HLNet Model and Application in Crop Leaf Diseases Identification

Yanlei Xu, Shuolin Kong, Zongmei Gao, Qingyuan Chen, Yubin Jiao and Chenxiao Li

Abstract: Crop disease has been a severe issue for agriculture, causing economic loss for growers. Thus, disease identification urgently needs to be addressed, especially for precision agriculture. As of today, deep learning has been widely used for crop disease identification combined with optical imaging sensors. In this study, a lightweight convolutional neural network model is designed and validated on two publicly available imaging datasets and one self-built dataset with 28 types of leaf and leaf disease images of 6 crops as the research object. This model is an improvement of the existing convolutional neural network, reducing the floating-point operations by 65%. In addition, dilated depth-wise convolutions were used to increase the network receptive field and improve the model recognition accuracy without affecting the network computational speed. Meanwhile, two attention mechanisms are optimized to reduce attention module computation, improving the capability of the model to select the correct regions of interest. After training, this model achieved an average accuracy of 99.86%, and the image calculation speed was 0.173 s. Comparing with 11 backbone models and 5 latest crop leaf disease identification studies, the proposed model achieved the highest accuracy. Therefore, this model with an advantage of balancing between the calculation speed and recognition accuracy. Furthermore, the proposed model provides a theoretical basis and technical support for the practical application and mobile terminal applications of crop disease recognition in precision agriculture.

Keywords: lightweight convolutional neural network; crop leaf disease identification; self-attention; deep learning

1. Introduction

Currently, the growth rate of global food production is much lower than the growth rate of the population. Food production and security are of great significance to ensure people’s living standards and the development of the national economy [1]. Crop disease is one of the main disasters affecting agricultural production, and the scope and extent of the disease seriously restrict the productivity, quality, and sustainable development of agricultural production. One of the main bases for discovering diseases and identifying the types of diseases is the symptoms of diseased leaves [2]. The timely grasp of plant diseases in type, severity and development can effectively reduce the economic losses to agricultural and forestry production, and at the same time reduce the environmental pollution caused by the abuse of pesticides and provide a basis for disease prevention and control strategy [3].

Traditional artificial identification of diseases mainly relied on experts by evaluating the visible features, such as symptoms [4]. Artificial identification is easily misled by subjective factors and is difficult to meet the requirements of efficient disease identification [5]. The methods of identifying plant diseases based on computer vision firstly...
extract the characteristics of the diseases using algorithms, and then identify the diseases using traditional identification methods [6-8]. However, the process of extracting features is complicated, and effectively extracting new disease features is challenging. So, it is not suitable for multi-region identification tasks [9]. In recent years, the rapid development of deep learning had achieved success in image analysis of plant leaf diseases [10]. Disease identification based on convolutional neural network (CNN) models, such as corn [11,12], apple [13], rice [14], and potato [15], had the advantages of high precision and wide versatility, providing effective technical means for accurate and rapid identification of plant diseases.

Although the model used in the above papers has high accuracy in identifying leaf diseases, identification speed is slow for large-scale datasets, which is not conducive to practical applications. In view of many calculations and calculation speed of CNN models, many lightweight models have been proposed, such as SqueezNet [16], MobileNet [17], Xception [18], and ShuffleNet [19]. Many studies identifying agricultural diseases are also based on lightweight models. Kamal, K. C. [20] proposed a lightweight convolutional neural network with depth-wise separable convolution. This study achieved 97.65% accuracy in the PlantVillage dataset. Wagle, S. A. et al. [21] proposed a more compact model based on AlexNet, which consumed the lowest 14.8 MB of memory among the three variants and achieved a maximum accuracy of 99.75% for the identification of nine plant diseases in PlantVillage. Wang, P. et al. [22] proposed a model based on ShuffleNetV2 with lightweight attention, collected images of five common leaf diseases of grape, and the results showed that it can achieve 99.66% accuracy. Bi, C. K. et al. [23] proposed a disease identification model based on MobileNet and the residual structure. They realized over 73.50% classification accuracy around two different types of apple diseases. Bhujel, A. et al. [24] built a lightweight convolutional neural network by combining with different attention modules, and then used tomato leaf diseases in the PlantVillage leaf disease database to train model. The best average accuracy of identifying diseases after training was 99.34%. Chao, X. F. et al. [25] fused the Squeeze-and-Excitation (SE) module with the Xception network, by reducing the depth and width of the model to build SE_miniXception. This model achieved 97.01% accuracy for six apple leaf diseases.

The above studies about diseases identification based on lightweight CNN model can identify the diseases with high accuracy. While some of them used one crop variety as study object, for example in studies [22–25]. In practical applications, the types of crops are complex, while the CNN model can only identify one type of crop, which cannot meet the actual requirements for identification of multiple crop types. While, some studies, for example references [20,21], take a variety of crops as study object, which can identify a wider type of crop leaf diseases. However, models for identifying multiple crop diseases can be more lightweight.

To resolve the above issues, we proposed a lightweight CNN model with fast identifying speed, high accuracy, which combined a channel attention mechanism (CA) and spatial attention mechanism (SA), and balanced the calculation speed and model size. It has a wider range of applications than single type crop disease identification models. It is lighter and faster than multiple type crop identification models. Specifically, the objectives of this study were to: (1) simplify the overall structure of the ShuffleNetV1 model to reduce the computational effort of the model and increase its computational speed; (2) supply dilated convolution to the model to expand the model perceptual field and improve the model identification accuracy; (3) improve the existing attention mechanism, which can increase its attention range and computational depth, with reducing attention mechanism arithmetic power consumption.

2. Materials and Methods

The study was mainly divided into two parts. In the data building part, multiple public datasets and self-built datasets were mixed to improve the robustness of the model and to expand data volume and unified data size through data pre-processing and data
expansion. The data was labelled with its corresponding disease and then divided into training data and test data. In the model improvement part, we improved the structure of the convolutional neural network and added the dilated depth-wise convolutions. The identification accuracy of each model was compared, and the optimal model was selected. Then, we improved the attention module based on the previously selected optimal model and designed two models incorporating the optimized and unoptimized attention modules, respectively. The comparisons with other CNN models (backbone models and crop leaf disease identification models) verified the effectiveness of our model. The flowchart of this study is shown in Figure 1.

![Figure 1. Research flowchart used in this study to identify plant leaf disease.](image)

**2.1. Dataset and Expansion**

In this study, the public datasets and the collected data were mixed to establish the research data sets, where the public datasets were from PlantVillage [26] and UC Irvine Machine Learning Repository [27]. The self-collected data includes 90 images of northern leaf blight for corn and 300 images of black spot for tomatoes from the experimental field of Jilin Agricultural University. The image capture device was a smartphone oneplus8P, the main camera lens was 48 million pixels, and the captured picture pixels were 3000 × 3000. The experimental data with a total of 20,490 images included 6 types of crops and 28 types of leaf diseases. Specifically, the images used in this study are digital RGB images and the type of images are in JPG format.

