Research on Fine-grained Disease Identification Network of Crop Leaves Based on Local Feature Enhancement

Liwei Fan*[a], Siyan Liu[a], Kai Cai[a], Sixing Lu[a], Ke Zhang[a], Huixian Chen[a]
[a]School of Electronic Information and Control Engineering, Guangzhou University of Software, Guangzhou, China

ABSTRACT

This paper addresses the challenges of large intra-class variation, inter-class ambiguity, and background interference in fine-grained classification of crop leaf diseases. We propose Light_asf_net, a lightweight dual-branch network based on local feature enhancement. The global branch (ASF-former Transformer) captures leaf morphology, while the local branch (MBConv module) extracts lesion microstructures. An adaptive feature fusion (AFF) module enables collaborative local-global enhancement. Experiments on the Kaggle cassava leaf disease dataset (21,367 images, 5 categories) demonstrate a mean accuracy (mAcc) of 70.45%, representing a 37.8% improvement over ResNet-101, with only 6.03M parameters (14.2% of ResNet-101). On the rice leaf disease dataset, Light_asf_net achieves 100% accuracy within 12 epochs. Moreover, the model supports real-time inference at 35 FPS on mobile devices, providing an efficient solution for field deployment.

Keywords:local feature enhancement, lightweight dual-branch network, adaptive feature fusion, crop leaf disease identification, fine-grained classification.

1. INTRODUCTION

Early and accurate identification of crop diseases is a key link in ensuring food security. Traditional disease detection methods are limited by the subjectivity and inefficiency of manual experience, while existing deep learning models face three major challenges in fine-grained classification tasks.

1. Fine-grained classification of crop diseases faces three major challenges: (1) large intra-class variation, where the same disease shows diverse morphological appearances across conditions, (2) inter-class ambiguity, as diseases may only differ subtly in lesion texture or density, and (3) background interference from field environments such as soil or stems.

2. Inter-class ambiguity: There are only small differences in lesion density and edge texture between different diseases. (such as wheat stripe rust and leaf rust)

3. Background interference: The complex environment in the field (stems, soil) makes it difficult to locate the lesions

To address the above challenges, we propose a local and global perception dual-branch network called Light_asf_net for plant disease identification. Specifically, to extract local and global disease features, we developed a hybrid two-branch network based on ASF module and MobileNet. ASF-former Transformer was used to capture the overall morphological characteristics of the leaves, and the microstructural features of the lesions were extracted based on the MBConv module that can be separated and convoluted. In addition, we designed an adaptive feature fusion (AFF) module and a multi-level feature fusion module for local and global disease feature perception and multi-scale feature fusion, respectively. Using a lightweight component, a 16-channel stem layer (75% less parameters than ResNet-50), deep separable convolution (only 1/9 of the standard convolution), and global average pooling instead of a fully connected layer (1.2M parameters reduction). Light_asf_net SOTA performance was achieved on 2 plant disease datasets. The above experiments have shown:

Accuracy improved : Light_asf_net up to 70.45% mAcc, surpassed 37.80% of single-branch ResNet-101 networks and 48.77% of dual-branch ASF-former Transformer networks.

Optimized computing efficiency: The number of parameters is only 6.03M (26% of ResNet-101), and the inference speed reaches 35 FPS (Snapdragon 865 mobile platform).

Strong generalization: In the cross-crop test, the accuracy of rice leaf disease recognition is as high as 100%, and the CIFAR-10 is 91.98% mAcc.

The main contributions of this study are as follows:

- Dual-branch feature extraction architecture: global branches capture the overall shape of leaves, and local branches focus on lesion microstructures.
- A local-global feature synergy enhancement mechanism is designed to solve the problem of inter-class confusion in fine-grained classification.
- Achieve a balance between accuracy and computational efficiency, and provide technical support for real-time diagnosis in the field.

2. RELATED WORK

Deep learning models perform well in visual tasks, but the increasing demand for computing resources has prompted researchers to explore lightweight network architectures to adapt to scenarios such as mobile devices, and at the same time, combining local and global information and integrating the advantages of different feature extractors has become the key to performance improvement.

2.1 Local feature enhancement

The convolutional layer of traditional CNN uses the same convolutional core for each channel, and the channels are independent, which limits the expression ability of the model.[2] The SENet proposed by the Hu, J., Shen, L., and Sun, G et al.[3] introduces the "SE module", which is inserted into the convolutional layer, obtains the global information of the channel through compression operations, and then generates channel attention through stimulation operations, dynamically adjusts the channel weights, and enhances the model expression ability. It reduces the top-5 error rate to 2.251% in the ImageNet dataset, which is better than the 2.991% at that time, which verifies the effectiveness of channel attention. However, the "SE module" will increase the amount of computation and parameters, ignore the spatial dimension dependence, and have poor generalization ability on small data.

