are grey-scale 28x28 pixel digit images. The MNIST training set is composed of 30,000 patterns from SD-3 and 30,000 patterns from SD-1. And SD-3 is much cleaner and easier to recognize than SD-1. SD-1 contains 58,527 digit images written by 500 different writers.

4. Applications

CNNs are implemented in widespread fields, because of its effectiveness and efficiency. In this section, two popular applications of CNNs will be introduced, including face recognition and license plate recognition.

4.1. Face recognition

One of the most popular fields in the ML is face recognition. Several algorithms have been attempted to achieve the goals, such as Local Face Analysis [12], Eigenface [13], Hidden Markov Model [14], etc. However, Neural Networks, especially CNNs has an obvious strength in self-training. And it could have a higher speed in the processing when the dimension reduction has been applied.

For example, the ORL database (http://www.cam-orl.co.uk) is a wide-use database for face recognition, containing images from 40 individuals, each providing 10 different images. There are totally 400 different pictures shot between 1992 and 1994 in the Olivetti Research Laboratory, Cambridge University. These pictures were taken respectively in different times. The main differences among them were light, expressions and facial details. Researchers utilize ORL to train or test their networks in face recognition issues.

4.2. License plate recognition

Vehicle control has becoming more and more important in civilized cities. It helps with understanding the demographic situation and traffic condition in a certain area. Comparing with the other scenes, the recognition of license plate always takes place on high-speed moving objects. Besides the ambient light or climate changes will also influence the results. Ou et al. [15] purposed an advanced algorithm with CNN for license plate recognition. They stated that the significant strength of CNN is good accuracy and high robustness, against the potential noise and jamming.

5. Conclusion

Dealing with the practical challenges in image recognition, Convolutional Neural Network provides a meaningful solution with high effectiveness and efficiency. The most significant features of CNNs are feature extraction, dimension reduction, and self-training process with other algorithms. The weight-sharing protocol and partial full-connection make CNN one of the most popular and useful structures for DL.

This paper briefly summarizes the basic layout of CNNs, as well as providing common training/test databases with introduction of realistic implementations. For the further researches, several challenges should be under considerations as follows. First, the training dataset should be specified by certain issue. Maybe it is not like one-to-one correspondence. However, choosing a proper training set will be important. Second, in some applications like silence plate recognition, the laboratory-level algorithm is insufficient. Research should better consider the working scene or testing environment, since some extra noises could be ignored.

References

[1] Zeiler, M. D., & Fergus, R. (2014, September). Visualizing and understanding convolutional networks. In European conference on computer vision (pp. 818-833). Springer, Cham.
[2] Lee, K., & Park, D. C. (2015). Image Classification using Fast Learning Convolutional Neural Networks. Advanced Science and Technology Letters, 113, 50-55.

[3] Koushik, J. (2016). Understanding Convolutional Neural Networks. arXiv preprint arXiv:1605.09081.

[4] Yi, C., Deng, Y. (2017). Image recognition with multi-channel Convolutional Neural Networks. Journal of Henan University of Science & Technology: Natural Science, (3), 41-44.

[5] Ding, X. (2014). Research on the applications of BP Neural Networks and Convolutional Neural Networks in character recognition (Master's thesis, Huazhong University of Science and Technology).

[6] Li, D. (2016). Research on the license plate recognition based on Convolutional Neural Networks (Master's thesis, Xiangtan University).

[7] Zhao, J. (2016). Research of Substation Monitoring Image Recognition Approach Based on Convolutional Neural Networks (Doctoral dissertation, North China Electric Power University).

[8] Cireşan, D. C., Meier, U., Masci, J., Gambardella, L. M., & Schmidhuber, J. (2011). High-performance neural networks for visual object classification. arXiv preprint arXiv:1102.0183.

[9] Jarrett, K., Kavukcuoglu, K., & LeCun, Y. (2009, September). What is the best multi-stage architecture for object recognition? In Computer Vision, 2009 IEEE 12th International Conference on (pp. 2146-2153). IEEE.

[10] LeCun, Y., Huang, F. J., & Bottou, L. (2004). Learning methods for generic object recognition with invariance to pose and lighting. In Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on (Vol. 2, pp. II-104). IEEE.

[11] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278-2324.

[12] Bartlett, M. S., Movellan, J. R., & Sejnowski, T. J. (2002). Face recognition by independent component analysis. IEEE Transactions on neural networks, 13(6), 1450-1464.

[13] Zhang, J., Yan, Y., & Lades, M. (1997). Face recognition: eigenface, elastic matching, and neural nets. Proceedings of the IEEE, 85(9), 1423-1435.

[14] Nefian, A. V., & Hayes, M. H. (1998, May). Hidden Markov models for face recognition. In Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on (Vol. 5, pp. 2721-2724). IEEE.

[15] Ou, X., Xiang, C., Zhan, X., Peng, X., and Shi, Y. (2016). License Plate Digital Character Recognition based on Convolution Neural Network. Journal of Chengdu Technological University, 19(4), 26-30.