Article

Classification of Fine-Grained Crop Disease by Dilated Convolution and Improved Channel Attention Module

Xiang Zhang [1,2], Huiyi Gao [1,3,*] and Li Wan [1,3]

1   Institute of Intelligent Machines, Hefei Institutes of Physical Science, CAS, Hefei 230031, China
2   Science Island Branch, Graduate School of USTC, Hefei 230026, China
3   Lu'an Branch, Anhui Institute of Innovation for Industrial Technology, Lu'an 237100, China
*   Correspondence: hygao@iim.ac.cn

Abstract: Crop disease seriously affects food security and causes huge economic losses. In recent years, the technology of computer vision based on convolutional neural networks (CNNs) has been widely used to classify crop disease. However, the classification of fine-grained crop disease is still a challenging task due to the difficult identification of representative disease characteristics. We consider that the key to fine-grained crop disease identification lies in expanding the effective receptive field of the network and filtering key features. In this paper, a novel module (DC-DPCA) for fine-grained crop disease classification was proposed. DC-DPCA consists of two main components: (1) dilated convolution block, and (2) dual-pooling channel attention module. Specifically, the dilated convolution block is designed to expand the effective receptive field of the network, allowing the network to acquire information from a larger range of images, and to provide effective information input to the dual-pooling channel attention module. The dual-pooling channel attention module can filter out discriminative features more effectively by combining two pooling operations and constructing correlations between global and local information. The experimental results show that compared with the original networks (85.38%, 83.22%, 83.85%, 84.60%), ResNet50, VGG16, MobileNetV2, and InceptionV3 embedded with the DC-DPCA module obtained higher accuracy (87.14%, 86.26%, 86.24%, and 86.77%). We also provide three visualization methods to fully validate the rationality and effectiveness of the proposed method in this paper. These findings are crucial by effectively improving classification ability of fine-grained crop disease by CNNs. Moreover, the DC-DPCA module can be easily embedded into a variety of network structures with minimal time cost and memory cost, which contributes to the realization of smart agriculture.

Keywords: fine-grained crop disease; convolutional neural networks; attention mechanism; classification

1. Introduction

Crop disease is one of the most serious problems affecting the quality and yield of agricultural production worldwide [1]. Manual disease control suffers from a lack of expertise, poor objectivity, visual fatigue and low efficiency [2]. Additionally, the likelihood of diseases developing and rapidly spreading has increased due to the current state of rising global temperatures [3]. With the development of big data and machine learning, agriculture has shifted from the mechanical stage to smart agriculture, and most of the younger generation has a positive attitude towards using smart agriculture [4]. Crop disease identification based on machine learning and other technologies is a part of smart agriculture. It can meet the growing demand for food by reducing agricultural losses through data modeling [5]. The crop losses could be aggravated due to wrong or excessive control, as well as pesticide residues can pose serious damage to human health and the ecological environment. Therefore, it is very necessary to design an automatic and precise control technology for fine-grained crop disease.

Since the 1990s, researchers have started to implement the classification of crop disease using traditional image processing methods, including image of pre-processing, disease part segmentation, feature selection, and classification [6]. Guan et al. [7] used Bayesian discrimination to identify three common diseases of rice after preprocessing and segmenting the images, and the highest recognition accuracy reached 97.2%; Jiang et al. [8] also reported a high accuracy of 95.91% using multiple features of plant leaf images and support vector machine (SVM). Correlation analysis and the random forest algorithm were coupled by Huang et al. to increase the precision of wheat stripe rust detection in the early and middle phases [9]. However, the traditional method of image classification is only applicable to simple disease classification tasks because it is difficult to manually design and extract effective features from a complex task [10]. To solve this problem, Hinton et al. [11] proposed deep belief networks (DBN). Since then, deep learning techniques have developed rapidly due to their powerful automatic feature extraction capability. Thus, deep learning techniques have gradually become a research hotspot in the field of crop disease classification [12]. Sladojevic et al. [13] used deep CNNs to classify 13 different types of plant diseases with a high average accuracy of 96.3%. Lu et al. [14] identified 10 common rice diseases by training a deep CNN model with an accuracy of 95.48%. Bhatt et al. [15] combined CNNs and decision tree classifiers to recognize four different types of corn leaf images with an accuracy of 98%, which was 8% higher than using CNNs only. However, since the differences between fine-grained images are relatively subtle, it is difficult for CNNs to find subtle features that fully represent the object [16]. In recent years, attention mechanism has been proposed to inject new inspiration for computer vision tasks, which mimic the human visual system and can automatically enhance positive information in images without additional component labeling information. Therefore, the attention mechanism is embedded in various CNNs to improve the classification ability of fine-grained crop disease. Gao et al. [17] proposed a crop disease identification method based on dual-channel effective attention. Chen et al. [18] improved the identification ability of crop disease networks by combining the spatial attention module with the efficient channel attention (ECA) module. Wang et al. [19] solved the problem of serial interference of two kinds of attention by connecting channel attention and spatial attention in parallel. However, these methods do not consider the characteristics of crop diseases themselves, and thus these methods perform unstably in the task of fine-grained classification of crop diseases. Therefore, fine-grained crop disease classification is still a challenging and realistic task.

In CNNs, receptive field is a very important concept, which is defined as the size of the region where the pixels on the feature map are mapped on the input image. The size of the receptive field represents the size of the range of input images that the network can see. The pixel information outside the receptive field is invisible to the network. CNNs expand the receptive field by stacking convolution operations. However, the effective receptive field is only a small fraction of the theoretical receptive field due to the negligible gradient of most pixels applied to the receptive field [20]. The receptive field can describe the maximum amount of information of feature points, and the effective receptive field can describe the effectiveness of information. Just like the human visual system, our eyes can see a large area, but changes in areas outside the center of vision do not attract much attention. Ding et al. [21] suggested that the traditional CNNs were generally inferior to Transformer in downstream tasks due to the small effective receptive field.

In order to improve the recognition ability of CNNs for fine-grained crop diseases, we focused on the effective receptive field of CNNs and the interaction between global and local information, and there are few reports on this aspect. We consider that one of the key factors affecting the ability of CNNs for fine-grained crop disease recognition is that the effective receptive field of CNNs is too small. This leads to the inability of CNNs to utilize the overall information of the image, and the modeling process of the network relies only on the local information of the image. Local information often fails to reflect the differences between fine-grained crop diseases. In addition, since global and local information are not

separated and closely related, we believe that learning the correlation between the two is important to improve the fine-grained crop disease identification capability of CNNs.

Based on the above analysis, in this paper, we propose a novel approach combining dilated convolution block and dual-pooling channel attention (DC-DPCA) to realize the expansion of the effective receptive field of CNNs and the interaction of global and local information. Extensive experiments and model visualization results validate the effectiveness of the DC-DPCA module. These findings can improve the classification ability of fine-grained disease by CNNs. Moreover, the DC-DPCA module has low computational and storage costs and can be used in agricultural terminals to help achieve smart agriculture.

