Research on crop disease recognition based on Multi-Branch ResNet-18

Chaofan Wang1, Peng Ni1 and Maoyong Cao1*

1 College of Electrical Engineering and Automation, Shandong University of Science and Technology, Qingdao, Shandong, 266000, PRC

*Corresponding author’s e-mail: cmy@qlu.edu.cn

Abstract. Traditional image processing has many problems in crop disease identification, such as complicated manual design and low efficiency. This article studies the application of deep learning algorithms in the identification of crop diseases. In this article, attention mechanism and feature fusion are introduced to optimize ResNet-18, and for the problem that the network has only a single output, based on the optimized ResNet-18, a Multi-Branch ResNet-18 (MB-ResNet-18) and a joint loss function are proposed to achieve the simultaneous classification of crop-type level, disease-type level, and disease-degree level. The experimental results show that, compared with ResNet-18, the network structure proposed in this article basically maintains the classification accuracy of crop-type level and disease-type level, and the classification accuracy on crop disease-degree level has increased by 2.11%.

1. Introduction

China is a large country with a population of 1.4 billion, food security is fundamental to the national economy, people’s livelihood. With the continuous development and progress of information technology, the development of big data technology and high-performance computing hardware has provided a new development direction for the development and transformation of modern agricultural technology, integrating artificial intelligence and big data and other high-tech in traditional agricultural production, "smart agriculture "It came into being. The basic process of traditional computer vision to identify the types of crop leaf diseases is: extracting crop leaf images or diseased areas from original images; extracting features such as edges, colors, and textures in the crop leaf images or diseased areas; manually selecting and designing effective feature variables; using traditional computer vision method builds a classifier to realize the recognition of crop leaf diseases. However, this method requires a lot of manpower, is inefficient, and has low accuracy.

With the rapid development of computer vision and the neural network, the research on crop leaf disease recognition algorithms based on computer vision and the neural network has begun to receive extensive attention from scholars at domestic and abroad[1]. In domestic research, Sun Jun et al.[2] used batch normalization and global average pooling to improve the AlexNet model and trained on the Plantvillage data set, which improved the problem of slow convergence and a large number of parameters of the AlexNet model and improve the accuracy of the model’s prediction. Zhang Jianhua et al. [3] froze part of the layers of the pre-trained VGG convolutional neural network and fine-tuned the remaining layers to classify the five diseases and normal leaves of cotton, and the recognition accuracy reached 89.51%. In abroad studies, Sjadojevic et al.[4] used a convolutional neural network
to identify crop leaf diseases by fine-tuning CaffeNet, and the accuracy rate reached 96.3%. Mohanty et al.[5] used pre-training fine-tuning AlexNet model to identify 26 diseases of 14 crops in PlantVillage and achieved a classification accuracy of more than 90% on the test set.

The existing algorithms for identifying crop leaf diseases based on neural networks are mostly only for identifying a few types of diseases, and cannot distinguish the degree of disease, and it is difficult to provide a reliable reference for intelligent pesticide application. Most neural network algorithms can only predict the label of one level of disease, and cannot achieve multi-level prediction of crop-type level, disease-level, and disease-degree level. In this article, contrary to the shortcomings of the existing research on crop leaf disease identification, based on ResNet-18[6], the MB-ResNet-18 network is proposed, which uses different feature maps to predict the crop-type, disease-type, and disease-degree respectively, and proposes a joint loss function, to train the MB-ResNet-18 network, achieved better classification accuracy in the final classification task.

### 2. Structural design of MB-ResNet-18

This article proposes a new ResNet-18 network model based on multi-branch, which simultaneously predicts the multi-level labels of the image through different feature maps in the network model. By constructing a joint loss function, while improving the accuracy of the model's recognition of the level of disease-degree, it maintains the model's ability to recognize the level of crop-type and disease-type, and speeds up network convergence. The algorithm flowchart of the MB-ResNet-18 network is shown in figure 1.

