IDENTIFICATION OF APPLE LEAF DISEASES BASED ON IMPROVED CONVOLUTIONAL NEURAL NETWORK
/
基于改进卷积神经网络的田间苹果叶片病害识别

Lili LI *1), Bin WANG , Zhiwei HU [1]
College of Information Science and Engineering, Shanxi Agricultural University, Taigu 030800/China
Tel: 18735431598; E-mail: wangbin1759@126.com
DOI: https://doi.org/10.35633/inmateh-67-54

Keywords: *attention mechanism, convolutional neural network, apple leaf disease, disease diagnosis*

ABSTRACT
In view of the obvious differences in the manifestations of the same diseases in apples at different stages of the disease, different diseases show certain similarities, and the early symptoms of the disease are not obvious. For these problems, a new model attention residual network (ARNet) was introduced based on the combination of attention and residual thought. The model introduces the multi-layer attention modules to solve the problems of early disease location dispersion and features that are difficult to extract. In order to avoid network degradation, a residual module was constructed to effectively integrate high and low-level features, and data augment technology was introduced to prevent the model from over-fitting. The proposed model (ARNet) achieved an average accuracy of 99.49% on the test set of 4 kinds of apple leaf diseases with real complex backgrounds. Compared with the models ResNet50 (99.19%) and MobileNetV2 (98.17%), it had better classification performance. The model proposed in this paper had strong robustness and high stability and can provide a reference for the intelligent diagnosis of apple leaf diseases in practical applications.

摘要
苹果同种病害在不同发病阶段表征差异明显，不同病害又表现出一定的相似性，且在病害早期症状不明显。针对该问题，在ResNet的基础上引入注意力机制提出一种新的模型（ARNet）。通过引入多层注意力模块，层次化提取病害分类信息，解决早期病害部位分散、特征难以提取等问题，为避免网络训练出现退化现象，构建残差模块有效融合高低阶特征，同时引入数据扩充技术以防止模型过拟合。研究表明，提出的模型(ARNet)在4种具有真实复杂背景的苹果叶病害测试集上的平均准确率达到99.49%，与现有模型ResNet50（99.19%）和MobileNetV2（98.17%）相比，ARNet具有更好的分类效果。本文提出的模型具有较强鲁棒性和较高稳定性，在实际应用中可为苹果病害智能诊断提供参考。

INTRODUCTION
With a high nutritional and medicinal value, the apple is one of the most productive types of fruit in the world. However, various diseases occur frequently on a large scale in apple production, such as apple Alternaria leaf blotch and apple rust, which affect the quality of fruits and thereby causing substantial economic losses.

We know that many diseases of apples start from the leaves and then spread to the entire plant, so timely and accurate identification of the types of leaf diseases is the key to disease prevention. In recent years, image-based pattern recognition technology has been widely used in the field of crop disease diagnosis *(LeCun et al., 2015)*. The traditional algorithm for crop disease identification is carried out by extracting and screening features such as color, texture, shape, etc. However, the symptoms of the same disease at different stages of the disease are obviously different, and multiple diseases may show similar pathological features. These complex natural feature factors make the traditional pattern recognition method less universal in solving the problem of apple disease recognition. Image-based disease recognition is essentially an image classification problem, and the application of deep convolutional neural networks in the field of image recognition is a research hotspot *(Szegedy et al., 2015; John et al., 2021; Mohanty et al., 2016; Yang et al., 2017)*.

[1] *Lili Li, Lec. Ph.D. Stud.; Bin Wang*, Lec. Ph.D. Eng.; Zhiwei Hu, Lec. Ph.D. Stud.*

