Research on Apple Leaf Disease Identification Method Based on Improved Deep Residual Shrinkage Network

Liu Pu, Wang Xun

School of Computing, Jiangsu University of Science and Technology, Zhenjiang, Jiangsu, 212000, China

869704298@qq.com, 302548426@qq.com

Keywords: Deep residual shrinkage network; Apple leaf disease; Dual transfer learning; Multiscale convolution

Abstract: In response to the problems of slow recognition speed and inaccurate recognition in traditional methods for identifying apple leaf diseases, this paper proposes an improved deep residual shrinkage network recognition method. This method takes the deep residual shrinkage network as the basic framework and introduces the Inception module to efficiently extract disease features at different scales, enhancing feature diversity. At the same time, double transfer learning is used to enhance the generalization ability of the model on the task of apple leaf disease with small samples, and reduce the impact of data on the model performance. The experimental results show that the proposed method achieves an accuracy of 98.8% in identifying apple leaf diseases. In addition, compared with traditional methods, the proposed method has significant improvements in many aspects such as convergence speed and robustness.

1. Introduction

China boasts the largest apple-producing region worldwide, leading the global ranks in apple production[1]. However, apple diseases have, to a certain extent, impeded the progress of China's apple industry[2]. Currently, the primary methods of apple disease diagnosis in China rely heavily on the experience of apple farmers and experts. While this method is quite common, it falls short in terms of accuracy and speed of identification, failing to meet the requirements for apple disease prevention and treatment[3]. With the rapid development of information technology, apple disease identification methods based on machine vision and deep learning have become a research hotspot. These methods, employing computer vision technology and extensive image data to train deep learning models, can accurately identify and classify apple diseases. This greatly enhances the efficiency and accuracy of apple disease diagnosis and prevention.

In recent years, Convolutional Neural Networks (CNNs)[4] have been widely applied in numerous fields such as image segmentation[5], image recognition[6], and object detection[7]. In the field of plant disease detection, CNNs have also demonstrated excellent performance. For instance, Meng Liang et al.[8] proposed a lightweight convolutional neural network model to identify vegetable disease images, which achieved a high recognition accuracy on the test dataset. Yu and colleagues[9] improved upon the base network, ResNet50, proposing a model for apple leaf disease detection. Experimental results showed that their model achieved commendable performance in terms of average precision, recall rate, and F1-score. It is evident that CNNs are capable of identifying apple leaf diseases, yet existing network models still have issues like lower recognition accuracy and slow convergence speed. Therefore, this paper proposes the Deep Residual Shrinkage Network based on Dual Transfer Learning and Inception Module (DTI-DRSN). This approach uses the deep residual shrinkage network as the base network, combined with a multi-scale module, to enhance the model's ability to capture disease features at different scales and obtain more abundant feature information. Simultaneously, the dual transfer learning training method is adopted to improve model performance.

DOI: 10.25236/icmmct.2023.013

2. Improved Deep Residual Shrinkage Network Model

2.1 Model Architecture

Figure 1 shows the overall architecture of the improved deep residual shrinkage network model, DTI-DRSN, proposed in this paper. The model first uses the Inception module to extract multi-scale features of apple leaf diseases. Then, with DRSN as the base structure, the embedded attention module, namely the Squeeze-and-Excitation (SE) module, allows the model to focus on task-related areas, thereby reducing the impact of noise information on apple leaf disease recognition. Finally, it employs global average pooling to aggregate features for classification.

