# TOMATO LEAVES DISEASE IDENTIFICATION MODEL BASED ON IMPROVED MobileNetV3
# /
# *基于改进 MobileNetV3 的番茄叶片病害识别模型*

**Cheng CHI$^{1,2,3)}$, Lifeng QIN*$^{1,2,3)}$**

$^{1)}$ College of Mechanical and Electronic Engineering, Northwest A&F University, Yangling, Shaanxi 712100, China
$^{2)}$ Key Laboratory of Agricultural Internet of Things, Ministry of Agriculture and Rural Affairs, Yangling, Shaanxi 712100, China
$^{3)}$ Shaanxi Key Laboratory of Agricultural Information Perception and Intelligent Service, Yangling, Shaanxi 712100, China
*Tel: +86 18009247416; E-mail: fuser@nwafu.edu.cn*

DOI: https://doi.org/10.35633/inmateh-75-89

***Keywords:*** *tomato leaf disease; improved MobileNetV3; attention mechanism; dilated convolution*

## ABSTRACT

*Aiming to address the issues of low accuracy and slow response in tomato leaf disease recognition models, an enhanced lightweight model for tomato leaf disease recognition was proposed. The SE attention module in the MobileNetV3-Large model was substituted with a CA attention module, and dilated convolution was incorporated to improve the model's recognition accuracy and response speed. The CA attention module enhances the perception and feature extraction capabilities of spatial coordinate information in images. Dilated convolution was introduced into the deep network architecture to expand the model's receptive field. The model was trained using a transfer learning approach that partially froze specific convolutional layers. Experimental results on a dataset comprising 10 common tomato leaf disease images and healthy leaf images demonstrated that the unimproved model achieved a recognition accuracy of 90.11% and an F1 score of 89.98%. After replacing the SE attention module with the CA attention module, the model's accuracy increased to 91.15%, with the F1 score rising to 91.08%. Furthermore, introducing the dilated convolution model improved the accuracy to 94.33% and the F1 score to 94.22%, while maintaining a parameter count of 2.79×10$^6$ and a validation set operation time of 11.76 seconds. Compared to other traditional lightweight models, this model exhibits significant advantages. The field test results show that the detection accuracy rate is 88.79% and the omission rate is 8.44%, which has practical application value. The DC-CA-MobileNetV3 tomato leaf disease recognition model proposed in this study can accurately and efficiently identify tomato leaf diseases, featuring a small number of parameters and ease of deployment in embedded systems.*

## 摘要

*针对番茄叶片病害识别模型的准确度低、响应速度慢的问题，提出了一种改进的轻量级番茄叶片病害识别模型，将 MobileNetV3-Large 模型中的 SE 注意力模块替换为 CA 注意力模块，并引入空洞卷积，以提高模型识别准确度与响应速度。利用 CA 注意力模块提高对图像空间坐标信息的感知能力和特征提取能力。在深层网络中引入空洞卷积模型，扩大模型感受野。使用只冻结部分卷积层的迁移学习方法对模型进行训练。在常见的 10 种番茄叶片病害图像与健康叶片图像构成的数据集上的实验结果表明，未改进的模型识别精准率为 90.11%，F1 值为 89.98%；改为 CA 注意力模块后，模型的精准率提高至 91.15%，F1 值提高至 91.08%；在此基础上引入空洞卷积模型后，精准率提高至 94.33%，F1 值提高至 94.22%，其模型参数量为 2.79×106，验证集运行耗时 11.76s，与其他传统轻量化模型相对比具有明显优势。在实际温室场地进行实地试验验证，检测准确率达 88.79%，漏检率为 8.44%，具有实际应用价值。本文提出的 DC-CA-MobileNetV3 番茄叶片病害识别模型可精准、高效地识别番茄叶片病害，同时具有参数量小、易搭载至嵌入式系统的优点。*

## INTRODUCTION

China ranks among the leading countries in global tomato production, and modern cultivation techniques such as greenhouse planting have significantly enhanced both the yield and quality of tomatoes. Nevertheless, during their growth cycle, tomatoes are susceptible to various diseases, including leaf mildew, yellow leaf curl virus, and powdery mildew. The absence of timely detection and intervention can severely compromise tomato yield and quality *(Zhang et al., 2020)*.

*Cheng Chi, M.S. Stud. Eng.; LiFeng Qin, Associate professor, Ph.D. Eng.*

However, prompt identification of diseases and implementation of preventive and control measures can effectively mitigate disease propagation, which holds substantial significance for enhancing tomato yield and quality while minimizing economic losses *(Tian et al., 2021)*.