2. Materials and Methods

2.1. Data Set Acquisition and Analysis

The data used in this paper is a partial crop disease dataset from the 2018 AI-Challenger competition. The dataset contains 27 diseases for 10 crops, and 10 healthy crop categories, for a total of 59 categories. Most of the diseases were subdivided into general and severe categories based on their degree of incidence. The dataset contains a total of 36,000 images, among which the training set and the test set account for 87.4% and 12.6%, respectively. This is a typical fine-grained classification dataset of crop diseases, where most of the classification errors are mainly from misclassification of disease severity and similar diseases of the same crop. In addition, the sample distribution of this dataset is extremely uneven, which may cause the network model to tend to fit categories with more data. The distribution of the sample images in the training set is shown in Figure 1.
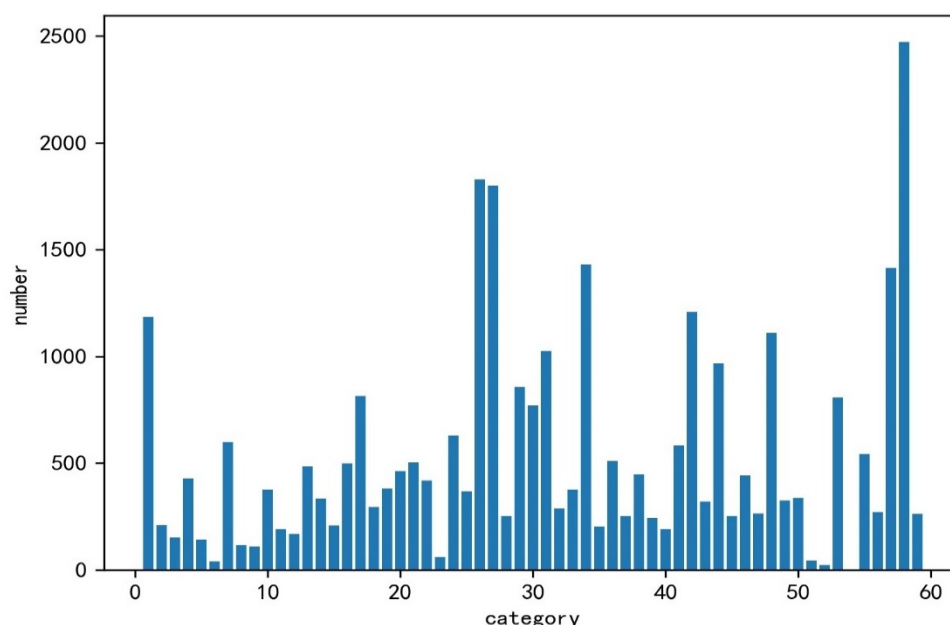
Figure 1. Distribution of sample images in the training set.

2.2. Loss Function for Uneven Sample Distribution

To solve the problem of uneven sample distribution, we design a cross-entropy loss function with a weighting factor (L-balance), which can be expressed as:

$$\mathrm{L-balance} = -\sum_{i=0}^{N} \left(1 - \frac{M_i}{M}\right)^{\beta} * P(X, i) * \log(Q(X, i)) \tag{1}$$

where X is the input data, N represents the total number of categories, M and M_i denote the number of samples in the training set and the number of samples of class i in the training set, respectively. The hyperparameter β smoothly adjusts the rate at which categories with larger sample sizes are downgraded. $P(X, i)$ represents the probability that X belongs to

class i in the labels, and $Q(X, i)$ represents the probability that X belongs to class i in the model output.

We derive the L-balance loss function to obtain the gradient of the network parameter update as $-\left(1 - \frac{M_i}{M}\right)^{\beta} * (Q(i) - P(i))$. This indicates that the weight factor $\left(1 - \frac{M_i}{M}\right)^{\beta}$ does not affect the computation of the gradient, maintaining the advantage of fast gradient computation of the cross-entropy loss function. In addition, categories with more samples have smaller values of the weight factor and therefore less updates to the parameters of the network, which prevents overlearning of the network for categories with more samples. The specific procedure for calculating the gradient is presented in Appendix A.

2.3. Dilated Convolution

In recent years, dilated convolution has been widely used in tasks, such as semantic segmentation [22] and target detection [23]. Dilated convolution is supposed to expand the convolution kernel by adding some zero elements between the elements of the convolution kernel, and the expansion rate is used to express the degree of convolution kernel expansion. The comparison between normal convolution and dilated convolution operations is shown in Figure 2. Dilated convolution can greatly increase the effective receptive field of the convolutional network. The advantages of dilated convolution include fewer parameters, no change in the feature map size, and high resolution of image. Global information is important for image understanding, especially for fine-grained disease classification. The dilated convolution was applied to the classification of fine-grained crop diseases in this study.
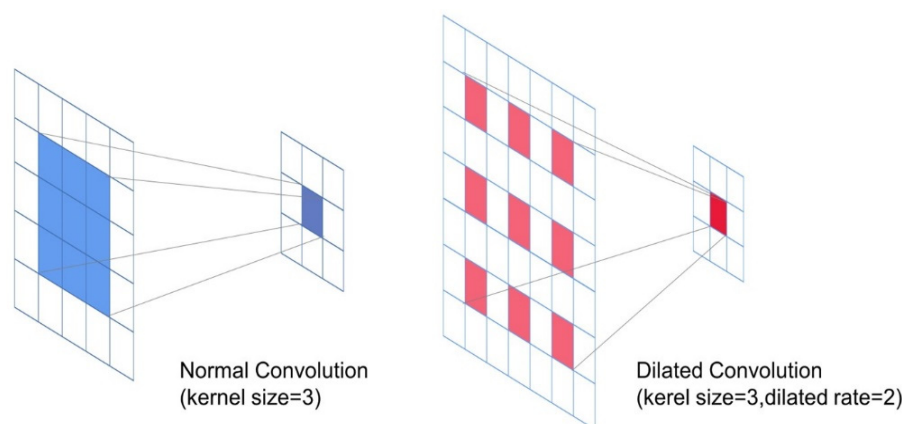

Figure 2. Comparison between normal convolution and dilated convolution.

However, the long-distance dilated convolution from the input layer will make the sampled signal sparse, destroy the local relevance of the image, and lose the detailed information learned in the shallow layer of the network, thus affecting the classification results. Accordingly, dilated convolution only in the deep convolution layer of the network was used to ensure that the effective receptive field of the network is expanded without destroying the local response properties in CNNs.

2.4. DC-DPCA Module

Different channels capture different characteristics, and channel attention is used to measure the importance of these channels. SE-net as a typical network of channel attention exhibits the strong performance that won the last ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [24]. Therefore, the channel attention module SE module in SE-net is employed as a prototype of the attention mechanism in this study. As shown in Figure 3, SE module consists of two main components, including squeeze and excitation. Global average pooling (GAP) is used to compress the features of a two-dimensional channel into a real number. Two fully connected layers allow the construction of interdependencies

between different channels while reducing the number of parameters. Finally, normalized weights are generated using a sigmoid function to reweight the features. In this way, the key features can be strengthened, and useless features can be suppressed.
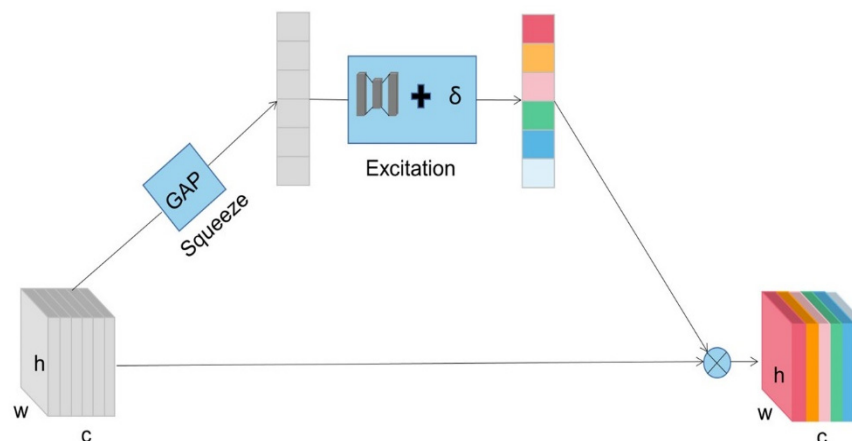

Figure 3. Structure diagram of SE module.

