

# Identification of Coffee Leaf Pests and Diseases based on Transfer Learning and Knowledge Distillation

Zi Wang<sup>1,\*</sup>, Zheng Ren<sup>1</sup>, Xue Li<sup>2</sup>

<sup>1</sup> School of Electrical and Information Engineering, Wanjiang University of technology, Maanshan, CO 243031, China

<sup>2</sup> School of Economics and Management, Yango University, Fuzhou, CO 350015, China

\* Corresponding author: Zi Wang (Email: 1043901879@qq.com)

**Abstract:** The yield of coffee has a significant effect on the development of the economy. It is important to monitor the health status of coffee plants. Leaves can represent the growth of crops. Analysis leaf image is an effective method to monitor crop growth status. With the advancement of artificial intelligence technology, neural networks with strong learning ability have been proposed. They have high accuracy in identifying leaf pests and diseases. However, the structure of these networks is complex and the speed of computing is slow. They are not conducive to real-time analysis. For simple networks, it is difficult to achieve high recognition accuracy directly. To solve this problem, a lightweight model is designed for leaf image analysis. Leaf images are learned by VGG network with pre-trained weights on ImageNet. Use the VGG network as a teacher network. Then design a lightweight student network. Train student network with knowledge distillation method. A lightweight model with high recognition accuracy can be obtained. This research explored the effect of the method on the coffee leaf data set. Experiment proved that the accuracy of the proposed method is 96.73%. The accuracy was 4.29% higher than directly training. Meantime, the calculation speed of the model is quick. The proposed method is of great practical significance for identifying coffee leaf pests and diseases.

**Keywords:** Coffee Leaf; Transfer Learning; Knowledge Distillation.

## 1. Introduction

The fruit of the coffee plant is nutritious. It has many industrial and product chains. So, it has an important impact on economic development [1]. Coffee leaves contain nutrients such as chlorophyll. Its content reflects the health status of the plant. Analysis plant leaf image is a method to monitor plant health. Judging the growth state of the leaf by human eyes and experience are labor-intensive. And it is also less efficient. With the promotion of agricultural intelligence, machine learning and other technologies have been introduced into agricultural production [2]. Deep learning methods have shown excellent efficiency in recent years. They can process images to achieve reasonable analysis results. Especially, deep and large neural networks have great learning abilities [3]. But their number of parameters makes calculation difficult. The calculation speed of the model is slow. Real-time analysis cannot be satisfied in production. Lightweight models are much faster to analyze [4]. However, it is difficult for lightweight models to achieve high recognition accuracy directly. They need knowledge base during training. The parameters are adjusted by iterative fitting. Knowledge distillation is a feasible way. It can make lightweight models learn from larger models. The results are usually better than independent study. Therefore, a rich knowledge base and an effective distillation method are necessary elements. This is also a research direction in current production and application [5].

## 2. Related Work

In recent years, researchers have carried out a series of work on crop leaf pests and diseases. Y Sun proposed a new algorithm combining SLIC (Simple Linear Iterative Cluster) with SVM (Support Vector Machine) [6]. The accuracy and

F-value are 96.8% and 97.7% respectively. JAAA Basavaiah used random forest and decision tree classification algorithms to classify leaf disease [7]. The classification accuracy is 90% for decision tree classifier and 94% for random forest classifier respectively. J Suto found an interesting deficiency and a key difference between studies [8]. This work offered an overall review about the efficient. In 2020, A Nigam used the PCA algorithm to extract specific features [9]. Moreover, the feature extraction and BFO-DNN method was proposed. And it achieved good results. In 2021, Z Jiang improved the Visual Geometry Group Network-16 model based on the idea of multi-task learning and then use the pre-training model on ImageNET for transfer learning and alternating learning [10]. The accuracy of such model is 97.22% for rice leaf diseases and 98.75% for wheat leaf diseases. In the same year, SK Noon proposed a simple yet efficient deep learning-based framework to recognize cotton leaf diseases [11]. The proposed model is capable of achieving the near ideal accuracy with early convergence to save computational cost of training. W Zeng proposed a lightweight dense-scale network for real-world corn leaf disease image identification and a new loss function [12]. Experimental results proved that the accuracy of the optimized model on the test data set reaches 95.4%. In 2022, Zhen Wang proposed a novel deep neural network structure named as MPF-Net for weed species identification [13]. The calculation speed of this model is fast and the accuracy is high. Among these methods, the structure of some models is complicated. Their calculation speed cannot meet the real-time requirement in production. Lightweight models are difficult to achieve high accuracy. After a comprehensive consideration, knowledge distillation is used to train a lightweight model in this research. The generated model has high calculation speed and high accuracy.