Sun et al. (2017) improved the AlexNet network by using methods such as Batch Normalization, Global Average Pooling, and reducing the size of the convolution kernel, which improved the accuracy and reduced the parameters required for the model. The average accuracy of identifying 26 diseases of 14 crops in a simple background is 99.56%, but the recognition rate of complex backgrounds in the field is low. *Wang et al. (2017)* applied transfer learning to the AlexNet network, and the average recognition accuracy of 10 categories of tomato leaves reached 95.62%. Rangarajan et al. *(2018)* used the original AlexNet and VGG 16 models, combined with transfer learning to obtain 97.29% and 97.49% accuracy in the recognition of 7 kinds of segmented tomato leaves. *Hu et al. (2019)* and *Zeng et al. (2018)* further introduced attention mechanisms, high-order residuals, and parameter sharing mechanisms on the residual network to improve the fine-grained level, accuracy, and robustness of disease recognition. *Bi et al. (2020)* proposed an apple leaf disease identification method based on MobileNet, and the accuracy achieved 73.5% for 164 Alternaria leaf spots and 170 rust. Due to the deficiency of AlexNet, *Guo et al. (2019)* combined with the characteristics of tomato diseased leaf images, designed a multi-receptive field recognition model (Multi-Scale AlexNet) based on AlexNet and developed a tomato leaf disease image recognition system using this model based on Android. The model has an average recognition accuracy of 92.7% for tomato leaf diseases (each disease in the early, middle, and late stages). The system had a recognition rate of 89.2% in the field. *Wang et al. (2020)* proposed an improved version (Multi-scale ResNet) based on ResNet18 by adding a multi-scale feature extraction module, changing the connection mode of the residual layer, decomposing the large convolution kernel, and performing group convolution operations.

Due to the complex and diverse symptoms of apple disease, the size of the lesions is different, and the color and location of the disease are different. For example, in the early stage of apple rust, the lesions are lighter in color and smaller in area, and in the later stage, the lesions are larger and darker in color. As a result, it is difficult to accurately extract disease features when identifying diseases.

Compared with the previous methods for processing all information, the attention mechanism will only focus on some significant or interesting information, and assign different weights to different information, thereby filtering out unimportant information, improving processing efficiency and model effects, the attention mechanism is already in the field of natural language processing *(Vaswani et al., 2017; Letarte et al., 2018)* and image segmentation *(Yu et al., 2018)* had made great progress, but it is still in the exploratory stage in the field of image classification. Other authors proposed a recurrent attention model (Recurrent attention model RAM) based on RNN (Recurrent neural network) and applied the attention mechanism to image classification tasks (https://arxiv.org/pdf/1406.6247). *Wang et al. (2019)* proposed to introduce attention-based residual learning into the field of image classification, which solved the problem of the inability to extract attention in the forward process, simplified the model structure, and accelerated the model training. *Fu et al. (2017)* used the attention mechanism in the task of fine-grained image classification, using the output of the previous layer of each layer as the input of the current layer, thus forming an inter-scale loop, which can fully learn fine-grained feature information. The attention mechanism has been proven to be able to effectively extract fine-grained feature information in the field of image classification. Besides current advances, the disease classification problem is far from being solved. The extensive work from the author *(*Barbedo *et al., 2016)* analyzes its current challenges in deep detail. These challenges comprise 1) The presence of multiple simultaneous disorders on a plant. 2) The existence of different disorders that present similar visual symptoms. 3) The high variability of symptoms for a specific disorder.

Aiming at the problem of apple leaf disease in the complex background and the presence of multiple simultaneous disorders on a leaf. This paper innovatively proposed a new model (ARNet) based on the combination of attention and residual mechanism to identify four kinds of apple leaf diseases (including multiple diseases on the same leaf) under complex background. In order to provide technical support for the apple leaf disease intelligent diagnosis system.

MATERIALS AND METHODS

Dataset

One part of the dataset used comes from the ‘Plant Pathology Challenge’ for CVPR 2020-FGVC7 (https://www.kaggle.com/c/plantpathology-2020 fgvc7) and the other part is self-acquired in the field. The former contains 1821 RGB images including 592 apple scabs, 622 apple rusts, 91 complex diseases (a picture contains multiple diseases), and 516 healthy leaves.

The latter was taken on the spot with a mobile phone at the Pomology Institute of Shanxi Agricultural University. After screening, 100 healthy apple leaves and 50 leaves with rust and scabs were obtained (no complex disease data was obtained).

