ISSN: 1813-4890

Crop disease identification based on MRLeNet model

Chaoqi Hu[1, a], Yaohui Li[2, b, *] and Zehao Fu[3, c]

[1]School of Information Technology and Engineering, Tianjin University of Technology and Education, Tianjin 300222, China

[2]School of Information Technology and Engineering, Tianjin University of Technology and Education, Tianjin 300222, China

[3]School of Information Technology and Engineering, Tianjin University of Technology and Education, Tianjin 300222, China

[a]2606126649@qq.com, [b]philliplee@163.com, [c]2290112047@qq.com

Abstract

In order to improve the accuracy of crop disease identification, a model called MRLeNet is proposed based on the LeNet-5 model. Firstly, the residual block and Squeeze-and-Excitation block are combined into a SE-ResNet Module to improve the fitting ability of the model. Secondly, the features extracted from the previous layers of the model are also fed into the fully connected layer for classification, so that the model can integrate high level features and low level features. Moreover, Dropout is added to the model to prevent overfitting, and the exponential decay learning rate is used as the learning rate. In the PlantVillage dataset, the recognition accuracy of the MRLeNet model reaches 97.93%, and the recognition is better than that of the LeNet-5 model.

Keywords

Crop disease identification, Residual block, LeNet-5 model.

1. Introduction

People cannot live without crops in their daily life, and the problem of crop diseases is a long-standing problem that has been troubling people, and it is easy to make mistakes only by manual identification. Reference [1] used image and near-infrared spectroscopy to propose a pest and disease leaf detection system based on image and spectral information fusion, which basically achieved nondestructive detection. Reference [2] used a plain Bayesian classifier and achieved classification and recognition of five maize leaf spots using this classifier with an accuracy of more than 83%. Reference [3] proposed a GoogLeNet model-based method for rice spike blight detection, using Inception blocks to stack the network, and the accuracy in spike blight disease prediction could reach 92.0%. Reference [4] proposed the combination of model batch normalization and global pooling, and the recognition accuracy of the improved model could reach more than 90% after three iterations. Reference [5] used a 5-layer network model and used median filtering method and histogram thresholding method to do pre-processing of images, and the recognition rate of six common diseases of wheat could reach more than 99%. Reference [6] proposed a neural network model with high-order residuals, which has high robustness and has a high accuracy rate when tested in a practical environment. Reference [7] improved the fully connected layers and SoftMax classifier of VGG-16, using migration learning shared to parameters, and the accuracy could reach 89.51% in disease identification of cotton. Reference [8] used the Faster-RCNN model to locate grape leaves in images and the accuracy of this model in grape leaf disease recognition was better than the algorithm that performed the detection directly. The MRLeNet model proposed in this paper is improved based on the LeNet-5 model by adding the SE-ResNet Module to the model to improve the fitting ability of the model,

ISSN: 1813-4890

and at the same time, the feature maps extracted by the previous layers and the feature maps extracted by the SE-ResNet Module at the very end are fed together into the fully connected layer for classification, so that the model can simultaneously The dropout is added in the fully connected layer to prevent overfitting. In the PlantVillage dataset, the accuracy of the MRLeNet model is higher than that of the LeNet-5 model on the testing set.

2. Traditional LeNet-5 model

LeNet-5 model is widely used in handwritten digit recognition and other fields. The structure of LeNet-5 model used in this paper is shown in Figure 1, which has one input layer, two convolution layer, two maximum pooling layer, one Flatten layer, two fully connected layer and one output layer. The activation functions used in the convolution layer are all relu activation functions, and softmax functions are used in the output layer.

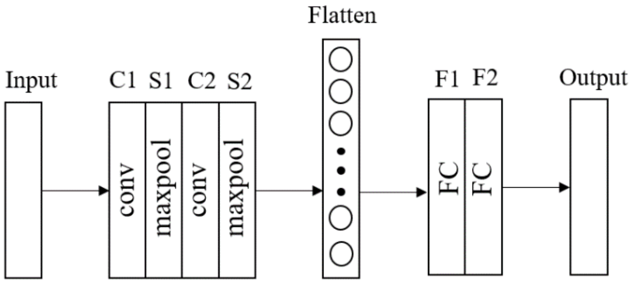

Fig. 1 LeNet-5 model

In the data preprocessing in this paper, the image size is uniformly normalized to 64×64 pixels, and the parameters of the whole LeNet-5 model are shown in Table 1. The Input layer is used to input image data. The step size of C1 is 1 and zero padding is used. The size of the image remains unchanged after passing through C1. The size of the pooled core is 2×2 and the step size is 2 without using zero padding. The step size is set to 2 without using zero padding. After passing S2, the size of the image becomes 16×16. F1 and F2 are both fully connected layers, and the number of neurons in F1 and F2 is set to 120 and 84. The dataset used in the experiment is the PlantVillage dataset, which contains 26 kinds of diseased leaves and 12 kinds of healthy leaves of 14 kinds of plants, which are divided into 38 categories in total. Therefore, the number of neurons in the Output layer is 38.

