Title

Local and global feature-aware dual-branch networks for plant disease recognition

Authors

Jianwu Lin [1,2], Xin Zhang [1, 2, 3] *, Yongbin Qin [1,2], Shengxian Yang [3], Xingtian Wen [3], Tomislav Cernava [6], Quirico Migheli [7], Xiaoyulong Chen [4,5] *

Affiliations

[1] Text Computing & Cognitive Intelligence Engineering Research Center of National Education Ministry, College of Computer Science and Technology, Guizhou University, Guiyang, 550025, China;

[2] State Key Laboratory of Public Big Data, College of Computer Science and Technology, Guizhou University, Guiyang, 550025, China;

[3] College of Big Data and Information Engineering, Guizhou University, Guiyang, 550025, China;

[4] College of Life Sciences, Guizhou University, Guiyang, 550025, China;

[5] Guizhou-Europe Environmental Biotechnology and Agricultural Informatics Oversea Innovation Center in Guizhou University, Guizhou Provincial Science and Technology Department, Guiyang, 550025, China;

[6] School of Biological Sciences, Faculty of Environmental and Life Sciences, University of Southampton, Southampton S017 1BJ, UK;

[7] Dipartimento di Agraria and NRD - Nucleo di Ricerca sulla Desertificazione, Università degli Studi di Sassari, Sassari, Italy;

* Corresponding author.

Email addresses: xzhang1@gzu.edu.cn (X. Zhang), chenxiaoyulong@sina.cn (X. Chen),

Abstract

Accurate identification of plant diseases is important for ensuring the safety of agricultural production. Convolutional neural networks (CNNs) and visual transformers (VTs) can extract effective representations of images and have been widely used for the intelligent recognition of plant disease images. However, CNNs have excellent local perception with poor global perception, and VTs have excellent global perception with poor local perception. This makes it difficult to further improve the performance of both CNNs and VTs on plant disease recognition tasks. In this paper, we propose a local and global feature-aware dual-branch network, named LGNet, for the identification of plant diseases. More specifically, we first design a dual-branch structure based on CNNs and VTs to extract the local and global features. Then, an adaptive feature fusion (AFF) module is designed to fuse the local and global features, thus driving the model to dynamically perceive the weights of different features. Finally, we design a hierarchical mixed-scale unit-guided feature fusion (HMUFF) module to mine the key information in the features at different levels and fuse the differentiated information among them, thereby enhancing the model's multiscale perception capability. Subsequently, extensive experiments were conducted on the AI Challenger 2018 dataset and the self-collected corn disease (SCD) dataset. The experimental results demonstrate that our proposed LGNet achieves state-of-the-art recognition performance on both the AI Challenger 2018 dataset and the SCD dataset, with accuracies of 88.74% and 99.08%, respectively.

Keywords

Plant disease recognition, visual transformers, convolutional neural networks, adaptive feature fusion, hierarchical mixed-scale unit-guided feature fusion.

MAIN TEXT

1. Introduction

Safeguarding the agricultural production process has become an important economic imperative, and is aligned with the growing demand for improved quality and improved yields of agricultural products. Plant diseases are among the most important factors threatening the security of agricultural production [1]. At present, pests and diseases have led to a decline in the yields of a large number of crops [2]. Precise management of plant diseases could substantially reduce crop losses in the future. Accurate identification of plant diseases is a prerequisite for precision prevention and control [3, 4, 5]. However, traditional methods of plant disease identification, which are usually based on farmers' long-standing experience, are time-consuming and have difficulty meeting the demands of large-scale cultivation [6, 7, 8, 9].

In recent years, image processing technology has developed rapidly, and image-based plant disease recognition tasks have become popular [4, 10, 11, 12, 13, 14, 15]. For example, identifying plant diseases using traditional machine learning methods achieved good performance in the early days [16, 17, 18, 19]. However, this method usually requires manual feature extractions, making the recognition accuracy low when coping with large-scale plant diseases or plant diseases in the field environment. With the rapid development of artificial intelligence, deep learning, with its strong feature representation capabilities, has been increasingly applied for image recognition tasks, such as image classification [20, 21, 22], object detection [23, 24, 25], and image segmentation [26, 27, 28]. Currently, deep learning-based methods, which can be categorized into convolutional neural network (CNN)-based methods and visual Transformer (VT)-based methods, have become the mainstream methods for identifying plant diseases [29, 30, 31]. The CNN-based method uses a sliding window to extract features from plant disease images and therefore has excellent local feature perception with poor global feature perception [32, 33, 34]. The VT-based method transforms the image into multiple patches and then uses multi-head self-attention to extract the features of the plant disease image and therefore has excellent global feature perception with poor local feature perception [11, 35, 36]. However, some of the existing state-of-the-art (SOTA) methods for plant disease identification are based on a single CNN or VT, which severely limits their performance [35, 37, 38, 39, 40, 41, 42, 43]. This is because the symptoms exhibited by plant diseases are usually diverse, i.e., some plants are less infested with the pathogen, so the spots are localized, while some plants are more infested with the pathogen, so the spots are global, thus making it impossible for a single CNN or VT to strike a balance between local and global feature extraction. Although fusion architectures based on CNNs and VTs have been proposed, they still ignore the weights between features from different architectures [44].

To address the above challenge, we propose a local and global feature-aware dual-branch network, named LGNet, for plant disease recognition. Specifically, to extract both local and global disease features, we develop a hybrid two-branch network based on a CNN and VT. In addition, we design an adaptive feature fusion module and a multilevel feature fusion module for local and global disease feature perception and multiscale feature fusion, respectively. LGNet achieves state-of-the-art performance on two plant disease datasets. Our contributions are as follows:

(1) We propose a hybrid dual-branch network based on a CNN and VT for plant disease recognition.

(2) An adaptive feature fusion (AFF) module is designed for the adaptive perception of local and global features.

(3) We design a hierarchical mixed-scale unit-guided feature fusion (HMUFF) module to mine effective information in features and fuse it at multiple scales.

(4) Our proposed LGNet achieves state-of-the-art performance on the AI Challenger 2018 dataset and the self-collected corn (SCD) dataset.

2. Related work

2.1 Traditional machine learning methods

Recognizing plant diseases using traditional machine learning methods can be classified in three steps: image preprocessing, feature design, and classification using the obtained features [45]. Omrani et al. (2014) [46] used k-means to detect diseased areas in apple images and utilized a support vector machine (SVM) for classification. Rumpf et al. (2010) [17] used SVM and hyperspectral techniques to achieve early disease identification in sugar beets. Experiments showed that their method reached an accuracy of 97%. Phadikar et al. (2013) [47] selected rice disease features using rough set theory (RST). The experimental results showed that the RST method outperformed traditional methods in terms of feature selection. Plant disease identification using traditional machine learning methods usually has strong subjectivities and poor generalization abilities and has been gradually replaced by deep learning-based methods.

2.2 Deep learning-based methods

Unlike traditional methods, deep learning-based methods can automatically extract effective features of plant diseases and therefore have stronger generalizability. Nawaz et al. (2024) [48] proposed an improved CenterNet for coffee plant leaf disease recognition by introducing an improved ResNet-50. The accuracy and mean average precision (map) of their proposed model were 98.54% and 97%, respectively. Salamai et al. (2023) [11] developed a lesion-aware visual Transformer for paddy leaf disease detection. Their model achieved an average accuracy of 98.74% and an average f1-score of 98.18% on public paddy disease datasets. Thai et al. (2023) [49] proposed the least important attention pruning (LeIAP) algorithm to improve the Transformer model. Furthermore, they also used sparse matrix–matrix multiplication (SPMM) to calculate matrix correlations. Their developed model was more accurate and had a smaller number of parameters than the other models. Moreover, Faisal et al. (2023) [50] developed DFNet for plant disease classification by using a double-pretrained CNN model. The DFNet model achieved 97.53% and 94.65% accuracy on the two datasets, respectively. The abovementioned studies verified the feasibility of the deep learning-based methods. However, almost all these methods are based on a single CNN or VT. These methods are limited in their performance when applied to diverse plant disease datasets. For this reason, we develop a hybrid architecture based on a CNN and VT to address this challenge.

3. Materials and Methods

3.1 Dataset

AI Challenger 2018 dataset: The AI Challenger 2018 dataset is an open-source large-scale plant disease dataset containing a total of 36,258 plant disease images, of which 32,660 are in the training set and 3,598 are in the test set. The dataset can be categorized into 10 categories by disease type and 61 categories by disease severity. Representative samples are shown in Figure 1 (a). Plant diseases with severe symptoms exhibit global symptoms, and plant diseases with general symptoms exhibit local symptoms. In addition, we further divided the training set into a new training set and a validation set at a ratio of 9:1. Thus, 29,394 images are used for model training, 3,266 images are used for model validation, and 3,598 images are used for testing the final performance of the model.