2.2 Dual-branch architecture

The research focuses on the CNN-Transformer hybrid architecture, which usually uses CNN to extract local features and then input them to the Transformer encoder for global semantic modeling.[4] For example, the LGNet proposed by Lin J et al.[5] is a CNN-ViT dual-branch network for plant disease identification, which integrates local and global features through the AFF module and multi-scale features with the help of the HMUFF module, which outperforms the single network and advanced models in AI Challenger 2018 and self-collected maize disease datasets. The ASF-former module proposed by Zixuan Su et al.[6] uses an adaptive fusion mechanism to calculate the weighted branch contribution (weighted summation of integrated features), and introduces jump connections to alleviate the disappearance of gradients when the weights are too small. The model achieves a top-1 accuracy of 83.9% on ImageNet-1K, and its performance is better than that of pure CNN, pure Transformer and other hybrid models at 12.9G MACs/56.7M parameters. However, this type of fusion architecture integrates two complex structures, resulting in an increase in the number of parameters and computational complexity, which limits its deployment in resource-constrained environments.

2.3 Lightweight networking with efficient model scaling

The demand for computing resources in deep learning models has driven the exploration of lightweight networks to adapt to scenarios such as mobile devices.

MobileNet-v1 achieves a top-1 accuracy of 70.6% on ImageNet, requiring only 569M FLOPs and 4.2M parameters, laying the foundation for lightweight networks.[7][8]

ShuffleNet-v1 achieves a top-1 accuracy of 67.4% at 140M FLOPs on ImageNet and 72.6% on 150M FLOPs, improving the inference efficiency of different hardware.[9][10]
Although these lightweight models achieve high efficiency, their trade-off between accuracy and model size often limits performance in agricultural fine-grained recognition tasks.

EfficientNet-B4 achieves a top-1 accuracy of 83.0% with 19.3G FLOPs and 17.5M parameters on ImageNet, but some models are not adaptable enough in high-resolution tasks or extremely lightweight scenarios, and there are problems of computational redundancy or compromise of feature expression capabilities. Some models rely on complex design, resulting in reduced readability, high-end versions have high computational intensity, making it difficult to deploy on low-end devices, and the adaptability to small data sets needs to be improved.[11]

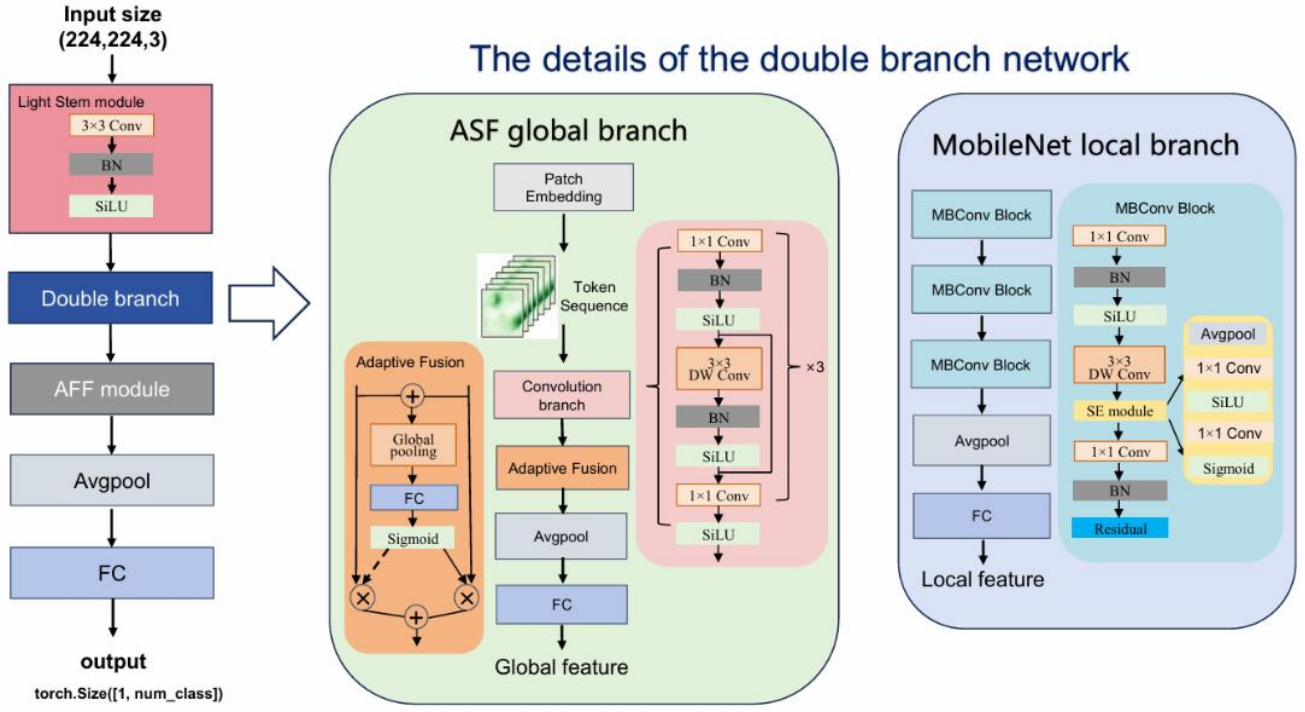

Figure 1. The architect of the light_asf_net framework and the details of the double branches network