The traditional LeNet-5 model can not be used to classify plant disease images well. The depth of LeNet-5 model is too shallow, and it can not identify complex images such as plant disease well. Moreover, LeNet-5 model uses a single scale to extract features, so it is easy to ignore some feature information. To solve the above problems, this paper improved the LeNet-5 model and proposed the MRLeNet model based on LeNet-5 model to alleviate the above problems and improve the accuracy of plant disease image classification.

Table 1 LeNet-5 model hyperparameters

Layer	Number of feature maps	Feature map size	Convolution kernel/pooling kernel size	Strides
Input	-	64×64	-	-
C1	6	64×64	5×5	1
S1	6	32×32	2×2	2
C2	16	32×32	5×5	1
S2	16	16×16	2×2	2
Flatten	4096 neurons			
F1	120 neurons			

F2	84 neurons
Output	38 neurons

3. Improvements to the LeNet-5 model

3.1. Construction of residual blocks

The structure of the residual block is shown in Figure 2. The residual block is constructed by the residual connection between the convolution layers. The use of residual connection can alleviate the problem of gradient disappearance. $H(x) = F(x) + x$, $F(x)$ is the output feature map of the input feature map after two convolution layers, x is the original input feature map. If the size and depth of the output feature map and the input feature map are the same, the matrix values of $F(x)$ and x can be directly added. If the size or depth of the output feature map and the input feature map are different, the input feature map needs to be first convolved with a convolution kernel size of 1×1. The convolution operation makes the size and depth of the input feature map consistent with that of the output feature map, and then the matrix values of the input feature map and the output feature map are added. Finally, the obtained $H(x)$ is passed through the relu activation function as the output of the residual block.

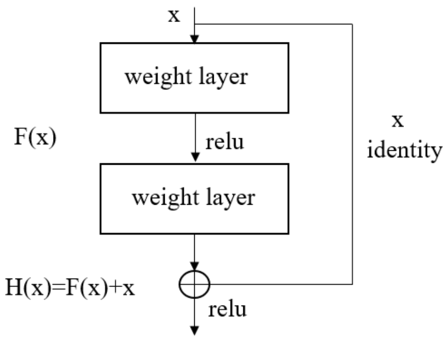

Fig. 2 Residual block

3.2. SE-ResNet Module

The Squeeze-and-Excitation block is the channel attention mechanism. The module will be involved in back propagation. The addition of Squeeze-and-Excitation block to the model can make the model allocate more weight in the more important feature channels. Less weight is assigned to less important feature channels. The Squeeze and-Excitation block will process the global mean pooling of the entered feature map for compression. The original H×W×C feature map will be changed to 1×1×C. The purpose of extrusion is to aggregate the information. The excitation operation is self-adaptive and will connect two fully connected layers. The number of neurons in the first fully connected layer is generally less than that of C, which is set as 1/16 of C in the experiment. The activation function uses relu activation function, and the number of neurons in the second fully connected layer is the same as that of C. Activation function Using sigmoid activation function, a vector of inter-channel weights is obtained. The effect of excitation is to capture channel dependent dependencies. Finally, the weights are multiplied with the original feature graph, and the model automatically focuses on the important channels according to the weights.

Reference [9] proposed the method of combining the Squeeze-and-Excitation block with the residual block, and named this module SE-ResNet Module. The structure of the SE-ResNet Module is shown in Fig.3. The Squeeze-and-Excitation block is added to the residual block. The entry feature map is firstly passed through the residual block, and the obtained feature map is then passed through the Squeeze-and-Excitation block. Finally, the matrix values of the input feature map and the output feature map are added, which are used as the output of the SE-ResNet Module after the relu activation function.