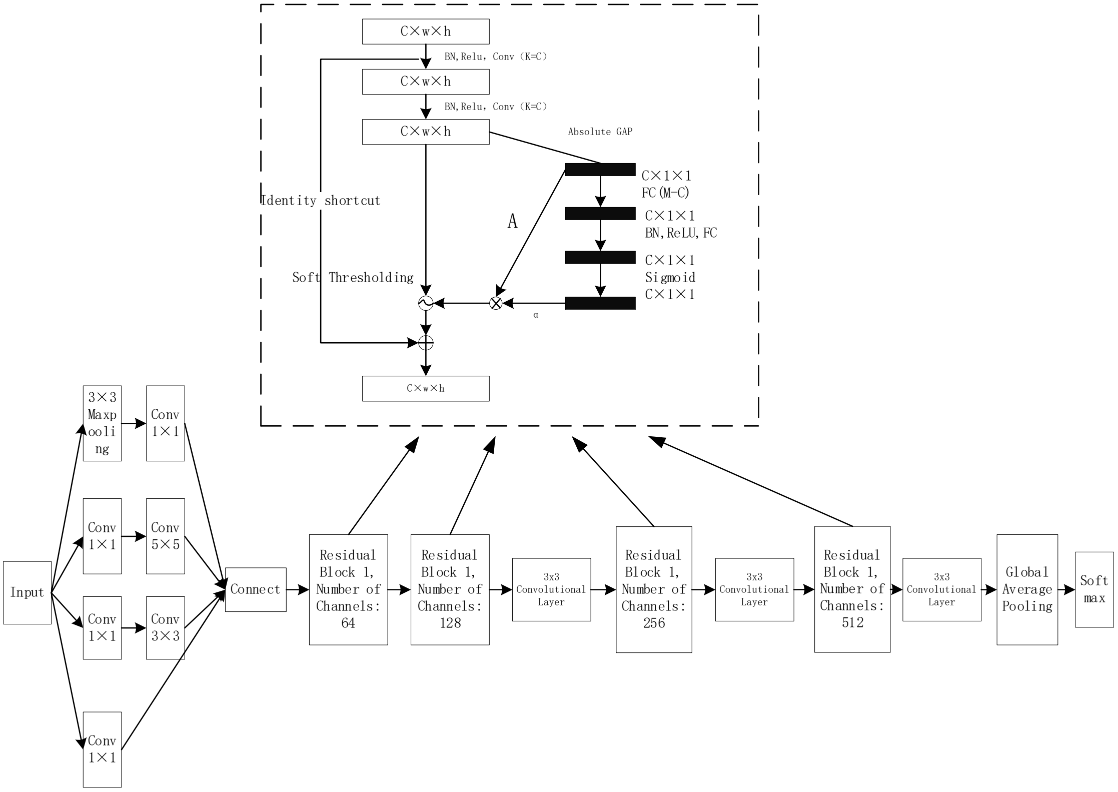


Figure 1 DTI-DRSN Network Structure Diagram

The model comprises one Inception module, four DRSN modules (specifically, a standard 3×3 convolution is applied for dimension reduction after the 2nd, 3rd, and 4th DRSN modules), and one global average pooling layer. The parameters of the relevant network layers are as shown in Table 1.

Table 1 Parameters of the DTI-DRSN Network Model

Layer Index	Operation	Output Size
Conv1_x	Inception Module, Maxpooling, s=2	112×112×64 56×56×64
Conv2_x	Deep Residual	56×56×64
Conv3_x	Deep Residual, Conv3×3, s=2,128	28×28×128
Conv4_x	Deep Residual, Conv3×3, s=2,256	14×14×256
Conv5_x	Deep Residual, Conv3×3, s=2,512	7×7×512
Average pooling	Global Average Pooling	1×1×512
Fully connected layer	Fully Connected Layer	5

2.2 Implementation Details

2.2.1 Inception Module

The Inception module[10], initially proposed by Google, operates on the core idea of performing convolution operations parallelly on the same feature map using convolution kernels of different scales. This effectively captures spatial features at different scales, enhancing the model's sensitivity to features of various scales. The Inception module employs 1x1 convolution, which retains local information while also serving the role of dimension reduction or elevation, thus avoiding overfitting of the network. The structure of the Inception module used in this paper is shown in Figure 2.

