International Journal of Science and Engineering Applications
Volume 14-Issue 11, 117 – 125, 2025, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1411.1021

Regional Segmentation Algorithm for Rice Leaf Diseases in Complex Backgrounds

Yu Xiaoyan
School of Information
Engineering

Wuhan Huaxia Institute of
Technology
Wuhan, China

Abstract: As a vital staple crop, rice is susceptible to leaf diseases such as bacterial leaf blight, rice blast, brown spot, and sheath blight during its growth cycle, which severely impact yield and quality. Precise disease segmentation is crucial for early prevention and control, yet it faces challenges including complex backgrounds, diverse lesion morphologies, and similar features. This paper proposes SwinDA-DeepLabv3+, a segmentation network based on Swin Transformer multi-scale feature fusion: it employs Swin Transformer as the backbone network to capture global disease features; replaces traditional ASPP with a dual-attention-enhanced dilated convolution module to enhance focus on sparse lesions; and mitigates sample imbalance through a Dice-Focal hybrid loss function. Experiments demonstrate that this network achieves an mIoU of 83.34% on a four-class rice disease dataset, representing a 4.68% improvement over the original DeepLabv3+.

Keywords: Rice Disease Segmentation; Swin Transformer; Attention Mechanism,;Multi-scale Feature Fusion

1. INTRODUCTION

As China's primary staple crop, the stability of rice production directly impacts food security and agricultural economic development.[1] Data from the National Bureau of Statistics indicates that in recent years, China's rice cultivation area, total yield, and yield per unit area have shown a fluctuating downward trend. Concurrently, changes in the population's age structure have led to a reduction in the agricultural labor force, weakening the level of precision management in rice cultivation and intensifying the difficulty of disease control. Deteriorating ecological conditions and pathogen mutations have further increased the frequency and transmission speed of rice diseases. Among these, leaf diseases—due to their early onset[2], observability, and direct impact on photosynthesis and subsequent panicle health—have become a critical focal point for disease control. Traditional manual diagnosis methods suffer from high subjectivity and low accuracy[3][5], failing to meet the demands of large-scale, precision-based control. At the policy level, documents such as the 14th Five-Year Plan and the National Smart Agriculture Action Plan (2024–2028)[6] emphasize agricultural technological innovation and smart agriculture development, providing a favorable policy environment for applying intelligent technologies like computer vision in disease monitoring. Computer vision technology can effectively address the low efficiency and poor accuracy of manual identification. However, segmenting rice leaf diseases still faces challenges such as complex backgrounds, diverse leaf morphologies, varying disease scales, and noise interference.

Chen et al. [7] combined GCLPSO with OTSU multi-threshold segmentation for maize disease images, incorporating non-local mean filtering to enhance noise resistance and achieving 94.58% segmentation accuracy. Rahmawati et al. [8] employed OTSU threshold segmentation with non-local mean denoising, achieving 88% accuracy in coffee leaf rust detection with good stability under noisy conditions. Traditional semantic segmentation methods are computationally simple but struggle to capture global information. Yang et al.[9] proposed a segmentation approach based on YOLOv8 and an improved DeepLabV3+, validated on public datasets. However, the samples lacked diversity in field environments, leaving the method's adaptability to real-world scenarios unknown. Shoaib et al. [10] enhanced U-Net with InceptionNet, achieving 98.66% accuracy and 98.73% Dice coefficient for tomato leaf disease segmentation. While excelling in classification tasks, such models heavily rely on clean datasets like PlantVillage, lacking real-world disturbances like field lighting and occlusion, thus limiting their generalization and practical applicability. CNN-based models exhibit high segmentation accuracy but struggle with modeling long-range dependencies and generalization. Yang et al.[11] enhanced MA-Net by introducing a Mix Vision Transformer to capture global information and ECANet for lightweight processing and noise reduction, achieving mIoU of 98.1% and Dice Loss of 0.9% for apple leaf spot segmentation. Elmessery et al. [12] proposed a SegFormer-based model that enhances strawberry disease segmentation accuracy through three Mix Transformer encoders. While Transformer architectures effectively model global context, they incur high computational costs.To address these challenges, this paper introduces SwinDA-DeepLabv3+, a segmentation network that integrates global context modeling with dynamic attention mechanisms.