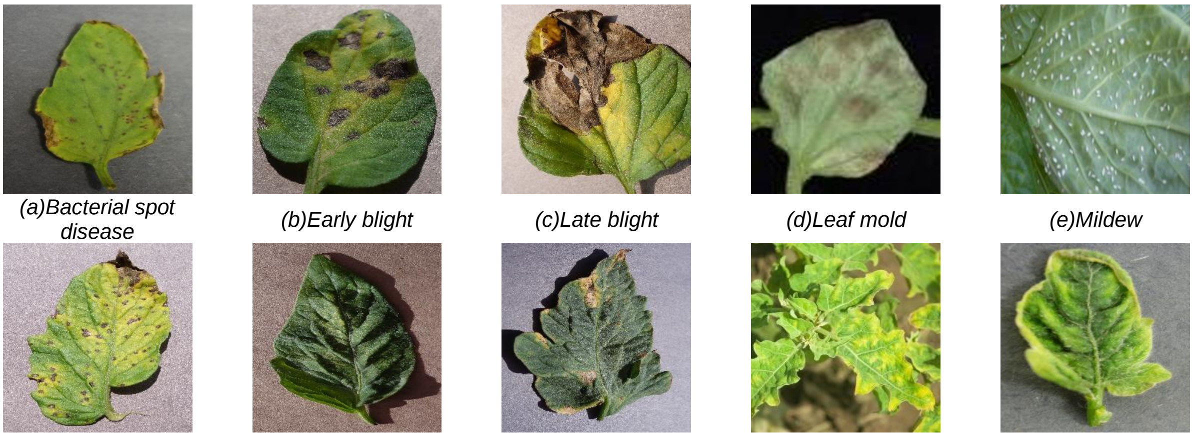
Vision-based technology for the detection and diagnosis of vegetable diseases has garnered significant attention. Xu developed an intelligent tomato disease diagnosis model using Bayesian optimization with LightGBM, preprocessed raw prescription data, and further extracted features from crop disease prescription data through a Wrapper-based recursive feature elimination method, achieving a comprehensive diagnostic accuracy of 89.11% *(Xu et al., 2022).* Sladojevic employed a deep CNN model along with the Stanford background dataset to classify plant diseases, identifying 13 distinct types of plant diseases from healthy leaves and distinguishing plant leaves from their surrounding environment. The model's accuracy ranged between 91% and 98%, with an average accuracy of 96.3% for individual class tests *(Sladojevic et al., 2016)*. Brahimi integrated AlexNet and GoogleNet models within a CNN framework, leveraging transfer learning and fine-tuning mechanisms to classify tomato leaf diseases, achieving accuracies of 98.66% and 99.18%, respectively *(Brahimi et al., 2017)*. Jiang utilized ResNet-50 to identify several tomato leaf diseases - primarily late blight, leaf mold, and yellow leaf curl virus - achieving a detection accuracy of 98% after multiple training iterations *(Jiang et al., 2020)*. Han employed infrared thermal imaging in conjunction with an improved version of YOLOv5 for the early detection of crop diseases. Their model achieved a mean Average Precision (mAP) exceeding 90% across various temperature gradients, enabling more precise and rapid identification of early-stage diseases *(Han et al., 2023)*. Ullah proposed the integration of EfficientNetB3 and MobileNet to detect tomato leaf diseases, achieving a detection success rate of 99.92% *(Ullah et al., 2023)*. Yang optimized the YOLOv5s-based crop yellowing and leaf bending detection model by applying trunk replacement, model pruning, and knowledge distillation techniques. Their approach reduced memory usage by 90% while maintaining an average accuracy reduction of less than 3% *(Yang et al., 2023)*.

The research on tomato leaf disease detection using deep learning models addresses the limitations of traditional manual inspection methods, significantly enhancing both the efficiency and accuracy of detection. However, several challenges remain. First, most deep learning models are characterized by their large size and complex architecture, requiring a substantial number of parameters to achieve high recognition accuracy. This results in heavy computational demands, poor real-time performance, and difficulties in deploying these models on embedded devices for practical applications. Second, given the diverse range of tomato leaf diseases and the similarity of symptoms among some conditions, there is still potential for improvement in the accuracy of existing recognition models.

To address the aforementioned issues, this paper proposes an enhanced algorithm based on the MobileNetV3 model, which is specifically designed for the detection of leaf diseases in tomato greenhouses. By leveraging transfer learning, the model is further optimized to extract both feature representations and their corresponding location information. This not only provides valuable data resources for subsequent inspection tasks but also effectively addresses the challenges of limited deployment on embedded devices.

## MATERIALS AND METHODS

### Data set construction



*(a)Bacterial spot disease* *(b)Early blight* *(c)Late blight* *(d)Leaf mold* *(e)Mildew*

*(f)Spot blight* *(g)Tetranychus* *(h)Target disease* *(i)Mosaic virus* *(j)Yellow leaf curl disease*

**Fig.1 - Sample images of different tomato leaf diseases**

In this study, tomato leaves were selected as the research subject, with healthy leaves and 10 common tomato leaf diseases chosen to construct a dataset. All images were sourced from a dataset published on Kaggle's official website. This dataset has been classified by domain experts, ensuring high reliability, which facilitates deep learning models in performing classification tasks effectively. Additionally, the original images underwent augmentation through rotation and mirroring techniques to expand the dataset size and enhance model generalization. Our dataset primarily comprises 10 types of tomato diseases, including bacterial spot, early blight, late blight, leaf mold, mildew, spot blight, spider mite damage, target spot, mosaic virus, yellow leaf curl virus, and healthy leaf images. A total of 32,534 images were split into training and validation sets at an 8:2 ratio. Disease names, corresponding labels, and sample counts are presented in Table 1, while partial image samples are illustrated in Figure 1.

**Table 1**

**Tomato leaf disease dataset composition**

| Label | Name of tomato leaf diseases | Number of sample images/piece |
|---|---|---|
| 0 | Healthy leaf | 3856 |
| 1 | Bacterial spot disease | 3558 |
| 2 | Early blight | 3098 |
| 3 | Late blight | 3905 |
| 4 | Leaf mold | 3493 |
| 5 | Mildew | 1256 |
| 6 | Spot blight | 3628 |
| 7 | Tetranychus | 2182 |
| 8 | Target disease | 2284 |
| 9 | Mosaic virus | 2737 |
| 10 | Yellow leaf curl disease | 2537 |

## Tomato leaf disease recognition model

(1) Lightweight network MobileNetV3 model

MobileNetV3 is a lightweight deep learning model introduced by the Google research team in 2019, and characterized by a reduced number of parameters, lower computational requirements, and shorter inference times *(Shi et al., 2024)*, which makes it particularly suitable for deployment on mobile and embedded devices. Building upon the MobileNetV2 architecture *(Sandler et al., 2018)*, MobileNetV3 incorporates the Squeeze-and-Excitation (SE) attention mechanism to enhance feature representation capabilities. Additionally, it introduces the h-swish activation function, which offers computational efficiency while maintaining high performance. The tail network structure has also been further optimized to improve overall accuracy *(Howard et al., 2020)*. MobileNetV3 provides two variants—MobileNetV3-Large and MobileNetV3-Small—to accommodate diverse hardware configurations and performance needs *(Wu 2024)*. This study focuses on the MobileNetV3-Large model, with its structural parameters detailed in Table 2.