3. METHODS

Our research route is shown in Figure 2:

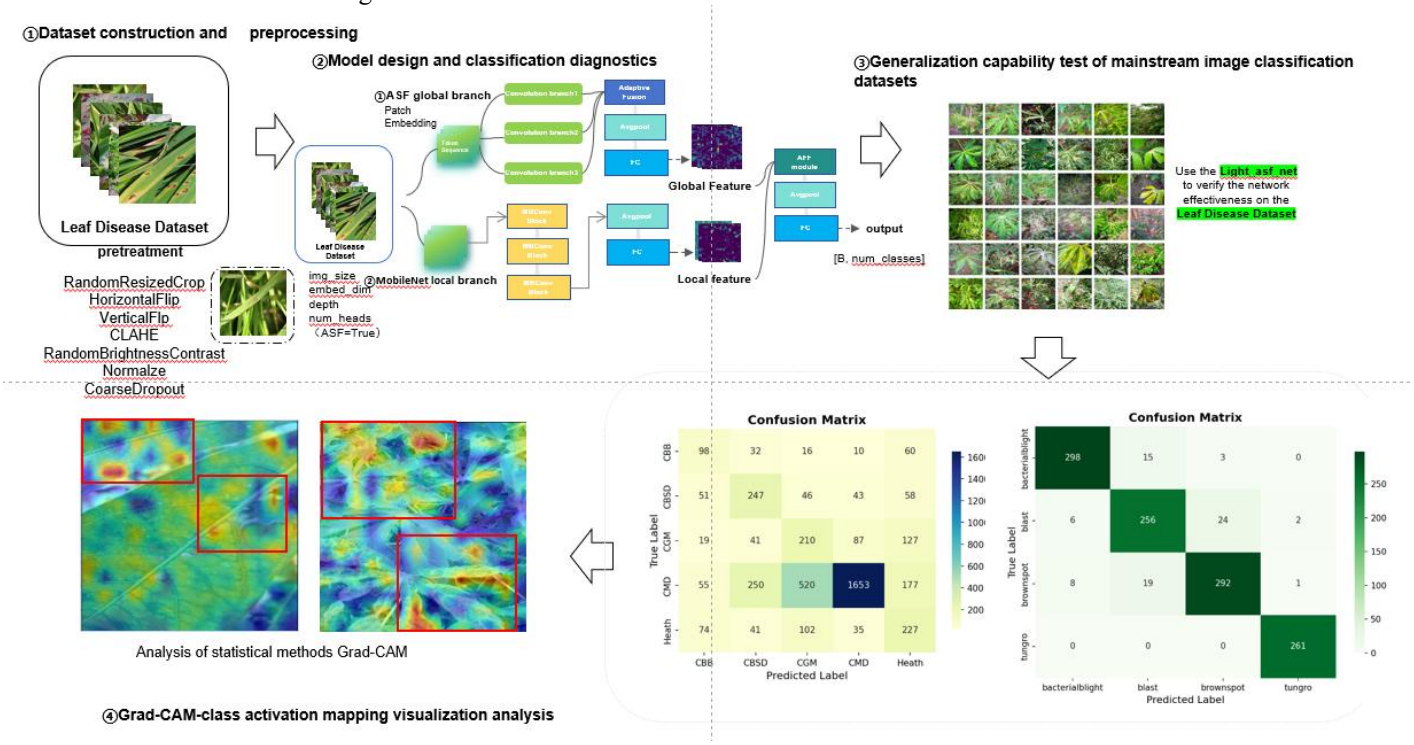

Figure 2. Our research pipeline figure

3.1 Dual-branch network architecture

Light_asf_net network adopts a two-branch structure (Figure 2).

3.1.1 Global branching

ASF-former Transformer was used to capture the overall morphological characteristics of the leaves. This ASF branch is a lightweight Transformer+ convolutional hybrid structure designed for efficient feature extraction. Although it uses ASF_former modules, compared with the original ASF_former$_P$, the light version greatly reduces the amount of parameters and computation, making it suitable for mobile terminals and resource-constrained scenarios. The input channel is reduced from 64 to 16, deep separable convolution is adopted, the embedding dimension is reduced to 256, the token dimension is reduced to 32, and the activation function is heavily used to replace ReLU.

3.1.2 Local branching

Extract the microstructural features of lesions based on the MBConv module based on deep separable convolution. Our MBConv module designed for local disease features adopts a large number of deep classifiable convolutions, greatly reduces the amount of computation and parameters on the basis of retaining spatial features and channel features, and introduces adaptive channel attention and residual connections. Then, the original information is preserved through residual connection to prevent feature degradation, and finally all spatial information is aggregated through global pooling to output the final feature.

3.2 Lightweight MBConv module

MBConv was first proposed by MobileNetV2 and has since been widely used in lightweight networks such as EfficientNet. Its core idea is inverted residual structure + deep separately convolution + SE attention.