2. Data Acquisition

2.1 Dataset Overview

Leaf diseases in rice are a major limiting factor affecting both yield and quality. These diseases are diverse in nature, including rice blast, sheath blight, bacterial leaf blight, and brown spot disease. This study addresses the issue of low segmentation accuracy for rice leaf diseases in complex field environments by conducting field surveys and integrating online rice leaf disease data. Four prevalent diseases—bacterial leaf blight, rice blast, brown spot, and sheath blight—were selected as research subjects.

The rice leaf disease images used in this study were sourced from the publicly available dataset by Zhao Jinfeng et al.[13] . All data were collected using a Xiaomi 10S smartphone in actual farmland settings. The dataset includes images captured during different time periods (7:00–11:30 AM and 4:00–6:30 PM) and under various weather conditions (sunny, cloudy, rainy) to enhance diversity. Images of rice leaf diseases under different weather and time conditions are shown in Figure 3-3. To avoid leaf distortion and clearly capture lesion details, the camera lens was maintained at a parallel distance of 10–50 cm above the canopy during shooting. A total of 3,122 rice leaf disease images with a resolution of 5,792 × 4,344 pixels were collected, including 761 images of bacterial leaf blight, 911 images of rice blast, 858 images of brown spot disease, 592 images of sheath blight, and 1,582 images of healthy leaves. These images were collected from 26 distinct rice field plots, as detailed in Table 1.

Table 1. Number of samples of 4 rice diseases

Disease Type	Sample Count/Image
Rice Bacterial Leaf Blight	761
Rice Blast Disease	911
Rice Brown Spot Disease	858
Rice Sheath Blight	592
Total	3122

The dataset for this study was captured under natural field conditions. To enhance the diversity of the dataset and improve the network's generalization capability, images were specifically collected under complex background interference conditions, including uneven brightness and lighting, folded leaves, blurred diseased areas, leaf edges blending with the background, water droplets on leaves, and soil. All collected image samples were captured in real agricultural environments, featuring intricate backgrounds, varying lesion scales, and numerous small targets that significantly increase recognition difficulty. However, research on disease recognition in complex backgrounds better prepares for the practical needs of disease segmentation in actual farmland settings.

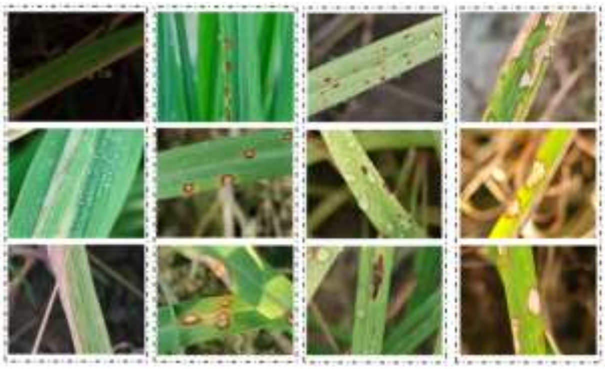

(a) Rice Bacterial Leaf Blight (b) Rice Blast Disease (c) Rice Brown Spot Disease (d) Rice Sheath Blight

Figure. 1 Some rice leaf disease pictures collected in actual field environment

During image preprocessing, the raw resolution of collected images (5792×4344) is large, with excessive background coverage. To highlight target areas, non-essential background interference must be minimized. Direct application of segmentation models would exhaust computer memory and GPU resources, necessitating image cropping. Under the guidance of plant protection experts, cropping was performed to preserve the complete outline of the entire leaf containing the disease lesion while removing other redundant parts. Images of different rice types were uniformly cropped. This approach effectively retained details of rice leaves and their diseased areas while reducing computational complexity for the recognition algorithm. It also enabled assessment of accuracy and uniqueness, forming the sample set as shown in Figure 1.

