

Research of Image Recognition Algorithm Based on Depth Learning

Zhang Jian, Ji Xinhao

Zhejiang Business College, Hangzhou, China, 310000

Keywords: Depth Learning, Image Recognition Algorithm, Deep Neural Network, Convolution Neural Network

Abstract: In recent years, deep learning has become more and more widely used in the fields of speech recognition, computer vision and bioinformatics. Especially in the field of image recognition, deep learning-based convolutional neural networks show a strong technical superiority that surpasses previous machine vision methods. The traditional methods of machine vision have not been able to meet the practicality and security requirements of image recognition in the big data era. At this stage, deep learning has become a hotspot in image recognition technology. To this end, the paper elaborates the background and basic ideas of deep learning, and analyzes common models of deep learning: deep belief networks, algorithm principles of convolutional neural networks, and image recognition efficiency.

1. Introduction

With the development of information technology, the amount of image information is getting larger and larger. Therefore, it is very important to extract the feature information we need from the huge amount of image information and identify the image accurately. year 2006. The deep belief network (DBN) proposed by Hinton et al. Has become a research hotspot in the field of image recognition because of its extraordinary ability of mathematical representation. Deep learning can build algorithm through deep neural network, so as to extract image features hierarchically and get the ability of automatic learning features through data training, which has greatly promoted the development of image recognition technology. In recent years, based on the depth of learning image recognition algorithm models continue to emerge. Mainly DBN (Deep Belief Network), CNN (Convolutional Neural Network), RNN (Recurrent Neural Networks). Practice has proved that convolution neural network is the most ideal structure in the current deep learning structure, convolution neural network in image recognition is widely used. Below, the text will explain the basic idea of deep learning.

2. Depth learning in the field of image recognition applications

Since 2006, when G.E. Hinton et al. Proposed deep learning, deep learning has shown superior performance advantages in many fields and has become a research hotspot in the field of Internet big data and artificial intelligence. In 2012, Hinton's research team used a deep learning model based on convolutional neural networks to win the ImageNet Image Classification Contest, reducing the error to 15.315%, compared with the lowest error rate of 26.172% for the traditional computer vision method. Many researchers have begun to study deep learning, especially for convolutional neural network model for a large number of studies. At the same time, deep learning has also brought tremendous business value. For example, some banks and enterprises started to provide online account opening service and brush face payment service, so that people can handle all kinds of business without leaving home, avoiding the long wait. Microsoft developed a deep neural network to replace the traditional GMM (Gaussian Mixture Model) algorithm, to speech recognition has brought a new technical framework, significantly reducing the error rate, and in 2012 officially launched the first deep neural network-based voice search system, becoming one of the first companies in the world to offer voice services. With the advent of the era of big data, major Internet companies in the world, such as Baidu and Google, are actively studying the application of

deep learning. For example, the use of deep neural networks to reduce the computational complexity of feature numbers now enhances the performance of search advertising systems. In addition, deep learning also has many applications in NLP (Natural Language Processing), IR (Information Retrieval) and visual scene annotation.

3. Research of Image Recognition Algorithm Based on Depth Learning

3.1 Reconstruction of MNIST Digital Images Based on Depth Neural Networks

MNIST, a sub-database of the National Institute of Standards and Technology's large data set, is a handwritten digital library containing 10,000 test samples and 60,000 training samples consisting of 0 to 9 digital samples with a resolution of $28 * 28$. These samples are not standard image formats but are stored in the idx file. The MNIST dataset basically does not require data preprocessing, it can be used immediately. As a result, many researchers have used MNIST as the preferred database for studying techniques and pattern recognition methods. Shown in Figure 1, the MNIST data sets have been converted to the image format part of the sample diagram.

