WEP Warwick Evans Publishing

International Journal of Computer Science and Information Technology

ISSN: 3005-9682 (Print), ISSN: 3005-7140 (Online) | Volume 4, Number 2, Year 2024
DOI: https://doi.org/10.62051/ijcsit.v4n2.27
Journal homepage: https://wepub.org/index.php/IJCSIT/index

Performance Optimization of Convolutional Neural Networks in Image Classification

Haoyu Ma

School of Computer Science, University of Nottingham, Nottingham, UK

ABSTRACT

With the development of deep learning, CNNs have become mainstream in image classification. However, CNN performance is affected by factors like loss function choice and optimization algorithm. This paper aims to improve CNN performance in image classification by optimizing SGD. Firstly, we analyze different loss functions and propose a suitable optimization strategy. Second, we explore SGD and its variants' roles in CNN training, focusing on accelerating convergence and improving accuracy. Experiments show the proposed scheme effectively enhances CNN performance, providing a reliable solution for image classification. This study supports CNN applications in image recognition and lays a foundation for exploring more efficient optimization methods.

KEYWORDS

Convolutional neural networks (CNN); Image classification; Loss function; Stochastic gradient Descent (SGD); Deep learning, Optimization algorithms; Feature extraction

1. INTRODUCTION

With the rapid development of computer vision technology, image area differentiation is a crucial part of many real-world application scenarios, such as medical image analysis, unmanned vehicle control and security monitoring, which can be seen in its wide application. Today, the rise of deep learning, especially convolutional neural networks (CNNS), with its strong feature representation learning power and excellent classification efficiency, has consolidated its dominant position in image discrimination and recognition tasks. By simulating the operation mechanism of human brain, supplemented by the delicate design of local perception domain and weight sharing, this model can efficiently mine multi-level features from image data, so that the model has the ability to acquire and understand complex visual patterns.

Although convolutional neural networks have shown remarkable superior performance in image classification tasks, several challenges still emerge in the process of putting them into practice. (1) Activation function plays a crucial role in the architecture of convolutional neural networks. However, the network types applicable to different activation functions are not the same, and the advantages and disadvantages of various activation functions are not the same at the same time. For example, Sigmoid is very easy to show ladder degree dispersion, affecting the ladder degree decline. In the end, the deep network training could not be completed, and Tanh solved the problem of Sigmoid gradient descent, but it could be the problem of gradient dispersion according to the old existence. Although the unsaturated active function Relu is a good solution to the problem of hierarchical dispersion of the transmission activation function, it has a very fast convergence rate and good sparsity, but it still exists in the "dead" phenomenon. For this reason, the choice of the activation function in the convolutional network is a problem [1]. (2) The learning method of the convolutional Sutra network

also has a crucial impact on the training of the network. There are two learning methods, one is supervised learning method, that is, the input training sample is labeled, the mapping function is obtained through learning, and then the mapping function should be used in the new data sample. The mapping function can give the exact classification of these data samples, but the supervised learning method has ladder dispersion, overfitting phenomenon, and it is difficult to learn the optimal value and other missing points. In order to make the trained convolutional sutra network have better generalization ability, the supervised learning method requires too many samples, but the previous data sample is less. However, the unsupervised learning method hopes to obtain more deep secondary characteristics, and the samples trained in the process of unsupervised learning are not marked. However, the training time without supervision and supervision is long and the process is very tedious. Before the next level of training and learning, the parameters must be determined through the previous training and learning. The characteristics learned at the same time do not apply well to the physical problems that need to be solved [2]. These two learning methods to a certain extent determine the application of the network of God's Scriptures in practice.

2. CONVOLUTIONAL NEURAL NETWORK

Convolutional neural network is an efficient recognition method. Based on the study on the primary visual cortex of cats, Hubei and WieseI have found a unique neural network structure and proposed the concept of receptive field for the first time. Effectively reduce the complexity of the divine channel network. The primary visual cortex is composed of simple and complex cells. Hubei simulates simple cell and complex cell into a two-layer channel network, which is connected by the office parts. Hubei proposes the office part receptor field, which can extract primary visual characteristics from the input image. Such as edges, turning points, etc., and then combine these primary features to get higher features. Previously, the full connection was used between the two layers, while the convolutional channel network used the office connection between the adjacent layers, and the office connection can reduce the parameters trained by the network.

Weight sharing means that the elements sharing the same parameter detect the same feature structure in different positions of the input map. By composing the elements sharing the same weight into a two-dimensional plane map, a feature map (Feature Map) is obtained. The reason is that the characteristics of some parts of the image are the same as the characteristics of other parts, so the characteristics learned before can also be used in other parts of the image [3]. For example, if you input a large-size picture, you can choose a 5 x 5 in the image as a sample, and learn some features by training this sample. The features learned with the sample can be convolved with any part of the large-size image, so that different eigenvalues can be learned, and the weight sharing amplitude reduces the parameters required for training

The subsampling operation is to carry out horizontal and vertical square feature mapping in the continuous region of the characteristic graph with the step size smaller than that of the characteristic graph [4]. If you have a 12 x12 feature graph and subsample the 3 x 3 continuous region in the feature graph with step size of 3, you will get a sampled feature graph of (1 2/3) x (12/3) = 4 x 4. Through the subsampling operation, the complexity of the network mode can be effectively reduced, the resolution of the input image can be reduced, the number of elements of the divine channel can be reduced, and the data operation can be reduced, so that the transformation of the divine channel network to the input map can be convolved. The transfer has a high degree of invariance.