The traditional SE module uses the real numbers obtained by GAP as the global response of the channel, but the feature representation capability of GAP is insufficient and different feature maps may get the same result after GAP [25]. More importantly, GAP completely ignores the effect of important local responses. Therefore, we introduced dual-pooling into the squeeze operation of the channel attention module (DPCA). Specifically, we combined GAP and global max pooling (GMP) to enrich the input of features. GMP and GAP focus more on important local features and global features, respectively. After the mapping of fully connected layer, the correlation between global and local features can be constructed autonomously. For fine-grained disease classification, the network learns feature maps with subtle differences. Thus, the DC-DPCA module can achieve more effective fine-grained semantic understanding by learning the interdependence of global and local information. Since the feature maps in the shallow layer of the network are large in size and small in number, the information obtained by GMP is not representative and has negligible impact on feature compression. So, we also apply dual-pooling only to the attention module in the deep layer of the network.

The DC-DPCA module is a combination of dilated convolution (DC) and DPCA. In this paper, we replace some of the standard convolutional kernels in convolution blocks with dilation convolution and embed the DPCA module into the network. Dilated convolution can provide effective information input to the attention mechanism, and the DPCA module can perform effective feature reweighting on the feature map to pick out representative features. It is noteworthy that they are both deployed in the deeper layers of the network. The structure diagram of embedding the DC-DPCA module in the residual module is shown in Figure 4. The residual module is mainly composed of two convolution blocks and a shortcut. We replace the standard convolution in the residual module with the dilated convolution and embed the DPCA module between the two residual modules.

The overall structure of the DC-DPCA module embedded in ResNet50 [26] is shown in Figure 5. It mainly contains several residual modules, several SE modules, and several DC-DPCA modules. The crop disease images are resized to 224 × 224 and fed into the network. After the features are extracted by the network, they are fed into the classifier (the fully connected layer) for classification.
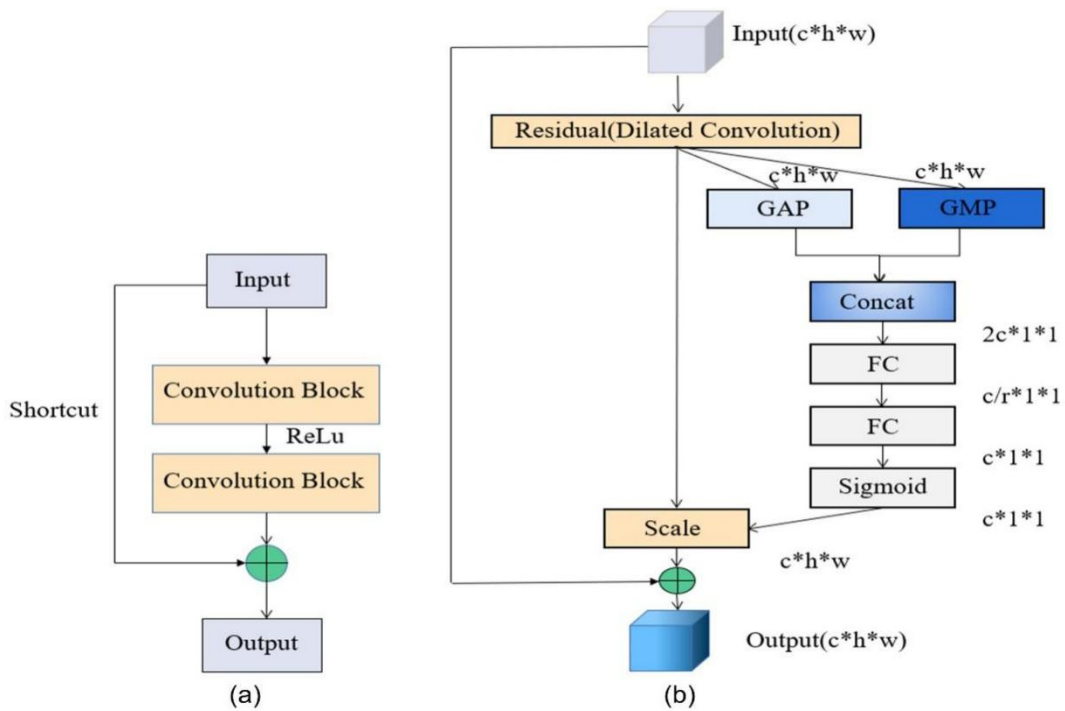
Figure 4. The DC-DPCA module. (**a**) The structure of the residual module. (**b**) The DC-DPCA module embedded in the residual module. FC represents the fully connected layer, c, h, and w represent the number of channels, height, and width of the input feature map, respectively, 2c*1*1, c/r*1*1, c*1*1, and c*h*w all represent the over size of different feature maps, and r represents the compression factor of the number of channels of the feature map.
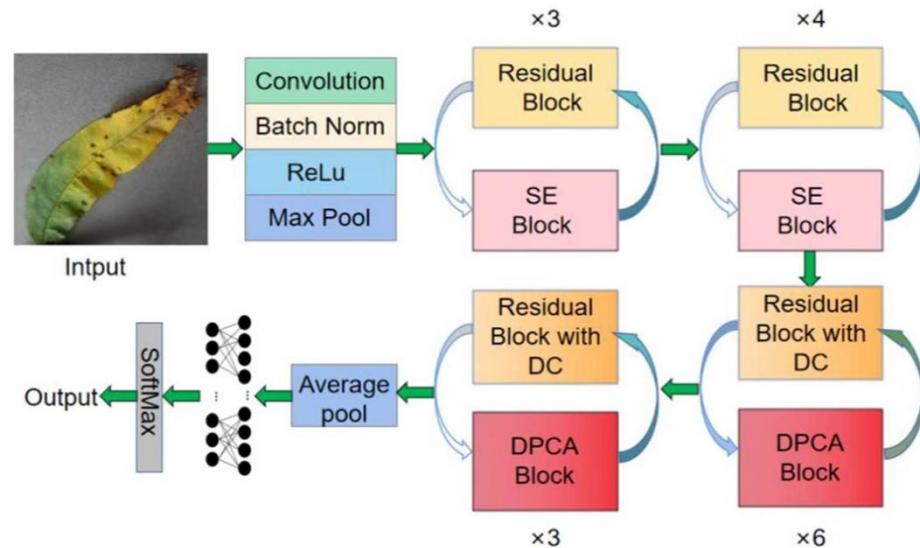
Figure 5. The overall structure of DC-DPCA module embedded in ResNet50.

