LAD-Net: A Novel Light Weight Model for Early Apple Leaf Pests and Diseases Classification

Xianyu Zhu, Jinjiang Li, Runchang Jia, Bin Liu, Zhuohan Yao, Aihong Yuan, Yingqiu Huo, and Haixi Zhang

Abstract—Aphids, brown spots, mosaics, rusts, powdery mildew and Alternaria blotches are common types of early apple leaf pests and diseases that severely affect the yield and quality of apples. Recently, deep learning has been regarded as the best classification model for apple leaf pests and diseases. However, these models with large parameters have difficulty providing an accurate and fast diagnosis of apple leaf pests and diseases on mobile terminals. This paper proposes a novel and real-time early apple leaf disease recognition model. AD Convolution is firstly utilized to replace standard convolution to make smaller number of parameters and calculations. Meanwhile, a LAD-Inception is built to enhance the ability of extracting multiscale features of different sizes of disease spots. Finally, the LAD-Net model is built by the LR-CBAM and the LAD-Inception modules, replacing a full connection with global average pooling to further reduce parameters. The results show that the LAD-Net, with a size of only 1.25MB, can achieve a recognition performance of 98.58%. Additionally, it is only delayed by 15.2ms on HUAWEI P40 and by 100.1ms on Jetson Nano, illustrating that the LAD-Net can accurately recognize early apple leaf pests and diseases on mobile devices in real-time, providing portable technical support.

Index Terms—Apple leaf pests and diseases, asymmetric convolution, dilated convolution, real-time inference, convolutional neural networks, deep learning

1 Introduction

China has become the world's largest apple producer, with apple planting areas and exportation that accounts for more than 50% of the world. In 2020, China's apple planting area exceeded 4.94 million acres and exports reached 41 million tons. During the apple growth process, due to the influence of the environment and climate, various pests and diseases often harm the apples, which seriously affects the yield and quality of the apples. Preventing pests and diseases is an urgent problem for reducing economic losses, and this problem is a concern of an increasing number of scholars.

Traditionally, apple leaf pests and disease types are visually determined by experts, which is time-consuming and relies on their level of expertise. In addition, once inappropriate pesticides are used, the growing environment of the apple will be destroyed [1]. Along with the remarkable advancement in smart agriculture, innovative machine learning algorithms have been applied to identify crop leaf pests and diseases, such as a BPNN (back propagation neural network), an SVM (support vector machine) and K-means clustering [2], [3], [4], [5], [6], [7], but conventional machine learning methods rely heavily on researchers' experience to select and extract features, and image preprocessing is complex. In the last few years, convolutional neural networks (CNNs) have shown outstanding performance as feature extractors and classifiers in crop leaf pest and disease recognition tasks [8], [9], [10], [11], [12], [13]. Because the learning process of CNNs does not require manually selecting features, this method has been widely applied to the smart agricultural field. However, most CNN-based models have deep layers, large parameters and calculations, which rely on the support of high-performance servers, and they are also difficult to deploy to resource-constrained devices for real-time recognition. Recently, an increasing number of lightweight models have been used to crop leaf disease recognition tasks. However, for the crop leaf disease recognition task, which involves identifying disease spots that vary in size and color, the existing lightweight models cannot adapt it well for accuracy, speed or size. Therefore, there is a lack of light weight models for recognizing early apple leaf pests and diseases accurately and quickly based on resource-constrained terminal devices, which have limited memory and computing resources.

- *Xianyu Zhu, Jinjiang Li, Runchang Jia, Zhuohan Yao, Aihong Yuan, Yingqiu Huo, and Haixi Zhang are with the College of Information Engineering, Northwest A&F University, Yangling, Shaanxi 712100, China. E-mail: {zhuxy, 2019_ljj, jiarunchang, yaozhuohan, ahyuan}@nwafu.edu.cn, , fallying@nwsuaf.edu.cn, zh.haixi@gmail.com.*
- *Bin Liu is with the College of Information Engineering, Key Laboratory of Agricultural Internet of Things, Ministry of Agriculture and Rural Affairs, Shaanxi Key Laboratory of Agricultural Information Perception and Intelligent Service, Northwest A&F University, Yangling, Shaanxi 712100, China. E-mail: liubin0929@nwsuaf.edu.cn.*

Manuscript received 9 November 2021; revised 19 May 2022; accepted 8 July 2022. Date of publication 18 July 2022; date of current version 3 April 2023.
This work was supported in part by Shaanxi Key Research and Development Program under Grant 2021NY-138, in part by CCF-Baidu Open Fund under Grant 2021PP15002000, in part by the Shaanxi Key Research and Development Program under Grant 2019ZDLNY07-06-01, in part by the Key Research and Development Project of Shaanxi province under Grant 2020NY098, and in part by China's National College Innovation and Entrepreneurship Training Program of Northwest A&F University under Grant S202110712607.
(Corresponding author: Bin Liu.)
Digital Object Identifier no. 10.1109/TCBB.2022.3191854

Focusing on the problems mentioned above, this paper proposes a novel real-time model, the LAD-Net (Light weight model using Asymmetric and Dilated convolution). The main contributions of the research are listed as follows:

- A novel light weight model, the LAD-Net, is proposed to achieve real-time diagnose on early apple leaf pets and diseases. Most of existing studies are performed on simple diseased images or taken by professionals for the academic purpose, and the existing classic CNN models with large parameters have difficulty providing a real-time diagnosis on mobile terminals, while the existing light weight models with low accuracy still needs further improvement for the complex disease images. Therefore, a novel light weight model is proposed and deployed to mobile terminals for real-time recognition of early apple leaf pests and diseases, providing a higher recognition accuracy under the natural environment.
- AD Convolution is presented to make parameters and calculations smaller for real-time inference of LAD-Net. To recognize apple leaf pests and diseases quickly, the parameters and calculations of LAD-Net deployed on resource-constrained devices must be reduced. Therefore, AD Convolution (Asymmetric and Dilated Convolution) is proposed to replace the standard convolution. It combines the advantage of asymmetric convolution and dilated convolution by adding the dilation rate to asymmetric convolution. Consequently, the number of parameters in AD Convolution is $\frac{6}{25}$ of that in standard 5×5 convolution.
- LAD-Inception (Inception module using Leaky-ReLU and AD Convolution) is established to improve recognition accuracy of LAD-Net. For apple leaf pests and diseases varying in size, the LAD-Inception is built by adding an extra branch and residual connection to improve the ability to extract multiscale features of disease spots. Meanwhile, channel attention is also utilized to allocate the importance of each branch reasonably, as well as to strengthen the capacity to extract more global information, which makes the accuracy of the model higher.
- The lightweight model LAD-Net is deployed and evaluated on mobile terminals in the natural environment for requirement of agricultural actual production, which achieves fast and accurate recognition for early apple leaf pests and diseases. The experimental results show that LAD-Net only has a delay of 15.2ms on HUAWEI P40 and 100.1ms on Jetson Nano, respectively, illustrating that the LAD-Net can reach real-time inference and providing portable technical support.

The remainder of the paper is organized as follows: Section 2 introduces the related work. Section 3 describes a novel recognition model for early apple leaf diseases and pests in the early stage. Section 4 presents the experiment, including the experimental settings and a performance comparison. Section 5 concludes the paper.

2 Related Work

As artificial intelligence continues to improve, breakthroughs in CNNs have occurred in computer vision. A considerable amount of literature has been published on the identification of crop pests and diseases.

In [14], the Kiwi-ConvNet was proposed based on the Kiwi-Inception structures and a dense connectivity strategy. The model could extract the features of three diseases and obtain an accuracy of 98.54%, reaching accuracy of 2.29% and 9.51% that are better than the GoogLeNet and the ResNet-20, respectively. In [15], Zeng proposed a self-attention convolutional neural network, which took full advantage of the extracted features of crop disease spots to identify crop diseases. The recognition accuracy outperformed the state-of-the-art method by 2.9%. In [16], a novel parallel real-time processing framework based on the MASK RCNN and transfer learning was proposed to achieve smart agriculture, and obtained the best accuracy of 96.6% on the Plant Village dataset. In [17], the paper introduced a novel image color histogram transformation technique for generating synthetic images for data augmentation in image classification tasks. The model could be easily deployed for recognizing and detecting cassava leaf diseases in lower quality images, which was a major factor in practical data acquisition. In [18], deep learning and IoT-based solutions were proposed in the literature for plant disease detection and classification. The classification of the 'Custom-Net' was comparable to state-of-the-art models, and it was effective in reducing the training time by 86.67%. The above work greatly promoted the process of smart agriculture, and improved the yield and quality of the crops. However, there is a lack of research on inference speed, and it is difficult to achieve real-time accurate monitoring on resource-constrained devices.

