Inception-Alexnet for Peony Recognition

Xi Deng [1], Jingrong He [1,*], Xiaoping Yi [2], Huiling Tian [3]

[1]College of Information Engineering, Northwest A & F University, Yangling, Shaanxi 712100, China

[2]Yangchun Junior High School of Nan Zheng District, Hanzhong, Shaanxi, 723100, China

[3]HuaYan School of Nan Zheng District, Hanzhong, Shaanxi, 723100, China

[a]dd@nwsuaf.edu.cn

Abstract

In order to more accurately identify the types of peony, a convolutional neural-network approach combined with Inception Module for peony recognition is proposed. It has a simpler structure, requires less training parameters, and makes network depth expansion easier. The complete image is used as the input and output of the network. The Inception Module is used to extract the original peony image and multiple features of different spatial scales, and various strategies are used to enhance the overall learning ability of the network. In order to prevent the gradient from disappearing, the rectified linear unit (ReLU) activation function is adopted; in order to speed up the training speed of the network, the batch normalization (referred to as BN) operation is added; to improve the recognition performance of the network, the jump structure is added for residual learning (abbreviated as RL).The verification was performed on the peony dataset, and the random gradient descent optimization algorithm was used to achieve an average accuracy of 94.93%.The experimental results show that compared with other image recognition methods, the model can reduce the storage space and improve the convergence speed, and the recognition effect is better. It has better robustness in image classification and is easier to expand in other application areas.

Keywords

Inception Module; convolutional neural network; peony recognition; rectified linear unit; batch normalization; residual learning.

1. Introduction

As the national flower of China, peony is not only the favorite of the literati, but also an important part of scientific research, which has great value. However, there are many kinds of peony, and it is very difficult for most people to accurately identify their types [1]. In general, users can perform category comparison by entering a name on the Internet, but it is very cumbersome and time consuming, and the result depends on the accuracy of the input text. If the user can quickly and accurately identify the peony type on the mobile side anytime and anywhere, it will be very convenient and practical.

Traditional machine learning requires a large amount of manual participation in feature design and extraction. It requires a huge amount of work and requires high image quality, which has poor robustness [2]. In recent years, compared with traditional machine learning, deep learning based on convolutional neural networks has developed rapidly, avoiding the complexity and blindness of large-scale manual extraction of features in traditional machine learning [3]. The emergence and development of classic LeNet [4], AlexNet [5], GoogleNet [6], VGG [7], ResNet [8] and other network models have prompted deep learning to achieve breakthroughs in image recognition and classification.

Song Xiaoru et al. [9] used LeNet for handwritten digit recognition. By improving the cost function and adding dropout to prevent over-fitting to improve recognition accuracy, it is less effective for other complex data sets such as flower data sets. Shen Ping et al. [10] proposed the recognition of flower species based on AlexNet. Compared with traditional neural networks and support vector machines, the accuracy rate is improved by more than 10%, but the accuracy is still not ideal, and the training time is long, the model occupies a large storage space and it is easy to overfit. Kong Yinghui et al. [11] applied the MobileNets model to flower recognition in complex background, and the recognition rate is greatly improved and the model occupies a small storage space, which can well meet the application requirements of mobile terminals, but the recognition rate is still not up to 90%.

Aiming at the above problems, this paper proposes a effective approach. The contributions of this work are summarized as follows:

1) We propose a convolutional neural-network based on Alexnet and replace its convolution-layers with inception module (that is Inception-Alexnet).

2) Compared with both classical Alexnet model and general Inception model, the proposed method achieves the best recognition effect, that is, the recognition accuracy is the highest and the convergence speed is the fastest.

3) We collected and labelled a peony dataset. The Peony dataset includes 12 species of peony, each with 1200 photos and a total of 14,400 photos.

2. Inception module structure

The Inception module was originally introduced by GoogLeNet, which is called Inception-v1. In the Imagenet Large Scale Visual Recognition Challenge 2014 (ILSVRC2014), 22-layer GoogleNet won the championship with a Top5 error rate of 6.67% on the ImageNet dataset. Prior to [6], the performance improvement of convolutional neural networks relied on increasing the depth and breadth of the network. The literature [6] started from the network structure, changed the network structure, and first proposed the inception convolution network module. In order to improve the utilization of network computing resources and increase the width and depth of the network in the case of constant calculation, the author believes that the full connection should be changed to a sparse connection, and the convolutional layer is also sparsely connected. But the asymmetric sparse data is computationally inefficient, because the hardware is optimized for dense matrices, so to find the convolutional network can be approximated. So we need to find the optimal local sparse structure that can be approximated by the convolutional network, and the structure can be implemented with the existing density matrix computing hardware, resulting in Inception. The advantage of the Inception module is the frequent use of 1×1 convolution kernels and multi-level feature transmission. On the one hand, the author uses 1*1 convolution kernel not only to reduce the dimension and reduce the computational bottleneck, but also to increase the number of network layers and improve the expression ability of the network. On the other hand, the Inception module contains subnets of different depths. These sub-networks provide different levels of features. Deeper sub-networks greatly increase the depth of the network, while shallower sub-networks allow features to reach the next Inception structure faster, alleviating the gradient disappearance and gradient explosion phenomenon due to the increase of network depth and better improving the generalization performance for scale of the network.

