Research on leaf image identification based on improved AlexNet neural network

Wenkun Zhang\(^1\) and Juanjuan Wen\(^2\)*

\(^1\)School of Mathematics, Taiyuan University of Technology, Jinzhong, Shanxi, 030600, China
\(^2\)Department of Information technology and engineering, Jinzhong university, Jinzhong, Shanxi, 030600, China
*Corresponding author’s e-mail: wenjuanjuan_2001@163.com

Abstract. The results of plant leaf classification can be used in many fields, such as plant protection. In addition, it also plays a certain role in the spread of plant diversity knowledge. At present, the recognition of plants by leaves has become a popular issue in many fields. This paper focuses more on what kind of network structure to choose to identify plant leaves. Convolution neural network is a popular network structure in recent years. Features are extracted by convolution layer, feature dimension reduction is completed by pooling layer, and finally a column vector is output by fully connected layer, and leaf recognition and classification are completed by classifier. Therefore, we construct an 11 layer convolution neural network, and select plant leaves to complete the five classification of plant leaves, and try to modify the different parameters of the network structure, choosing various optimizer, the average accuracy of the final test set is more than 99%. From the experimental results, the model has high recognition rate and high accuracy for the existing blades, which has a certain practical significance.

1. Introduction
The most crucial component of precise plant leaf identification, whether it’s for plant protection or plant knowledge, is the proper identification of plant leaves. Manual recording was once employed to gather and identify plant leaves, but this method is not conducive to preservation, and human eye identification is also inaccurate. The initial step in the classification of plant leaves is to gather images of plant leaves. Plants have roots, stalks, leaves, flowers, fruits, and other parts, to name a few.
Many researchers have put forward some excellent methods for leaf feature extraction. These methods play an important role in the classification and recognition of plant leaves, and also improve the recognition rate significantly. Gao Liang\(^{[1]}\) proposed a multi-feature fusion method based on plant leaf recognition. The plant leaves are collected firstly, secondly, based on the extraction of leaf texture and geometric features, the corner matrix describing the contour is added. Based on these three features, the comprehensive similarity of the leaves is calculated, and the recognition rate can reach 97.5%. Zheng Yili\(^{[2]}\) proposed plant leaf recognition based on multi-feature dimension reduction. Principal component analysis and discriminant analysis were used to reduce the dimension of the extracted plant features. The leaf images of Flavia database and ICL database were classified and recognized by support vector machine. The recognition accuracy was 92.52% and 89.97% respectively. Jiao Zhizhao\(^{[3]}\) et al. have compared four kinds of network structures,(VGG16, ResNet50, DenseNet121, cGAN respectively) and applied them to leaf recognition, and the recognition rate is between 92% and
97%. Hu Shengqiang uses ensemble classifier and random forest, and uses BP neural network to realize the recognition of leaves. His overall recognition accuracy for 20 types of leaves is 91%, while the recognition accuracy for single texture feature is about 95%.

2. Task and methods
This section mainly introduces that the main task of this paper is to deal with the recognition rate of blade five classification problem, and explains that the method used to deal with this problem is neural network method.

2.1. Task of this article
Computers are widely employed in a variety of disciplines due to the rapid expansion of society, science, and technology. The use of a neural network improves the accuracy of image identification even more. The above-mentioned research approaches for leaf feature extraction, as well as the recognition and classification of leaf features using neural networks, are more directly tied to the topic of computer vision. The basic idea and steps of feature extraction are to grayscale the input color image, select different filters to denoise the gray image, transform the denoised image into binary image, and perform a series of rotation, corrosion, and expansion processing on the binary image, in order to remove the influence of random factors in the process of leaf acquisition. Finally, the blade has a vector feature.

However, because this study focuses on the network structure, how to create a convolution neural network, and how to use the network to improve the recognition rate of plant leaves, the objective of this study is to use the network to increase the recognition rate of plant leaves. The image is merged into a three-channel image with a resolution of 100 * 100 pixels.

In this paper, an 11-layer convolution neural network is built. To extract features and lower the dimension of the extracted features, the first to eighth layers cycle between convolution and pooling layers. Three full connection layers make up the final three layers. Finally, a column vector is generated, and the work of recognizing and classifying plant leaves is completed.

We choose five kinds of leaves, each with 40 pictures, and divide them into training set and test set. At the same time, we adjust the network parameters to achieve the goal of minimum loss and highest recognition rate.