![Algorithm flowchart](image)

**Figure 1. Algorithm flowchart.**

#### 2.1. Network structure design

The convolutional neural network has several convolution kernels in each convolutional layer, which can learn the local spatial connection mode of all channels. That is to say, all the convolution kernels
in the convolutional layer extract the fusion information of the space and channel in the local receptive area, plus the nonlinear activation layer and the down-sampling layer, CNN can obtain the hierarchical pattern with the global receptive area as the description of the image. Some works have shown that the performance of the network can be improved by adding a learning mechanism that helps to obtain the correlation of feature maps. In the process of evaluating disease-degree, it is often necessary to evaluate the disease-degree with tiny features. Therefore, in this article, we choose to add the attention mechanism in SE-Net[7]. By integrating the global information of each channel of the feature map extracted by the convolutional layer, the weights of each channel of the feature map are rescaled, and the importance of the feature map channels is distinguished by different weights, obtain the channels that the network model needs to pay attention. The channel feature can adaptively change the weight of each channel. By rescaling the importance of the feature map channel, the classification accuracy of the assessment task of the severity of crop diseases can be improved. The structure of the attention mechanism is shown in figure 2.

![Figure 2. Structure of attention mechanism.](image)

In addition, the feature maps extracted by layer2, layer3, layer4, and layer5 are merged across layers through a channel splicing method to make the final feature representation as rich as possible, and further improve the classification accuracy of the crop disease-degree assessment task. On this basis, additionally selects the feature output of layer3 and layer4 as nodes, adding two longitudinal connections. After global pooling, the two groups of feature outputs are sent to the classifier to perform classification tasks of crop-type and disease-type. The specific structure of MB-ResNet-18 is shown in figure 3.

![Figure 3. MB-ResNet-18 structure.](image)

2.2. Joint loss function design

Literature[8] pointed out: "Multi-task learning improves generalization performance by using domain knowledge contained in the supervision signals of related tasks." The hard sharing mechanism of parameters is the most common way of multi-task learning in neural networks. And literature[9]
proves that the more tasks learn at the same time, the more tasks the trained model can capture the same representation, which leads to the smaller the risk of overfitting on the original task.

Based on this, this article combines the loss functions of the three classification tasks into a joint loss function and uses the correlation between tasks to promote each other. While the network can use the feature maps of different convolution groups to predict different levels of labels, the recognition accuracy of the model is improved, and speed up the convergence of the model. The expression form of the joint loss function is shown in formula (1).

\[
\begin{align*}
F_{\text{loss}} &= \alpha f_s + \beta f_d + \gamma f_i \\
f_s &= -\frac{1}{n} \sum_{y_s=1}^{n} \left[ y_s \ln(a_s) + (1 - y_s) \ln(1 - a_s) \right] \\
f_d &= -\frac{1}{m} \sum_{y_d=1}^{m} \left[ y_d \ln(a_d) + (1 - y_d) \ln(1 - a_d) \right] \\
f_i &= L_{\text{dis}} - \lambda \times L_{\text{div}} \\
L_{\text{dis}}(F) &= L_{\text{CE}} \left( y, \left[ e^{g(F_0)} e^{g(F_1)}, \ldots, e^{g(F_{c-1})} \right]^T \right) \\
g(F_i) &= \frac{1}{WH} \sum_{k=1}^{WH} \max_{j=1,2,\ldots,c} \left[ M_i \times F_{i,j,k} \right] \\
L_{\text{div}}(F) &= \frac{1}{c} \sum_{i=0}^{c-1} h(F_i) \\
h(F_i) &= \sum_{k=1}^{WH} \max_{j=1,2,\ldots,c} \left[ \sum_{k'=1}^{WH} e^{F_{i,j,k'}} \right]
\end{align*}
\]

Among them, \(F_{\text{loss}}\) represents the proposed joint loss function, \(\alpha, \beta, \gamma\) are the hyperparameters of the improved joint loss function, which are used to adjust the proportion of the loss function; \(f_s\) and \(f_d\) are the cross-entropy loss functions, which are used to calculate the loss of output of Layer3 and Layer4 output; \(f_i\) is the MC-Loss loss function, \(L_{\text{dis}}\) and \(L_{\text{div}}\) are the discriminative component and the diversity component, respectively, see formula (2) for details.

3. Experiment and result analysis

3.1. Data sets introduction

This article uses part of the data of the crop leaf disease data set in the 2018 AI_Challenger, which contains a total of 9 kinds of plants, 15 kinds of diseases, and the disease distinguishes two kinds of disease levels and a total of 39 types of data. The training set contains 16,931 images, and the validation set contains 2716 images. The training set and the validation set are mixed and randomly divided into ten parts, 9 of which are used as the training set, and the remaining one is used as the validation set. Part of the experimental data is shown in figure 4.