Self-collected corn disease (SCD) dataset: The SCD dataset is available from four sources, namely, CD&S [51], PlantDoc [52], and websites. More specifically, we collected images of northern leaf blight, gray leaf spot, and northern leaf spot from the CD&S dataset, and corn embroidery disease images from PlantDoc. In addition, we obtained images of healthy corn and some diseased plants from the website. The SCD dataset contained a total of 3258 images of six corn leaf diseases. An 8:1:1 ratio is randomly used for division to obtain 2609 images in the training set, 324 images in the validation set, and 325 images in the test set. Representative samples of the SCD dataset are shown in Figure 1 (b). Corn leaf diseases are characterized by both local and global symptoms and are affected by complex conditions.

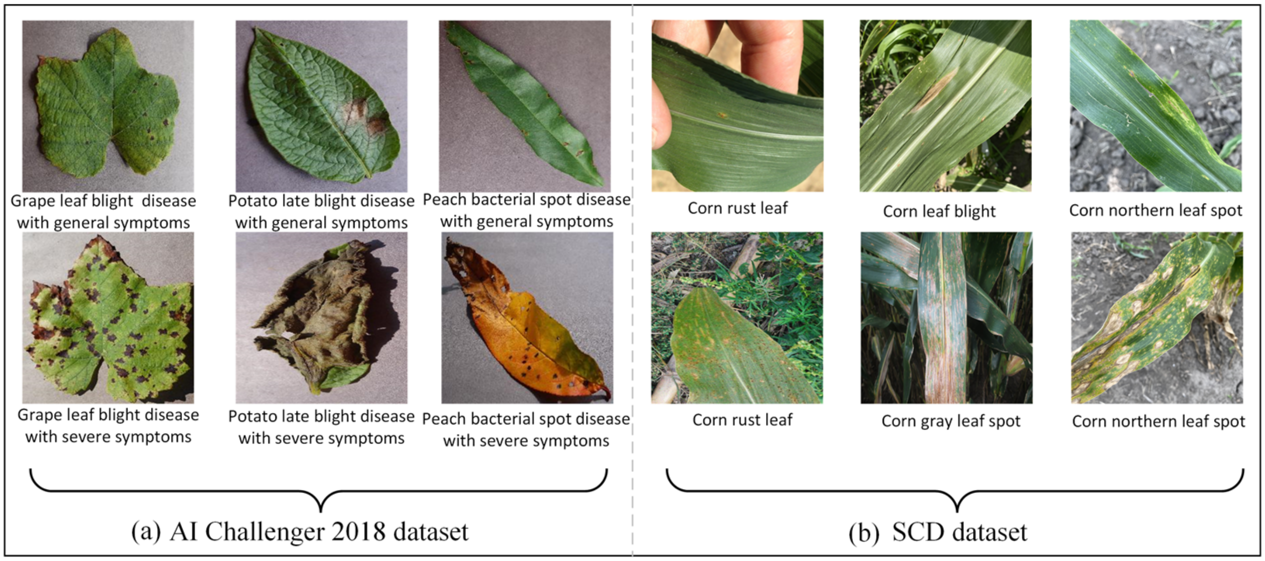
Grape leaf blight disease with general symptoms
Potato late blight disease with general symptoms
Peach bacterial spot disease with general symptoms
Corn rust leaf
Corn leaf blight
Corn northern leaf spot
Grape leaf blight disease with severe symptoms
Potato late blight disease with severe symptoms
Peach bacterial spot disease with severe symptoms
Corn rust leaf
Corn gray leaf spot
Corn northern leaf spot
(a) AI Challenger 2018 dataset
(b) SCD dataset

Figure 1. Representative examples from the AI Challenger 2018 and SCD datasets.

3.2 Proposed LGNet

Figure 2 shows the overall structure of LGNet, which consists of a dual-branch backbone network, adaptive feature fusion (AFF) modules, and hierarchical mixed-scale unit-guided feature fusion (HMUFF) modules. Specifically, we first crop the size of the input image to 224×224. Then, a dual-branch backbone network is used for multiscale feature extraction. In the dual-branch backbone network, we use ConvNext-Tiny [53] to extract the multiscale local features and Swin Transformer-Tiny [54] to extract the multiscale global features. Subsequently, the AFF modules are designed to achieve adaptive perceptual fusion of the local and global features, driving the model to dynamically perceive the local and global features. In addition, the HMUFF modules are designed to extract vital information from the features to guide efficient fusions between the multiscale features. More specifically, the HMUFF module mines differential features between different layers and then fuses them, thus providing each layer with a strong representation. Finally, the three scales of feature representations are fed into the pooling layers and classifiers to complete the identification of the plant disease.

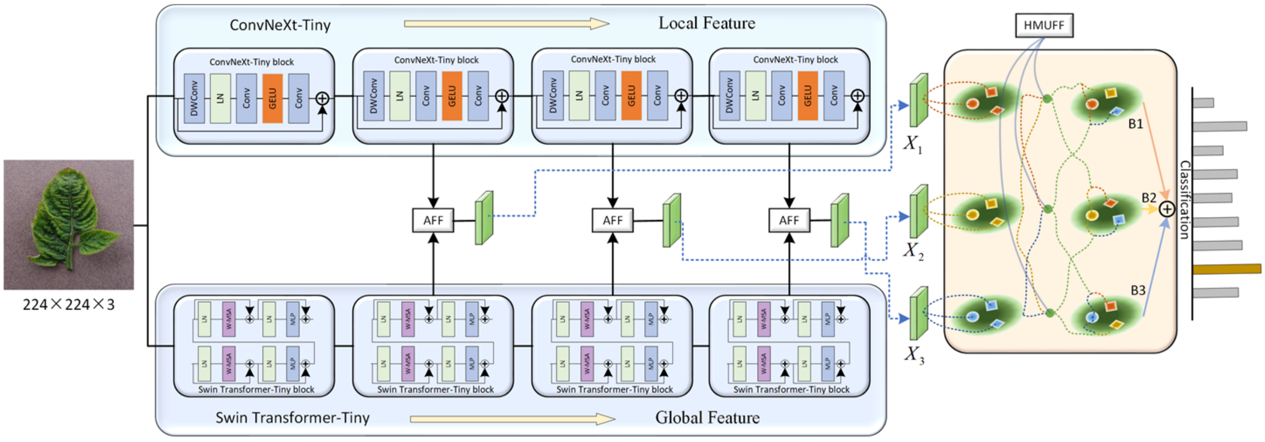

Figure 2. The overall architecture of LGNet consists of a dual-branch backbone network, adaptive feature fusion modules, and hierarchical mixed-scale unit-guided feature fusion modules.

3.2.1 Adaptive feature fusion (AFF) module

Usually, CNNs can capture local disease information, and VTs can capture global disease information. However, the diverse symptom representations of plant diseases make it necessary for the model to learn more appropriate feature representations for different plant diseases. To address this challenge, as described in the previous section, we designed a hybrid two-branch backbone network to extract the local and global features of plant diseases. Therefore, an efficient fusion module for fusing local and global features is needed. Here, we design the AFF module to implement the adaptive perceptual fusion of the local and global features, allowing the model to adaptively adjust the weights of the different features. Figure 3 illustrates the structure of the AFF module, which learns the adaptive weights of the two features and then weights the obtained weights on the original feature maps, enabling the model to adaptively perceive the local and global disease features, thus enhancing the model's feature representation capability.