2.2 Data Augmentation Techniques

Due to the uneven sample sizes across different categories of rice leaf diseases during data collection, the scarcity of sample data for rice sheath blight can easily lead to overfitting during model training. In this scenario, data augmentation[14] serves as an effective method to enhance the quality of the training set, thereby improving the model training process. Increasing dataset diversity enables the model to comprehensively learn disease characteristics. Through geometric transformations such as image flipping, translation, scaling, and rotation, the model achieves sample augmentation for identifying leaf diseases on rice plants at different growth stages and from various angles. This enhances the model's generalization performance in real-world scenarios. Image transposition adjusts spatial relationships by swapping rows and columns, altering the orientation of disease patterns within images. This expands the model's learning of disease characteristics across multiple angles, enabling better handling of rice leaf disease images captured at varied perspectives. Consequently, it enhances the model's stability when processing field-based rice leaf disease imagery. Under natural field conditions, lighting environments vary significantly due to weather and time, resulting in differences in brightness and contrast across collected images. Randomly adjusting brightness and contrast simulates conditions across multiple lighting scenarios, improving the model's adaptability to varying light intensities. In practice, a 30% adjustment range for brightness and contrast is employed. This ensures the disease detection module maintains stable performance in identifying affected areas even under drastically changing light conditions. In real-world field conditions, leaves may be obscured by soil, other crops, or agricultural machinery, leading to loss of disease information in the measured area. The random field occlusion method simulates actual field scenarios by randomly discarding rectangular regions in field images. This ultimately endows the network with the ability to accurately identify disease areas even when rice leaf disease information is lost, demonstrating strong interference resistance. This study employs diverse data augmentation techniques with probabilistic settings to ensure randomness and variability in augmentation, as illustrated in Figures 2-7 (augmented images). This approach not only expands the dataset but also exposes the model to a wider range of transformed disease images during training, enhancing its adaptability. These augmentations increase training complexity, enabling the model to learn features in more challenging environments. Ultimately, the model's segmentation performance in complex backgrounds and varying lighting conditions was optimized, improving its adaptability to real-world agricultural scenarios.

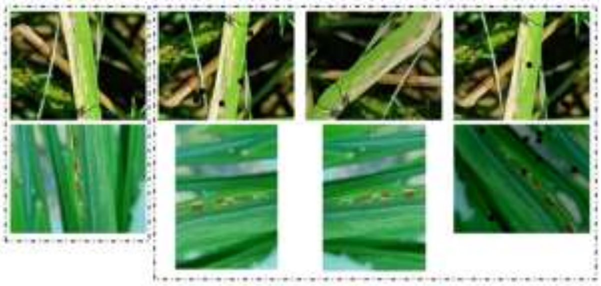


Figure. 2 Image after data enhancement

2.3 Dataset Partitioning

The data-enhanced and deduplicated sample dataset was randomly sampled from each rice leaf disease at a ratio of 7:2:1 to form the training set, test set, and validation set, respectively. The training set comprised 9,548 samples, the test set contained 2,728 samples, and the validation set included 1,365 samples, with no overlap among the three sets. The specific distribution is shown in Table 2. The training set is used for supervised learning of the model, serving as the data samples for model fitting. The validation set is used to adjust the model's hyperparameters and conduct a preliminary evaluation of the model's performance. The test set is used to test the classification capability of the trained model and to evaluate the overall performance of the final model.

Table 2. Rice disease image sample set partitioning

Disease Type	Training set/image	Test set/image	Validation set/image
Rice Bacterial Leaf Blight	2128	608	304
Rice Blast Disease	2543	727	363
Rice Brown Spot Disease	2398	685	343
Rice Sheath Blight	2479	708	355
Total	9548	2728	1365

3. Segmentation Network Based on Multi-Scale Feature Fusion Using Swin Transformer

3.1 Model Architecture

This paper addresses the challenge of semantic segmentation for rice leaf diseases in complex environments by proposing an improved method based on the Swin Transformer backbone network and a dual-attention-enhanced dilated convolution module. Rice leaf diseases in complex environments are often disrupted by factors such as lighting, shadows, and cluttered backgrounds, making segmentation tasks more challenging. To address this, this paper adopts the Swin Transformer as the backbone network, fully leveraging its efficient extraction of local and global features to enhance segmentation accuracy of diseased areas in complex scenarios. By strengthening the capture of multi-scale contextual information, it facilitates the isolation of diseased regions within complex backgrounds.

The incorporation of Swin Transformer[15] during feature extraction significantly enhances the detection of disease regions on rice leaves. The dual attention mechanism—channel attention and positional attention—strengthens dimensional information and spatial context, respectively. The expanded dilated convolution mechanism enhances the recognition of rice blast and brown spot lesions. This approach more effectively extracts relevant information between image pixels while suppressing the influence of background information on lesion features, thereby improving segmentation accuracy and yielding more stable results. The decoder stage reconstructs features extracted by the encoder. Encoder feature maps are upsampled to restore the original image size, and cascaded Concat operations fuse features from different layers to fully propagate information. A 3×3 convolution refines the image features, and a Softmax operation yields the segmentation result. Loss functions were optimized for brown spot and rice blast lesions to enhance segmentation accuracy for both diseases. Incorporating the Swin Transformer provides multi-scale feature extraction capabilities, while channel and positional self-attention mechanisms enhance the fusion of deep features.