Using multiple image expansion methods can effectively improve model robustness and avoid over-fitting, this paper expanded the original sample data to three times size using two random 40-degree rotations and horizontal flips. In the expansion process, we name the corresponding expanded image for each original image in the format “image name a, b, c”. In the process of separating the validation data from the training data, each original image and the corresponding expanded images are considered as a whole and moved to training data or validation data overall. They are not divided to different data to avoid confusion of the images. Then, the sample data was divided into a training set and validation set according to the ratio of 8:2. For obtaining a higher training speed while maintaining a good identification rate, an interpolation algorithm was used to uniformly compress the input images to 224 × 224 pixels. The information of each crop leaf disease can be obtained in Table A1 in Appendix A. The display image of crop diseases is shown in Figure 2.
2.2. HLNet

HLNet was a disease identification model with high computational speed and lightweight structure. This model was improved based on the backbone model ShuffleNetV1. Firstly, HLNet had a more reasonable structure compared to ShuffleNetV1, which allowed HLNet to achieve a higher number of blocks and faster computation speed. Secondly, this study added dilated convolution layers with a larger perceptual field, which was less computationally intensive than traditional convolution layers of the same size and allowed the model to achieve higher recognition accuracy. Finally, the attention mechanism acting on the spatial and channel domains had been improved. The range of regions of interest and the number of feature extractions were increased, and the computational power consumption caused by the attention mechanism was reduced. The improved attention mechanism was added to HLNet, resulting in a significant improvement in the model’s ability to identify similar crop diseases. The structure of the HLNet model is shown in Figure 3.
2.3. ShuffleNetV1

The CNN model had achieved excellent results in various fields due to its excellent identification accuracy [28–30]. CNN had gradually developed from a high precision and slow calculation speeds to a high precision and fast calculation speeds [31]. ShuffleNetV1, proposed in 2018, had attracted great attention due to its low computational load, which reduced the computational complexity using depth-wise convolution (the number of groups in convolutional layers are the same as the input channels) and group pointwise convolution. Group pointwise convolution would induce inefficient interaction in the information between channels, resulting in weaker model generalization ability. In order to overcome the side effects caused by group pointwise convolution, ShuffleNetV1 proposed the channel shuffle operation to help information flowing between characteristic channels, the specific operation is shown in Figure 4, with the advantage of a substantial reduction in model parameters while maintaining high-precision performance. However, ShuffleNetV1 still has the following problems: (1) When the number of groups is 3 or higher, although the identification accuracy is the highest, the number of calculations is large. (2) Group pointwise convolution is frequently used in Block, and a large number of group pointwise convolutions with feature fusion function would cause computational redundancy.

![Figure 4. The process of channel shuffle operation.](image)

2.4. Improvements of ShuffleNetV1 Model

In view of the shortcomings of ShuffleNetV1, this study improved ShuffleNetV1. Firstly, two parallel convolutional layers, $3 \times 3$ and $5 \times 5$ size convolution kernels, were added to the first layer of the model. Such layers can provide multi-scale feature information for the model and improve the extraction effect of different size lesions. Secondly, this study optimized the block structure of ShuffleNetV1 and reduce the calculation number of each block. Thirdly, $5 \times 5$ dilated depth-wise convolutions replaced the $3 \times 3$ depth-wise convolution, which cannot increase the number of model calculations while improving the accuracy of model identification. Fourth, we improved the attention module, reduced the calculation number of the attention module, and improved the identification accuracy of the model for similar leaf diseases. Finally, the feature map was compressed to a size of $1 \times 1$ by the global average pooling layer, and the identified category was output after input to the fully connected layer of the model. The model structure of this paper is shown in Figure 5.

![Figure 5. Structure of the proposed ShuffleNetV1 model.](image)

2.4.1. Block Improvement

There were two block types in ShuffleNetV1, one is the block (A) for feature extraction from the feature map, the size of the output feature map was the same as the size of the input feature map, shown in Figure 6a. The other was the block (B) with a down-sampling function, as shown in Figure 6b.
This study firstly improved the Block (A) by deleting the first group pointwise convolution and putting the channel rearrangement module at the end and constructed the block in the order from depth-wise convolution (feature extraction)–group pointwise convolution (information circulation)–channel shuffle (circulation compensation). Compared with the original block, the block structure constructed in this way was more reasonable and the information circulation is more comprehensive. The improved structure is shown in Figure 7a. The study also improved the Block (B) by removing group pointwise convolutions and reordering the layers and replacing $3 \times 3$ average pooling layers in shortcut connection of matching sizes between different layers with $1 \times 1$ point convolution. The improved Block(B) is shown in Figure 7b. After the above optimization, the problems of excessive calculation number and calculation redundancy of the model are all resolved.

In addition to the above improvements, this paper replaced the ReLU activation function with ReLU6 in ShuffleNetV1. ReLU6 limits the maximum output of ReLU to less
than 6. This activation function would speed up the activation calculation and the model would be easy to transplant to small mobile devices.

2.4.2. Dilated Convolution

The key of improving model convergence speed was to speed up the aggregation of feature information. When the convolution kernel performs convolution operations, the size of the feature map would be reduced, which promotes information aggregation. With a larger size convolutional kernel, the size of the feature map was reduced more, and the speed of information aggregation was faster. Therefore, in order to increase the speed of model information aggregation and maintain the number of model calculations, we dilated the 3 × 3 depth-wise convolution (dilation is 2) for improving the accuracy of model identification and speeding up model training. Figure 8 shows the normal convolution and the dilated convolution with double expansion.

![Figure 8. Normal convolution and dilated convolution. (a) Normal convolution with a kernel size of 3 × 3; (b) Convolution with a kernel size of 3 and dilation of 2.](image)

The holes caused by the dilated depth-wise convolution kernel would lead to incoherent information of extracted features. When the size of the feature map was reduced to a certain extent, model identification accuracy might decrease when using dilated depth-wise convolutions. In order to analyze the influence of dilated depth-wise convolutions on the model, this paper designed three models with different numbers of dilated depth-wise convolutions, namely HLNet (A), HLNet (B), HLNet (C). HLNet (A) was the original model that had not been replaced, HLNet (B) was the model in which half of the depth-wise convolutions were replaced with dilated depth-wise convolutions, and HLNet (C) was the model that replaced all depth-wise convolutions with dilated depthwise convolutions. (HLNet: high-speed and lightweight network)

2.5. Attention Module Improvements

In actual production, many leaf diseases are very similar, dilated depth-wise convolutions alone cannot improve the accuracy of the model’s identification for similar diseases. Thus, in order to improve the accuracy of similar disease identification, this study added a channel attention module and a spatial attention module to the model. The attention module obtained the weight of the feature information through the neural network, so that the model allocated the computing power resources reasonably and learned the key features, in which the channel attention of SENet [32] and the spatial attention of CBAM [33] were excellent.