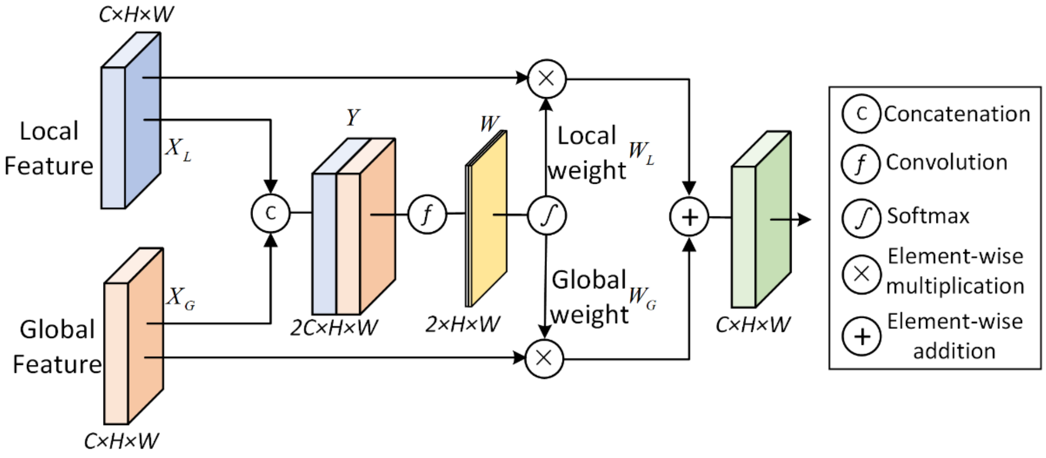

Figure 3. Structure of the AFF module.

Given a local feature input $X_L \in ¡^{C\times W\times H}$ and a global feature input $X_G \in ¡^{C\times W\times H}$, C, W, and H denote the channel number, width, and height of the feature map, respectively. First, we cascade the two feature maps (X_L and X_G) along the channel direction to obtain the input feature Y . It can be written as:

$$Y = Concat\{X_L; X_G\} \tag{1}$$

where $Y \in ¡^{2C\times W\times H}$. Then, we map the number of channels of the output to 2 via a convolution operation and use softmax to obtain the weighted feature map W. It can be written as:

$$W = Softmax(f_{3\times 3}(Y)) \tag{2}$$

where $f_{3\times 3}$ denotes a 3 × 3 convolution. Subsequently, we separate the weight feature map W along the channel directions to obtain the local weights W_L and global weights W_G. Finally, we apply the obtained weights to the original feature map and perform elementwise addition to obtain the adaptive output A:

$$A = X_L * W_L + X_G * W_G \tag{3}$$

3.2.2 Hierarchical mixed-scale unit-guided feature fusion (HMUFF) module

In general, different scales of feature maps contain different key information, so fusing the multiscale features facilitates the performance of the model. In LGNet, we design three branches in the backbone network to extract three scales of feature maps. Therefore, an efficient feature fusion method is needed for fusing feature information at these three scales. Here, we design a hierarchical mixed-scale unit-guided feature fusion (HMUFF) module for the fusion of different scale features. Specifically, the key ground information in the different scale features is first extracted by the hierarchical mixed-scale unit (HMU) module, thus guiding the well-designed module to realize the fusion of the multiscale features. As shown in Figure 4, an HMU module is first used to mine key information in the feature map, and then a feature fusion (FF) module is employed to fuse the differential features between the different scales.

Given two input feature maps $X_1 \in ¡^{C\times W_1\times H_1}$ and $X_2 \in ¡^{C\times W_2\times H_2}$, where *C*, *W*, and *H* denote the channel number, width, and height of the input feature map, respectively. First, we use an HMU to identify and highlight key disease features:

$$H_{X_1} = HMU(X_1) \tag{4}$$

$$H_{X_2} = HMU(X_2) \tag{5}$$

Then, H_{X_1} and H_{X_2} are flattened through the spatial direction to obtain $FH_{X_1} \in ¡^{C\times S_1}$ and $FH_{X_2} \in ¡^{C\times S_2}$. Here, $S_1 = W_1 * H_1$ and $S_2 = W_2 * H_2$. Subsequently, we obtain the differentiated feature matrix of the two feature maps via matrix multiplication, inverse, and softmax operations:

$$M_d = Softmax(-Z(Tanspose(FH_{X_1}), FH_{X_2})) \tag{6}$$

where Z denotes matrix multiplication. Next, we perform matrix multiplication to obtain the fusion information corresponding to the feature map:

$$F_1^2 = Reshape(Z(M_d, Tanspose(FH_{X_2}))) \tag{7}$$

$$F_2^1 = Reshape(Z(M_d, FH_{X_1})) \tag{8}$$

where $F_1^{\,2}$ denotes the fusion information corresponding to X_1, and F_2^1 denotes the fusion information corresponding to X_2. Therefore, the output of the HMUFF module is as follows:

$$X_1^2 = X_1 + F_1^2 \tag{9}$$

$$X_2^1 = X_2 + F_2^1 \tag{10}$$

The HMU module was originally proposed in the field of camouflaged object detection to mine the discriminative semantic features of camouflaged objects [32]. In this study, we design three multiscale branches based on a hybrid backbone network, as shown in Figure 2. Therefore, we first use the HMU module to refine the multiscale features and mine the discriminative feature information in different branches. For example, mining detailed information such as textures and edges in shallow features, and semantic information in deep features, can guide the FF module for efficient multiscale feature fusion. As shown in Figure 4, the structure of the HMU module can be divided into two stages, namely, groupwise iteration and channelwise modulation. For the groupwise iteration stage, a convolution operation is first used to expand the number of channels of the input feature map X. Then, we divide the feature maps into G groups $\{g_j\}_{j=1}^{G}$ along the channel direction. Next, we divide the first group $\{g_j\}$ into three parts $\{g_1'^k\}_{k=1}^3$ using the convolution operation. The first part, $g_1'^1$, is used for feature interaction with the next group, and the other two parts are used for channel modulation. With this continuous iterative approach, the key feature information in the channel is mined, enabling the module to be more discriminative with the semantic information. For the channelwise modulation stage, the features $[\{g_j'^2\}_{j=1}^G]$ are nonlinearly transformed to obtain the weights α, which are then weighted on the features $[\{g_j'^3\}_{j=1}^G]$ to further highlight the key information in the features. Finally, the output H of the HMU module can be written as:

$$H = \mathrm{A}\left(X + \mathrm{N}\left(\mathrm{T}\left(\alpha \cdot \left[\left\{ g_j'^3 \right\}_{j=1}^{G} \right] \right) \right) \right) \tag{11}$$

where A represents the activation function, N represents the normalization layer, and T represents the convolution operation.

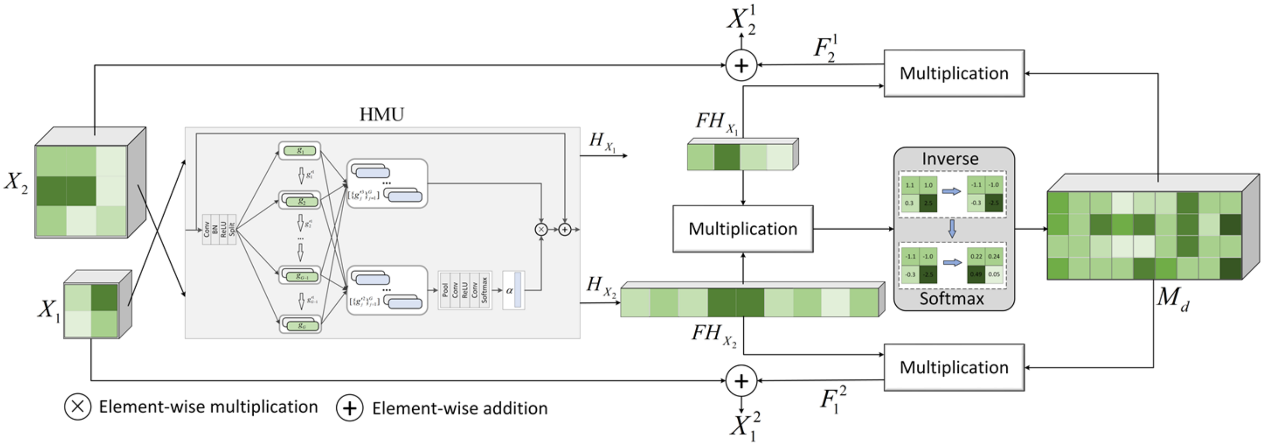

Figure 4. Structure of the HMUFF module.

4. Results

4.1 Experimental environment and evaluation indices

We divided the parameters of LGNet into two parts for training. First, we initialized the weights of the dual-branch backbone network using the official model weights on ImageNet 1k. Therefore, for the parameters of the backbone network, we simply needed to fine-tune them using an initial learning rate of $2e^{-5}$. On the other hand, for parameters that were not loaded with the corresponding pretrained weights, we used an initial learning rate of $2e^{-4}$. SGD was adopted as the optimizer with momentum=0.9 and weight decay=$5e^{-6}$. The batch size and number of iterations were set to 32 and 80, respectively. During the training process, we used online data augmentation with random rotation angles and random horizontal flips to enhance the generalization ability of the model. The cross-entropy function was used as the loss function for the three branches in the LGNet model. For all the experiments, we carried out training on a Windows 11 system with an NVIDIA GeForce RTX 3090 GPU and PyTorch.