A total of 2021 apple diseases were obtained in the dataset. The images after cropping with a rectangle containing leaf named as dataset 1, and the original images collected without any processing named as dataset 2. Partial sample images in dataset 2 are shown in Fig. 1 and partial sample images in dataset 1 and dataset 2 are shown in Fig. 2.

a) Healthy *b) Multiple-diseases* *c) Rust* *d) Scab*

Fig. 1 - Sample image of the dataset

a) the rust in dataset 1 *b) the rust in dataset 2*

Fig. 2 - Two datasets of samples

Image Preprocessing

Image Preprocessing

The preprocessing of the picture includes redefining the image size, image de-averaging, and normalizing each batch of samples, etc. This experiment redefined the image as 224×224×3, which reduced the pixel value and removed a lot of redundant information, thereby reducing the amount of calculation. De-averaging the image is to standardize the image and remove the average brightness value of the image. The specific operation is as follows: each pixel value in the image subtracts the average value of the image pixel. In many cases, it is not sensitive to the brightness of the image, and more attention is paid to its content. In the image classification task in this experiment, the overall brightness of the image does not affect what objects exist in the image. At this time, it makes sense to remove the mean value of pixels for each data point. For CNN, the normalization of data is required for gradient descent. After data normalization, gradient explosion can also be prevented, so as to accelerate network convergence and further reduce the number of feature maps. The batch Normalization method is adopted in this experiment, and each sample image in the batch needs to be normalized. The processing method for a sample is shown in Equations (1) and (2):

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{1}$$

$$\sigma = \frac{1}{n}\sum_{i=1}^{n} (x_i - \mu)^2 \tag{2}$$

Where x_i represent the value of the i-th pixel of the sample, n is the total number of pixels of the sample, μ represents the mean value, and σ represents the variance. The normalization formula is shown in formula (3).

$$\hat{x}_i = \frac{x_i - \mu}{\sqrt{\sigma^2 + \varepsilon}} \tag{3}$$

Where $\hat{x}_i$ is the normalized of the i-th pixel value of the sample, ε is a tiny value greater than 0 to ensure that the denominator is greater than 0.

Image Augmentation

The image enhancement must meet the actual application conditions and cannot be amplified at will (for example, color is one of the main signs of different diseases, and the color of the original image cannot be changed when doing image enhancement). In order to simulate the impact of the complex environment under natural shooting conditions and enrich the dataset, the following data enhancement methods are mainly adopted: random rotation angle is performed on datasets 1 and dataset 2 (the maximum rotation angle is 30), random horizontal or vertical translation (the maximum distance of translation is 0.1×image width or height) and random horizontal flipping, and by introducing slight disturbances to achieve data expansion. After the data enhancement operation, the missing part of the pixel is completed by filling the adjacent pixel. In deep learning, uneven sample distribution will affect the accuracy of model recognition. The enhanced sample dataset is 2960 apple scabs, 3095 apple rusts, 2275 complex diseases, and 2580 healthy samples. According to the ratio of 8:1:1, it is randomly divided into the training set, validation set, and test set. Among them, the training set contains 8728 images, the validation set contains 1091 images, and the test set contains 1091 images.

Convolutional Neural Network Models

ARNet *(Hu et al. 2019)* was proposed based on the combination of attention mechanism and residual network ResNet *(He et al., 2016)*, which was used for disease identification of apple leaves. The model structure is shown in **Error! Reference source not found.**.