**Table 2**

**MobileNetV3-Large structure**

| Input | Operations | SE | Activation function | Stride |
|---|---|---|---|---|
| $224^2$×3 | Conv2d | — | HS | 2 |
| $112^2$×16 | Bneck,3×3 | — | RE | 1 |
| $112^2$×16 | Bneck,3×3 | — | RE | 2 |
| $56^2$×24 | Bneck,3×3 | — | RE | 1 |
| $56^2$×24 | Bneck,5×5 | √ | RE | 2 |
| $28^2$×40 | Bneck,5×5 | √ | RE | 1 |
| $28^2$×40 | Bneck,5×5 | √ | RE | 1 |
| $28^2$×40 | Bneck,3×3 | — | HS | 2 |
| $14^2$×80 | Bneck,3×3 | — | HS | 1 |

| Input | Operations | SE | Activation function | Stride |
|---|---|---|---|---|
| 14²×80 | Bneck,3×3 | — | HS | 1 |
| 14²×80 | Bneck,3×3 | — | HS | 1 |
| 14²×80 | Bneck,3×3 | √ | HS | 1 |
| 14²×112 | Bneck,3×3 | √ | HS | 1 |
| 14²×112 | Bneck,5×5 | √ | HS | 2 |
| 7²×160 | Bneck,5×5 | √ | HS | 1 |
| 7²×160 | Bneck,5×5 | √ | HS | 1 |
| 7²×160 | Conv2d,1×1 | — | HS | 1 |
| 7²×960 | Pool,7×7 | — | — | 1 |
| 1²×160 | Conv2d,1×1, NBN | — | HS | 1 |
| 1²×1280 | Conv2d,1×1, NBN | — | — | 1 |

(2) Improved MobileNetV3 model

Since the existing MobileNetV3 network model exhibits limited sensitivity to location information and insufficient feature learning capability for similar diseases, this paper proposes an enhanced approach for this network model. The improved network architecture is illustrated in Figure 2. First, the SE (Squeeze-and-Excitation) attention module in the bottleneck layer of the network is substituted with the CA attention module, enabling the network to better capture image location information, which facilitates further processing of such information. Second, dilated convolution is incorporated into the last two bottleneck layers, effectively expanding the receptive field without increasing computational complexity, thereby enhancing feature extraction.

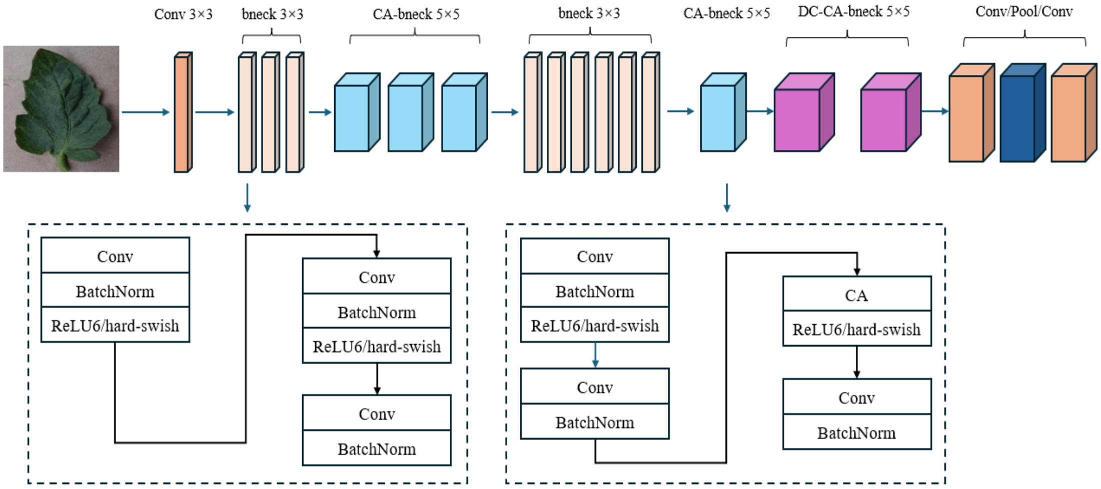


**Fig. 2 - Structure diagram of improved MobileNetV3**

Since the SE module primarily emphasizes the relationship between channels *(Li et al., 2020)* without accounting for spatial position information, its capability to acquire image information of tomato disease leaves is constrained. The CA attention module integrates decomposed coordinate coding to embed the diseased area's coordinate information into the feature map, thereby enhancing the model's spatial perception of the diseased region. Consequently, the CA attention module, which captures positional information, was employed to replace the SE attention module. This substitution only marginally increased the computing resources required and facilitated seamless integration into mobile inspection devices.

The CA attention mechanism effectively captures image features by modeling inter-channel relationships and long-range spatial dependencies *(Hou et al., 2021)*. Specifically, channel attention is decomposed into two one-dimensional feature encoding processes. Global average pooling of input feature maps is performed along both the vertical and horizontal directions to generate perception maps in each direction. Subsequently, these perception maps are combined, and a series of convolutional layers along with nonlinear activation functions are employed to produce a pair of direction-aware and position-sensitive attention maps. Ultimately, the generated attention maps are applied to the input feature maps via element-wise multiplication, thereby enhancing the representation of the object of interest while preserving accurate positional information *(Zheng et al., 2024)*.