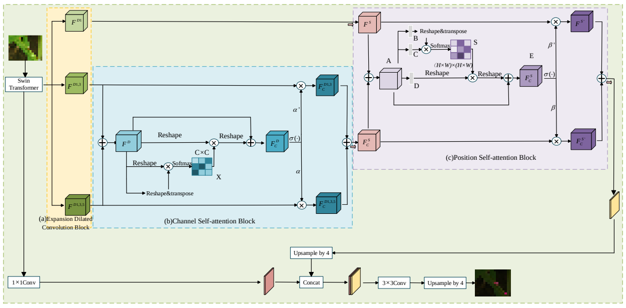


Figure. 3 SwinDA-DeepLabv3+ network structure

3.2 Backbone Network

The Swin Transformer adopts a hierarchical structure, with the network divided into four stages: Stage 1, Stage 2, Stage 3, and Stage 4. As the network deepens, the height and width of feature maps progressively decrease while the number of channels continuously increases. This enables each stage to extract multi-scale features, enhancing the model's ability to represent image information across different resolutions. This architecture leverages self-attention mechanisms to extract image features both within local windows and across windows. This approach not only reduces computational overhead but also strengthens the model's feature modeling capabilities. Traditional Transformers suffer from computational complexity that scales exponentially with image size when computing global self-attention, making high-resolution image processing inefficient. The Swin Transformer addresses this by introducing a sliding window mechanism, confining attention calculations to non-overlapping local windows while establishing cross-window information flow through inter-window interactions. This enables the model to maintain computational efficiency while effectively capturing connections between local and global features. Compared to CNNs, Swin Transformer inherits the advantages of hierarchical structures while offering greater flexibility in feature representation. Through module stacking, the model constructs multi-scale feature maps that adapt to more complex image information and demonstrate stronger learning capabilities when data exhibits hierarchical features.

First, the input image undergoes segmentation, dividing the H×W×3 image into 4×4 image patches. These patches are then flattened along the channel dimension, resulting in a channel dimension of 4×4×3=48 and an image size of H/4×W/4×48. Stage 1 adjusts the feature dimension of each image block to C via a Linear Embedding operation, then feeds it into two Swin Transformer Blocks for processing. Stage 2 halves the resolution through image block fusion while increasing the channel count to 2C, followed by another pass through Swin Transformer Blocks to compute self-attention features. Stage 3 repeats the same process, yielding features of size H/16×W/16×4C, which undergo further processing through six Swin Transformer Blocks. Finally, Stage 4 reduces the resolution to H/32×W/32 while increasing the channel count to 8C. Two Swin Transformer Blocks generate multi-scale feature maps, providing hierarchical feature representations for subsequent tasks.

3.3 Dual Attention Expansion Inflated Convolution Module

To enhance segmentation accuracy for rice leaf diseases, this paper proposes the Dual Attention Expanded Dilated Convolution Module (DAEDCM) to address feature extraction errors caused by field soil, water droplets, light reflection, and other disturbances. This module integrates multi-scale features with a dual attention mechanism, comprising three components: the Extended Dilated Convolution Module (EDCB), the Channel Self-Attention Module (CSB), and the Position Self-Attention Module (PSB). Through their collaborative focus on disease features and suppression of background interference, they ensure segmentation accuracy in complex environments.

Specifically, the EDCB achieves multi-scale feature capture by employing three 3×3 convolutional kernels with distinct dilation rates (1, 3, 5): At dilation rate 3, the receptive field expands to 11×11, simultaneously capturing disease region information and contextual data from adjacent areas; At dilation rate 5, the receptive field further expands to 19×19, integrating local and global features to capture long-range dependencies between lesions and background within small areas. Fusion of these three outputs significantly enhances the network's ability to extract information from complex rice leaf disease images in the field.

CSB is employed to mitigate complex background interference. Its core mechanism involves enhancing disease-related channels and suppressing irrelevant background channels through channel-wise similarity calculations. The computation proceeds as follows: input feature maps undergo weighted summation to generate preliminary processed features. These features are then optimized via reshaping and transposition operations to adapt the structure for similarity calculations. The reshaped feature maps are then multiplied to generate a similarity matrix measuring inter-channel correlations. Softmax normalization constrains channel attention values within 0–1 to ensure reasonable weights. This correlation matrix is multiplied by the original feature map to redistribute attention weights and produce an enhanced feature map. Finally, sigmoid activation yields a channel attention map quantifying channel importance.