It can be seen that the advantage of Inception is that the number of cells in each step is significantly increased. The computational complexity is not unrestricted, the dimension is reduced before the convolution of the larger block, and the visual information is processed and aggregated on different scales, so that features can be extracted from different scales in the next step. Inception's role is to replace the manual determination of the type of filter in the convolutional layer or whether to create a convolutional layer and a pooling layer, so that the network can learn what parameters it needs. The basic structure of Inception is shown in Figure 1.

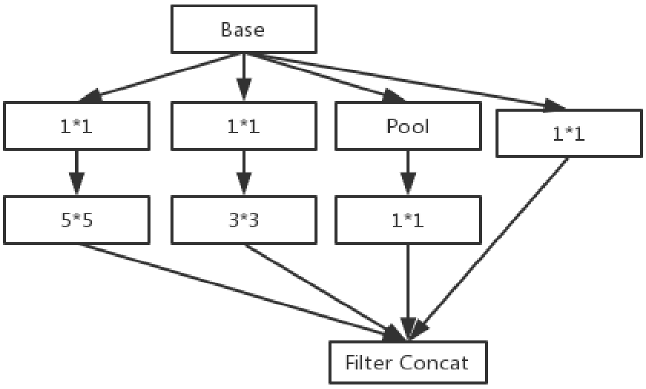

Figure 2.The kind of Inception module

Literature [13] proposed Inception-v2, adding the Batch Normalization (BN) layer, which reduced the error rate of the model to 4.9% on the ImageNet dataset. The main functions of BN not only speed up network training, but also prevent gradient disappearance. If the activation function is sigmoid, for each neuron, the input distribution that gradually moves closer to the saturated region of the nonlinear mapping can be forcibly pulled back to the standard normal distribution of 0 mean unit variance, that is, the excitation region of the activation function. It has a large gradient in the sigmoid excitation region, that is, accelerates the network training, and also prevents the gradient from disappearing. Based on this, BN has a large effect on the sigmoid function.

Literature [14] proposed Inception v3, which introduced decomposition. The n×n convolution kernel is split into two convolution kernels, n×1 and 1×n, which not only saves a lot of parameters, but also improves the expression ability of the model. At the same time, it can process more abundant spatial features and increase diversity of features. The model's Top5 error rate on the ImageNet dataset is reduced to 3.5%.

The literature [15] draws on the ideas in the Microsoft ResNet paper, combining the Inception module with the ResNet residual network structure. The Top5 error rate on the ImageNet data set is reduced to 3.08%. The residual network is the stack of residual blocks, which can make the network design very deep. With the increase of network depth, the training error of Resnet will always decrease, which greatly improves the speed of network training and accelerates the convergence of the network.

This article simplifies and improves the Inception module while preserving the advantages of Inception, making network depth development easier and improving network performance. Combined with the Inception module, this paper proposes a new convolutional network structure, which improves the recognition performance of the network and avoids the situation that the parameter quantity of the network increases greatly with the number of categories. In order to prevent the gradient from disappearing, the rectified linear unit (ReLU) activation function is adopted; in order to speed up the training speed of the network, the batch normalization (referred to as BN) operation is added; to improve the recognition performance of the network, the jump structure is added for residual learning (abbreviated as RL).

3. Convolutional neural network structure combined with inception structure

The overall structure of the 8-layer image recognition model of the convolutional neural network combined with the Inception module is shown in Figure 2, Table 1.

The specific structure of the 8-layer image recognition model of the convolutional neural network combined with the Inception(that is Inception-Alexnet) is as follows.

The size of input image X_0 is set as 227*227*3.

