Improved Algorithm Based on The Deep Integration of Googlenet and Residual Neural Network

Huang xuehua, Chen weihong, Yang wanghai*

*School of information and Electronic Engineering, Hunan City University, yiyang city 43100, china

*107531852@qq.com

Abstract: In this paper, we propose a new improved algorithm based on the deep integration of GoogleNet and Residual neural network and we call it GRSN. The new improved algorithm has the new advantages of multi-size and small convolution kernel in the same layer in the network and the advantage of interlayer hop connection (bypass) to reduce information loss. The algorithm is applied to the general image data set cifair10, and the experimental results are compared with that of GoogleNet, the accuracy is improved, the convergence is accelerated, and the stability is better.

Keywords: deep neural network, Google inception, residual network, ResNet, GRSN, recognition

1. Introduction

Deep learning is a research hotspot at present, and its research results have been widely used in image recognition, speech recognition, and other fields. We mainly study the application of deep neural network algorithms in image classification in this paper. We propose an improved neural network structure based on the Google net and residual neural network. Firstly, we introduce the mainstream algorithms of deep neural networks in image recognition in this paper and we analyze the advantages and disadvantages of the GoogleNet series algorithm and residual neural network algorithm. Based on the advantages of the two kinds of algorithms, an improved algorithm GRSN is proposed by introducing the hop line (bypass) into the GoogleNet inception algorithm. Secondly, we show the detailed network structure of GRSN. In addition, the experimental analysis is carried out. Then, the overfitting situation is optimized. Finally, the conclusion and prospects are given.

2. Related work
In 1986, the BP neural network was first proposed by Rumelhart and McClelland and applied to the classification of complex patterns. As the fully connected network structure is adopted in the BP neural network, the performance and accuracy of image recognition are not ideal. According to the fact that the human eye is more sensitive to the main features of objects, a CNN deep convolution neural network was proposed, which is a feed-forward neural network with depth structure.

Convolution is used to retain the important features and filter out the unimportant features on CNN. Pooling (downsampling) is adopted to compress code. The structure of the fully connected network is modified and the structure of shared weight is adopted, which greatly reduces the weight parameters and improves the performance of image recognition; at the same time, it also greatly improves the recognition accuracy. CNN convolution neural network is widely used in image classification by obtaining image features. LeNET[1] is a typical convolutional neural network structure, which was proposed by Lecun in 1998. Lent is composed of two pairs of convolution pooling and two layers of full connection. And it was applied to American handwritten digit recognition. The test error rate is less than 1%. But its recognition rate of Chinese characters and more complex images is not good. Therefore, in the 2012 international computer vision competition, the ALEXNET [2] was proposed by Hinton and his student Alex Krichevsky and they won the championship. The number of layers is increased in the algorithm, and it is run on the GPU with parallel structure, relu nonlinear activation function is adopted and dropout is used to prevent overfitting, thus its image recognition performance has been improved rapidly. Convolution groups and small convolution kernels are used in VGG[3]. The number of layers and channels is increased. The recognition performance and accuracy have been further improved. It won second place in the 2014 computer vision competition.

The winner of the 2014 computer vision competition was GoogleNet[4] algorithm and inception block was introduced, which used convolution kernels of different sizes in the same layer of the network to extract features of different sizes, and it improves the perception of the model and the accuracy of image recognition was improved. Moreover, on the premise of the same performance, small convolution is used to reduce the weight parameters, the amount of calculation, the cost of calculation, and the network complexity. After the 1 * 1 convolution kernel is introduced, the depth of the output feature map will be reduced and the dimension will be reduced. There is V1, V2, V3, and V4 version in GoogleNet. It is mainly reflected in the difference in the inception module. However, the width of GoogleNet is wide, the depth is not enough, and the efficiency of parameter operation is not high.

The recognition accuracy can be improved by deepening the number of network layers in a deep neural network. There are 5 layers for LeNet, 8 layers for AlexNet, 16 layers and 19 layers for VGG, and 22 layers for inception v1. But, He Kaiming, the author of RESNET [5], did an experiment on the cifair10 dataset and found that the error rate of image recognition of 56 layers convolution network is higher than that of 20 layers convolution network. Therefore, simply stacking the network layers will degrade the neural network model; the back features will lose the original appearance of the previous features. Thus he used a jump line to connect the front feature directly to the back. As shown in Fig. 1, the output result H (x) is composed of the upper stacked convolution output F(x) and the output of the direct upper layer X. It effectively alleviates the model degradation caused by the increase of the
number of layers of the neural network model, and makes the neural network develop to a deeper level.

Therefore, the RESNET algorithm, the representative of the residual neural network algorithm, won the champion of computer vision competition in 2015, and its top 5 error rate was 3.57%. The residual jump (bypass) is introduced to reduce the information loss and alleviate the loss of gradient, which makes it possible to increase the number of layers and improve the accuracy of image recognition. In addition to the RESNET algorithm, the residual neural network algorithm also includes Densenet and Darknet, the main difference between both is the hop line and residual block. A hop line is added between every two blocks in the deep residual network RESNET, and a hop line is added to every block in the deep residual network Densenet. The accuracy of Densenet has been further improved. Although the number of layers of the residual neural network is deeper, the width is narrow, and the multiscale feature extraction is worse than the inception block.

