Image Classification And Recognition Based On The Deep Convolutional Neural Network

Yuan-yuan WANG¹, Long-jun ZHANG²,Yang XIAO¹, Jing XU¹,* ,You-jun ZHANG¹

1. The College of Information and Electric Engineering, Shenyang Agricultural University, Shenyang, 110866, China
2. STATE GRID XINJIANG INFORMATION & TELECOMMUNICATION COMPANY, 830000, China
*Corresponding author

Keywords: Machine Learning; Computer Vision; Deep Learning; Convolutional Neural Network.

Abstract

With the development of the information age, there were a lot of data whose features couldn't be extracted or predicted effectively in real life. One of the core function of computer vision technology is to classify and recognize, with classification and recognition as its summit mission of object detection and object positioning. Due to image data were affected by multiple factors such as illumination, environment, angle, certain object features couldn't be established by manual coding, and it is hard for high latitude data in a computer to realize real-time object detection, object localization, classification and recognition. Therefore the higher accuracy in classification could be obtained with the help of GPU of high performance and the large scale pre-training on the super database Image Net, based on the most advanced deep convolution neural network algorithm. The complete optimization training was conducted on data sets Pascal Vision Object Classes (VOC), and the real-time object detection, object localization, classification and recognition were realized by high performance GPU of NVIDIA.

1 Introduction

Machine learning is one of the most excited technology in recent years. It is an algorithm that is used to predict the unknown data by the features that is analyzed and attained automatically from data, and it is widely used in such fields as speech and handwriting recognition, robots, strategy games and computer vision.

LeCun has published the convolutional neural network “LeNet-5” in 1998, which succeeded in the function of writing recognition, and has become an example in the deep convolutional neural network learning nowadays. The rapid increase of the performance of GPUs makes it possible to realize the large-scale training on deep convolutional neural network. The most advanced computer vision technology, namely the convolutional neural network, can classify the image data intelligently.

The computer vision technology can also be used in other fields such as robots, automatic driving, unmanned aerial vehicle, etc.

2 Main Content

Image data belong to high latitude data, the characteristics of learning technology in deep learning has solved the problem in feature extraction of engineering. In the framework of deep learning, convolutional neural network is used to make effective feature extraction. The unknown data can't be predicted due to the over-fitting of the basic convolutional neural network [4]. The high latitude data like image must be optimized in order to quicken the calculation speed by better feature extraction, as its amount of computing is increasing in multiples of magnitude. After the feature extraction is finished on the optimized network, it is sent into the classifier for classification and recognition. It is necessary to classify and recognize the image and judge the probability of the object that meet the conditions. The use of the complex cross entropy loss function in the classifier SoftMax [5] not only accords with the sparse connectivity of brain, but also provides the intuitive output, which is just the probability of the object that is judged. Therefore SoftiMax is chosen as the classifier. Back Propagation (BP) is introduced to assess the gradient of network so that it can decrease the value of the loss function to the least to achieve the more accurate classification. The gradient descent technology [6, 7] is used to update and adjust the network weights in order to minimize loss value.

3 Architecture Core

The core of system architecture is deep convolutional neural network. Convolution layer, in the whole network framework, is used in the feature extraction of the original image, and full-connect layer is used to predict the probability of the output classification and to route the coordinates.

Google developed the deep convolutional neural network named GoogleNet with 19 layers in 2014 to fulfill the function of classification and recognition, which is characterized by few parameters, easy training, good performance and high accuracy. The core architecture of the system derives from GoogLeNet, Since Shaoqing Ren put forward that more convolution layers and full-connect layers add to the promotion of performance, the core framework of the system is designed into 24 convolution layers, with two full-connect layers. The convolution layers of the Filter 1X1 are used alternately for dimension reduction operation among
the convolution layers of the Filter 3X3, instead of the Inception Framework in GoogLeNet. The core architecture is shown in Figure 1.

Figure 1. Deep convolutional neural network architecture

4 The Core Principle of Deep Convolutional Neural Network Design

For a computer image, it is a 3 dimensional array, and the array of numerical range is [0,255], which is shown in Figure 2.

Figure 2. Image Shown in Computer

Convolutional neural network algorithm could be effective feature extraction. The internal structure of the neural network based on the depth of the neural network is composed of neurons, which is shown in Figure 3. Deep convolution neural network internal structure is shown in Figure 4.

Figure 3. Neuron Structure

In the core of the neuron lies the activation function and convolution kernel. The essence of convolution is represented by vector dot product in linear algebra, and the schematic diagram of convolution operation is shown in Figure 5. In order to speed up the calculation, the 3 dimensional image data are transferred into a plurality of the vector in the actual encoding according to the convolution kernel size, and the weight set of the convolution kernel is calculated.

Figure 4. Internal Structure of Basic Deep Convolution Neural Network

Figure 5. Schematic Diagram of Convolution Operation

5 Optimized Deep Convolutional Neural Network

The convolution computing of image data are extremely expensive, for the deep convolutional neural network, it is growing by the number of magnitude, to result in the unfinished training in a short period of time, even to cause learning stop due to non-convergence. Therefore it is necessary to optimize the convolutional layer and the whole of the deep convolutional neural network in many ways. In order to reduce the amount of parameters and the computing cost, the local connection technique is introduced to replace the whole connection technique in Fukushima, which is shown in Figure 6.
For an image, the weight sharing technology is used in the condition of using the local connection technology, if the internal memory consumption can't be completely reduced. In the actual program code, only one filter is used to slide the set step size on the whole image and generate an active feature map, and as a result the internal memory consumption and computing cost are greatly reduced, the higher qualified extracted characteristics are obtained at the same time.

From the analysis of the whole network structure, Pooling Layer is added after the convolutional layer for better management of feature data. Pooling Layer is as in Figure 7. Pooling Layer is the subsample acquisition, including Max Pooling Layer, as shown in Figure 8.

In the training of the deep convolutional neural network, Dropout is introduced to solve the problem of over fitting. The core principle of Dropout technology is to make some neurons zero at random in order to achieve sparse connectivity when the feed forward is calculated. Figure 9 shows the network structure after Dropout is used.

6 Summary

Based on the deep convolutional neural network, the system architecture has realized the object detection, object localization and object classification, and it has achieved the higher speed of detection, and basically realized the real-time detection and classification recognition, with the help of NVIDIA and GPU of high performance.

Deep convolutional neural networks can extract the extremely complex characteristics, and enhance the depth and width of the network, using a variety of advanced optimization algorithm. With the help of advanced computing equipment, large-scale training can be carried out on such large scale database as ImageNet to obtain a higher accuracy of classification.

Feature extraction is conducted on deep convolutional neural network though the accuracy is low for one-time object detection, object localization and object classification, the speed is high. If the classification scale is reduced and applied in single professional field for higher accuracy of detection and localization, the usability of the deep convolutional neural network will be greatly improved.
Acknowledgements

This research was financially supported by the science foundation of Liaoning Province (L2014265).

References