2.2. Methods of this article
The plant leaf image database, as well as the recognition procedure, comprising image preprocessing, network structure, data training and testing, and various parameters, will be detailed in this part. Five leaves were selected and 40 images were selected for each leaf. The five plants were Plantain, Chinese parasol, Rhododendron, Magnolia and Ginkgo biloba.

There are 200 photos in the database, each with a resolution of 416 * 416 pixels. To begin, the blade's pixels are modified to 100 * 100. The reason for this is that the data can be decreased and the network structure's operation speed can be enhanced, enhancing the algorithm's efficiency; second, because the image is a color image, the number of channels is set to three (RGB three channels).

The actual image processing procedure is as follows:
To begin, read the image, define a function, read each leaf image in the path in turn, and put the read image into the list, simultaneously labeling the read image. Finally, use the Numpy tool to convert the images list and labels list generated after the reading work into matrices, and return two matrices.

The image sequence is then scrambled, with the number of samples determined by the data matrix read in the previous phase. The goal is to make picture extraction more random so that the recognition rate is more dependable and convincing, and so that the classification problem may be applied to bigger sample sizes in the future.

Finally, a portion of the image serves as the training set, while the remainder serves as the validation set. The experiment reveals that when the training set is 60 percent and the validation set is
40 percent, the model has the best effect and can achieve a better recognition rate.

The following are the experimental results for varied proportions of training set, validation set, and loss function (35 percent, 40 percent, 60 percent, and 65 percent, respectively):

Figure 1. The accuracy of training and validation set under Ratio=0.35

Figure 2. The value of loss Function under Ratio=0.35

Figure 3. The accuracy of training and validation set under Ratio=0.4

Figure 4. The value of loss Function under Ratio=0.4

Figure 5. The accuracy of training and validation set under Ratio=0.6

Figure 6. The value of loss Function under Ratio=0.6
According to the above experimental results, regardless of whether proportion of training set and validation set is chosen, the pattern of the loss function is that it rises to progressively reduce and tends to zero from the beginning of the learning stage. However, after 30 times of training, the selection proportion of 0.6 reaches a stable level. The accuracy of the training set and the validation set tend to be the same, while any other proportion is oscillatory.

Therefore, this paper chooses 0.6 as the ratio of training set and validation set.

3. Experiment

In this section, we begin the experiment, mainly from the network structure, data training, situation discussion and experiments under different parameters. The experiments under different parameters include the selection of convolution layer parameters, the selection of pooling layer parameters, the selection of optimizer and the selection of iteration times.

3.1. Data training and testing

Different training times are chosen for training in this paper (50, 100, 150, 200 times respectively).

To begin training data, we must first extract data. Instead of reading all data at once, we define the function mini batches, which is used to acquire data in batch order. The goal is to increase the speed with which data is read; throughout the experiment, set the batch_size=64.

Finally, we get the value of each parameter under the optimal recognition rate by altering different parameters in the network hierarchy, resulting in a recognition accuracy of more than 99.5 percent. The model can accurately estimate the species of plant leaves in this scenario.

3.2. Network structure

To begin, an 11-layer convolution neural network with four convolution layers, four pooling layers, and three fully connected layers is built.

The convolution layer all uses the complement of 0, so the convolution layer length and width remain unchanged, only the depth is deepened. The length and width of the pooling layer decrease and the depth remains unchanged.

The convolution layer extracts features from the input feature graph, whereas the pooling layer compresses it. On the one hand, it reduces the size of the feature graph and reduces the network's computational cost; on the other hand, feature compression extracts the essential features.

All of the characteristics are connected via the full connection layer, which also transmits the output value to the classifier.

The forth layer (pool layer) in this study selects maximum pooling because average pooling makes
the picture blurry and maximum pooling can maintain more edge information and increase the accuracy of the validation set.

Dropout is utilized in the fully connected layer to avoid overfitting, causing some nodes to be output to 0 at random during the training process. Besides, the fully connected layer's weight is regularized to prevent overfitting.

The structure of the entire network will then be depicted in the form of a chart:

![Network Structure Diagram](image)

3.3. Discussion

The experimental findings of various parameters are studied and compared in this section, and the corresponding values of each parameter are produced in the case of the maximum recognition rate of plant leaves. The proportion of training and test sets, the size of the convolution kernel in the convolution layer, the size of the convolution kernel in the pooling layer, the number of training times, and the optimizer selection are among these parameters. By comparing the experimental results of various parameters, the parameters that result in the highest recognition rate and the smallest number of training sessions are identified. And the loss function, the accuracy of training set and the accuracy of test set are plotted to get a visual image, which can more intuitively depict the training process and the final result.