2.5.1. Channel Attention and Spatial Attention

The channel attention in SENet model reduces the feature map F_i of the i channel to a size of 1 × 1 through global pooling φ, and then passes through two fully connected
layers $\sigma$. After the first fully connected layer, ReLU activation is performed, and finally the feature maps input to the sigmoid activation function, and each group of parameters is compressed to between 0–1 to obtain the channel attention weight $CW_i$. The calculation formula of channel attention weight is displayed in Equation (1).

$$CW_i = \text{sigmoid}(\sigma(\text{ReLU}(\sigma(\phi(F_i))))) \quad (1)$$

After obtaining the channel attention weight, $CW_i$ multiplies the original feature map $F_i$ to obtain the feature map $CF_i$ with channel attention. The calculation formula of feature map with channel attention is displayed in Equation (2).

$$CF_i = CW_i \cdot F_i \quad (2)$$

The spatial attention in CBAM model compresses the same set of feature maps $F_i$ into one feature map using maximum pooling and average pooling, respectively, and fuses these two feature maps by concatenate operation. After that, a convolutional kernel of size 7 $\times$ 7 is used to convolve the group of feature maps so that the two feature maps are fused into one, and finally the feature maps are fed into the function to obtain the spatial attention weights $SW_i$. The calculation formula of spatial attention weight is displayed in Equation (3).

$$SW_i = \text{sigmoid}(\text{Conv} (\text{MaxPool}(F_i) + \text{AvgPool}(F_i))) \quad (3)$$

After obtaining the spatial attention weight, $SW_i$ multiplies the original feature map $F_i$ to obtain the feature map with spatial attention. The calculation formula of feature map with spatial attention is displayed in Equation (4).

$$SF_i = SW_i \cdot F_i \quad (4)$$

### 2.5.2. Model Based on Improved Attention

A large number of attention modules would increase the number of calculations sharply, and the calculation speed would drop sharply. In order to balance the identification accuracy and the calculation speed, we improved the two kinds of attentions.

For channel attention (CA), the fully connected layer consumes a significant number of arithmetic powers. However, the convolutional layer had fewer computational parameters and can replace the fully connected layer. Therefore, the two fully connected layers in the CA module were replaced with two pointwise convolutional (group 5 and 3) layers and point convolutional layers. After the replacement, the CA module had a deeper structure, and the number of calculations was also reduced, and the channel weight can be extracted more accurately and quickly.

For spatial attention (SA), since the feature map size was reduced to 7 $\times$ 7 at the end of the model, the leaf disease feature information was basically reduced to within one pixel, and the use of large-scale convolution kernels would cause computational redundancy. Therefore, the 7 $\times$ 7 convolution in SA was replaced with two 3 $\times$ 3 convolutions. Through this improvement, the SA module structure was deepened, the accuracy was improved, and the calculation parameters were reduced to 50% of the original.

In addition, we added two attention modules based on the HLNet (B) model with the highest identification accuracy in Section 2.4.2 (the comparison results of the identification accuracy experiment were included in Section 3.1). Since multiple attention modules can increase the number of calculations, this study only added attention modules to the last two stages of the model. The block arrangement method of adding attention module is shown in Figure 9.

For evaluating the effectiveness of the improved attention module, this paper constructed HLNet (BA) model using the unimproved attention module, and the HLNet (BB) model using the improved attention module. The model was obtained by stacking the above block 17 times, and the parameters of each layer of the model are shown in Table 1.
For channel attention (CA), the fully connected layer consumes a significant number of arithmetic powers. However, the convolutional layer had fewer computational parameters and can replace the fully connected layer. Therefore, the two fully connected layers in the CA module were replaced with two pointwise convolutional (group 5 and 3) layers and point convolutional layers. After the replacement, the CA module had a deeper structure, and the number of calculations was also reduced, and the channel weight can be extracted more accurately and quickly.

For spatial attention (SA), since the feature map size was reduced to 7 × 7 at the end of the model, the leaf disease feature information was basically reduced to within one pixel, and the use of large-scale convolution kernels would cause computational redundancy. Therefore, the 7 × 7 convolution in SA was replaced with two 3 × 3 convolutions. Through this improvement, the SA module structure was deepened, the accuracy was improved, and the calculation parameters were reduced to 50% of the original.

In addition, we added two attention modules based on the HLNet (B) model with the highest identification accuracy in Section 2.4.2 (the comparison results of the identification accuracy experiment were included in Section 3.1). Since multiple attention modules can increase the number of calculations, this study only added attention modules to the last two stages of the model. The block arrangement method of adding attention module is shown in Figure 9.

![Figure 9. Blocks added to spatial attention (SA) and channel attention (CA).](image)

### Table 1. Improved ShuffleNetV1 model layer parameters.

| Layer               | Input Size | Output Size | Repeat | Stride | Remarks |
|---------------------|------------|-------------|--------|--------|---------|
| Image               | 224 × 224 × 3 | 224 × 224 × 3 |        |        |         |
| Conv 1              | 224 × 224 × 3 | 112 × 112 × 12 | 1       | 2      |         |
| Conv 2              | 224 × 224 × 3 | 112 × 112 × 12 | 1       | 2      |         |
| Stage 0             | 112 × 112 × 24 | 56 × 56 × 48  | 1       | 2      |         |
| Stage 1             | 56 × 56 × 48  | 56 × 56 × 72  | 2       | 1      |         |
| Stage 2             | 56 × 56 × 72  | 28 × 28 × 144 | 1       | 2      |         |
| Stage 3             | 28 × 28 × 144 | 14 × 14 × 288 | 3       | 1      |         |
| Stage 4             | 14 × 14 × 288 | 7 × 7 × 576  | 3       | 1      | Join Att |
| Stage 5             | 7 × 7 × 576  | 7 × 7 × 1152 | 2       | 1      | Join Att |
| Adaptive/AvgPool    | 7 × 7 × 1152 | 1 × 1 × 1152 |        |        |         |
| FC                  | 1 × 1 × 1152 | 28          |        |        |         |

### 2.6. Experimental Environment Parameter Settings

In this experiment, the computer operating system based on windows 10, equipped with NVIDIA Titan X GPU and Intel Xeon E5-2696 v3 CPU, and adopted Python 3.6 programming language. The deep learning framework used is Pytorch. The tools used for experimental validation are integrated in Pytorch. In this paper, we use the Adam training optimizer [34] and a quadruple cross-entropy function to calculate the loss values.