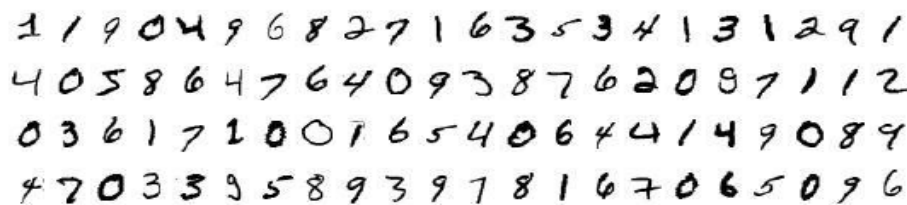


Figure 1. A partial sample diagram of the MNIST dataset that has been converted to image format

Reconstruction of samples in MNIST dataset with four - layer depth belief network using RBM. Firstly, the 300-dimensional effective features of the image are extracted, the parameters are adjusted, and the error of the initial image between the reconstructed image and the input network is reduced. This method can significantly reduce the dimensions of the image, compress the data efficiently, and reduce the space required to store the image.

3.2 MNIST Digital Image Recognition Based on Convolutional Neural Network

3.2.1 Transform layer

The image has a fixed feature, and part of the image statistics is the same as the other parts. In other words, the features learned in this part can also be the same as the other part. Therefore, we can use the same learning feature to handle the same image anywhere. Convolutional neural network is the use of this feature of the image to achieve weight sharing, thereby reducing the network parameters. We can regard the image as a plane, and based on preserving the two-dimensional characteristic of the image, we can process the input image by two kinds of transformations: linear transformation and non-linear transformation. The above-mentioned non-linear operation refers to the so-called incentive function, shown in Figure 2, where mainly three kinds of nonlinear.

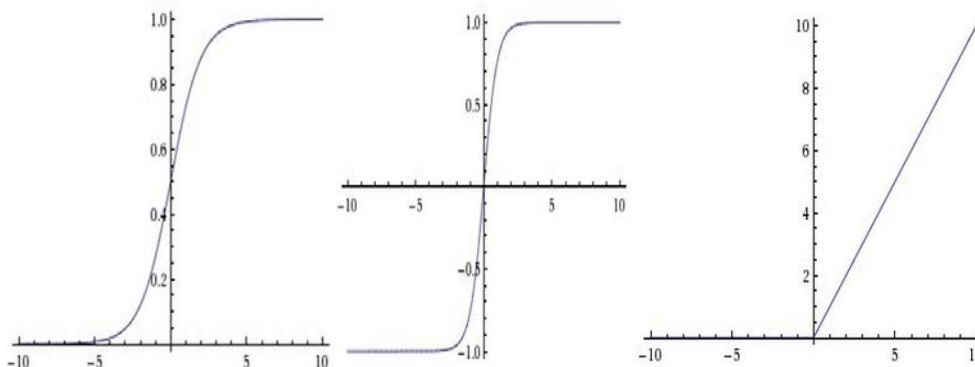


Figure 2. Nonlinear function

The left one is the sigmoid function. According to the image, we can see that the range of the input value is (0, 1), which is a common nonlinear transformation function of neural network. It can well explain nearly fully saturated neurons. At present, less used in the application, mainly because the neuronal activation value in the vicinity of 0 or 1, the gradient of the region is almost 0, when the back propagation, the first few layers of weight change is minimal, if the initial weight is too large, Neurons will soon be saturated, the network will stop learning. In addition, neurons handle data centers that are not zero, adversely affecting the dynamics of the gradient decline.

The middle is a hyperbolic tangent function, which limits the input value within the range of (-1, 1), but there is also a saturation problem. The first one is a non-linear correction function. Compared with the sigmoid function and the hyperbolic tangent function, the calculation of the number of nonlinear correction lines is easier and accelerates the convergence of stochastic gradient descent. Therefore, it has become very popular in recent years, but it also has the disadvantage that ReLU neurons are not activated at any one data point when large gradient values pass through ReLU neurons, so ReLU cells are more vulnerable.