In the recent years, more and more researches focus on the lightweight model, which has small parameters and well accuracy. The existing lightweight model, such as MobileNet [19] and SqueezeNet [20], uses depthwise separable convolution and fire module to reduce the parameters of the model efficiently. In other hand, ShuffleNet [21] introduces the group convolution and proposes the channel shuffle to reduce the calculations of the model. The existing lightweight models have reached the well performance and some lightweight models have proposed to be used to crop diseases recognition. In [22], Xie proposed a deep-CNN-based model Faster DR-IACNN, combining the Inception-v1, SE-blocks and Inception-ResNet-v2 modules. The experimental result showed that the accuracy of the model reached 81.1% with an inference speed of up to 15.01 FPS. In [23], Sun proposed a novel model MEAN by using Apple-Inception, which is built by reconstructing 3×3 convolutions and replacing the 5×5 convolution with an improved 3×3 convolution. The accuracy of the MEAN reached 97.07% and the speed is 12.53 FPS. In [24], an improved YOLOv3 introduced convolution factorization, made parameters smaller and the detection time was only 55ms. In [25], a new low-rank CNN model (LR-Net) based on the CANDECOMP tensor decomposition is developed, which as trained on a specific dataset and implemented in a low-cost Python programmable machine vision camera for real-time classification. These models gained marvelous

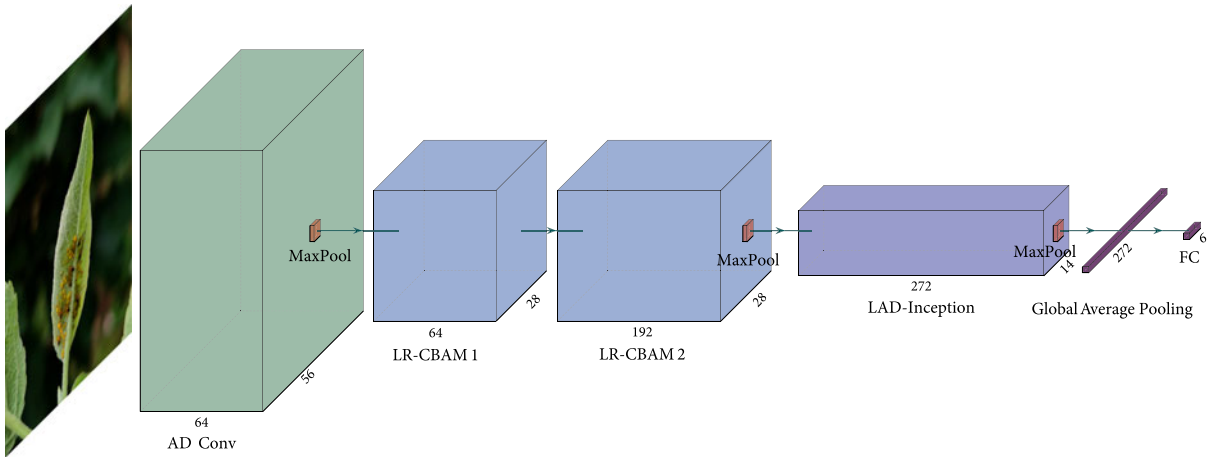


Fig. 1. Model structure of LAD-Net.

performance in terms of speed, but they failed in terms of size and accuracy. A novel lightweight convolutional neural network for mobile resource-constrained devices is urgently needed.

3 LAD-Net: A Novel Lightweight Model

In this section, a novel model, the LAD-Net, is described in detail. First, in Section 3.1, the overall framework is presented to show the structure of the LAD-Net. Then, a new convolution operation, AD Convolution, is proposed in Section 3.2. Next, the improved basic module, the LR-CBAM module, is introduced in Section 3.3. Finally, the core module, the LAD-Inception, is described in Section 3.4.

3.1 The Overall Structure

A novel model named the LAD-Net is built to ensure that the model can be deployed in resource-constrained devices with high accuracy. The overall framework is shown in Fig. 1, showing that the AD Convolution, the LR-CBAM, the LAD-Inception and global average pooling are used to build the model. The detailed parameters are shown in Table 1.

3.2 Asymmetric and Dilated Convolution

Too many convolution kernels result in large model sizes, which restricts models with parameters to be deployed on resource-constrained terminal devices to provide a fast diagnosis of early apple leaf pests and diseases. Therefore, this is the key to reducing the parameters and calculations of the model, as well as make the model adapt to mobile terminal devices. To decrease the parameters and calculations, the AD convolution is proposed to replace the standard convolution.

Asymmetric convolution [26] can ensure that the receptive field is invariant, and can reduce the model's size. The 3×3 convolution is decomposed to 1×3 and 3×1, which is capable of reducing the model's parameters to $\frac{2}{3}$. The asymmetric convolution is shown in Fig. 2a. The asymmetric convolution is divided into two tandem steps. First, $1 \times n$ kernels are used to operate the feature map. Because the size of the convolution kernels is asymmetric, the intermediate feature map is asymmetric. Then, using $n \times 1$ kernels to operate makes the size of the final feature map become $n \times n$.

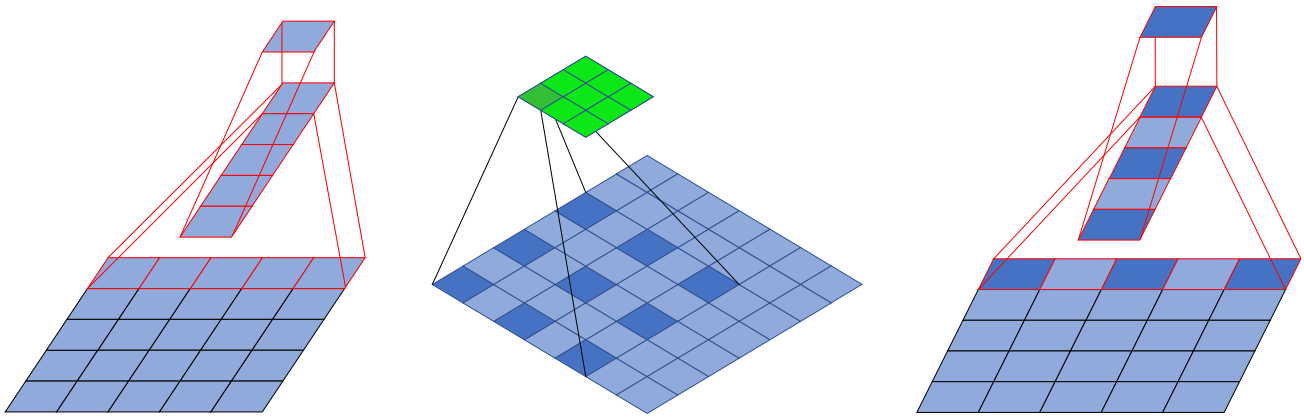
Brown spots, mosaic and powdery mildew spread over the whole leaf, which needs to be recognized according to global information. The dilated convolution [27] is shown in Fig. 2b. Compared with standard convolution, dilated convolution has a hyperparameter dilation rate referring to the number of kernel intervals. Dilation is injected into the standard convolution map to maintain a high resolution without adding a pooling layer as well as to improve the small object detection performance. Therefore, the dilated convolution can incorporate more useful information to reach high accuracy, and thus, to reduce the parameters. As Fig. 2b shows, the parameters of the dilated convolution are reduced to $\frac{9}{25}$ compared with the standard 5×5 convolution.

