

Lightweight Ghost Dense Network for Tomato Leaf Disease Identification

Yinan Wang

*Faculty of Electrical Engineering and
Computer Science, Ningbo University*
Ningbo, China
2111082385@nbu.edu.cn

Zhijun Xie*

*Faculty of Electrical Engineering and
Computer Science, Ningbo University*
Ningbo, China
xiezhijun@nbu.edu.cn

Libo Zhuang

*Research and Development Center
Ningbo Shixin Agro-Ecology Company*
Ningbo, China
1366268766@qq.com

Kewei Chen

*Faculty of Mechanical Engineering and Mechanis
Ningbo University*
Ningbo, China
chenkewei@nbu.edu.cn

Yuntao Xie

*School of Computer Science and Engineer
The University of New South Wales*
Sydney, Australia
z5411128@ad.unsw.edu.au

Abstract—Timely and accurate identification of tomato leaf disease types can effectively improve the quality and yield of tomatoes, increase farmers' economic returns, and promote intelligent and modernized tomato production. To address the problems of intra- and inter-class multi-scale variation, complex background interference, and difficulty in mobile model deployment faced by tomato leaf disease identification, we propose a lightweight Ghost Dense network (LGDNet) to identify diseases of tomato leaves. First, we replace the standard convolution of the bottleneck layer in DenseNet with the Ghost module, which compresses the network size while maintaining the model's adaptability to the multi-scale variation of tomato leaf diseases. Then, we propose a lightweight and efficient coordinate multidimensional information fusion attention (CMIFA) module that enhances feature extraction for tomato leaves and enables the model to locate the diseased areas more accurately. The experimental results indicate that LGDNet reaches optimal recognition performance in both the PlantVillage dataset with a simple background and the Dataset of Tomato Leaves with natural scenes. Moreover, LGDNet achieves the minimum number of parameters among the compared models. In summary, LGDNet provides an excellent solution to the problem of accurately identifying tomato leaves in complex environments and provides a reference for deployment on mobile.

Index Terms—Tomato leaf disease identification, natural scenes, Lightweight model, Ghost module, CMIFA module

I. INTRODUCTION

Tomato is rich in nutrients and is widely grown around the world. According to recent statistics, the global production of tomatoes is about 18666 million tons, with an export

This research was supported by National Natural Science Foundation of China (Grant No.U20A20121); Ningbo public welfare project (Grant No.202002N3109, 2022S094); Natural Science Foundation of Zhejiang Province (Grant No.LY21F020006); Science and Technology Innovation 2025 Major Project of Ningbo (Grant No.2019B10125, 2019B10028, 2020Z016, 2021Z031, 2022Z074, 20223ZDYF020006); Ningbo Fenghua District industrial chain key core technology "unveiled the commander" project (Grant No.202106206).

*Corresponding author: Zhijun Xie, Associate professor, E-mail: xiezhijun@nbu.edu.cn

value of 1052 million. However, tomatoes are susceptible to various diseases during growth and development, which can reduce quality, affect the edible taste, lower yields, and even cause extinction, resulting in severe economic losses [1]. Therefore, rapid and accurate detection of tomato diseases contributes to the sustainable and appropriate development of agriculture and the prevention of unnecessary waste of financial and other resources [2]. Traditional methods of crop disease identification require manual inspection or chemical analysis of diseased areas, often requiring a trained expert to identify and diagnose them [3]. However, specialists are susceptible to personal factors such as fatigue and emotional changes during diagnosis, which can cause misjudgment of the disease or untimely diagnosis and bring substantial economic losses to farmers. Therefore, accurate, rapid, and automatic identification of crop disease types is essential for tomato disease control.

With machine learning methods continuing to be applied in agriculture, researchers have conducted in-depth studies on identifying tomato leaf diseases [4], [5]. However, Machine identification methods for tomato leaf diseases usually require manual labeling of tomato disease features, which vary from disease. In addition, even for the same disease, the symptoms exhibited at different stages of onset vary greatly and are susceptible to noise perturbations such as lighting and complex backgrounds, making the extraction of disease symptom features very difficult.

In recent years, deep learning techniques, especially convolutional neural networks (CNNs), have made good progress and wide application in the research of tomato disease identification due to their ability to automatically extract disease leaf features for identification and avoid manual extraction of features [6]. However, most of these solutions propose network model methods with more parameters and complex designs, which limit their capabilities on mobile devices.

With the wide application of Internet of things (IoT) tech-

nology in various industries and the increasing popularity of mobile devices, researchers are beginning to investigate tomato leaf disease identification algorithms that can use in the mobile terminal and tackle the problem of deep networks with more layers and complex structures. Although most of the current lightweight tomato leaf disease identification networks have achieved good results in identifying tomato leaf diseases in simple backgrounds, they cannot obtain rich features in complex backgrounds due to their limited parameters [7]. Therefore, solving the interference of complex backgrounds is a challenge in identifying tomato leaf diseases.

Attention mechanism is an excellent way to solve complex background interference. In image classification tasks, attention assigns different weights to different information to focus attention on important or interesting information, thus filtering out the unimportant information and improving the recognition ability of the network. There has been much research on attention mechanisms with promising results [8]–[10]. Recently, researchers have also applied attentional mechanisms to plant disease recognition. Tang *et al.* [11] added Channel-Wise attention mechanisms to ShuffleNet V1 and V2 models to improve the real-time accuracy of recognition. Sheng *et al.* [12] proposed a novel CNN that integrates Squeeze-and-Excitation attention mechanisms [13] and residual blocks, which can better adapt to diagnose many tomato leaf diseases.