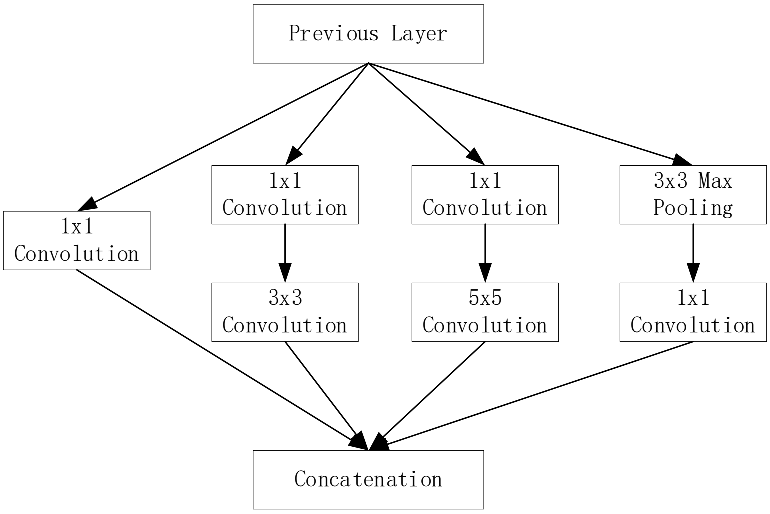


Figure 2 Structure of the Inception Module

Traditional standard convolutions utilize single-scale convolution kernels for feature extraction, often resulting in relatively monotonous feature information. However, apple leaf diseases present significant variations in disease spots across different periods and regions. Therefore, this paper employs the Inception module at the initial stage to extract features from the input images, thoroughly considering the disease feature information at different scales, hence resolving this issue.

2.2.2 Deep Residual Shrinkage Block

The Deep Residual Shrinkage Network (DRSN)[11] is a network composed of two classic neural networks (ResNet and SENet). In this network, unlike the original implementation of SENet, DRSN replaces the Reweigh operation in the SE block with soft thresholding, and adds a branch for learning a set of thresholds needed for soft thresholding, endowing different features with unique weights. The basic module structure is shown in Figure 3.

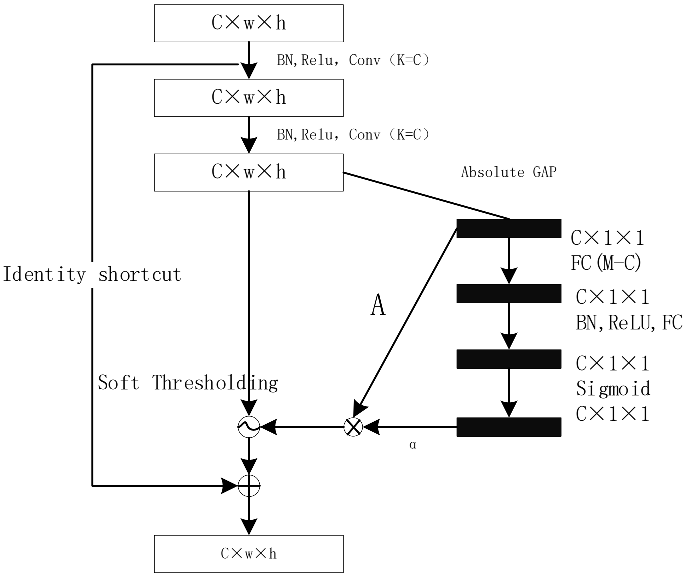


Figure 3 Subnetworks of the Deep Residual Shrinkage Network

After the convolution operation, features enter the residual shrinkage module. Specifically, one branch aggregates spatial features of various channels through a global average pooling operation, yielding the compressed feature A. Then, a set of weights α is calculated through two fully connected layers and is multiplied with each channel of feature A, obtaining the threshold. This threshold is then applied to feature A for soft thresholding, resulting in a new feature. Another path directly transfers the feature to the output location through a skip connection. Ultimately, the network mapping after soft thresholding and the identity mapping are weighted, forming the basic module of DRSN. The key technologies involved in DRSN will now be briefly outlined.

(1)Residual Module

Traditional convolutional neural networks often face the issue of network degradation, where network performance declines instead of improving as the number of layers in the network model increases. Residual networks can effectively solve this problem.