2.5. Experimental Setup

Transfer learning can leverage existing knowledge to train models with better generalization performance faster [27] and using transfer learning can compensate for the lack of data volume in the dataset of this paper. We use the CNN models trained on the ImageNet dataset as pre-trained models, and to retain the already learned generalized shallow features, such as color, texture, and edges, we freeze the pre-trained weights of the shallow layers in the CNNs models and only fine-tune the weights of the deep layers.

The configuration of the experimental environment is shown in Table 1. The hyperparameter settings are shown below. A total of 40 epochs and Adam (as the optimizer) were

selected in this study. The batch size was set as 32. The initial learning rate is 0.0001 and a cosine annealing strategy was adopted to periodically adjust the learning rate to help the model get rid of saddle points [28]. The dilated rate of the dilated convolution was designed as a sawtooth structure (i.e., dilation rate = [1, 2, 3, 1, 2, 3 ... ...]) to avoid the gridding effect [29].

Table 1. Configuration of the experimental environment.

Name	Parameter
System	Windows 10
CPU	Intel(R) Core (TM) i5-6200U CPU
GPU	NVIDIA GeForce RTX 1080Ti
Deep learning framework	Pytorch 1.10.0 + cuda toolkit 10.1
Programming language	Python 3.7.0
Environment construction	Anaconda 3

2.6. Evaluation Metrics

In order to comprehensively evaluate the performance of the module, we used the following evaluation metrics: accuracy, precision, recall, and F1-score. The formulas are as follows:

$$\text{accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \tag{2}$$

$$\text{precision} = \frac{TP}{TP + FP} \tag{3}$$

$$\text{recall} = \frac{TP}{TP + FN} \tag{4}$$

$$\text{F1} - \text{score} = \frac{2TP}{2TP + FP + FN} \tag{5}$$

where TP, FP, TN, and FN denote the number of true-positive samples, the number of false-positive samples, the number of true-negative samples, and the number of false-negative samples, respectively.

3. Results

3.1. The Impact of L-Balance Loss Function

Using the traditional cross-entropy loss function (Figure 6a), we found that after the 35th epoch, the loss on the training set decreases, but the loss on the test set increases instead, which is a typical phenomenon of overfitting and seriously affects the learning ability of the model. The reason for this problem is the unbalanced distribution of the dataset. An unbalanced data distribution leads to an unbalanced loss distribution. Categories with more samples will generate a large percentage of losses, and the network will overlearn these categories with more attention to minimize the overall loss. This leads to poor generalization of the network and increasing errors on the test set (overfitting phenomenon). From Figure 6b, we can see that the L-balance loss function can well solve the overfitting problem caused by unbalanced data distribution. Since the L-balance loss function adds a weighting factor to the cross-entropy loss function, this weighting factor gives less weight to the category with a larger number of samples. This treatment balances the distribution of losses and does not lead to overlearning of the network for certain categories, which affects the generalization ability of the model.

As shown in Table 2, the accuracy of the ResNet50 with the L-balance loss function was 85.38%, which is 1.72% higher than that of the ResNet50 with cross-entropy loss function.
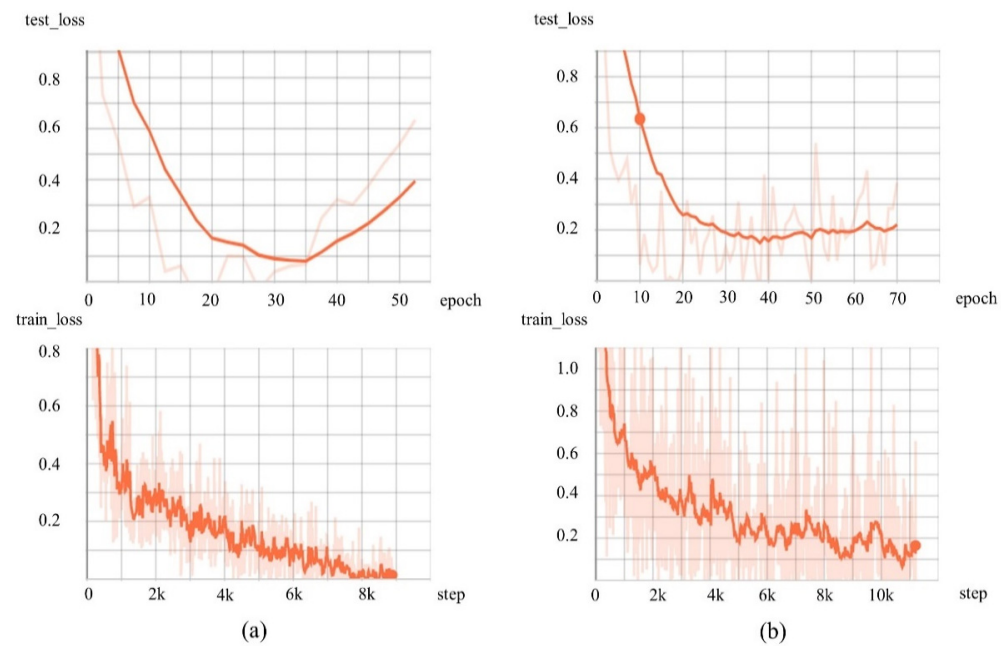
Figure 6. Relative trends of loss. (**a**) with cross-entropy. (**b**) with L-balance. Step represents the number of iterations during network training. One batch of data finished training represents one iteration.

Table 2. Comparison of the accuracy of two loss functions.

Loss Function	Accuracy
Cross-entropy	83.66%
L-balance	85.38%

3.2. Ablation Experiments

The traditional channel attention module (SE module), channel attention with dilated convolution module (DC-CA), DPCA module, and DC-DPCA module were embedded into ResNet50 to demonstrate the superiority of the method. DC-DPCA module is deployed in the last two stages of ResNet50 with dilated rates of [1, 2, 3, 1, 2, 3, ...]. Figure 7 shows the accuracy of embedding different modules in ResNet50. For the traditional channel attention module, no significant improvement of accuracy of fine-grained classification was observed.

The specific results are shown in Table 3, and the ResNet50 + DC-DPCA module has the best performance in terms of accuracy, precision, recall, and F1-score (87.14%, 87.17%, 87.07%, and 87.10%, respectively). Moreover, the DC-CA module and the DPCA module are not coupled, and they can both improve the classification accuracy independently.

Table 3. Results of ablation experiments.

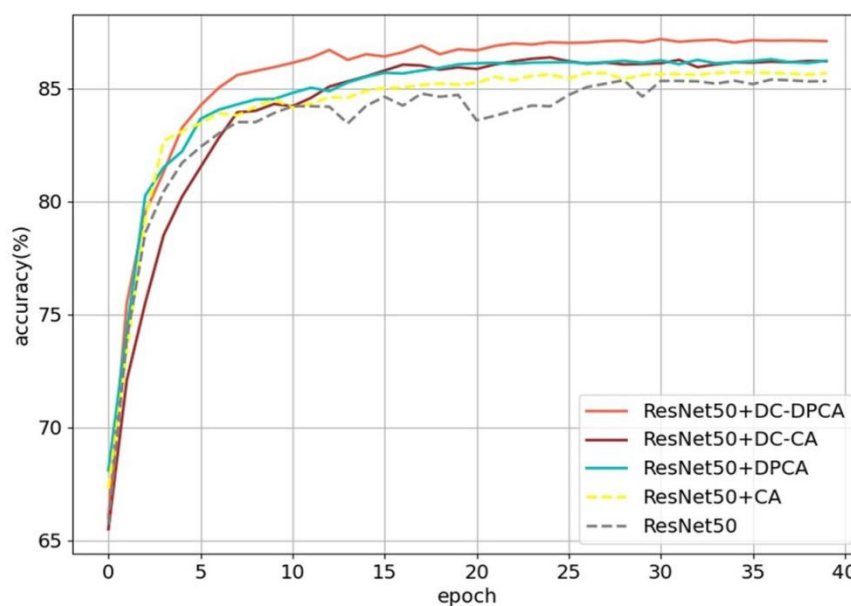| Model	Accuracy	Precision	Recall	F1-Score
ResNet50	85.38%	85.13%	84.80%	85.06%
ResNet50 + SE	85.70%	85.21%	85.77%	85.48%
ResNet50 + DC-CA	86.25%	86.20%	86.43%	86.33%
ResNet50 + DPCA	86.28%	85.54%	86.32%	86.13%
ResNet50 + DC-DPCA	87.14%	87.17%	87.07%	87.10%

Figure 7. Accuracy line graph for different modules embedded in ResNet50.