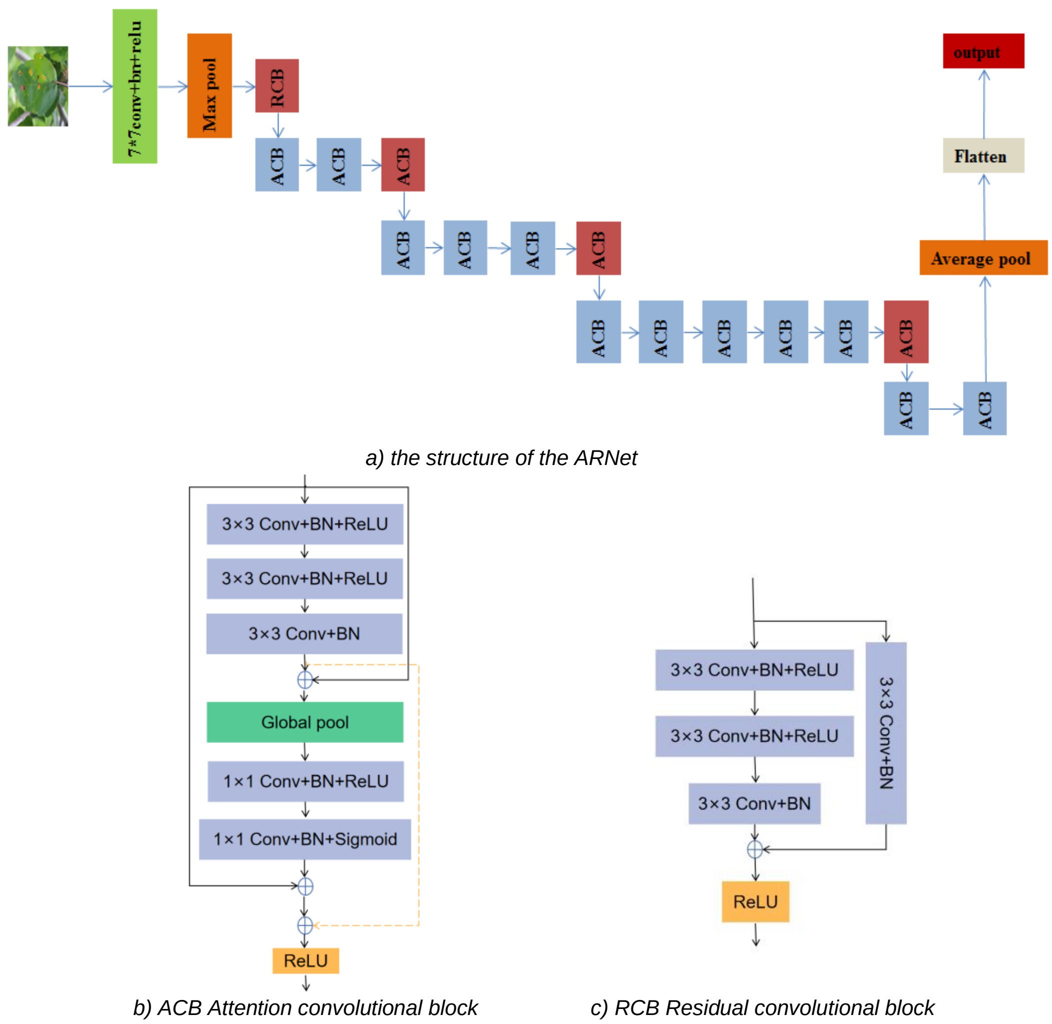


Fig. 3 - The structure of the ARNet

It consists of attention convolutional block ACB and residual convolutional block RCB. The model contained a total of 5 layers of which the 2-5 layers stack with the above two types of modules, and use the output of the previous layer as the input of the next layer. Among them, the deep network structure was constructed through the interconnection between ACB blocks in the layers to extract multi-dimensional image feature information, and the output of the last ACB module was used as the input of the RCB module.

Among the layers, the output of the ACB block is spliced through the output of the RCB block to reduce the size of the feature map. Finally, the model used a 2-layer fully connected network to classify the input of diseased images. The number of convolutional channels in the second layer of ARNet was 256. And as the number of network layers increases, the number of convolutional channels doubles until the fifth layer reached 2048.

Performance Evaluation Metrics

There are multiple evaluation indicators for evaluating the performance of a disease recognition model. Different evaluation indicators can be used to evaluate the model recognition effect from different perspectives. This study uses the following two evaluation methods to evaluate the performance of the network model.

Average Accuracy (AA)

AA refers to the ratio of the predicted correct sample to all observations in the model, and is the most important indicator for testing model performance and is calculated as in Equation (4):

$$\mathrm{AA} = \frac{1}{n_s}\sum_{i=1}^{n_s}\frac{n_{ii}}{n_i} \tag{4}$$

Where n_S is the total number of sample categories (for this article the $n_s = 4$, $i \in [1,4]$ is the sample category label, n_i represents the total number of samples of the i-th type, n_{ii} represents the total number of samples whose true class is i and the predicted class is also i. (that is, the number of correct predictions in each type of sample).

Highest accuracy

The difference between the highest recognition accuracy (%) of a model and the average recognition accuracy (%) is that it can measure the best performance of a model in order to save the model parameters when the model reached the best performance.