Compared to traditional convolution, the residual module adds a cross-layer identity path, which is composed of batch normalization layers, convolution layers, and activation function layers. Figure 4 shows a schematic diagram of the residual module. Within the residual module, identity mapping and residual mapping are used to build the network structure. Here, the residual mapping represents a mapping that processes the input to a certain extent, while the identity mapping indicates a mapping that does not transform the input at all. When the residual mapping is close to zero, the network performance is optimal. If the number of network layers is increased further, the network performance will remain at this optimal level.

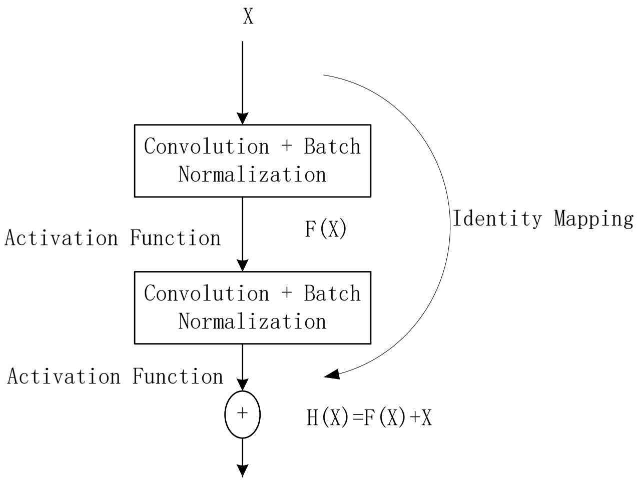


Figure 4 Diagram of Residual Block Structure

Let's assume the target optimal solution is $H(x)=x$. The residual mapping refers to the difference between the mappings of $H(x)$ and x, which can be represented as $F(x)$, i.e., $F(x)=H(x)-x$. When $F(x)$ approaches zero infinitely, the network performance reaches its best. When the input to the residual block is x_n, the computed output can be obtained as:

$$x_{n+1}=f(x_n+F(x_n,W_n)), \tag{1}$$

In the formula, $F(\cdot)$ represents the residual mapping, W_n is the corresponding weight parameters, and $f(\cdot)$ is the activation function. When the dimensions between different residual blocks do not match, it is necessary to perform a linear transformation W_s on the identity mapping x_n to obtain:

$$x_{n+1}=f(W_S x_n+F(x_n,W_n)), \tag{2}$$

In the formula, W_s represents the weight parameters.

(2)SE Module

In the process of image recognition, there are often noise and redundant information that can have adverse effects on the model's performance [12]. As a pioneering work of channel attention, SENet can evaluate the importance of each feature channel and adjust the weights of each channel accordingly, helping the model focus on beneficial information for the recognition task, reducing the weight of interfering factors, and enhancing the model's expressive capability [13].

The SE module mainly consists of two sub-modules: squeezing and excitation. Specifically, in the squeezing module, in order to highlight the important information associated with the task, the information of all channels is compressed. That is, the input feature map of size $W \times H \times C$ is compressed to output $1 \times 1 \times C$, thereby obtaining the overall information of C feature channels in this layer. The formula is as follows:

$$z_c = F_{sq}(u_c) = \frac{1}{WH} \sum_{i=1}^{W} \sum_{j=1}^{H} u_c(i, j) \tag{3}$$

In the formula, z_c represents the output feature map, $\mathrm{F}_{sq}()$ represents the compression operation function, u_c denotes the input feature map, W and H represent the height and width of the feature map, and (i, j) stands for the coordinates' position on the feature map.