Inter- and intra-class multi-scale variation is another difficulty with tomato leaf diseases. Different tomato leaf diseases and the same tomato leaf disease under different periods have different phenotypes with varying sizes and sparseness of spots. If only single-scale models use for identification, it will significantly reduce the identification performance of the model.

Considering the above discussion, we propose a lightweight Ghost Dense network (LGDNet) for identifying tomato leaf diseases in this study. First, we use lightweight Ghost Dense (LGD) blocks to extract tomato leaf disease features at different scales (e.g., Septoria Leaf Spot and Target Spot). Then, to reduce the interference of complex backgrounds on recognition, we use coordinate multidimensional information fusion attention (CMIFA) modules to improve the extraction of tomato leaf disease features and enable the model to locate the disease site more accurately. Comparative experiments demonstrate that our designed model achieves the highest recognition performance and the least number of parameters, which can provide a reference for deploying mobile tomato disease detection models in the natural environment. In summary, this paper has the following three contributions:

- We designed a novel lightweight CNN model (LGDNet) that significantly compresses the network size while guaranteeing the performance of identifying tomato leaf diseases. Our model is constructed from LGD blocks based on densely connected that enhance the identification of tomato leaf diseases at multiple scales. LGDNet has less number of parameters and Floating point Operations (FLOPs) and is well-suited for deployment on mobile or embedded devices.

- CMIFA module enables the model to fuse spatial information with channel features in a weighted manner to more accurately locate the disease area while adding only a small number of model parameters and FLOPs.
- We validate the effectiveness and robustness of the proposed architecture by comparing it with other advanced models on datasets with simple and complex backgrounds, respectively. The experimental results indicate that LGDNet performs better in tomato leaf disease identification.

II. RELATED WORK

Nowadays, intelligent methods for identifying tomato leaf diseases have been a great success. Researchers have extensively used machine learning or deep learning methods to identify diseases from tomato leaves.

A. Machine learning methods

Earlier approaches for tomato leaf disease identification first required manual extraction of features such as texture and shape of the disease image and then input into a machine learning-based classifier for identification. Xie *et al.* [4] combined the wavelength and the texture features of hyperspectral images to identify tomato diseases and achieved 97.1% accuracy. Xie *et al.* [5] used tomato image spectrum and texture features in a study of early blight of tomatoes with an accuracy of 88.46%. Basavaiah *et al.* [14] first fused the Local Binary Feature, Haralick, Hu Moments, and Color Histogram pattern features of disease images. Then the fused features were used for classification using Decision Tree and Random Forest classification algorithms and achieved 94% and 90% accuracies, respectively. Machine learning methods rely on individually extracted features and are susceptible to noise, which makes them time-consuming and labor-intensive.

B. Deep learning methods

In recent years, researchers have proposed various image recognition models to identify tomato leaf disease. Tm *et al.* [6] used LeNet, a simple CNN, to identify ten tomato diseases, and the average recognition accuracy reached 94.85%. Kaur *et al.* [15] proposed a DAG-ResNet-based model and used Error Correction Output Code (ECOC) to classify seven tomato diseases in the PlantVillage dataset, and the accuracy achieved 98.8%. Mkonyi *et al.* [16] applied three transfer learning models to identify tomato diseases. The experimental results indicated that the VGG 16 model based on transfer learning had the best recognition accuracy of 91.9%. These models are pre-trained on massive datasets and automatically extract relevant features without manual labeling, which is superior to machine learning-based models. However, these models are enormous and computationally expensive for mobile devices.

To reduce the model size and computational costs, researchers designed lightweight and faster network models. Durmuş *et al.* [17] detected ten types of tomato leaf diseases using AlexNet and SqueezeNet and conducted the experiments on an embedded device. Elhassouny *et al.* [18] developed

an innovative mobile application that used MobileNetV1 to detect nine tomato plant diseases on the PlantVillage dataset achieving 90.3% accuracy. However, the recognition accuracy achieved by these two methods seems insufficient compared to the deeper models. SABBIR AHMED *et al.* [3] used a pre-trained MobileNetV2 network and classifier to extract features from tomato leaf images in the PlantVillage dataset, ultimately achieving an accuracy of 99.30%. However, this author only identified tomato leaves under a simple background and did not consider the effect of complex background interference. All of the above issues place a higher demand on lightweight models that can fully consider the disease features of tomato leaf diseases and achieve state-of-the-art performance.

III. METHOD

A. LGDNet model structure design

Fig. 1 shows the overall structure of LGDNet. LGDNet first uses a Feature Block to downsample the input image, reduce the computational parameters of the model and obtain rich image features. Then, four LGD blocks are connected for feature extraction to obtain multi-scale tomato leaf disease features. Each block consists of three lightweight Ghost bottleneck layers and three CMIFA modules. The CMIFA module enables the model to locate the disease area of tomato leaf diseases more accurately by fusing spatial and channel information, reducing the interference of complex backgrounds, and improving disease identification. Meanwhile, a transition layer is added between each two LGD blocks to collect all scale features and reduce the number of channels and the size of feature maps to improve computational efficiency. Finally, LGDNet connects a Global Average Pooling (GAP) layer and an n -class classifier (n depends on the number of classifications) to obtain the category output.