Based on the description above, it is obvious that the asymmetric convolution and dilated convolution can efficiently reduce the parameters. AD convolution is proposed, which combines the advantages of asymmetric convolution and dilated convolution. The operation of AD convolution can be seen in Fig. 2c. The convolution method is divided into two steps. The first step is to use 3×1 kernels to make

TABLE 1
The LAD-Net and its Parameters

Name	Kernel Size/ Stride	Dilation/ Padding	Output
Input[1]			$224 \times 224 \times 3$
AD Convolution	$3 \times 3/2$	3/3	$56 \times 56 \times 64$
Max Pooling	$3 \times 3/2$	1/1	$28 \times 28 \times 64$
LR-CBAM_1	$1 \times 1/1$	1/1	$28 \times 28 \times 64$
LR-CBAM_2	$3 \times 3/1$	1/1	$28 \times 28 \times 192$
Max Pooling	$3 \times 3/2$		$14 \times 14 \times 192$
LAD-Inception			$14 \times 14 \times 272$
Max Pooling	$3 \times 3/2$		$7 \times 7 \times 272$
CBAM			$7 \times 7 \times 272$
GAP[2]	7×7/-		$1 \times 1 \times 272$
FC[3]			6

[1]*Input : Before being input into the network, the size of the image is normalized to 224 x 224.*
[2]*GAP : Global Average Pooling.*
[3]*The dropout is set to 0.5, effectively reducing the overfitting.*



(a) Asymmetric convolution (b) Dilated convolution (c) AD convolution

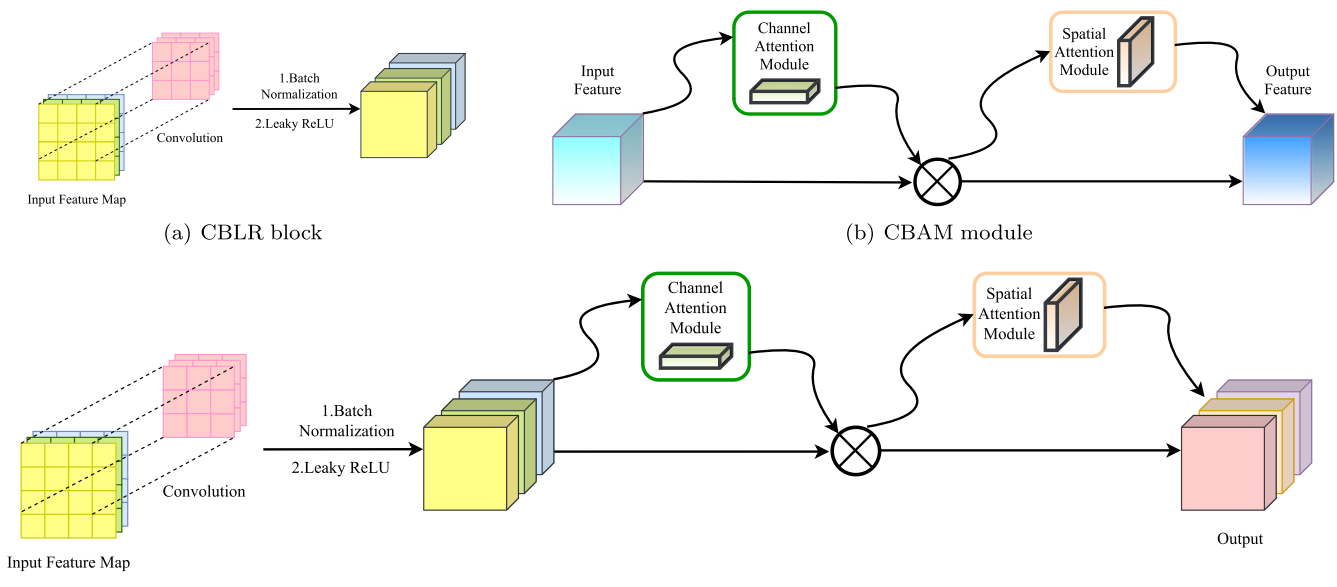
Fig. 2. The operation method of asymmetric convolution, dilated convolution and AD convolution.

convolution operations with the original feature map. The second step uses 1×3 kernels to calculate the generated feature map. The steps are the same as the asymmetric convolution, but the calculations and the parameters are smaller for using the improved kernels that are known as the AD convolution kernels, which adds the dilation rate to the asymmetric convolution kernel for constructing the different-scale kernels. The parameters of AD convolution in Fig. 2c are reduced to $\frac{6}{25}$ compared with the common 5×5 convolution. At the same time, the parameters of AD convolution are less than those of asymmetric convolution and dilated convolution. Using AD convolution can make the size of the model smaller to be easily transplanted to the device with a limited memory. In addition, less parameters aid the generation of the model.

3.3 LR-CBAM Module

Due to the complexity of crop leaf pest and disease recognition, to improve the accuracy of model, a convolutional block attention module [28] (CBAM) is introduced, which combines channel attention and spatial attention. The channel attention mechanism focuses on global information, while spatial attention focuses on local information. The channel attention mechanism assigns different weights to each channel, allowing the network to focus on important features and suppress unimportant features. It enhances the ability of extracting the multiscale diseases spots by using different branches to extract disease spots which have different sizes. Spatial attention mainly focuses on which parts of the input diseased image have more effective information to extract spatial feature of multiscale diseases spots. Besides, the spatial attention can further reduce the calculations of the model, which aids the real-time recognition for early apple leaf pest and disease. The structure of the CBAM is shown in Fig. 3b.

The convolution layer, batch normalization layer, Leaky-ReLU activation layer and CBAM block are combined into an improved module, namely, the LR-CBAM. The structure of the LR-CBAM is shown in Fig. 3c. The batch-normalization layers are used to achieve local normalization, which is



(a) CBLR block (b) CBAM module

(c) LR-CBAM structure

Fig. 3. The structure of the CBLR, CBAM and LR-CBAM, which are basic blocks for constructing the LAD-Net.

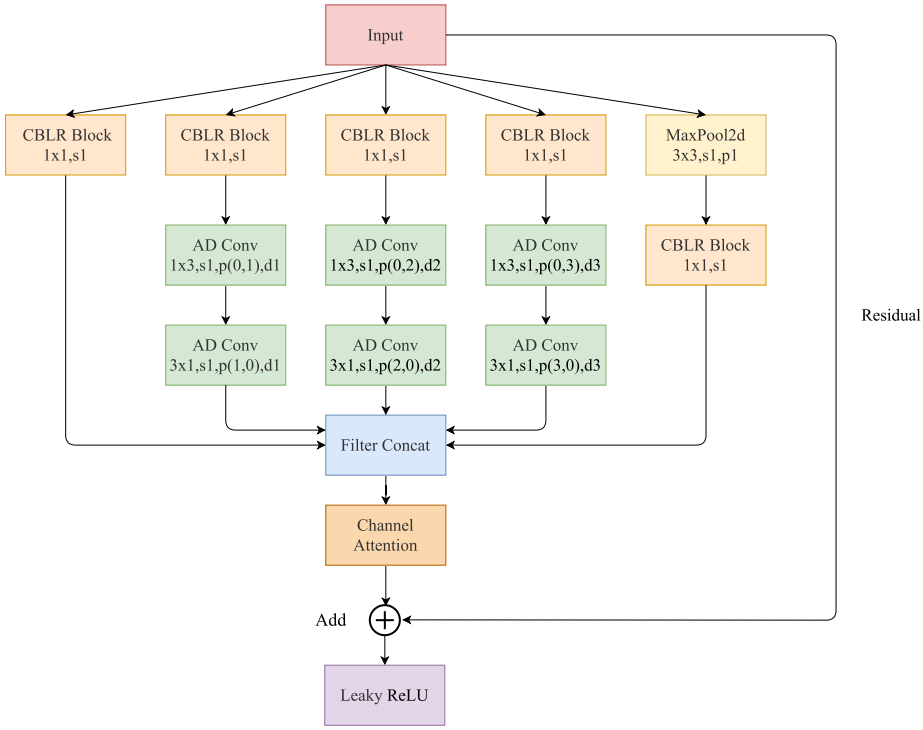


Fig. 4. The structure of the LAD-Inception, introducing AD convolution, the attention mechanism and the residual connection.

also thought to be an effective way to resolve the overfitting problem.

3.4 LAD-Inception

Due to the various sizes and shapes of the pests and diseases, the multiscale feature extraction module LAD-Inception is proposed. The construction of the LAD-Inception is shown in Fig. 4.

3.4.1 Multiscale Branch

Early apple leaf pests and diseases are known to appear in various sizes in images. Alternaria blotch spots and rust spots tend to have a small size compared to the whole leaf, while powdery mildew, brown spot or mosaic spots are spread over the full leaf. The aphids concentrate in part of the leaf. Therefore, a multiscale convolution kernel is provided for feature extraction. The original Inception structure only contains 3×3 convolutions and 5×5 convolutions, which can efficiently extract the local features. To enhance the feature extractability of powdery mildew, brown spots and mosaics, an extra branch of 7×7 convolution is added. In addition, to break symmetry and enhance the generalizability of the network and the gradient propagation, a residual connection is introduced. Based on the multiscale branch introduced, the LAD-Inception has multiscale feature extractability to fit the needs of early apple leaf pests and diseases recognition.

3.4.2 The CBLR Block

The activation function can increase the nonlinearity of the model and improve the expression ability. Furthermore, to restrain neuron death, the activation function, the ReLU, is replaced with the Leaky-ReLU activation function. The expression of the Leaky-ReLU function is shown in (1) as follows:

$$y_i = \begin{cases} x_i & if \quad x_i \geq 0 \\ \dfrac{x_i}{a_i} & if \quad x_i < 0 \end{cases} \tag{1}$$

When x is less than 0, the ReLU activation function directly changes the value of x_i to 0. However, the Leaky-ReLU activation function gives the negative value a nonzero slope, helping to solve the problem of neuron death, which can save more information.