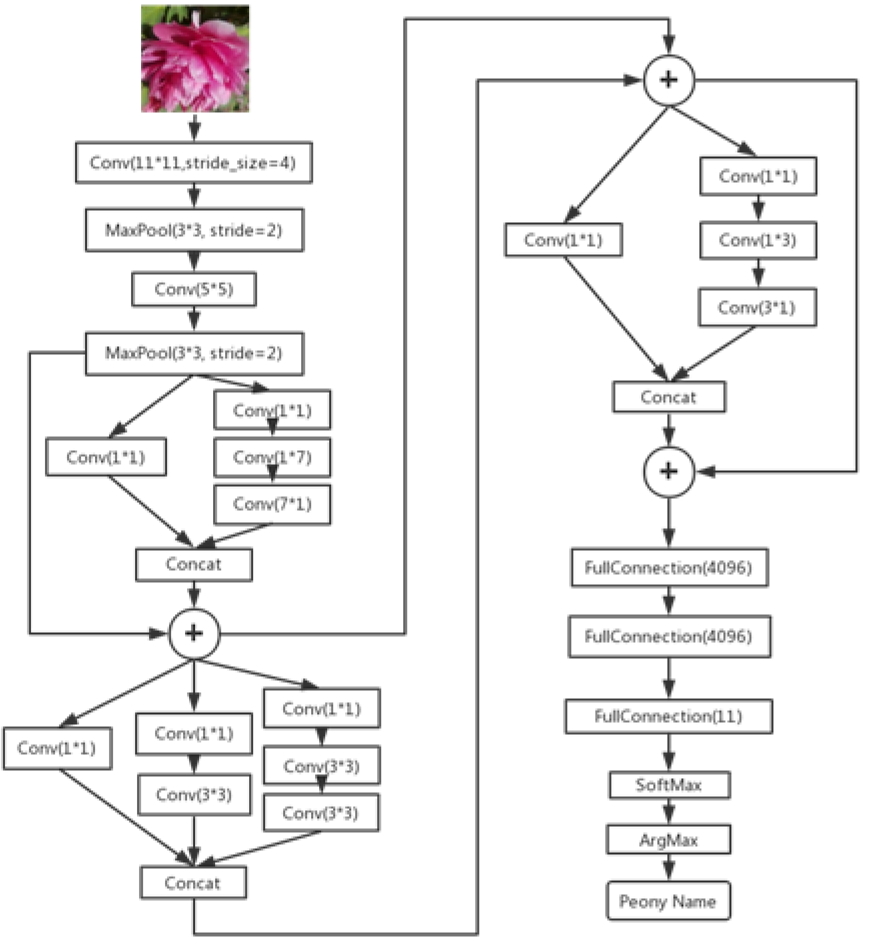

Figure 2. The overall structure of Inception-Alexnet

(1) The first layer is Conv+BN+Relu+MaxPool layer, convolution kernel W_1: the size is 11*11*96, the stride is 4 , the output of the first layer is expressed as X_1=max(0, W_1 * X_0 + b_1)。The large-scale convolution kernel of the first layer extracts the large-scale features of the image, and obtains the feature activation value at each position of the image by convolution operation, and then obtains the activation value through the Relu activation function. The activation value passes through the max pooling layer with a size of 3*3 and a stride of 2. The size of output X_1 is 27*27*96。

(2) The second layer is Conv+BN+Relu+MaxPool layer, convolution kernel W_2: the size is 5*5*256, the output of the second layer is expressed as X_2=max(0, W_2 * X_1 + b_2). Further extract image features and obtains the activation value through the Relu activation function. The activation value passes through the max pooling layer with a size of 3*3 and a stride of 2. The size of output X_2 is 13*13*384.

(3) The third layer is based on the Inception module.

Branch 1: 1*1*192 Convolution kernel W_{31} extract local correlation:

$$X_{31} = W_{31} * X_2 + b_{31}$$

Branch 2

1*1*128 Convolution kernel W_{321}; 1*7*160 Convolution kernel W_{322};

7*1*192 Convolution kernel W_{323}.

The third layer is expressed as

$$X_3 = \frac{X_{31} + X_{32}}{2}$$

(4) The fourth layer is based on the Inception module.

Branch 1: 1*1*32Convolution kernel W_{41}

$$X_{41} = W_{41} * X_3 + b_{41}$$

Branch 2

1*1*32 Convolution kernel W_{421}; 3*3*32 Convolution kernel W_{422}.

Branch 3

1*1*32 Convolution kernel W_{431}; 3*3*48 Convolution kernel W_{432};

3*3*64 Convolution kernel W_{433}.

The fourth layer is expressed as

$$X_4 = \frac{X_{41} + X_{42} + X_{43}}{3}$$

(5) The fifth layer is based on the Inception module.