B. Structure of LGD block

To address the problem of multi-scale variation in tomato leaf disease, a network architecture that captures as much multi-scale variation as possible is needed. As shown in Fig. 2, Dense blocks connect all layers directly, with each layer taking input from the output of all previous layers before it. It ensures maximum information transfer between the layers of the network so that the input of each layer is the sum of the outputs of all previous layers, enhancing the transfer of features and making more efficient use of them. The increased diversity of scales obtained with dense connectivity facilitates the extraction of different scales of tomato leaf disease features.

Each bottleneck layer in the Dense block consists of a 1×1 convolution and a 3×3 convolution, where the 1×1 convolution is used for dimensionality reduction. As shown in Fig. 3, to further reduce the model's number of parameters and FLOPs, we replace the standard convolution in the bottleneck layers with the Ghost module. Unlike traditional CNN architectures, a study in GhostNet pointed out that generating many "ghost" feature maps using a series of cheap cost linear transformations on top of a set of eigenfeature maps to

reveal information about the intrinsic features can also achieve competitive performance [19]. As shown in Fig. 4, the Ghost module first uses a 1×1 pointwise convolution to obtain a new feature map $F1$. Then it uses linear operation with a 3×3 depthwise convolution to produce a feature map $F2$ cascade with feature map $F1$ to produce the final output. This way, the Ghost module can maintain fewer parameters and FLOPs when generating the same number of feature maps compared to standard convolution. Meanwhile, we use the traditional post-activation (Conv-Bn-Relu) as a composite function instead of the pre-activation (Bn-Relu-Conv) in DenseNet to improve the actual speed [20].

In addition, to improve feature extraction of tomato leaf disease while maintaining lightweight advantages, the CMIFA module is introduced into the LGD block for effective feature learning.

C. CMIFA module

CMIFA is constructed based on coordinate attention (CA) [21]. The coordinate attention mechanism embeds location information into the channel attention mechanism to enable the module to focus on the region of interest and obtain information in a larger region to distinguish background from foreground more effectively and achieve better target classification results. The mechanism first performs average pooling of the input feature maps in horizontal and vertical directions. Then it implements spatial information encoding through dimensional shift, dimensional splicing, and dimensionality reduction operations to reveal whether Region of Interest (ROI) targets exist in these two directions. Finally, it fuses spatial information with channel features in a weighted manner, considering both channel and location information, helping the model better localize to the object of interest. Fig. 5 illustrates the coordinate attention.

As shown in Fig. 6, the CMIFA module is an improvement of the CA module. Each channel of the input feature map X in the CA module is encoded along the width and height directions using an average pooling of kernel size $(1, W)$ and $(H, 1)$, respectively, to obtain the feature maps in the width and height directions. Thus, the output $F_A(z)$ of the t -th channel with width W and height H formulate as follows:

$$\begin{cases} F_A(z_t^w(w)) = \frac{1}{H} \sum_{0 \leq j < H} X_t(j, w) \\ F_A(z_t^h(h)) = \frac{1}{W} \sum_{0 \leq i < W} X_t(h, i) \end{cases} \quad (1)$$

Average Pooling (AP) reduces the estimates' variance caused by the restricted neighborhood size when extracting features with convolution and preserves more image background information. Maximum pooling (MP) reduces the estimation's mean value offset caused by the convolution layer's parameter error and retains more texture information. Therefore, adding MP-based aggregation features in CA can better preserve the texture information of images and obtain the aggregation feature $F_M(z)$ as follows:

$$\begin{cases} F_M(z_t^w(w)) = \max_{0 \leq j < H} X_t(j, w) \\ F_M(z_t^h(h)) = \max_{0 \leq i < W} X_t(h, i) \end{cases} \quad (2)$$