For the quantitative evaluation of models, the evaluation indices used in this study are as follows:

$$\mathrm{Acc} = \frac{TP + TN}{TP + FP + TN + FN} \tag{12}$$

$$\mathrm{Pre} = \frac{TP}{TP + FP} \tag{13}$$

$$\mathrm{Rec} = \frac{TP}{TP + FN} \tag{14}$$

$$\mathrm{F1} = \frac{2TP}{2TP + FP + FN} \tag{15}$$

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

4.2 Comparisons with single models

To verify the better performance of LGNet compared to a single deep learning model, we conduct comparison experiments using ConvNeXt-Tiny and Swin Transformer-Tiny (the backbone networks in LGNet). Figure 5 illustrates the accuracy of each epoch on the validation set. All three models are initialized with pretraining weights during the training process, so the accuracy at the beginning of training can also be maintained at a high level. For example, the accuracy of the first epoch of all three models on the AI Challenger 2018 dataset is greater than 84%, and the accuracy of the first epoch of all three models on the SCD dataset is greater than 88%. However, more parameters need to be fine-tuned during the training of the LGNet model; thus, the initial accuracy of LGNet is slightly lower than that of the other two single models. More importantly, when the model converges to fit, LGNet's accuracy substantially improves, with its accuracy being approximately 1-2% higher than that of the single models on both datasets. These findings validate that LGNet has a more powerful feature extraction capability compared to a single model.

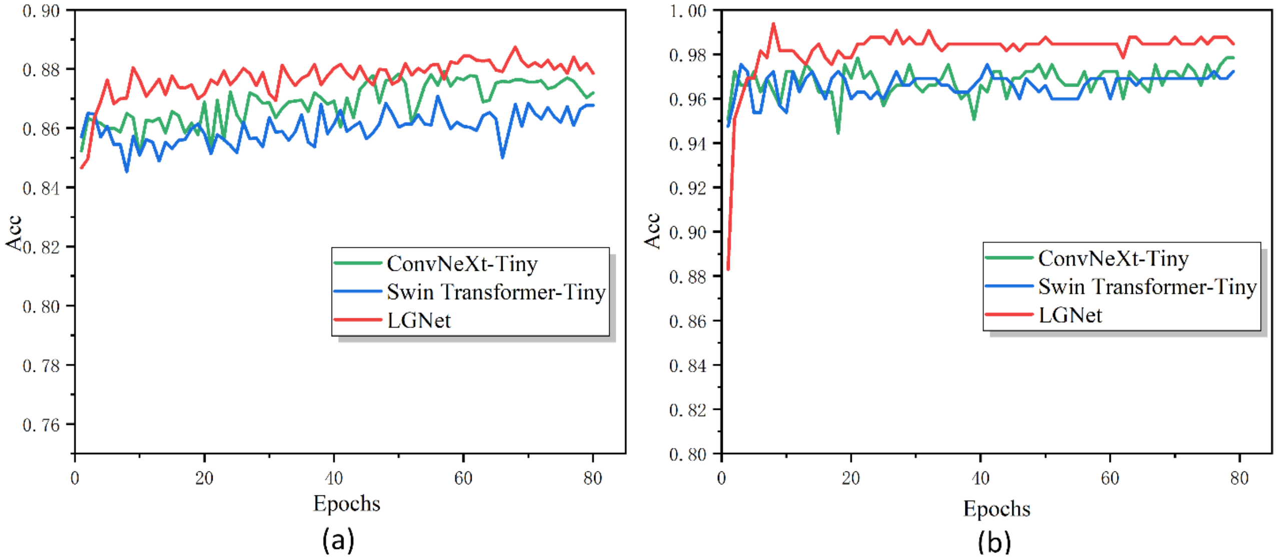

Figure 5. The accuracy of each epoch. (a) Results on the AI Challenger 2018 dataset. (b) Results on the SCD dataset.

Table 1 illustrates the comparison results with single deep learning models on the test set. For the AI Challenger 2018 dataset, the recognition accuracies of LGNet are 1.34% and 1.65% higher than those of ConvNeXt-Tiny and Swin Transformer-Tiny, respectively. For the SCD dataset, the recognition accuracy of LGNet is 1.55% and 1.88% higher than that of ConvNeXt-Tiny and Swin Transformer-Tiny, respectively. These results are almost the same as those analyzed on the validation set, further validating that our proposed LGNet has better performance advantages for plant disease identification tasks.

Table 1. Comparison results with single deep learning models on the test set

Model	Acc (%)	
	AI Challenger 2018 dataset	SCD dataset
ConvNeXt-Tiny	87.4	97.53
Swin Transformer-Tiny	87.09	97.2
LGNet	**88.74**	**99.08**

4.3 Ablation analysis

To validate the impact of the AFF modules and the HMUFF modules on the performance of the model, we perform ablation experiments on the AI Challenger 2018 and SCD datasets. Table 2 exhibits the comparison results of the ablation experiments. Both the use of the AFF modules and the use of the HMUFF modules improved the performance. In contrast, using the AFF modules on the benchmark model provides a better performance improvement than using the HMUFF modules on the benchmark model, with accuracies of 88.33% and 98.46% on the AI Challenger 2018 and SCD datasets, respectively. This may be because the dual-branch backbone network extracts rich disease features, and the AFF module allows LGNet to adaptively extract key global and local features, whereas the model using only the HMUFF module extracts a large amount of redundant information, which limits the performance of the model. Overall, the model that includes both the AFF modules and the HMUFF modules achieves a state-of-the-art performance in most of the metrics, with accuracies of 88.74% and 99.08% on the AI Challenger 2018 and SCD datasets, respectively, which are 0.77% and 1.23% better than those of the benchmark model.

Table 2. Comparison of the results of ablation experiments

AFF	HMUFF	AI Challenger 2018 dataset				SCD dataset			
		Acc (%)	Pre (%)	Rec (%)	F1 (%)	Acc (%)	Pre (%)	Rec (%)	F1 (%)
		87.97	84.03	82.08	82.68	97.85	97.85	97.76	97.79
✓		88.33	**85.71**	84.03	84.55	98.46	98.6	98.48	98.53
	✓	88.07	82.59	84.38	82.97	98.15	98.23	98.18	98.18
✓	✓	**88.74**	85.58	**84.68**	**84.94**	**99.08**	**99.31**	**99.18**	**99.23**

To show the effectiveness of the HMUFF modules, we visualize the feature maps of the three branches in LGNet. The visualization results are shown in Figure 6. B1, B2, and B3 represent the three scales of the branches. The model with the HMUFF module not only captures more detailed texture information but also retains richer disease information than the model without the HMUFF module. On the one hand, the HMU module in the HMUFF module is able to mine more detailed texture information; on the other hand, we utilize the differentiated matrix to fuse features at different scales, which enables different branches to fuse key information they do not have; therefore, the obtained feature maps are richer in disease information. These findings demonstrate the effectiveness of our designed HMUFF module.