The basic block is called the convolutional layer, the batch normalization layer and the Leaky-ReLU activation function layer (CBLR) block. The construction of the CBLR block is shown in Fig. 3a.

3.4.3 The Channel Attention Mechanism

The importance of multiscale branches in the inception structure should differ, highlighting some branches to further strengthen the feature extractability of this module. A channel attention mechanism is introduced, which enables the network to determine the importance of each output channel in the recognition process. Each layer of the model has many convolution kernels, and each convolution generates a channel. Compared with the spatial attention mechanism, channel attention allocates weights among each convolution channel. Because the inception module has multiscale branches, the different weights of the channels correspond to the importance of the different scale branches, which can allocate greater weight to the important convolution branch.

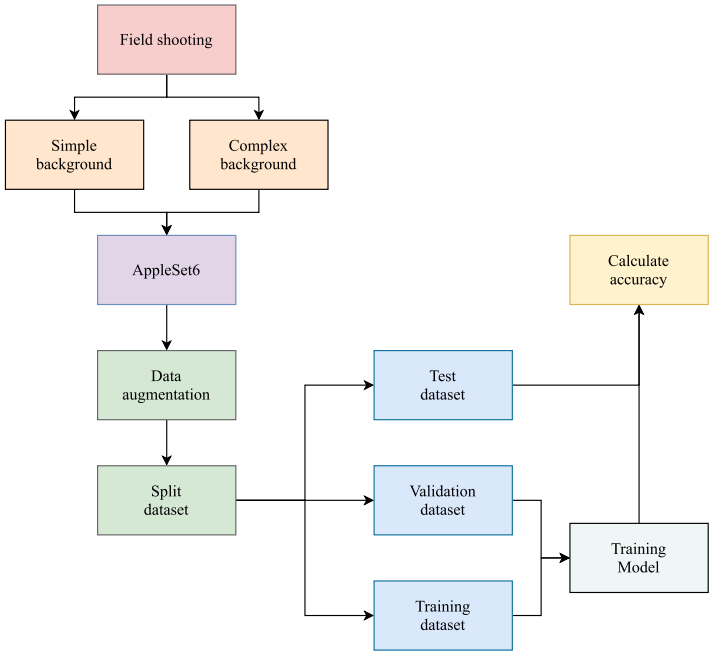


Fig. 5. Flow chart of the LAD-Net for early apple leaf pest and disease recognition.

4 The Experimental Results and Analysis

The experiment is described in this section, including the experimental settings, the performance comparison, the ablation experiments, the performance curve, the feature-based visualization deployment and the generalization of the model.

4.1 Experimental Settings

In this section, the dataset, experimental platform, training strategy and evaluation index are introduced.

4.1.1 Dataset Construction and Augmentation

The process of early apple leaf pest and disease recognition is shown in Fig. 5. First, the original images are captured from apple orchards. Second, the images are expanded through data augmentation. Finally, AppleSet6 is divided into three parts: a training set for training the LAD-Net, a validation set for model evaluation, and a test set for testing the generalization of the model and the rate of partition is 4 : 1 : 1.

In Fig. 6, six common early apple leaf pests and diseases, including the aphid, brown spots, powdery mildew, mosaics, rusts and Alternaria blotches are chosen as the research objects, and their lesions are more widespread to the detriment of the apples quality and quantity. In addition, existing research focuses primarily on the datasets of apple leaf pests and diseases in the late stage. However, much less is known about the early stage. Therefore, those six pests and diseases are shot in QianXian's Apple Monitoring Station to construct the dataset required by the project, which is called AppleSet6. The process of shooting for an image is different views from near and far to simulate the actual application scene. The data is also shot separately in different weather conditions, such as during sunny, cloudy, and rainy days to improve the robustness of the model.

In Fig. 7, various digital image processing technologies, such as image rotation, mirror symmetry, brightness adjustment, salt noise and blur have been applied to the natural training images to simulate the real acquisition environment and to increase the diversity and quantity of the apple diseased training images, which can prevent the overfitting problem, as well as improve the proposed model's generalizability during the training process. The details are illustrated in Table 2. In addition, 5-fold cross-validation is used to ensure replicability of the models.

4.1.2 Experimental Platform

The experimental environment includes an Ubuntu 16.04.2 server, a mobile phone(HUAWEI P40) and an embedded

(a) Aphid (b) BrownSpot (c) Mosaic (d) Rust (e) Alternaria (f) PowderyMildew

Fig. 6. Six common classes of early apple leaf pests and diseases.

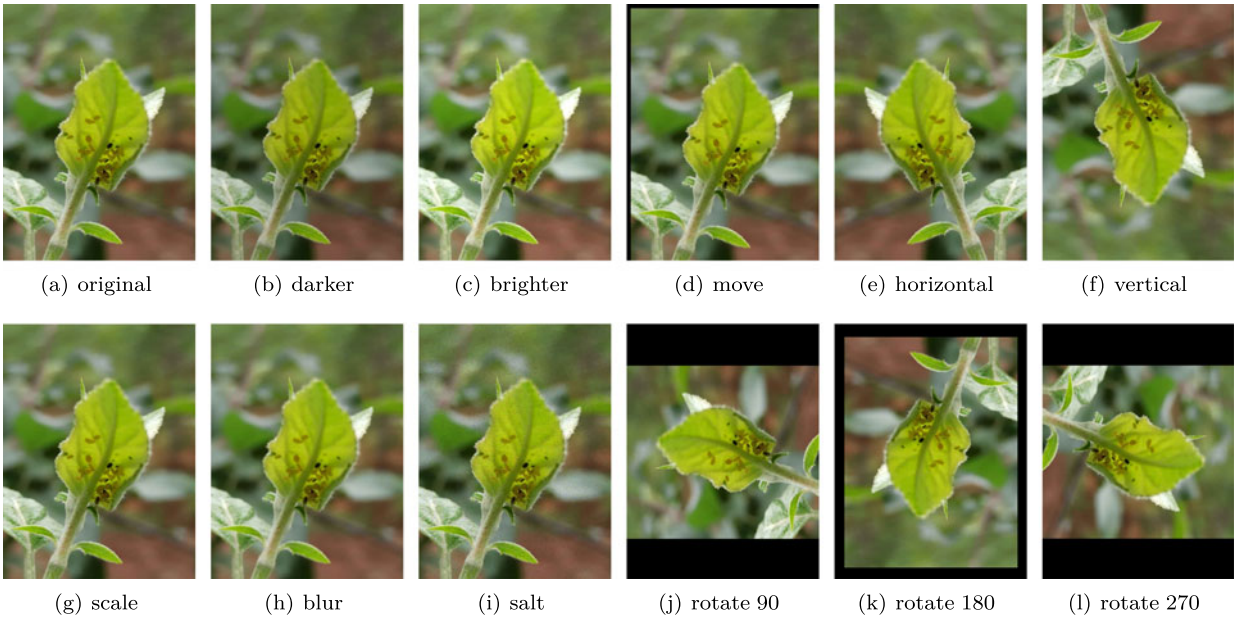


Fig. 7. Data augmentation of an aphid image.

device(Jetson Nano). The detailed configuration of the high-performance server is shown in Table 3. Additionally, PyTorch is utilized to deploy the LAD-Net on the Ubuntu server and on the embedded device.

To perform the application of the model in practical application scenarios, the PaddleLite tool is used to package the model into an app. First, the PyTorch model is converted to the $ONNX$ format. Then, the $ONNX$ format is converted to $pdparams$ supported by Paddle with X2Paddle. Finally, the PaddleLite tool is used to generate an $apks$ file based on Android Studio and deploy it on the mobile terminal. The configuration parameters of the mobile device are listed in Table 4. The application interface is shown in Fig. 8.

Embedded devices are widely used in practical application scenarios. Jetson Nano is a popular embedded device from Nvidia. The Ubuntu operation system is installed, and the PyTorch framework is configured. The detailed information of Jetson Nano, the embedded device, is shown in Table 5.