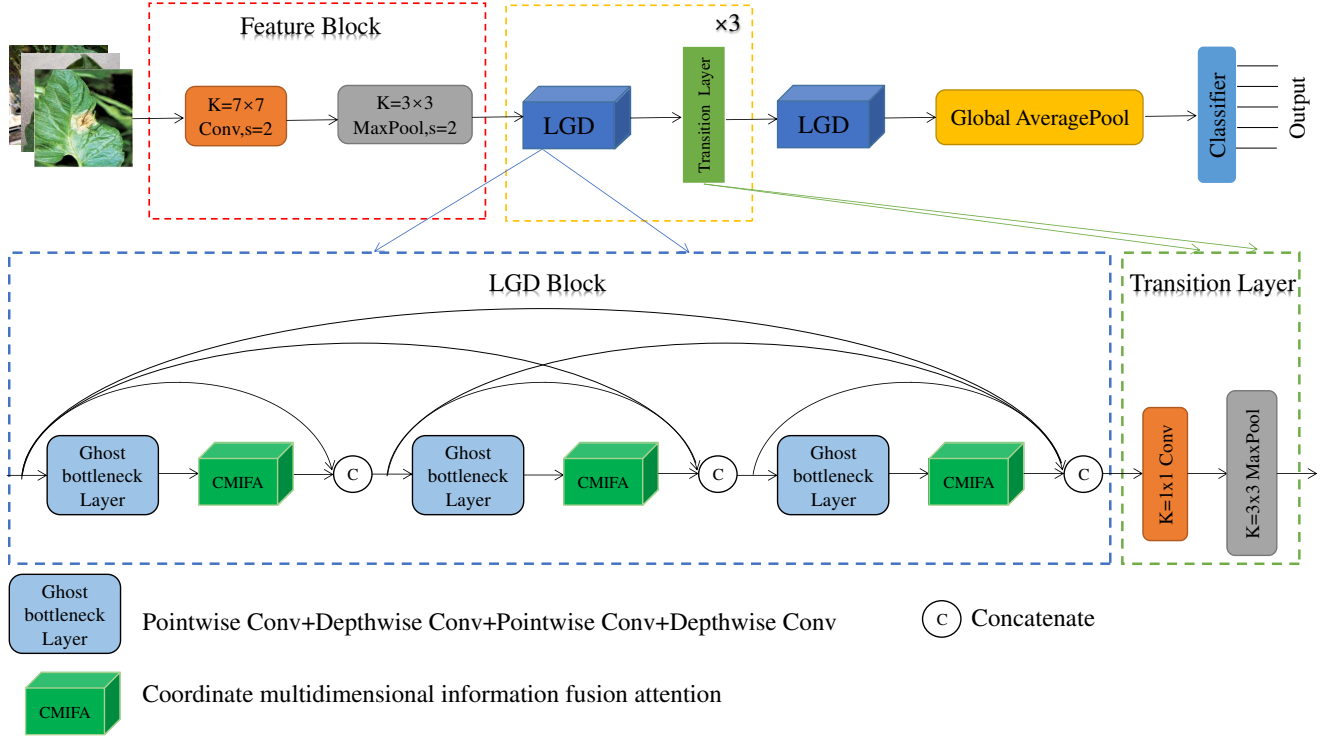


Fig. 1. The overall structure of LGDNet model.

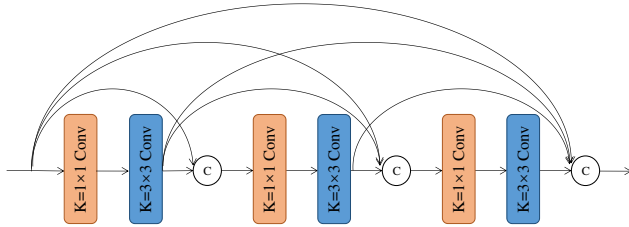


Fig. 2. Structure of Dense block.

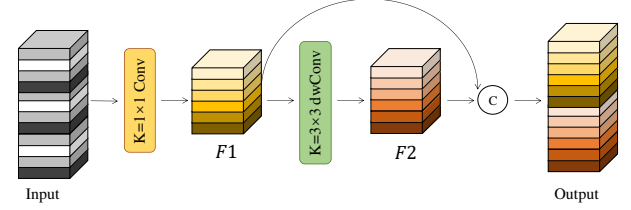


Fig. 4. The Ghost module.

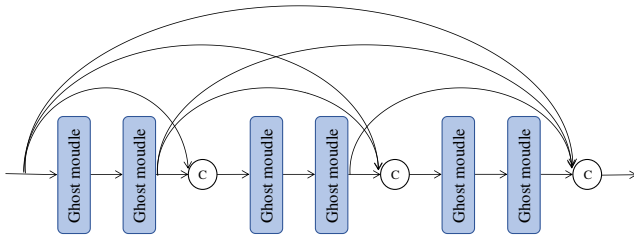


Fig. 3. Structure of LGD block without CMIFA module.

To fully use the global perceptual field obtained by the above transformations to encode precise location information and accurately localize the region of interest, we first reshape

and concatenate the feature maps of the obtained global perceptual field in the width and height directions along the spatial dimension. Consider that CA uses a fully connected layer to map features when computing channel attention, and the fully connected layer is computationally intensive. Even if the channel features are compressed using a factor of r , the number of parameters is still proportional to the square of the number of channels for the input feature map. In addition, in intermediate feature maps, where there is a more significant correlation between adjacent channels, mapping channel features using a fully connected layer generates many redundant computations [22]. Therefore, we choose to use the $k_1 \times 1$ convolution $F1$ to fuse the features of k_1 channels in the vicinity of the channels. Also, we choose to use the $1 \times k_2$ convolution $F2$ to fuse the contextual spatial information of the image so that the module can locate the lesion region more

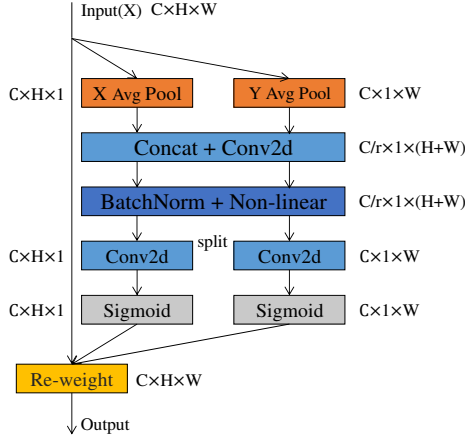


Fig. 5. CA module.