### 3. Method

#### 3.1. Data Preprocessing

The experimental object of this research is coffee leaves. There are 5 kinds of pests and diseases on coffee leaves. They are respectively healthy, browned, withered, perforated and spotted. The number of 5 kinds of images are 327, 286, 295, 351 and 332 respectively. As shown in Figure 1.



Figure 1. Images of coffee leaf

There may be noise in the original images. Noise will affect the change of image gray value and interfere with the display of detail features. In the preprocessing stage, the mean filtering method is used to suppress the noise on the image [14]. Mean filtering replaces the value of the center pixel with the mean of all pixels in a fixed window. Figure 2(a) and Figure 2(b) respectively show the histogram distribution of the image before and after filtering [15]. After filtering, the noise of the image is reduced and the gray value is stable.

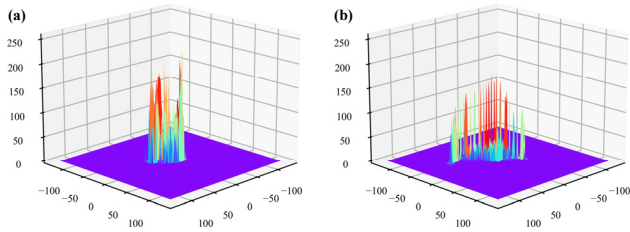


Figure 2. Histogram distribution of the image before and after filtering

Resize all images to 64×64. The input image contains three color channels such as R, G, and B. Output category for each image. Use one-hot to encode the category of each image.

#### 3.2. Principle

The method proposed in this research is to train and obtain a large network with excellent recognition effect by transfer learning firstly. Then, design a small network. Use knowledge distillation to make the small network learn information extracted by the large network. Finally, A lightweight model with fast computing speed and high recognition accuracy is generated.

Transfer Learning is a machine learning method. Start with the model developed in Task A. Develop the model in Task B again. Transfer learning can use knowledge from a related task that has been learned to learn a new task [16]. It is often difficult for a model to learn all the feature representations in the same task. This requires different models to learn and adjust from other models. This research used transfer learning. Its flow is shown in Figure 3. Prepare the pre-trained weight of VGG on ImageNet. The parameters of the VGG network were adjusted on the coffee leaves. This allows the model to be better optimized. Use the trained VGG network as a teacher network.

People often use complex models to get the best results. This may lead to severe redundancy of parameters. For example, VGG has about 100 million parameters. In the case of forward propagation, it includes convolution computation, compression and full connection. The process of calculation

is complicated. It has poor performance in practical engineering application. Teach what a complex model (Teacher) has learned to another lightweight model (Student). The lightweight model can often learn more than by its own. And the model works a lot faster. This method is called knowledge distillation [17]. As shown in Figure 4. It is an effective model compression method. The model generated by this method has good practical application value.

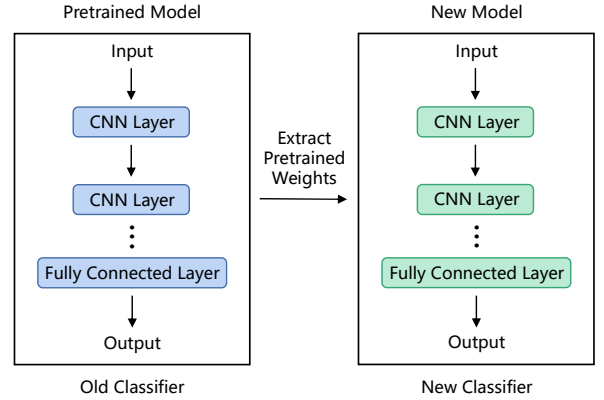


Figure 3. Flow chart of transfer learning

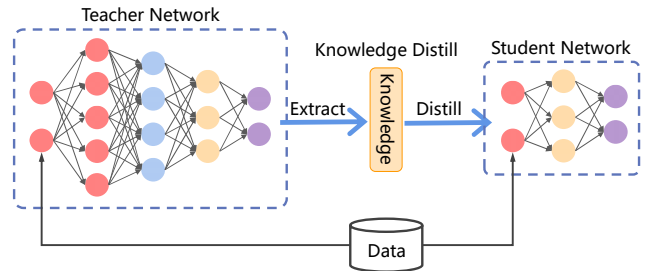


Figure 4. Flow chart of knowledge distillation

In neural network model, convolution layer is used to extract image feature information [18]. The calculation formula of convolution is shown in (1).