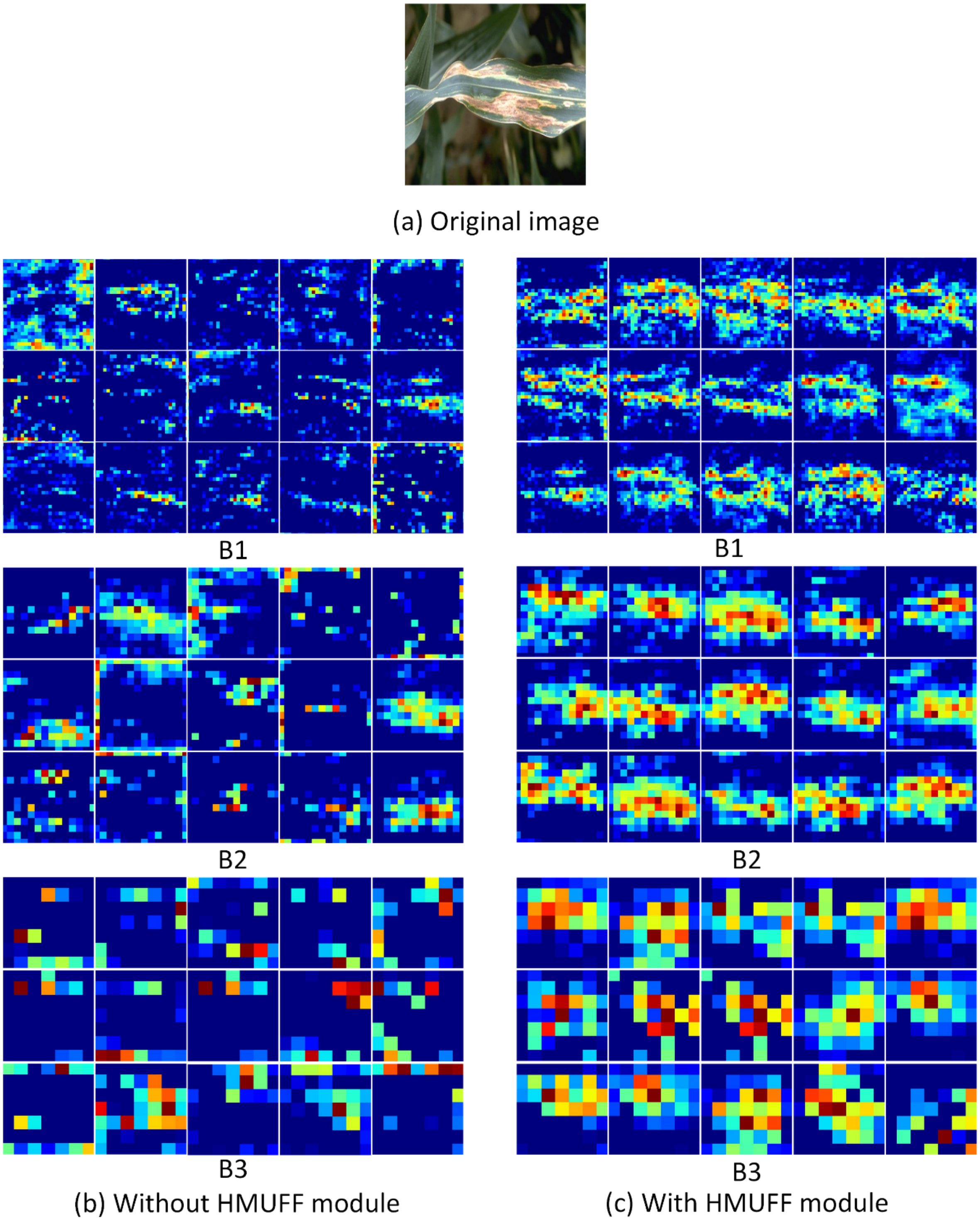

Figure 6. Visualization of feature maps for different branches. (a) represents the original image. (b) Model without the HMUFF module. (c) Model with HMUFF modules

The AFF module is designed to adaptively fuse the weights of the local and global features. Therefore, we validate the effectiveness of the AFF module using Grad-cam [55]. Figure 7 illustrates the obtained results. The model with the AFF module is able to localize the lesion area better and has a strong lesion perception ability. The model without the AFF module can only focus on part of the lesion area and has a poor lesion perception ability. These findings demonstrate the effectiveness of our designed AFF module.

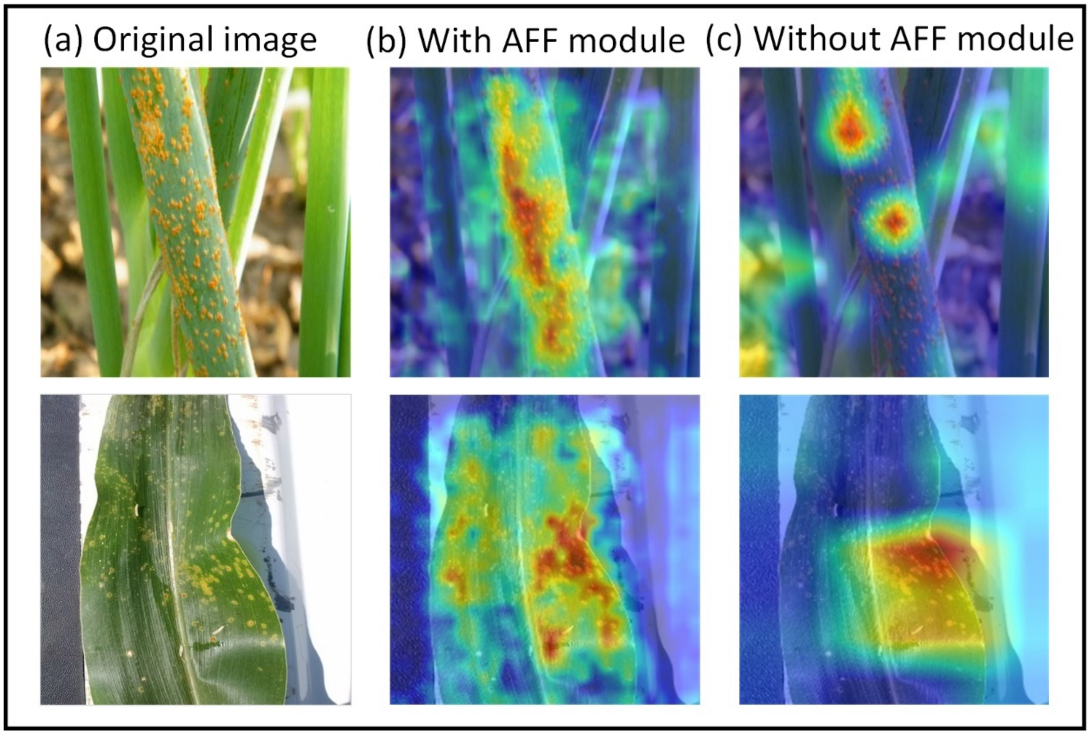

Figure 7. Class-activation mapping of the two models. (a) represents the original image. (b) Model without AFF modules. (c) Model with AFF modules

4.4 Impact of multiscale branches on LGNet

As shown in Figure 2, we use the multiscale features in LGNet to enable the model to learn features of different granularities. Specifically, there are three scales of branches. Therefore, to verify the impact of the multiscale branches on LGNet, we conduct comparative experiments using different combinations of the three branches, and the results on the test set are shown in Table 3. The performance of the combination "B1+B3" is slightly better than that of the combination "B2+B3". This may be because the B1 branch represents shallow features, which can provide a more subtle feature representation. In addition, both the "B1+B3" and "B2+B3" combinations had greater recognition performance than did the "B1+B2" combination, which indicates the greater importance of the F3 branch in the disease recognition task. Finally, the combination of B1+B2+B3 achieves the highest performance for most of the metrics, with accuracies of 88.74% and 99.08% on the AI Challenger 2018 and SCD datasets, respectively. These findings demonstrate the effectiveness of using the multiscale branches on LGNet.

Table 3. Experimental results using the multiscale branches on the test set

	AI Challenger 2018 dataset				SCD dataset			
	Acc (%)	Pre (%)	Rec (%)	F1 (%)	Acc (%)	Pre (%)	Rec (%)	F1 (%)
B1 + B2	87.99	84.76	83.26	83.72	98.15	98.23	98.18	98.18
B2 + B3	88.52	84.92	83.82	84.1	98.46	98.46	98.45	98.44
B1 + B3	88.55	**86.08**	84.38	84.91	98.77	98.93	98.87	98.88
B1 + B2 + B3	**88.74**	85.58	**84.68**	**84.94**	**99.08**	**99.31**	**99.18**	**99.23**

4.5 Visualization Analysis

In this section, we visualize the class activation mapping for some of the samples using Grad-CAM to demonstrate the regions of interest for different models. As shown in Figure 8, we chose the two samples with local disease features and global disease features from the AI Challenger 2018 dataset and SCD dataset, respectively. It can be seen that ConvNeXt-Tiny focuses on inaccurate or incomplete lesion areas and has weak lesion feature perception in complex scenes. Although Swin Transformer-Tiny can localize the lesion area, it also focuses on a large amount of redundant information. In contrast, our proposed LGNet not only focuses on global disease features but also accurately captures local features while suppressing complex background information.