TABLE 2
Early Apple Leaf Pest and Disease Dataset : AppleSet6

Disease	Total (original)	Total (augmented)	Train	Test	Validation
Aphid	121	1742	1162	290	290
Mosaic	188	2707	1805	451	451
Rust	241	3470	2314	578	578
Brown spot	38	547	365	91	91
Alternaria blotch	136	1958	1306	327	327
Powdery mildew	97	1396	932	232	232
Total	821	11820	7884	1968	1968

TABLE 3
Configuration of High-Performance Server

Configuration	Value
CPU	Intel(R) Xeon(R) E5-2650 v4 @ 2.20 GHz (X2)
GPU	NVIDIA GP100GL [Tesla P100 PCIe 16GB]
RAM	512 GB
Hard Disk	2 TB
OS	Ubuntu 16.04.2 LTS (64-bit)
Language	Python 3.6.1
Cuda version	10.0
Framework	Pytorch 1.4.0, PaddlePaddle 2.1.2

4.1.3 Training Strategy and Evaluation Metrics

In terms of the training strategy, the Adam algorithm dynamically adjusts the learning rate of each parameter using the first and second moment estimation of the gradient. The advantage of the Adam algorithm is that after bias correction, the learning rate of each iteration has a certain range, which makes the parameters relatively stable. Meanwhile, to jump out of the local optimal value, the cosine annealing learning rate is introduced. The settings of the related parameters are shown in Table 6. The larger the batch size is, the better the performance of the gradient descent. So based on past experience and experiment conditions, the batch size is set to 32.

The accuracy, size, latency, computational complexity and parameters are selected for the evaluation metrics.

TABLE 4
Mobile End Configuration

Configuration	Value
Phone type	HUAWEI P40 Pro 5 G
CPU	HUAWEI Kirin 990 5 G
RAM	8G
ROM	256G
Camera resolution	8192×6144
Operating system	Harmony OS 2.0.0

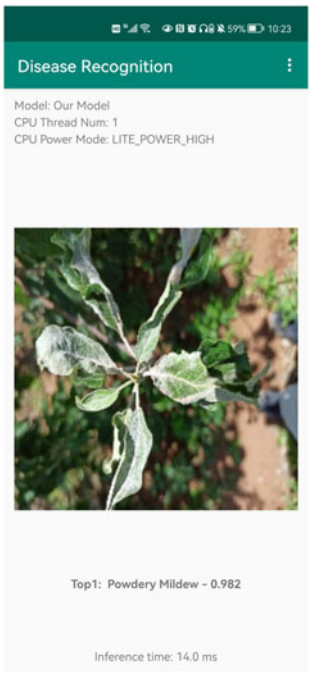


Fig. 8. Early apple leaf pest and disease recognition results on the mobile phone.

Accuracy is an important metric to measure the quality of a model. Apart from high accuracy, the size and the parameter are important factors in determining whether the model can be deployed to the device. Latency and computational complexity are important indicators of whether the model can make an inference in real-time.

4.2 Performance Comparison

Compared with the classical identification network, the accuracy of the LAD-Net is high, and the model size is the smallest, which can be well-deployed to resource-constrained mobile devices. With the purpose of ensuring accuracy, the LAD-Net can maintain a balance between size and latency to meet the needs of the application in crop leaf pest and disease recognition. The experimental results are shown in Table 7.

Although classical CNNs, such as the AlexNet, the VGG and the GoogLeNet, are capable of reaching good accuracy, they cannot meet the need to be deployed to resource-constrained devices and realize real-time inference. For example, the size of the VGG is approximately 500MB, which is too large to be deployed to the device with a small memory.

TABLE 5
Configuration of the Jetson Nano

Configuration	Value
Name	NVIDIA Jetson Nano
CPU	ARM A57 1.43GHZ
Width	32 b
RAM	4 GB
Operating system	Ubuntu 18.04
Language	Python 3.6.4
Deep learning framework	Pytorch 1.6.0

TABLE 6
The Training Parameters and the Training Strategy

Parameter	Value
Optimizer	Adam
Batch size	32
m Initial learning rate	5×10^{-4}
Decay method	Cosine Annealing Learning Rate
Epoch	200

Deeper neural networks, such as the ResNet and the ResNeXt [29], have a nonlinear expression ability and reach a high accuracy. With the deepening of the network layers and the improvement of network expression ability, its accuracy also increases. For example, the accuracy of the ResNet101 is over the ResNet50 and higher than the ResNet34. However, the latency of these deep neural networks is too slow to reach real-time recognition. Some more useful models, such as the DenseNet [30], the EfficientNet [31], the RegNet [32], the InceptionV4 [33] and the DarkNet53 [34], have high accuracy, even reaching 99%. Although their accuracy is slightly higher than that of the LAD-Net, their size is larger than that of the LAD-Net, and their inference speed is slower. For resource-constrained devices being sensitive to the size and speed of the model, it makes them inappropriate for real-time inference on terminal devices.

Recently, an increasing number of lightweight networks have been proposed that have faster speeds and smaller sizes to adapt to a resource-constrained device, such as the Xception [35], the MobileNet, the SqueezeNet, the GhostNet [36], the AsymmNet [37], and the RepVGG [38]. An increasing number of tricks have been proposed to reduce the parameters and calculations, such as group convolution and bottleneck layers. For various sizes of early apple leaf diseases and pests, the light weight networks above cannot recognize early apple leaf pests and diseases under natural complex environment well. When using AD convolution and introducing the LAD-Inception module, the LAD-Net has a smaller size, faster speed and high accuracy, which makes it more adaptive to mobile terminal devices for apple leaf pests and diseases recognition.

Based on the description above, the LAD-Net achieves outstanding and balanced performance, and the recognition accuracy and latency outperform the state-of-the-art methods.

4.3 Ablation Experiments

In this section, ablation experiments are shown in Table 8. Four ablation experiments are conducted to validate the performance of the attention mechanism, the LAD-Inception, the residual connection and the Leaky-ReLU. Additionally, another two ablation experiments are performed to investigate the effects of AD convolution and global average pooling on the model's performance. Fig. 9 shows the performance of the model, and the size of the circle shows the size of the model.

Removing the attention mechanism can slightly accelerate the computation speed of the model, but the accuracy of the model is greatly lost by approximately 0.92%. At the same time, beneficial to multiscale image information

TABLE 7
Performance Comparison

Type	Model	↓Size(MB)	↓Latency(ms)[1]	↑Accuracy(%)[2]	↓CC(GMac)[3]	↓Parameters(M)
Classical[4]	AlexNet	55.68	22.81	90.95	0.31	14.59
	GoogLeNet	39.42	73.0	97.35	1.59	10.32
	VGG-11	491.3	362.3	97.51	7.63	128.79
	VGG-13	492.01	276.2	96.59	11.34	128.98
	VGG-16	512.26	352.1	94.81	15.5	134.29
	ResNet-34	81.34	104.2	98.47	3.68	21.29
	ResNet-50	90.03	198.4	98.83	4.12	23.52
	ResNet-101	162.68	355.9	99.64	7.85	42.51
	ResNeXt-50	88.06	271.7	98.47	4.27	22.99
	ResNeXt-101	331	1041.7	99.79	16.51	86.75
	RegNet	15.13	36.4	98.11	0.25	2.32
	Inception-V4	157	344.8	**99.69**	6.15	41.15
	DenseNet-121	27.12	147.7	97.56	2.88	7.98
	EfficientNet	15.55	62.0	97.30	0.4	4.02
	EfficientNet-V2	77.8	182.1	97.71	2.87	20.19
	DarkNet-53	155.08	248	96.03	7.13	40.6
Light[5]	Xception	79.68	259.1	99.59	4.59	20.82
	MobileNet-V2	8.74	25.6	97.28	0.32	2.23
	MoblieNet-V3	16.66	27.03	97.03	0.31	4.32
	SqueezeNet-V1	2.84	47.8	97.16	0.74	0.74
	ShuffleNet-V2	4.97	**13.8**	97.04	0.15	1.26
	GhostNet	15.11	48.6	99.49	0.15	3.91
	RepVGG	30.05	57.0	94.07	1.52	7.84
	AsymmNet	7.19	54.0	94.07	**0.06**	1.86
	LAD-Net	**1.25**	15.2	98.58	0.17	**0.32**

[1]*Latency : The latency of the model recognizing an image using a single large core on HUAWEI P40.*
[2]*Accuracy : Accuracy on AppleSet6.*
[3]*CC : Computational complexity.*
[4]*Classical : Classical deep CNNs.*
[5]*Light: New Light weight CNNs.*

extraction, the accuracy increases by 0.82% and the size decreases by 1.11MB due to the use of the LAD-Inception. The residual module is used to help improve the expressiveness of the model and simplify the training process. The Leaky-ReLU activation function is used, which can slightly increase the accuracy. It is obvious that these four measures are useful to increase the accuracy of the model.