The calculation process of the CA attention mechanism primarily consists of three stages *(Gu et al., 2025)*. Given an input feature map of size C×H×W, where C denotes the number of channels, H represents the height, and W indicates the width of the feature map, the CA module initially applies two pooling kernels, (H,1) and (1,W), to perform global average pooling along the horizontal and vertical directions, respectively. This operation yields two feature maps of dimensions C×H×1 and C×1×W, which are subsequently concatenated to produce a feature map of size C×(H+W). The concatenated feature map is then transformed via 1×1 convolution, followed by batch normalization and a nonlinear activation function. Subsequently, the transformed feature map is split into horizontally and vertically independent feature maps, which undergo another 1×1 convolution operation. The resulting feature maps have the same number of channels as the input feature map. Finally, the Sigmoid activation function is applied to generate the attention map, which is used to emphasize feature regions through element-wise multiplication, thereby producing the final output feature map. Figure 3 illustrates the detailed calculation process of the CA attention module.

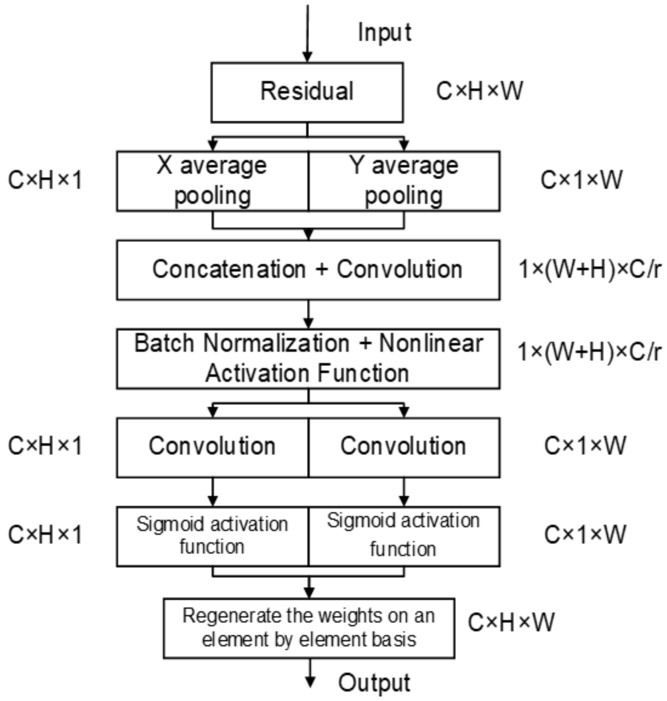

Input
Residual
C×H×W
C×H×1
X average pooling
Y average pooling
C×1×W
Concatenation + Convolution
1×(W+H)×C/r
Batch Normalization + Nonlinear Activation Function
1×(W+H)×C/r
C×H×1
Convolution
Convolution
C×1×W
C×H×1
Sigmoid activation function
Sigmoid activation function
C×1×W
Regenerate the weights on an element by element basis
C×H×W
Output

**Fig. 3 - Flowchart of calculation of coordinate attention**

Dilated Convolution is a technique that incorporates intervals into the convolution kernel, thereby expanding the receptive field of the kernel without increasing computational complexity or introducing additional parameters *(Kumar et al., 2022)*. This facilitates more effective extraction of image features and enhances model learning and training. Given the diversity of tomato leaf diseases and the similarity among different diseases, it is crucial to preserve detailed texture information of disease spots. By leveraging its sparse sampling and resolution-preserving mechanism, dilated convolution effectively addresses detail loss issues caused by traditional down-sampling operations. Furthermore, it establishes correlations between the pathology of lesions and the entire leaf through a large receptive field, achieving a balance between detail enhancement and global perception in multi-scale feature fusion.

Figure 4 illustrates the cavity convolution expansion process. The critical aspect of this process involves the expansion coefficient r, which determines the number of intervals between elements within the convolution kernel. This mechanism enables the establishment of convolution kernels of varying sizes across different bottleneck layers, facilitating the extraction of image features at a deeper level and enhancing their alignment with the target object. In this study, cavity convolution is incorporated into the last two bottleneck modules of the MobileNetV3 network, with expansion coefficients *r* set to 2 and 4, respectively. The corresponding convolution kernels are also depicted in Figure 4.

Initially, cavities are inserted into the convolution kernel based on the specified expansion coefficient, followed by the computation of feature graph convolution. Finally, the resulting feature graph is output. The size of the feature map can be determined using the following formula:

$$\mathrm{o}=\left[\frac{i+2p-(k-1)r+1}{s}\right]+1 \tag{1}$$

$o$ - Output the dimension (width or height) of the feature map.
$i$ -- Input the dimension (width or height) of the feature map.
$p$ --The extent of edge padding
$k$ --The size of the convolution kernel
$r$ --Coefficient of expansion
$s$ -- Stride. In a hollow convolution, the value is consistently set to 1.

The initial convolution kernel size is set to 3×3. When the expansion coefficient r=2, the convolution kernel expands to 5×5; when r=4, it further expands to 9×9. It is evident that as the expansion coefficient r increases, both the size of the convolution kernel and its receptive field grow significantly, without introducing any additional operational parameters.

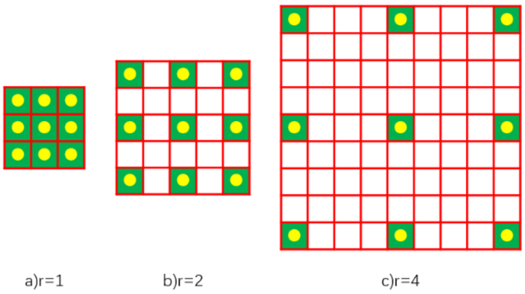


**Fig. 4 - Schematic diagram of dilated convolution expansion**

(3) Transfer Learning

Transfer learning *(Jain et al., 2021; Li et al., 2023)* is a machine learning technique that enables the adaptation of a pre-trained model to a new yet related problem, thereby enhancing the model's performance. By leveraging transfer learning, the amount of annotated data required for new tasks can be significantly reduced. In scenarios where existing data resources are limited, transfer learning can significantly boost model accuracy and generalization. This approach enables the model to leverage knowledge from related tasks, thereby learning more robust and transferable features. Consequently, this paper employs transfer learning to optimize the model for deployment and usage. Specifically, MobileNetV3 pre-trained on the ImageNet dataset serves as the source domain model, while the target domain is the tomato leaf disease dataset constructed in this study. To fine-tune the model, a transfer learning optimization strategy involving the freezing of selected convolutional layers is adopted.

