Tobacco-disease Image Recognition via Multiple-Attention Classification Network

Y Sun¹, H Q Wang¹,³, Z Y Xia², J H Ma² and M Z Lv³,*

¹School of Software, Beihang University, Beijing 100083, China
²Yunnan Academy of Tobacco Agricultural Science, Kunming, Yunnan 650021, China
³Beihang University Yunnan Innovation Institute, Kunming 650000, China

*Corresponding author’s E-mail: lvmingzhan@163.com

Abstract. The recognition of disease images is of great importance in modern tobacco agriculture. Compared with other plant diseases, the early disease spots of tobacco are highly similar, and the quantity of tobacco disease images is scarce due to the difficulty of data collection. As such, the image recognition of early tobacco diseases is challenging. In response to these problems, this study proposes a novel image-classification method, namely, Multiple-Attention Classification Network (MACN). It improves the existing image classification network from two aspects: 1) the acquisition of low-level features based on transfer-learning and 2) the extraction of high-level features by modelling the multiple dependencies in low-level features. The pretrained model is used to obtain the initial feature map, thereby breaking the limitation of small-scale dataset. A novel multiple-attention module (MAM) is designed to learn fine-grained differences between classes. To evaluate the proposed method, this paper conducts experiments on field-collected images. Results show that this method is superior in identifying early tobacco diseases and has demonstrated better robustness when dealing with the actual field images with complex factors.

1. Introduction

Tobacco is an important economic crop. At present, the diagnosis of diseases relies on the traditional method in accordance with tobacco farmers’ experience, which causes economic losses frequently due to untimely and inaccurate diagnosis. With the development of computer vision technology, agricultural-disease image recognition is attracting extensive research interest. However, the studies related to tobacco images are limited.

This article focuses on the realisation of early image recognition of common tobacco diseases. Through this work, tobacco farmers can obtain leaf lesion information by taking pictures with mobile phones in the field, thereby avoiding the large-scale spread of the disease. As shown in Figure 1, the initial plaques of tobacco target spot, tobacco brown spot and tobacco wildfire are very similar. The tobacco weather fleck and the tobacco mosaic disease are relatively insignificant and highly similar to healthy tobacco leaves. The image recognition of early tobacco diseases remains challenging due to the following problems. 1) The early disease spots of some diseases are highly similar. 2) The collection work in tobacco fields is restricted by season, and diseases develop quickly. Therefore, the scale of early disease images cannot meet the needs of training the fine-grained classification model.
The main contributions of this study can be summarised as follows.

1) A novel MAM is designed to model multiple dependencies in low-level features.
2) A novel classification network structure is proposed to extract fine-grained features from a small sample set, which combines transfer learning with the above MAM to learn high-level features.

The rest of the article is organised as follows. In section 2, the related work about recent disease image classification and attention mechanism are reviewed briefly. In section 3, the details of our method are presented. In section 4, the experimental results on the actual early tobacco-disease dataset are shown. Finally, in section 5, the summary of this work is presented.

2. Related Work

In the study of the fixed disease categories, manual [1][2] or semiautomatic [3] feature-extraction methods are often used to model the features directionally. But the manual methods need to solve the problems caused by environmental factors [4]. Deep learning technology provides the possibility of automatic feature-extraction [5] [6]. While images in public plant datasets are usually collected under controlled conditions. Studies [7] have shown that the accuracy declines rapidly when use the model trained on PlantVillage dataset in actual environment. Given the difficulty of field-collection, transfer learning has been used to make up for the lack of data [8].

In addition, the visual attention mechanism is effective for weakly supervised fine-grained classification. The attention proposal network (APN) [9] has been proposed to locate the distinguishing regions in the image. The squeeze-and-excitation (SE) networks (SE Nets) [10] have been designed to explore the inter-channel dependencies. Wang et al. [11] have designed residual attention network (RAN) composed of a series of residual attention modules (RAM), which has received mixed attention constraints based on residual learning.

The fine-grained classification on a small-scale dataset remains challenging because the extraction of high-level feature from limited data is difficult for CNNs. In addition, the quality of field-collected images is easily affected by environment factors. The method proposed in this article is dedicated to solving the above problems.

