Tomato Leaf Disease Recognition Based on Improved Lightweight Convolutional Neural Network MobileNetV3 (Postprint)

Authors: Zhou Qiaoli, Ma Li, Cao Liying, Yu Helong

Date: 2023-02-17T00:00:00+00:00

Abstract

Timely detection of tomato diseases can effectively improve the quality and yield of tomatoes. To achieve real-time non-destructive detection of tomato diseases, this study proposes a tomato leaf disease classification and recognition method based on improved MobileNetV3. First, the lightweight convolutional neural network MobileNetV3 is selected for pre-training on the ImageNet dataset, and the shared parameters obtained from pre-training are transferred to the model for tomato leaf disease recognition and fine-tuned. Using the same training method, transfer learning is also performed on three deep convolutional network models—VGG16, ResNet50, and Inception-V3—for comparison. The results show that MobileNetV3 achieves the best overall learning performance, with an average test recognition accuracy of 94.68% for 10 classes of tomato diseases under Mixup data augmentation and the focal loss function. Building upon transfer learning, the MobileNetV3 model is further improved by introducing dilated convolutions and perceptron structures into the convolutional layers and adopting the GLU (Gated Liner Unit) gating mechanism activation function. The trained optimal tomato disease recognition model achieves an average test recognition accuracy of 98.25%, with a model size of 43.57 MB and a detection time of only 0.27 s per tomato disease image. Through 10-fold cross-validation, the model demonstrates good robustness. This study can provide theoretical foundation and technical support for real-time detection of tomato leaf diseases.

Full Text

Identification of Tomato Leaf Diseases Based on Improved Lightweight Convolutional Neural Networks MobileNetV3

ZHOU Qiaoli, MA Li*, CAO Liying, YU Helong*
(College of Information Technology, Jilin Agricultural University, Changchun

130118, China)

Abstract

Timely detection of tomato diseases can effectively improve tomato quality and yield. To achieve real-time, non-destructive detection of tomato diseases, this study proposes a classification and recognition method for tomato leaf diseases based on improved MobileNetV3. First, the lightweight convolutional neural network MobileNetV3 was selected and pre-trained on the ImageNet dataset. The shared parameters obtained from pre-training were then transferred to the tomato leaf disease recognition model and fine-tuned. Under the same training methodology, comparative experiments were conducted with VGG16, ResNet50, and Inception-V3 deep convolutional network models using transfer learning. The results showed that MobileNetV3 achieved the best overall learning performance, with an average test recognition accuracy of 94.68% for ten types of tomato diseases under Mixup hybrid augmentation and focal loss. Building upon transfer learning, the MobileNetV3 model was further improved by introducing dilated convolution and perceptron structures in the convolutional layers and adopting the Gated Linear Unit (GLU) activation function. The resulting optimal tomato disease recognition model achieved an average test recognition accuracy of 98.25%, with a model data size of only 43.57 MB and a detection time of just 0.27 seconds per tomato disease image. Ten-fold cross-validation demonstrated the model' s robustness and good performance. This research provides a theoretical foundation and technical support for real-time detection of tomato leaf diseases.

Keywords: tomato disease identification; convolutional neural networks; transfer learning; MobileNetV3; activation function; identification and classification

---

1 Introduction

During tomato cultivation and fruit production, plants are susceptible to various common diseases due to factors such as weather, temperature, and humidity. These diseases exhibit different damage characteristics at different stages. Leaf diseases mainly include bacterial spot, leaf mold, early blight, late blight, septoria leaf spot, target spot, two-spotted spider mite, mosaic virus, and yellow leaf curl disease. Tomato diseases are diverse and their impact on leaves is complex, requiring early identification of disease types and severity for targeted pesticide treatment. Otherwise, fruit development will be affected, ultimately impacting tomato quality and yield.