Table 2 shows the hyperparameters used in model training. The training was divided into 40 epochs. The learning rate of the optimizer is $1 \times 10^{-3}$, the weight decay was 0.001, the learning rate was reduced by 1/10 every 10 epochs, batch size is 10, input image size was 224 × 224. The pointwise convolutions were grouped 3. A total of 28 categories in this study.

### Table 2. Hyperparameters used in model training.

| Name                  | Num         |
|-----------------------|-------------|
| Adam learning rate    | $1 \times 10^{-3}$ |
| Weight decay          | 0.001       |
| Epoch                 | 40          |
| Batch size            | 10          |
| Image size            | 224 × 224   |
| Group                 | 3           |
| Classes               | 28          |
3. Experimental Results and Analysis

This section analyzed the effectiveness of the improved model with the dilated depth-wise convolutions and the attention module. In Section 3.1, ShuffleNetV1 was compared with the initially improved model, and the impact of adding dilated depth-wise convolutions on the model was analyzed. In Section 3.2, the networks incorporating the improved attention module were compared and the effectiveness of the attention module was demonstrated by analyzing the confusion matrix. In Section 3.3, the model with the highest identification accuracy was compared with backbone CNN models. In Section 3.4, we compared our study with latest lightweight crop leaf disease identification studies.

3.1. Improved ShuffleNetV1 Experimental Results and Analysis

Through training ShuffleNetV1, HLNet (A), HLNet (B), HLNet (C), the evaluation indicators of each model are shown in Table 3. Evaluation indicators include:

1. Best Acc (%): The best accuracy of top-1 identification was obtained during the model testing. This represents the accuracy of model identification and is the most important evaluation metric. The formula for calculating the accuracy rate is displayed in Equation (5), where TP is true positive, FP is false positive, TN is true negative, FN is false negative.

   \[
   \text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)
   \]

2. FLOPs (M): Floating point operations. Computational memory consumed by each operation in convolution layers and fully connected layers during forward propagation of the model can measure the complexity of the model. The formula for calculating the FLOPs is displayed in Equation (6), where \(C_i\) is convolutional layers Input channel, \(K\) is convolution kernel size, \(H, W\) is height and weight of convolutional layers output feature map, \(C_O\) is output channel. \(I, O\) is input and output numbers in fully connected layers.

   \[
   \text{FLOPs} = \left[ \sum_{l_{conv}=1}^{n} (2C_iK^2 - 1)HWCO \right] + \left[ \sum_{l_{full}=1}^{n} (2I - 1)O \right] \quad (6)
   \]

3. Model Size (K): Memory requirements for saving models. It can measure model preservation and application costs.

4. Computation time (s): Computation time for each 224 × 224 size image computed by the model in the CPU. It measures the model’s computational efficiency.

| Name          | Best Acc (%) | FLOPs (M) | Model Size (K) | Computation Time (s) |
|---------------|--------------|-----------|----------------|----------------------|
| HLNet (A)     | 98.98        | 200.59    | 6375           | 0.166                |
| HLNet (B)     | 99.56        | 200.59    | 6375           | 0.166                |
| HLNet (C)     | 99.46        | 200.59    | 6375           | 0.166                |
| ShuffleNetV1  | 99.00        | 579.5     | 6891           | 0.236                |

The testing results in Table 3 indicated that after added multi-scale feature extraction and deleted redundant pointwise convolution, FLOPs (200.59 M vs. 579.5 M), model size (6375 K vs. 6891 K) and computation time (0.166 s vs. 0.236 s) of HLNet (A) were less than ShufflenetV1, which proved that most pointwise convolution was redundant, and removed pointwise convolution and then deepened network was an effective means to maintain model identification accuracy, reduce model computation and model calculation time.

Due to the addition of dilated depth-wise convolutions, HLNet (B), HLNet (C) maintained a small computational time and model size, the identification accuracy was greatly improved compared with HLNet (A). It proved that the dilated depth-wise convolution
can improve the identification accuracy without affecting the calculation time, which was an effective method to improve the efficiency of the model.

Figure 10 shows the HLNet (A), HLNet (B), HLNet (C), ShuffleNetV1 accuracy of the model in the test dataset at each epoch during the training process. As shown in the figure, HLNet (B) and HLNet (C) quickly achieved the highest accuracy due to the addition of dilated depth-wise convolutions. However, the model HLNet (C) with more dilated depth-wise convolutions kept oscillating in the follow-up training, and the best accuracy was not as high as HLNet (B). This proved that the holes formed by the expansion would lead to the learning bias in the extraction of small-sized feature maps, which reduced the accuracy to a certain extent.

Figure 10. Testing curves of HLNet (A), HLNet (B), HLNet (C), ShuffleNetV1 models.

To further observe and evaluate the impact of dilated depth-wise convolutions on the model, we used the CAM heat map to observe the regions learned by the model [35]. The color close to red in the heat map was the area where the model extracted more features, and the color close to blue in the heat map was the area where the model extracted less features. Figure 11 shows the CAM heat map of HLNet (A), HLNet (B) and HLNet (C) models for three leaf diseases. As shown in the figure, HLNet (B) (Figure 11b) with a proper number of dilated depth-wise convolutions had more effective feature information than HLNet (A) (Figure 11a) with no dilated convolution. In Figure 11c, HLNet (C), which replaced all depth-wise convolution with dilated depth-wise convolutions, had a better information extraction effect on potato bacterial spot, but it was worse for the other two leaf diseases. Therefore, although the accuracy of HLNet (C) had improved, it was not as accurate as HLNet (B).

Figure 11. CAM heat map of HLNet (A), HLNet (B) and HLNet (C) for Potato_bacterial_spot, Corn northern_leaf_blight, Grape_black_rot. (a) HLNet (A); (b) HLNet (B); (c) HLNet (C); (d) RGB Image.
In summary, by improving ShuffleNetV1 and adding an appropriate number of dilated depth-wise convolutions to build HLNet (B), the best accuracy was improved from 98.98% to 99.56%, the model floating point operations were reduced from 579.5 M to 200.59 M, and the model computation time was reduced from 0.236 s to 0.166 s. Considering several factors together, the HLNet (B) model had higher identification accuracy, less model floating-point operations and computation time. Therefore, HLNet (B) was used as a model for further improvement.