$$v_{ij}^x = b_{ij} + \sum_m \sum_{p=0}^{p_i-1} w_{im}^p v_{(i-1)m}^{x-p} \quad (1)$$

Where  $i$  represents the  $i_{th}$  convolution layer,  $j$  represents the  $j_{th}$  channel of the current convolution layer,  $p_i$  is the width of the convolution kernel,  $m$  indicates the feature mapping from the previous layer to the current layer,  $b$  is bias,  $w$  is the weight matrix. The batch normalization layer speeds up the iterative convergence of the network. The pooling layer are used to compress the size of feature map. It speeds up the calculation. The role of the dropout layer is to prevent the model from overfitting and make the model generalize. The fully connected layer gets the mapping of the feature data. Its forward propagation [19] formula is shown in (2). It can also be used as the final forecast output.

$$y_j = \sum_{i=1}^n w_{ij} x_i + b_j \quad (2)$$

Where  $x_i$  is the paved feature vector,  $w_{ij}$  is the weight,  $b_j$  is the bias.

#### 3.3. Model

In this research, the VGG network [20] was used to learn the information of coffee leaf images. Pre-trained weights on ImageNet were added at the beginning of training. Keep training until the loss function converged, the VGG network

with good recognition effect is obtained. Use it as a teacher network. A lightweight convolution neural network is also designed. Use it as a student network. The knowledge distillation method is used to make student network learn the information extracted by the teacher network. Finally, the lightweight network with fast computing speed and high accuracy was obtained. The complete flow of the proposed method is shown in Figure 5.

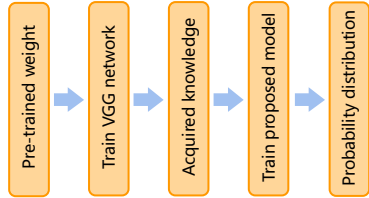


Figure 5. The complete flow of the proposed method

The student network has two convolution layers. The first convolution layer has 16 filters. The second convolution layer has 32 filters. Each convolution layer is followed by batch normalization and pooling operations. Pooling layer reduces the length and width of the feature map to a quarter of its original size. The goal is to speed up the computational efficiency. The multi-dimensional feature map is mapped to the fully connected layer containing 32 neurons. Finally, the activation function is the SoftMax function of the 5 neurons. They output probability. It represents the probability of belonging to a certain class. The cross-entropy loss [21] function is used in the training. The formula is shown in (3).

$$L = -\sum y_i * \log(p_i) \tag{3}$$

Where  $y_i$  is the true label for which class the sample belongs to,  $p_i$  is a prediction probability [22] belonging to class  $i$ . Dropout operation has been added. To prevent the model from overfitting. And make the model generalize. The structure diagram of teacher network and student network is shown in Figure 6.

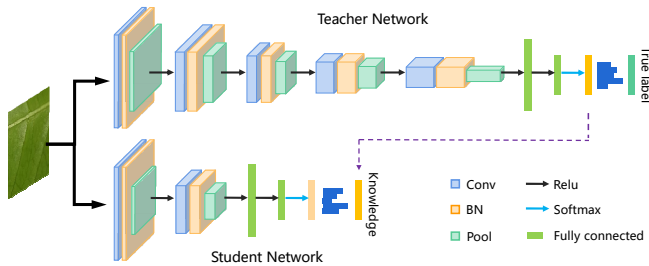


Figure 6. The structure diagram of teacher network and student network

After knowledge distillation, the student network will have the recognition ability of the teacher model. The calculation speed of the model is also improved while the accuracy is guaranteed.

## 4. Results

The calculation framework of this research is TensorFlow. The data set is divided into the train set and the test set in a ratio of 3:1. Use 10% of the train set as the verification set. The network structure of the proposed model is shown in Table 1.

The proposed model is lightweight. Compared to large models, its speed of calculation is very fast. The experiment also compares the recognition efficiency of different models.

As shown in Table 2.