Extension layer（1×1 Conv）：$C_{in} \times C_{mid}$

Depths can separate convolutions（3×3 Depthwise）：$C_{mid} \times 3 \times 3$

SE module（compression ratio r）：$C_{mid} \times (\frac{C_{mid}}{r}) + (\frac{C_{mid}}{r}) \times C_{mid}$

Projection layer（1×1 Conv）：$C_{mid} \times C_{out}$

Total parameter amount（Not included BN/SE）

$$Params\_MBConv = C_{in} \times C_{mid} + C_{mid} \times 3 \times 3 + C_{mid} \times C_{out} \quad (1)$$

3.3 Lightweight SE module

The module achieves channel attention by compressing the spatial information of each channel into 1 number (global average pooling), outputting the shape [B, C, 1, 1], then outputting the weight of each channel (range 0~1) through Excitation (channel relationship modeling), and finally recalibrating (scaling) to multiply each channel of the original input feature map by the corresponding weight.

$$Params\_SE = in\_channels \times (in\_channels//r) + (in\_channels//r) \times in\_channels = 2 \times in\_channels^2/r \quad (2)$$

3.4 AFF adaptive feature fusion

The AFF Module is one of the core innovations of this network architecture, which cleverly solves the problem of feature fusion in dual-branch networks. It dynamically adjusts the weights for local detail features such as textures and edges extracted from the MBConv branch and long-distance dependencies and global semantic information captured from the ASF-former branch. Only one linear layer (in_features*2→2) is used to generate weights. The weights are initialized to zero first, and the simple average two-branch weights are gradually learned at the beginning of training, and then gradually learn the optimal fusion strategy, dynamically adjust the importance of local/global features for different samples, and realize feature complementarity while maintaining lightweight, so that the model can adaptively adjust the weights of different features. Figure 2 shows the structure of the AFF module, which learns the adaptive weights of two features and then weights the obtained weights on the original feature map, so that the model can adaptively perceive local and global disease features, thereby enhancing the model's feature representation ability.

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

$$Y = Concat\{X_L; X_G\} \quad (3)$$

Thereinto. We then map the output number of channels to 2 through a convolution operation and use the softmax function to obtain the weighted feature plot W. This can be expressed as:$X_G \in \mathbb{R}^{2C \times W \times H}$

$$W = Softmax(f_{3\times3}(Y)) \quad (4)$$

where represents a 3x3 convolutional operation. Subsequently, we separate the weight feature graph W along the channel direction to obtain the local weight W and the global weight. Finally, we apply the obtained weights to the original feature map and perform element-level addition operations to obtain the adaptive output A:$f_{3\times3}W_G$

$$A = X_L * W_L + X_G * W_G \tag{5}$$

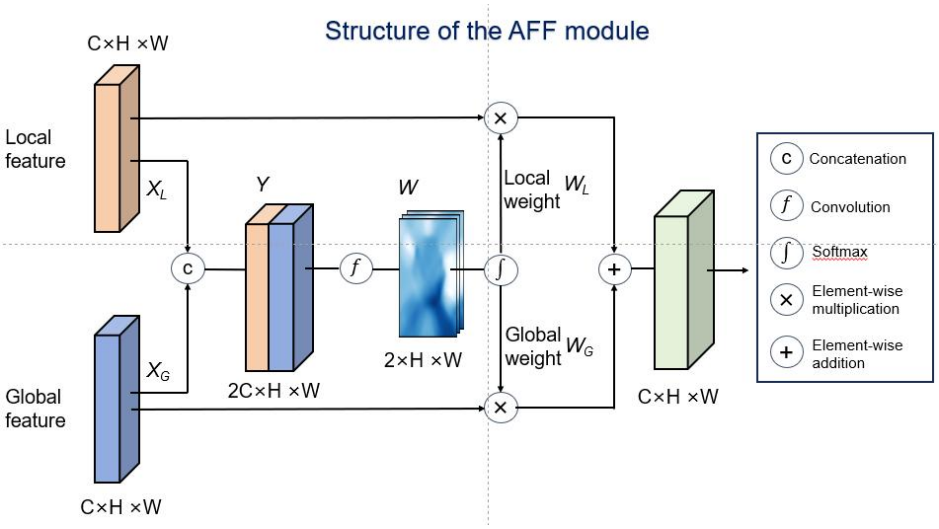

Figure 3. The Structure of the AFF module

4. EXPERIMENT

4.1 Datasets and settings

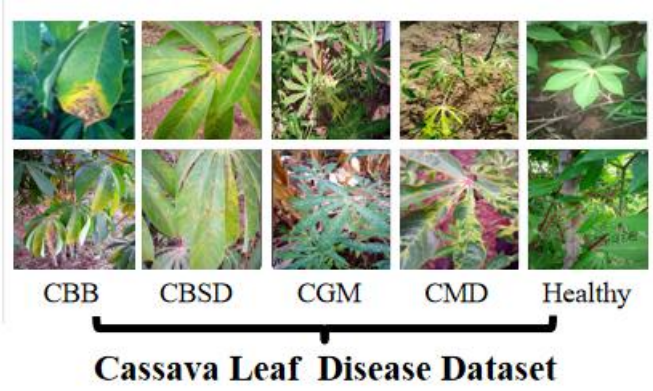

Figure 4. Cassava Leaf Diease Dateset Case

Table I Phenotypic information summary of the participants from the Cassava Leaf Diease Dateset.