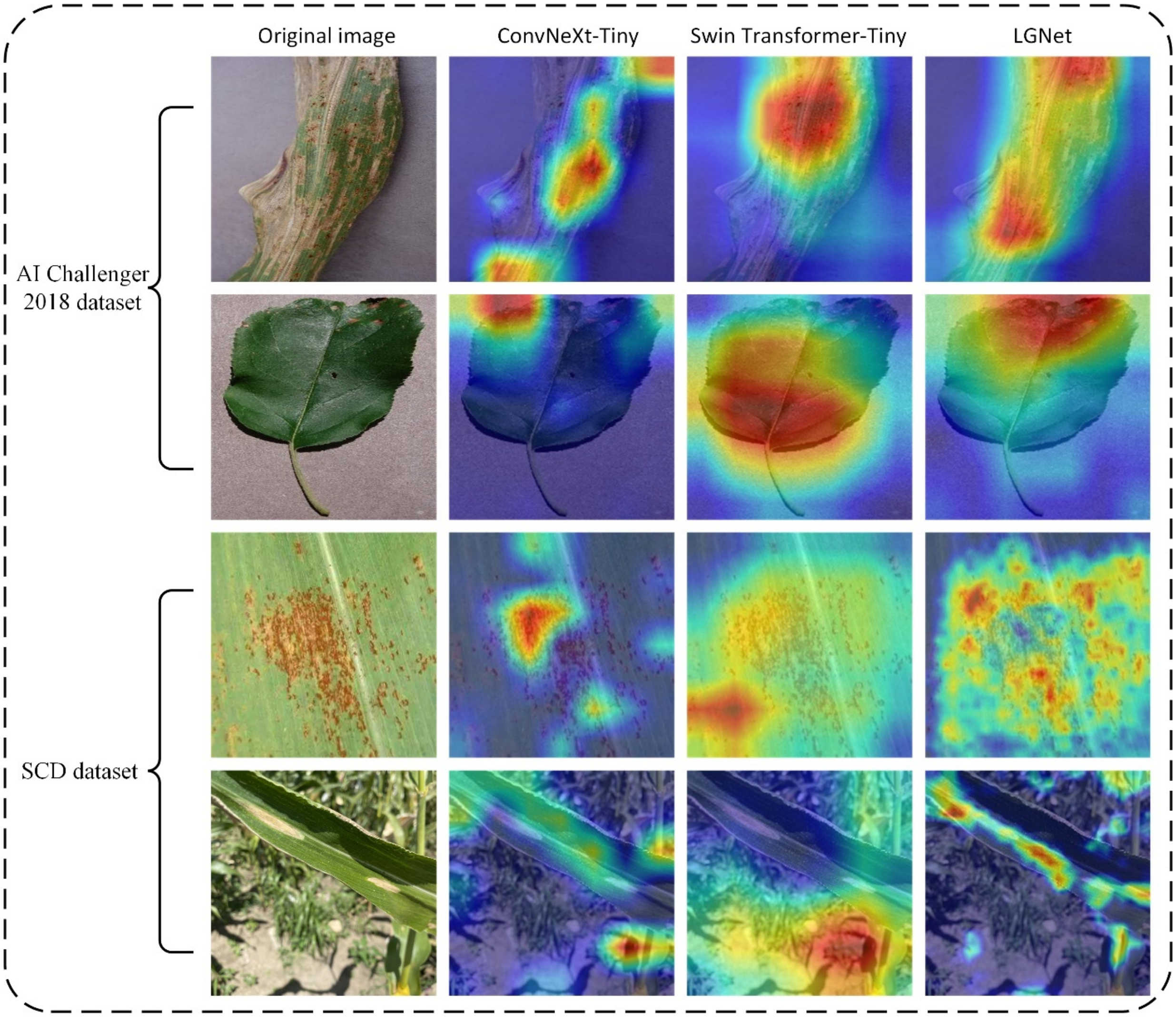

Figure 8. Class activation visualization results on the AI Challenger 2018 dataset and SCD dataset

To demonstrate the intra- and interclass distances in the test samples. We extract the last layer of the model and use t-distributed stochastic neighbor embedding (t-SNE) [56] to reduce the test set on the AI Challenger 2018 dataset to two dimensions. The t-SNE visualization results for the three models are shown in Figure 9. All three models can effectively distinguish between different disease types. However, for subtypes of the same disease, it is usually difficult for ConvNext-Tiny and Swin Transformer-Tiny to categorize them accurately. In contrast, LGNet not only implements precise classifications between disease categories but also improves the metric distance between different subcategories. This is because LGNet can adaptively perceive different symptomatic disease areas and introduces well-designed modules to extract key features, thus resulting in stronger feature representations.

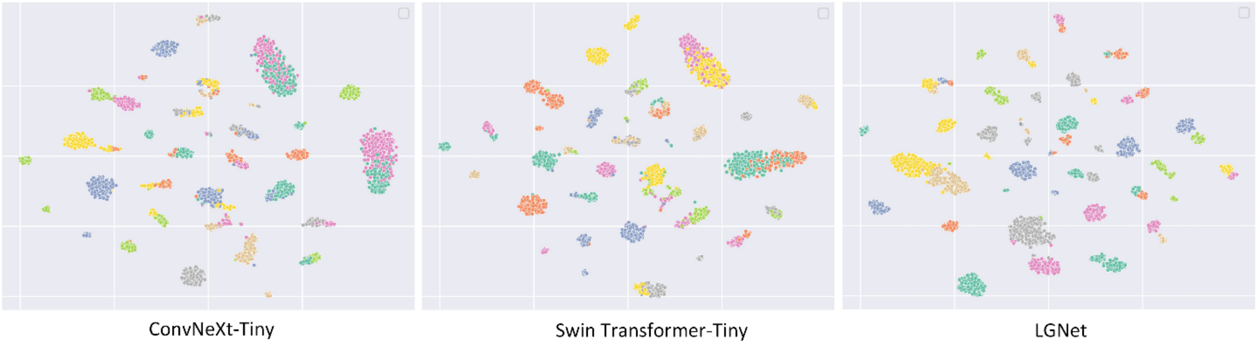

Figure 9. The t-SNE visualization results for the three models on the AI Challenger 2018 dataset

4.6 Comparisons with SOTAs on the AI Challenger 2018 dataset

The comparison results between our proposed LGNet and the state-of-the-art models on the AI Challenger 2018 dataset are shown in Table 4. LGNet achieves a state-of-the-art performance with an accuracy of 88.74%. In contrast, previous studies have almost always been designed based on a single CNN or VT model, making their performances limited to diverse plant disease datasets. Although the hybrid structure of CNN and VT is also used in the ConvVIT-Ti+ model, they simply fused the structure of the convolutional structure and the structure of the self-attention module and did not consider the trade-off between the global disease features and the local disease features. This makes their accuracy only 86.17%, which is 2.57% lower than that of LGNet, verifying that our proposed hybrid architecture based on a CNN and VT is effective.

Table 4. Comparison results on the AI Challenger 2018 dataset

Model	Acc (%)
Inception-ResNet-v2 [57]	86.1
DECA_ResNet [30]	86.35
Swin Transformer-Tiny [54]	87.09
ConvViT-Ti + [44]	86.17
ConvNeXt-Tiny [53]	87.4
SMLP_ResNet [58]	86.93
Improved attention model [59]	87.11
HCNet [60]	88.42
LGNet	**88.74**

4.7 Comparisons with SOTAs on the SCD dataset

We verify that LGNet also has a state-of-the-art performance on the plant disease recognition task in a field environment. We conduct comparative experiments on the SCD dataset using several classical models and several previous corn disease recognition models. Table 5 exhibits the comparison results on the SCD dataset. LGNet achieves the highest recognition accuracy of 99.08%, which is 0.61% and 0.31% higher than that of DFCANet and FCA-EfficientNet, respectively, further demonstrating the effectiveness of our proposed method.

Table 5. Comparison results on the SCD dataset

Model	Acc (%)
ResNet50 [61]	96.92
MobileNetV2 [62]	96.3
Swin Tranformer-Tiny [54]	97.2
ConvNeXt-Tiny [53]	97.53
DFCANet [63]	98.47
FCA-EfficientNet [64]	98.77
LGNet	**99.08**

5. Discussion

We propose a dual-branch network, LGNet, based on a CNN and VT for plant disease recognition. In LGNet, the AFF module is designed to efficiently fuse CNNs and VTs for local and global feature extraction of plant diseases. In addition, we design the HMUFF module to fuse multiscale disease features to further enhance the disease sensing capabilities. Subsequently, extensive experimental results validate that our method is more effective than a single deep network, and it outperforms the existing state-of-the-art plant disease recognition models on both the AI Challenger 2018 dataset and the SCD dataset.

Moreover, through our study, we found that there are still some key issues that need to be addressed, such as efficient plant disease recognition models and robust plant disease recognition models. We further discuss them in the following subsection.