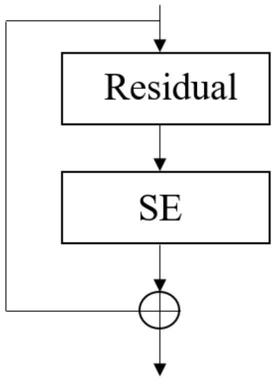

Fig. 3 SE-ResNet Module

3.3. The MRLeNet model

The structure of MRLeNet model proposed in this paper is shown in Figure 4. The traditional LeNet-5 model extracts features on a single scale, and the output feature map of each layer is only input to the next layer. In MRLeNet model, the output feature map of the last layer is not only sent to the fully connected layer, but also the output feature map of the previous part of the layer is sent to the fully connected layer for classification. In this way, the model can integrate the low-level feature information and high-level feature information, which can effectively improve the recognition effect of the model. Add a Squeeze-and-Excitation block behind the maximum pooling layer S2 to give more weight to the important channels. After SE1, the feature map is maximized in one of the routes. In addition, by connecting SE-ResNet1 all the way, the network depth can be increased, and more weights can be assigned to important channels and less weights to unimportant channels, so as to improve the accuracy of the model. After passing through SE-ResNet1, the feature map is divided into two ways. One way is also maximized for the feature map, and the other way is connected to SE-ResNet2. The feature map after SE-ResNet2 is spliced with the previous feature map after S3 and S4 maximum pooling, and the spliced feature map is Flatten. It then feeds into the fully connected layer, uses Dropout in the fully connected layer to prevent overfitting of the model, and finally connects to the output layer.

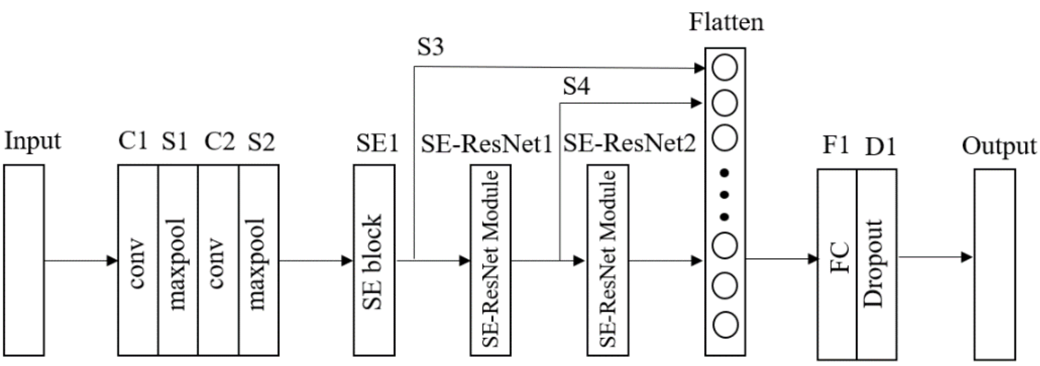

Fig. 4 MRLeNet model

In data preprocessing, the image size is uniformly normalized to 64×64 pixels, and the parameters of the whole MRLeNet model are shown in Table 2. The convolution kernel size of the convolution layer C1 is 5×5, the step size is 1, the number of output feature maps is 40, and the size is 64×64. The size of the pooled core of the maximum pooling layer S1 is 2×2 and the step size is 2. The number of output feature maps is 40 and the size is 32×32. The convolution kernel size of the convolution layer C2 is 5×5, the step size is 1, the number of output feature maps is 40, and the size is 32×32. The size of the pooling core of the maximum pooling layer S2 is 2×2, and the step size is 2. The number of output feature maps is 40 and the size is 16×16. SE1 is a Squeeze-and-Excitation block. The number of feature maps after passing through SE1 is 40. The size is 16×16. The size of the pooling core of the maximum pooling layer S3 is 4×4, and the step size is 4. The number of output feature maps is 40, and the size is 4×4. SE-ResNet1 consists of a residual block and a Squeeze-and-Excitation block. The convolution kernel size of the convolution layer in SE-ResNet1 is 3×3, the step size of the first convolution layer is 2, the step size of the second convolution layer is 1, the number of output feature maps is 80, and the size is 8×8. The size of the pooling core of the maximum pooling layer S4 is 2×2 and the step size is 2. The number of output feature maps is 80 and the size is 4×4. The convolution kernel size of the convolution layer in SE-ResNet2 is 3×3, the step size of the first convolution layer is 2, the step size of the second convolution layer is 1, the number of output feature maps is 160, and the size is 4×4. The feature map after the maximum pooling layer S3 and S4 was spliced with the feature map after SE-ResNet2, and then flattened by Flatten, a one-dimensional vector with 4480 neurons was obtained. The number of neurons in the fully connected layer F1 is set to 320, and use Dropout to prevent overfitting. The number of neurons in the output layer is set to 38.