Class	***Total number***
Cassava Bacterial Blight (CBB)	*1087*
Cassava Brown Streak Disease (CBSD)	*2189*
Cassava Green Mottle (CGM)	*2386*
Cassava Mosaic Disease (CMD)	*13158*
Healthy	*2577*

Kaggle cassava leaf dataset: Current research focuses on the identification of fine-grained diseases in crops, especially using the Kaggle cassava leaf disease classification dataset. Data volume: 21,367 images, covering 5 categories (4 diseases + healthy leaves) Features: Contains real field images taken by crowdsourced farmers, annotated by agricultural experts, and there are label noise and category imbalances. The summary information of this dataset is shown in Table I.

Indicators: mAcc, pre, recall, Spec, Parameter quantity

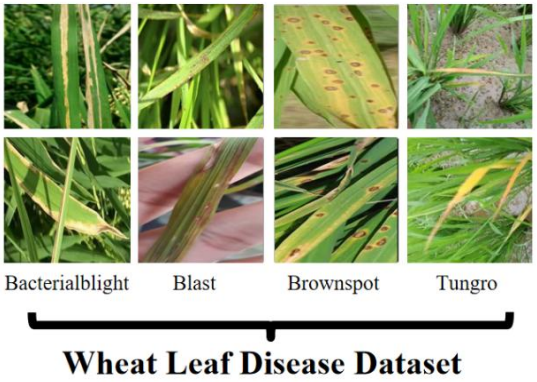

Figure 5. Wheat Leaf Diease Dataset case

Table II Phenotypic information summary of the participants from the Wheat Leaf Diease Dataset.

Class	*Total number*
Rice Bacterialblight (Bacterialblight)	*1584*
Rice Blast (Blast)	*1440*
Rice BrownSpot(Brownspot)	*1600*
Rice Tungro(Tungro)	*1308*

Rice leaf disease image dataset: This dataset contains 5932 digital images, including four rice leaf diseases: bacterial blight, rice blast, brown spot, and leaf disease. Cited from: Sethy, P. K., Barpanda, N. K., Rath, A. K., & Behera, S. K. (2020). A support vector machine was used for depth characteristic-based disease identification of rice leaves. The summary information of this dataset is shown in Table II.

Indicators: mAcc, parameter quantity, convergence rounds

4.2 Data preprocessing

The training set uses rich data enhancement, which mainly includes: random cropping and scaling: random cropping and scaling to 224×224 to enhance the robustness of the model to scale and position. Random horizontal flip, random vertical flip, adaptive histogram equalization, enhanced image contrast, improved lighting influence, random brightness contrast adjustment, normalization, normalization of pixels using ImageNet-style mean and variance, random occlusion, random occlusion of part of the area according to 50% probability, improve model robustness, etc.

The validation set only does the most basic preprocessing and does not do random enhancement to ensure the fairness of the evaluation: the training set randomly scrambles the original data as actual training data to improve the generalization ability and training efficiency.

4.3 Experiment Settings

In both of our experiments, we divided into training and validation sets according to an 8:2 ratio and used resnet101, googlenet, DenseNet-121, mobilenet_v3, shufflenet_v2_x1_0 , efficientnet_b0, and other networks for training and verification. In the cassava leaf dataset, the training set and the validation set are divided according to various proportions, and there is also an imbalance in the samples. To enhance the learning accuracy and ability of the network, we use methods such as increasing Gaussian noise and random occlusion. In order to uniformly evaluate the effect of each model, the main parameters of our experiment are as follows: the initial learning rate is 0.001, and the learning process is divided into [10, 20, 30] three decreases, and the AdaW optimizer and cosine annealing learning rate scheduler are used, and the BATCH_SIZE is 8EPOCH is 50. Considering the imbalance of the sample, we mainly evaluate the effect of the model by means of accuracy, precision, sensitivity, specificity and other indicators.

4.4 Loss function

In order to solve the problem of sample imbalance in the cassava leaf dataset, we use the FocalLoss loss function, use the reciprocal of the sample number as the weight, and normalize it. It is improved on the basis of cross entropy loss, which can make the model pay more attention to hard-to-classify samples (i.e., samples that are easy to be misclassified) and pay less attention to easy-to-classify samples. The specific formula for calculating FocalLoss is as follows:

$$FocalLoss = -\alpha(1-p_t\quad)^{\gamma} log(p_t\quad) \tag{6}$$

Thereinto:

- $p_t$is the model's predicted probability of the real category (i.e., the probability of the true label corresponding to the soft max).
- The α is the category balance factor (which can adjust the weight of different categories and solve the problem of category imbalance).
- γ is the focusing parameter, which adjusts the focus on difficult-to-classify samples, usually by 2.
- $(1-p_t\quad)^{\gamma}$It is a regulator, the smaller it is (that is, the more difficult it is to distinguish pairs), the larger the factor, the greater the loss; The larger it is (i.e., the easier it is to pair), the smaller the factor and the smaller the loss.$p_t p_t$