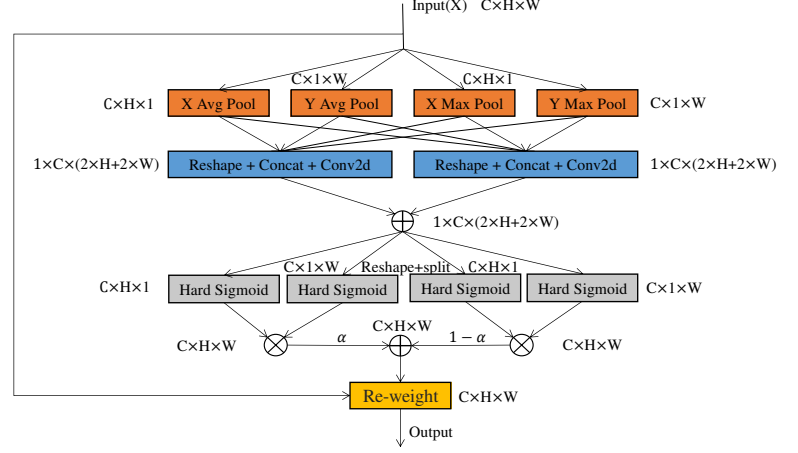


Fig. 6. CMIFA module.

accurately:

$$Q = F_1([F_A(z^w), F_A(z^h), F_M(z^w), F_M(z^h)]) + F_2([F_A(z^w), F_A(z^h), F_M(z^w), F_M(z^h)]) \quad (3)$$

For varying numbers of channels C and feature map height H within the model, the convolution kernel sizes k_1 and k_2 are calculated as follows:

$$\begin{cases} k_1 = \rho(C) = \text{odd}(|\log_2(C)|) \\ k_2 = \rho(H) = \text{odd}(|\log_2(H)|) \end{cases} \quad (4)$$

where $\rho(\alpha)$ denotes the nonlinear mapping and $\text{odd}(|\beta|)$ denotes the odd number closest to β . Thus, high-dimensional channels and low-dimensional space have a longer range of interactions. In contrast, low-dimensional channels and high-dimensional space undergo a shorter range of interactions.

Then, Q is reshaped and divided into four separate tensors $Q_A^w \in R^{C \times W}$, $Q_A^h \in R^{C \times H}$, $Q_M^w \in R^{C \times W}$, and $Q_M^h \in R^{C \times H}$ along the spatial dimension, and use σ (Hard Sigmoid activation function) to obtain the attention weights of g_w^A and g_w^M in the width direction and g_h^A and g_h^M in the height direction of the feature map, respectively. Then, fuse them using multiplicative and additive operations. Moreover, the adaptive parameter α is added to help the model obtain critical features, and ω_t is calculated as follows:

$$\begin{cases} g_A^w = \sigma(Q_A^w) & g_A^h = \sigma(Q_A^h) \\ g_M^w = \sigma(Q_M^w) & g_M^h = \sigma(Q_M^h) \\ \omega_t = \alpha \times g_A^w \times g_A^h + (1 - \alpha) \times g_M^w \times g_M^h \end{cases} \quad (5)$$

Finally, the output feature weights ω_t are multiplied by the input features X_t to form the final output features U_t .

$$U_t = X_t \cdot \omega_t \quad (6)$$

CMIFA overcomes the disadvantages of redundant computation and many parameters brought by the fully connected layer in CA. Moreover, contextual spatial information fusion and adaptive aggregation of dual pooling feature information are added to more accurately locate the disease areas on tomato leaves.

IV. EXPERIMENTS

A. Experimental datasets

We use the PlantVillage dataset and the Dataset of Tomato Leaves dataset as the datasets for this paper. The PlantVillage dataset contains 18,160 images of tomato leaves, many taken under laboratory conditions with simple backgrounds and only one leaf in each image, divided into ten categories with nine diseased and healthy leaves. The Dataset of Tomato Leaves dataset is a collection of images of tomato leaves in natural scenes from the public website (<https://data.mendeley.com/>), consisting of single leaves, multiple leaves, simple backgrounds, and complex backgrounds. It has 622 images in six categories, containing one type of healthy leaves and five types of diseased leaves.

Considering that fewer samples of natural scenes will produce data overfitting, thus producing the issue of higher recognition accuracy of the model in the training set and lower recognition accuracy in the test set, which is unfavorable to the recognition of tomato leaf disease. Therefore, we use an offline data enhancement method (e.g., flip, random rotation, luminance transformation, Gaussian blur, add agitation, random cropping, random panning) to expand the tomato leaf dataset to 7464 images to avoid model training overfitting, enhance the model's generalisation ability, and make the final tomato leaf disease dataset in natural scenes.

Before the experiments, we used images from each category of the above two datasets in the ratio of 6:2:2 for training, validation, and testing, respectively, and then resized all images to 224×224 pixels and did normalization.

B. Experimental environment

All models are trained, validated, and tested on NVIDIA RTX 3080Ti GPUs, Ubuntu 20.04 systems, and PyTorch version 1.8.0 framework with the same hyperparameters (not necessarily the best ones). In the training phase of the network model, we optimize the network weights using the SGDM

optimizer. The initial learning rate is 0.1, the optimizer momentum is 0.9, the weight decayed is $1e^{-4}$, the batch size is 16, and the learning rate is dynamically adjusted to $1e^{-5}$ using the cosine annealing method. The number of epoch elements for all network training is set to 60. Accuracy (Acc), Recall, Precision, F1 Score, Parameters, and FLOPs are used in this paper to measure the performance of the network model. Flops and Parameters are used to measure the computational complexity and the consumption of computational memory resources during model training.

To reduce the impact of the overfitting problem on recognition accuracy due to the imbalance between easy and hard samples and positive and negative samples in model training, we replace the traditional Cross-Entropy loss with Reduced Focal loss [23]. For abnormal sample labels in the dataset, label smoothing is used to mitigate the impact of mislabeling on model training [24].