The objective of PSB is to generate attention weights in the spatial dimension, enhancing the locational features of diseased regions. First, the outputs from the previous stage are normalized using 1×1 convolution kernels and processed through LeakyReLU activation functions to yield two feature maps. These are then weighted and fused to produce a new feature map. This feature map undergoes processing through three convolutional layers, generating matrices B, C, and D, which are reshaped to suit spatial similarity calculations. The transpose of B is multiplied by C to form a spatial similarity matrix, normalized via softmax to produce a spatial attention map; This attention map weights spatial positions in the feature map, enhancing disease regions while suppressing background noise. Transposing the attention map and multiplying it by matrix D yields weighted features, which are then added to the original feature map to generate the final output feature map incorporating positional information.

Ultimately, the channel attention map generated by CSB and the spatial attention map generated by PSB work synergistically. The spatial attention map calibrates the spatial dimension information of the feature map, while the channel attention map calibrates the channel dimension information. This enables the model to adaptively adjust its feature selection strategy based on the spatial distribution of rice leaf diseases, further enhancing the accuracy of disease area identification in complex field environments.

3.4 Loss Function Optimization

For semantic segmentation tasks, particularly when dealing with complex and imbalanced structures in rice leaf disease regions, the choice of loss function significantly impacts both training time and segmentation accuracy of the final model. To better detect blast disease and brown spot disease areas while minimizing unnecessary errors caused by background noise, a well-chosen loss function plays a crucial role in model training.

Dice Loss[16] addresses challenges in handling sample imbalance and the segmentation of small-area rice leaf disease regions. By detecting overlap between predicted and ground-truth regions, the DiceLoss function encourages the model to thoroughly learn detailed boundary features of the disease. Small field lesions are difficult to extract effectively against complex backgrounds. The Dice Loss function enhances the model's sensitivity to small lesions by maximizing the intersection between predicted and actual regions, thereby mitigating misclassifications caused by category imbalance and enabling precise segmentation of minute areas. The Dice Loss formula is:

$$DiceLoss = 1 - \frac{2 \times \sum_{i=1}^{N} p_i \times g_i}{\sum_{i=1}^{N} p_i^2 + \sum_{i=1}^{N} g_i^2}$$

Among them, N is the total number of image pixels, p_i represents the predicted value of the i pixel by the model, and g_i represents the true label value of this pixel.

Focal Loss[17] is a loss function proposed to address class imbalance issues. For disease-affected areas that are difficult to distinguish, this loss function applies weighting to better focus on these regions. This enables the model to prioritize segmentation in such areas, thereby enhancing detection capabilities for small targets. In field rice paddies, rice blast and brown spot disease areas are small and easily obscured by other objects. Consequently, models may overlook these disease patches during training, leading to misclassification. FocalLoss reduces the weight assigned to easily classified samples, compelling the model to concentrate on the small, difficult-to-classify disease spots. This approach significantly improves segmentation accuracy, particularly in scenarios with complex backgrounds.

$$FocalLoss = -\alpha_t (1 - p_t)^{\gamma} \log(p_t)$$

Where p_t is the probability of the disease category, α_t is the balancing factor, $(1 - p_t)^{\gamma}$ is the modulation factor, and α ≥0 is an adjustable focusing parameter.

In summary, the combined use of Dice Loss and Focal Loss effectively addresses the challenges of sample imbalance and small target recognition in rice leaf disease segmentation. By optimizing the loss function, the model achieves higher segmentation accuracy in complex backgrounds, balances performance across global and local regions, and ultimately enhances overall segmentation quality.

$$TotalLoss = \alpha DiceLoss + \beta FocalLoss$$

Where α and β are weighting coefficients that determine the relative importance of the two loss functions.

4. Experimental Results and Analysis

4.1 Parameter Settings

The experimental environment configuration is listed in Table 3. This chapter implements the SwinDA-DeepLabv3+ segmentation network using Python 3.7 and PyTorch 1.11.0. The hardware environment includes an NVIDIA RTX 3060 Ti GPU with CUDA 10.0 and CUDNN 7.6.0 support, running on Windows 10. Input images are uniformly set to dimensions of 512×512×3. The Adam optimizer is employed with a momentum parameter of 0.9, a maximum learning rate of 1e-4, and a minimum learning rate set to 1% of the maximum learning rate. Cosine annealing is used for learning rate decay.