3. Methodology

3.1 Structure of the model

We developed a novel MACN to achieve fine-grained classification due to the limited samples of tobacco-disease images. The overall structure of this model is shown in Figure 2.

MACN contains two key components which are pretrained model and multiple-attention module. Considering that the lesions of some diseases are distributed extensively on the leaves, we believe that the attention mechanism [11] is more suitable for this work than the method based on local ROI [9]. On the basis of the initial feature map provided by pretrained model, the lightweight MAM extract fine-grained features and export the multi-weighted feature map. And then a global average pooling layer is added to compress it into a one-dimensional feature vector. Finally, a fully connected layer realises the mapping between the feature and the category vectors.
In this work, the InceptionV3 [12] model trained on the ImageNet is used due to the higher accuracy of it on our dataset than the others. The relevant experimental data are provided in Section 4.2.

The MAM refers to the structure of attention module in RAN [11], which includes two branches, namely, the channel- and the spatial-attention branches. Unlike the RAM [11], MAM model the channel- and spatial-domain dependencies in the initial feature map, rather than extract new features. Thus, lightweight MAM without redundant convolution calculation units is designed.

In section 3.2, we have reviewed the structure of the InceptionV3 [12] and introduced the initial feature map. In section 3.3, we have described the structure and design ideas of the MAM.

3.2 InceptionV3 and initial feature map
The InceptionV3 network constructs sparse structural units and deconstructs the large convolution integral into multiple small convolution kernels. This design increases the nonlinear mapping capability of the network whilst reducing the parameters. Moreover, to avoid the sharp attenuation of the feature-map size, it uses a parallel dimensionality reduction structure. The three main types of grid size in this model are 35×35, 17×17 and 8×8, and parallel dimensionality reduction plays a role in the changes amongst them. At the stage where the grid size is 8×8, the output modules of the filter are added, which enhances the feature expressiveness. In the InceptionV3 model, the lower-dimensional embedding finishes the spatial aggregation, and the higher-dimensional representation explains the more complex features.

To fully exploit the advantages of pretrained model, we use the final feature map of the 8×8 grid stage as the initial feature map. It utilises the strong feature-extraction capability which enables rapid training on small-scale dataset and increases the features’ effectiveness. A good basis is also provided for extracting fine-grained features.

3.3 MAM
The branches of lightweight MAM generate the channel- and spatial-attention constraints respectively, and output high-level features based on low-level features.

The SE module in channel-attention branch compresses the global spatial information in every channel and then generates the global-distribution embedding of the channel domain. The operation of global average pooling provides basis for the excitation operation to perform feedforward calculation and two consecutive fully connected layers fully capture the channel correlation. Finally, the output $F_{se}(X)$ is calculated by multiplying the weights generated and input feature map $X$. The structure of the channel-attention branch is shown in Figure 2.

The spatial-attention branch includes bottom-up feed-forward scanning and top-down feedback steps. The first step completes a quick scan of global information and extracts key information from the input feature map. The latter operation restore the key information and generates the weight value $S_{sc}(X)$ for each position of the input feature map $X$. Considering that the initial feature map is highly
abstracted by InceptionV3 model, the spatial-attention branch we design is more lightweight than the Mask Branch in RAM [11]. First, max pooling is performed to obtain a \(4 \times 4\) feature map containing global information. Then, bilinear interpolation is conducted to upsample it and maintain the same output size as the input feature map. The weights \(S_{c,i}(X)\) are subsequently normalised from each channel and then adjusted into interval \([0, 1]\) as the output of spatial-attention branch. In the bottom-up and top-down structures, we retain the residual unit [11] between sampling operation to increase the flexibility of mask calculation.

Finally, the spatial-domain weight matrix \(S(X)\) generated by the spatial-attention branch is applied to the channel-weighted feature map \(F_{sq}(X)\). To avoid information loss, we borrow the idea of attention residual learning [11]. Thus, the output \(F_{ma}(X)\) of MAM is calculated as follows:

\[
F_{ma}(X) = (1 + S(X)) \times F_{sq}(X).
\]

Based on the initial feature map, a MAM is used in this work to obtain fine-grained features. We believe that the initial feature map in this work fully extracts the low-level features, so we do not continue to add the convolution unit. If the features in initial feature map cannot sufficiently deal with more complex problems, continued addition of feature-extraction units would be necessary that include convolution calculations whilst adding other MAMs behind it to filter new features.