In the excitation module, a fully connected layer with a dimension of $\frac{C}{r} \times C$ is used to reduce the dimension of the compressed global information. Here, r is a scaling parameter used to reduce the number of channels, thus reducing the amount of network parameters and computational complexity while activating through the ReLU function. This not only maintains the performance of the model but also improves the computational efficiency of the model. Then, the number of channels is restored by sending it to the second fully connected layer. Subsequently, the Sigmoid function is used to map the input to the range of 0 to 1, generating channel weights. The formula is as follows:

$$s = F_{ex}(z_c, W_i) = \delta(W_2 \sigma(W_1 z_c)) \tag{4}$$

In the formula, S represents the channel-wise weight adjustment parameters, $\mathrm{F}_{ex}()$ denotes the Excitation operation function, σ stands for the ReLU activation function, and δ refers to the Sigmoid function. Then, the feature map is adjusted. By multiplying the original feature map with the corresponding weight parameters, a new weighted feature map is obtained. This can emphasize feature maps containing more important information, ultimately enhancing the model's performance. The formula is as follows:

$$\widetilde{X_c} = F_{scale}(u_c, s_c) = u_c s_c \tag{5}$$

In the formula, $\widetilde{X_{\mathrm{c}}}$ represents the adjusted output features, $F_{scale}()$ denotes the recalibration of the feature map, and s_c represents the weight parameters of the C-th feature map.

(3)Soft Thresholding

Soft thresholding is an effective denoising method. This technique eliminates irrelevant features by removing those whose absolute values are below a certain predetermined threshold, and bringing the features whose absolute values exceed this threshold closer to zero. The choice of soft threshold can be adjusted according to specific circumstances to achieve the best denoising effect. The soft threshold is given by:

$$y = \begin{cases} x - r, x > r, \\ 0, -r \le x \le r, \\ x + r, x < r, \end{cases} \tag{6}$$

In the formula, r is the threshold and is a positive number. The derivative of the output with respect to the input is:

$$\frac{\partial y}{\partial x} = \begin{cases} 1, x > r, \\ 0, -r \le x \le r, \\ 1, x < -r, \end{cases} \tag{7}$$

2.3 Dual Transfer Learning

Transfer learning is a solution for similar problems across different domains. Compared to traditional deep learning, it can reduce the model's data requirements, avoid overfitting, and decrease computation time [14,15,16]. This paper uses a dual transfer learning training method to alleviate the issue of insufficient sample volume in the dataset. Firstly, the DRSN-18 model is trained on ImageNet, yielding a set of weight parameters. The I-DRSN model loaded with these parameters is then trained on the augmented Plant Village dataset, obtaining a second set of weight parameters. Finally, the model loaded with the new weights is trained on the apple leaf disease dataset, implementing the second transfer learning. The training process is shown in Figure 5.

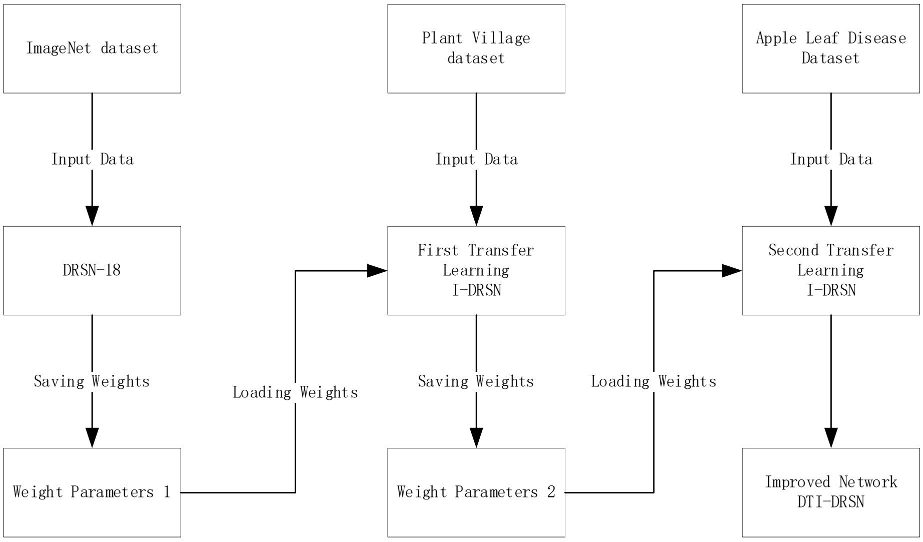


Figure5 Flowchart of Dual Transfer Learning

3. Experiment

3.1 Wild Plant Dataset and Its Preprocessing

The research data used in this paper comes from the Plant Village project and the 2020 Plant Pathology Challenge dataset [17], including 8674 images of apple leaf diseases. The research focuses on five conditions of apple leaves: rust, frog-eye leaf spot, black spot, powdery mildew, and healthy. The data is divided into a training set and a test set at a ratio of 4:1.