3.2. Attention Module Experimental Results and Analysis

To further improve the ability of the model to identify similar leaf diseases, we improved the attention module and put it on the HLNet (B) model in Section 3.1. For analyzing the effectiveness of the attention module, based on HLNet (B), we defined HLNet (BA) with the unimproved attention module and HLNet (BB) models with the improved attention module. We performed the following comparison experiments. After training the HLNet (B), HLNet (BA), and HLNet (BB) models, the evaluation metrics of these models were shown in Table 4. Compared to HLNet (B) without the attention module, HLNet (BA) with the unimproved attention module showed a slight increase in identification accuracy (99.58% vs. 99.56%), but a significant increase in model computation time (0.166 s vs. 0.208 s), floating point operations (265.15 M vs. 200.59 M), and model size (9096 K vs. 6375 K). The best accuracy of HLNet (BB) with the optimized attention module was improved (99.86% vs. 99.58%). Compared to HLNet (BA), HLNet (BB) greatly reduce computation time (0.268 s vs. 0.173 s), model size (8239 K vs. 9096 K), floating point operations (248.07 M vs. 264.15 M).

Table 4. Results of HLNet (B), HLNet (BA), HLNet (BB) models.

| Name       | Best Acc (%) | FLOPs (M) | Model Size (K) | Computation Time (s) |
|------------|--------------|-----------|----------------|----------------------|
| HLNet (B)  | 99.56        | 200.59    | 6375           | 0.166                |
| HLNet (BA) | 99.58        | 264.15    | 9096           | 0.208                |
| HLNet (BB) | 99.86        | 248.07    | 8239           | 0.173                |

Figure 12 shows the training curves of the three models. As shown in the figure, although HLNet (BA) achieved a faster fit than HLNet (B), the identification accuracy was not improved. The optimized HLNet (BB) obtained a smoother training curve and achieved a faster fit.

Figure 12. Testing curves of HLNet (B), HLNet (BA), HLNet (BB) models.
Similarly, we also used the CAM heat map to observe the regions where the model extracted features. As shown in Figure 13, it was obvious that the HLNet (BA) (Figure 13b) and HLNet (BB) (Figure 13c) models with the attention module learned more leaf features in maize and grape disease identification. While for potatoes and grapes, HLNet (BA) with the unimproved attention module also learned many wrong features. By improving the attention module, the interesting region of HLNet (BB) was more precisely focused on the leaf. Therefore, the improvement of the attention module enabled the model to learn the correct feature information quickly and accurately, which increased the accuracy of the model in the classification of similar diseases.

![CAM heat map of HLNet (B), HLNet (BA) and HLNet (BB) for Potato_bacterial_spot, Corn northern_leaf_blight, Grape_black_rot. (a) HLNet (B); (b) HLNet (BA); (c) HLNet (BB); (d) RGB Image.](image)

The purpose of adding the attention module in this paper was to improve the accuracy of the model in classifying similar crop leaf diseases, so the accuracy of single class disease identification was very important. Therefore, this paper adopted the confusion matrix to show and analyze the number of model misidentifications intuitively. In order to observe the improvement effect of each step of the model, we choose HLNet (A), HLNet (B), HLNet (BA) and HLNet (BB) to generate the confusion matrix. Figure 14 were the confusion matrixes of HLNet (A), HLNet (B), HLNet (BA), HLNet (BB). The confusion matrix produced in this paper select 500 pieces from each category of the data for identification. In the confusion matrix, each column represented the predicted category and each row represented the true attribution category of the data, and the shade of the color in each square represented the number of recognized categories. The closer the color of the square was to red, the more often it was incorrectly identified. As shown in the confusion matrix in Figure 14a, HLNet (A) without the dilated depth-wise convolutions and attention module have the highest number of misidentifications. There were more errors in identification among for different diseases of the same crops such as apple, corn, and rice. Also, HLNet (A) had big identification errors in different but similar diseases of different crops, such as apples and potatoes. HLNet (A) without the dilated depth-wise convolutions and attention module had the highest number of misidentifications. In Figure 14b, we can see that the HLNet (B) model had improved the accuracy of identifying similar diseases in different crops, but it was less effective in identifying diseases in the same crop. In Figure 14c, we can see the misidentification of the HLNet (BA) model was not improved obviously for adding the attention module, which demonstrated the model’s ability for identifying similar diseases did not improve. In Figure 14d, the HLNet (BB) model with the optimized attention module
achieved a significant improvement in similar diseases identification. Therefore, it can be concluded that: (1) The addition of large-sized convolution kernels, such as dilated depth-wise convolutions, can only marginally improved the accuracy of single-class identification. (2) Deepening the attention module can help the model effectively distinguishing similar diseases and improving the accuracy of model identification.

![Confusion Matrix](image)

Figure 14. Confusion matrix for the HLNet (A), HLNet (B), HLNet (BA), HLNet (BB) models. (a) HLNet (A); (b) HLNet (B); (c) HLNet (BA); (d) HLNet (BB).

3.3. Comparison of Various Backbone Models

From the analysis above, HLNet (BB) is the model with the highest identification accuracy. To demonstrate that HLNet (BB) model can effectively balance arithmetic power and identification accuracy, we compared this model with backbone CNN models. Backbone CNN models compared in this paper are VGG16 [36], ResNet101 [37], DenseNet161 [38], InceptionV3 [39], AlexNet [40], MobileNetV1, MobileNetV2 [41], MobileNetV3 [42], Shuf-
fleNetV1, ShuffleNetV2 [43], SqueezeNet Xception models. Since the training epoch in this paper was 40, some models with too many parameters cannot be trained to the best identification effect in this case. To demonstrate the advantage of the HLNet (BB) model in identification accuracy, this paper introduced transfer learning to initialize the models’ parameters for models with floating point operations over 1G (VGG16, ResNet101, DenseNet161, InceptionV3, Xception), which ensured that the models achieved the highest identification accuracy after 40 epochs of training. The ImageNet [44] dataset is used for transfer learning.

For training HLNet (BB) and above 11 models, same hyperparameters were used. The optimizer and loss function are shown in Section 2.5. The evaluation metrics of each model are shown in Table 5. It clearly showed that the HLNet (BB) model had the highest accuracy. Although the VGG16, ResNet101, DenseNet161, InceptionV3, Xception models used transfer learning, the accuracy of these models was lower than that of HLNet (BB) which did not use transfer learning. HLNet (BB) had a smaller model size than other models except ShuffleNetV2 (238.4 M vs. 198.73 M), however, the computation time of HLNet (BB) was shorter than ShuffleNetV2. In practical applications, computation time was more important than floating point operations. The model size of HLNet (BB) was smaller than other models except for MobileNetv2, ShuffleNetV1 and SqueezeNet models (8235 K vs. 7353 K vs. 6891 K vs. 4850 K). While the floating-point operations of the three models were much bigger than HLNet (BB), the computation time of the HLNet (BB) model was shorter than the other models except the SqueezeNet model. However, SqueezeNet had a low accuracy; less than 50%. Thus, the HLNet (BB) model outperformed the other models in comprehensive consideration of accuracy, FLOPs, model size and calculation time.