Parameter Setting

This paper divided the dataset into three parts: training set, validation set, and test set. In order to avoid memory overflow, the batch training method was adopted. The ARNet, MobileNetV2, and ResNet 50 were used in the training set and validation set for comparative experiments. Each batch was trained with 16 images, and the batch size of the validation set was the same as the batch size of the training set. This paper set the number of iterations to 50 and used the Categorical-cross entropy in Keras as the cost function. In order to solve the problem of gradient disappearance and explosion in the backpropagation process, Batch Normalization was introduced to standardize the input of the hidden layer of the network. In order to improve the efficiency of tuning, adopt the adaptive matrix estimation algorithm Adam optimization algorithm, and its initial learning rate was set to 0.0005. In order to automatically modify the learning rate, the learning rate scheduler in Keras (Reduce LR on the plateau) was introduced. This article set the learning rate to be reduced to 0.9 of the current learning rate when the loss function value of the validation set did not decrease during the 5 iterations. The value of dropout was 0.5. In order to save the optimal model parameters, the current model was decided to be saved by observing whether the value of the validation set loss function decreases, after each iteration. Finally, the saved model structure and parameters were used to predict the disease images of the test set.

Training methods

The three models used the same experiment parameters setting, and the difference was the training method. ARNet model used Glorot uniform distribution to initialize the weight parameters. In order to speed up the convergence of the network, ResNet50 and MobileNet adopt two training methods of transfer learning. The transfer learning 1 was the weights of different deep learning network models that were initialized by using the model parameter files pre-trained on the ImageNet dataset, instead of the original random initialization operation, and the global fine-tuning was carried out. The transfer learning 2 was to keep the underlying parameters of the network model unchanged and only fine-tune its full-connection layer parameters.

RESULTS AND DISCUSSION

The results of the models

Table 1 summarized the highest recognition accuracy on the validation set and average recognition accuracy of the test set achieved by the 3 models on the two datasets.

Table 1

Results of datasets 1 and 2 by different models

Model	Training methods	Highest recognition accuracy of the validation set (%)		Average recognition accuracy of the test set (%)	
		Dataset 1	Dataset 2	Dataset 1	Dataset 2
ResNet50	Transfer learning 1	99.77	99.81	99.13	99.19
	Transfer learning 2	93.11	84.72	92.96	84.76
MobileNet	Transfer learning 1	99.70	99.53	98.91	98.179
	Transfer learning 2	90.167	81.45	90.59	79.679
ARNet	Training from scratch	99.50	99.63	99.35	99.49

From the contrast of the two training methods of MobileNetV2 and ResNet50 in Table 1, it was found that for dataset 1 the highest accuracy on the test set using transfer learning 1 was (99.13%), 8.54% higher than transfer learning 2 (90.59%), and the difference was 19.5% for dataset 2. It can be concluded that for the models, fine-tuning all the parameters of a pre-trained neural network architecture achieved higher classification accuracy as compared to using the neural network architecture with feature extraction only. The reason was that the image features can be extracted by the convolution layer module with the weight of the transfer parameters, but there are great differences between the apple leaf disease images in this study and ImageNet. Only training and changing the full connection module cannot achieve the ideal effect, and transfer learning training in all layers can significantly improve the accuracy of the test set. It showed that the training classification layers alone cannot make the model adapt to our data well.

From Table 1, it can be seen that the recognition effect of the ARNet model was better than the other two models. The highest accuracy of the ARNet on the test set of dataset 1 (99.35%) and dataset 2 (99.49%) with little difference, and MobileNet and ResNet get the same result also. It indicated that the clipping operation cannot well improve the performance of the model. Since dataset 2 can better reflect the authenticity and diversity of the field disease images, the performance of the classification model on dataset 2 will be studied later.

The classification accuracy of the Model

Table 1 showed the prediction accuracy of each category in the test set. The results showed that: 1) The ARNet model proposed can achieve the best classification accuracy for all four apple leaf diseases. Compared with the other two models, the classification accuracy of the ARNet model had increased by 0.30% and 1.31%, respectively. Specifically, the classification accuracy for multiple diseases, scabs, and healthy was higher than ResNet50 and MobileNet models. 2) The other two models (ResNet50, MobileNet) had the worst predictions for multiple diseases, for the reason that some multiple diseases had obvious symptoms of one disease while the symptoms of the second disease were not obvious, leading to misjudgment of single diseases. However, ARNet can capture fine-grained information very well. This was because even if the early lesions of apple leaf diseases are concealed or there are unobvious symptoms, the ACB module in the ARNet model can still hierarchically self-learn attention information during training, and give higher weight to the effective disease feature area, while the RCB module can obtain low-level information. 3) The obtained disease features were effectively transmitted to the high levels through the jumping mechanism, and the high-levels were guided to perform feature selection, and then the subtle key disease features were fully extracted, which greatly improved the recognition ability of the model when the disease features were not obvious. As shown in Fig. 4, the leaf had multiple diseases, but the symptom of rust was obvious and the symptom of other diseases was not obvious, so it is easy to be misjudged as a single rust disease.