Given the impact of sample size on the effectiveness of classification experiments and the issue of data imbalance in the dataset, this study enhances the data to improve the generalization performance of the model. This is achieved by combining methods such as translation, flipping, shear transformation rotation, and adjusting contrast and brightness [18]. The distribution of the data after processing is shown in Table 2.

Table 2 Distribution of Apple Leaf Image Quantities

Types of Diseases	Number of Training Samples	Number of Test Samples
Healthy	1481	370
Rust Disease	1488	372
Black Spot Disease	1516	378
Powdery Mildew Disease	1896	472
Frog Eye Leaf Spot Disease	1508	377
Total	7889	1969

3.2 Experimental Environment and Parameter Settings

In this study, all experiments were conducted in a Windows 10 operating system environment, with the Intel(R)Xeon(R)W-2235 as the processor, the NVIDIA RTX2080Ti as the graphics processor, and using the Tensorflow 2.4.0 deep learning framework. In order to optimize the model's performance and training effect, it was found through comparison that the model performs optimally when the Adam optimizer is used with a learning rate of 0.0001. The number of iterations is set to 100.

3.3 Performance Evaluation Metrics

This study uses Accuracy (A), Precision (P), Recall (R), and the weighted F1 score as four evaluation metrics to assess the classification performance of the model. The calculation formulas are as follows:

$$A=\frac{T_{\mathrm{p}}+T_N}{T_P+T_N+F_N}\times 100\% \tag{8}$$

$$P=\frac{T_P}{T_P+F_P}\times 100\% \tag{9}$$

$$R=\frac{T_P}{T_P+F_N}\times 100\% \tag{10}$$

$$F_1=2\times\frac{P\times R}{P+R}\times 100\% \tag{11}$$

In the formula: T_P is the number of positive samples predicted to be positive; T_N is the number of negative samples predicted to be negative; F_P is the number of positive samples predicted to be negative; F_N is the number of negative samples predicted to be positive.

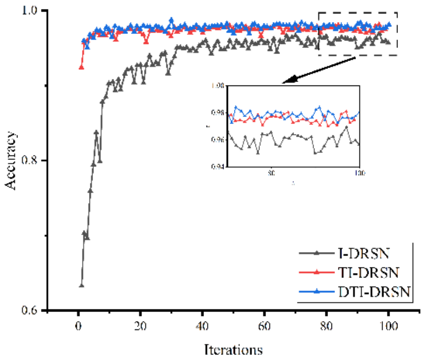
3.4 Experimental Results and Analysis

This paper conducted ablation experiments to validate the impact of the proposed improvements on the model's classification performance. Using the DRSN-18 network model as the baseline, it was observed through comparison that integrating the Inception module and adopting dual transfer learning resulted in accuracy improvements of 0.51% and 1.74%, respectively. Furthermore, improvements were also observed in precision, recall, and F1 score metrics. The introduction of the Inception module in the I-DRSN model with dual transfer learning led to a 2.33% increase in accuracy. Overall, compared to the baseline model, the proposed improved model exhibited better performance in terms of accuracy and F1 score, leading to an overall improvement in performance.

To evaluate the performance of the models in recognizing different categories of images, we compared the F1 scores of each model in recognizing different categories of images. The experimental results demonstrate that DTI-DRSN exhibits good recognition capabilities for all categories of images.

3.5 Impact of Dual Transfer Learning on the Experiment

To evaluate the impact of the dual transfer learning strategy on model performance, this paper compared the model based on the I-DRSN model with the Inception module using the dual transfer learning training method to network models that underwent 0 and 1 transfer learning iterations. Observing Figures 6 and 7, it is evident that the use of single transfer learning positively affects the network model in terms of convergence speed, fitting performance, accuracy, and model loss. However, the adoption of dual transfer learning further improves the performance of the network model in these aspects. This confirms that dual transfer learning indeed enhances the model's recognition accuracy and improves its overall performance.