The architecture of MobileNetV3 was optimized through techniques such as cavity convolution, coordinate attention module embedding, and other methods. Subsequently, the features of the ImageNet pre-trained model were transferred. The model was trained using a feature transfer approach based on freezing shallow common features and a model transfer approach involving fine-tuning of deep parameters. As a result, the tomato disease recognition model DC-CA-MobileNetV3 was developed.

## RESULTS AND ANALYSIS

### Experimental platform

The cloud platform provided by Featurize was utilized in this study. The hardware configuration comprised a 6-core Intel E5-2680 v4 processor, an NVIDIA RTX 3060 graphics card with 12GB of memory, and 32GB of system RAM. The software environment included Python 3.7 and the deep learning framework Tensorflow 2.7.

### Performance index

Precision (P), Recall (R), mean Average Precision (mAP), and F1 Score (F1) *(Sun et al., 2025; Chen et al., 2024)* were employed as evaluation metrics to assess the performance of the DC-CA-MobileNetV3 model proposed in this study.

The formulae for precision (P) and recall (R) are:

$$P = \frac{TP}{TP + FP} \times 100\% \tag{2}$$

$$R = \frac{TP}{TP + FN} \times 100\% \tag{3}$$

In the context of mean Average Precision (mAP), P represents the accuracy rate, AP denotes the Average Precision for a single label (the average of the maximum precision values at each recall level), and mAP signifies the mean of the Average Precision across all labels.

The formula is as follows:

$$AP = \frac{\sum P}{N\left(TotalImages\right)} \tag{4}$$

$$mAP = \frac{\sum AP}{N\left(Classes\right)} \tag{5}$$

Taking bacterial spot disease as an example, TP denotes the count of samples that correctly identify bacterial spot disease among all samples; FP indicates the count of samples that erroneously classify another category or background as bacterial spot disease; FN refers to the count of bacterial spot disease samples misclassified as other categories or unidentified. AP represents the average of the highest precision values achieved for a single disease species across varying recall rates, while mAP signifies the mean value of AP across all disease categories.

The F1 score is a metric employed to comprehensively assess the performance of a classification model. It integrates the precision rate (P) and the recall rate (R), serving as their harmonic mean. This metric allows for the evaluation of model performance while maintaining equilibrium between precision and recall, and its calculation formula is as follows:

$$F1 = \frac{2PR}{P + R} \tag{6}$$

### A Comparative Experiment on Transfer Learning Methods

Transfer learning can be categorized into three approaches, namely: Full Migration, which involves freezing all convolutional layers and training only the fully connected layer; Reuse Model, which employs only the model architecture without utilizing pre-trained parameters; and Fine-Tuning, which freezes only a portion of the convolutional layers *(Cui et al., 2023)*. In this study, the Fine-Tuning method was adopted. To evaluate the effectiveness of the transfer learning approach used in this study, the improved MobileNetV3 model was employed as a benchmark, and the same experimental dataset was utilized. The three models were trained using the aforementioned transfer learning methods, respectively, and a tomato disease classification experiment was conducted. The results are presented in Table 3, while the training and validation accuracy curves for the three transfer learning methods are illustrated in Figure 5.

**Table 3**

**Experimental results of test-set of the three migration methods**

| **Transfer Methodology** | **Precision/%** | **Recall/%** | **F1 Score/%** |
|---|---|---|---|
| Fine tuning | **94.33** | **94.12** | **94.22** |
| Full migration | 93.59 | 92.94 | 93.28 |
| Reuse model | 93.88 | 93.69 | 93.78 |

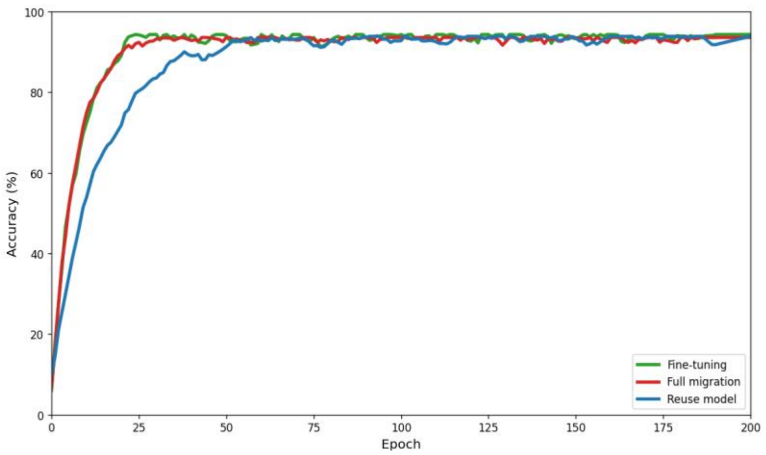


**Fig. 5 - Comparison of accuracy of three migration methods**

## Ablation Study

To validate the effectiveness of the improvement measures, an ablation study was designed. Specifically, the CA attention module and cavity convolution techniques were integrated into the original model to enhance its performance. A classification experiment was subsequently conducted to compare the results, as detailed in Table 4.