Using AD convolution to replace standard 5×5 and 7×7 convolution is beneficial to accelerate the operation speed and reduce the size. When using AD convolution, the size of the model decreases by 0.24MB and the accuracy is lost slightly. In addition, using global average pooling to replace the fully connected layer can reduce the parameters and accelerate the inference speed. The size of the model, without global average pooling, increases to 15.08MB. Global average pooling can efficiently extract the global feature and save the spatial information compared to flatten the feature map to a vector. Thus, the accuracy of using the GAP is higher than without it. In

TABLE 8
Ablation Experiment

AM[1]	Inception	Residual	LR[2]	AD[3]	GAP[4]	↑Accuracy(%)[5]	↓Size(MB)	↓Latency(ms)[6]	↓CC(GMac)[7]
✗	✓	✓	✓	✓	✓	97.66	1.16	**12.8**	0.173
✓	✗	✓	✓	✓	✓	97.76	2.36	16.0	0.233
✓	✓	✗	✓	✓	✓	97.86	**1.04**	14.4	0.163
✓	✓	✓	✗	✓	✓	98.11	1.25	15.8	0.173
✓	✓	✓	✓	✗	✓	**98.78**	1.49	51.4	0.700
✓	✓	✓	✓	✓	✗	97.86	15.08	16.8	0.177
✓	✓	✓	✓	✓	✓	98.58	1.25	15.2	0.173

[1]*AM: Attention Mechanism.*
[2]*LR : Leaky-ReLU activation function.*
[3]*AD : Asymmetric and Dilated Convolution.*
[4]*GAP : Global Average Pooling.*
[5]*Accuracy : Accuracy on AppleSet6.*
[6]*Latency : Latency on HUAWEI P40.*
[7]*CC : Computational complexity.*

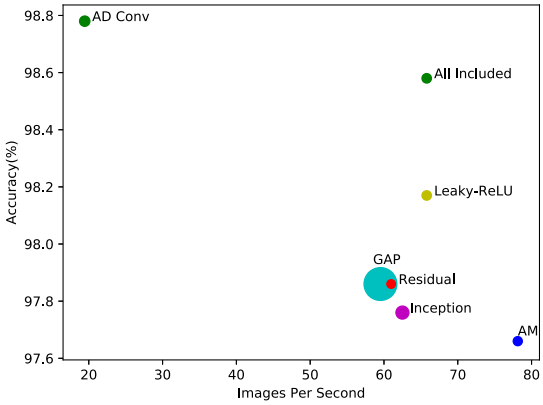


Fig. 9. The ablation experiment results. The size of the circle represents the size of the model.

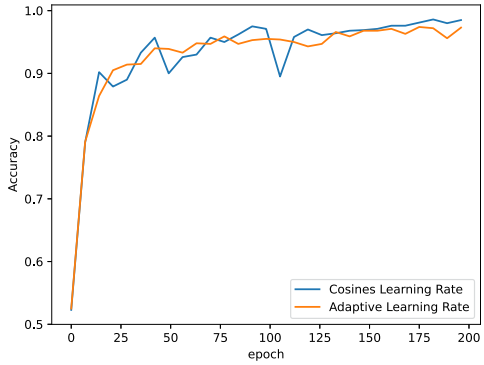


Fig. 11. The performance curve of the cosine annealing learning rate and the adaptive learning rate.

conclusion, the AD convolution and the global average pooling can efficiently decrease the size of the model.

4.4 Performance Curve

The cosine annealing learning rate can jump out of the local optimum, and can possibly be close to the global optimum to obtain higher accuracy. The equation of the cosine learning rate is in (2) and shown in Fig. 10. As in the equation, the learning rate is increased to μ_{max} when reaching the fixed round, which is at 7, 21, 49, 105 and 217 epochs.

The curve of accuracy is shown in Fig. 11. According to the figure, the curve with the cosine annealing learning rate fluctuates greatly, while the curve with the adaptive learning rate fluctuates gently. In addition, the highest accuracy when using the cosine annealing learning rate is higher than that when using the adaptive learning rate. Without using the cosine annealing learning rate, the accuracy is 98.11%, which is less than that obtained using the cosine annealing learning rate. The equation can be seen as follows:

$$\mu_t = \mu_{min} + \frac{1}{2}(\mu_{max} - \mu_{min})\left(1 + cos\left(\frac{T_{cur}}{T_{max}}\pi\right)\right) \tag{2}$$

- μ_t : Learning rate
- μ_{min} : Minimum learning rate
- μ_{max} : Maximum learning rate

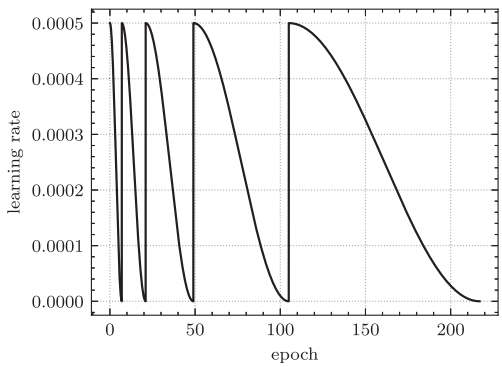


Fig. 10. The curve of the cosine annealing learning rate.

- T_{cur} : Current iteration
- T_{max} : Maximum iteration

4.5 The Evaluation Index of the Models Performance

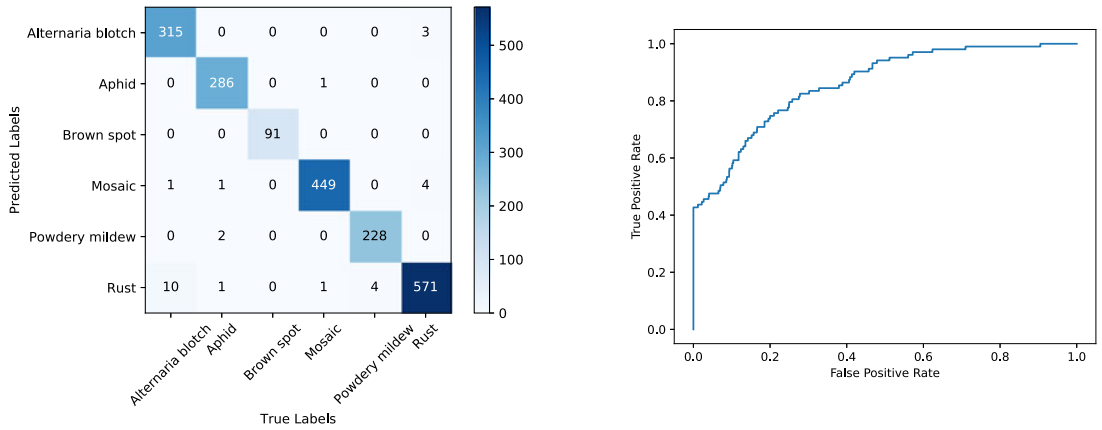
The result of the confusion matrix is shown in Fig. 12a. The precision, recall, specificity and F1 score are shown in Table 9. It can be seen that the recognition of the brown spot is the best, with the highest $F1$ score of 1.0. The $F1$ scores of the Aphid, Mosaic and Powdery mildew reach 0.9914, 0.9914 and 0.9869, respectively. The recognition accuracy for the rust and the Alternaria blotch is slightly lower, with $F1$ scores of 0.9804 and 0.9783, respectively. Early rust appears in small orange spots, resulting in difficult recognition. In the early stage of the Alternaria blotch, the size varies, and thus, it is difficult to extract the unified characteristics. In addition, the features of the Alternaria blotch spots are similar to those of the Rust spots. Therefore, eleven images of the Alternaria blotch are recognized as rust.

The ROC curve is shown in Fig. 12b, and the AUC value is approximately 0.8602, which can effectively solve the problems of category imbalance. The AUC value shows that the LAD-Net is capable of identifying most kinds of early apple leaf pests and diseases, and it is useful and effective for use in actual application scenarios.

4.6 Feature-Based Visualization

Based on the experimental results above, the LAD-Net achieves a good balance among accuracy, size and latency. To determine whether the features extracted from the network are effective and observe the features of regions that have the greatest influence on the network prediction results, the Grad-CAM is used to visualize the images. If the temperature of a certain region in the input image is higher, it indicates that the region has a greater influence on the output of the network results.

According to the visualization results shown in Fig. 13, all areas of leaf pests and diseases show higher temperatures than other areas; hence, the LAD-Net can extract and utilize the characteristics of leaf pests and diseases well. The rust and Alternaria spots are two indistinguishable small spot diseases, and thus, the high temperature areas are mainly concentrated in the spot of the disease, as Figs. 13a and 13d show. Rust spots are prominent, while the



(a) Confusion matrix (b) The ROC curve

Fig. 12. Evaluation index of the LAD-Net performance.