Traditional plant leaf disease identification methods rely on manual feature extraction. Wang et al. [2] segmented apple leaf lesions and combined color and texture features with Support Vector Machine (SVM) for disease recognition. Qin et al. [3] extracted color, shape, and texture features of leaf lesions to establish an alfalfa leaf disease recognition model using naive Bayes and linear discriminant analysis. Xia et al. [4] extracted texture and color features of wheat diseases for recognition using SVM. These approaches depend on manual annotation of color, texture, and shape features, resulting in insufficient classification accuracy and poor generalization.

Convolutional neural networks enable automatic feature extraction for crop diseases, with GoogLeNet, AlexNet, and ResNet achieving excellent results. Liu et al. [5] used AlexNet to recognize rice sheath blight with 97% accuracy, but with limited disease categories and insufficient data. Wu [6] adjusted VGG16 and ResNet parameters for corn leaf disease recognition, achieving 93.33% accuracy, but with high image resolution and large model parameters requiring further performance improvement. Ding and Zhou [7] built a convolutional neural network based on AlexNet with transfer learning, achieving 96.18% test accuracy. Chen et al. [8] used data augmentation and transfer learning with improved Inception-V3 for corn disease recognition, reaching 96.6% average accuracy. Waheed et al. [9] proposed an optimized dense convolutional neural network for corn leaf disease recognition with 98.06% accuracy. While combining transfer learning with neural networks improves recognition accuracy, these convolutional neural network algorithms have large parameter counts, high image resolution requirements, and long running times, making them unsuitable for rapid real-time detection on mobile devices.

This study focuses on tomato disease leaves as the primary research object, using the lightweight convolutional neural network MobileNetV3 [10] as the backbone. The model structure was improved based on transfer learning and compared with VGG16, ResNet50, and Inception-V3 models.

2 Materials and Methods

2.1 Experimental Dataset

The experimental dataset consisted of tomato leaf disease images from the publicly available Plant Village dataset [11], which includes plant leaf images of multiple disease categories across different crops, totaling 38 classes categorized by species and disease. The tomato samples included ten common tomato disease leaf types: early blight (1,000 images), late blight (1,909 images), bacterial disease (1,320 images), leaf mold (952 images), septoria leaf spot (1,771 images), target spot (1,404 images), two-spotted spider mite (1,676 images), mosaic virus (373 images), yellow leaf curl disease (5,351 images), and healthy leaves (1,591 images). All images were in *.jpg format and uniformly resized to 64$×$64 pixels.

2.2 Data Augmentation

2.2.1 Conventional Data Augmentation To ensure data balance, sample diversity, and sufficient training samples for the convolutional neural network model [12], batch normalization and data augmentation [13] were applied. Disease images underwent Gaussian noise addition, brightness enhancement, contrast transformation, random cropping, and random rotation to enhance sample diversity and simulate natural environments. The augmented dataset contained 18,521 disease images. All ten disease categories received identical augmentation processing to improve sample quality, quantity, and model generalization capability. Figure 1 [Figure 1: see original paper] illustrates the augmentation process for septoria leaf spot, showing adjustments to angle, brightness, blur, and local lesion detail enhancement.

2.2.2 Mixup Hybrid Augmentation Mixup augmentation [14] generates new hybrid images through linear interpolation of two original images at a specified ratio, producing samples closely resembling real ones. This study applied intra-class Mixup augmentation to tomato samples, increasing diversity while enhancing learning of important features. Mixup synthesizes new images by linearly superimposing feature vectors of two original images, improving model adaptability to out-of-sample predictions and prediction smoothness. During Mixup augmentation, two original images from the same disease category were randomly selected to synthesize new tomato images. Different λ values [15] (0.3, 0.5, and 0.8) were tested, with Figure 2 [Figure 2: see original paper] showing examples at λ ranging from 0 to 1. The Mixup-augmented training set contained 15,974 images. While both methods achieve data expansion, Mixup better increases data diversity.