Table 2

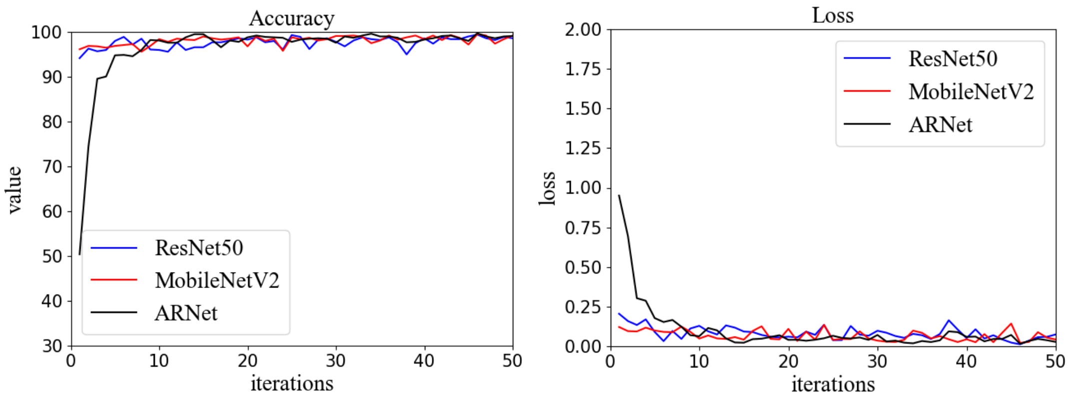
The accuracy of models on dataset 2

Type of diseases	ARNet (%)	ResNet50 (%)	MobileNet (%)
Rust	98.25	99.65	99.65
Scab	100	99.25	97.37
Multiple-diseases	100	98.03	96.55
Healthy	100	99.56	98.68
Average	99.49	99.19	98.17

Fig. 4 – Multiple-diseases

Model stability

As shown in Fig. 5, the three models had reached the convergence state, but different models showed different performances during the iteration. ResNet50 and MobileNetV2 had high accuracy in the initial stage because of the use of initial weights pre-trained on the ImageNet dataset. ARNet had low accuracy at the beginning for the reason that it used the randomly generated initial weight. For ARNet and MobileNetV2 with the number of iterations increased, the accuracy curve showed an upward trend, and the shaking phenomenon was not obvious, indicating that the ARNet and the MobileNetV2, compared with the ResNet50, had higher stability and can steadily improve the accuracy of the model.