Branch 1: 1*1*192 Convolution kernel W_{51}

$$X_{51} = W_{51} * X_4 + b_{51}$$

Branch 2

1*3*224 Convolution kernel W_{421}; 3*1*256 Convolution kernel W_{422}.

The fifth layer is expressed as

$$X_5 = \frac{X_{41} + X_{42}}{2}$$

(6) Through a MaxPool layer with a size of 3*3 and a stride of 2, the output size is 6*6*256. Flatten the output to a vector of length 9216.The sixth layer is a fully connected layer with an output size of 4096, and the log l2 norm multiplied by 0.001 is included in the loss function.

(7) The seventh layer is a fully connected layer with an output size of 4096, and the log l2 norm multiplied by 0.001 is included in the loss function.

(8) The eighth layer is a fully connected layer with an output size of 11, and the log l2 norm multiplied by 0.001 is included in the loss function.

Finally, the vector with output 11 is obtained by the softmax layer, which is the predicted value of each bit.

Table 1. The layers of Inception-Alexnet

algorithm	Convolution kernel size	Input channel	Output channel	Pooling layer
Conv1	11*11 (stride 4)	3	96	
Pool1	3*3 (stride 2)	96	96	MaxPool
Conv2	5*5	96	256	
Pool2	3*3 (stride 2)	256	256	MaxPool
Conv3	1*1 or 1*1, 1*7, 7*1	256	192	
Conv4	1*1 or 1*1, 3*3 or 1*1, 3*3, 3*3	192	32	
Conv5	1*1 or 1*1, 1*3, 3*1	32	192	
Fc6		9216	4096	
Fc7		4096	4096	
Fc8		4096	11	

4. Experiment

4.1 Dataset

In order to verify the validity of the convolutional neural network structure model combined with Inception structure in the identification of peony species, this paper selected the peony dataset that we shot. The Peony dataset includes 12 species of peony, each with 1200 photos and a total of 14,400 photos. The photos are taken by Samsung S7 (resolution 2560*1440), Huawei Honor 7 (resolution

5152*3888), Huawei nova2s (resolution 4608*3456), iPhone 6 (resolution 1334*750). four The four kinds of equipment are used in the cloudy and sunny weather and they are shot at different times in the morning, afternoon and evening. The dataset belongs to the large-scale pattern recognition sample set. Reduce all the pictures in the dataset to 256 × 256 × 3 for recognition experiments.

4.2 Experimental

The system used in this paper is Ubuntu16.04. Under the system, the deep learning framework Tensorflow is built for training, and the experimental test is carried out on Pycharm. The computer hardware is configured as Intel(R) Xeon(R) CPU E5-2620 v3 @ 2.40GHz. The size of RAM is 64 GB and the kind of GPU is NVIDIA Corporation GM107GL [Quadro K2200].

4.3 Experimental strategy

We trained our network multiple times on the training set using different parameters, and tested it with the test set. The parameters and training strategy of the network with the highest accuracy and the fastest convergence are shown below.

Training for 15 epochs

The batch size is 32

Learning rate adjustment strategy: initial learning rate lr = 0.0001, adjusted to 0.00001 after 10 epochs

Optimizer: adam, optimizer beta1=0.9, beta2=0.999, epsilon=1e-08

The data is randomly divided into 10, of which 7 are used as training, 2 are used as test, and 1 is used as verification set (eval)

The data was enhanced during training, including random cropping to 227*227 size, random horizontal flip, random flip up and down, random rotation, random brightness, random contrast, random hue, random noise, random jpeg quality, normalization.

4.4 Experimental results and analysis

We visualize the output structure of the first layer under tensorflow and find that the network can extract the features of the input image very well. The effect of feature extraction is shown in Fig.3.

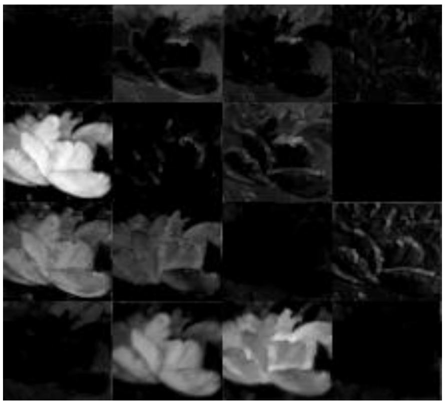

Figure 3The effect of feature extraction

In order to verify the characteristics of the algorithm compared with other algorithms, the experiment compares the recognition effect of the proposed algorithm with Inception-Alexnet using sgd optimizer and Alexnet using adam optimizer. The recognition effect is shown in Figure 4. It can be seen from the figure that the Inception-Alexnet network of the adam optimizer is significantly better than other networks.