C. Comparative experiment on different network models

We compare LGDNet with various other advanced recognition models in the simple background of the PlantVillage dataset to verify the performance advantages of LGDNet. As shown in Table I, our proposed model reaches optimal Acc, Recall, Precision, and F1 scores of 99.72%, 99.63%, 99.68%, and 99.65%, respectively. Also, the number of parameters of our model reaches a minimum of 1.03M. These experimental results show that LGDNet improves tomato disease recognition rate and reduces model parameters' memory requirement. Table II compares the proposed model's recognition accuracy with other state-of-the-art models on the PlantVillage dataset. The model proposed in this paper achieves the minimum number of parameters and the highest recognition accuracy.

To further validate the recognition performance of the proposed model in natural scenes, we trained our model again on a tomato leaf dataset in a natural environment for experimental comparison. As shown in Table III, the F1 score of the proposed model is 99.00%, which is 3.65% better than the next best model, MobileNetV3 Large. The results indicate that our proposed model can overcome complex background interference on recognition well and achieve the best recognition results compared with other models.

In summary, the dense connectivity mechanism employed by LGDNet can effectively address the challenge of multi-scale variation of tomato leaf diseases. The introduction of the attention mechanism (CMIFA module) also makes it more focused on the diseased areas of the leaves, reducing the interference of complex backgrounds, increasing the extraction of tomato leaf disease features, and enhancing the robustness of the model. LGDNet can effectively enhance the recognition of tomato leaf disease with only a few parameters and can provide a reference for tomato leaf disease recognition on mobile.

D. Ablation experiments

Previous experiments have shown that LGDNet achieved the highest F1 scores in two tomato disease datasets with different

difficulty background interference. To demonstrate the impact of our proposed CMIFA module on the model performance, we conduct the corresponding ablation experiments on each of the above two datasets. Table IV represents the results of tomato leaf disease identification incorporating different pooling operations of the CMIFA module. From the results, we can obtain that the single-pool attention modules LGDNet (only AP) and LGDNet (only MP), including only AP or MP, have similar F1 scores and are lower than the dual-pool attention module LGDNet, which includes both AP and MP. This finding suggests that using only MP or AP to reduce the dimensionality of the feature mapping can lead to feature loss. As shown in Table V, adding an adaptive parameter α to LGDNet improves the F1 scores in the two datasets by 0.39% and 0.78%, respectively. The α helps the model select the weights of each pooling feature independently, avoiding the average direct addition of two pooling features and reducing the loss of valuable features due to dimensionality reduction, thus improving the model's performance. Table VI presents the recognition results of the spatial information fusion ablation experiment. The F1 scores of LGDNet improved by 0.34% and 0.75% in the two datasets, respectively. The fusion of spatial information allows the model to determine the class of a pixel by considering not only the grayscale value of the pixel at a specific location but also its close pixels, thus enhancing the recognition performance of the model.

Finally, we compare LGDNet with the addition of the CMIFA module, CA module, and no attention mechanism, and the results are presented in Table VII. Compared with the LGDNet model without the attention mechanism, the LGDNet model with the CMIFA module adds only a small number of parameters and FLOPs and obtains the highest experimental results among the three models. This result shows that the CMIFA module, which considers both spatial information and channel features, can better raise the recognition of tomato leaf diseases and balance the model's efficiency and complexity. As shown in Fig. 7, the identification results of tomato leaf diseases using Class Activation Mapping (CAM) visualization demonstrate that the CMIFA module effectively improves the model's ability to accurately identify disease areas while minimizing the impact of complex backgrounds on recognition.

V. CONCLUSION

In this paper, a lightweight CNN (LGDNet) is designed for identifying tomato leaf diseases. The network improves the recognition of tomato leaf diseases at different scales by linking LGD blocks together and can accurately locate tomato leaf disease features in complex backgrounds. The proposed CMIFA module helps the model to locate the diseased areas of tomato leaves more accurately and improve the performance of tomato leaf disease identification while only increasing the number of FLOPs and parameters by a small amount. The experimental results demonstrate that our proposed model achieves the best results in tomato leaf disease identification in simple and complex backgrounds and is suitable for

TABLE I
COMPARISON OF DIFFERENT MODELS ON THE PLANTVILLAGE DATASET

Method	Accuracy(%)	Recall(%)	Precision(%)	F1 Score(%)	Params(M)	FLOPs(M)
ResNet50	98.43	97.88	97.99	97.93	23.53	4142.83
DenseNet121	99.28	98.83	98.97	99.17	6.96	2911.54
MobileNetV2 1.0×	98.59	98.03	98.17	98.09	2.24	332.96
MobileNetV3_Large 1.0×	98.26	97.39	97.62	97.48	4.22	237.96
EfficientNet B0	98.90	98.41	98.60	98.5	4.02	420.61
MixNet_M 1.0×	99.09	98.73	98.95	98.83	5.01	360.12
Res2Net50	97.71	96.98	97.13	97.03	23.67	4316.27
GhostNet 1.0×	98.87	98.63	98.44	98.53	3.91	158.12
AsymmNent_Large 1.0×	97.35	96.47	96.25	96.33	4.81	237.19
Mobile VIT Small	99.17	98.94	98.79	98.86	4.94	1576.05
Ours	99.72	99.63	99.68	99.65	1.03	242.12