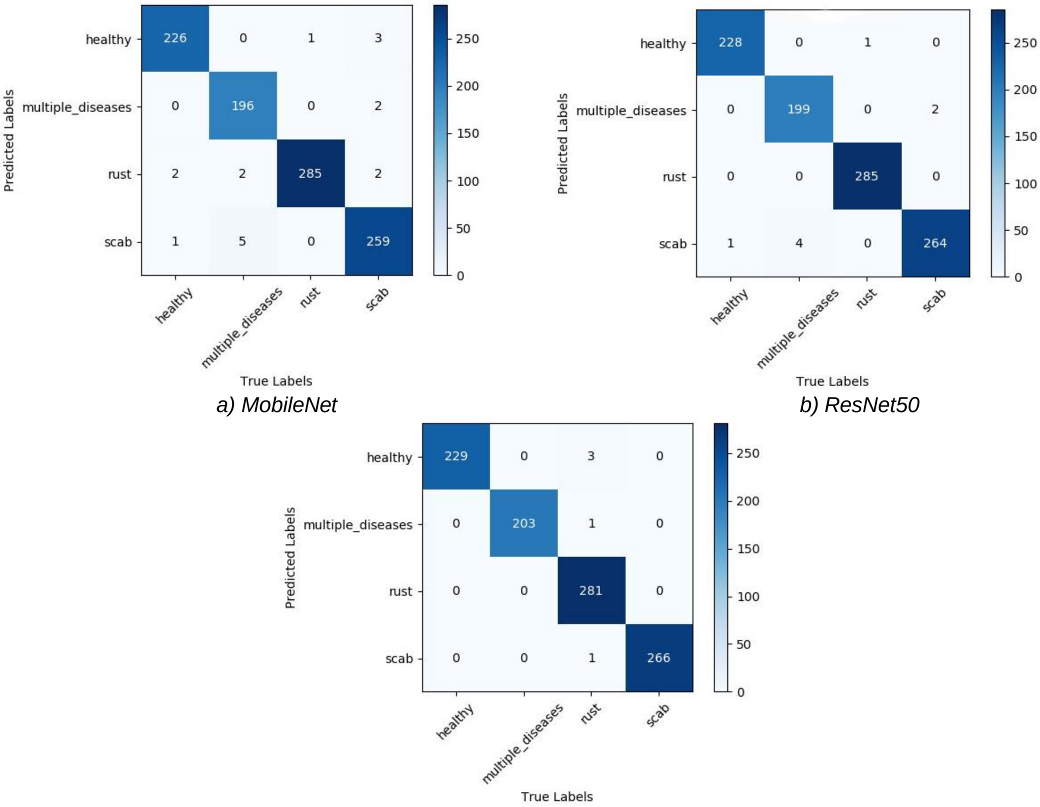


a) accuracy curve *b) loss function curve*

Fig. 2 - The accuracy and loss function value of different models

The results

In order to further observe the performance of the model, visualize the confusion matrix of the three models (MobileNet, ResNet50, and ARNet) on the test set. As depicted in Fig. 6, each column in the figure represents the predicted labels, and its total number indicates the number of samples predicted for that category. Each row represents the real labels, and the total number indicates the number of samples of the true data belonging to the category. The value at the intersection of the row and column represents the number of data predicted as the corresponding row category, and the sum of values in the diagonal line was the predicted correct result. The results showed that the three models differ greatly in the classification of different types of diseases. MobileNet and ResNet50 were better than ARNet for the detection of rust, but they were not as effective as ARNet for the other three kinds of diseases. The overall prediction effect of ARNet was the most ideal, with an accuracy of 99.49%.



a) MobileNet *b) ResNet50*

c) ARNet

Fig. 3 - The confusion matrix of the three models

CONCLUSIONS

In this study, the ARNet based on the combination of attention and the residual mechanism was proposed to solve the problems of concealment, small area, and unobvious symptoms of the early lesions of apple leaf diseases. The results showed that the image clipping operation cannot well improve the performance of the models. And the proposed ARNet model had the best performance, with an accuracy rate of 99.49%, which was higher than the other two models MobileNetV2 (98.17%) and ResNet50 (99.19%). This was because even if the early lesions of apple leaf diseases were concealed or there were unobvious symptoms, the ACB module in the ARNet model can still hierarchically self-learn attention information during training, and give higher weight to the effective disease feature area, while the RCB module can obtain low-level information. Thus, this work can provide a reference for the timely prevention and control of apple leaf diseases.

In the next step, the proposed model will be deployed to smartphones and optimization techniques will be proposed to improve model performance and realize agricultural modernization and intelligence.

ACKNOWLEDGEMENT

This work was supported by the Innovation Project of Shanxi for Postgraduate Education, China (2020BY060); "Teaching and Research Management" Fund Project of Shanxi Agricultural University, China (XK19207)

REFERENCES

[1] Barbedo, J. G. A. (2016). A review on the main challenges in automatic plant disease identification based on visible range images. *Biosystems Engineering*, *144*, 52-60. https://doi.org/10.1016/j.biosystemseng.2016.01.017

[2] Bi, C., Wang, J., Duan, Y., Fu, B., Kang, J. R., & Shi, Y. (2020). MobileNet based apple leaf diseases identification. *Mobile Networks and Applications*, 1-9. https://doi.org/10.1007/978-3-319-46493-0-38

[3] Fu, J., Zheng, H., & Mei, T. (2017). Look closer to see better: Recurrent attention convolutional neural network for fine-grained image recognition. In Proceedings of the *30th IEEE Conference on Computer Vision and Pattern Recognition*, Hawaii, USA. https://openaccess.thecvf.com/content_cvpr_2017/html

