Journal of Research in Science and Engineering (JRSE)
ISSN: 1656-1996 Volume-7, Issue-1, January 2025
DOI: 10.53469/jrse.2025.07(01).18

MAU-Net: Full-period Mango Leaf Disease Image Segmentation Algorithm Based on an Improved UNet Network

Wenkang Tang[1], Bibo Lu[2,*], Peipei Zhou[3], Jie Yang[4], Aiqing Song[5]

[1,2,3]School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo, Henan, China
[4,5]Jiaozuo Agricultural and Forestry Research Institute, Jiaozuo, Henan, China
[1]212209020058@home.hpu.edu.cn, [2]lubibo@hpu.edu.cn, [3]212209020056@home.hpu.edu.cn, [4]13069442244@163.com, [5]jzny@126.com
**Correspondence Author*

Abstract: ***Mango leaf disease segmentation is an essential foundation for accurate disease diagnosis and intelligent grading. The size and shape of mango leaf diseases vary significantly at different times, making it difficult for mainstream semantic segmentation methods to segment disease areas accurately. Therefore, this paper proposes a method called MAU-Net for fine segmentation of mango leaf diseases over the whole period. The MAU-Net is based on the traditional Unet architecture, integrates the Self-Aligning Attention Feature Fusion (SAFF) module and the Multiscale Feature Enhancement (MFE) module, and designs a new loss function DF_Loss. Specifically, the designed SAFF module changes the traditional Unet's skip-connection approach by fusing the global and local two-branch attention mechanisms. It enhances the attention to crucial leaf and disease features at different levels and thus retains richer semantic information about mango leaf diseases. The designed MFE module aims to solve the problem of complex multi-scale disease segmentation in different periods of mango leaves by introducing different scales of cavity convolution to enhance the extraction of disease features at different scales. The designed DF_Loss combines the idea of the similarity measure in Dice Loss and the advantages of the attentional conditioning mechanism in Focal Loss with an additional conditioning factor. It allows the model to focus more on pixels that are difficult to categorize during the learning process, thus improving the segmentation accuracy. MAU-Net achieved 99.21%, 84.33%, 97.1%, and 96.94% of leaf IoU, disease IoU, F1, and mPA metrics on the mango leaf disease dataset. It improved 0.36%, 4.88%, 3.9%, and 1.91% over UNet, and 5.59%, 0.19%, 1.6%, and 2.26% over DeepLabv3+, respectively. Therefore, the present study may provide an accurate method for segmenting mango leaf spots over the whole period and provide a sufficient basis for the accurate analysis of mango leaf diseases.***

Keywords: Full-period mango leaf disease, MAU-Net, Leaf Segmentation, Disease Segmentation**.**

1. Introduction

Mango is a plant of the genus named Sucky fruit in the Lacertidae family, an important cash crop in the tropics and subtropics. It is famous for its sweet flesh and high nutritional value and is known as the “king of tropical fruits” [1]. Mangoes are now a significant source of income for tropical countries in Asia, Africa, Central America, and the Caribbean [2]. According to the Food and Agriculture Organization of the United Nations (FAO), China is currently the world’s second-largest producer of mangoes, second only to India. The annual production is about 4,351,593 tons, accounting for 11.2% of the world’s mango production [3]. Mango, a tropical fruit tree, is not only delicious in terms of its fruits for direct consumption, but its leaves also harbor a wealth of medicinal value. The ethanol extract of mango leaves has analgesic, anti-inflammatory, and antibacterial properties [4]. However, diseases such as mango leaf spot and anthracnose often threaten its growth process. These diseases can seriously affect the health of mango trees, leading to yellowing of the leaves and, in severe cases, even a total loss of leaves [5]. Failure to recognize and manage the disease effectively and promptly can trigger the widespread spread of the disease. Mango yield and quality may suffer a significant blow, thus causing severe economic losses to the fruit growers [6]. However, traditional spraying methods often fail to accurately consider the actual severity of the disease. This practice may lead to the underuse of pesticides in some areas and ineffective disease control. Inaccurate application of agrochemicals may pollute the environment and prevent effective disease prevention and treatment [7][8]. In the practicalities of agricultural production, growers usually identify the type of spot and determine the severity of the disease manually. However, this manual discrimination method is not only a considerable workload and time-consuming but also susceptible to subjective perceptual bias, leading to erroneous diagnosis [9]. Computer vision technology plays a crucial role in pest and disease identification. Through methods such as image processing and semantic segmentation, growers are empowered to accurately assess disease severity and guide the rational use of agrochemicals. This improves crop yields and reduces environmental loads.

In recent years, with the rapid development of computer vision technology, the research on crop disease recognition through computer vision has attracted much attention. Merchant et al. [10] utilized digital image processing techniques to extract multiple dimensional features such as RGB values and leaf texture of mango leaves. These feature data were then fed into an unsupervised machine-learning model that was analyzed by clustering to separate and identify different disease features. Eko Prasetyo et al. [11] used a color space transformation strategy to convert the image from RGB color space to HSV color space and combined the luminance dimensions, blue and red, to achieve a detailed segmentation of mango leaf disease using a thresholding method. In addition, Gina S. Tumang [12] introduced a combined multiple SVM and GLCM image processing method for identifying specific diseases such as anthracnose, fruit moth, and sooty mold. The method successfully located and

segmented the disease areas by analyzing the statistical properties of the images, such as contrast, kurtosis, skewness, and entropy. Researchers have proposed a variety of methods for mango leaf and disease segmentation based on traditional image processing techniques. However, these methods often involve complex preprocessing steps, which limits their ability to be applied in real-world scenarios. In addition, since these algorithms are usually designed for specific types of diseases, they often need more flexibility and generalization ability to adapt to multiple disease scenarios when dealing with diverse lesion features.

With the rapid advancement of deep learning technology, it exhibits high accuracy and powerful migration capabilities. It has attracted an increasing number of research scholars into mango disease segmentation. It achieves significant progress in mango leaf disease segmentation and recognition. D. Lita Pansy et al. [13] used Logistic U-net to segment mango leaves. Then, the leaves were subjected to MD-FCM clustering to cluster diseased leaves and pests separately. Finally, LFD-BOA was utilized to retrieve pest and disease characteristics for further classification. However, they only segmented the leaves of mangoes and did not achieve accurate segmentation of the pests. Saleem et al. [14] segmented the diseased portion by considering the vein pattern of the leaf. This method of leaf vein segmentation segments the vein pattern of the leaf. Then, the features are extracted and fused using a typical correlation analysis (CCA) based method. The segmentation results work for the following classification step and are unacceptable for segmenting the diseased part. Consideration of early-stage diseases of mango leaves needs to be included. Chouhan et al. [15] used the Radial Basis Function Neural Network (RBFNN) to extract lesion regions from mango leaf images and achieved good extraction results for anthracnose in mango leaves. Vinay Gautam et al. [16] by segmenting the diseases into regions of interest and feeding them into a stack of various deep neural networks. Then, machine learning was used to recognize various mango foliar diseases, such as powdery mildew and anthracnose. Pham et al. [17] used an artificial neural network (ANN) approach to detect early-stage diseases on plant leaves with small spots that can only be detected with higher-resolution images. After a preprocessing step using a contrast enhancement method, the infected spots were segmented, and the results of the ANN outperformed those of the CNN using a simpler network structure. The algorithm is effective, though, for early disease of mango leaves. However, it still lacks in segmentation results for many spots.

Significant progress has been made with these methods in the mango leaf disease image segmentation research field. Satisfactory segmentation results can be achieved with distinctive features for intermediate to advanced disease stages. However, in the early mango leaf disease stage, the above models encountered challenges in segmenting such early, small-targeted diseases due to tiny and unrecognizable lesion features. In addition, timely identification of early-stage diseases is critical for effectively controlling disease spread, rapidly implementing control measures, and ensuring mango yield and quality. To solve the above problems, we propose a full-period mango leaf disease segmentation algorithm named MAU-Net in this paper, which can detect early minor target diseases and various mid- and late-stage diseases. The main contributions of this paper are as follows:

1) To solve the problem of early tiny lesion feature information loss in deep convolutional networks. In this study, the SAFF attention mechanism is designed to fuse global and local attention to improve the jump connection of Unet. SAFF automatically learns the attention coefficients of different levels of features and strengthens the weights of crucial mango leaf disease features. Thus, it realizes delicate feature fusion and effectively prevents the loss of small disease information.

2) Because the characteristics of mango leaf diseases in different periods exhibit significant scale differences, the MFE module was designed in this study. The MFE aims to capture the characteristics of full-cycle diseases more effectively. The MFE module can capture the leaf disease information of each period in detail by concurrently performing multiple cavity convolution operations at different scales. This multi-scale analysis strategy significantly improves the model's accuracy in segmenting the disease across different periods.

3) Aiming at the problem of unbalanced mango leaf disease samples, this paper proposes a loss function DF_Loss based on dynamic weight adjustment. Compared with the traditional cross-entropy loss function, DF_Loss can dynamically adjust its weight in model training according to the different sample categories. By reducing the weights of easy-to-recognize samples, the model pays more attention to pixels that are difficult to classify during the learning process, effectively improving the segmentation accuracy of leaf and disease regions.

