Deep Learning Based Classification for Tomato Diseases Recognition

Tao Zhang¹, Xiankun Zhu¹, Yiqing Liu¹, Kun Zhang¹, Azhar Imran¹

¹School of Software Engineering, Beijing University of Technology, Beijing, 100124, China

*Corresponding author’s e-mail:iizxkii@qq.com

Abstract. The rapid and accurate identification of tomato diseases is the basis of crop disease control. In order to achieve accurate identification of tomato diseases, this paper first explores the impact of model depth and the presence or absence of mixup data enhancement on the ResNet model. Experimental results show that using the data set enhanced by mixup can effectively make the model more robust. At the same time, ResNet34 depth model recognition accuracy is higher. Considering the differences in classification accuracy and calculation speed between ResNet and SE-ResNet, the paper chooses the SE-ResNet model as the basis of the network structure. We tuned the model and built a SE-ResNet network that is more suitable for tomato disease identification. The experimental results show that the accuracy of training SE-ResNet using datasets such as mixup is 88.83%. Our model can effectively identify various tomato diseases and the severity of each tomato disease.

1. Introduction

Disease is one of the biggest factors affecting the sustainable and stable development of tomato production. There are many types of tomato diseases, which have large impacts and outbreaks. Accurate identification of tomato diseases is the key to preventing diseases and improving tomato yield and quality. But accurate disease identification requires a high level of literacy and disease experience in agricultural practitioners. Agricultural practitioners must understand the causes of all diseases and insect pests, as well as the symptoms and characteristics of diseases. Agricultural practitioners are even required to inspect crops around the clock. Traditional tomato disease recognition is limited to laboratories and has the disadvantages of no failure, low efficiency, and high cost. Using computer information technology to automatically identify and diagnose tomato diseases provides a new solution to the above problems.

When crops are infected with different diseases, the disease will be displayed in different forms in leaf size, color, texture, and veins. Therefore, by processing tomato leaves, accurate identification of tomato diseases can be achieved. Existing deep learning methods have made great efforts to improve the accuracy of the data set, but ignored the important anti-interference ability of the method. When these methods are applied to real-life scenarios, they may not be able to deal with diseased leaves with various noise disturbances. However, improving the robustness of the recognition algorithm is the key to practical applications.

This paper uses mix-up [1] and other methods to expand and enhance the original image data set, effectively avoiding the risk of overfitting the model, and the model's accuracy and robustness are significantly higher than existing methods. This paper uses different network models and optimizes them based on the original network, and finally obtains the optimal network model. We provide a
large number of experiments to evaluate the effectiveness of our method. Our method achieves state-of-the-art results with an average accuracy of 88.83% on the tested data set.

2. Related Works

In the past, many crop disease methods have been proposed. According to the way of extracting features, we can divide them into two categories: methods based on manual feature extraction and deep learning-based methods.

In terms of image recognition of crop diseases based on hand-designed features: Li Wang et al. [2] used different kernel function vector machines to compare the color features of the extracted HIS space of cucumber leaves, and the results showed that the radial basis kernel has the best classification effect on the three diseases of cucumber downy mildew, corner spot disease, and powdery mildew; Du Haishun et al.[3] proposed a dual-cooperative representation classification method for wheat, and the 22 leaf diseases of corn, peanut and cotton were classified and identified.

Aspects of crop disease image recognition based on deep learning: Mohanty SP et al. used Google Net [4] and AlexNet [5] to draw an average classification effect of GoogleNet [4] for Plant Village, which is slightly better than AlexNet [5]. Nachtigall LG et al. [6] used the AlexNet [5] model to identify six types of apple disease images totaling more than 2,000, and the recognition accuracy rate exceeded the accuracy rate of expert recognition. Funentes et al. [7] tested tomato pests and diseases. The three deep learning original structures of R-CNN [8], Faster-RCNN [9], and SSD [10] are combined with VGGNet [11] and ResNet [12] networks, and performance comparison is performed to find suitable deep learning structures.

However, under the real environment with a complex background and various noise disturbances, the accuracy of recognition is often greatly reduced, making it difficult to identify crop diseases accurately in the real environment. Therefore, how to improve the accuracy and the anti-interference ability of crop disease image recognition has become the key to the research of crop disease recognition.

3. Implementation details

3.1 The Source of Data Set

Tomato disease pictures used in this article are mainly from the AI Challenger platform. The crop disease data set under the AI Challenger platform have nearly 50,000 pictures of crop diseases and insect pests, 10 species, 27 diseases, and each plant contains several kinds of health and Image of a disease state. Each disease contains two pictures with different degrees. For example, tomato leaf mold includes serious fungus and general fungus. This article selected 11 types of tomato disease pictures, a total of 11,476 pictures. As shown in Figure 1, it shows part of the tomato disease dataset:
3.2 Data Set Construction

Tomato disease training data provided by the AI Challenger platform have problems such as insufficient image data of disease categories and imbalance between different categories, which will reduce the classification accuracy of the classification model. However, selecting suitable features from the imbalanced data set has an important role in improving the accuracy of the classification of machine learning and deep learning models. In the real world, imbalanced data sets are common. The amount of data in the sample is insufficient, and the categories are not balanced, which will greatly affect the accuracy of the deep model classifier.