Table 1. The network structure of the proposed model

Name	Kernel size	Stride	Output size
Conv1	3×3	1×1	64×64×16
BN1	--	--	64×64×16
Pooling1	4×4	4×4	16×16×16
Conv2	3×3	1×1	16×16×32
BN2	--	--	16×16×32
Pooling2	4×4	4×4	4×4×32
Flatten	--	--	512
FC	--	--	32

Table 2. The recognition efficiency of different models

Method	Accuracy	Time
SVM	87.91	0.023s
VGG	97.23	1.314s
VGG with pre-trained weights	97.98	1.297s
Lightweight model	92.44	0.065s
Proposed model	96.73	0.069s

It can be seen from Table 2 that the accuracy of the method proposed is 4.29% higher than the model directly trained with light weight. It was only 1.25% less accurate than the teacher network. It shows that the proposed method not only has high computation speed, but also has good recognition effect. The region of interest of the model to the leaf image is visualized [23]. Verify the ability of the model to extract features. As shown in Figure 7.

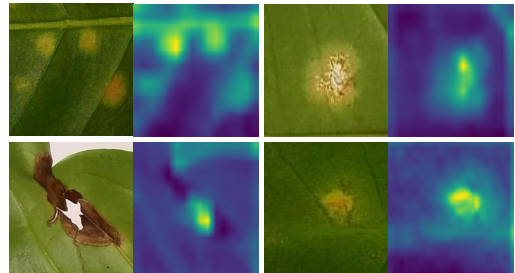


Figure 7. Visualization of the feature region of the images

For leaves infected with pests and diseases. The model can extract the feature region of the images correctly. This is helpful for analysis of leaf pathology. Properly monitor crop health and prevent the spread of pests and diseases. Plot the confusion matrix [24] and T-SNE visualization [25] results. As shown in Figure 8.

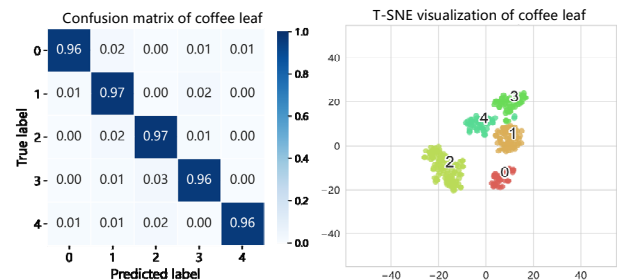


Figure 8. Confusion matrix and T-SNE visualization

For most categories of samples, the model has a high recognition accuracy. After the data of the fully connected layer are compressed into two dimensions, the scatter plot is drawn. The distribution of samples with extracted features

showed obvious regularity. Samples of the same kind are close together and samples of different kinds are far apart. This helps identify pests and diseases in the sample.

## 5. Conclusion

This research took into account the slow calculation speed of large models. And the direct learning efficiency of lightweight models is not good. This research proposed a method based on transfer learning and knowledge distillation to identify diseases in coffee leaves. The large network with pre-trained weights on ImageNet was used to learn information about coffee leaf images. Use it as the teacher network. Design a lightweight convolution neural network. Use it as the student network. The knowledge distillation method is used to make the student network learn the information extracted by the teacher network. Finally, a lightweight network with fast computing speed and high accuracy is obtained. Experimental results show that the proposed method is fast in computation and high in recognition accuracy. The model can be used for real-time analysis. It has good practical significance for monitoring crop health status.