3.4. Results under various parameters

In this subsection, we choose four aspects to do the experiments. Through choosing different parameters, we finally give the best parameters when this network has the highest ratio of recognition. Then we compare the various results and analysis them.

3.4.1. Different convolution layers kernel-size

We change the convolution kernel size of the four convolution layers to see which one extracts features the best and achieves a consistent recognition rate with less training time. 3 * 3 and 5 * 5 are the most popular convolution kernel sizes. We are attempting to use these two convolution kernels for distinct convolution layers in this example. As a result, the following experiments were conducted:

1. Change the convolution kernel size of layer3-conv2 to 3 * 3;
2. Change the convolution kernel size of layer5-conv3 to 5 * 5;
3. Change the convolution kernel size of layer3-conv2, layer5-conv3 and layer7-conv4 to 5 * 5;
4. Change the convolution kernel size of layer3-conv2 to 3 * 3 and change the convolution kernel size of layer5-conv3 to 5 * 5;

The following are the experimental results for four scenarios, as well as their related loss functions:
Figure 10. The accuracy of training and validation set under situation 1

Figure 11. The value of loss Function under situation 1

Figure 12. The accuracy of training and validation set under situation 2

Figure 13. The value of loss Function under situation 2

Figure 14. The accuracy of training and validation set under situation 3

Figure 15. The value of loss Function under situation 3
3.4.2. Different pooling layers kernel-size

The pooling layer's main purpose is to minimize the dimension of features. We test 1 * 1 kernel and 2 * 2 kernel for different pooling layers, which can reduce the dimension and information loss. We run the following three experiments to change the kernel size of the pooling layer:

1. Change the size of convolution kernel of layer2-pool1 to 1*1;
2. Change the size of convolution kernel of layer2-pool1 and layer4-pool2 to 1*1;
3. Change the size of convolution kernel of layer2-pool1, layer4-pool2 and layer6-pool3 to 1*1;

In this part, we change the size of convolution kernel of pooling layer, and do the above three experiments respectively.

According to the three experimental results, when changing the convolution kernel size of the first pooling layer to 1 * 1, when the training times is more than 60, the recognition rate of the training set and the validation set tends to be the same, the number of training times is too large, the time is too long, which increases the time complexity of the algorithm.

Then the second experiment, we try to change the convolution kernel size of the second pooling layer from 2 * 2 to 1 * 1. Although the pooling layer is used to reduce the dimension of the image features extracted from the convolution layer, the convolution kernel size of 1 * 1 has no significant effect on the reduction of the feature dimension, but the convolution kernel size fuses the information of other channels and reduces the amount of calculation. It can also be seen from the experimental results that the recognition rate of the training set and the validation set reaches 1 when the number of training is about 10 times by using the 1 * 1 convolution kernel, and the recognition rate of the training set and the validation set reaches 1, and in the follow-up training process, it also present a relatively stable state.

In the last experiment, the convolution kernel size of the first, second and third pooling layer is changed to 1 * 1, but according to the experimental results, this change has no good effect, on the contrary, the recognition accuracy of the validation set can not reach the recognition rate of the training set. Therefore, when 1, 2 and 3 pooling layers are changed to 1 * 1 convolution kernel, the recognition rate does not achieve good results because there is no feature dimension reduction.

Based on all the above discussion and analysis, we finally choose the kernel size of 1 * 1 both in the first and second pooling layers and choose the kernel size of 2 * 2 both in the third and forth pooling layers, which show a stability of the internet structure.

The experimental results of three cases and their respective loss functions are given:
Figure 18. The accuracy of training and validation set under situation 1

Figure 19. The value of loss Function under situation 1

Figure 20. The accuracy of training and validation set under situation 2

Figure 21. The value of loss Function under situation 2

Figure 22. The accuracy of training and validation set under situation 3

Figure 23. The value of loss Function under situation 3
3.4.3. Different optimizer

In the common neural network structure, there are many kinds of optimizer to choose from, and each optimizer has its own unique characteristics, so the use of the situation is not the same. In this paper, we select four kinds of common optimizer for experiments, and we can judge two points through the experimental results. One is to judge whether the optimizer is suitable for the current network model. The second is to judge whether the optimization effect of the optimizer can improve the recognition rate of the training set and the validation set or reduce the training times, so as to reduce the running time, and according to the previous analysis, when the proportion of training set and validation set is 60%, the recognition rate is the highest, so the selection of optimizer is based on the premise of ratio = 0.6, the experimental results of each optimizer and its loss function are shown in the figures below:

Figure 24. The accuracy of training and validation set under Adadelta

Figure 25. The value of loss Function under Adadelta

Figure 26. The accuracy of training and validation set under Adagrad

Figure 27. The value of loss Function under Adagrad
3.4.4 Different epoch times
In terms of training times, a larger number of training sessions, in our experience, increases the correctness of the training and validation sets. However, due to the temporal complexity and the number of training times required to attain stability, the following epoch values for training were selected:
Figure 3.2. The accuracy of training and validation set under Epoch=50

Figure 3.3. The value of loss Function under Epoch=50

Figure 3.4. The accuracy of training and validation set under Epoch=100

Figure 3.5. The value of loss Function under Epoch=100

Figure 3.6. The accuracy of training and validation set under Epoch=150

Figure 3.7. The value of loss Function under Epoch=150
From the above experimental results, when the number of training is 50 times, the accuracy of training set and validation set does not reach a good state until the end of training, and the highest loss value is more than 8000. When the number of training times is 100, the accuracy of training set and validation set is basically stable after 40 times of training, but the peak value of loss is more than 14000. When the number of training times is 150, after iteration of 30 times, the accuracy of training set and validation set tends to be consistent and stable, but the peak value of loss function is more than 8000. Finally, when the number of training times is set to 200, the number of training is basically stable after 25, and the peak value of loss function is about 7000.

According to the analysis, increasing the number of training would enhance the recognition rate. However, if the number of training is too large, the training time of the entire network structure will be too long, increasing the time complexity. Hence, the final number of training times is 200.

4. Conclusion

The classification of plant leaves is of great significance for the protection of rare species, the classification of new species and the judgment of genetic relationship between plants. However, if the traditional recording methods are used to record the morphological characteristics of leaves, it will not only take time and effort, not easy to preserve, but also have errors caused by eye observation. Since the convolution neural network was proposed, after continuous development, convolution neural network has been applied to many fields, such as image processing, face recognition, data classification etc.

4.1. Result description

Combined with a leaf recognition method based on convolution neural network proposed in this paper, a network structure with higher recognition rate is obtained by changing the parameters and layers in the network structure based on the classical network Alex net. The network is a neural network with 11 layers. By constantly adjusting the parameters, the final recognition rate is more than 99%.

4.2. Deficiency and future research

The leaf recognition and classification model is limited to five different leaf classifications. In actual, five classifications are insufficient for classifying and identifying leaves for plant protection or plant diversity research. As a result, it is important to supplement the plant leaf database.

The leaf database used in this study consists of five different types of leaves with distinct physical properties. The recognition rate for leaves with comparable morphology may be low.

The experimental results demonstrate that the recognition rate of leaves is greater when the number of training times is 200, but as the number of training times increases, the program's running time becomes excessively long. This study does not address the issue of long running time.
However, the function of the network is limited, and it needs to be continuously adjusted and optimized and tested many times. The network in this paper only classifies the leaves with obvious morphological differences, and it is a five classification problem. But in nature, we need to distinguish and classify thousands of images, but blindly increasing the number of network layers will lead to too much computation and too slow running speed. Therefore, in order to better use the network in practice, the multi classification of leaves with large morphological differences is the key problem to be solved in the future research.

If the multi classification problem with a large amount of data can be realized, there will be better solutions to plant protection and a series of problems, and the network can be applied to more classification problems.

Acknowledgments
I would like to thank the provider of the database for giving me a data base for my experiment. I would also like to thank the previous scholars who research on leaf classification for providing me with a reference method.

References
[1] Gao, L. (2017) Plant leaf recognition based on fusion of multiple features. Acta Agriculturae Zhejiangensis, 29(4).
[2] Zheng, Y.L. (2017) Method of Leaf Identification Based on Multi-feature Dimension Reduction. Journal of agricultural machinery, 48(3).
[3] Jiao, Z.H. (2019) Plant Leaf Recognition Based on Conditional Generative Adversarial Nets. In: ICIC 2019, LNCS 11643.
[4] Hu, S.Q. (2016) Plant Recognition Method Based on Leaf Images. Henan University of Technology.
[5] Liu, K. (2019) Research on Multi-feature Fusion Neural Network in Plant Leaf Recognition. Lanzhou University.
[6] Bo, Q.W. (2017) Study on Plant Leaf Recognition Based on Convolutional Neural Network. Zhejiang A&F University.
[7] Ye, P. (2010) The classification and identification of plants based on the Bp neural network[D]. Suzhou University.
[8] Wu, H.S. (2017) Automatic Leaf Recognition Based on Deep Convolutional Networks. In: ICONIP 2017, Part III, LNCS 10636.