Different convolution kernels can be extracted by extracting different eigen structures. Suppose that the convolution operation with shared weights is extracted by 100 parameters as features, which is equivalent to having 10 x 10 convolution kernels. The extracted information can only be incomplete, so multiple convolution kernels can be added to make it learn more and more full features [5]. Convolution network is generally composed of convolution layer and pooling layer, convolution layer main feature extraction layer, through a series of column operations to extract the local feature of the

input image, pooling layer main feature mapping layer, The transfer function in the characteristic mapping layer is the activation function, which is the height invariance of the image with deflection, rotation and transformation after processing. The weights of a general scriptures are the same, and there is usually a fully connected layer of one or two layers at the end as the classifier. In order to enhance the performance of the convolutional channel network, it is normal to have a pool layer after the rolling layer.

3. NEURAL NETWORK OPTIMIZATION

3.1. Loss Function

Loss-loss function is the reflection of the fitting degree of the model trained through the convolutional channel network to the test data. The lower the fitting degree, the smaller the value of the loss-loss function, and the worse the fitting degree, the larger the value of the loss-loss function. When the value of the loss-loss function increases, the variable is updated faster, which requires its corresponding gradient to be large, because the minimum square error criterion (MSE) is usually used in the loss-loss function, see formula (1):

$$\min\big(y, G(x)\big) = \sum_i \big(G(x_i - y_i)\big)^2 \tag{1}$$

Where x represents the input image matrix, G represents the trained model, and $G_x$represents the output predictive vector. We can penetrate clearly to the point that the greater the Euclidean distance between yand G_x, the greater the value of the loss function.

3.2. Random Gradient Descent Method

Stochastic Gradient Descent (SGD) iteratively updates each sample, whereas batch descent iterates through each sample once, if there are a large number of samples, Stochastic gradient descent method can update the optimal solution more quickly and efficiently, because it uses a small number of samples to learn the text optimal solution. If the batch on-machine descent method is used, each sample needs to be traversed once. This will take a long time to increase the time complexity significantly.

3.3. Droput

In the convolutional channel network, droput can make the channel node of the hidden layer always come alive with a certain probability when the weight iteration is updated. The effective avoidance of the co-production of the two divine elements is not dependent on their co-operation. Since droput requires new parameters during the training and learning of each iteration, the structural model of the convolutional channel network is almost different for each input sample. At the same time, these structure models can be shared with weights. It is also said that droput is a kind of equilibrium model. droput can effectively avoid overfitting situations, and at the same time, in the field of image classification, it can reduce the classification error rate.

3.4. DropConnect

DropConnect is the result of improvements and optimizations to droput to make its anti-overfitting effectiveness even more noticeable. It can effectively reduce the classification error rate.

4. CONVOLUTIONAL NEURAL NETWORK OPTIMIZATION

The convolutional channel network has a certain advantage, but it also has poor performance of the time, one of the key elements is the choice of activation function. The Sigmoid activation function presents the phenomenon of gradient dispersion and non-zero output regardless of any input value. Tanh activation function is similar to Sigmoid and is also non-E-saturated activation function, but the difference is that Tanh can compress the input value to -1 to 1. The missing point that the output of Sigmoid function is not zero mean value is solved, it is basically zero mean value, the convergence speed is faster than Sigmoid, and the fault tolerance and performance ratio are better. But the Tanh function still exists in the problem of gradient dispersion. As shown in Table 1.

Relu is a non-linear unsaturated activation function, its mathematical form is expressed as:

$$f(x) = max(0, x) \tag{2}$$

The values obtained in the negative part are all 0, and the values obtained in the positive part are normal values.

However, the activation functions of Relu also have a certain number of missing points, and these missing points will also severely affect the convolution of the network training. In the network training process, Relu - not small heart will "die",

Softplus is a function similar to the ReIu function, which can carry out all non-linear mapping of the input data. It will not hide some valuable information, but its convergence rate is very slow, as shown in Table 2. The formula is as follows:

$$f(x) = ln(1 + e^{x}) \tag{3}$$

Relu-Softplus is an activation function that combines Relu and Softplus. The Softplus function is excited by the Softplus function in the part less than zero, and the Relu function is excited by the RELU function in the part greater than zero to shift the SOFTPLUS function curve down by $Ig_e 2$ units. As shown in Figure 1.