**Table 4**

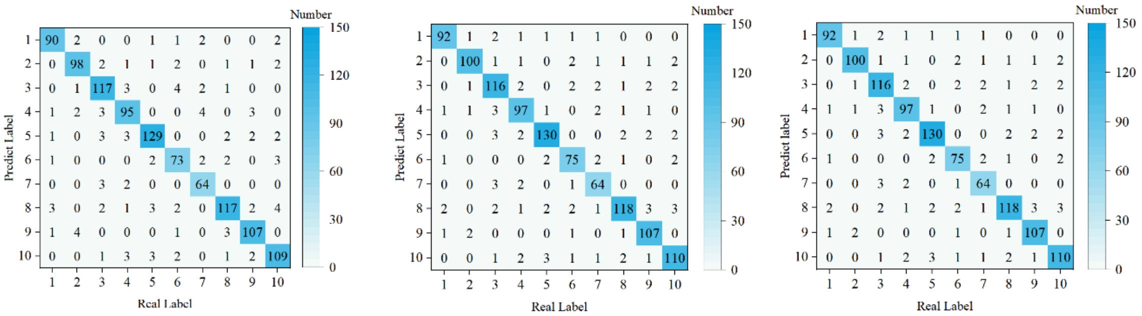
**MobileNetV3-based ablation experiments**

| Serial Number | CA | DC | Precision (%) | Recall (%) | mAP (%) | F1 Score (%) |
|---|---|---|---|---|---|---|
| 1 | | | 90.11 | 89.86 | 90.02 | 89.98 |
| 2 | √ | | 91.15 | 91.02 | 91.34 | 91.08 |
| 3 | | √ | 92.36 | 92.30 | 91.91 | 92.33 |
| 4 | √ | √ | **94.33** | **94.12** | **93.64** | **94.22** |

As shown in Table 4, the introduction of the CA attention module and the cavity convolution module individually led to a certain degree of improvement in model performance. When both methods were combined, the overall effect was optimal. The accuracy of tomato leaf disease identification reached 94.33%, and the F1 score reached 94.22%, surpassing the results obtained by either method alone as well as those of the unimproved model. Simultaneously, the improved model contains 2.79×10$^6$ parameters, enabling deployment on mobile devices. The increase in required training time is minimal, satisfying the demands for immediate application. In conclusion, the enhanced DC-CA-MobileNetV3 model demonstrates superior comprehensive performance compared to the original model.

## A Comparative Study on Various Attention Mechanisms

In this study, the original SE module in MobileNetV3 was replaced with the CA module, which is capable of recognizing spatial positions. To compare and validate the performance of the CA module, MobileNetV3 models incorporating SE, ECA, and CA modules were evaluated under identical experimental data, model architecture, and training conditions. Under these consistent conditions, a fine-tuned transfer learning approach was employed for training over 200 epochs. Figure 6 presents the confusion matrix of classification results under different attention mechanisms, where the values on the main diagonal indicate the number of correctly classified samples. Detailed experimental data are provided in Table 5.



a) Use the SE attention module b) Use the ECA attention module c) Use the CA attention module

**Fig. 6 - The confusion matrix of the MobileNetV3 model using different attention modules**

Table 5

**Comparison experiments of different attention mechanisms based on MobileNetV3**

| Methodology | P<br>[%] | R<br>[%] | mAP<br>[%] | F1 Score<br>[%] |
|---|---|---|---|---|
| SE | 90.20 | 89.97 | 90.27 | 90.08 |
| CA | 91.15 | 91.02 | 91.34 | 91.08 |

## Comparative experiment of different models

To validate the effectiveness and practicality of the model proposed in this study, comparative experiments were conducted with ShuffleNetV2 *(Ma et al., 2018)*, AlexNet *(Krizhevsky et al., 2012)*, VGG-16 *(Simonyan et al., 2015)*, and ResNet-50 *(He et al., 2014),* respectively.

Among these models, ShuffleNetV2 is a classic lightweight convolutional neural network that shares similar applications with MobileNetV3. Additionally, AlexNet, VGG-16, and ResNet-50 have gained widespread recognition for their performance in visual classification tasks and thus provide valuable references for comparison. Under identical experimental datasets and model parameter configurations, the proposed model was evaluated against the four aforementioned models with the number of epochs set to 200. The results are presented in Table 6.

Table 6

**Experiments of five models on test set**

| Name | Parameter number/number | Time / s | P / % | R / % | mAP / % | F1 Score / % |
|---|---|---|---|---|---|---|
| AlexNet | 6×10$^7$ | 18.72 | 91.62 | 91.42 | 90.03 | 92.52 |
| ShuffleNetV2 | **1.31×10$^6$** | 19.5 | 92.92 | 92.19 | 92.33 | 92.55 |
| VGG-16 | 1.39×10$^8$ | 35.34 | 92.52 | 92.25 | 92.16 | 92.38 |
| ResNet-50 | 2.21×10$^7$ | 24.12 | 93.56 | 93.28 | 93.27 | 93.42 |
| Textual method | 2.79×10$^6$ | **11.76** | **94.33** | **94.12** | **93.64** | **94.22** |

As shown in Table 6, the accuracy P of the enhanced MobileNetV3 network model proposed in this paper reaches 94.33%, which is respectively 2.71%, 1.41%, 1.81%, and 0.77% higher than that of the other four networks. This improvement can be attributed to the incorporation of CA (Coordinate Attention) and void convolution mechanisms, which enable the model to learn image features more effectively. The number of parameters in the model is only marginally higher than that of the lightweight ShuffleNetV2 model, while its recognition time is significantly shorter at 11.76 seconds, representing a reduction of 7.74 seconds compared to ShuffleNetV2. This performance ensures the real-time and high-efficiency requirements necessary for mobile devices and enables the processing of video stream image information. Additionally, the F1 score of the model's recognition capability reaches 94.22%, which is the highest among the compared models, indicating an excellent balance between precision and recall. In conclusion, the improved model presented in this study not only enhances recognition accuracy, but also maintains the characteristics of a lightweight architecture and fast response speed, thereby demonstrating its strong application potential.