2. Material and Methods

2.1 Dataset

The dataset used in this study was obtained from Shri Mata Vaishno Devi University [18]. As shown in Figure 1, this dataset covers examples of several types of mango leaf diseases, including healthy leaves, rust, leaf spot, and anthracnose. In order to improve the accuracy of model training, we used Labelme software to finely label the diseases in these images under the guidance of experts in agriculture. The corresponding labeled files were created by carefully outlining the edges of each leaf and the diseased areas. The labeled images' pixel values representing the background, leaf, and disease are explicitly assigned as 0, 1, and 2, respectively. The labeling process is shown in Figure 2.

This study uses data enhancement and image preprocessing techniques to improve the model's performance, robustness, adaptability, and accuracy in a variable test environment. The dataset was extended to 1384 images through data enhancement operations. In order to facilitate training and evaluation, the dataset is divided into training, testing, and validation sets in the ratio of 7:2:1 in this study to enhance the generalization ability of the model in different data instances. The specific division of the dataset is shown in Table 1.

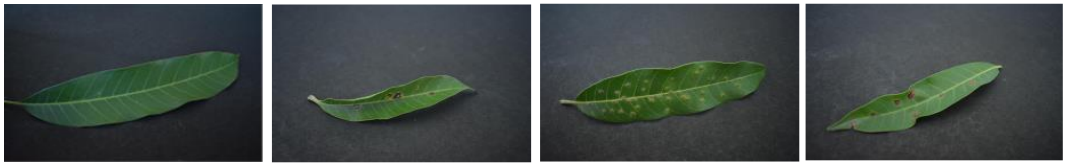


Figure 1: Examples of partial images from a plant leaf dataset

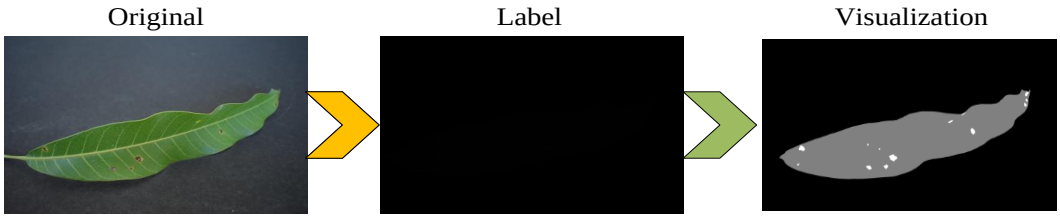


Figure 2: Labeling flowchart. (a)Original. (b)Label. (c) Visualization

Table 1: Statistics on the number and size of data sets

	Original image	*Expanded*	*Training*	*Test*	*Validation*
Quantity	173	1384	969	277	138
Size	6000×4000	1200×800	1200×800	1200×800	1200×800

2.2 U-Net

During the segmentation of mango leaf diseases, leaves, and diseases must be finely segmented. Annotation of mango leaf disease images is difficult and time-consuming with the presence of a small number of samples. UNet can get accurate segmentation even on small sample datasets. Olaf Ronneberger et al. [19] proposed the UNet network model in 2015. It is a deep-learning model designed for image segmentation tasks. As shown in Figure 3, UNet achieves accurate segmentation of images by fusing global contextual information with local detail information. The model uses an encoderdecoder architecture, which makes it perform well in the field of image segmentation.

UNet is a well-designed image segmentation network that skillfully fuses global contextual information with local detailed features through its encoder-decoder structure and hopping connection mechanism. However, the UNet model also has some things that could be improved. Its jump-connection mechanism may lead to too much difference in feature information between different layers. It will lead to the problem of mismatch between the extracted features and the actual features during feature fusion. In addition, the single convolutional structure of UNet shows some limitations in extracting multi-scale target features. During the training process, the UNet algorithm uses the cross-entropy loss function to measure the difference between the generated segmentation results and the actual labels. The cross-entropy loss function can effectively measure the classification error between different pixel classes, thus accurately guiding the model optimization. Due to the effectiveness of this loss function, UNet can produce accurate segmentation results even on small sample datasets. However, the cross-entropy loss function also has shortcomings. For the case of significant differences in target features and large differences in the amount of data between different targets, the cross-entropy loss function needs to learn better for features with small amounts of data. To address the above problems, this study makes targeted improvements and proposes a novel MAU-Net network architecture.

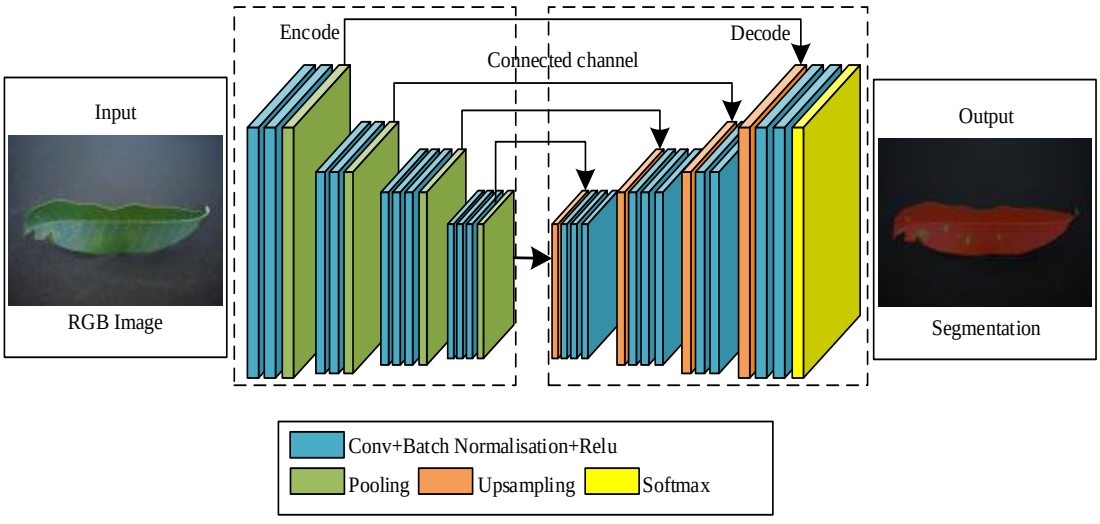


Figure 3: The network structure of Unet

2.3 Design for MAU-Net

To overcome the limitations of the traditional UNet model in capturing multi-scale disease features and fusion of features in different levels, a novel MAU-Net network architecture is proposed in this study. As shown in Figure 4, the SAFF module assigns attentional weights to feature information by combining global and local attentional mechanisms, and these weights are subsequently applied to the fusion process of the feature map. SAFF significantly enhances the network's ability to focus on disease-related features at different levels, facilitating effective information transfer and feature fusion. In order to enhance the model's ability to extract features of objects at different scales, we have specially designed the MFE module. The module focuses on capturing rich multi-scale contextual information and effectively improves the model's overall performance through multi-scale feature enhancement. To address the problem of the number of disease samples in the mango leaf disease dataset being much less than the number of healthy leaf samples, we designed the DF_Loss function in the training process. DF_Loss aims to solve the sample imbalance problem to enhance the learning weights for disease features. By allowing the model to focus more on disease categories that are difficult to recognize, DF_Loss helps the model to adapt better and learn the disease characteristics of mango leaves.

Specifically, the image size of the input network is 512×512×3. The downsampling operation through a series of 3×3 and 1×1 convolutions in the encoder stage yields the feature maps of the five layers E1, E2, E3, E4, and E5, respectively. These five feature maps are 512 × 512 × 64, 256 × 256 × 128, 128 × 128 × 256, 64 × 64 × 512, and 32 × 32 × 512 in size. In the decoder stage, the feature maps of the four layers E1, E2, E3, and E4 are fed into the MFE module for a multi-scale feature enhancement to obtain E1', E2', E3', and E4' respectively. The learning of multi-scale features by the network is enhanced by augmenting the semantic information of different scale features in each level. Then, E5 is up-sampled by a factor of 2 and then input with E4' into the SAFF module for feature fusion to obtain D4 (64×64×512). The SAFF module is used to reduce the difference of semantic information in different layers to prevent the loss of target feature information that cannot be easily recognized. By performing the above decoding operation with different layers of feature maps, the final feature map size of 512×512×64 is obtained as D1. Finally, the acceptable segmentation result is produced by two depth separable convolutions.