Table 2 MRLeNet model hyperparameter

Layer	Convolution kernel/pooling kernel size	Number of feature maps	Feature map size
Input	-	-	-
C1	5×5	40	64×64
S1	2×2	40	32×32
C2	5×5	40	32×32
S2	2×2	40	16×16
SE1	-	40	16×16
S3	4×4	40	4×4
SE-ResNet1	-	80	8×8
S4	2×2	80	4×4
SE-ResNet2	-	160	4×4
Flatten		4480 neurons	
F1+D1		320 neurons	
Output		38 neurons	

4. Experiment

4.1. Data preprocessing

The dataset used in this paper is PlantVillage dataset, which is used to classify images of crop diseases and insect pests. It contains 26 kinds of diseased leaves and 12 kinds of healthy leaves of 14 kinds of crops, which are divided into 38 categories and contain 54,305 images, some of which are shown in Figure 5. In the experiment, 60% of the dataset was used as the training set, 20% of the dataset was used as the validation set, and the remaining 20% was used as the testing set. The image of the dataset is normalized to 64×64 size using bilinear interpolation method, using the exponential decay learning rate as the learning rate in the experiment, set the initial learning rate is 0.001, the learning rate decay rate is set to 0.88, every 1000 steps for a decay.

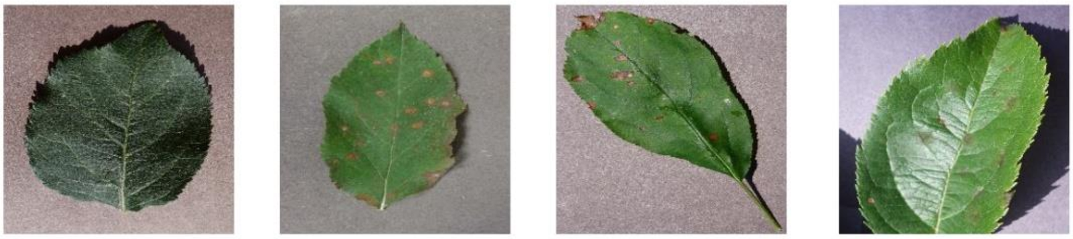

Fig. 5 Some images in PlantVillage dataset

4.2. Experimental hyperparameter setting

4.2.1. Number of neurons in F1

This paper proposes to increase the number of neurons in F1 of the fully connected layer of the MRLeNet model. By selecting the number of neurons as 160, 320 and 640 for comparative experiments, the experimental results are shown in Table 3. It can be found that when the number of neurons in F1 of the fully connected layer is 320, the recognition accuracy of the model is the highest.

Table 3 Comparison of the number of neurons in F1 layer

Number of neurons in F1 layer	Recognition accuracy rate (%)
160	97.56
320	97.93
640	97.68

4.2.2. Determination of Dropout parameters

Dropout is added behind the fully connected layer F1. Using Dropout can prevent the model from overfitting. In the experiment, the Dropout parameters are set to 0.1, 0.2, and 0.5 respectively for comparative experiments. The experimental results are shown in Table 4. The experiment found that when the Dropout parameter is 0.5, the model identification accuracy is the highest.

Table 4 Dropout parameter comparison results

Dropout parameter	Recognition accuracy rate (%)
0.2	97.34
0.3	97.53
0.5	97.93

4.2.3. Determination of the activation function

Different activation functions affect the training effect of the model. Using inappropriate activation functions will make the model appear gradient disappearance and long training time in the training process. At present, the commonly used activation functions are Sigmoid, ReLU and Tanh. The above activation functions are used for comparative experiments, and the experimental results are shown in Table 5. It can be seen from the table that the recognition accuracy of the model is the highest when the ReLU activation function is used.