To verify the MB-ResNet-18 proposed in this article, the image’s label is expanded from a single label to a form of multi-level classification of crop-type level, disease-type level, and disease-degree level. The data set at the crop-type level consists of 9 types of labels, excluding the disease-type; the data set at disease-type level is composed of 24 types of labels, including 15 types of disease labels and 9 types of health labels, which do not distinguish the degree of disease; The disease-degree level consists of two types of labels: general and severe, which do not distinguish disease-type level.

The original data set has a total of 19647 images, but the data distribution is extremely unbalanced. Among them, the potato health category with the largest number has a total of 1493 images, while the cherry powdery mildew general and severe categories have only 119 images respectively. To alleviate
the impact of category imbalance on the model training performance, this article uses rotation, flip, and normalization methods to preprocess the original training set.

![Crop disease images](image1)

**Figure 4.** Crop disease images.

Horizontal flip uses the vertical axis of the image as the inversion line to flip the image left and right. The image after horizontal flip is shown in figure 5(a); the vertical flip uses the horizontal axis of the image as the inversion line to flip the image up and down. The image after vertical flip is shown in figure 5(b); Rotation is to randomly rotate a certain angle clockwise with the focus on the center of the image. The randomly rotated image is shown in figure 5(c).

![Data pre-processing schematic](image2)

**Figure 5.** Data pre-processing schematic.

### 3.2 Model parameter setting

The computer operating system for this experiment is a 64-bit system of Ubuntu 20.04, and all programs are written in the Python language under the Pytorch framework. Computer memory is 32GB, equipped with Interl® Core™ i5-8500 CPU @ 3.00GHz processor and NVIDIA RTX3090 24GB graphics card.

Hyperparameters play a very important role in the training of the model. In this experiment, Adam optimizer is selected to update and calculate network parameters. Use L2 regularization and Dropout method to reduce the over-fitting problem of the network model, where the regularization parameter is 0.0001, and the Dropout parameter is set to 0.5; The initial learning rate is set to 0.001, and the learning rate drops ten times every 15 rounds; The network model uses a small batch training method, and the batch size is set to 64, and the number of epochs is set to 100; α, β, and γ are set to 1 in the joint loss function.

### 3.3 Experimental results

This article uses accuracy to evaluate the performance of the model. The accuracy index is defined as follows:

$$\text{Acc} = \frac{N_r}{N} \times 100\%$$

where:
\(N_p\) — the number of validation samples that are predicted to be correct

\(N\) — the total number of validation samples

To verify the classification performance of MB-ResNet-18, choose to set up three groups in the same environment, in which experiments 1 and 2 use MCLoss as the loss function, and experiment 3 uses the joint loss function proposed in this article as the loss function. The experimental results are shown in table 1.

Table 1. Experimental result.

| Experiment number | Model                     | Crop-type level | Disease-type level | Disease-degree level |
|-------------------|---------------------------|-----------------|-------------------|---------------------|
| 1                 | ResNet-18                 | 96.14%          | 93.46%            | 87.88%              |
| 2                 | ResNet-18 + Attention module + Feature fusion | 95.21%          | 90.93%            | 89.05%              |
| 3                 | MB-ResNet-18              | 95.47%          | 93.97%            | 89.99%              |

Analysis of the experimental results shows that the prediction accuracy rate of the MB-ResNet-18 network at the crop-type level is slightly lower than the original ResNet-18 network. This is because MB-ResNet-18 is used the feature output of the shallow convolution group when predicting the crop-type level, and the prediction accuracy of the disease-type level and the disease-degree level are higher than ResNet-18. BS-ResNet-18 has a prediction accuracy rate of 93.97% and 89.99% for the two levels of disease-type level and disease-degree level, respectively. Compared with the original ResNet-18, the prediction accuracy rates of the two levels are increased by 0.51% and 2.11%, respectively.

To verify the stability of the prediction accuracy of BS-ResNet-18, an experiment was performed using ten-fold cross-validation. The data set samples are randomly divided into ten parts with the same amount of data. Each training model uses nine of the data sets as the training set, and the last one as the validation set. The experimental results in table 1 use data from data set 0 to data set 8 as the training set, and data set 9 as the validation set.