Table 5. Results of Backbone CNN model models and HLNet (BB).

| Name            | Best Acc (%) | FLOPs (M) | Model Size (K) | Computation Time (s) |
|-----------------|--------------|-----------|----------------|----------------------|
| HLNet (BB)      | 99.86        | 238.44    | 8235           | 0.173                |
| AlexNet         | 92.40        | 715.54    | 238,690        | 0.182                |
| VGG16           | 86.62        | 15.5 × 1024 | 540,463       | 0.255                |
| ResNet101       | 96.92        | 7.84 × 1024 | 100,100       | 0.389                |
| DenseNet161     | 96.89        | 7.82 × 1024 | 113,019       | 0.562                |
| MobileNet       | 98.70        | 581.7     | 8261           | 0.187                |
| MobileNetV2     | 98.90        | 318.99    | 7353           | 0.210                |
| MobileNetV3     | 98.96        | 265.15    | 10,833         | 0.312                |
| ShuffleNetV1    | 98.96        | 579.5     | 6891           | 0.236                |
| ShuffleNetV2    | 98.96        | 198.73    | 9003           | 0.198                |
| SqueezeNet      | 41.92        | 355.69    | 4850           | 0.149                |
| InceptionV3     | 98.91        | 2.85 × 1024 | 95,748        | 0.293                |
| Xception        | 98.89        | 4.58 × 1024 | 81,808        | 0.306                |

For direct observations, we compared the HLNet (BB) model to the models with transfer learning (VGG16, ResNet101, DenseNet161, InceptionV3, Xception) using training curves, as depicted in Figure 15. We also used the same way to compare the HLNet (BB) model with the models that used transfer learning (AlexNet, MobileNet, MobileNetV1, ShuffleNetV1, ShuffleNetV2, SqueezeNet), as depicted in Figure 15. It was necessary to note that the HLNet (BB) model without transfer learning was used in two comparisons.

Figure 15 shows that all large models reached fit at 20 epochs and that the training curve for HLNet (BB) was as smooth as most models using transfer learning. Even without using transfer learning, lightweight and addition of attention modules promoted the HLNet (BB) to achieve the fit at a very fast rate.

Figure 16 shows the training curves for the HLNet (BB) model and the lightweight models. In Figure 16, the accuracy of HLNet (BB) proposed in this paper was the highest, and the training curve of HLNet (BB) was more stable than other models after the first
epoch. The comparison results demonstrate that the weak pooling effect of large size convolutional kernels accelerate the model learning speed, and the improved attention module allowed the model to learn more effective features fast.

![Figure 15. Large model testing curve.](image)

In summary, HLNet (BB) had high classification accuracy (99.86%), lightweight calculation amount (238.44 M) and memory requirements (8235 K), and required less calculation time (0.173 s). It successfully reduced the computational effort of the attention module while enhancing the identification of similar diseases. HLNet (BB) had better identification results and a lighter model size compared to large models, and faster fitting and identification compared to lighter models, which indicated that our improvement of the model was successful, and this model surpassed most of the common backbone CNN models.

### 3.4. Comparison of Latest Lightweight Models

To further demonstrate the advanced performance of HLNet, this study also compared HLNet with various latest lightweight leaf disease identification models. The results of each model are shown in Table 6, and the results of these models are taken from the original papers.

The HLNet (BB) model with a novel lightweight structure, dilated convolution and improved attention mechanism achieved the highest accuracy of 99.86% compared with studies that identified variety of crops. Although the study by Wagle, S. A et al. [21] identified nine crops, more than our study, their model size was much larger than our model’s, and the accuracy was lower than our model’s.
Table 6. Comparison of the accuracy of our proposed models with that of the latest published crop leaf diseases identification models.

| Name                        | Crop Number | Years | Model                  | Best ACC (%) | Flops (M) | Model Size (K) | Computation Time (s) |
|-----------------------------|-------------|-------|------------------------|--------------|-----------|---------------|----------------------|
| Our study                   | 6           | 2022  | HLNet (B)              | 99.56        | 200.59    | 6375          | 0.166                |
|                             |             |       | HLNet (BB)             | 99.86        | 238.44    | 8235          | 0.173                |
| Wagle, S. A. et al. [21]    | 9           | 2021  | N1 model               | 99.45        | -         | 15,155        | -                    |
|                             |             |       | N2 model               | 99.65        | -         | 30,412        | -                    |
|                             |             |       | N3 model               | 99.55        | -         | 15,155        | -                    |
| Wang, P. et al. [22]        | 1           | 2021  | ECA-SNett_0.5×         | 98.86        | 37.4      | -             | -                    |
|                             |             |       | ECA-SNett_1.0×         | 99.66        | 125.6     | -             | -                    |
| Bhujel, A. et al. [24]      | 1           | 2022  | lw_resnet20            | 98.85        | 499.5     | 16,998        | 0.795                |
|                             |             |       | lw_resnet20_cham       | 99.69        | 450.56    | 17,203        | 0.914                |
|                             |             |       | lw_resnet20_se         | 98.85        | 450.56    | 17,203        | 0.927                |
|                             |             |       | lw_resnet20_sa         | 99.32        | 566.27    | 18,739        | 0.961                |
|                             |             |       | lw_resnet20_da         | 98.90        | 601.09    | 18,022        | 0.984                |
| Chao, X. et al. [25]        | 1           | 2021  | SE_Xception            | 99.40        | 48.15     | -             | -                    |
|                             |             |       | SE_miniXception        | 97.01        | 6.67      | -             | -                    |

The HLNet (BB) model still achieved the highest recognition accuracy compared with 3 studies that identified one crop. Wang, P et al. [22] obtained 37.4 M, 125.6 M floating point operations and 98.86%, 99.66% accuracy. Chao, X. F et al. [25] obtained 48.15 M, 6.67 M floating point operations and 99.40%, 97.01% accuracy. Compared with these models, HLNet (BB) had more crop identification types and higher accuracy, although the floating-point operations was slightly higher, which can solve the problem of practical application. Compared with the model proposed by Bhujel, A et al. [24], our model was lighter and had higher accuracy.

In summary, our models achieved a higher accuracy compared with that of the crop leaf disease identification models in the last two years. Compared with the models that identified only one crop, the universality of our model was stronger. Compared with the models that identified a variety of crops, our model was lighter. In addition, our study had abundant and objective evaluation indicators of light-weighting, which further proved that our model not only required less parameters and memory, but also had faster processing speed.