4.5 Comparison experiment

Table III Comparsion with different methods on Cassava Leaf Disease Dataset

Model	mAcc	Precision	Recall	Specificity	Params(M)
light_asf_net (ours)	0.7045	0.7953	0.7953	0.9488	6.03
resnet101	0.3265	0.1762	0.1762	0.7941	42.51
googlenet	0.3955	0.6611	0.6611	0.9153	5.61
DenseNet-121	0.3634	0.2384	0.2384	0.8096	6.96
mobilenet_v3	0.3305	0.1559	0.1559	0.789	1.52
shufflenet_v2	0.3382	0.1545	0.1545	0.7886	1.26
efficientnet_b0	0.3409	0.1554	0.1554	0.7889	4.01

Table III shows that the light_asf_net network has the highest accuracy and precision in both balanced and non-equilibrium cases, which demonstrates the effectiveness of the network in crop disease identification. For most of the classical networks on the cassava leaf dataset, the performance of various diseases on cassava leaves is very similar, and the above networks do not work on fine-grained recognition, and the ability of single-branch networks to distinguish background interference and local features at the same time is limited, and it can also be found that the general network can only distinguish 3-4 types of diseases and is affected by the imbalance of sample distribution, and our light_asf_net The network solves such difficult fine-grained recognition tasks well through local + global feature modeling and adaptive fusion weight adjustment.

Table IV Transferability to CIFAR-10/100.

Dataset type	Model	*mAcc*	*Precision*	*Recall*	*Specificity*
CIFAR-10	**light_asf_net**	**0.9198**	**0.9198**	**0.9198**	**0.9911**
	ASF_former_S	0.7839	0.7839	0.7839	0.7839
	asf_former_b	0.1895	0.1895	0.1895	0.9099
CIFAR-100	**light_asf_net**	**0.6859**	**0.6859**	**0.6859**	**0.9968**
	ASF_former_S	0.3548	0.3548	0.3548	0.9934
	asf_former_b	0.2895	0.2895	0.2895	0.7632

Table IV shows that this light_asf_net still achieves a very high accuracy on the CIFAR-10/100 dataset, which has a strong generalization ability compared with other ASF series networks, and is not only for the identification of the above two crop leaf diseases, but also for the application of the network to generalized image classification.

Table V Different network results on the Rice Leaf Disease Dataset.

Model	mAcc	Convergence epochs	Params(M)
light_asf_net (ours)	1	12	6.03
resnet101	99.34%	42	42.51
googlenet	98.41%	17	5.61
DenseNet-121	99.91%	39	6.96
mobilenet_v3	1	42	1.52
shufflenet_v2	1	37	1.26
efficientnet_b0	1	37	4.01

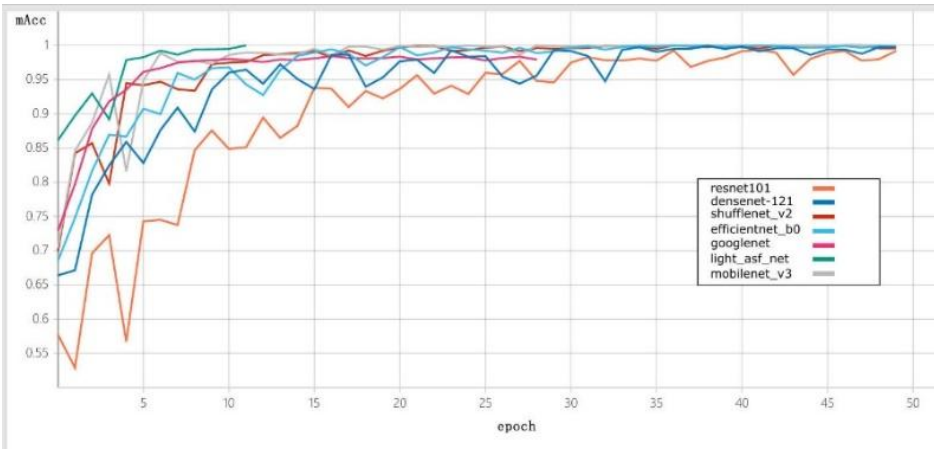

Figure 6. The mAcc and epoch result about these model on Rice Leaf Disease Images.

Through the above Figure 6, it can be found that most of the networks have extremely high accuracy on rice datasets, and some networks can reach 100% mAcc, but there are differences in the learning efficiency and convergence batch of each network, while the light_asf_net network has the fastest convergence speed, reaching 100% mAcc in only 12 rounds of training, indicating that the learning ability of this network is beyond that of ordinary networks.

Table VI Compare with the different part of model ON Cassava Leaf Dataset.