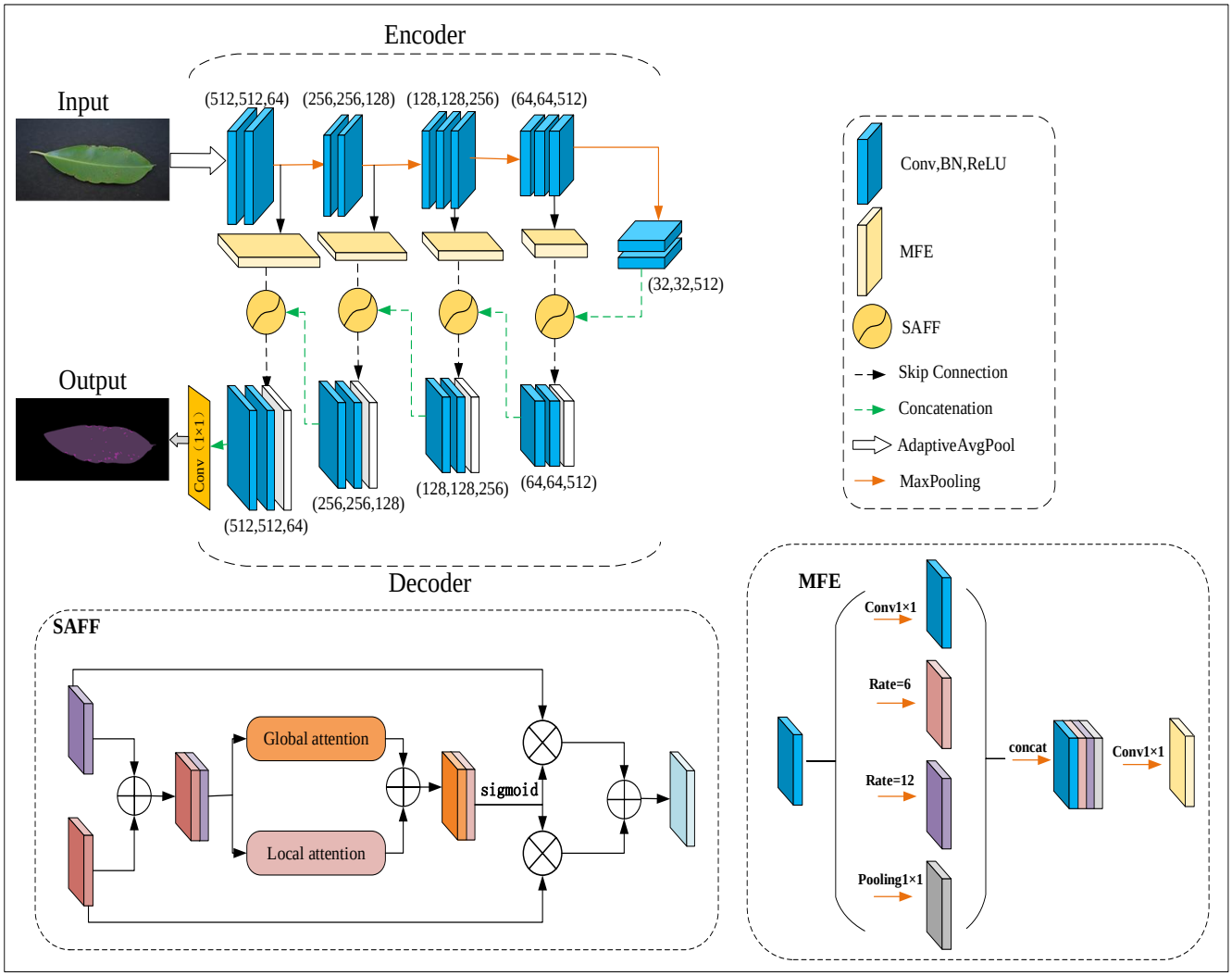


Figure 4: The network structure diagram of MAU-Net

2.3.1 Self-balancing attention feature fusion

In UNet network architectures, feature fusion during hopping connectivity is often challenged by a mismatch between scale and semantics, and these discrepancies can adversely affect network performance. In mango leaf disease feature extraction, this mismatched variability can cause the loss of small disease feature information of mango leaves. To overcome these challenges, the study proposes an improved SAFF attention module based on the Multiscale Channel Attention Mechanism (MS-CAM) [20]. In the encoding phase, the generated features often contain too much detail and noise, which may interfere with the model's judgment ability. In addition, while the decoding phase generates feature-rich semantic information, they may carry redundant or contradictory detail data. The SAFF Attention Module efficiently integrates feature information from different layers in a UNet network in a skip connection. By fine-tuning and optimizing the original simple design of the skip connection, it is ensured that the most relevant information about the disease characteristics is passed between the network layers. In turn, the key features are effectively fused to significantly improve network performance.

As shown in Figure 5, the SAFF module is divided into two branches: global attention and local attention. The global attention branch uses a global average pooling technique to discriminate attention from global features, while the local attention branch extracts channel attention from local features by two 1×1 convolutions. This strategic design allows the network to selectively focus on the feature information at each level, thus improving the model's perceptual ability and overall performance.

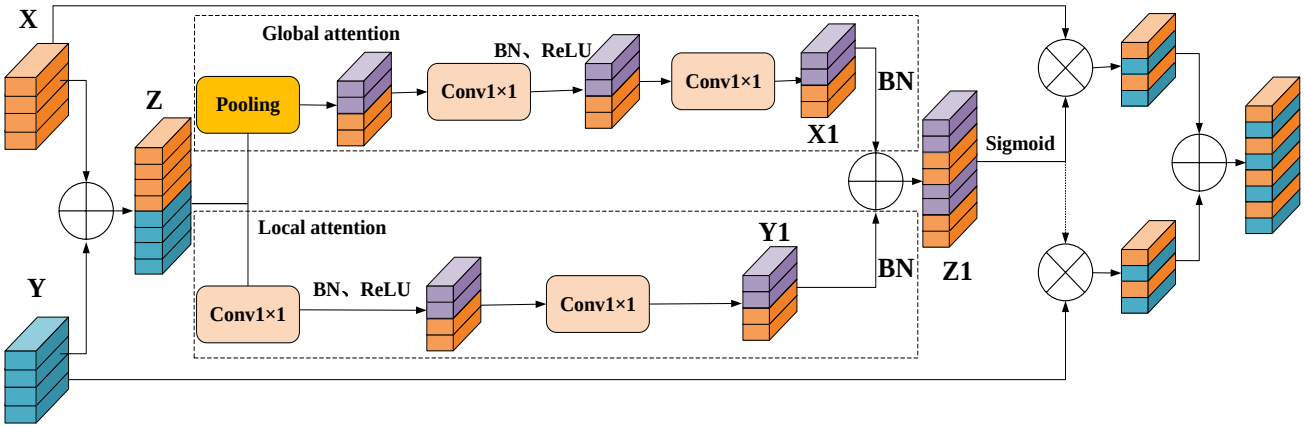


Figure 5: The module structure diagram of SAFF

In particular, the stitching operation is first performed on features X and Y to get the feature map Z. The feature map X1 is obtained by global average pooling and two 1×1 convolution operations in the global attention branch. The local attention branch then generates the feature map Y1 by two pointwise convolution operations. The X1 and Y1 feature maps are spliced to generate Z1. Finally, Z1 is passed through a sigmoid function to obtain the attention coefficient ω. The coefficients ω are used to weight the input features X and Y for fusion.

The formula for the SAFF attention mechanism is as follows:

$$X_1 = Conv\left(Conv\left(Pooling(X \oplus Y)\right)\right) \quad (1)$$

$$Y_1 = Conv\left(Conv(X \oplus Y)\right) \quad (2)$$

$$\omega = sigmoid(X_1 \oplus Y_1) \quad (3)$$

$$\mathrm{SAFF}(X, Y) = \omega \otimes X + (1 - \omega) \otimes Y \quad (4)$$

where X and Y denote the features of different layers, respectively, and ω denotes the learnable attention coefficients, where the fusion weights ω and 1 - ω are composed of real numbers between 0 and 1.

SAFF addresses the lack of accuracy and robustness in semantic segmentation tasks by reducing different degrees of semantic gaps in the feature fusion process. It achieves the effect of improving computational efficiency. It provides an effective strategy for solving the feature fusion problem in Unet networks and lays a solid foundation for improving the model's performance.

2.3.2 Multi-scale feature enhancement module

The size of the spots of mango leaf disease varies significantly in different stages of development. In semantic segmentation, traditional UNet models rely on single-scale convolutional operations to extract features. However, the segmentation accuracy could be better when using a single-scale convolution for learning training of multi-scale disease features. The feature information of large-scale diseases will cover the minor target diseases. To address the above problems, we designed an MFE module to improve the model's accuracy for full-period multiscale mango leaf disease segmentation by using a parallel structure of null convolution without three expansion rates. The structure of the MFE module is shown in Figure 6.

By introducing dilated convolution in the Deeplab algorithm [21]. Dilated convolution can efficiently extract multi-scale feature information without changing the size of the feature map. When different expansion coefficients are set, their receptive fields are different. The larger the expansion coefficient setting is, the better its feature extraction effect on the target comparison is. However, cavity convolution with a single expansion coefficient is set to lose critical information for extracting small target features when dealing with multi-scale targets. The principle of cavity convolution is shown in Figure 7.