In order to understand the specific classification of each disease category, we analyzed the confusion matrix of the classification results for the data in the partial test set, as shown in the left part of Figure 8. Each column of the confusion matrix represents the predicted category and each row represents the true category, and only the elements on the diagonal of the confusion matrix are the elements that are correctly classified. Further, we chose the 25th category with the highest error rate (general citrus greening) to compare the classification ability of the network for difficult samples, as shown in the right part of Figure 8, where TP, TN, FP, and FN represent the number of samples with both labeled and predicted values of 25, the number of samples with both labeled and predicted values of 26, the number of samples with labeled values of 26 but predicted values of 25, and the number of samples with labeled values of 25 but predicted values of 26, respectively. Compared with the original network, the accuracy of ResNet50 with the embedded DC-DPCA module is greatly improved. This indicates that the DC-DPCA module has good classification ability for difficult samples.
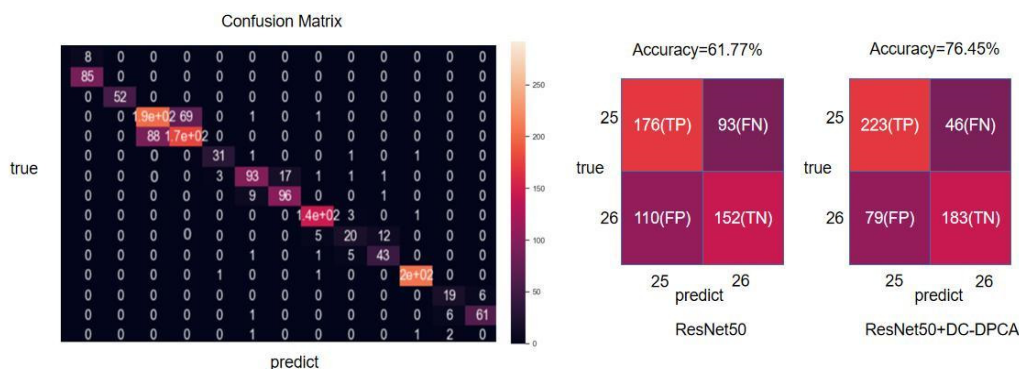

Figure 8. Comparison of class 25, 26 confusion matrices.

3.3. Experiments on Different Networks

To further demonstrate the robustness of our method, three classical deep CNNs (i.e., VGG16 [30], MobileNetV2 [31] and InceptionV3 [32]) were compared to avoid the effect of a single network structure. The DC-DPCA module was deployed in the last six convolutional layers of VGG16, the last six inverted residuals structures of MobileNetV2, and the last five inception structures of InceptionV3. As shown in Table 4, the classification

accuracies of embedding DC-DPCA modules on VGG16, MobileNetV2, and InceptionV3 were improved by 0.84%, 1.06%, and 0.93%, respectively, over those of embedding the conventional SE modules. These findings demonstrate the strong generalization ability the DC-DPCA module proposed in this paper.

Table 4. Experimental results on different networks.

| Original Model	Attention Module	Accuracy	Precision	Recall	F1-Score
VGG16	—	83.22%	82.93%	83.25%	82.96%
	SE	85.42%	84.72%	85.40%	85.22%
	DC-DPCA	86.26%	85.76%	86.41%	86.20%
MobileNetV2	—	83.85%	83.77%	83.93%	83.80%
	SE	85.18%	85.02%	85.29%	85.16%
	DC-DPCA	86.24%	86.35%	86.22%	86.23%
InceptionV3	—	84.60%	84.30%	84.58%	84.34%
	SE	85.84%	85.83%	85.48%	85.59%
	DC-DPCA	86.77%	86.73%	86.70%	86.72%

3.4. Visual Verification

We visualized the model from three different perspectives to verify the rationality of the method proposed in this paper.

3.4.1. Visualization of the Effective Receptive Field

We backpropagated the mean of the feature map to obtain the absolute value of the gradient of the input tensor, and then display it in a heat map (the yellow area represents the larger value, i.e., the effective receptive field, and the blue area represents the smaller value). The larger the gradient value, the greater the influence of the input region changes on the feature map, i.e., the region is in the effective receptive field of the network. A smaller gradient value indicates that the region has little impact on the network's judgment.

We visualized the effective receptive field of the last convolutional block of ResNet50, as shown in Figure 9, and the yellow area after visualization of ResNet50 with dilated convolution is larger compared to the original network, indicating that the effective receptive field of the network is larger. This can help the network to obtain a larger range of information for more effective judgments.


(a)   (b)
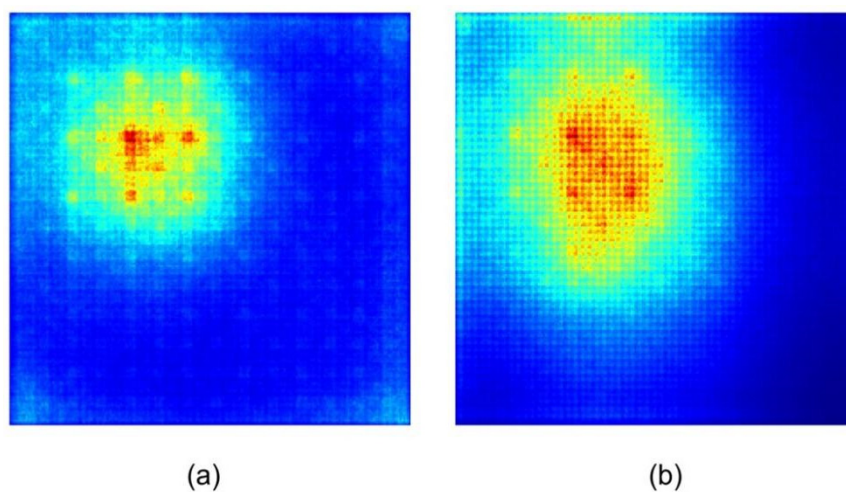
Figure 9. Comparison of the effective receptive field. (**a**) with standard convolution. (**b**) with dilated convolution.

3.4.2. *T*-SNE Visualization

We visualized the deep features adopting the t-SNE method [33]. The t-SNE is a dimensionality reduction method suitable for visualizing high-dimensional data. We use the t-SNE method to reduce the high-dimensional features extracted by the network into two dimensions, and different colors represent different kinds.