[4] Guo, X. Q., Fan, T. J., & Shu, X. (2019). Tomato leaf diseases recognition based on improved multi-scale AlexNet. *Transactions of the Chinese Society of Agricultural Engineering*, *35*(13), 162-169. https://doi.org/10.11975/j.issn.1002-6819.2019.13.018

[5] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the 2016 *29th IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, USA. https://openaccess.thecvf.com/content-cvpr-2016/html

[6] He, K.M., Zhang, X.Y., Ren, S.Q., & Sun, J. (2016). Identity mappings in deep residual networks. In Proceedings of the 2016 *16th European Conference on Computer Vision*, Amsterdam, Netherlands. https://doi.org/10.1007/978-3-319-46493-0-38

[7] Hu, Z., Yang, H., Huang, J., & Xie, Q. (2019). Fine-grained tomato disease recognition based on attention residual mechanism. *Journal of South China Agricultural University*, *40*(6), 124-132. https://doi.org/10.7671/j.issn.1001-411X.201812048

[8] John, S., Rose, A. L. (2021). Machine learning techniques in plant disease detection and classification-a state of the art. *INMATEH-Agricultural Engineering*, 65(3), 362-372. https://doi.org/10.35633/inmateh-65-38

[9] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436-444. https://doi.org/10.1038/nature14539

[10] Letarte, G., Paradis, F., Giguère, P., & Laviolette, F. (2018). Importance of self-attention for sentiment analysis. In Proceedings of the 2018 *EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*, Brussels, Belgium. https://aclanthology.org/W18-5429

[11] Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). Using deep learning for image-based plant disease detection. *Frontiers in plant science*, *7*, 1419. https://doi.org/10.3389/fpls.2016.01419

[12] Rangarajan, A. K., Purushothaman, R., & Ramesh, A. (2018). Tomato crop disease classification using pre-trained deep learning algorithm. *Procedia computer science*, *133*, 1040-1047. https://doi.org/10.1016/j.procs.2018.07.070

[13] Sun, J., Tan, W., Mao, H., Wu, X., Chen, Y., & Wang, L. (2017). Recognition of multiple plant leaf diseases based on improved convolutional neural network. *Transactions of the Chinese Society of Agricultural Engineering*, *33*(19), 209-215. https://doi.org/10.11975/j.issn.1002-6819.2017.19.027

[14] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Boston, USA.

[15] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N.; Kaiser, L., & Polosukhin, I. (2017). In Proceedings of the 2017 *31st Conference on Neural Information Processing Systems*, Long Beach, CA, USA. https://arxiv.org/abs/1706.03762

[16] Wang, C. S., Zhou, J., Wu, H. R., Teng, G. F., Zhao, C. J., & Li, J. X. (2020). Identification of vegetable leaf diseases based on improved multi-scale ResNet. *Transactions of the Chinese Society of Agricultural Engineering*, *36*(20), 209-217. https://doi.org/10.11975/j.issn.1002-6819.2020.20.025

[17] Wang, F., Jiang, M., Qian, C., Li, C., Zhang, H., & Tang, X. (2017). Residual attention network for image classification. In Proceedings of the *30th IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, Hawaii, USA. https://openaccess.thecvf.com/content_cvpr_2017/html

[18] Wang, Y., Zhang, H., Liu, Q., & Zhang, Y. (2019). Image classification of tomato leaf diseases based on transfer learning. *Journal of China Agricultural University*, 24(6), 124-130. https://doi.org/10.11841/j.issn.1007-4333.2019.06.14

[19] Yang, G., Bao, Y., & Liu, Z. (2017). Localization and recognition of pests in tea plantation based on image saliency analysis and convolutional neural network. *Transactions of the Chinese Society of Agricultural Engineering*, *33*(6), 156-162. https://doi.org/10.11975/j.issn.1002-6819.2017.06.020

[20] Yu, C.; Wang, J.; Peng, C.; Gao, C.; Yu, G.; & Sang, N. (2018). Bisenet: Bilateral segmentation network for real-time semantic segmentation. In Proceedings of the *18th European Conference on Computer Vision*, Munich, Germany. https://openaccess.thecvf.com/content_ECCV_2018/html

[21] Zeng W. H. (2018). Research on Robust Recognition of Crop Leaf Diseases based on Deep Convolution Neural Network. University of Science and Technology of China: Hefei, China.