4. Conclusions

Fast and accurate identification of crop leaf diseases is the advantage of automatic disease identification. Compared with traditional CNN models, lightweight CNN models had faster identification speed and lower memory requirements, which were more suitable for practical applications. In this paper, we developed a deep learning model HLNet (BB) specifically for fast and efficient identification of crop leaf disease images based on lightweight convolutional neural network. Firstly, we improved the ShuffleNetV1 structure by deleting some of the pointwise convolutions and rearranging the positions of the layers in the block. By doing so, it successfully maintained the computational accuracy of the model and reduced the number of floating-point operations by 65% and the computation time by 30%. Secondly, we replaced the depth-wise convolution in the model with the dilated depth-wise convolution, which improved the model identification accuracy without affecting the computational speed and memory requirements (99.56% vs. 98.98%). Finally, we improved the existing attention module to reduce the number of floating-point operations and enhanced the ability to correctly extract interesting regions. The addition of the improved attention module resulted in a slight 4% improvement in model computation speed but a significant increase in the ability to identify similar diseases.

We constructed the dataset based on several publicly available datasets and self-built dataset with a total of 20,490 images, including 28 types of leaf and leaf disease images of 6 crops. Based on the dataset, we performed a series of validation experiments, and the
results showed that the HLNet (BB) model had high accuracy, low computational effort and short computational consumption time compared with many backbone CNN models. Meanwhile, compared with the latest crop disease identification study, our study evaluated the lightness of the model from a wider range of perspectives. This study is helpful for providing a theoretical basis and technical support for the practical application of automatic identification of crop diseases.

Author Contributions: Conceptualization, Y.X. and S.K.; Methodology, Y.X. and Q.C.; Validation, S.K. and C.L.; Writing—Original Draft Preparation, Z.G.; Writing—Review & Editing, Q.C. and C.L.; Supervision, Y.J. and Y.X.; Funding Acquisition, Y.X. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China [Grant number 31801753]; the JiLin provincial science and technology department international exchange and cooperation project [Grant number 20200801014GH]; the Jilin Province Science and Technology Development Plan Project [Grant number 20210101157JC]; the Key scientific and technological research projects of Changchun Science and Technology Bureau [Grant number 21ZGN28].

Data Availability Statement: Publicly available datasets were analyzed in this study. This data can be found here: https://archive.ics.uci.edu/ml/datasets/Rice+Leaf+Diseases.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. Plant leaf datasets.

| Classes Num | Variety | Disease               | Image Num |
|-------------|---------|-----------------------|-----------|
| 1           | Apple   | Apple_scab            | 2016      |
| 2           |         | Black_rot             | 1987      |
| 3           |         | Cedar_apple_rust      | 1760      |
| 4           |         | Healthy               | 2008      |
| 5           | Corn    | Gray_leaf_spot        | 1642      |
| 6           |         | Common_rust           | 1907      |
| 7           |         | Northern_leaf_Blight  | 1907      |
| 8           |         | Healthy               | 1859      |
| 9           | Potato  | Early_blight          | 1939      |
| 10          |         | Late_blight           | 1939      |
| 11          |         | Healthy               | 1824      |
| 12          | Grape   | Black_rot             | 1888      |
| 13          |         | Black_Measles         | 1920      |
| 14          |         | Leaf_blight           | 1722      |
| 15          |         | Healthy               | 1692      |
| 16          | Tomato  | Bacterial_spot        | 1702      |
| 17          |         | Early_blight          | 1920      |
| 18          |         | Late_blight           | 1851      |
| 19          |         | Leaf_mold             | 1882      |
| 20          |         | Septoria_leaf_spot    | 1745      |
| 21          |         | Spider_mites          | 1741      |
| 22          |         | Target_Spot           | 1827      |
| 23          |         | Tomato_mosaic_virus   | 1790      |
| 24          |         | Yellow_Leaf_Curl_Virus| 1961      |
| 25          |         | Healthy               | 1920      |
| 26          | Rice    | Bacterial_leaf_blight | 1438      |
| 27          |         | Brown spot            | 1072      |
| 28          |         | Leaf smut             | 1016      |
References

1. Tilman, D.; Balzer, C.; Hill, J.; Befort, B.L. Global food demand and the sustainable intensification of agriculture. *Proc. Natl. Acad. Sci. USA* 2011, 108, 20260–20264. [CrossRef]

2. Pantazi, X.E.; Moshou, D.; Tamouridou, A.A. Automated leaf disease detection in different crop species through image features analysis and One Class Classifiers. *Comput. Electron. Agric.* 2019, 156, 96–104. [CrossRef]

3. Schwarzenbach, R.P.; Egli, T.; Hofstetter, T.B.; Von Gunten, U.; Wehrli, B. Global water pollution and human health. *Annu. Rev. Environ. Resour.* 2010, 35, 109–136. [CrossRef]

4. Bock, C.H.; Poole, G.H.; Parker, P.E.; Gottwald, T.R. Plant disease severity estimated visually, by digital photography and image analysis, and by hyperspectral imaging. *Crit. Rev. Plant Sci.* 2010, 29, 59–107. [CrossRef]

5. Ghosal, S.; Blystone, D.; Singh, A.K.; GanapathySubramanian, B.; Singh, A.; Sarkar, S. An explainable deep machine vision framework for plant stress phenotyping. *Proc. Natl. Acad. Sci. USA* 2018, 115, 4613–4618. [CrossRef]

6. Kumar, S.D.; Esakkirajan, S.; Bama, S.; Keerthiveena, B. A microcontroller based machine vision approach for tomato grading and sorting using SVM classifier. *Microprocess. Microsyst.* 2020, 76, 103090. [CrossRef]

7. Goel, N.; Sehgal, P. Fuzzy classification of pre-harvest tomatoes for ripeness estimation—An approach based on automatic rule learning using decision tree. *Appl. Soft Comput.* 2015, 36, 45–56. [CrossRef]

8. Ai, L.; Fang, N.F.; Zhang, B.; Shi, Z.H. Broad area mapping of monthly soil erosion risk using fuzzy decision tree approach: Integration of multi-source data within GIS. *Int. J. Geogr. Inf. Sci.* 2013, 27, 1251–1267. [CrossRef]

9. Marconi, T.G.; Oh, S.; Ashapure, A.; Chang, A.; Jung, J.; Landivar, J.; Enciso, J. Application of unmanned aerial system for management of tomato cropping system. In *Autonomous Air and Ground Sensing Systems for Agricultural Optimization and Phenotyping IV*; International Society for Optics and Photonics: Bellingham, WA, USA, 2019; p. 11008. [CrossRef]