The result in bold indicates the best result for each row

TABLE II
PERFORMANCE COMPARISON WITH THE STATE-OF-THE-ART MODELS ON THE PLANTVILLAGE DATASET

Study	Method	Number of images	Accuracy(%)	Params(M)
Tm <i>et al.</i> [6]	LeNet	18160	94.85	52.40
Elhassouny <i>et al.</i> [18]	MobileNetv1	7176	90.03	3.22
Bir <i>et al.</i> [25]	EfficientNet-B0	15000	98.60	4.02
Zhao <i>et al.</i> [12]	SE-ResNet50	20925	96.81	33.59
Abbas <i>et al.</i> [26]	DenseNet121 + Synthetic images	16012	97.11	6.96
Kibriya <i>et al.</i> [27]	GooLeNet	10735	99.23	10.83
Moussafir <i>et al.</i> [28]	ResNet50+EfficientNet B0	14526	98.10	28.53
Ahmed <i>et al.</i> [3]	MobileNetv2+Classifier network	18160	99.30	2.40
Ours	LGNet	18160	99.72	1.03

TABLE III
COMPARISON OF DIFFERENT MODELS ON THE DATASET OF TOMATO LEAVES

Method	Accuracy(%)	Recall(%)	Precision(%)	F1 Score(%)	Params(M)	FLOPs(M)
ResNet50	83.40	83.02	82.56	82.78	23.53	4142.83
DenseNet121	92.30	91.50	92.83	92.07	6.96	2911.54
MobileNetV2 1.0×	86.61	85.70	86.11	85.82	2.24	332.96
MobileNetV3_Large 1.0×	95.72	95.31	95.41	95.35	4.22	237.96
EfficientNet B0	82.13	80.92	81.72	81.21	4.02	420.61
MixNet_M 1.0×	90.50	90.07	89.74	89.85	5.01	360.12
Res2Net50	91.43	90.73	91.31	90.91	23.67	4316.27
GhostNet 1.0×	81.59	80.35	80.78	80.49	3.91	158.12
AsymmNent_Large 1.0×	85.43	84.31	84.18	84.20	4.81	237.19
Mobile VIT Small	93.64	93.48	93.07	93.22	4.94	1576.05
Ours	99.00	99.02	98.99	99.00	1.03	242.12

TABLE IV
ABLATION EXPERIMENT FOR THE DIFFERENT POOLING OPERATIONS OF THE CMIFA MODULE

Method	PlantVillage	Dataset of Tomato Leaves
LGNet(Only AP)	99.37%	98.31%
LGNet(Only MP)	99.45%	98.41%
LGNet	99.65%	99.00%

TABLE V
ABLATION EXPERIMENT FOR THE SPACE INFORMATION FUSION

Method	PlantVillage	Dataset of Tomato Leaves
LGNet	99.26%	98.22%
LGNet+SIF	99.65%	99.00%

The "SIF" denotes Space Information Fusion.

tomato leaf disease prediction in real scenarios. In addition, our model achieves a minimal number of parameters, which makes the proposed model suitable for building mobile device applications. Moreover, in our experiments, only one disease was found per sample. However, having multiple diseases in a single leaf is often possible, so this would be another challenging task. Finally, our model can be applied to classify more crop diseases and become a promising application for farmers.

TABLE VI
ABLATION EXPERIMENT FOR ADAPTIVE PARAMETER α

Method	PlantVillage	Dataset of Tomato Leaves
LGNet	99.31%	98.25%
LGNet+α	99.65%	99.00%

TABLE VII
COMPARATIVE EXPERIMENT FOR THE DIFFERENT ATTENTION OF LGDNet

Method	PlantVillage	Dataset of Tomato Leaves	Params(M)	FLOPs(M)
LGDNet(WA)	98.93%	93.78%	1.025086	240.64
LGDNet(CA)	99.33%	95.54%	1.138667	243.02
LGDNet(CMIFA)	99.65%	99.00%	1.025254	242.12

The "WA" denotes without attention mechanism.

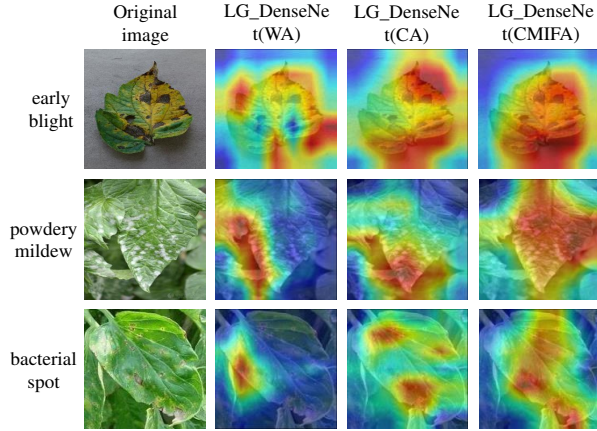


Fig. 7. Visualized results for the different attention of LGDNet in recognizing tomato leaf disease.