3 Model Architecture and Improvements

3.1 MobileNetV3 Model

MobileNetV3 builds upon MobileNetV1 [16] and MobileNetV2 [17], combining their advantages to achieve higher efficiency as a lightweight convolutional neural network. Since ReLU functions operate inefficiently in low-dimensional spaces, hindering feature extraction and causing information loss, this study employs linear bottleneck structures that use linear layers instead of ReLU in convolutional layers with few channels to ensure more adequate feature information.

Based on MobileNetV2, MobileNetV3 incorporates the SE (Squeeze-and-Excitation) attention module [18] into bottleneck structures to emphasize prominent features and suppress insignificant ones. The final 1$×1*convolutionallayerwasremoved, withtheaveragepoolinglayermovedforward.Theh−swishactivationfunctionsignificantlyreducescomputationalcost.Thelightweightframeworkrequiresminima*
depthwise convolution.

3.2 Transfer Learning

This study applied transfer learning [19] using VGG16, ResNet50, Inception-V3, and MobileNetV3 for tomato disease image recognition. Each algorithm offers distinct advantages for effective disease classification and detection time savings. First, the large-scale ImageNet dataset [20] served as the source domain for pre-training, with model weights transferred to the tomato disease recognition model and fine-tuned [21] during training. In the transfer process, all convolutional layers were frozen except the final output layer, with the remaining network acting as a feature extractor. Extracted features were input to a classifier for disease prediction, with network output adapted to ten tomato disease categories. Figure 3 [Figure 3: see original paper] illustrates the transfer learning workflow.

3.3 Model Improvements

3.3.1 Multilayer Perceptron A Multilayer Perceptron (MLP) [22] can be embedded as a small network within a deep architecture, functioning as a classifier for categorical output. This study added a 1$×1*convolutionallayerafterthe*5×5*convolutioninbottleneckmo*
linearactivationenhancesnetworkexpressivecapability, *whiletheperceptronstructureenablesfeaturereuse*, *im*
convolutions that reduce parameters and computational cost compared to traditional fully connected layers. Figure 4 [Figure 4: see original paper] shows the MLP structure.

3.3.2 Dilated Convolution Dilated convolution [23] introduces expansion rate r into traditional convolution, expanding the kernel while maintaining parameter count to extract more feature information. Tomato diseases are diverse with similarities between types, and early-stage lesions are small with subtle details. Lesion colors closely resemble healthy tissue, making boundary and contour features difficult to distinguish. Different kernel sizes in bottleneck structures facilitate deeper abstract feature extraction. Therefore, this study introduced dilated convolution in the last two MobileNetV3 bottleneck modules. All initial kernels were 3$×3, *withrvaluesof*1, 2, *and*4*producingreceptivefieldsof*3×3, 5×5, *and*9×$9, respectively. Increasing r significantly enlarges the receptive field without additional computational cost. Figure 5 [Figure 5: see original paper] illustrates dilated convolution expansion.

3.3.3 GLU Function GLU (Gated Linear Unit) and GTU units [24] are activation functions based on gate mechanisms. GLU' s linear channel prevents gradient vanishing during backpropagation, accelerating convergence and preventing gradient dispersion. The GLU expression is:

$$h_l = (x * W + b) \otimes \sigma(x * V + c)$$

where x represents layer l input, h_l is output information, W and V are convolutional kernel parameters, and c is bias. At layer l, input x processed by kernel W and bias b is controlled by an output gate processed by kernel V and bias c through sigmoid activation, preserving more important information for the next layer and enhancing disease feature learning. Figure 6 [Figure 6: see original paper] shows the improved MobileNetV3 structure.

4 Experiments and Results

4.1 Experimental Environment

Model training and testing were conducted using the PyTorch deep learning framework. Hardware included an Intel(R) Celeron(R) CPU N3150 @ 1.60GHz, 4GB RAM, and NVIDIA GeForce GPU with 4GB VRAM. The system ran Windows 7 with PyCharm and Python 3.7. Each experiment ran for 60 epochs with batch size 64, considering memory constraints and generalization. Learning rates of 1$×10^{\{-5\}, 1\times 10}${-4}, $1\times 10^{\{-3\}, and 1\times 10}${-2}$ were tested, with 1$×10^{-3}$ achieving optimal results. The Adam optimizer and ReLU activation were used.