Alternaria spots are depressions, which our model can still classify correctly. In Fig. 13e, areas of high temperature are concentrated where aphids congregate. Mosaic, powdery mildew and brown spots are three diseases that usually infect the whole leaf. Leaves with mosaics have yellow spots, leaves with powdery mildew are covered with white powder and become curly, and leaves with early brown spots are inconspicuous, but the whole leaf turns yellow. At the same time, the visualization results in Figs. 13b, 13c and 13f show that for brown spots, mosaics, powdery mildew and the whole leaf with all three diseases show higher temperatures.

By visualizing the images input into the network, the diseased areas of the apple leaves have a greater impact on the output of the network. This more intuitively shows that the LAD-Net can effectively extract the features of the diseased areas and accurately classify the disease through the features.

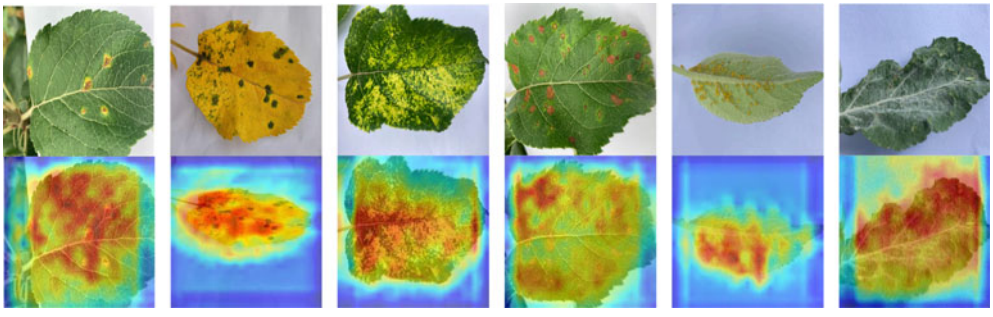
TABLE 9
F1 Score of Different Diseases

Diseases	Precision	Recall	Specificity	F1[1]
Alternaria blotch	0.991	0.966	0.998	0.9783
Brown spot	1.0	1.0	1.0	1.0
Mosaic	0.987	0.996	0.996	0.9914
Rust	0.973	0.988	0.988	0.9804
Aphid	0.997	0.986	0.999	0.9915
Powdery mildew	0.991	0.983	0.999	0.9869

[1] *F1 : harmonic mean.*

4.7 Deployment on the Embedded Equipment

The model is transplanted to the embedded device, Jetson Nano. On the embedded device, the recognition latency is only 100.1ms, which can meet the practical application requirements. Meanwhile, the device is assembled into a small robot. To improve the ability for automatic inspection, the camera, battery, etc. are equipped. The robot can automatically move in a simulated orchard environment and recognize leaf pests and diseases. A simulation experiment is conducted to verify whether the LAD-Net can be used for automatic inspection on the robot. The robot connects to the WLAN and remotely logs in to the robot by shell using SSH. Then, the robot follows the planned line around the apple tree while capturing an early apple leaf disease image via its camera. At the same time, the LAD-Net recognizes the image in real-time, and the result is shown in the terminal. The shape of the robot and the terminal is shown in Fig. 14. The development environment is based on the Ubuntu 18.04 operating system, PyTorch framework and Python language.



(a) (b) (c) (d) (e) (f)

Fig. 13. Class activation maps of six kinds of early apple leaf diseases via gradient- based localization. (a) Rust. (b) Brown spot. (c) Mosaic. (d) Alternaria blotch. (e) Aphid. (f) Powdery mildew.

(a) Robot (b) Terminal

Fig. 14. Simulation robot inspection experiment.

4.8 The Generalization of the LAD-Net

To evaluate the generalization of the LAD-Net, two public and existing datasets are chosen, including LateAppleSet, Tomato DataSet of AI Challenge. The performance of the LAD-Net on these datasets is shown in Table 10.

LateAppleSet is a public dataset shared on Baidu's AI Studio that includes five diseases, such as rust, Alternaria boltch, brown spot, mosaic and gray spot. The accuracy of the LAD-Net is the highest among other lightweight models, proving that the model can not only be used to recognize early apple pests and diseases but can also recognize late apple pests and diseases accurately.

The Tomato DataSet of AI Challenge Data Set is a public dataset that includes 9 tomato leaf diseases. In [39], RRDN is proposed and obtains an accuracy of 95%, which verifies its satisfactory performance. However, the LAD-Net can reach a high accuracy of 97.66%, which is more accurate than the other networks. In conclusion, the LAD-Net has good performance on public tomato leaf diseases, proving that the LAD-Net can effectively recognize the various kinds of crop leaf diseases.

Based on the experimental results above, it is clear that the LAD-Net has the well generalization and robustness to adapt complex tasks of recognizing various crop leaf diseases accurately and easily be deployed to the mobile terminal devices.

TABLE 10
The Performance of the LAD-Net on a Public Data Set

Dataset	Model	Accuracy(%)	Size(MB)
LateAppleSet[1]	LAD-Net	**97.72**	1.25
	MobileNet-V2	97.44	8.70
	SqueezeNet	97.11	2.83
	ShuffleNet	97.23	4.92
	GhostNet	97.55	15.18
	RepVGG	97.00	30.04
	AsymmNet	97.51	7.22
Tomato9[2]	LAD-Net	**97.92**	1.26
	RRDN	95.00	–
	MobileNet-V2	97.13	8.72
	SqueezeNet	96.09	2.84
	ShuffleNet	96.27	4.99
	GhostNet	96.15	15.20
	RepVGG	97.04	30.06
	AsymmNet	97.67	7.24

[1]*LateAppleSet : Apple leaf disease in the late stage including five categories.*
[2]*Tomato9 : Data set including nine kinds of tomato leaf diseases from the AI Challenge 2018.*

5 Conclusion

In this paper, a novel light weight model, the LAD-Net with high accuracy, small size and fast speed, is proposed and deployed to smart terminals for real-time recognition of early apple leaf pests and diseases. To reduce the calculation and parameter, a new convolution operation AD Convolution is presented to take the place of all standard convolutions. Meanwhile, the LAD-Inception is built to enhance the ability to extract the features of multiscale diseases and pests, as well as to utilize a channel attention mechanism to determine the importance of each channel. Finally, the LAD-Net is constructed by using CBLR block, LAD-Inception and global average pooling.

The LAD-Net is trained on the Ubuntu Server and transplanted to HUAWEI P40 and Jetson Nano. The recognition performance can reach 98.58% on AppleSet6, the size is only 1.25MB, and the speed of the LAD-Net is fast. The latency only reaches 15.2ms on HUAWEI P40 and 100.1ms on Jetson Nano. The experimental results show that the LAD-Net can accurately recognize early apple leaf diseases and pests, and provide a new method for real-time inference on resource-constrained devices.

Although our work has achieved some success, due to the influence of objective factors, there is still room for improvement. When training the network, hyperparameter optimization tools will be introduced to further improve the accuracy and speed of the model. On the another hand, the more applications in the actual agricultural production by robot inspection will be taken to push the smart agriculture forward.

Author Contributions

Xianyu Zhu carried out the primary work, including proposing the idea, designing the experiment, implementing and verifying the method, and writing and revising the manuscript. **Jinjiang Li** helped to perform the experiment and write the paper. **Bin Liu** made a significant contribution by giving constructive advice, providing the research environment, offering funding and correcting the draft. **Runchang Jia** and **Zhuohan Yao** joined to transfer the model to the Android and Linux

platforms. **HaiXi Zhang**, **Aihong Yuan** and **Yingqiu Huo** read the paper and provided advice.

ACKNOWLEDGMENTS

The authors are thankful for Ziyue You and Cong Yu for helping to modify the paper and giving constructive advice.

REFERENCES

[1] X. Wang, Z. Xu, S. Tang, and R. A. Cheke, "Cumulative effects of incorrect use of pesticides can lead to catastrophic outbreaks of pests," *Chaos, Solitons Fractals*, vol. 100, pp. 7–19, 2017.

[2] S. Zhang, W. Huang, Y.-A. Huang, and C. Zhang, "Plant species recognition methods using leaf image: Overview," *Neurocomputing*, vol. 408, pp. 246–272, 2020.

[3] R. R. R. F. M. Anu S. and Nisha T., "Fruit disease detection using GLCM and SVM classifier," *Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol.*, vol. 5, pp. 365–371, 2019.

[4] M. R. Kale and M. S. Shitole, "Analysis of crop disease detection with SVM, KNN and random forest classification," *Inf. Technol. Ind.*, vol. 9, no. 1, pp. 364–372, 2021.