![Figure 2. Distribution map of a training set](image)

From Figure 2, we can see that there is only one picture each for severe tomato scab and tomato scab, so this article first deletes the single pictures of severe tomato scab and tomato scab categories. There are only 21 training samples for the tomato spot disease category, and 2473 training samples for the tomato yellow leaf curl virus disease severe category. This will lead to serious bias problems, which will reduce the generalization ability of the learned deep model. In order to improve the accuracy of the model and enhance the robustness of the model, this paper uses a variety of data preprocessing and data augmentation techniques to expand the dataset. These data enhancement methods can well address the issue of insufficient datasets and uneven data categories.

Pictures are processed as follows:

- **Gaussian noise**: Adds Gaussian distributed additive noise to the picture
- **Brightness change**: Randomly enhance the lightness with the factor of 1.5 or decrease it with the factor of 0.5
- **Random flip**: The image can be flipped in any way from left to right or up and down
- **Contrast change**: Randomly enhance the Contrast with the factor of 1.5 or decrease it with the factor of 0.7
- **Mix-up data enhancement**: Randomly draw two samples from the training data, and use the ratio of $\lambda$ to 0.4 to mix the two pictures to get a brand-new picture.

3.3 Implementation Details

We adjusted all networks to the specific task of disease pictures. A dense layer is used as a fully connected layer, and feature vectors of different dimensions are utilized to train an 18 layers network classifier. Finally, we obtained results based on the confidence of the 18 categories. The network architecture of the classification task is shown in Figure 3.
Since the pictures in the dataset cannot be used directly for network model training, this article first compresses the images to a size of 256 * 256, and then all the input pictures are resized according to the resolution of 224 × 224 and 3 channels. We use ResNet [12], Residual Attention Network [13], SE-ResNet networks [15], and Shufflenet_v2[15] for training. Due to the uneven distribution of sample classes in the crop disease image dataset, we use the loss function focal loss [7] to fully mine the positive sample data, improve the training weight, and accuracy of network recognition. The mathematical expression of focal loss is given in Equation 1:

\[
L_f = \begin{cases} 
-\alpha (1 - y') \log y', & y = 1 \\
-(1 - \alpha) y' \log (1 - y'), & y = 0 
\end{cases}
\]  

(1)

The network uses the standard Adam optimization algorithm to optimize the loss function. The batch size is set to 32 and the batch size is set to 120. After each epoch, the model will verify the validation set, record the accuracy of the model verification, and generate the final model as the ultimate result. The initial learning rate is set at 0.001. Under the current learning rate, when the number of training rounds of the validation dataset reaches 12 and the standard evaluation stops improving, the learning rate is reduced, and the learning rate decay to 0.1 times the current learning rate.

### 3.4 Comparison of Recognition Accuracy

We divide the original data set into a training set and validation set, and the ratio of the training set to the validation set is about 9:1. In order to solve the problem of imbalanced datasets, we perform data augmentation on the original data set. The number of pictures in the original dataset is 11476 and the number of pictures in the enhanced training set is 20158, and the number of pictures in the validation set is 1638.

In order to evaluate the effectiveness of these measures, we trained the enhanced dataset and the original dataset on the network separately. The experimental results are shown in Table 1.

| The type of Network | mixup | train_acc | train_loss | validation_acc | validation loss |
|---------------------|-------|-----------|------------|----------------|-----------------|
| ResNet18            | no-mixup | 0.8998 | 0.3416 | 0.873 | 0.4558 |
|                     | mixup   | 0.9067 | 0.2991 | 0.8771 | 0.414 |
| ResNet34            | no-mixup | 0.9043 | 0.2263 | 0.8809 | 0.3463 |
|                     | mixup   | 0.9369 | 0.1576 | 0.8812 | 0.3358 |
| ResNet50            | no-mixup | 0.9013 | 0.2508 | 0.8793 | 0.3708 |
|                     | mixup   | 0.8986 | 0.3367 | 0.8771 | 0.488 |
| ResNet101           | no-mixup | 0.8815 | 0.3683 | 0.8674 | 0.4497 |
|                     | mixup   | 0.902  | 0.3261 | 0.8751 | 0.4275 |

Figure 3. The network architecture of the classification task
It can be observed that increasing the number of network layers can improve the classification performance of the network. For example, the accuracy of the ResNet34 network is higher than that of ResNet18, but when the depth of the network continues to increase, the performance of the network decreases. It may be due to the increase of the network depth that during the process of transmitting the input features of the convolutional layer back, some features are suppressed, which affects the classification effect of the fully connected layer.