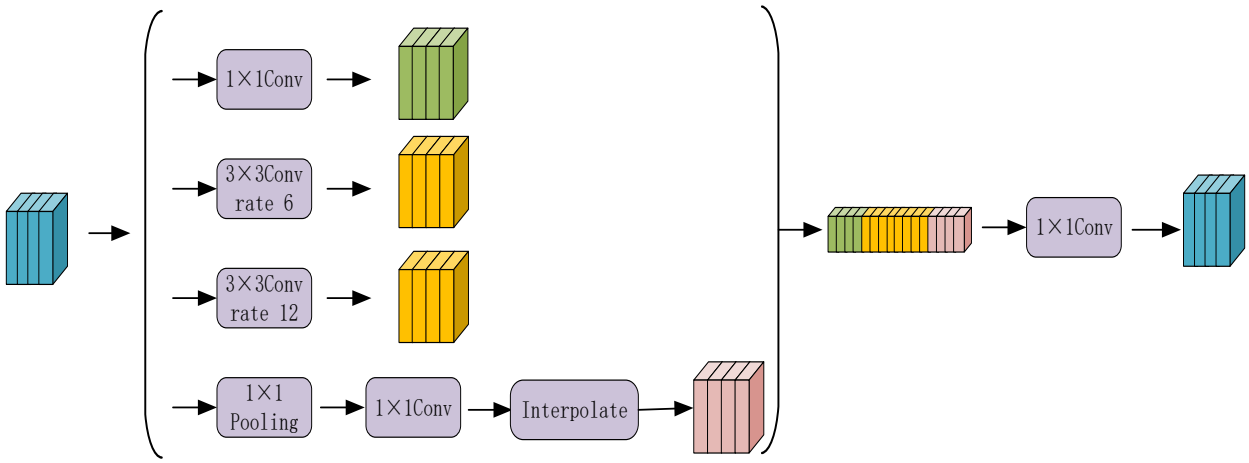


Figure 6: The module structure diagram of MFE

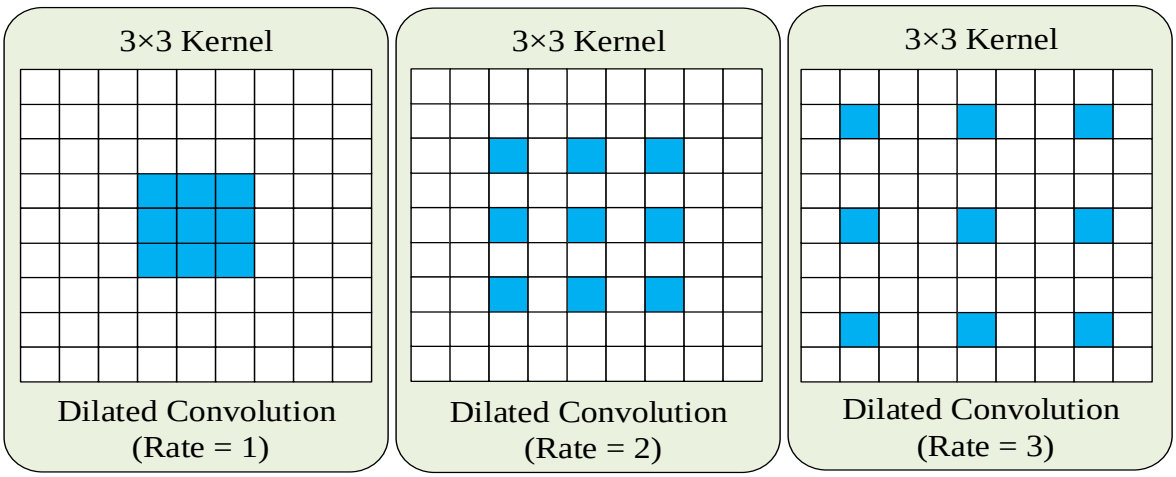


Figure 7: Convolution blocks of different expansion rate sizes

Specifically, the MFE module enhances the feature representation with three sizes of dilation convolution operations and pyramid pooling operations to improve the perceptual capabilities of the model. The dilation convolution operation increases the size of the receptive field by introducing different expansion rates to capture contextual information at different scales. The small scale convolutional kernel is mainly responsible for extracting the lesion region, while the large scale convolutional kernel is mainly used to enhance the diversity of features. This approach enables the network to better adapt to the different scales of the lesions in the image. The pyramid pooling operation, on the other hand, performs feature pooling by using pooling kernels of different sizes to obtain feature representations at different scales.

As shown in Figure 6, the MFE module contains four branches. The first branch is a 1×1 convolution operation. It can not only be used to adjust the number of channels of input features, but also able to preserve small target mango leaf disease features. The second and third branches are two 3×3 dilation convolution operations with expansion rates of 6 and 12, respectively. These two branches capture more connections between pixels by enlarging the size of the receptive field to obtain information about the disease characteristics of mango leaves at different scales. The fourth branch first goes through a pooling layer and then a 1×1 convolutional layer. Finally it goes through a 2-fold bilinear upsampling operation to get a feature map of the same size as the input features. The input features go into four branches respectively, and the high dimensional feature information is obtained after the splicing operation. After the dimensionality reduction operation of 1×1 convolutional layer, it is reduced to the same number of channels as the input features. Finally, the obtained enhanced features contain richer contextual information and stronger perceptual ability, which helps to improve the accuracy of multi-scale disease feature segmentation in mango leaves. The formula for the MFE module is as follows:

$$T_1 = Conv_{1\times1}(x) \tag{5}$$

$$T_2 = Conv_{3\times3,r=6}(x) \tag{6}$$

$$T_3 = Conv_{3\times3,r=12}(x) \tag{7}$$

$$T_4 = Interpolate\left(Conv_{1\times1}\left(Pooling(x)\right)\right) \tag{8}$$

$$T = concat(T_1, T_2, T_3, T_4) \tag{9}$$

The MFE module enhances the extraction of multi-scale features through parallel multi-branch convolutional operations. Make the model have stronger generalization ability when dealing with multi-scale features. This in turn improves the model's segmentation of multi-scale targets.

2.3.3 DF_Loss function

In the dataset of mango leaf diseases, there is a challenge of imbalance in the number of pixels. Figure 8 presents statistical data on the percentage of pixels in each category in the mango

leaf disease dataset. The disease category only accounts for 0.47% of the overall dataset, and its pixel number is much less than that of healthy leaves. Facing this situation of uneven sample distribution, the traditional cross-entropy loss function is often difficult to cope with it efficiently. When there are significant quantitative differences between categories, the model tends to predict those categories that are numerically superior. This usually leads to a decrease in the model's performance in the identification of diseased areas and fails to achieve accurate segmentation. To address the impact of unbalanced data on the model segmentation performance, a DF_Loss that incorporates Dice Loss and Focal Loss is proposed. It allows the model to focus more on pixels that are difficult to categorize during the learning process, thus improving the segmentation accuracy for mango leaf diseases.

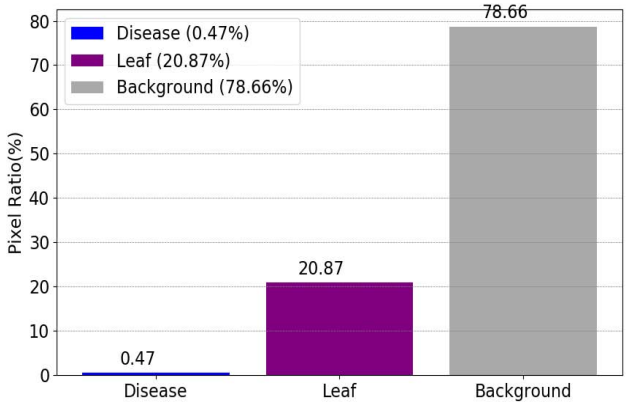


Figure 8: Pixel ratios for each category in the dataset, with only 0.47% of diseases, 20.87% of leaf parts and 78.66% of backgrounds

DF_Loss enhances the learning weights for difficult-to-categorize samples during training by integrating the attention regulation mechanism and balancing factor of Focal Loss. DF_Loss not only motivates the model to target the features of a few categories more accurately, but also significantly enhances the model's ability to identify and classify rare categories when dealing with unbalanced data. DF_Loss is compatible with the idea of similarity measure in Dice Loss. This strategy maps the degree of pixel-level match between predicted and actual segmentation results more realistically than the traditional cross-entropy loss function. The introduction of this metric allows DF_Loss to show excellent adaptability in segmentation tasks, especially when rare categories are present in the dataset. When confronted with the challenge of sample imbalance, the traditional cross-entropy loss function tends to bias the model in favor of the majority category prediction. However, DF_Loss effectively corrects this bias by incorporating the Focal Loss strategy. It empowers the model with greater ability to focus on sample categories that are less numerous and difficult to categorize. During the training process, DF_Loss demonstrated excellent learning ability to capture and learn information from all categories in a balanced manner, thus significantly improving the overall performance of the model.

The formula for calculating CE losses is as follows:

$$CELoss = -\frac{1}{N}\sum_i^N \sum_j^N \left(Y_{ij}\, log\left(X_{ij}\right)\right) \quad (10)$$

where N is the number of samples and Y_{ij} is the unique code of the actual label indicating the probability that the i_{th} sample belongs to the j_{th} category. X_{ij} is the probability that the i_{th} sample belongs to the j_{th} category as predicted by the model. The formula for the Dice loss is as follows:

$$\text{Dice}Loss = 1 - \frac{2\times\sum_i^N p_i \times g_i}{\sum_i^N p_i^2 + \sum_i^N g_i^2} \quad (11)$$

where pi denotes the pixel value in the predicted segmentation result, g_i denotes the pixel value in the actual segmentation result, and N is the pixel sum.

The formula for the Focal loss is as follows:

$$\text{FocalLoss} = -\alpha \times (1 - p_t)^{\beta} \times log(p_t) \quad (12)$$

where pt denotes the probability of a pixel in the predicted segmentation result, α is a balancing factor, and β is a tuning parameter.

Specifically, this paper uses Dice Loss as the basic loss function to measure the segmentation accuracy of the model at pixel level. Meanwhile, Focal Loss is introduced as a weighting term to increase the model's focus on a few categories. The DF_Loss function is calculated as follows:

$$\text{DF}\_Loss = DiceLoss + \lambda \times FocalLoss \quad (13)$$

where λ is a balancing parameter that adjusts the weights of the Dice loss and the Focal loss.