(1) Efficient plant disease recognition models

Efficient plant disease recognition models are those that can accurately identify various plant diseases while maintaining low computational complexity, fast inference speeds, and minimal resource utilizations. In agricultural production environments, the development of efficient plant disease recognition models is particularly important due to the scarcity of computing power and poor equipment. Given that LGNet is designed based on the parallel structure of CNN and VT, the number of parameters and computations of LGNet will also increase significantly. Therefore, the development of efficient models using machine learning methods is an important direction for our future studies [65]. Knowledge distillation could be an effective means to solve this problem. Knowledge distillation aims to transfer knowledge from a larger, high-performing teacher model to a smaller, more efficient student model that can help retain the recognition capabilities of the former while inheriting the compactness of the latter [66]. In the future, we will develop LGNet with a larger backbone network (ConvNeXt-Base and Swin Transformer-Base) to gain more in-depth knowledge, and then use it as a teacher model to guide lightweight student models to learn more effective representations, thus designing highly efficient disease recognition models that can be deployed on mobile devices to help users identify plant diseases.

(2) Robust plant disease recognition models

In real agricultural environments, plant diseases not only have complex backgrounds but also suffer from other environmental factors, such as light, rain, and fog. This is a challenge for existing deep learning models. From the dataset perspective, obtaining more samples from real environments or mimicking real environments for data augmentation is extremely effective. From a model structure perspective, in general, large-size models have more complex network results and therefore greater robustness. In addition, model training using the current state-of-the-art training methods, such as the meta-learning-based methods and self-supervised learning-based methods, can improve the robustness of the model. Overall, the development of robust plant disease recognition models, and improving the generalization ability of these models in real-world environments, is highly important for agricultural production.

Acknowledgments

Funding: This study was supported by the National Key Research and Development Program of China (2021YFE0113700), the National Natural Science Foundation of China (32360705; 31960555), the Guizhou Provincial Science and Technology Program (2019-1410; HZJD[2022]001), the Outstanding Young Scientist Program of Guizhou Province (KY2021-026), and the Program for Introducing Talents to Chinese Universities (111 Program; D20023).

Author contributions: Jianwu Lin: Writing-original draft, funding acquisition, Writing-review & editing, Investigation. **Xin Zhang**: Resources, Conceptualization, Methodology, Writing-original draft, Supervision, Writing-review & editing. **Yongbin Qin**: Funding acquisition, Formal analysis. **Shengxian Yang:** Validation, Investigation. **Xingtian Wen**: Investigation, Visualization. **Tomislav Cernava**: Writing-original draft, Investigation, Writing-review & editing, Investigation. **Quirico Migheli**: Writing-original draft, Data curation. **Xiaoyulong Chen**: Writing-original draft, Funding acquisition, Supervision.

Competing interests: The authors declare that there are no conflicts of interest regarding the publication of this article.

Data availability: The experimental results obtained in this study are available from the corresponding author.

References

[1] Thakur, P.S., Khanna, P., Sheorey, T., Ojha, A., 2022. Trends in vision-based machine learning techniques for plant disease identification: A systematic review. Expert Systems with Applications, 118117.

[2] Gandhi, R., Nimbalkar, S., Yelamanchili, N., Ponkshe, S., 2018. Plant disease detec- tion using cnns and gans as an augmentative approach, in: 2018 IEEE International Conference on Innovative Research and Development (ICIRD), IEEE. pp. 1–5.

[3] Lin, J., Chen, X., Pan, R., Cao, T., Cai, J., Chen, Y., Peng, X., Cernava, T., Zhang, X., 2022. Grapenet: A lightweight convolutional neural network model for identification of grape leaf diseases. Agriculture 12, 887.

[4] Lin, J., Chen, Y., Pan, R., Cao, T., Cai, J., Yu, D., Chi, X., Cernava, T., Zhang, X., Chen, X., 2022. Camffnet: A novel convolutional neural network model for tobacco disease image recognition. Computers and Electronics in Agriculture 202, 107390.

[5] Khanna, M., Singh, L.K., Thawkar, S., Goyal, M., 2024. Planet: a robust deep convolutional neural network model for plant leaves disease recognition. Multimedia Tools and Applications 83, 4465–4517

[6] Pal, A., Kumar, V., 2023. Agridet: Plant leaf disease severity classification using agriculture detection framework. Engineering Applications of Artificial Intelligence 119, 105754.

[7] Sahu, S.K., Pandey, M., 2023. An optimal hybrid multiclass svm for plant leaf disease detection using spatial fuzzy c-means model. Expert Systems with Applications 214, 118989.

[8] Yu, S., Xie, L., Huang, Q., 2023. Inception convolutional vision transformers for plant disease identification. Internet of Things 21, 100650.

[9] Shi, T., Liu, Y., Zheng, X., Hu, K., Huang, H., Liu, H., Huang, H., 2023. Recent advances in plant disease severity assessment using convolutional neural networks. Scientific Reports 13, 2336.

[10] Zhang, S., Zhang, C., 2023. Modified u-net for plant diseased leaf image segmentation. Computers and Electronics in Agriculture 204, 107511.

[11] Salamai, A.A., Ajabnoor, N., Khalid, W.E., Ali, M.M., Murayr, A.A., 2023. Lesion-aware visual transformer network for paddy diseases detection in precision agriculture. European Journal of Agronomy 148, 126884.

[12] Parthiban, S., Moorthy, S., Sabanayagam, S., Shanmugasundaram, S., Naganathan, A., Annamalai, M., Balasubramanian, S., 2023. Deep learning based recognition of plant diseases, in: Computer Vision and Machine Intelligence Paradigms for SDGs: Select Proceedings of ICRTAC-CVMIP 2021. Springer, pp. 83–93.

[13] Dong, X., Wang, Q., Huang, Q., Ge, Q., Zhao, K., Wu, X., Wu, X., Lei, L., Hao, G., 2023. Pddd-pretrain: A series of commonly used pre-trained models support image-based plant disease diagnosis. Plant Phenomics 5, 0054

[14] Dong, X., Zhao, K., Wang, Q., Wu, X., Huang, Y., Wu, X., ... & Hao, G. (2024). PlantPAD: a platform for large-scale image phenomics analysis of disease in plant science. Nucleic Acids Research, 52(D1), D1556-D1568.

[15] Wu, X., Deng, H., Wang, Q., Lei, L., Gao, Y., & Hao, G. (2023). Meta-learning shows great potential in plant disease recognition under few available samples. The Plant Journal, 114(4), 767-782.

[16] Sankaran, S., Mishra, A., Ehsani, R., Davis, C., 2010. A review of advanced techniques for detecting plant diseases. Computers and electronics in agriculture 72, 1–13.

[17] Rumpf, T., Mahlein, A.K., Steiner, U., Oerke, E.C., Dehne, H.W., Plu¨mer, L., 2010. Early detection and classification of plant diseases with support vector machines based on hyperspectral reflectance. Computers and electronics in agriculture 74, 91–99.

[18] Arivazhagan, S., Shebiah, R.N., Ananthi, S., Varthini, S.V., 2013. Detection of un- healthy region of plant leaves and classification of plant leaf diseases using texture features. Agricultural Engineering International: CIGR Journal 15, 211–217.

[19] Mahlein, A.K., Oerke, E.C., Steiner, U., Dehne, H.W., 2012. Recent advances in sensing plant diseases for precision crop protection. European Journal of Plant Pathology 133, 197–209.

[20] Ning, X., Tian, W., Yu, Z., Li, W., Bai, X., Wang, Y., 2022. Hcfnn: high-order coverage function neural network for image classification. Pattern Recognition 131, 108873.

[21] Song, J., Yang, R., 2021. Feature boosting, suppression, and diversification for fine-grained visual classification, in: 2021 International Joint Conference on Neural Net- works (IJCNN), IEEE. pp. 1–8.

[22] Cheng, G., Lai, P., Gao, D., Han, J., 2023. Class attention network for image recognition. Science China Information Sciences 66, 132105

[23] Li, Y., Mao, H., Girshick, R., He, K., 2022. Exploring plain vision transformer backbones for object detection, in: European Conference on Computer Vision, Springer. pp. 280–296.

[24] Zou, Z., Chen, K., Shi, Z., Guo, Y., Ye, J., 2023. Object detection in 20 years: A survey. Proceedings of the IEEE.

[25] Wang, C.Y., Bochkovskiy, A., Liao, H.Y.M., 2023. Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 7464–7475

[26] Jain, J., Li, J., Chiu, M.T., Hassani, A., Orlov, N., Shi, H., 2023. Oneformer: One transformer to rule universal image segmentation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2989–2998.