## Mobile Deployment and Experimentation of the Model

To verify the performance of the improved tomato leaf disease recognition model in practical scenarios, it was applied to the ROS framework and mounted on a patrol device to conduct a patrol experiment for tomato leaf diseases.

The ROS system selected the ROS2 Foxy version, where a workspace and module package were created, and a detection node was designed. The camera image topic /camera/image_raw was subscribed as the model input, and the labeled image /detection_results was used as the model output. The package.xml and setup.py and other dependency files were modified, and the module package was compiled using the colcon command.

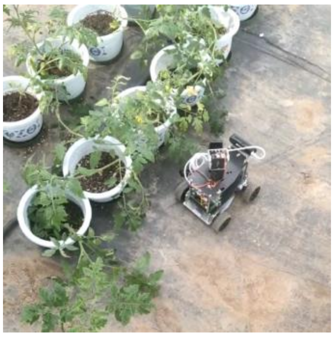


**Fig. 7 - Inspection device for field experiment drawing**

In the experimental glass greenhouse in Wuquan Town, Yangling District, Xianyang City, Shaanxi Province, China, an Ackermann intelligent car equipped with a Raspberry Pi 4B development board was used as the patrol device, as shown in Figure 7.

A total of 121 tomato leaf images were collected to form the target detection validation set. The improved DC-CA-MobileNetV3 model combined with the SSD algorithm module was used to detect tomato leaf disease targets, and the results are shown in Figure 8, where only the rectangular boxes with a confidence level greater than 0.8 are retained. The comparison of the average detection accuracy and missed detection rate of the improved model and the original model on 121 images is shown in Table 7.

**Table 7**

**Comparison of model detection results before and after improvement**

| Detection method | Precision / (%) | Omission Rate / (%) |
|---|---|---|
| Before improvement | 84.55% | 10.29% |
| After improvement | 88.79% | 8.44% |

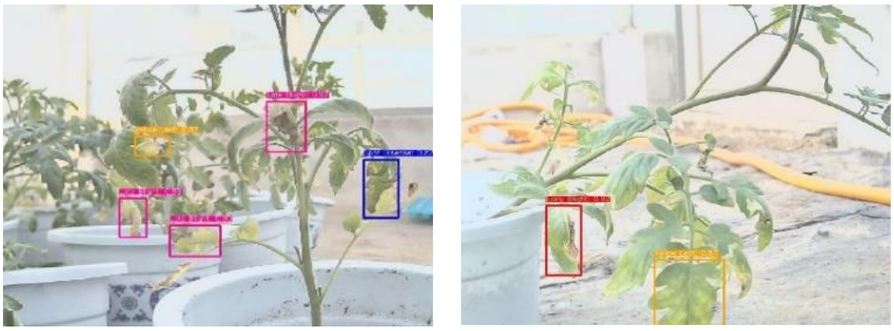


**Fig. 8 - Model testing results in the actual scenario**

## CONCLUSIONS

In this study, we proposed an enhanced MobileNetV3-based model for tomato leaf disease recognition. Specifically, the CA attention module is integrated to replace the SE attention module in the original architecture, enabling more effective extraction of image target feature location information and facilitating subsequent localization tasks on mobile recognition devices. Additionally, dilated convolution is incorporated to expand the model's receptive field and improve its feature recognition capabilities. Experimental results demonstrate that the proposed DC-CA-MobileNetV3 model achieves a recognition accuracy of 94.33%, a recall rate of 94.12%, an average accuracy of 93.64%, and an F1 score of 94.22%, with a parameter count of $2.79\times10^6$ and an operation time of 11.76 seconds. These performance metrics indicate that the model satisfies the requirements for stability, rapidity, and accuracy, providing robust technical support for the development of intelligent monitoring and inspection systems for tomato leaf disease information. The model was deployed to mobile devices in a greenhouse environment. In actual environmental scenarios, the accuracy rate reached 88.75%, and the missed detection rate was 8.44%, demonstrating practical application value.

In practical application scenarios, factors such as illumination and occlusion can interfere with the performance of the tomato leaf disease recognition model. In future work, we will further optimize the model architecture by incorporating training with images captured under diverse complex conditions, thereby improving accuracy in real-world scenarios and enhancing network robustness.

## ACKNOWLEDGEMENT

This research was supported by the Agricultural Science and Technology Innovation and Research Project of Shaanxi Province (Project No.: 2020NY-101 and 2024NC-ZDCYL-05-07).

## REFERENCES

[1] Brahimi M., Boukhalfa K., Moussaoui A. (2017), Deep Learning for Tomato Diseases: Classification and Symptoms Visualization. *Applied Artificial Intelligence,* 31(4-6): 1-17. Algeria. DOI: 10.1080/08839514.2017.1315516.

[2] Chen Z. C. (2024). *Research on tomato leaf disease recognition based on deep learning* (基于深度学习的番茄叶片病害识别研究) [Master's dissertation, Heilongjiang University];

[3] Cui J. R., Wei W. Z., Zhao M. (2019). Rice disease recognition model based on improved MobileNetV3 (基于改进 MobileNetV3 的水稻病害识别模型). *Transactions of the Chinese Society for Agricultural Machinery*, 54(11): 217-224+276, Guangdong/China;

[4] Gu R., Gu J. L., Song C. L.,Qian C. H. (2019). Multi-scale lightweight apple leaf disease recognition model based on coordinate attention (基于坐标注意力的多尺度轻量级苹果叶片病害识别模型). *Chinese Journal of Agricultural Mechanization*, 46(02): 173-180+186, Jiangsu/China;

[5] Han X., Xu Y. X., Feng R. Z., Liu T. X., Bai J. B., Lan Y. B. (2019). Early identification of crop diseases based on infrared thermal imaging and improved YOLO v5 (基于红外热成像和改进 YOLO v5 的作物病害早期识别). *Transactions of the Chinese Society for Agricultural Machinery*, Shandong/China, 54(12): 300-307+375.