The DF_Loss loss function incorporates the ideas of both Dice Loss and Focal Loss loss functions. During the training process, it gives the model a greater ability to focus on sample categories that are less numerous and difficult to classify. It can improve the model's ability to segment mango leaf diseases during the task of mango leaf disease segmentation.

3. Result

The hardware configuration of the experimental environment in this study is as follows: Intel(R) Core (TM) i9-10900k, 64 G RAM, NVIDIA® GeForce RTX4080ti, 64-bit Windows operating system. The model is built by the Pytorch framework; the version of PyTorch is 1.10.0. After several trials, the hyperparameters are set as follows: the optimizer is Adamw, the momentum is 0.9, the weight decay is 1e-2, the batch size is 4, the initial learning rate is 1e-4, the minimum learning rate is 1e-7, the learning rate decay strategy is cos, the decay rate is 0.1, and the epoch is 100. We divided the mango leaf disease dataset into training set, testing set and validation set according to the ratio of 7:2:1 for cross-validation in the training stage.

3.1 Evaluation Indicators

We chose the following three evaluation metrics to measure the segmentation effectiveness of the model: intersection over union (IoU), mean pixel accuracy (mPA), and overall accuracy F1. In the segmentation task, IoU represents the ratio of the intersection and union between the prediction results of a category and the actual values of that category.PA represents the ratio of all correctly predicted pixels to all pixels. F1 Accuracy represents the accuracy of all targets detected by the semantic segmentation algorithm. IoU, mPA and F1 are calculated as follows:

$$\text{IoU} = \frac{TP}{TP+FP+FN} \quad (14)$$

$$PA = \frac{TP+TN}{TP+TN+FP+FN} \quad (15)$$

$$mPA = \frac{sum(PA_i)}{m} \quad (16)$$

where TP denotes the true case, i.e., the number of pixels where both the prediction result and the true label are positive cases. TN denotes the number of pixels where both the prediction result and the true label are negative cases. FP denotes the number of pixels where the prediction result is positive but the true label is negative. FN denotes the number of pixels where the prediction result is negative but the true label is positive.

$$P = \frac{TP}{TP+FP} \times 100\% \quad (17)$$

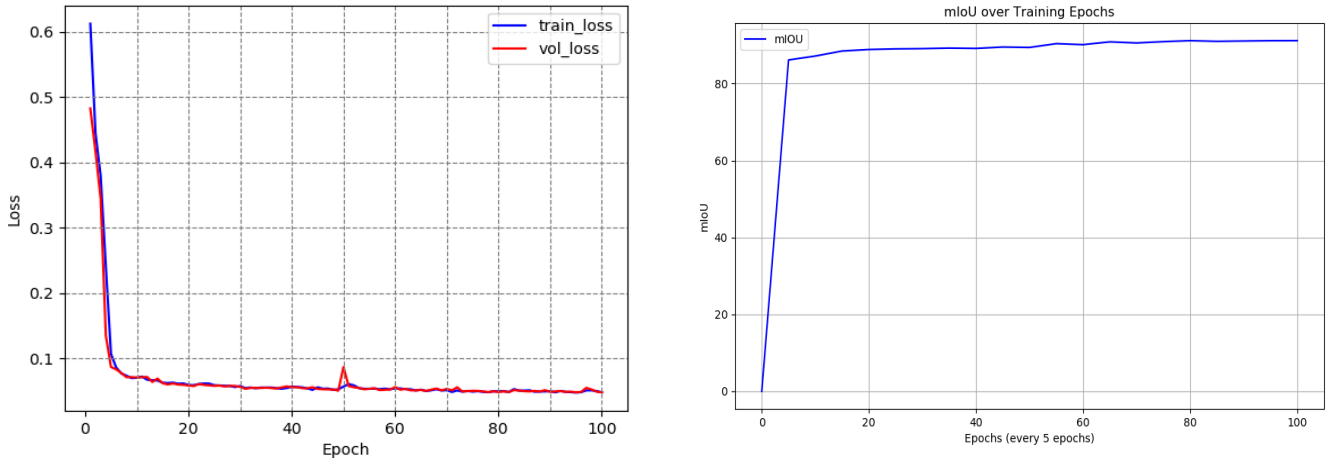
$$R = \frac{TP}{TP+FN} \times 100\% \quad (18)$$

$$F1 = 2 \div \left(\frac{1}{P} + \frac{1}{R}\right) \quad (19)$$

where P denotes the inspection accuracy rate and R denotes the inspection completeness rate. F1 The accuracy rate takes values ranging from 0 to 1, with higher values indicating a better output from the model.

3.2 Experimental Analysis of Comparative Loss Functions

The segmentation results of the UNet algorithm using the DF_Loss loss function are shown in Figure 9, where the network model is rapidly fitted and the loss values are significantly reduced in the first 30 iteration cycles. In the subsequent 70 iteration cycles, the loss value remains basically smooth which indicates that the training of the model has been optimized and the loss function of the model has converged. There is a small fluctuation in the fiftieth training cycle. The reason for this is that the experiment used freeze training for the first fifty rounds and thaw training for the last fifty rounds. It is normal to have fluctuations at the beginning of thawing.



(a) (b)

Figure 9: Plot of loss values for DF_Loss training

During the experiments, four different loss functions are used in the paper: CE loss, Dice Loss, Focal Loss, and DF_Loss function. The performance of these loss functions in practical applications is analyzed in detail below. The resultant data using the individual loss functions are shown in Table 2. To begin with, the experiment is benchmarked against cross-entropy loss with a disease IoU of 79.45%. However, it did not perform optimally for our disease segmentation task. Compared to cross-entropy, Dice loss has a 0.43% improvement in disease IoU to 79.88%. The results show that Dice loss has better adaptation to the dataset and model of the present study. It is relatively insensitive to the initial state of the model and the quality of the data, showing strong generalization ability. The model with the Focal loss function improved the disease IoU by 0.8 percentage points to 80.25% compared to the model with the crossentropy loss function. The main reason for this is due to the strong ability of Focal loss to regulate the problem of positive and negative sample imbalance, which successfully improves the focus of the model on diseases. The result suggests that Focal loss can show significant advantages when there is a category imbalance problem. Finally, the Unet model with DF Loss achieves 81.85% in disease IoU, which is 2.4 percentage points higher than the baseline model, and this improvement outperforms the model with Dice and Focal losses. Meanwhile, the F1 score of the Unet model with DF Loss also reached 0.945, which is an improvement of 1.3 percentage points compared to the cross-entropy model before improvement. The results show that DF Loss achieves significant improvement in handling unbalanced data. It effectively combines the advantages of Dice and Focal losses and successfully solves the category imbalance problem while focusing on pixel-level information. The innovative loss function provides new ideas and directions for solving the problem of segmentation of unbalanced data.

Table 2: Results of segmentation accuracy of models using different loss functions on a test set of mango leaf diseases

Methods	*IoU/%*		*F1/%*	*mPA/%*
	leaf	Disease		
Unet	98.85	79.45	93.2	95.03
UNet+Dice_loss	99.05	79.88	93.8	95.31
UNet+Focal_loss	99.10	80.25	94.1	95.66
UNet+DF_loss	**99.12**	**81.85**	**94.5**	**95.64**

3.3 Ablation Experiments

In order to verify the validity of the MAU-Net model, ablation experiments were conducted on the MFE and SAFF modules

of the MAU-Net model. The Mango Leaf Disease dataset was used in the experiments and DF_Loss was used as the loss function for the Unet network, which was used as the baseline model for the ablation experiments. The experimental results are shown in Table 3. We can find that each part contributes to the final performance. We introduced only the MFE feature enhancement module, which improved the performance over the baseline model by 0.04% and 0.49% for leaf and disease IoU, respectively. The combined F1 score improves performance by 1.8%. We believe that by introducing the MFE module, the model is more robust in extracting and recognizing disease features and more accurate in segmenting leaf and disease edge information. We introduced only the SAFF module, who improved performance over the baseline model by 0.05% and 0.91% for leaf and disease IoUs, respectively. the F1 score composite score improved by 2.0 %. This suggests that the application of the SAFF module helps the model to better balance the feature information of different layers, thus improving the accuracy of segmentation. In addition, when the MFE module and SAFF module were introduced, it increased the performance over the baseline model by 0.09% and 2.48% for leaf and disease IoUs, respectively. The F1 score composite score increased by 2.6%. These results show that the MFE and SAFF mechanisms provide proof for improving the performance of mango leaf disease segmentation and achieving better feature representation learning.

Table 3: The results of segmentation accuracy for different test models.

Methods	*IoU/%*		*F1/%*	*mPA/%*
	leaf	Disease		
Biseline	99.12	81.85	94.5	95.64
Biseline+MFE	99.16	82.36	96.3	96.15
Biseline+SAFF	99.17	82.76	96.5	96.21
MAU-Net	**99.21**	**84.33**	**97.1**	**96.94**

In addition, as shown in Figure 10(c), the prediction results of the baseline model are not clearly segmented for the edge detail information of leaf diseases, and there are cases of false-negative misclassification of small spots. As shown in Figure 10(d), the model prediction results after the introduction of the SAFF module are more accurate in terms of disease edge information. However, there is also a false negative misclassification, which is due to the limitation of a single-size convolutional kernel to extract features, resulting in insufficient semantic information extraction. The introduction of the SAFF module for feature fusion weakens the background information in the shallow features. It can lead to irreversible loss of semantic information causing errors in some pixel predictions. Figure 10(e) shows the model prediction results of introducing the MFE module, and the edge information is clearer compared to the other model prediction results. Figure 10(f) shows the prediction results made by the method proposed in this paper. It can be seen that the segmented disease and leaf edge information are close to the real information. And it can accurately segment the void.