4.2 Transfer Learning Validation

To validate transfer learning effectiveness, recognition performance and experimental results of VGG16, Inception-V3, ResNet50, and MobileNetV3 were compared. As shown in Figure 7 [Figure 7: see original paper], initial performance improved from 19%-35% to 55%-74%. MobileNetV3 converged within 10 epochs compared to 20 epochs originally, while VGG16, Inception-V3, and ResNet50 converged after 25 epochs with significantly improved accuracy. Transfer learning curves demonstrated greater stability during training.

Figure 8 [Figure 8: see original paper] compares the four algorithms' performance on tomato leaf disease recognition. MobileNetV3 transfer learning achieved higher accuracy, converging at 10 epochs—15 epochs earlier than other methods—with smaller oscillation amplitude and more stable training. These results indicate MobileNetV3 transfer learning is most suitable for tomato leaf disease recognition.

4.3 MobileNetV3 Transfer Model Performance Under Different Experimental Schemes

After transfer learning validation showing MobileNetV3' s optimal performance, the impact of loss functions and data augmentation was evaluated. Four combination schemes were tested using conventional augmentation versus Mixup hybrid augmentation with focal loss versus cross-entropy loss functions. Table 1 shows accuracy and loss variations.

Table 1 Identification accuracy of four comparison schemes on tomato disease identification

Scheme	Loss Function	Augmentation	Accuracy (%)	Loss (%)
F	Focal loss	Mixup	94.68	8.47

Scheme	Loss Function	Augmentation	Accuracy (%)	Loss (%)
FX	Focal loss	Conventional	94.31	9.94
C	Cross-entropy	Mixup	94.57	13.98
CX	Cross-entropy	Conventional	94.26	14.59

Results show similar recognition accuracy across schemes but more significant loss variation. Schemes F and FX achieved 0.05%-0.11% higher accuracy and approximately 5% lower loss than C and CX, indicating focal loss outperforms cross-entropy. Comparing FX vs F and CX vs C, Mixup hybrid augmentation improved accuracy by 0.37% and 0.31% respectively while reducing loss by 1.47% and 0.61%, demonstrating Mixup' s effectiveness for deep network performance.

4.4 Comparison Before and After MobileNetV3 Improvement

4.4.1 Model Recognition Performance Enhancement Table 2 compares recognition results of four transfer learning algorithms. VGG16 achieved only 86.62% accuracy, while ResNet50 and Inception-V3 reached 89.95% and 90.26% respectively. MobileNetV3 attained 94.68%, which improved to 98.25% after structural modifications—a 3.57% enhancement.

Table 2 Recognition results of four transfer learning algorithms for tomato leaves

Model	Accuracy (%)	Detection Time (s)
VGG16 Transfer	86.62	0.57
ResNet50 Transfer	89.95	1.54
Inception-V3 Transfer	90.26	0.55
MobileNetV3 Transfer	94.68	0.39
Improved MobileNetV3	98.25	0.27

Figure 9 [Figure 9: see original paper] shows test curves: MobileNetV3 (transfer learning), dilated (with dilated convolution and perceptron), and dilated+GLU (adding GLU function). The improved model achieved higher accuracy, with dilated convolution and perceptron improving recognition by 2.62% over transfer learning alone, and GLU adding another 0.85% improvement. ResNet50 required 1.54s per image, while VGG16 and Inception-V3 needed 0.57s and 0.55s respectively. MobileNetV3 averaged 0.39s, which reduced to 0.27s with GLU—a 0.12s improvement—making it most suitable for tomato disease detection.