3.2.2 Pooling layer

Convolution feature extraction dimension is too high, and there is redundancy, not suitable for classification, we must first reduce the dimension. You can aggregate features in different locations of the image. For example, the average or maximum value of a feature in a region of an image is calculated, so as to reduce the feature dimension and make the network difficult to overfit. This type of aggregation is called pooling. As shown in Figure 3, pooling uses the local correlation principle of the image to sample the feature map and reduces the amount of data to be processed while retaining useful structural information.

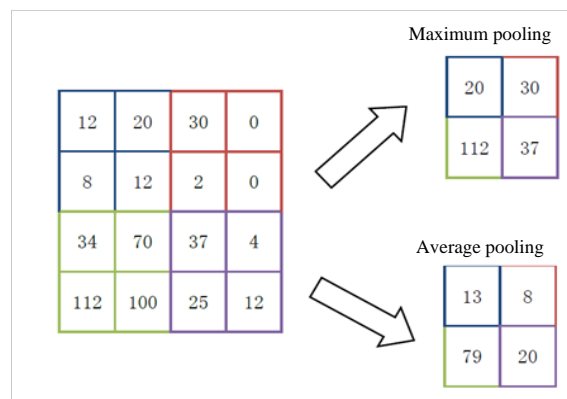


Figure 3. 2*2 two kinds of pool type diagram

4. Analysis of Algorithms

In this paper, CNN, a convolutional neural network with depth structure, is used to automatically learn image features and perform image recognition. By combining the feature extraction and classifier training in depth learning, the readiness of image recognition can be significantly improved. It should be pointed out that the traditional image recognition algorithms need to preprocess images to extract better image features, and then select the appropriate classifier for more features. The traditional image recognition algorithm with a lot of uncertainty, requires a wealth of experience, and vulnerable to human factors, resulting in accurate recognition is not high, lack of stability. In addition, the traditional image recognition algorithm also requires complex parameter adjustment, thus significantly increasing the training time. The traditional image recognition algorithms different neural network CNN is the convolution of the two-dimensional image can be directly input into the network, neural network recognizes visual patterns in the original picture, the required preprocessing little affected by human factors Small, breaking the limitations of human design features. CNN, as an end-to-end learning network, has an error rate of 0.84% on the recognition result and its recognition performance is comparable to or even better than

the traditional method, which proves the effectiveness of CNN method. The value of the gradient in the image recognition algorithm based on convolutional neural network comes from the number of input samples. Here we will discuss how to calculate the gradient.

$$\begin{cases} W_{ij}^1 = W_{ij}^1 - \alpha \frac{\partial}{\partial W_{ij}^1} J(w, b) \\ b_i^1 = b_i^1 - \alpha \frac{\partial}{\partial b_i^1} J(w, b) \end{cases} \quad \square$$

The most primitive method is based on the sample gradient, as shown in the above formula, if the sample is small, the program can run normally, but if the sample number is large, the running time will be very long, the computing speed is very busy, you need to consume A large number of computing resources, easily lead to insufficient hardware space resources. The random selection of a sample update parameter is called SGD (Stochastic Gradient Descend), SGD will lead to more serious cost penalty function oscillation, will result in greater data error. \square

$$\begin{cases} W_{ij}^1 = W_{ij}^1 - \alpha \frac{\partial}{\partial W_{ij}^1} J(w, b, x_i, y_i) \\ b_i^1 = b_i^1 - \alpha \frac{\partial}{\partial b_i^1} J(w, b, x_i, y_i) \end{cases} \quad \square$$

In this paper, we use the Mini-batch SGD to update the parameters. For the gradient of the random sample calculator, the parameters are updated as follows:

$$\begin{cases} W_{ij}^1 = W_{ij}^1 - \alpha \frac{1}{N} \sum_{i=1}^N \frac{\partial}{\partial W_{ij}^1} J(w, b, x_i, y_i) \\ b_i^1 = b_i^1 - \alpha \frac{1}{N} \sum_{i=1}^N \frac{\partial}{\partial b_i^1} J(w, b, x_i, y_i) \end{cases}$$

The algorithm steps are as follows: for $i = 1$: MaxIters (maximum number of iterations)

(a). SGD randomly samples; (b). Calculate the residuals; (c). Calculate the gradient using the BP algorithm; (d). Network Parameter Update: Update the network parameters with the gradient calculated by the BP algorithm. Softmax regression is an extension of logistic regression, which is mainly used to deal with the second-class classification problems, while the softmax regression can be used for multi-class classification tasks mutually exclusive. The training set label can take k values and output a k -dimensional vector for representing the probability values that the sample belongs to k categories.