![Fig. 1 core composition of resnet algorithm](image)

3. An improved algorithm GRSN is proposed

3.1 GRSN theory
Combined with the advantages and disadvantages of GoogleNet and residual neural network, an improved network structure algorithm based on GoogleNet and residual neural network is proposed. As the advantage of GoogleNet is a small convolution kernel and the information loss of deep neural network occurs with the increase of layers. Therefore, we propose an algorithm GRSN that is with the improved structure of the deep neural network, that is, hop lines are added on the basis of the inception block, which not only retains the advantages of the GoogleNet algorithm but also has the advantages of residual neural network algorithm.

As the convolution kernel of different sizes is used in the same layer to extract features of different sizes, the perception of the model is improved, and the accuracy of image recognition is improved. On the premise of the same performance, the weight parameters are reduced as small convolution is used, the amount of calculation is reduced, and the complexity of the network is reduced. At the same time,
the residual Hop (bypass) between layers is added to reduce the information loss and it is possible to increase the number of layers. Thus, the accuracy of image recognition is further improved.

Compared with GoogleNet, the improved algorithm GRSN has better performance and accuracy than GoogleNet.

3.2 Structure of improved algorithm GRSN

The core structure of the deep neural network RESNET algorithm is shown in Fig. 1. In the improved algorithm GRSN, the residual block is replaced by the inception module, which is shown in Fig. 2.

![Diagram](image)

**Fig. 2** the core module of gsn

In other words, the hop line is added to the inception block, and the information of the upper layer is added to the inception block through the hop line. What needs to be discussed is that the hop line comes from and goes to. In fact, there are $1 \times 1$ convolution kernels in each inception block, as shown in Fig. 3, which is equivalent to the input information of the previous block is added to each inception block, so as to reduce the loss.

In this paper, we only study the simplest model, that is, jump lines from the input layer are added into the inception block of each layer. The model is shown in Fig 4. X represents the input of the image, which can also be the middle layer. Here we only consider the simple case. At the same time, the research on the residual block and hop line will be expanded in the following article. Since the inception block of GoogleNet has V1, V2, V3, V4 versions, and our improved algorithm is carried out on inception V1. The name of the improved algorithm is google-v1-resnet, abbreviated as GRSN.
Fig. 3 the core module of google net inception v1

Since the size of the input image may change after through the inception block, the size of the input image may not match with the output image when the image is superimposed between input and output. When the size of the image between input and output is not consistent, the image size can be adjusted by using the $1 \times 1$ convolution kernel.

Fig. 4 structure 2 of the improved algorithm grsn
3.3 Parameter setting of network structure
Standardization operation (BN) can accelerate data convergence and alleviate overfitting, which is generally located after the convolution layer and before the activation layer. Dropout leakage operation is mainly to alleviate the overfitting phenomenon. During the neural network training, some neurons are discarded from the neural network according to a certain probability. When the neural network is used, the abandoned neurons resume the connection. Pooling operation is used to reduce the amount of feature data in a convolution neural network.

4. Experimental analysis

4.1 Experimental data
We use the typical general data set cifar10 for the experiments comparison between GRSN and GoogleNet.

The cifar10 dataset contains 60000 images and tags, of which 50000 are for training and 10000 are for testing. The data set contains 10 categories of dogs, cats, birds, airplanes, frogs, horses, boats, and so on. Each category has 6000 pictures.

4.2 The process of experimental
In the process of experimental, we use the data enhancement of Keras to handle the data slightly. In this experiment, the accuracy of the cifair10 data set is tested on one inception block, two, three, four, five, and six inception blocks of GoogleNet v1, and the accuracy rate are on the rise. At the same time, the accuracy of the improved algorithm GRSN is also increasing with the increase of the number of blocks.

We use six blocks for comparison between the improved GRSN algorithm and the GoogleNet inception v1. And in the GRSN algorithm, the dropout parameter is set to 0.5 for the last layer and 0.05 for other layers.

4.3 The result analysis
After 5000 rounds of training on the cifar10 dataset, the accuracy curve is shown in Fig 5.
**Fig. 5** accuracy curve of grsn algorithm on cifar10 with data enhancement

After the data enhancement is removed, the accuracy rate is further improved. The curve is shown in Fig 6, and the maximum can be increased to 87.24%

![Training and validation accuracy](image1)

**Fig. 6** accuracy curve of grsn algorithm on cifar10 without data enhancement

However, the accuracy is achieved by nearly 85.57% by using the GoogleNet V1 algorithm with the six inception blocks. The improved algorithm GRMN has higher accuracy, faster convergence, and without an increase in training parameters.

5. **Conclusion**

After a comprehensive study of CNN, LeNet, AlexNet, VGG, GoogleNet, ResNet, DenseNet algorithm, we propose the GRMN algorithm. Compared with the GoogleNet V1 algorithm on the cifair10 data set, the accuracy is improved, the convergence is faster and the performance is more stable. However, where the hop line is from and to, and the parameter setting of the hop line is the future research directions.

**References**

[1] Krizhevsky A, Sutskever I and Hinton G 2012 ImageNet classification with deep convolutional Neural Networks NIPS( New York:Curran Associates Inc.) 25

[2] Krizhevsky A, Sutskever I and Hinton G E 2017 Imagenet classification with deep convolutional neural networks Communications of the ACM 60 pp 6 84

[3] Simonyan K and Zisserman A 2014 Very deep convolutional networks for large-scale-image recognition Computer ence 14091556

[4] Szegedy C, Ioffe S, Vanhoucke V and Alemi A 2016 Inception-v4, inception-resnet and the impact of residual connections on learning
[5] He K, Zhang X, Ren S and Sun J 2016 Deep residual learning for image recognition Proc. of 2016 IEEE Conference on Computer Vision and Pattern Recognition