4.4.2 Per-Class Disease Recognition Accuracy Improvement Figure 10 [Figure 10: see original paper] shows the confusion matrix of the improved model, where 0-9 represent bacterial spot, early blight, late blight, leaf mold, septoria leaf spot, target spot, two-spotted spider mite, mosaic virus, yellow leaf curl disease, and healthy leaves. The improved model significantly reduced misclassifications.

Table 3 details per-class accuracy before and after improvement. Recognition rates for healthy leaves, two-spotted spider mite, and leaf mold improved by 6%-7%, while other categories improved by approximately 3%. The average accuracy across ten classes reached 98.25%, up from 94.68%.

Table 3 Test accuracy of improved MobileNetV3 for tomato leaf diseases

Disease Class	Before (%)	After (%)
Bacterial Spot	92.31	95.87
Early Blight	93.45	96.12
Late Blight	94.22	97.85
Leaf Mold	89.76	96.34
Septoria Leaf Spot	95.67	98.43
Target Spot	93.89	97.21
Two-spotted Spider Mite	88.54	95.12
Mosaic Virus	96.78	99.05
Yellow Leaf Curl	94.56	97.88
Healthy Leaves	87.92	94.67

4.5 Model Recognition Performance Evaluation

Cross-validation [25] was used to evaluate model performance. Ten-fold cross-validation divided tomato leaf images into ten parts, using nine parts for training and one for testing in each iteration. The average of ten test results served as the robustness metric. Accuracies for folds 1-10 were 96.78%, 98.51%, 97.31%, 98.22%, 98.17%, 99.05%, 98.66%, 98.93%, 97.72%, and 99.17%, averaging 98.25%. The consistent results demonstrate the improved MobileNetV3 model' s stability and reliability.

5 Conclusion

This study improved the activation function and bottleneck convolutional layers of MobileNetV3 based on transfer learning to construct a tomato disease recognition model classifying ten disease types. Experimental analysis yielded the following conclusions:

1. Lightweight networks offer advantages for deep learning recognition tasks. Compared with VGG16, ResNet50, and Inception-V3 deep transfer models, the lightweight MobileNetV3 achieved better tomato disease recognition with 98.25% average accuracy, 0.27s per-image recognition time, and 43.57 MB model size, enabling real-time mobile APP-based detection.

2. Transfer learning enhances network recognition performance but has limitations. Combining transfer learning with Mixup hybrid augmentation

and appropriate loss functions improved MobileNetV3 accuracy from 94.57% to 98.25% with more stable training.

3. Addressing challenges of similar leaf colors and lesion shapes among different tomato diseases, model convolutional layer improvements with dilated convolution and perceptron structures enhanced feature extraction, improving accuracy by 2.62%. The GLU gating mechanism activation function accelerated convergence, reducing recognition time by 0.12s.
4. Ten-fold cross-validation demonstrated model stability and reliability with 98.25% average test accuracy.

Future work will collect natural environment images of multiple crop diseases for further model lightweight improvements toward developing a mobile-based crop disease recognition system.

References

[1] ZHANG C, YUAN S, HUO K. Factors affecting tomato production efficiency[J]. Hubei Agricultural Mechanization, 2020(12): 33-34.

[2] WANG J, NING F, LU S. Study on apple leaf disease identification method based on support vector machine[J]. Shandong Agricultural Science, 2015, 47(7): 122-125, 141.

[3] QIN F, LIU D, SUN B, et al. Recognition of four different alfalfa leaf diseases based on image processing technology[J]. Journal of China Agricultural University, 2016, 21(10): 65-75.

[4] XIA Y, LI Y, LI C. Recognition of wheat leaf disease based on image processing technology[J]. Science and Technology Bulletin, 2016, 32(4): 92-95.

[5] LIU T, WANG T, HU L. Rhizocotonia Solani recognition algorithm based on convolution neural network[J]. China Journal of Rice Science, 2019, 33(1): 90-94.