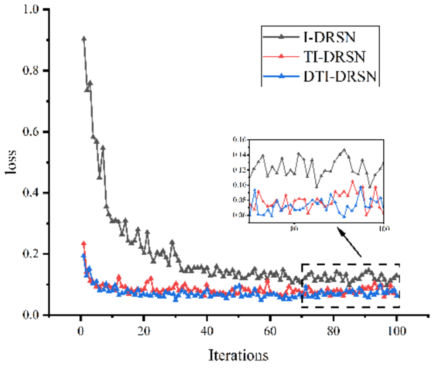


Figure 6 Comparison of Accuracy Curves Figure 7 Comparison of Loss Curves

3.6 Impact of Multiscale Structures on the Experiment

As shown in Table 3, the I-DRSN model, which includes the Inception module, performs better than the DRSN-18 model. The key factors include:The Inception module uses its ability to capture image features on different spatial scales, synchronously extracting multi-scale features of the image, which enhances the receptive field of the model and optimizes the network model's performance when dealing with disease spot features from different times and regions. This allows the model to acquire feature information from more dimensions, improving the accuracy of the model's classification.

Table 3 Comparative Results of Ablation Experiments

Model	Inception Module	DTL	Accuracy	Average Precision	Average Recall	Weighted F1 Score
DRSN-18	—	—	0.9647	0.9640	0.9643	0.9638
I-DRSN	√	—	0.9698	0.9701	0.9701	0.9693
DT-DRSN	—	√	0.9821	0.9818	0.9820	0.9818
DTI-RSN	√	√	0.9880	0.9876	0.9872	0.9872

Table 4 Comparison of F1 Scores for Different Classes Across Models

Model	Healthy	Rust Disease	Powdery Mildew Disease	Black Spot Disease	Frog Eye Leaf Spot Disease
DRSN-18	0.9502	0.9736	0.9824	0.9430	0.9740
I-DRSN	0.9508	0.9819	0.9836	0.9562	0.9779
DT-DRSN	0.9736	0.9883	0.9924	0.9801	0.9781
DTI-DRSN	0.9843	0.9896	0.9942	0.9901	0.9901

3.7 Comparative Experiments with Other Network Models

This paper's improved model is compared with five classic networks (ResNet18, VGG16, AlexNet, MobileNetV2, GoogLeNet) [18-21]. The test accuracy curves of each model are shown in Figure 8. Compared with other models, the recognition accuracy of the improved model proposed in this paper is higher. This can be mainly attributed to the following three aspects:The Inception module extracts disease spot features of different scales, enhancing the model's receptive field.Double transfer learning helps reduce the model's dependence on data, maintaining accurate and efficient model performance even when data volume is limited.The Deep Residual Shrinkage Network includes the SENet module, which reduces the risk of the model being affected by irrelevant information, and enhances the expressive power of the network model.

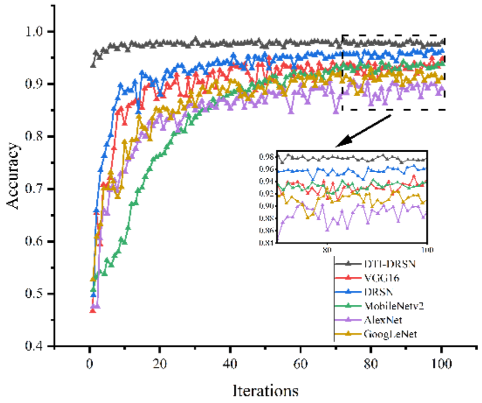


Figure 8 Accuracy Curves of Different Models on the Test Set

4. Summary

An improved Deep Residual Shrinkage Network model for apple leaf disease detection is proposed in this paper. By integrating the Inception module, the model's receptive field is enhanced, thereby improving the model's recognition accuracy. The double transfer learning method is used to reduce the model's dependence on data and accelerate the convergence speed of the model. The experimental results confirm that the proposed improved model performs well overall, providing a valuable reference for further research on intelligent plant disease leaf recognition.