5. Deep Learning in the field of image recognition

With the increasing demand for image recognition and the development of deep learning, the application of deep learning will become more and more widespread and more intelligent in the future field of image recognition. Now we look at the direction of the development of deep learning in the field of image recognition.

5.1 The model hierarchy is more and more, the model structure is more complicated

Depth learning needs to model the characteristics of the image layer by layer. If the network model does not have enough depth, it will lead to the exponential increase of the required computing unit, and the image recognition complexity is increased. By learning the multi-layer image features, the characteristics of the network model are getting more and more global, and the restored images are more realistic. For example, AlexNet acquired the ImageNet Image Recognition Contest champion in 2102 using five convolutional layers, three pool layers, and two fully

connected layers. The network model used by GoogLeNet to win the ILSVRC Championship in 2014 used 59 Convolutional layer, 16 pool layers and 2 fully connected layers. The network model used by Microsoft's ResNet in 2015 uses 152 convolutional layers. This shows that the depth and complexity of the model are increasing.

5.2 The scale of training data increased rapidly

The complexity of depth learning model is getting higher and higher, more and more features are needed, which requires larger training data to improve the classification or prediction accuracy in image recognition. At present, the training data of deep learning algorithms are usually in the scale of hundreds of thousands and millions. For large enterprises such as Google and Baidu, the training data scale has reached ten million or even hundreds of millions of levels. So far, ImageNet has collected more than 20,000 categories and a total of about 14.2 million images, but still cannot meet the rapidly increasing demand for training.

5.3 The accuracy of model recognition is getting higher and higher

With the continuous development of the deep learning model and continuous improvement, the accuracy and efficiency of image recognition have obviously improved. Early R-CNN models took 13 seconds to process an image on the GPU and 53 seconds to the CPU, yielding an accuracy of 53.7% in recognition. The YOLO model of 2016 achieves the recognition speed of 45FPS without losing the accuracy, and the model recognition efficiency and detection precision are greatly improved.

6. Conclusions

To sum up, with the continuous development of deep learning, the development of image recognition technology ushered in a leap-forward development. Deep learning has effectively promoted the development of image recognition technology and promoted the research and application of image recognition technology by leaps and bounds. At present, all major science and technology enterprises in the world are investing heavily in research on image recognition as the main issue. In the coming years, we can foresee that deep learning will enter a period of rapid development in theory, algorithms and applications, and will also have a profound impact on the field of image recognition and other related fields. This paper first describes the basic idea of deep learning, and analyzes the method of reconstructing MNIST dataset based on depth belief network in order to get the most effective characterization of the image set. By constructing a 5-layer convolution neural network to recognize MNIST images, the performance of each image recognition algorithm model based on depth learning is contrasted. The deeper the network level, the more accurate the recognition of the essential features of images.

References

- [1] Hinton G E, Osindero S, Teh Y W. A fast learning algorithm for deep belief nets [J]. Neural Computation, 2006.
- [2] LeCun, Y., Bottou, L., Bengio, Y. & Haffner, P. Gradient-based learning applied to document recognition. Proc. IEEE 86, 2278-2324 (1998)
- [3] Yann, LeCun, Yoshua, Bengio, &, Geoffrey, Hinton. Review of Deep Learning [J]. Nature, 2015, (521): 436-444
- [4] Krizhevsky, A., Sutskever, I. & Hinton, G. ImageNet classification with deep convolutional neural networks. In Proc. Advances in Neural Information Processing Systems 25 1090-1098 (2012)