The Relu-Softplus active function formula is as follows:

$$f(x) = max\big(max(ln(1 + e^{x}), x)\big) \tag{4}$$

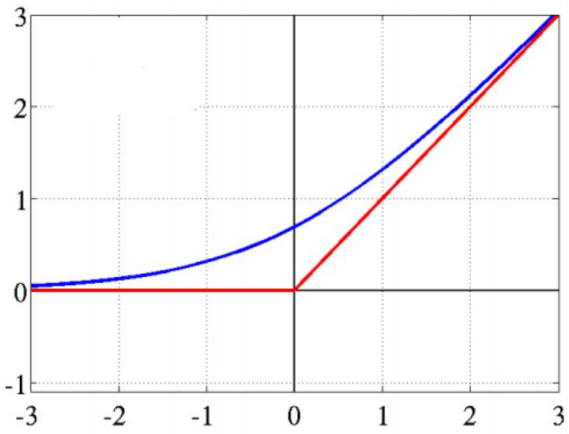

Figure 1. Relu-Softplus function curve

5. EXPERIMENT

5.1. Experimental Setup

Called the Cafe ConvoJIutionArchitectureForFeatureEmbedding. It is a deep learning algorithm framework with fast, clear network structure, high code efficiency and high readability. In Cafe, all data calculation is expressed in the form of Iayer. The purpose of Iayer is to obtain data, calculate and then input calculation results. The advantages of Cafe framework: fast hand, fast speed, openness, good community, modularity.

CUDA (Compu ter Uni fi ed Device Architecture) is a general parallel computing frame structure, which combines the advantages of COU and GPU. The use of CUDA allows the GPU to solve complex computing problems, greatly improving the performance of the computer. The use of CUDA-supported Gpus allows users to better learn Caffe, The ano and other deep learning frames. Therefore, CUDA has the advantages of simple and easy programming, small difficulty to open, and a direct interface can be conveniently and quickly deployed in the server.

5.2. Experimental Results and Analysis

Table 1. Experimental results

Activation function	Accuracy rate (%)
Sigmoid	53.24
Tanh	66.23
Relu	76.34
Softplus	77.23
Re 1u-Softplus	78.78

The experimental results are shown in Table 1. When Sigmoid is used as activation function, the accuracy rate of image classification is the lowest, which is 53.24%, and the convergence rate of Sigmoid is slow at the same time. When Tanh is used as the activation function, the accuracy of image classification is improved to 66.23%, and the convergence rate is obviously increased. When Relu is used as the activation function, the image classification accuracy is 76.34%, and the convergence rate is also very fast. When Softplus is used as activation function, the accuracy of image classification is high (77.23%), but its convergence rate is not fast enough. The accuracy rate of the improved activation function Relu-Softplus is the highest in the training process, reaching 78.78%.

Therefore, the improved activation function can not only improve the accuracy of image classification, but also accelerate its convergence speed. The simultaneous experiment also verifies the time required for different activation function training.

Table 2. Different from the activation function training time

Active function	Time (h)
Sigmoid	3.5
Tanh	3.2
Relu	2.6
Softplus	3.0
Relu-Softplus	2.3

It can be seen from Table 2 that the training time of the improved activation function RELU-SOFTPLUS is shorter than that of Relu.

6. CONCLUSION

This chapter first describes the use of activation functions in the convolutional channel network, and then describes several common activation functions, such as Sigmoid, Tanh, Relu and Softplus. At the same time, the advantages and disadvantages of each are analyzed respectively. Furthermore, it is proposed to solve the problem of stepwise dispersion of the activation function generation of the transmission, and too strong sparsity will mask some useful information gaps. Based on the advantages of Relu and Softplus activation functions, a convolutional channel network based on RELU-Softplus activation functions is proposed. By setting up the experimental environment and selecting the reliable and scientific verification set MNIST data set, the experimental verification is carried out and compared with other activation functions at the same time. The final verification results show that the improved activation function successfully takes into account the good convergence efficiency, and also confirms the improvement of image classification accuracy.

REFERENCES

[1] Sinha T, Verma B, Haidar A. Optimization of convolutional neural network parameters for image classification[C]//2017 IEEE symposium series on computational intelligence (SSCI). IEEE, 2017: 1-7.

[2] Aamir M, Rahman Z, Abro W A, et al. An optimized architecture of image classification using convolutional neural network [J]. International Journal of Image, Graphics and Signal Processing, 2019, 11(10): 30.

[3] Guo T, Dong J, Li H, et al. Simple convolutional neural network on image classification[C]//2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA). IEEE, 2017: 721-724.

[4] Ciresan D C, Meier U, Masci J, et al. Flexible, high performance convolutional neural networks for image classification[C]//Twenty-second international joint conference on artificial intelligence. 2011.

[5] Sun Y, Xue B, Zhang M, et al. Evolving deep convolutional neural networks for image classification [J]. IEEE Transactions on Evolutionary Computation, 2019, 24(2): 394-407.