The experimental results show that by introducing the feature enhancement module of MFE, the model's ability to perceive diseases at different scales can be enhanced. It also enhances various types of features during extraction, which effectively avoids the overfitting phenomenon caused by the excessive depth of the network. The SAFF module can help the network select the feature information of the diseases in different layers to be fused in the process of feature fusion. The module can ignore the features of less impact on the current task. The allows the model to focus more on processing those features that are useful for the segmentation task. In summary, by introducing the MFE feature enhancement module and the SAFF feature fusion module, the performance of the model in extracting target features in the image semantic segmentation task has been significantly improved. It provides a new effective way to solve the leaf disease segmentation problem and provides useful insights for research in related fields.

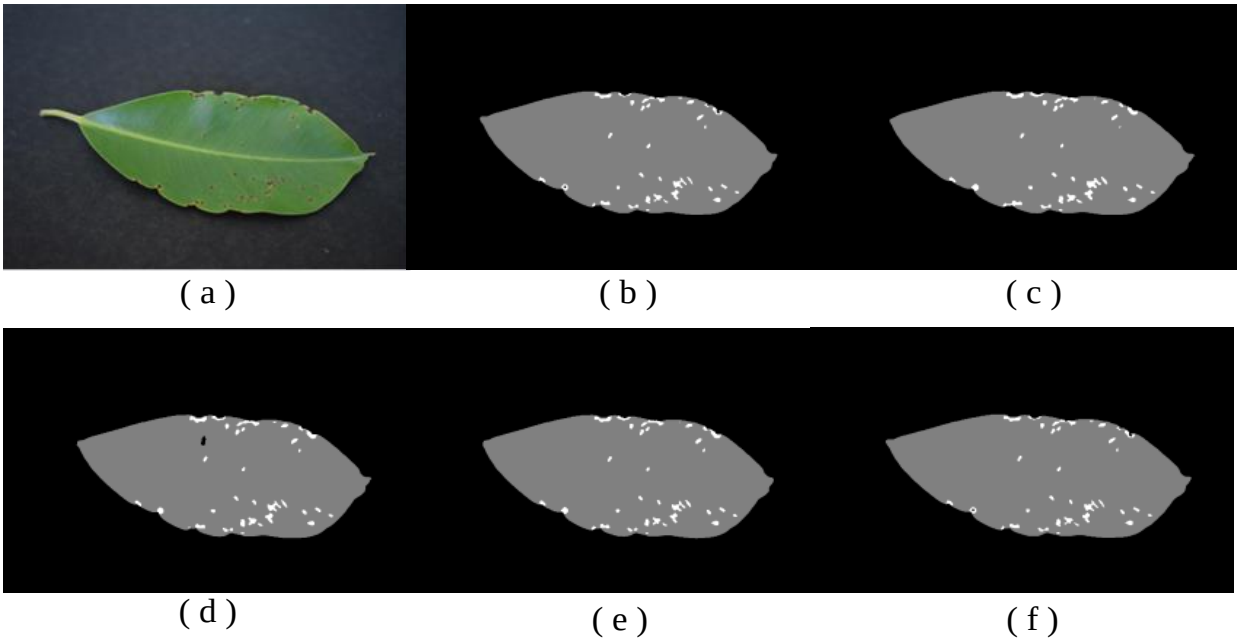


Figure 10: Comparison of segmentation results from ablation experiments, where (a) is the original image, (b) is the labeled image, (c) is the baseline model Unet with the DF_Loss loss function replaced, (d) is the addition of the SAFF module, and (e) is the addition of the MFE module. (f) is our proposed MAU-net. The red boxes show locations that differ from the actual labeling

3.4 Comparison of Different Models

In this subsection, MAU-Net is compared with popular deep-learning semantic segmentation models to demonstrate the advantages of the MAU-Net model. These models are DeepLabv3+ [22], SENet [23], FCN [24], Swin-Unet [25], PSPNet [26] and ENet [27]. DeepLabv3+ employs an ungauged spatial pyramid pooling (ASPP) technique which effectively enhances the sensory field and mitigates the problem of information loss due to pooling. FCN replaces the fully connected layer with a fully convolutional layer which allows the network to accept input images of arbitrary size and output a dense feature map of corresponding size. SENet introduces an attention mechanism that focuses on enhancing the model's attention to important features. By introducing the Squeeze-and-Excitation block, SENet can adaptively adjust the weights of each channel to highlight attention to the important parts of the feature map. Swin-Unet enables the model to better capture the global contextual information of the image by introducing the Transformer block. PSPNet uses the Pyramid Pooling block to effectively utilize contextual information. ENet reduces model complexity while maintaining high performance by using lightweight designs such as 1x1 convolution and depth-separable convolution. All models are placed in the same experimental environment as MAU-Net and are trained on the mango leaf disease image dataset. The segmentation accuracies of different models are shown in Table 4.

As shown in Table 4, MAU-Net performed best on the mango leaf disease dataset. On the IoU of disease segmentation, its segmentation accuracy was improved by 5.59%, 27.62%, 11.32%, 25.39%, 27.56% and 12.57% compared to DeepLabv3, FCN, ENet, Swin-Unet, PSPNet and SENet models, respectively. In the IoU of blade segmentation improved by 0.19%, 1.83%, 0.38%, 2.64%, 0.67%, and 0.44%, respectively. In F1 composite scores were improved by 1.6%, 5.9%, 2.4%, 3.6%, 6.1%, and 3.1%, respectively. In mPA accuracy improved by 2.26%, 6.1%, 3.27%, 4.07%, 6.51%, and 3.98%, respectively.

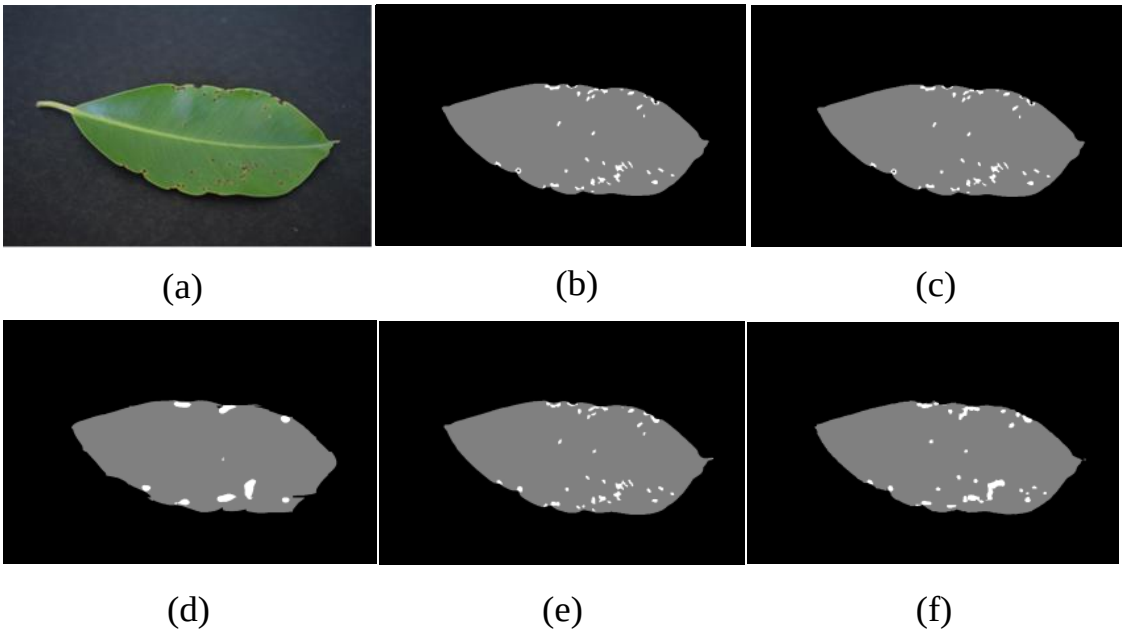
Table 4: The results of segmentation accuracy of different models on the test set of mango leaf diseases.