The results of the BS-ResNet-18 network ten-fold cross-validation are shown in table 2. The lowest accuracy rate of crop-type level is 95.47% appear in the data set 9, the highest accuracy rate of crop-type level is 96.55% appear in data set 2. The lowest accuracy rate of disease-type level is 92.79% appear in data set 0, the highest accuracy rate of disease-type level is 94.44% appear in data set 1. The lowest accuracy rate of disease-degree level is 89.14% appear in data set 5, the highest accuracy rate of disease-degree level is 91.45% appear in data set 8. The fluctuation of the accuracy rate is relatively stable.

Table 2. Ten-fold verification of recognition accuracy

| Ten-fold cross-validation data set | Crop-type level | Disease-type level | Disease-degree level |
|-----------------------------------|-----------------|-------------------|---------------------|
| data set 0                        | 95.88%          | 92.79%            | 90.13%              |
| data set 1                        | 95.78%          | 94.44%            | 90.74%              |
| data set 2                        | 96.55%          | 94.23%            | 89.90%              |
| data set 3                        | 96.40%          | 94.39%            | 91.06%              |
| data set 4                        | 95.83%          | 94.28%            | 91.05%              |
| data set 5                        | 95.67%          | 93.97%            | 89.14%              |
| data set 6                        | 96.19%          | 94.38%            | 90.94%              |
| data set 7                        | 95.93%          | 93.92%            | 90.35%              |
| data set 8                        | 95.67%          | 93.82%            | 91.45%              |
| data set 9                        | 95.47%          | 93.97%            | 89.99%              |
4. Conclusion
In this article, aiming at the problem of ResNet-18's low recognition accuracy of crop disease-degree, firstly, the attention mechanism, and feature fusion are introduced to optimize the network, which improves the recognition accuracy of crop disease-degree. Secondly, based on the optimized ResNet-18, BS-ResNet-18 is proposed for the problem that the network can only output a single output. In this way, the network can classify at the three levels of crop-type, disease-type, and disease-degree at the same time, and by constructing the joint loss function, the model can not only improve the recognition accuracy of disease-degree level but also maintain the recognition ability of the model for crop-type level and disease-type level. The final experimental results show that BS-ResNet-18 is compared to the original ResNet-18, the prediction accuracy rate at the crop-type level only decreased by 0.67%, and the prediction accuracy rates at the disease-type level and disease-type level increased by 0.51% and 2.11%, respectively. Finally, ten-fold cross-validation is used to verify the performance stability of BS-ResNet-18. The experimental results show that the classification accuracy of BS-ResNet-18 is stable at three levels.

References
[1] Rastogi A, Arora R, Sharma S. (2015) Leaf disease detection and grading using computer vision technology & fuzzy logic. In: 2015 2nd international conference on signal processing and integrated networks, 500-505.
[2] Sun J, Tan Wj, et al. (2017) Recognition of multiple plant leaf diseases based on improved convolutional neural network. J. Transactions of the Chinese Society of Agricultural Engineering, 33: 209-215.
[3] Zhang Jh, Kong Ft, et al. (2018) Cotton disease identification model based on improved VGG convolution neural network. J. Journal of China Agricultural University, 23:161-171.
[4] Srdjan S, Marko A, Andras A, et al. (2016) Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification. J. Computational Intelligence and Neuroscience, 2016: 1-11.
[5] Mohanty S P, Hughes D P, Salathé M. (2016) Using Deep Learning for Image-Based Plant Disease Detection. J. Frontiers in Plant Science, 7: 1419-1425.
[6] He Km, et al. (2016) Deep Residual Learning for Image Recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, Las Vegas. 770-778.
[7] Hu J, Shen L, Sun G. (2018) Squeeze-and-excitation networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, Salt Lake City, 7132-7141.
[8] Caruana. R. (1998) Multitask Learning. J. Autonomous Agents and Multi-Agent Systems. 27: 95-133.
[9] Baxter, J. (1997) A Bayesian/ Information Theoretic Model of Learning to Learn via Multiple Task Sampling. J. Machine Learning. 28: 7-39.