As shown in Figure 10a, the features of different kinds cannot be separated effectively, and the features of the same kind are not concentrated enough. This is due to the insufficient extraction ability of traditional CNNs and channel attention module in fine-grained image recognition. As shown in Figure 10b, the representations of ResNet50 embedded with DC-DPCA module are more compact and separable than those of the traditional channel attention module, proving that the DC-DPCA module allows the network to learn more discriminative features, which is helpful for fine-grained crop classification.
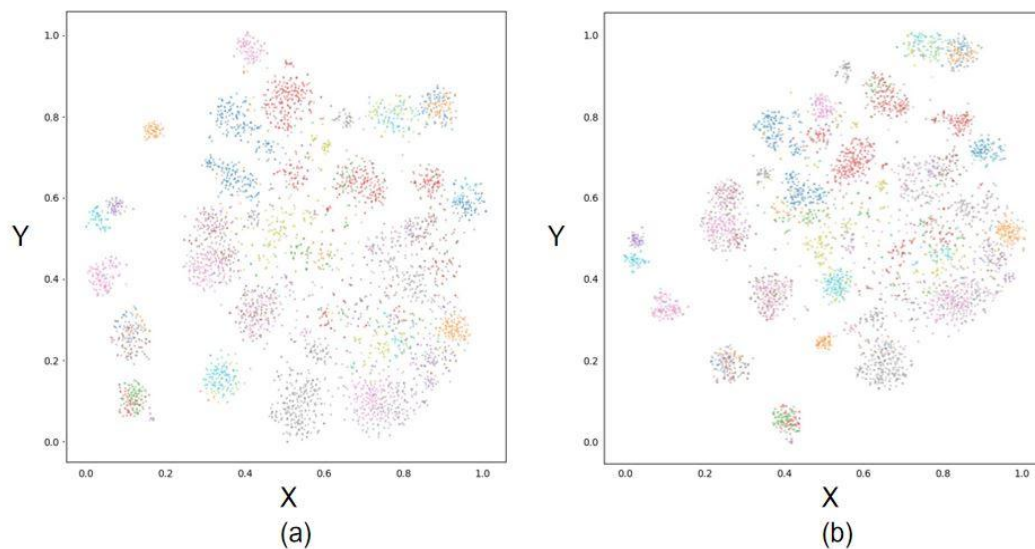
Figure 10. *t*-SNE of features learned by networks. (**a**) ResNet50 + SE. (**b**) ResNet50 + DC-DPCA.

3.4.3. Grad-CAM Visualization

The regions of interest of the network for a given category can be visualized using Grad-CAM [34], which can be used to know whether the network has learned the correct features or information. In order to further analyze the difference of accuracy, some images in this experiment were visualized using Grad-CAM (Figure 11). After embedding the DC-DPCA module, the areas of the network concerned are larger and more continuous, which can locate the disease areas in the images more accurately.
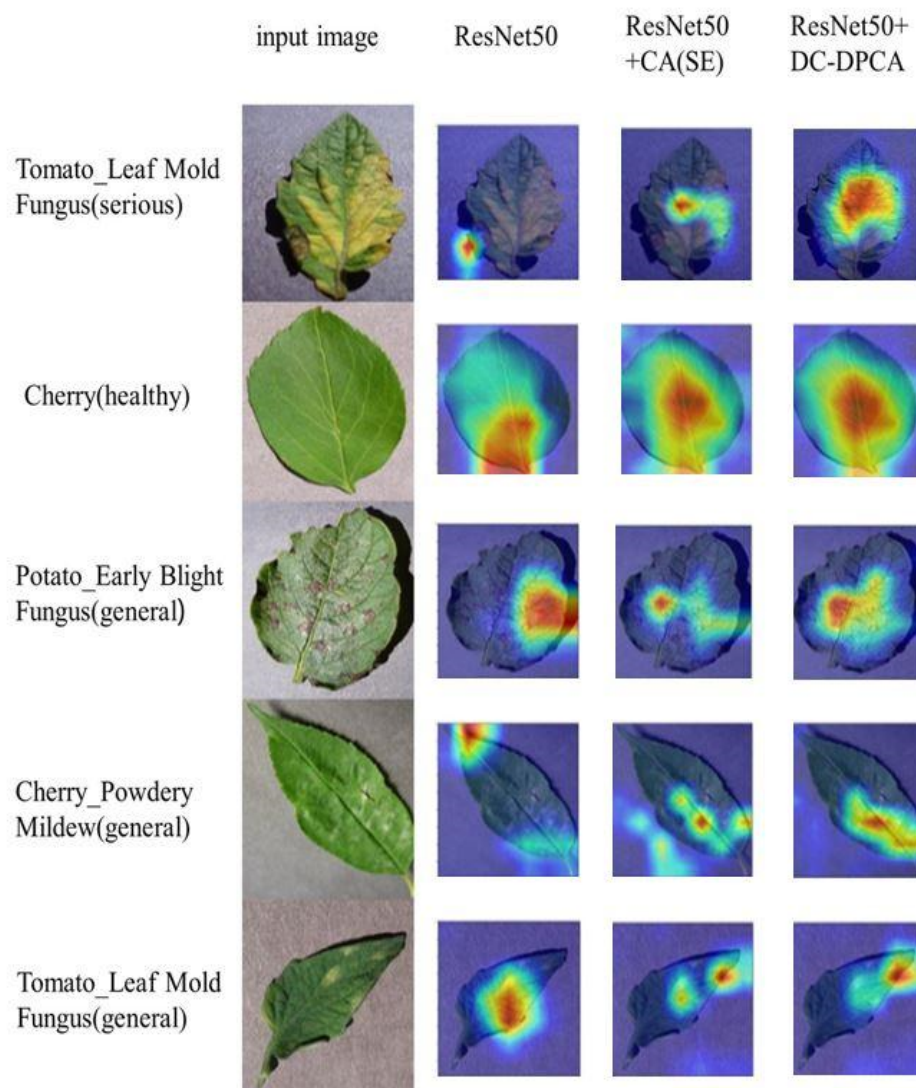
Figure 11. Grad-CAM comparison for embedding different attention modules in ResNet50.

4. Discussion

Crop disease is one of the main threats affecting agricultural production. Thus, achieving the accurate identification of crop disease is a very meaningful task. Deep CNNs have strong representation learning capabilities, and the integration of deep CNN techniques for crop disease identification has been widely used.

Because CNNs cannot accurately extract key features from subtle differences, the identification of fine-grained crop diseases has always been a challenging problem. Global information is very important for fine-grained crop image recognition, but the effective receptive field of CNNs is too small to obtain global information due to convolution operation. Therefore, we propose to expand the effective receptive field of the network by dilated convolution. In addition, the channel attention mechanism can improve the performance of the network, but the effect is not obvious in the fine-grained crop disease classification task. We consider that the reason for this may be that it ignores the response of local information. Therefore, we propose a dual-pooling channel attention mechanism to realize the interaction between global information and local information. We combine the two to form the DC-DPCA module, which is also more in line with the human visual judgment process, i.e., we first judge what the object is roughly from a whole, and then combine some important local features to make a more detailed judgment.