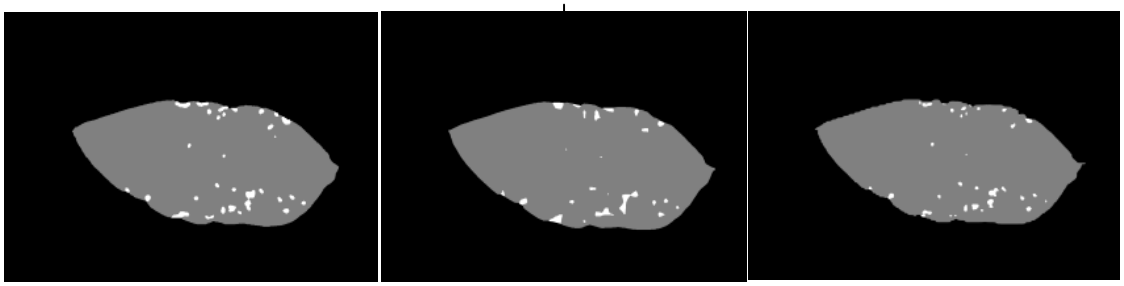
Methods	*IoU/%*		*F1/%*	*mPA/%*
	leaf	Disease		
Deeplabv3-Plus	99.02	78.74	95.3	94.68
FCN	97.38	56.61	91	90.84
ENet	98.83	73.01	94.5	93.67
Swin-UNet	96.57	68.94	93.3	92.87
PSPNet	98.54	56.81	90.8	90.43
SENet	98.77	71.76	93.8	92.96
MAU-Net	**99.21**	**84.33**	**97.1**	**96.94**

3.5 Visualization of Mango Diseased Leaf Segmentation Results

To validate the segmentation performance of MAU-Net for mango leaf disease, all models were compared using two sets of mango leaf disease test images. The resultant images were predicted by segmentation, with black as the background, gray as the leaf, and white as the disease. Differentiation using the above colors allows for a more precise comparison of the segmentation results of each method.

Figure 11 demonstrates the segmentation results of each model under the test set of mango anthracnose leaves. FCN segmented both leaves and spots poorly. FCN extracted an incomplete leaf area and missed some small spots. Deeplabv+ segmented leaves and spots better than the other models, which had background misclassification and directly added small voids to the spots. SENet is not accurate enough in segmenting the edges of the spots, and multiple individual spots are split directly into one spot. Net needs to be more precise in extracting the leaf area and miss small spots. PSPNet has better segmentation of leaves, but PSPNet misses many small spots. Swin-Unet has a relatively complete segmentation of leaves but has poorer completeness in segmenting spots, and some spots are missed. It can be seen that MAU-Net has better leaf and disease segmentation than Deeplabv3+. The tiny spots missed by other methods can also be extracted clearly, and segmentation is the finest for leaf edges.

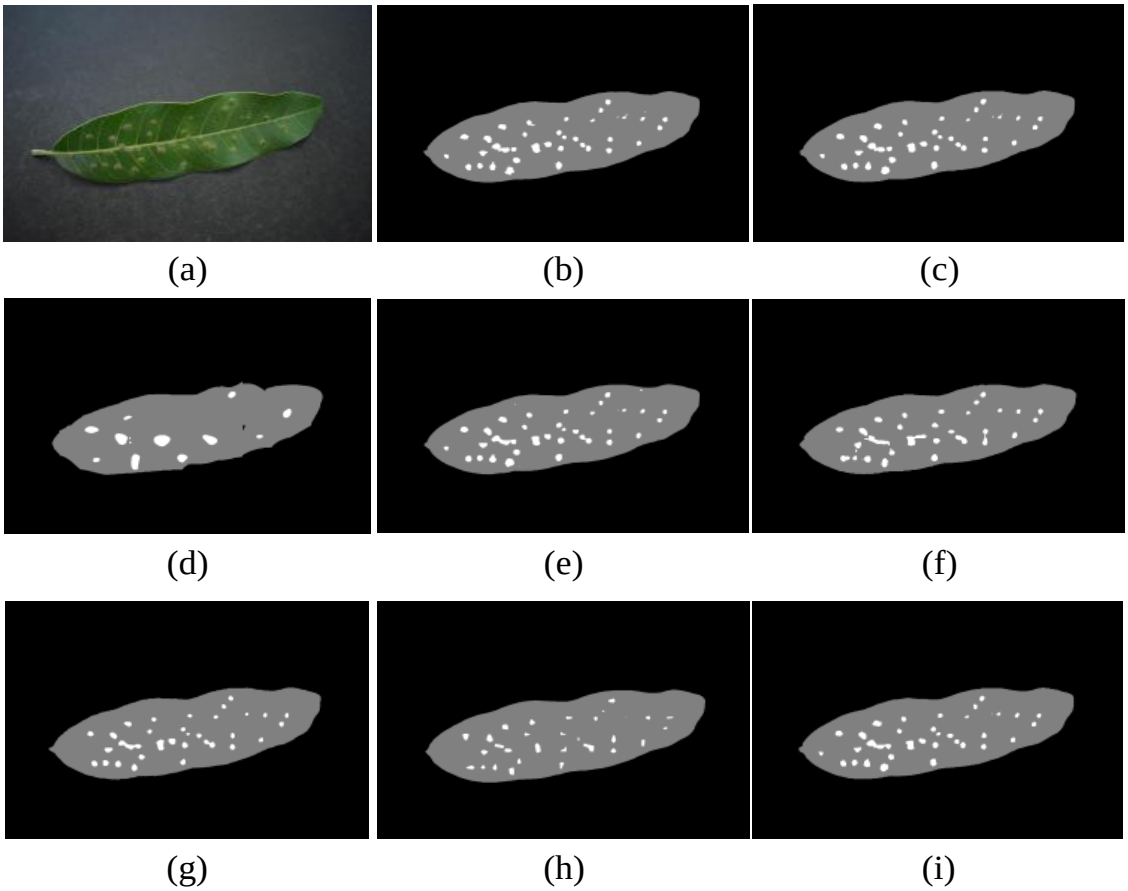




(g) (h) (i)

Figure 11: Segmentation results of each model under the test set of mango anthracnose leaves, where the red boxes show locations that differ from the actual labeling. (a)Original. (b)Label. (c)MAU-Net. (d)FCN. (e)Deeplabv3+. (f)SENet. (g)Enet. (h)PSPNet. (i)Swin-Unet.

Figure 12 shows the segmentation results of each model under the test set of mango leaf spot disease leaves. FCN segmented both leaves and spots poorly, with a large number of misclassifications. Deeplabv+ had omissions of small spots and segmented disconnected spots as one spot. The SENet model had misclassifications of small disease targets. ENet missed some small spots, and extracting prominent target diseases needs to be completed. PSPNet omitted small spots and extracted large targets of broken diseases incompletely. Swin-Unet was used to extract small spots. PSPNet has incomplete extraction of small lesions. Swin-Unet splits disjoint lesions into one lesion. In contrast, MAU-Net can extract the area of lesions more accurately and can segment each lesion precisely.



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Figure 12: Segmentation results for each model under the test set of mango leaf spot disease leaves, where the red boxes show locations that differ from the actual labeling. (a)Original. (b)Label. (c)MAU-Net. (d)FCN. (e)Deeplabv3+. (f)SENet. (g)Enet. (h)PSPNet. (i)Swin-Unet.

In order to compare the differences in the segmentation performance of different models more comprehensively, the segmentation results of different models are compared. It can be seen from Figure 13 that, compared with MAU-Net, the other models have poorer segmentation results due to the differences in the structure of the feature extraction and the different ways of jump connections. For example, the pyramid pooling and spatial pyramid structure of PSPNet and DeepLabv3+ enlarges the perceptual area but tends to miss minor points in downsampling. Swin-Unet uses self-attention for global modeling and a jump-connected encoder-decoder structure but cannot encode absolute locations. Therefore, it cannot accurately recover disease location information during up-sampling, which leads to poor segmentation results. By comprehensive comparison, the segmentation effect of MAU-Net is superior to other methods.