REFERENCES

- [1] S. Panno, S. Davino, A. G. Caruso, S. Bertacca, A. Crnogorac, A. Mandić, E. Noris, and S. Matić, "A review of the most common and economically important diseases that undermine the cultivation of tomato crop in the mediterranean basin," *Agronomy*, vol. 11, no. 11, p. 2188, 2021.
- [2] Ü. Atila, M. Uçar, K. Akyol, and E. Uçar, "Plant leaf disease classification using efficientnet deep learning model," *Ecological Informatics*, vol. 61, p. 101182, 2021.
- [3] S. Ahmed, M. B. Hasan, T. Ahmed, M. R. K. Sony, and M. H. Kabir, "Less is more: Lighter and faster deep neural architecture for tomato leaf disease classification," *IEEE Access*, vol. 10, pp. 68868–68884, 2022.
- [4] C. Xie, Y. Shao, X. Li, and Y. He, "Detection of early blight and late blight diseases on tomato leaves using hyperspectral imaging," *Scientific reports*, vol. 5, no. 1, pp. 1–11, 2015.
- [5] C. Xie and Y. He, "Spectrum and image texture features analysis for early blight disease detection on eggplant leaves," *Sensors*, vol. 16, no. 5, p. 676, 2016.
- [6] P. Tm, A. Pranathi, K. SaiAshritha, N. B. Chittaragi, and S. G. Koolagudi, "Tomato leaf disease detection using convolutional neural networks," in *2018 eleventh international conference on contemporary computing (IC3)*, pp. 1–5, IEEE, 2018.
- [7] L. Abrahamyan, V. Ziatichin, Y. Chen, and N. Deligiannis, "Bias loss for mobile neural networks," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 6556–6566, 2021.
- [8] P. Anderson, X. He, C. Buehler, D. Teney, M. Johnson, S. Gould, and L. Zhang, "Bottom-up and top-down attention for image captioning and visual question answering," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 6077–6086, 2018.
- [9] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in *Proceedings of the European conference on computer vision (ECCV)*, pp. 3–19, 2018.
- [10] H. Zhao, J. Jia, and V. Koltun, "Exploring self-attention for image recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10076–10085, 2020.
- [11] Z. Tang, J. Yang, Z. Li, and F. Qi, "Grape disease image classification based on lightweight convolution neural networks and channelwise attention," *Computers and Electronics in Agriculture*, vol. 178, p. 105735, 2020.
- [12] S. Zhao, Y. Peng, J. Liu, and S. Wu, "Tomato leaf disease diagnosis based on improved convolution neural network by attention module," *Agriculture*, vol. 11, no. 7, p. 651, 2021.
- [13] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7132–7141, 2018.
- [14] J. Basavaiah and A. Arlene Anthony, "Tomato leaf disease classification using multiple feature extraction techniques," *Wireless Personal Communications*, vol. 115, no. 1, pp. 633–651, 2020.
- [15] M. Kaur and R. Bhatia, "Development of an improved tomato leaf disease detection and classification method," in *2019 IEEE Conference on Information and Communication Technology*, pp. 1–5, IEEE, 2019.
- [16] L. Mkonyi, D. Rubanga, M. Richard, N. Zekeya, S. Sawahiko, B. Maiseli, and D. Machuve, "Early identification of tuta absoluta in tomato plants using deep learning," *Scientific African*, vol. 10, p. e00590, 2020.
- [17] H. Durmuş, E. O. Güneş, and M. Kırıcı, "Disease detection on the leaves of the tomato plants by using deep learning," in *2017 6th International conference on agro-geoinformatics*, pp. 1–5, IEEE, 2017.
- [18] A. Elhassouny and F. Smarandache, "Smart mobile application to recognize tomato leaf diseases using convolutional neural networks," in *2019 International Conference of Computer Science and Renewable Energies (ICCSRE)*, pp. 1–4, IEEE, 2019.
- [19] K. Han, Y. Wang, Q. Tian, J. Guo, C. Xu, and C. Xu, "Ghostnet: More features from cheap operations," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 1580–1589, 2020.
- [20] R. J. Wang, X. Li, and C. X. Ling, "Pelee: A real-time object detection system on mobile devices," *Advances in neural information processing systems*, vol. 31, 2018.
- [21] Q. Hou, D. Zhou, and J. Feng, "Coordinate attention for efficient mobile network design," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 13713–13722, 2021.
- [22] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "Eca-net: Efficient channel attention for deep convolutional neural networks," in *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, IEEE, Seattle, WA, USA*, pp. 13–19, 2020.
- [23] T. He, Z. Zhang, H. Zhang, Z. Zhang, J. Xie, and M. Li, "Bag of tricks for image classification with convolutional neural networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 558–567, 2019.
- [24] N. Sergievskiy and A. Ponamarev, "Reduced focal loss: 1st place solution to xview object detection in satellite imagery," *arXiv preprint arXiv:1903.01347*, 2019.
- [25] P. Bir, R. Kumar, and G. Singh, "Transfer learning based tomato leaf disease detection for mobile applications," in *2020 IEEE International Conference on Computing, Power and Communication Technologies (GUCON)*, pp. 34–39, IEEE, 2020.
- [26] A. Abbas, S. Jain, M. Gour, and S. Vankudothu, "Tomato plant disease detection using transfer learning with c-gan synthetic images," *Computers and Electronics in Agriculture*, vol. 187, p. 106279, 2021.
- [27] H. Kibriya, R. Rafique, W. Ahmad, and S. Adnan, "Tomato leaf disease detection using convolution neural network," in *2021 International Bhurban Conference on Applied Sciences and Technologies (IBCAST)*, pp. 346–351, IEEE, 2021.
- [28] M. Moussafir, H. Chaibi, R. Saadane, A. Chehri, A. E. Rharras, and G. Jeon, "Design of efficient techniques for tomato leaf disease detection using genetic algorithm-based and deep neural networks," *Plant and Soil*, vol. 479, no. 1, pp. 251–266, 2022.