[27] Mazurowski, M.A., Dong, H., Gu, H., Yang, J., Konz, N., Zhang, Y., 2023. Segment anything model for medical image analysis: an experimental study. Medical Image Analysis 89, 102918

[28] Heidari, M., Kazerouni, A., Soltany, M., Azad, R., Aghdam, E.K., Cohen-Adad, J., Merhof, D., 2023. Hiformer: Hierarchical multi-scale representations using trans-formers for medical image segmentation, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 6202–6212

[29] Liu, H., Zhan, Y., Xia, H., Mao, Q., Tan, Y., 2022a. Self-supervised transformer- based pre-training method using latent semantic masking auto-encoder for pest and disease classification. Computers and Electronics in Agriculture 203, 107448.

[30] Gao, R., Wang, R., Feng, L., Li, Q., Wu, H., 2021. Dual-branch, efficient, channel attention-based crop disease identification. Computers and Electronics in Agriculture 190, 106410.

[31] Wu, X., Fan, X., Luo, P., Choudhury, S.D., Tjahjadi, T., Hu, C., 2023. From laboratory to field: Unsupervised domain adaptation for plant disease recognition in the wild. Plant Phenomics 5, 0038.

[32] Pang, Y., Zhao, X., Xiang, T.Z., Zhang, L., Lu, H., 2022. Zoom in and out: A mixed-scale triplet network for camouflaged object detection, in: Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition, pp. 2160– 2170.

[33] Haar, L.V., Elvira, T., Ochoa, O., 2023. An analysis of explainability methods for convolutional neural networks. Engineering Applications of Artificial Intelligence 117, 105606.

[34] Ding, X., Zhang, X., Ma, N., Han, J., Ding, G., Sun, J., 2021. Repvgg: Making vgg-style convnets great again, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 13733–13742.

[35] Liu, Y., Zhang, Y., Wang, Y., Hou, F., Yuan, J., Tian, J., Zhang, Y., Shi, Z., Fan, J., He, Z., 2023. A survey of visual transformers. IEEE Transactions on Neural Networks and Learning Systems.

[36] Han, K., Wang, Y., Chen, H., Chen, X., Guo, J., Liu, Z., Tang, Y., Xiao, A., Xu, C., Xu, Y., et al., 2022. A survey on vision transformer. IEEE transactions on pattern analysis and machine intelligence 45, 87–110.

[37] Zhang, D., Huang, Y., Wu, C., Ma, M., 2023. Detecting tomato disease types and degrees using multi-branch and destruction learning. Computers and Electronics in Agriculture 213, 108244.

[38] Zhang, S., Zhang, C., 2023. Modified u-net for plant diseased leaf image segmentation. Computers and Electronics in Agriculture 204, 107511.

[39] Liu, Y., Wang, Z., Wang, R., Chen, J., Gao, H., 2023. Flooding-based mobilenet to identify cucumber diseases from leaf images in natural scenes. Computers and Electronics in Agriculture 213, 108166.

[40] Thakur, P.S., Chaturvedi, S., Khanna, P., Sheorey, T., Ojha, A., 2023. Vision transformer meets convolutional neural network for plant disease classification. Ecological Informatics 77, 102245.

[41] Li, G., Wang, Y., Zhao, Q., Yuan, P., Chang, B., 2023. Pmvt: a lightweight vision transformer for plant disease identification on mobile devices. Frontiers in Plant Science 14, 1256773.

[42] Pan, R., Lin, J., Cai, J., Zhang, L., Liu, J., Wen, X., Chen, X., Zhang, X., 2023. A two-stage feature aggregation network for multi-category soybean leaf disease identification. Journal of King Saud University-Computer and Information Sciences 35, 101669.

[43] Lin, J., Yu, D., Pan, R., Cai, J., Liu, J., Zhang, L., Wen, X., Peng, X., Cernava, T., Oufensou, S., et al., 2023b. Improved yolox-tiny network for detection of tobacco brown spot disease. Frontiers in Plant Science 14, 1135105.

[44] Li, X., Chen, X., Yang, J., Li, S., 2022. Transformer helps identify kiwifruit diseases in complex natural environments. Computers and Electronics in Agriculture 200, 107258.

[45] Al-Hiary, H., Bani-Ahmad, S., Reyalat, M., Braik, M., Alrahamneh, Z., 2011. Fast and accurate detection and classification of plant diseases. International Journal of Computer Applications 17, 31–38.

[46] Omrani, E., Khoshnevisan, B., Shamshirband, S., Saboohi, H., Anuar, N.B., Nasir, M.H.N.M., 2014. Potential of radial basis function-based support vector regression for apple disease detection. Measurement 55, 512–519.

[47] Phadikar, S., Sil, J., Das, A.K., 2013. Rice diseases classification using feature selection and rule generation techniques. Computers and electronics in agriculture 90, 76–85.

[48] Nawaz, M., Nazir, T., Javed, A., Amin, S.T., Jeribi, F., Tahir, A., 2024. Coffeenet: A deep learning approach for coffee plant leaves diseases recognition. Expert Systems with Applications 237, 121481.

[49] Thai, H.T., Le, K.H., Nguyen, N.L.T., 2023. Formerleaf: An efficient vision transformer for cassava leaf disease detection. Computers and Electronics in Agriculture 204, 107518.

[50] Faisal, M., Leu, J.S., Avian, C., Prakosa, S.W., K¨oppen, M., 2023. Dfnet: Dense fusion convolution neural network for plant leaf disease classification. Agronomy Journal.

[51] Ahmad, A., Saraswat, D., Gamal, A.E., Johal, G., 2021. Cd&s dataset: Handheld imagery dataset acquired under field conditions for corn disease identification and severity estimation. arXiv preprint arXiv:2110.12084 .

[52] Singh, D., Jain, N., Jain, P., Kayal, P., Kumawat, S., Batra, N., 2020. Plantdoc: A dataset for visual plant disease detection, in: Proceedings of the 7th ACM IKDD CoDS and 25th COMAD, pp. 249–253.

[53] Liu, Z., Mao, H., Wu, C.Y., Feichtenhofer, C., Darrell, T., Xie, S., 2022b. A convnet for the 2020s, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 11976–11986.

[54] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B., 2021. Swin transformer: Hierarchical vision transformer using shifted windows, in: Proceedings of the IEEE/CVF international conference on computer vision, pp. 10012–10022.

[55] Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D., 2017. Grad-cam: Visual explanations from deep networks via gradient-based localization, in: Proceedings of the IEEE international conference on computer vision, pp. 618– 626.

[56] Van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. Journal of machine learning research, 9(11).

[57] Ai, Y., Sun, C., Tie, J., Cai, X., 2020. Research on recognition model of crop diseases and insect pests based on deep learning in harsh environments. IEEE Access 8, 171686–171693.

[58] Wen-bin, S., Rong, W., Rong-hua, G., Qi-feng, L., Hua-rui, W., Lu, F., 2022. Crop disease recognition based on visible spectrum and improved attention module. Spectroscopy and Spectral Analysis 42, 1572–1580.

[59] Wang, X., Cao, W., 2023. Bit-plane and correlation spatial attention modules for plant disease classification. IEEE Access.

[60] Lin, J., Chen, X., Cai, J., Pan, R., Cernava, T., Migheli, Q., Zhang, X., Qin, Y., 2023a. Looking from shallow to deep: Hierarchical complementary networks for large scale pest identification. Computers and Electronics in Agriculture 214, 108342.

[61] He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778.

[62] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.C., 2018. Mobilenetv2: Inverted residuals and linear bottlenecks, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4510–4520.

[63] Chen, Y., Chen, X., Lin, J., Pan, R., Cao, T., Cai, J., Yu, D., Cernava, T., Zhang, X., 2022. Dfcanet: A novel lightweight convolutional neural network model for corn disease identification. Agriculture 12, 2047.

[64] Cai, J., Pan, R., Lin, J., Liu, J., Zhang, L., Wen, X., Chen, X., Zhang, X., 2023. Improved efficientnet for corn disease identification. Frontiers in Plant Science 14.

[65] Kang, R., Huang, J., Zhou, X., Ren, N., & Sun, S. (2024). Toward Real Scenery: A Lightweight Tomato Growth Inspection Algorithm for Leaf Disease Detection and Fruit Counting. Plant Phenomics, 6, 0174.

[66] Huang, Q., Wu, X., Wang, Q., Dong, X., Qin, Y., Wu, X., ... & Hao, G. (2023). Knowledge distillation facilitates the lightweight and efficient plant diseases detection model. Plant Phenomics, 5, 0062.