We have done abundant experiments to prove the superiority of DC-DPCA module. ResNet50, VGG16, MobileNetV2, and InceptionV3 embedded with the DC-DPCA module

have achieved an average accuracy of 87.14%, 86.26%, 86.24%, and 86.77%, respectively, which is 1.44%, 0.84%, 1.06%, and 0.87% higher than the networks embedded with the conventional SE module, respectively. At the same time, the classification metrics of precision, recall and F1-score are also improved. Moreover, we conducted ablation experiments on the DC-DPCA module, and the experimental results showed that compared with the original network, the accuracy of ResNet50 with embedded DC module and DPCA module increased by 0.87% and 0.90%, respectively, and the combination of the two was the best with 1.76% accuracy improvement. In the identification of difficult samples (general citrus greening), the DC-DPCA module can help the network achieve higher accuracy. In addition, we use a variety of visualization methods to fully prove that the DC-DPCA module can indeed help the network learn representative features in the case of fine-grained images with subtle differences.

In this study, we compared the results of our experiments on crop disease classification with those of some other literature, as shown in Table 5. With the same dataset, the accuracy of our method was 0.16%, 0.21%, and 0.79% higher than that of Wang et al. [19], Sun et al. [35], and Gao et al. [36], respectively. In addition, the dataset used by Lin et al. [37] is only a part of the dataset we used, with fewer types of crop diseases, but the accuracy of our method was still 0.85% higher than theirs, which indicates that our method has good generalization performance and can be used for large-scale crop disease identification.

Table 5. Comparison of our results with those of other literatures.

Paper	Model	Classification	Accuracy	Parameter	Time
Wang et al. [19]	InResV2 + I_CBAM	61-class	86.98%	122.47 MB	13.4 ms
Sun et al. [35]	SMLP_ResNet18	61-class	86.93%	48.6 MB	4.8 ms
Gao et al. [36]	DECA_ResNet50	61-class	86.35%	26.16 MB	2.3 ms
Lin et al. [37]	GrapeNet	7-class	86.29%	2.15 MB	1.9 ms
Ours	DC-DPCA + ResNet	59-class	87.14%	26.13 MB	2.2 ms

We adopt the dilated convolution as a way to expand the effective receptive field because it does not increase the number of parameters. Moreover, since the DC-DPCA module is only deployed in the deeper layers of the network, the additional computational cost is small. Therefore, we tested different models for the number of parameters and the average time to predict a picture and present the results in Table 5. Our method is not only more accurate, but also has lower time and storage costs compared to the methods of Wang et al., Sun et al., and Gao et al. This shows that our method has lower time complexity and space complexity and can be well applied in agricultural terminals to help achieve smart agriculture.

Overall, the DC-DPCA module is simple but effective, and successfully enhances CNNs' capacity to recognize fine-grained crop diseases.

5. Conclusions

In this study, a DC-DPCA module was proposed to collect a larger range of information, providing more reasonable input information to the attention module, and DPCA enriches the feature inputs to the channel attention module and allows the network to construct correlations between global and local features. The results of comparison experiments and ablation experiments demonstrated that the method proposed in this paper can improve the accuracy of fine-grained disease identification and has strong generalization performance. The visualization results also showed that the DC-DPCA module can help the network pick out key features that are more discriminative. In addition, the time cost and storage cost of our model are low, which is conducive to applications in mobile terminals for precision agriculture.

In the future, we intend to further optimize our approach by realizing the automatic adjustment of some hyperparameters, such as dilated rate, to accommodate the differences between different datasets. Meanwhile, we will do some research on model compression

and embed the model into mobile terminals such as smartphones to achieve crop disease identification in real agricultural environments.

Appendix A

The specific procedure for calculating the gradient of L-balance is as follows:

$$Q(i) = \frac{e^{Z_i}}{\sum_N e^{Z_N}} \text{ (SoftMax function)}$$

$$\text{L} - \text{balance} = -\sum_{i=0}^{N}\left(1 - \frac{M_i}{M}\right)^{\beta} * P(X,i) * \log(Q(X,i))$$

$$\text{gradient} = \frac{\partial(\text{L}-\text{balance})}{\partial Z_i} = \frac{\partial(\text{L}-\text{balance})}{\partial Q(j)} * \frac{\partial Q(j)}{\partial Z_i}$$

$$\frac{\partial(\text{L}-\text{balance})}{\partial Q(j)} = -\left(1 - \frac{M_i}{M}\right)^{\beta} * P(j) * \sum_N \frac{1}{Q(j)}$$

$$if\ i = j,\ \frac{\partial Q(i)}{\partial Z_i} = \frac{\sum_N e^{Z_N} * e^{Z_{i-(e^{Z_i})^2}}}{\sum_N (e^{Z_N})^2} = \frac{e^{Z_i}}{\sum_N e^{Z_N}} * \left(1 - \frac{e^{Z_i}}{\sum_N e^{Z_N}}\right) = Q(i) * (1 - Q(i))$$

$$if\ i \neq j,\ \frac{\partial Q(j)}{\partial Z_i} = -e^{Z_j} * \frac{1}{\sum_N e^{Z_N}} * e^{Z_i} = -Q(i) * Q(j)$$

$$\therefore \text{gradient} = -\left(1 - \frac{M_i}{M}\right)^{\beta} * P(j) * \sum_N \frac{1}{Q(j)} * \frac{\partial Q(j)}{\partial Z_i}$$

$$= -\left(1 - \frac{M_i}{M}\right)^{\beta} * \left(-\frac{P(i)}{Q(i)} * (1 - Q(i)) + \sum_{j \neq i} \frac{P(j)}{Q(j)} * Q(i) * Q(j)\right)$$

$$= -\left(1 - \frac{M_i}{M}\right)^{\beta} * \left(-P(i) + Q(i) * \sum_j P(j)\right)$$

$$= -\left(1 - \frac{M_i}{M}\right)^{\beta} * (Q(i) - P(i))$$

References

1. Lu, J.Z.; Tan, L.J.; Jiang, H.Y. Review on Convolutional Neural Network (CNN) Applied to Plant Leaf Disease Classification. *Agriculture* **2021**, *11*, 707. [CrossRef]
2. Zhao, S.Y.; Peng, Y.; Liu, J.Z.; Wu, S. Tomato Leaf Disease Diagnosis Based on Improved Convolution Neural Network by Attention Module. *Agriculture* **2021**, *11*, 651. [CrossRef]
3. 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]
4. Das, V.J.; Sharma, S.; Kaushik, A. Views of Irish Farmers on Smart Farming Technologies: An Observational Study. *AgriEngineering* **2019**, *1*, 164–187.

5. Yadav, S.; Kaushik, A.; Sharma, M.; Sharma, S. Disruptive Technologies in Smart Farming: An Expanded View with Sentiment Analysis. *AgriEngineering* **2022**, *4*, 424–460. [CrossRef]

6. Kaur, S.; Pandey, S.; Goel, S. Plants Disease Identification and Classification Through Leaf Images: A Survey. *Arch. Comput. Method Eng.* **2019**, *26*, 507–530. [CrossRef]

7. Guan, Z.; Tang, J.; Yang, B.; Zhou, Y.; Fan, D.; Yao, Q. Study on Recognition Method of Rice Disease Based on Image. *Chin. J. Rice Sci.* **2010**, *24*, 497–502.

8. Jiang, L.; Lu, S.; Feng, R.; Guo, Y. A Plant Pests and Diseases Detection Method Based on Multi-Features Fusion and Svm Classifier. *Comput. Appl. Softw.* **2014**, *31*, 186–190.