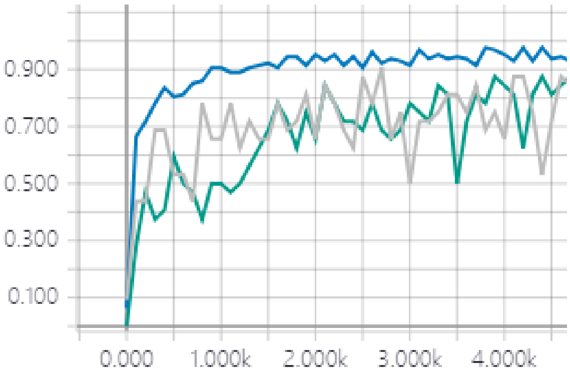

Figure 4.Curves of accuracy in the test dataset

As the Figure 4 shows: The blue curve is Inception-Alexnet using the adam optimizer; The grey curve is Inception-Alexnet using the sgd optimizer; Green curve for Alexnet using adam optimizer

Table 2. Accuracy in the eval dataset

network	Top1-Accuracy
Inception-Alexnet (adam)	94.93%
Inception-Alexnet (adam)	90.63%
Alexnet	87.50%

As can be seen from Table 2, the identification method of this paper shows the best recognition effect. Compared with the Inception-Alexnet using the sgd optimizer, the accuracy of our proposed recognition method in the Peony test dataset exceeds 4.30% on average. Compared with the classic Alexnet in the deep neural network identification method, our network also better completes the image recognition task.

5. Conclusion

This paper proposes an 8-layer image recognition model of convolutional neural network combined with Inception module, combining BN and RL, which increase the depth of the network and reduce the parameters of the network. This not only improves the recognition ability of the model, but also improves the speed of network convergence and improves the robustness and generalization ability of the network. The experimental results show that compared with both classical Alexnet model and general Inception model, the proposed method achieves the best recognition effect, that is, the recognition accuracy is the highest and the convergence speed is the fastest. In the experiment, the successful application in the fine identification of peony species shows that the model has certain practical value.

Acknowledgments

This paper is the research result of the Teaching research and reform project of Northwest A&F University: Teaching Reform and Practice of Program Design Foundation (VB) based on CDIO idea(No.JY1504049), and also partly result of the MOOC Project of Northwest A&F University: Fundamentals of College Computer.

References

[1] M. -. Nilsback, A. Zisserman. A Visual Vocabulary for Flower Classification, 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), New York, NY, USA, 2006, 1447-1454.

[2] Wu Kaili et al. Recognition of marine vessels under complex weather conditions based on deep learning, Science Technology and Engineering, vol. 19(2019), 130-135.

[3] Fu Tianju, Zheng Chang'e, Tian Ye, Qiu Qimin. Lin Sijun. Forest Fire Recognition Based on Deep Convolutional Neural Network Under Complex Background, Computer and Modernization, 2016, 52-57.

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

[5] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural network, Neural Information Processing Systems, 2012, 1097-1105.

[6] C. Szegedy et al. Going deeper with convolutions, 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, 2015, 1-9.

[7] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition, 2014, arXiv preprint arXiv:1409.1556.

[8] K. He, X. Zhang, S. Ren, J. Sun. Deep Residual Learning for Image Recognition, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, 770-778.

[9] Song Xiaoru, Wu Xue, Gao Song, Chen Chaobo. Simulation Study on Handwritten Numeral Recognition Based on Deep Neural Network, Science Technology and Engineering, vol.19(2019), 193-196.

[10] Shen Ping, Zhao Bei. Automatic Classification of Flowers Based on Deep Learning Model, Bulletin of Science and Technology, vol.33(2017), 115-119.

[11] Kong Yinghui, Zhu Chengcheng, Che Linlin. Flower Recognition in Complex Background and Model Pruning Based on MobileNets, Science Technology and Engineering, vol. 18(2018), 84-88.

[12] Wu Xiaoxin, Gao Liang, Yan Min, Zhao Fang. Flower species recognition based on fusion of multiple features, Journal of Beijing Forestry University, vol. 39(2017), 86-93.

[13] offe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift, International Conference on Machine Learning, 2015, 448-456.

[14] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna. Rethinking the Inception Architecture for Computer Vision, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, 2818-2826.

[15] zegedy C, Ioffe S, Vanhoucke V, et al. Inception-v4, Inception-ResNet and the impact of residual connections on learning, arXiv: 1602.07261, 2016.