Table 5 Comparison of different activation functions

Activation function	Recognition accuracy rate (%)
Sigmoid	97.40
ReLU	97.93
Tanh	96.45

4.3. Comparison of different models

The MRLeNet model proposed in this paper is compared with the improved LeNet-5 model. According to the previous comparison experiments, the hyperparameters of the MRLeNet model are selected, the number of neurons in the fully connected layer F1 is set to 320, the Dropout parameter is set to 0.5, and the activation function is selected as the ReLU activation function. On the basis of LeNet-5 model, the MRLeNet model improves the fitting ability of the model by adding SE-ResNet Module, and sends the features extracted from the previous part of the layer into the fully connected layer for classification. Compared with the single scale feature extraction of LeNet-5 model, the MRLeNet model has the advantage of multi-scale feature extraction, and the MRLeNet model also adds Dropout to prevent overfitting of the model. Through these improvements, the recognition accuracy of MRLeNet model in PlantVillage dataset is higher than that of LeNet-5 model. The specific results are shown in Table 6. The number of parameters of MRLeNet model is about 1.94 million, while that of LeNet-5 model is about 0.51 million. The number of parameters in MRLeNet model is about 3.8 times that in LeNet-5 model. The iteration time of each step of the MRLeNet model is about 49ms, and the iteration time of each step of the LeNet-5 model is about 39ms. The MRLeNet model takes 10ms more time than the LeNet-5 model. The recognition accuracy of MRLeNet model on PlantVillage dataset is 97.93%, and the recognition accuracy of LeNet-5 model on PlantVillage dataset is 90.07%. The recognition accuracy of MRLeNet model is 7.86% higher than that of LeNet-5 model.

Table 6 The results of the comparison of different models

Model	Parameter quantity/million	Time of iteration/(ms/step)	Loss	Accuracy of recognition(%)
LeNet-5	0.51	39	0.4751	90.07
MRLeNet	1.94	49	0.2008	97.93

5. Conclusion

This paper proposes a model called MRLeNet, which is improved on the basis of LeNet-5. It retains the convolution layers of C1 and C2 and the maximum pooling layers of S1 and S2 of LeNet-5 model. After S2, SE-ResNet Module is constructed to increase the depth of the model. The SE ResNet Module is composed of a residual block and a Squeeze-and-Excitation block. The residual block can increase the depth of the model while mitigating the degradation of the model. The Squeeze-and-Excitation block can be assigned different weights to the channels. The different weights can be used to get the model to focus on the important channels. Different from the single-scale LeNet-5, MRLeNet is multi-scale. MRLeNet will send the feature maps after the maximum pooling layer S3 and S4 and the feature maps after SE-ResNet2 into the fully connected layer F1. 60% of PlantVillage dataset is taken as the training set, 20% as the validation set, and the remaining 20% as the testing set. Through experimental verification, the recognition accuracy of MRLeNet model can reach 97.93%, 7.86% higher than that of LeNet-5 model.

References

[1] C.Y. Xie, D.K. Wu, C.Y. Wang, et al: Insect Pest Leaf Detection System Based on Information

[2] Fusion of Image and Spectrum, Transactions of the Chinese Society for Agricultural Machinery, Vol. 44 (2013) No.S1, p.269-272.

[3] Y.X. Zhao, K.R. Wang, Z.Y. Bai, et al: Bayesian classifier method on maize leaf disease identifying based images, Computer Engineering and Applications, Vol. 43 (2007) No.5, p.193-195.

[4] S.P. Huang, C. Sun, L. Qi, et al: Rice panicle blast identification method based on deep convolution neural network, Transactions of the Chinese Society of Agricultural Engineering, Vol. 33 (2017) No.20, p.169-176.

[5] J. Sun, W.J. Tan, H.P. Mao, et al: Recognition of multiple plant leaf diseases based on improved convolutional neural network, Transactions of the Chinese Society of Agricultural Engineering, Vol. 33 (2017) No.19, p.209-215.

[6] H. Zhang, Q. Cheng, Y.J. Wu, et al: A Method of Wheat Disease Identification Based on Convolutional Neural Network, Shandong Agricultural Sciences, Vol. 50 (2018) No.3, p.137-141.

[7] W.H. Zeng, M. Li, Z. Li, et al: High-Order Residual and Parameter-Sharing Feedback Convolutional Neural Network for Crop Disease Recognition, Acta Electronica Sinica, Vol. 47 (2019) No.9, p.1979-1986.

[8] J.H. Zhang, F.T. Kong, Z.J. Wu, et al: Cotton disease identification model based on improved VGG convolution neural network, Journal of China Agricultural University, Vol. 23 (2018) No.11, p.161-171.

[9] T.Y. Liu, Q. Feng, S. Yang: Detecting grape diseases based on Convolutional neural network, Journal of Northeast Agricultural University, Vol. 49 (2018) No.3, p.73-83.

[10] J. Hu, L. Shen, G. Sun: Squeeze-and-excitation networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7132-7141.