## References

- [1] M. Zaccardelli, G. Roscigno, C. Pane. Essential oils and quality composts sourced by recycling vegetable residues from the aromatic plant supply chain[J]. *Industrial Crops and Products*, 2021, 162:113255.
- [2] S. A. A, J. S. B, M. R. A. Demystifying artificial intelligence amidst sustainable agricultural water management[J]. *Current Directions in Water Scarcity Research*, 2022, 7:17-35.
- [3] C. C. Lee, H. M. Fazalul R, L. P. Leow. A deep neural networks-based image reconstruction algorithm for a reduced sensor model in large-scale tomography system [J]. *Flow Measurement and Instrumentation*, 2022.
- [4] J. He, W. Zhang, R. Shang. Multi-Angle Models and Lightweight Unbiased Decoding-Based Algorithm for Human Pose Estimation [J]. *International Journal of Pattern Recognition and Artificial Intelligence*, 2023, 37(08).
- [5] Z. Tian, Y. Wang, X. Zhen. The effect of methanol production and application in internal combustion engines on emissions in the context of carbon neutrality: A review[J]. *Fuel: A journal of fuel science*, 2022(Jul.15):320.
- [6] Y. Sun, Z. Jiang, L. Zhang. SLIC\_SVM based leaf diseases saliency map extraction of tea plant[J]. *Computers and Electronics in Agriculture*, 2019, 157:102-109.
- [7] A. A. Basavaiah J A. Tomato Leaf Disease Classification using Multiple Feature Extraction Techniques[J]. *Wireless personal communications: An International Journal*, 2020, 115(1).
- [8] J. Suto. Plant leaf recognition with shallow and deep learning: A comprehensive study[J]. *Intelligent data analysis*, 2020 (6):24.
- [9] A. Nigam, A. K. Tiwari, A. Pandey. Paddy leaf diseases recognition and classification using PCA and BFO-DNN algorithm by image processing [J]. *Materials Today: Proceedings*, 2020, 33(3).
- [10] Z. Jiang, Z. Dong, W. Jiang, et al. Recognition of rice leaf diseases and wheat leaf diseases based on multi-task deep transfer learning [J]. *Computers and Electronics in Agriculture*, 2021, 186:106184.
- [11] S. K. Noon, M. Amjad, M. A. Qureshi, et al. Computationally light deep learning framework to recognize cotton leaf diseases [J]. *Journal of Intelligent & Fuzzy Systems: Applications in Engineering and Technology*, 2021(6):40.
- [12] W. Zeng, H. Li, G. Hu, et al. Lightweight dense-scale network (LDSNet) for corn leaf disease identification[J]. *Computers and Electronics in Agriculture*, 2022.
- [13] Z. Wang, J. Guo, S. Zhang. Lightweight Convolution Neural Network Based on Multi-Scale Parallel Fusion for Weed Identification[J]. *International Journal of Pattern Recognition and Artificial Intelligence*, 2022.
- [14] M. Jia, M. Dong. Analysis and comparison of Gaussian noise denoising algorithms[J]. *Journal of Physics Conference Series*, 2021, 1846(1):012069.
- [15] M. Suresha, D. S. Raghukumar, S. Kuppa. Kumaraswamy Distribution Based Bi-histogram Equalization for Enhancement of Microscopic Images[J]. *International Journal of Image and Graphics*, 2022(1):22.
- [16] P. Kora, C. P. Ooi, O. Faust. Transfer learning techniques for medical image analysis: A review[J]. *Biocybernetics and biomedical engineering /*, 2022(1):42.
- [17] K. Borup, P. Kidmose, H. Phan, et al. Automatic sleep scoring using patient-specific ensemble models and knowledge distillation for ear-EEG data[J]. *Biomedical signal processing and control*, 2023.
- [18] L. Xu, B. Zhou, X. Li. Gaussian process image classification based on multi-layer convolution kernel function[J]. *Neurocomputing*, 2022(Apr.1):480.
- [19] X. Liu, Y. Lei, X. Zheng. Reconfigurable Spoof plasmonic Coupler for Dynamic Switching between Forward and Backward Propagations[J]. *Advanced Materials Technologies*, 2022.
- [20] J. Qiu, X. Lu, X. Wang, et al. Research on Rice Disease Identification Model Based on Migration Learning in VGG Network[J]. *IOP Conference Series: Earth and Environmental Science*, 2021, 680(1):012087 (10pp).
- [21] J. Zhang, L. Zhao, J. Zeng. Deep MRI glioma segmentation via multiple guidances and hybrid enhanced-gradient cross-entropy loss[J]. *Expert Systems with Application*, 2022 (Jun.): 196.
- [22] H. Q. Mu, J. H. Shen, Z. T. Zhao, et al. A novel generative approach for modal frequency probabilistic prediction under varying environmental condition using incomplete information [J]. *Engineering Structures*, 2022, 252:113571-.
- [23] G. E. Yan, L. Zuan-Ran, L. I. Hai-Tao. Research on Mesoscale Vortex Feature Extraction Algorithm and Visualization[J]. *Computer Systems & Applications*, 2018.
- [24] Y. Wang, Y. Jia, Y. Tian, et al. Deep reinforcement learning with the confusion-matrix-based dynamic reward function for customer credit scoring[J]. *Expert Systems with Application*, 2022(Aug.):200.
- [25] H. Liu, J. Yang, M. Ye, et al. Using t-distributed Stochastic Neighbor Embedding (t-SNE) for cluster analysis and spatial zone delineation of groundwater geochemistry data[J]. *Journal of Hydrology*, 2021, 597(43):126146.