Model components	mAcc	generalization Acc	Params(M)
asf_former_s	21.68%	36.67%	18.86M
hmcb_former_s	24.08%	19.33%	22.26M
asf_former_b	31.94%	38.67%	56.12M
efficientnet_b0	34.09%	37.33%	4.01M
light_asf_net	70.45%	58.67%	6.03M

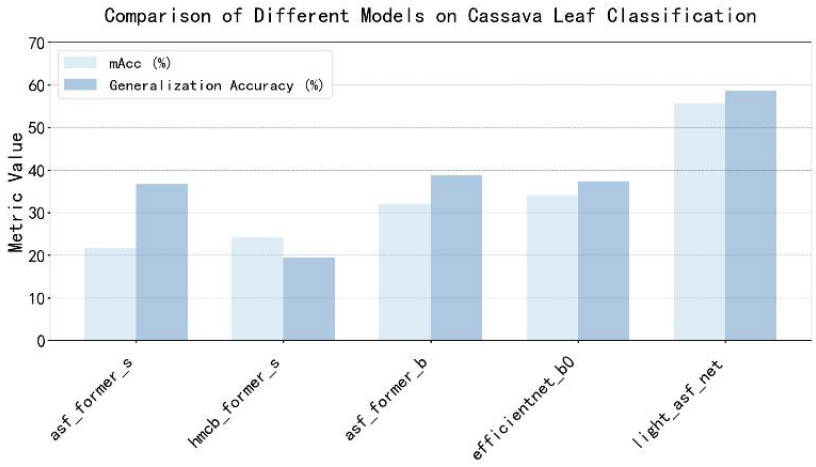

Figure 7. Visualization of the result of the mAcc and generalization Accuracy in different model.

According to the above experimental parameters, we performed ablation experiments with ASF series networks and effcientnet networks based on cassava leaf 5 classification according to the two-branch architecture, and the results are shown in Figure 7. The generalization accuracy of mAcc is to randomly select 30 images and a total of 150 images from the original dataset to test the generalization ability of the model.

According to Table VI, we can find that the mAcc and generalization accuracy light_asf_net on the cassava leaf 5 classification dataset is due to the asf series network and efficientnetInternet. This lightweight ASF network achieves the improvement of network accuracy and generalization while greatly reducing the number of ASF network parameters, and provides a solution for the efficient identification and treatment of crop diseases in subsequent deployment on edge devices.

4.6 Visualization analysis

In this section, we use Grad-CAM to visualize the class activation mapping of some samples to demonstrate regions of interest for different models.[12] As shown in Figure 8. We selected two samples with local disease characteristics and global disease characteristics from the Kaggle cassava leaf dataset and the rice leaf disease dataset, respectively. It can be seen that ConvNeXt-Tiny focuses on inaccurate or incomplete lesion regions and has weak lesion feature perception in complex scenarios. While Swin Transformer-Tiny can locate lesion areas, it also focuses on a lot of redundant information. In contrast, our proposed light_asf_net not only focuses on global disease signatures, but also accurately captures local features while suppressing complex background information.

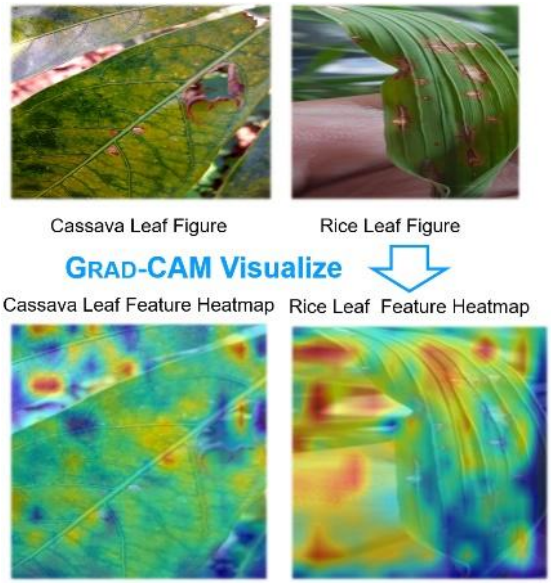

Figure 8. The Grad-CAM Visualize the cassava leaf and the rice leaf

5. CONCLUSION AND FUTURE WORK

5.1 Conclusion

This study proposed Light_asf_net, a lightweight dual-branch network for fine-grained crop disease identification. By integrating a global branch (ASF-former Transformer) for morphological analysis and a local branch (MBConv-based CNN) for lesion microstructure extraction, together with an Adaptive Feature Fusion (AFF) module, the model achieves synergistic local-global feature enhancement. The experimental results confirm that the approach effectively addresses intra-class variation, inter-class similarity, and background interference, while maintaining computational efficiency suitable for edge deployment.

5.2 Future Work

To advance fine-grained crop disease identification, future research should focus on scalability, robustness, and integration with agricultural ecosystems. The following directions are proposed:

5.2.1 Model Optimization for Edge Deployment

Hardware-Aware Compression: Integrate quantization (FP16/INT8) and pruning techniques to reduce model size below 3M parameters, targeting ultra-low-power devices like plant protection drones or IoT sensors. Explore dynamic inference mechanisms that skip redundant branches (e.g., bypassing local feature extraction for healthy leaves), reducing average latency by 20–40%.Cross-Platform Compatibility: Optimize the model for heterogeneous hardware (e.g., NVIDIA Jetson, ARM Cortex-M) using TensorRT Lite or ONNX Runtime, ensuring consistent performance across farm-edge devices.[13]

5.2.2 Generalization Enhancement

Unsupervised Domain Adaptation (UDA): Leverage contrastive learning or adversarial training to bridge domain gaps between lab-collected and field images, minimizing annotation dependency for unseen crops or environmental conditions.