4. Experiment

4.1 Dataset

The image data of the tobacco diseases we use are all taken in the field using mobile phones. They contain early images of five common tobacco disease leaves and healthy tobacco leaves.

To solve the problem of insufficient data in the dataset, we used data-enhancement methods, including cropping, random rotation, random flipping, Gaussian noise, to expand the training set. Considering that colour feature is a distinguishing feature of lesions, we do not use colour-dithering enhancement methods. Furthermore, the amount of data in different categories varies greatly in the original dataset. And we maintain data balance by controlling the amount of enhanced data used actually.

In the experiment, we divide each category of the original dataset into 10% as the test set and then expand the remaining data. The amount of data before and after dataset expansion is shown in Table 1.

| Table 1. Number of images in dataset. |
|--------------------------------------|
| **Original Training Set** | **Expanded Set** | **Testing Set** |
| Tobacco Target Spot | 313 | 601 | 32 |
| Tobacco Brown Spot | 509 | 655 | 54 |
| Tobacco Weather Fleck | 287 | 683 | 30 |
| Tobacco Mosaic Disease | 726 | 726 | 81 |
| Tobacco Wildfire | 81 | 486 | 9 |
| Healthy Tobacco Leaf | 87 | 522 | 9 |

4.2 Results and discussion

We initially evaluate the learning ability of various classic CNNs on the tobacco-disease dataset and then select the suitable model as the initial feature-map providing unit. To prove the importance of MAMs, we design verification experiments with single attention unit and compared with existing attention networks on tobacco-disease dataset to show the advantages of our method. In the experiment, we use the early-stop rule to save optimal model, i.e., training is stopped when the accuracy of the verification set is not improved after a few epochs.

1) **Model selection and transfer-learning advantages.** To select the suitable basic model, we compare three classic CNNs (VGG16, ResNet50 and InceptionV3) on the tobacco-disease dataset. The data in Table 2 show that models without transfer learning are overfitting, and amongst the three
transfer learning models trained on the tobacco-disease dataset, InceptionV3 has higher test accuracy than the other models. Thus, the InceptionV3 model based on transfer learning play an important role in improving the accuracy and robustness of the model for identifying tobacco diseases.

Table 2. Test accuracy of basic CNNs.

| Top-1 accuracy | VGG16  | ResNet50 | InceptionV3 |
|---------------|--------|---------|-------------|
| Regular Learning (%) | 26.07  | 16.67   | 19.24       |
| Transfer Learning (%) | 76.27  | 67.87   | 80.3        |

2) Advantages of MAMs. We verify its effectiveness from two aspects: 1) comparing MAM with SEM and the lightweight-RAM separately to prove the importance of mixed weights, and 2) by combing residual attention module [11] with pretrained model to verify the role of the lightweight design.

The experimental data in Table 3 show that compared with the transfer-learning model, the accuracy increases by 1.4% after adding the SEM and by 1.91% after adding lightweight-RAM. Compared with the lightweight MAM, superimposing the RAM containing multiple convolution units on pretrained model induces overfitting. Therefore, the lightweight MAM not only exploits multiple dependencies but also prevents overfitting effectively.

Table 3. Accuracy of models with different attention modules.

| Added Module | SEM[10] | Simple-RAM | RAM[11] | MAM     |
|--------------|---------|------------|---------|---------|
| Val-Accuracy (%) | 93.75   | 89.06      | 93.75   | 95.31   |
| Testing-Accuracy (%) | 81.72   | 82.21      | 31.48   | 85.56   |

3) Advantages of MACN. Finally, we compared MACN with mature attention networks such as SE-ResNet[10], SE-Inception-ResNet-V2[10], and Residual Attention Network[11]. In the study of tobacco-disease identification, our proposed MACN has better recognition ability and robustness. Owing to the small size of the dataset and the small differences amongst classes, deep attention classification networks are not applicable in this work. SENets [10] got the optimal models less than 30 epochs, while they all overfitted on the test set. However, the optimal RAN model was obtained after 50 epochs which performed similar accuracy on testing set and validation set. As shown in Table 4, MACN demonstrates significant superiority over these methods.