[6] He K., Zhang X., Ren S., Sun J. (2014). Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition.*IEEE Transactions on Pattern Analysis & Machine Intelligence*, 37(9): 1904-16. Shaanxi/China;

[7] Hou Q., Zhou D., Feng J. (2021) Coordinate Attention for Efficient MobileNet work Design. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Singapore,13708-13717.

[8] Howard A., Sandler M., Chen B., Wang W., Chen L. C., Tan M., Chu G., Vasudevan V., Zhu Y., Pang R. (2020). Searching for MobileNetV3. *IEEE/CVF International Conference on Computer Vision (ICCV)*, Korea (South), DOI:10.1109/ICCV.2019.00140.

[9] Jain S., Gour M. (2021). Tomato plant disease detection using transfer learning with C-GAN synthetic images.*Computers and Electronics in Agriculture*, 187(2021), India; DOI: 10.1016/j.compag.2021.106279

[10] Jiang, D.; Li, F.; Yang, Y.; Yu, S. (2020). A tomato leaf diseases classification method based on deep learning. *In Proceedings of the Chinese Control and Decision Conference*, 22–24 August, pp.1446–1450. Hefei/China;

[11] Krizhevsky A., Sutskever I., Hinton G.E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, USA;

**[12]** Kumar V., Singh R., Dua Y. (2022). Morphologically dilated convolutional neural network for hyperspectral image classification. *Signal Processing: Image Communication*, 101: ID 116549. India; DOI: 10.1016/j.image.2021.116549

[13] Li J., Wang C., Ma Z. Y., Wang G. W., Liao C. Y., Wang H.R., Guan .L. (2023). MobileNet-CAL: Classification of tomato pests and diseases based on transfer learning and attention mechanism (MobileNet-CAL：基于迁移学习和注意力机制的番茄病虫害分类方法). *Journal of Jilin University (Engineering and Technology Edition)*, pp. 1-9, Jilin/China;

[14] Li X., Shen X., Zhou Y., Wang X., Li T. Q. (2020). Classification of breast cancer histopathological images using interleaved DenseNet with SENet (IDSNet). *PLoS ONE*, 15(5), UK, DOI: 10.1371/journal.pone.0232127

[15] Ma N. N., Zhang X. Y., Zheng H. T.,Sun J. (2018). ShuffleNet V2: Practical Guidelines for Efficient CNN Architecture Design.*Springer, Cham*, Beijing/China. DOI:10.1007/978-3-030-01264-9_8.

[16] Sandler M.., Howard A., Zhu M., Zhmoginov A., Chen L. C. (2018).MobileNetV2: Inverted Residuals and Linear Bottlenecks. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, USA, DOI:10.1109/CVPR.2018.00474.

[17] Shi B. M., He Y. X., Zhao X. (2024). Identification method of maize disease by migrating MobileNetV3 (迁移 MobileNetV3 的玉米病害识别方法). *Journal of Ningxia University (Natural Science Edition)*,Gansu/China, 1-8.

[18] Simonyan K., Zisserman A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. *International Conference on Learning Representations.Computational and Biological Learning Society*, UK.

[19] Sladojevic, S.; Arsenovic, M.; Anderla, A.; Culibrk, D. (2016), Stefanovic, D. Deep neural networks based recognition of plant diseases by leaf image classification. *Computational Intelligence and Neuroence*, Republic of Serbia, (3). DOI:10.1155/2016/3289801.

[20] Sun S. S., Li X. K. , Zhang H. L., He L., Zhao L. (2025). Tomato leaf pest detection model with embedded platform (嵌入式平台的番茄叶片病虫害检测模型). *Computer Engineering and Applications*, 1-12. Xinjiang/China;

[21] Tian P. J., Li S. J., Zhang Y., Liu Y., Zhang S. B., Zhang D. Y.(2021), Detection and genetic variation analysis of leaf curl virus in infected tomato (侵染番茄的曲叶病毒检测及遗传变异分析). *Chinese Vegetables* , (05):28-32, Hunan/China;

[22] Ullah, Z., Alsubaie, N., Jamjoom, M., Alajmani, S.H., Saleem, F. (2023). EffiMob-Net: A Deep Learning-Based Hybrid Model for Detection and Identification of Tomato Diseases Using Leaf Images. *Agriculture*, Kingdom of Saudi Arabia, 13, 737.

[23] Wu X. X.(2024). *Research on rice disease image recognition method based on improved MobileNet* (基于改进 MobileNet 的水稻病害图像识别方法研究) [Master's dissertation, Heilongjiang Bayi Agricultural University].

[24] Xu C., Ding J. Q., Zhao D. T., Qiao Y., Zhang L. X. (2022). Tomato disease diagnosis method based on LightGBM and prescription data (基于 LightGBM 和处方数据的番茄病害诊断方法). *Transactions of the Chinese Society for Agricultural Machinery*, 53(09):286-294, Beijing/China;

[25] Yang J.H., Zuo H.X., Huang Q.C., Sun Q., Li S.E., Li L. (2023). Lightweight method of crop leaf disease detection model based on YOLO v5s (基于 YOLO v5s 的作物叶片病害检测模型轻量化方法). *Transactions of the Chinese Society for Agricultural Machinery*, 54(S1):.222-229 Beijing/China,.

[26] Zhang C.C., Yuan S.K., Huo K.K. (2020). Factors affecting tomato production efficiency (影响番茄生产效益的因素). *Hubei Agricultural Mechanization*, (12):33-34, Henan/China;

[27] Zheng C. J., Li S.B., Pu R.Q., Zhang T. (2019). Tomato leaf disease recognition based on lightweight convolutional neural network (基于轻量化卷积神经网络的番茄叶片病害识别). *Jiangsu Agricultural Sciences*, 52(11): 225-231, Guizhou/China.