[5] M. Bhagat, D. Kumar, I. Haque, H. S. Munda, and R. Bhagat, "Plant leaf disease classification using grid search based SVM," in *Proc. Int. Conf. Data, Eng. Appl.*, 2020, pp. 1–6.

[6] A. Almadhor, H. T. Rauf, M. I. U. Lali, R. Damaševičius, B. Alouffi, and A. Alharbi, "Ai-driven framework for recognition of guava plant diseases through machine learning from DSLR camera sensor based high resolution imagery," *Sensors*, vol. 21, no. 11, 2021, Art. no. 3830.

[7] S. Zhang, W. Huang, and Z. Wang, "Combing modified grabcut, k-means clustering and sparse representation classification for weed recognition in wheat field," *Neurocomputing*, vol. 452, pp. 665–674, 2021.

[8] D. Agarwal, M. Chawla, and N. Tiwari, "Plant leaf disease classification using deep learning: A survey," in *Proc. Int. Conf. Inventive Res. Comput. Appl.*, 2021, pp. 643–650.

[9] R. Patil and S. Kumar, "Bibliometric survey on diagnosis of plant leaf diseases using artificial intelligence," *Int. J. Modern Agriculture*, vol. 9, no. 3, pp. 1111–1131, 2020.

[10] P. Deepika and S. Kaliraj, "A survey on pest and disease monitoring of crops," in *Proc. Int. Conf. Signal Process. Commun.*, 2021, pp. 156–160.

[11] P. D. Kalwad, S. G. Kanakaraddi, T. Preeti, S. Ichalakaranji, S. Salimath, and S. Nayak, "Apple leaf disease detection and analysis using deep learning technique," in *Information and Communication Technology for Competitive Strategies*. Berlin, Germany: Springer, 2022, pp. 803–814.

[12] S. Zhang, W. Huang, and Z. Wang, "Plant species identification based on modified local discriminant projection," *Neural Comput. Appl.*, vol. 32, no. 21, pp. 16 329–16 336, 2020.

[13] S. Zhang, C. Zhang, and X. Wang, "Plant species recognition based on global-local maximum margin discriminant projection," *Knowl.-Based Syst.*, vol. 200, no. 2, 2020, Art. no. 105998.

[14] B. Liu, Z. Ding, L. Tian, D. He, and H. Wang, "Grape leaf disease identification using improved deep convolutional neural networks," *Front. Plant Sci.*, vol. 11, pp. 1082–1099, 2020.

[15] W. Zeng and M. Li, "Crop leaf disease recognition based on self-attention convolutional neural network," *Comput. Electron. Agriculture*, vol. 172, pp. 105 341–105 357, 2020.

[16] Z. U. Rehman *et al.*, "Recognizing apple leaf diseases using a novel parallel real-time processing framework based on mask RCNN and transfer learning: An application for smart agriculture," *IET Image Process.*, vol. 15, no. 10, pp. 2157–2168, 2021.

[17] O. O. Abayomi-Alli, R. Damaševičius, S. Misra, and R. Maskeliūnas, "Cassava disease recognition from low-quality images using enhanced data augmentation model and deep learning," *Expert Syst.*, vol. 38, no. 7, 2021, Art. no. e12746.

[18] N. Kundu *et al.*, "IoT and interpretable machine learning based framework for disease prediction in pearl millet," *Sensors*, vol. 21, no. 16, pp. 5386–5408, 2021.

[19] A. G. Howard *et al.*, "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*.

[20] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "SqueezeNet: Alexnet-level accuracy with 50x fewer parameters and < 0.5 mb model size," 2016, *arXiv:1602.07360*.

[21] X. Zhang, X. Zhou, M. Lin, and J. Sun, "Shufflenet: An extremely efficient convolutional neural network for mobile devices," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 6848–6856.

[22] X. Xie, Y. Ma, B. Liu, J. He, S. Li, and H. Wang, "A deep-learning-based real-time detector for grape leaf diseases using improved convolutional neural networks," *Front. Plant Sci.*, vol. 11, 2020, Art. no. 751.

[23] H. Sun *et al.*, "MEAN-SSD: A novel real-time detector for apple leaf diseases using improved light-weight convolutional neural networks," *Comput. Electron. Agriculture*, vol. 189, pp. 106 379–106 390, 2021.

[24] X. Wang, J. Liu, and X. Zhu, "Early real-time detection algorithm of tomato diseases and pests in the natural environment," *Plant Methods*, vol. 17, no. 1, pp. 1–17, 2021.

[25] L. Falaschetti *et al.*, "A low-cost, low-power and real-time image detector for grape leaf esca disease based on a compressed CNN," *IEEE Trans. Emerg. Sel. Topics Circuits Syst.*, vol. 11, no. 3, pp. 468–481, Sep. 2021.

[26] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 2818–2826.

[27] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," in *Proc. Int. Conf. Learn. Representations*, 2016, pp. 1–13.

[28] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 3–19.

[29] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 5987–5995.

[30] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 4700–4708.

[31] M. Tan and Q. V. Le, "Efficientnetv2: Smaller models and faster training," in *Proc. Int. Conf. Mach. Learn.*, 2021, pp. 10 096–10 106.

[32] I. Radosavovic, R. P. Kosaraju, R. Girshick, K. He, and P. Dollár, "Designing network design spaces," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 10 428–10 436.

[33] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Proc. 31st AAAI Conf. Artif. Intell.*, 2017, pp. 4278–4284.

[34] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," 2018, *arXiv:1804.02767*.

[35] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 1251–1258.

[36] K. Han, Y. Wang, Q. Tian, J. Guo, C. Xu, and C. Xu, "GhostNet: More features from cheap operations," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 1580–1589.

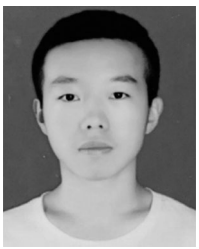
[37] H. Yang, Z. Shen, and Y. Zhao, "AsymmNet: Towards ultralight convolution neural networks using asymmetrical bottlenecks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 2339–2348.

[38] X. Ding *et al.*, "": Making VGG-style convnets great again," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 13 733–13 742.

[39] C. Zhou, S. Zhou, J. Xing, and J. Song, "Tomato leaf disease identification by restructured deep residual dense network," *IEEE Access*, vol. 9, pp. 28 822–28 831, 2021.

Xianyu Zhu was born in Anhui. He currently majors in computer science and technology with Northwest A&F University, China. His research interests include computer vision, object detection, and deep learning.

Jinjiang Li was born in Yunnan. He currently majors in software engineering with Northwest A&F University, China. His research interests include deep learning, object detection, and image recognition.



Runchang Jia was born in Shaanxi. He currently majors in information management and information systems with Northwest A&F University, China. His research interests include deep learning and computer vision.

Bin Liu was born in Shaanxi in 1981. He received the BS degree in computer science and technology from the Shaanxi University of Science and Technology, China, in 2004, the MSc degree in technology with a major in parallel computing and cloud computing from Yunnan University, China, in 2010, and the PhD degree in electronic and information engineering from Xi'an Jiaotong University, China, in 2014. Since 2018, he has been an associate professor and doctoral supervisor with the College of Information Engineering, Northwest A&F University, China, where he has obtained a postdoctoral fellow with the College of Mechanical and Electronic Engineering. His research interests include deep learning and computer vision. He currently serves as a reviewer for the *IEEE Transactions on Computers*, *Computers and Electronics in Agriculture*, and *Biosystems Engineering*, among other journals.

Zhuohan Yao was born in Shaanxi. She currently majors in computer science and technology with Northwest A&F University, China. Her research interests include deep learning and computer vision.

Aihong Yuan received the PhD degree from the Center for Optical Imagery Analysis and Learning (OPTIMAL), Xian Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xian, Shaanxi, China, in 2019. He is a lecturer with the Computer Science Department, Northwest A&F University. His research interests include machine learning, image/video content understanding, and deep learning

Yingqiu Huo was born in Tangshan, China, in 1978. He received the BEng degree in computer science and technology, the MEng, and PhD degrees in agricultural electrification and automation from Northwest A&F University Yangling, China, in 2003, 2006 and 2015, respectively. Since 2016, he has been an associate professor with the College of Information Engineering, Northwest A&F University. His research interests include machine learning and computer vision.

Haixi Zhang received the BE degree in electronic and information engineering and the ME degree in signal and information processing from Northwestern Polytechnical University in 2011 and 2014, respectively, and the PhD degree in electronic science and technology from Northwestern Polytechnical University in 2020. He is currently an assistant professor with the College of Information Engineering, Northwest A&F University, Yangling, China. From 2015 to 2017, he was a visiting scholar with UNC-Charlotte. His research interests include machine learning, computer vision, and agriculture information.