We conducted experiments on three networks: Residual Attention Network [13], SE-ResNet [9], and Shufflenet_v2[10]. The results are shown in Table 2. The model based on the SE-ResNet achieved 88.83% accuracy on the validation set, which is higher than other networks but SE-ResNet network model complexity and model footprint are higher than other networks.

| The type of Network     | mixup     | train_acc | train_loss | validation_acc | validation_loss |
|-------------------------|-----------|-----------|------------|----------------|-----------------|
| Residual Attention Network | no-mixup  | 0.8919    | 0.2688     | 0.8809         | 0.3206          |
|                         | mixup     | 0.9166    | 0.2065     | 0.8852         | 0.2929          |
| SE-ResNet               | no-mixup  | 0.9223    | 0.2099     | 0.8786         | 0.3551          |
|                         | mixup     | 0.9444    | 0.1578     | 0.883          | 0.3515          |
| Shufflenet_v2           | no-mixup  | 0.9112    | 0.216      | 0.8599         | 0.4033          |
|                         | mixup     | 0.9348    | 0.1572     | 0.8649         | 0.3818          |

4. Conclusion and Further Research
In this study, we have proposed the deep learning approach to build a classifier for disease classification. Our model also can be fast, accurate and robust recognition target of crop disease image in a practical environment. However, crop diseases sometimes occur at the same time. For this situation, it is necessary to study the disease identification method when multiple diseases coexist. In the next work, we will consider related research on the diagnosis and early diagnosis of multiple diseases.

Acknowledgement
Some of the systems described in this paper is the work of many researchers now at school of software engineering, Beijing University of Technology. We would like to express my heartfelt gratitude to Prof. Tao Zhang, who led me into the world of artificial intelligence. We also wish to thank Prof. Tao Zhang, Yiqing Liu, and Kun Zhang for helpful discussions, and Azhar Imran for providing the English help, and Prof. Tao Zhang for relentless support and encouragements.

References
[1] Zhang, H., Cisse, M., Dauphin, Y. N., Lopez-Paz, D. (2017) mixup: Beyond Empirical Risk Minimization. In: ICLR 2018: International Conference on Learning Representations. Vancouver.
[2] Li, W., Tang, S., Chen, R. (2014) Recognition of Cucumber Disease Based on Color Features and Support Vector Machine. Journal of Agricultural Mechanization Research, 2014(1):73-75.
[3] Du, H., Jiang, M., Wang, J., Wang, S. (2017) Double Weighted Collaborative Representation Based Classification for Crop Leaf Disease Image Recognition. Computer Science, 44(10): 302-306.
[4] Krizhevsky, A., Sutskever, I., Hinton, G. E. (2012) ImageNet Classification with Deep Convolutional Neural Networks. In Advances in neural information processing systems. Lake Tahoe. pp. 1097-1105.
[5] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A. (2015) Going Deeper with Convolutions. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston. pp. 1-9.
[6] Nachtigall, L. G., Araujo, R. M., Nachtigall, G. R. (2016) Classification of Apple Tree Disorders Using Convolutional Neural Networks. In: 2016 IEEE 28th International Conference on Tools with Artificial Intelligence (ICTAI). San Jose. pp. 472-476.

[7] Fuentes, A., Yoon, S., Kim, S., Park, D. (2017) A Robust Deep-Learning-Based Detector for Real-Time Tomato Plant Diseases and Pests Recognition. Sensors, 17: 2022.

[8] Girshick, R., Donahue, J., Darrell, T., Malik, J. (2014) Rich Feature Hierarchies for Accurate Object Detection and semantic segmentation. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Columbus. pp. 580-587.

[9] Ren, S., He, K., Girshick, R., Sun, J.,(2015) Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Transactions on Pattern Analysis & Machine Intelligence,39(6):1137-1149.

[10] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., Berg, A. C. (2015) SSD: Single Shot MultiBox Detector. In: 2016 European Conference on Computer Vision (ECCV). Amsterdam. pp. 21-37.

[11] Simonyan, K., Zisserman, A. (2014) Very Deep Convolutional Networks for Large-Scale Image Recognition. In: International Conference on Learning Representations 2015 (ICLR). San Diego.

[12] He, K., Zhang, X., Ren, S., Sun, J. (2016) Deep Residual Learning for Image Recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas.pp. 770-778.

[13] Wang, F., Jiang, M., Qian, C., Yang, S., Li, C., Zhang, H., Wang, X., Tang, X. (2017) Residual Attention Network for Image Classification. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu.pp. 3156-3164.

[14] Hu, J., Shen, L., & Sun, G. (2018) Squeeze-and-Excitation Networks. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City. pp. 7132-7141.

[15] Zhang, X., Zhou, X., Lin, M., & Sun, J. (2018) Shufflenet: An Extremely Efficient Convolutional Neural Network for Mobile Devices. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City. pp. 6848-6856.

[16] Lin, T. Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017) Focal Loss for Dense Object Detection. In: 2017 IEEE International Conference on Computer Vision (ICCV). Venice. pp. 2999-3007.