10. Mohanty, S.P.; Hughes, D.P.; Salathé, M. Using deep learning for image-based plant disease detection. *Front. Plant Sci.* 2016, 7, 1419. [CrossRef]

11. Zhang, X.; Qiao, Y.; Meng, F.; Fan, C.; Zhang, M. Identification of maize leaf diseases using improved deep convolutional neural networks. *IEEE Access* 2018, 6, 30370–30377. [CrossRef]

12. Zhang, Y.; Wa, S.; Liu, Y.; Zhou, X.; Sun, P.; Ma, Q. High-Accuracy Detection of Maize Leaf Diseases CNN Based on Multi-Pathway Activation Function Module. *Remote Sens.* 2021, 13, 4218. [CrossRef]

13. Liu, B.; Zhang, Y.; He, D.; Li, Y. Identification of apple leaf diseases based on deep convolutional neural networks. *Symmetry* 2018, 10, 11. [CrossRef]

14. Deng, R.; Tao, M.; Xing, H.; Yang, X.; Liu, C.; Liao, K.; Qi, L. Automatic diagnosis of rice diseases using deep learning. *Front. Plant Sci.* 2012, 11, 1691. [CrossRef]

15. Oppenheim, D.; Shani, G.; Erlich, O.; Tsror, L. Using deep learning for image-based potato tuber disease detection. *Phytopathology* 2019, 109, 1083–1087. [CrossRef]

16. Iandola, F.N.; Han, S.; Moskewicz, M.W.; Ashraf, K.; Dally, W.J.; Keutzer, K. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5 MB model size. arXiv 2016, arXiv:1602.07360.

17. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Adam, H. Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv 2017, arXiv:1704.04861.

18. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Detection (CVPR)*; Honolulu, UT, USA, 22–25 July 2017; pp. 1251–1258.

19. Zhang, X.; Zhou, X.; Lin, M.; Sun, J. Shufflenet: An extremely efficient convolutional neural network for mobile devices. In *Proceedings of the CVPR*, Salt Lake City, UT, USA, 18–22 June 2018; pp. 5848–6856.

20. Kamal, K.C.; Yin, Z.; Wu, M.; Wu, Z.L. Depthwise separable convolution architectures for plant disease classification. *Comput. Electron. Agric.* 2019, 165, 104948. [CrossRef]

21. Wagle, S.A.; Harikrishnan, R.; Ali, S.H.M.; Faseehuddin, M. Classification of plant leaves using new compact convolutional neural network models. *Plants* 2022, 11, 24. [CrossRef]

22. Wang, P.; Niou, T.; Mao, Y.R.; Liu, B.; Yang, S.Q.; He, D.J.; Gao, Q. Fine-Grained Grape Leaf Diseases Recognition Method Based on Improved Lightweight Attention Network. *Front. Plant Sci.* 2021, 12, 738042. [CrossRef]

23. Bi, C.K.; Wang, J.M.; Duan, Y.L.; Fu, B.F.; Kang, J.R.; Shi, Y. MobileNet Based Apple Leaf Diseases Identification. *Mob. Netw. Appl.* 2022, 27, 172–180. [CrossRef]

24. Bhujel, A.; Kim, N.E.; Arulmozhi, E.; Basak, J.K.; Kim, H.T. A lightweight Attention-based convolutional neural networks for tomato leaf disease classification. *Agriculture* 2022, 12, 228. [CrossRef]

25. Chao, X.F.; Hu, X.; Feng, J.Z.; Zhang, Z.; Wang, M.L.; He, D.J. Construction of apple leaf diseases identification networks based on xception fused by SE module. *Appl. Sci.* 2021, 11, 14. [CrossRef]

26. Hughes, D.; Salathé, M. An open access repository of images on plant health to enable the development of mobile disease diagnostics. arXiv 2015, arXiv:1511.08060.

27. UC Irvine Machine Learning Repository. Available online: https://archive.ics.uci.edu/ml/datasets/Rice+Leaf+Diseases (accessed on 14 April 2019).

28. Peng, D.; Zhang, Y.; Guan, H. End-to-end change detection for high resolution satellite images using improved UNet++. *Remote Sens.* 2019, 11, 1382. [CrossRef]
29. Xu, G.; Liu, M.; Jiang, Z.; Söffker, D.; Shen, W. Bearing fault diagnosis method based on deep convolutional neural network and random forest ensemble learning. Sensors 2019, 19, 1088. [CrossRef] [PubMed]

30. Shujaat, M.; Wahab, A.; Tayara, H.; Chong, K.T. pcPromoter-CNN: A CNN-Based Prediction and Classification of Promoters. Genes 2020, 11, 1529. [CrossRef] [PubMed]

31. Tian, C.; Zhuge, R.; Wu, Z.; Xu, Y.; Zuo, W.; Chen, C.; Lin, C.W. Lightweight image super-resolution with enhanced CNN. Knowl. Based Syst. 2020, 205, 106235. [CrossRef]

32. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the CVPR, Salt Lake City, UT, USA, 18–22 June 2018; pp. 7132–7141.

33. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.

34. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. arXiv 2014, arXiv:1412.6980.

35. Zhou, B.; Khosla, A.; Lapedriza, A.; Oliva, A.; Torralba, A. Learning deep features for discriminative localization. In Proceedings of the CVPR, Las Vegas, NV, USA, 27–30 June 2016; pp. 2921–2929.

36. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image detection. arXiv 2014, arXiv:1409.1556.

37. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image detection. In Proceedings of the CVPR, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–777.

38. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the CVPR, Honolulu, HI, USA, 22–25 July 2017; pp. 4700–4708.

39. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the inception architecture for computer vision. In Proceedings of the CVPR, Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826.

40. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. Adv. Neural Inf. Processing Syst. 2012, 25, 1097–1105. [CrossRef]

41. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.C. Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the CVPR, Salt Lake City, UT, USA, 18–22 June 2018; pp. 4510–4520.

42. Howard, A.; Sandler, M.; Chu, G.; Chen, L.-C.; Chen, B.; Tan, M.; Wang, W.; Zhu, Y.; Pang, R.; Vasudevan, V.; et al. Searching for mobilenetv3. In Proceedings of the ICCV, Seoul, Korea, 27 October–2 November 2019; pp. 1314–1324.

43. Ma, N.; Zhang, X.; Zheng, H.T.; Sun, J. Shufflenet v2: Practical guidelines for efficient cnn architecture design. In Proceedings of the ECCV, Munich, Germany, 8–14 September 2018; pp. 116–131.

44. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Fei-Fei, L. Imagenet: A large-scale hierarchical image database. In Proceedings of the CVPR, Miami Beach, FL, USA, 20–24 June 2009; pp. 248–255.