Multi-Crop Harmonization: Develop a unified framework for disease identification across crops (e.g., cassava, rice, wheat) using meta-learning or modular neural networks, enabling knowledge transfer while preserving task-specific accuracy.

5.2.3 Advanced Feature Fusion and Explainability

Hierarchical AFF Modules: Design multi-stage AFF blocks that progressively fuse features across scales (e.g., lesion-level → leaf-level → canopy-level), enhancing context-aware feature integration.Uncertainty Calibration: Integrate Bayesian neural networks to quantify prediction confidence, critical for high-stakes agricultural decisions (e.g., pesticide application).Interactive Visualization Tools: Extend Grad-CAM to generate actionable insights for farmers, such as lesion severity maps and treatment recommendations, improving usability and trust.

5.2.4 Ecological Integration

Edge-Cloud Collaborative Systems: Deploy Light_asf_net on edge devices for real-time inference, with cloud-based retraining using federated learning to preserve data privacy across farms.Multimodal Sensor Fusion: Incorporate non-visual data (e.g., hyperspectral imaging, soil moisture) to enhance diagnostic accuracy under occluded field conditions.Global Crop Disease Atlas: Collaborate with agricultural institutions to build a federated database for model fine-tuning across geographical regions, addressing climate-specific disease patterns.

6. ACKNOWLEDGMENT

This research is partially supported by Guangzhou University of Software 2025 Research Project(Grant No. KY202523), Guangzhou University of Software 2025 "Quality Engineering" Construction Project (Grant No. JYS202502), 2024 Guangdong Provincial Teaching Quality and Teaching Reform Project (Grant No. 000320000104).

REFERENCES

[1] Lilhore, U. K., Imoize, A. L., Lee, C.-C., et al., "Enhanced convolutional neural network model for cassava leaf disease identification and classification," Mathematics 10(4), 580 (2022). https://www.mdpi.com/2227-7390/10/4/580

[2] Li, M. and Yao, H., "FMVP: Fine-grained meta visual prompt enabled domain-specific few-shot classification," Neurocomputing 633, 129688 (2025). https://www.sciencedirect.com/science/article/abs/pii/S0925231225003601

[3] Hu, J., Shen, L., and Sun, G., "Squeeze-and-excitation networks," Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 7132–7141 (2018). https://ieeexplore.ieee.org/document/8578843

[4] Dosovitskiy, A., Beyer, L., Kolesnikov, A., et al., "An image is worth 16x16 words: Transformers for image recognition at scale," arXiv:2010.11929 (2020). https://arxiv.org/abs/2010.11929

[5] Lin, J., Zhang, X., Qin, Y., et al., "Local and global feature-aware dual-branch networks for plant disease recognition," Plant Phenomics 6, 0208 (2024). https://spj.science.org/doi/10.34133/plantphenomics.0208

[6] Su, Z., Chen, J., Pang, L., et al., "Adaptive split-fusion transformer," Proc. IEEE Int. Conf. Multimed. Expo (ICME), 1169–1174 (2023). https://arxiv.org/abs/2204.12196

[7] Hamilton, W. L., Ying, R., and Leskovec, J., "Inductive representation learning on large graphs," arXiv:1706.02216 (2017). https://arxiv.org/abs/1706.02216

[8] Jahin, Md A., Shahriar, S., Mridha, M. F., et al., "Soybean disease detection via interpretable hybrid CNN-GNN: Integrating MobileNetV2 and GraphSAGE with cross-modal attention," arXiv:2503.01284 (2025). https://www.arxiv.org/pdf/2503.01284

[9] Zhang, X., Zhou, X., Lin, M., et al., "ShuffleNet: An extremely efficient convolutional neural network for mobile devices," Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 6848–6856 (2018). https://ieeexplore.ieee.org/document/8578814

[10] Ma, N., Zhang, X., Zheng, H.-T., et al., "ShuffleNet V2: Practical guidelines for efficient CNN architecture design," arXiv:1807.11164 (2018). https://arxiv.org/abs/1807.11164

[11] Tan, M. and Le, Q. V., "EfficientNet: Rethinking model scaling for convolutional neural networks," Proc. Int. Conf. Mach. Learn. (ICML) (2019). https://arxiv.org/abs/1905.11946

[12] Sandler, M., Howard, A., Zhu, M., et al., "MobileNetV2: Inverted residuals and linear bottlenecks," arXiv:1801.04381 (2018). https://arxiv.org/abs/1801.04381

[13] Howard, A. G., Zhu, M., Chen, B., et al., "MobileNets: Efficient convolutional neural networks for mobile vision applications," arXiv:1704.04861 (2017). https://arxiv.org/abs/1704.04861