[6] WU Y. Identification of maize leaf diseases based on convolutional neural network[J]. Journal of Physics: Conference Series, 2021, 1748(3): ID 032004.

[7] DING R, ZHOU P. Identification of typical crop leaf diseases based on convolutional neural network[J]. Packaging Journal, 2018, 10(6): 74-80.

[8] CHEN G, ZHAO S, CAO L, et al. Corn plant disease recognition based on migration learning and convolutional neural network[J]. Smart Agriculture, 2019, 1(2): 34-44.

[9] WAHEED A, MUSKAN G, GUPTA D, et al. An optimized dense convolutional neural network model for disease recognition and classification in corn leaf[J]. Computers and Electronics in Agriculture, 2020, 175: ID 105456.

[10] CHANG J, GUAN S, SHI H, et al. Strip defect classification based on improved generative adversarial networks and MobileNetV3[J]. Laser & Optoelectronics Progress, 2021, 58(4): 221-226.

[11] FAYE M, CHEN B, KANE A S. Plant disease detection with deep learning and feature extraction using plant village[J]. Journal of Computer and Communications, 2020, 8(6): 10-22.

[12] WANG X, ZHANG C, ZHANG S. Multiscale convolutional neural networks with attention for plant species recognition[J]. Computational Intelligence and Neuroscience, 2021, 2021: ID 5529905.

[13] ZOHAIB M, SHUN F, QUOC V. Spectral images based environmental sound classification using CNN with meaningful data augmentation[J]. Applied Acoustics, 2021, 172: ID 107581.

[14] LU J, LIN J, HUANG Z, et al. Identification of citrus fruit infected with Huanglongbing based on Mixup algorithm and convolutional neural network[J]. Journal of South China Agricultural University, 2021, 42(3): 94-101.

[15] GE L. Research on modeling method of mixed beta distribution[J]. South China Review, 2019(4): 166.

[16] YI Q, LI C, LI B, et al. Field vehicle recognition based on improved MobileNetV1 network[J]. Industrial Control Computer, 2020, 33(7): 104-107.

[17] CHEN Z, JIAO H, YANG J, et al. Garbage image classification algorithm based on improved MobileNetV2[J]. Journal of Zhejiang University (Engineering Edition), 2021, 55(8): 1490-1499.

[18] LI X, SHEN X, ZHOU Y, et al. Classification of breast cancer histopathological images using interleaved DenseNet with SENet (IDSNet)[J]. PLoS One, 2020, 15(5): ID e0232127.

[19] LI M, WANG J, LI H, et al. Method for crop disease identification based on CNN and transfer learning[J]. Smart Agriculture, 2019, 1(3): 46-55.

[20] MOON J, HOSSAIN M B, CHON K H. AR and ARMA model order selection for time-series modeling with ImageNet classification[J]. Signal Processing, 2021, 183(10): ID 108026.

[21] YAO Y, HU L, GUO J. Terrain recognition based on deep migration network MobileNetV3[J]. Journal of Guangxi University (Natural Science Edition), 2021, 46(4): 996-1007.

[22] AMAHAN P A, VILLARICA M V, VINLUAN A A. Technical analysis of Twitter data in preparation of prediction using multilayer perceptron algorithm[C]// Proceedings of 2021 4th International Conference on Data Science and Information Technology (DSIT 2021). New York, USA: Association for Computing Machinery, 2021: 120-124.

[23] KUMAR V, SINGH R, DUA Y. Morphologically dilated convolutional neural network for hyperspectral image classification[J]. Signal Processing: Image Communication, 2022, 101: ID 116549.

[24] GENG H. Research on mono singing sound separation based on U-Net model[D]. Urumqi: Xinjiang University, 2020.

[25] MABUNI D, BABU S A. High accurate and a variant of k-fold cross validation technique for predicting the decision tree classifier accuracy[J]. International Journal of Innovative Technology and Exploring Engineering, 2021, 10(2): 105-110.

Note: Figure translations are in progress. See original paper for figures.