References

[1] HU Qingyu, HU Tongle,WANG Yanan,et al.Survey on the occurrence and distribution of apple diseases in China[J].Plant Protection, 2016,42(1):175-179.

[2] HE Zifen, HUANG Junxuan,LIU Qiang,et al.High precision identification of apple leaf diseases based on asymmetric shuffle convolution[J].Transactions of the Chinese Society for Agricultural Machinery,2021,52(8):221-230.

[3] LIU Bin,JIA Runchang,ZHU Xianyu,et al. Light weight identification model for apple leaf diseases and pests based on mobile terminals[J].Transactions of the CSAE, 2022,38(6):130- 139.(in Chinese)

[4] LECUN Y, BOSER B, DENKER J S, et al. Backpropagation applied to handwritten zip code recognition[J]. Neural Computation, 1989, 1(4) : 541 - 551.

[5] Fan, Q., Zhu, G., & Huang, W., et al. (2021). Research on visual tracking strategy of execution end based on image segmentation model. Journal of Instrument and Equipment, 42(9), 62-70.

[6] Liu, K., Dong, M., & Wang, P., et al. (2022). A method of traffic light recognition based on image enhancement. Electronic Measurement Technology, 45(7), 137-145.

[7] Lu, X., Cao, Y., & Zhou, X., et al. (2021). Two-stage salient object detection based on deep reinforcement learning. Journal of Electronic Measurement and Instrumentation, 35(6), 34-4.

[8] Meng, L., Guo, X., & Du, J., et al. (2021). A lightweight CNN crop disease image recognition model. Jiangsu Agricultural Journal, 37(5), 1143-1150.

[9] YU H,CHENG X,CHEN C, et al. Apple leaf disease recognition method with improved residual network[J].Multimedia Tools and Applications,2022,81(6):7759-7782.

[10] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions[C]∥Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 1 - 9.

[11] ZHAO M, ZHONG S, FU X, et al. Deep Residual Shrinkage Networks for Fault Diagnosis [J]. IEEE Transactions on Industrial Informatics, 2020, 16(7) : 4681-4690.

[12] ZHU H Y, XIE C, FEI Y Q, et al. Attention mechanisms in cnn-based single image super-resolution :a brief review and a new perspective[J].Electronics, 2021, 10(10):1187--11188.

[13] Liu, X., Li, Y., & Liu, L., et al. (2019). Improved YOLOV3 target recognition algorithm embedded with SE-Net structure. Computer Engineering, 45(11), 243-2248.

[14] Zhao, L., Hou, F., & Lv, Z., et al. (2020). Cotton leaf pest and disease image recognition based on transfer learning. Journal of Agricultural Engineering, 36(7), 184-191.

[15] Yu, X., Yang, M., & Zhang, H., et al. (2020). Research and application of crop pest and disease detection methods based on transfer learning. Journal of Agricultural Machinery, 51(10), 252-25.

[16] THAPA R,ZHAHNG K,SNAVELY N,et al. The Plant Pathology Challenge 2020 data set to classify foliar disease of apples[J]. Applications in Plant Sciences,2020,8(9):e11390.

[17] Jiang, Y., Zhang, H., & Chen, L., et al. (2019). Image data augmentation algorithm based on convolutional neural network. Computer Engineering and Science, 41(11), 2007-2016.

[18] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolution neural networks[J].Advances in Neural Information Processing Systems.2012,25.

[19] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. Computer Science ,2014, arXiv;1409.1556.

[20] SZEGEDY C, LIU W, JIA Y ,et al. Going deeper with convolutions[C].Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.2015:1-9.

[21] SANDLER M, HOWARD A,ZHU M, et al.Mo-bilenetv2: Inverted residual and linear bottlenecks[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018;4510-4520.