Table 4. Comparison of the proposed method with baselines.

|               | SE-ResNet[10] | SE-Inception-ResNet-V2[10] | RAN[11] | MACN     |
|---------------|---------------|----------------------------|---------|----------|
| Val-Accuracy (%) | 83.13         | 85.94%                     | 65.24   | 95.31    |
| Testing-Accuracy (%) | 19.14         | 18.33                      | 64.6    | 85.56    |

5. Conclusion
This paper proposes to solve the image-recognition problem of early tobacco diseases through MACN. According to the characteristics of the small-scale tobacco-disease dataset and the small differences amongst classes, the following solutions are designed specifically: 1) using transfer learning to quickly obtain the initial feature map and thus break the limitation of the small-scale dataset, and 2) modelling the multiple dependencies in the initial feature map and presenting a novel MAM. Experiments conducted on field-collected tobacco-disease dataset prove the effectiveness of transfer learning and MAM, thereby proving that this method has significantly better than existing methods.

Acknowledgement
This research project was funded by the Yunnan Tobacco Company Science and Technology Project’s “Research and Application of Intelligent Identification of Tobacco Diseases and Insect Pests” (2019530000241007). We are also grateful to the Yunnan Academy of Tobacco Agricultural Science for providing guidance and data support in the processing of tobacco-disease images.
References

[1] Hlaing C S, Maung Zaw S M. Plant diseases recognition for smart farming using model-based statistical features. 2017 IEEE 6th Global Conference on Consumer Electronics (GCCE). IEEE, 2017: 1-4.

[2] Hlaing C S, Maung Zaw S M. Tomato Plant Diseases Classification Using Statistical Texture Feature and Color Feature. 2018 IEEE/ACIS 17th International Conference on Computer and Information Science (ICIS), Singapore, 2018, pp. 439-444.

[3] Kaur S, Pandey S, Goel S. Semi-automatic leaf disease detection and classification system for soybean culture[J]. Iet Image Processing. 2018, 12(6):1038-1048.

[4] Priya C M, Tamilselvi P R. Primary Pitfalls in Detection and Analysis of Diseased Chunk of Leaf Images. 2017 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), Coimbatore, 2017, pp. 1-4.

[5] H. TANI, R. KOTANI, S. KAGIWADA, H. UGA and H. IYATOMI. Diagnosis of Multiple Cucumber Infections with Convolutional Neural Networks. 2018 IEEE Applied Imagery Pattern Recognition Workshop (AIPR), Washington, DC, USA, 2018, pp. 1-4.

[6] Verma S, Chug A, Singh A P. Prediction models for identification and diagnosis of tomato plant diseases[C]. 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI). IEEE, 2018: 1557-1563.

[7] Arnal, Barbedo J G. Impact of dataset size and variety on the effectiveness of deep learning and transfer learning for plant disease classification. Computers and Electronics in Agriculture 153(2018):46-53.

[8] Hyun K S, Joris IJ, Jan W H, et al. 2018. Transfer learning for the classification of sugar beet and volunteer potato under field conditions, Biosystems Engineering, Volume 174, Pages 50-65.

[9] Fu J, Zheng H, Mei T. 2017. Look Closer to See Better: Recurrent Attention Convolutional Neural Network for Fine-Grained Image Recognition. IEEE Conference on Computer Vision and Pattern Recognition. 4476-4484.

[10] Hu J, Shen L, Sun G. Squeeze-and-excitation networks. CVPR, pp. 7132-7141, 2018.

[11] Wang F, Jiang M, Qian C, Yang S, Li C. 2017. Residual attention network for image classification. CVPR, pp. 3156-3164.

[12] Szegedy C, Vanhoucke V, Ioffe S, Shlens J and Wojna Z. Rethinking the Inception Architecture for Computer Vision. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 2818-2826.