9. Huang, L.S.; Liu, Y.; Huang, W.J.; Dong, Y.Y.; Ma, H.Q.; Wu, K.; Guo, A.T. Combining Random Forest and XGBoost Methods in Detecting Early and Mid-Term Winter Wheat Stripe Rust Using Canopy Level Hyperspectral Measurements. *Agriculture* **2022**, *12*, 74. [CrossRef]

10. Min, S.; Lee, B.; Yoon, S. Deep learning in bioinformatics. *Brief. Bioinform.* **2017**, *18*, 851–869. [CrossRef] [PubMed]

11. Hinton, G.E.; Salakhutdinov, R.R. Reducing the dimensionality of data with neural networks. *Science* **2006**, *313*, 504–507. [CrossRef]

12. Liu, J.; Wang, X.W. Plant diseases and pests detection based on deep learning: A review. *Plant Methods* **2021**, *17*, 18. [CrossRef] [PubMed]

13. Sladojevic, S.; Arsenovic, M.; Anderla, A.; Culibrk, D.; Stefanovic, D. Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification. *Comput. Intell. Neurosci.* **2016**, *2016*, 11. [CrossRef] [PubMed]

14. Lu, Y.; Yi, S.J.; Zeng, N.Y.; Liu, Y.R.; Zhang, Y. Identification of rice diseases using deep convolutional neural networks. *Neurocomputing* **2017**, *267*, 378–384. [CrossRef]

15. Bhatt, P.; Sarangi, S.; Shivhare, A.; Singh, D.; Pappula, S. Identification of Diseases in Corn Leaves using Convolutional Neural Networks and Boosting. In Proceedings of the 8th International Conference on Pattern Recognition Applications and Methods (ICPRAM), Prague, Czech Republic, 19–21 February 2019; Scitepress: Prague, Czech Republic, 2019; pp. 894–899.

16. Yang, G.F.; He, Y.; Yang, Y.; Xu, B.B. Fine-Grained Image Classification for Crop Disease Based on Attention Mechanism. *Front. Plant Sci.* **2020**, *11*, 600854. [CrossRef] [PubMed]

17. Gao, R.; Wu, H.; Sun, X.; Gu, J. Crop Disease Recognition Method Based on Improved Channel Attention Mechanism. In Proceedings of the 2021 International Conference on Intelligent Computing, Automation and Applications, ICAA 2021, Nanjing, China, 25–27 June 2021; Institute of Electrical and Electronics Engineers Inc.: Nanjing, China, 2021; pp. 537–541.

18. Chen, Z.; Cao, M.; Ji, P.; Ma, F. Research on Crop Disease Classification Algorithm Based on Mixed Attention Mechanism. In Proceedings of the 2021 International Conference on Computer Engineering and Innovative Application of VR, ICCEIA VR 2021, Guangzhou, China, 11–13 June 2021; IOP Publishing Ltd.: Guangzhou, China, 2021.

19. Wang, M.; Wu, Z.; Zhou, Z. Fine-grained Identification Research of Crop Pests and Diseases Based on Improved CBAM via Attention. *Trans. Chin. Soc. Agric. Mach.* **2021**, *52*, 239–247.

20. Luo, W.J.; Li, Y.J.; Urtasun, R.; Zemel, R. Understanding the Effective Receptive Field in Deep Convolutional Neural Networks. In Proceedings of the 30th Conference on Neural Information Processing Systems (NIPS), Barcelona, Spain, 5–10 December 2016; Neural Information Processing Systems (Nips): Barcelona, Spain, 2016.

21. Ding, X.; Zhang, X.; Zhou, Y.; Han, J.; Ding, G.; Sun, J. Scaling Up Your Kernels to 31x31: Revisiting Large Kernel Design in CNNs. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 19–24 June 2022.

22. Huang, Y.; Wang, Q.Q.; Jia, W.J.; Lu, Y.; Li, Y.X.; He, X.J. See more than once: Kernel-sharing atrous convolution for semantic segmentation. *Neurocomputing* **2021**, *443*, 26–34. [CrossRef]

23. Li, Y.H.; Chen, Y.T.; Wang, N.Y.; Zhang, Z.X. Scale-Aware Trident Networks for Object Detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea, 11–17 October 2021; IEEE: Seoul, Korea, 2019; pp. 6053–6062.

24. Hu, J.; Shen, L.; Albanie, S.; Sun, G.; Wu, E.H. Squeeze-and-Excitation Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 2011–2023. [CrossRef]

25. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. ECA-Net: Efficient channel attention for deep convolutional neural networks. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle, WA, USA, 14–19 June 2020; IEEE Computer Society: New York, NY, USA, 2020; pp. 11531–11539.

26. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the 29th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, 26 June–1 July 2016; IEEE Computer Society: Las Vegas, NV, USA, 2016; pp. 770–778.

27. Pan, S.J.; Yang, Q.A. A Survey on Transfer Learning. *IEEE Trans. Knowl. Data Eng.* **2010**, *22*, 1345–1359. [CrossRef]

28. Loshchilov, I.; Hutter, F. SGDR: Stochastic gradient descent with warm restarts. In Proceedings of the 5th International Conference on Learning Representations, ICLR 2017, Toulon, France, 24–26 April 2017; ICLR: Toulon, France, 2017.

29. Wang, P.Q.; Chen, P.F.; Yuan, Y.; Liu, D.; Huang, Z.H.; Hou, X.D.; Cottrell, G. Understanding Convolution for Semantic Segmentation. In Proceedings of the 18th IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Tahoe, NV, USA, 12–15 March 2018; IEEE: Manhattan, NY, USA, 2018; pp. 1451–1460.

30. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. In Proceedings of the 3rd International Conference on Learning Representations, ICLR, San Diego, CA, USA, 7–9 May 2015; ICLR: San Diego, CA, USA, 2015.

31. Sandler, M.; Howard, A.; Zhu, M.L.; Zhmoginov, A.; Chen, L.C. MobileNetV2: Inverted Residuals and Linear Bottlenecks. In Proceedings of the 31st IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; IEEE: Salt Lake City, UT, USA, 2018; pp. 4510–4520.
32. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the Inception Architecture for Computer Vision. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 27–30 June 2016; IEEE: Seattle, WA, USA, 2016; pp. 2818–2826.
33. van der Maaten, L.; Hinton, G. Visualizing Data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.
34. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. In Proceedings of the 16th IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; IEEE: Venice, Italy, 2017; pp. 618–626.
35. Sun, W.; Wang, R.; Gao, R.; Li, Q.; Wu, H.; Feng, L. Crop Disease Recognition Based on Visible Spectrum and Improved Attention Module. *Spectrosc. Spectr. Anal.* **2022**, *42*, 1572–1580.
36. Gao, R.H.; Wang, R.; Feng, L.; Li, Q.F.; Wu, H.R. Dual-branch, efficient, channel attention-based crop disease identification. *Comput. Electron. Agric.* **2021**, *190*, 10. [CrossRef]
37. Lin, J.W.; Chen, X.Y.; Pan, R.Y.; Cao, T.B.; Cai, J.T.; Chen, Y.; Peng, X.S.; Cernava, T.; Zhang, X. GrapeNet: A Lightweight Convolutional Neural Network Model for Identification of Grape Leaf Diseases. *Agriculture* **2022**, *12*, 887. [CrossRef]