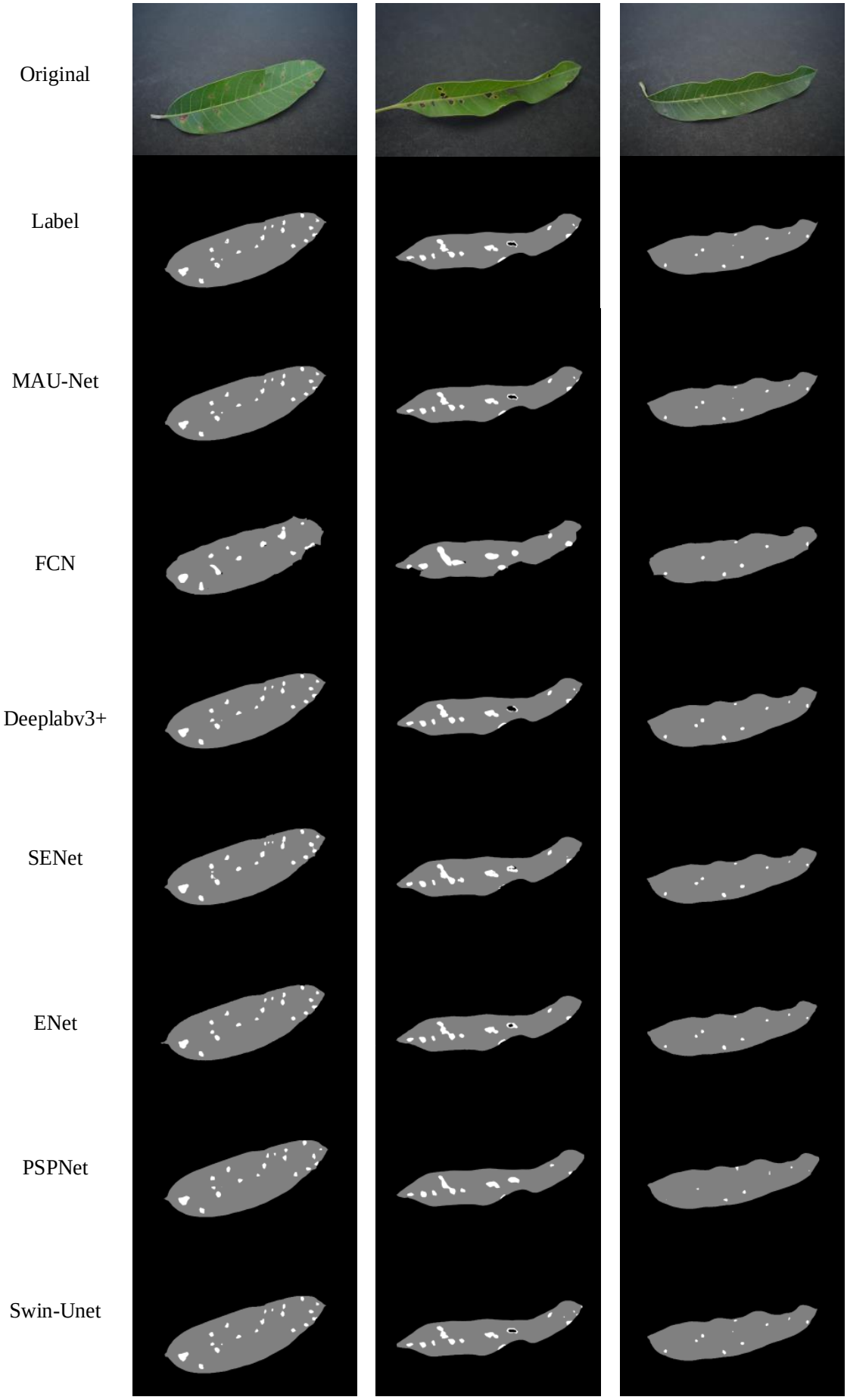


Figure 13: Segmentation results of different models for mango leaf disease image dataset.

4. Discussion

The MAU-Net model proposed in this study is an innovative improvement of the classic U-Net architecture, integrating the MFE and SAFF modules explicitly designed to segment mango leaves and their diseases. MAU-Net demonstrates superior accuracy and segmentation performance compared to current leading segmentation models. Ablation experiments further validate the impact of different modules and loss functions on MAU-Net's segmentation performance. The study investigated the feasibility of applying MAU-Net to segment the disease of mango leaves over a total period. The robust feature representation capability of MAU-Net in capturing complex features of target leaves and complex diseases is demonstrated. While the study focused on mango leaf segmentation, the core principles and technical framework of MAU-Net exhibit generalizability, suggesting potential applicability to other crop leaves. To enhance MAU-Net's practical versatility, future work will expand datasets to include diseased leaves from diverse crop varieties and natural outdoor scenes, refining its generalization capabilities. However, potential limitations may arise when applying MAU-Net to other crop leaf categories, given variations in morphology, texture, and disease characteristics. Thus, future studies extending MAU-Net to different crops will require thorough research and understanding of the specific crop leaf and disease characteristics. This involves necessary optimizations and adjustments to ensure the method's accuracy and robustness in diverse agricultural contexts.

5. Conclusion

In this paper, a new segmentation network, MAU-Net, is proposed. MAU-Net solves the problem of difficulty in recognizing small targets and low accuracy in segmenting mango leaf disease images. First, the SAFF module is proposed to solve the problem of the large gap between semantic features at different levels of features in the process of skip connection. This module refines the fusion process using reliable weights from global and local attention mechanisms. Secondly, the MFE module was proposed to solve the problem of low accuracy of disease segmentation for small targets with significant differences in the characteristics of mango leaf diseases in various periods. This module facilitates multi-scale feature extraction and enhancement by capturing contextual information at different scales, thereby improving the model's ability to recognize objects of different sizes. Finally, the study designed the DF_Loss loss function to replace the original cross-entropy loss function to solve the problem of imbalance between positive and negative samples in the mango leaf disease dataset. This design enables the model to learn disease features better, especially when the number of samples is limited. Experimental results show that MAU-Net performs best on the mango leaf disease dataset compared to mainstream semantic segmentation methods. This indicates that MAU-Net has excellent potential for broad application in mango leaf disease segmentation. It is also a significant reference value for the design of full-period disease segmentation models for other plants. Future research will optimize the model to adapt to different crop disease leaves.

References

[1] Lebaka, V.R., Wee, Y.-J., Ye, W., Korivi, M.: Nutritional composition and bioactive compounds in three different parts of mango fruit. International Journal of Environmental Research and Public Health 18(2), 741 (2021)

[2] Saravanan, T., Jagadeesan, M., Selvaraj, P., Aravind, M., Raj, G.D., Lokesh, P.: Prediction of mango leaf diseases using convolutional neural network. In: 2023 International Conference on Computer Communication and Informatics (ICCCI), pp. 1–4 (2023). IEEE

[3] Huan Cao, R.F.: Deep learning based classification and identification of mango pests and diseases (in chinese). Computer Technology and Development 33, 115–119 (2023)

[4] Islam, M., Mannan, M., Kabir, M., Islam, A., Olival, K.: Analgesic, antiinflammatory and antimicrobial effects of ethanol extracts of mango leaves. Journal of the Bangladesh Agricultural University 8, 239–244 (2010)

[5] Gulavnai, S., Patil, R.: Deep learning for image based mango leaf disease detection. International Journal of Recent Technology and Engineering (IJRTE) 8(3S3), 54–56 (2019)

[6] Sharma, R.K., Ghandi, P.: Reliability estimation and optimization: A neuro fuzzy based approach. International Journal of Computer Science and Information Security (IJCSIS) 16, 128–132 (2018)

[7] Mu, H., Wang, K., Yang, X., Xu, W., Liu, X., Ritsema, C.J., Geissen, V.: Pesticide usage practices and the exposure risk to pollinators: a case study in the north china plain. Ecotoxicology and Environmental Safety 241, 113713 (2022)

[8] Luo, Y., Sun, J., Shen, J., Wu, X., Wang, L., Zhu, W.: Apple leaf disease recognition and sub-class categorization based on improved multi-scale feature fusion network. IEEE Access 9, 95517–95527 (2021)

[9] Arya, S., Singh, R.: A comparative study of cnn and alexnet for detection of disease in potato and mango leaf. In: 2019 International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT), vol. 1, pp. 1–6 (2019). IEEE

[10] Merchant, M., Paradkar, V., Khanna, M., Gokhale, S.: Mango leaf deficiency detection using digital image processing and machine learning. In: 2018 3rd International Conference for Convergence in Technology (I2CT), pp. 1–3 (2018). IEEE

[11] Prasetyo, E., Adityo, R.D., Suciati, N., Fatichah, C.: Mango leaf image segmentation on hsv and ycbcr color spaces using otsu thresholding. In: 2017 3rd International Conference on Science and Technology - Computer (ICST), pp. 99–103 (2017). IEEE

[12] Tumang, G.S.: Pests and diseases identification in mango using matlab. In: 2019 5th International Conference on Engineering, Applied Sciences and Technology (ICEAST), pp. 1–4 (2019). IEEE

[13] Pansy, D.L., Murali, M.: Uav hyperspectral remote sensor images for mango plant disease and pest identification using md-fcm and xcs-rbfnn. Environmental Monitoring and Assessment 195(9), 1120 (2023)

[14] Saleem, R., Shah, J.H., Sharif, M., Yasmin, M., Yong, H.-S., Cha, J.: Mango leaf disease recognition and

classification using novel segmentation and vein pattern technique. Applied Sciences 11(24), 11901 (2021)
[15] Chouhan, S.S., Singh, U.P., Jain, S.: Web facilitated anthracnose disease segmentation from the leaf of mango tree using radial basis function (rbf) neural network. Wireless Personal Communications 113, 1279–1296 (2020)
[16] Gautam, V., Ranjan, R.K., Dahiya, P., Kumar, A.: Esdnn: A novel ensembled stack deep neural network for mango leaf disease classification and detection. Multimedia Tools and Applications, 1–27 (2023)
[17] Pham, T.N., Van Tran, L., Dao, S.V.T.: Early disease classification of mango leaves using feed-forward neural network and hybrid metaheuristic feature selection. IEEE Access 8, 189960–189973 (2020)
[18] Chouhan, S.S., Singh, U.P., Kaul, A., Jain, S.: A data repository of leaf images: Practice towards plant conservation with plant pathology. In: 2019 4th International Conference on Information Systems and Computer Networks (ISCON), pp. 700–707 (2019). IEEE
[19] Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18, pp. 234–241 (2015). Springer
[20] Dai, Y., Gieseke, F., Oehmcke, S., Wu, Y., Barnard, K.: Attentional feature fusion. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 3560–3569 (2021)
[21] Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L.: Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE transactions on pattern analysis and machine intelligence 40(4), 834–848 (2017)
[22] Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H.: Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 801–818 (2018)
[23] Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7132–7141 (2018)
[24] Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3431–3440 (2015)
[25] Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., Wang, M.: Swinunet: Unet-like pure transformer for medical image segmentation. In: European Conference on Computer Vision, pp. 205–218 (2022). Springer
[26] Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J.: Pyramid scene parsing network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2881–2890 (2017)
[27] Paszke, A., Chaurasia, A., Kim, S., Culurciello, E.: Enet: A deep neural network

Author Profile

Bibo Lu male, Doctor, professor, master supervisor, graduated from Jilin University in June 2008, mainly engaged inartificial intelligence, image and video processing and analysis and other aspects of research.