Table 3. Test environment configuration

Type	Experimental Parameters
Operating System	Windows 10
Programming Language	Python 3.7
Training Framework	PyTorch 1.11.0
Hardware GPU Model	NVIDIA RTX 3060 Ti
GPU	CUDA 10.0, CUDNN 7.6.0

4.2 SwinDA-DeepLabv3+ Performance

4.2.1 Comparison of Segmentation Metrics Across Different Disease Types

Table 4 presents the segmentation results comparison for the rice leaf disease dataset using UNet, DeepLabv3+, PSPNet, HRNet, and SwinDA-DeepLabv3+ networks. The experiments employed mean Intersection over Union (mIoU) as the evaluation metric, which reflects the overlap between the model's predicted segmentation regions and the ground truth segmentation regions.

As shown in Table 4, the SwinDA-DeepLabv3+ network achieved the highest disease segmentation performance with an mIoU of 83.34%. Its mIoU surpassed DeepLabv3+ by 4.68%, PSPNet by 10.66%, and HRNet by 11.45%. SwinDA-DeepLabv3+ demonstrated particularly outstanding performance in segmenting four types of diseases: bacterial leaf blight, rice blast, brown spot disease, and sheath blight. This indicates the network's advantage in handling complex rice leaf disease segmentation tasks. By integrating multi-scale feature information, SwinDA-DeepLabv3+ significantly enhances the feature representation capabilities of the segmentation model. This enables the network to better capture and utilize detailed characteristics of the diseases, thereby improving overall network performance.

Table 4. Comparison of Part Segmentation IoU Results on ShapeNet Part Dataset (Unit: %)

Network Model	Bacterial Leaf Blight	Blast	Brown spot	Sheat blight
UNet	70.24	70.01	56.68	78.18
DeepLabv3+	78.86	71.80	60.85	82.56
PSPNet	69.47	68.02	50.19	76.56
HRNet	68.23	65.16	52.45	74.54
SwinDA-DeepLabv3+	83.92	79.97	68.14	85.34

Figure 3 illustrates the IoU variation trends across different deep learning models in the rice leaf disease segmentation task. The horizontal axis represents different categories, including “background,” “bacterial blight,” “rice blast,” “brown spot,” and “sheath blight.” The vertical axis denotes IoU (%), indicating the segmentation accuracy of the model for each category.

Figure 4 reveals that SwinDA-DeepLabv3+ achieves the highest IoU across all categories, demonstrating superior performance particularly in complex environments. DeepLabv3+ follows closely with overall strong results, especially in segmenting “bacterial blight” and “sheath blight.” In contrast, HRNet and PSPNet exhibit lower IoU across multiple categories, with notably poor segmentation performance in “rice blast” and “brown spot” categories. Collectively, this chart visually illustrates the relative strengths and weaknesses of different models in rice disease segmentation tasks. SwinDA-DeepLabv3+ stands out due to its outstanding performance across all categories.

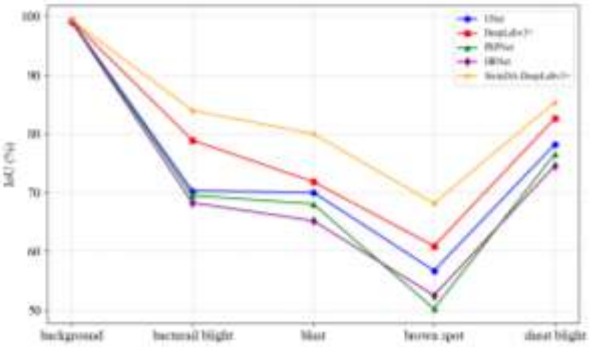


Figure. 4 Comparison of IoU results for spots of different diseases

4.2.2 Comparison of Segmentation Evaluation Metrics Across Different Network Models

Table 5 presents experimental results comparing the SwinDA-DeepLabv3+ network with mainstream networks on segmentation tasks.

As shown in Table 5, SwinDA-DeepLabv3+ outperforms other networks across mIoU, mPA, Precision, and Recall metrics, demonstrating superior segmentation performance. Specifically, SwinDA-DeepLabv3+ achieves an mIoU of 83.34%, representing a 4.68 percentage point improvement over the best baseline model, DeepLabv3+. Furthermore, its mPA reaches 87.53%, surpassing DeepLabv3+ by 1.51 percentage points and outperforming UNet, PSPNet, and HRNet. Regarding segmentation accuracy, SwinDA-DeepLabv3+ achieved a Precision of 91.20%, surpassing DeepLabv3+ by 2.06 percentage points and significantly outperforming UNet, PSPNet, and HRNet. This indicates lower misclassification rates in identifying diseased regions. Furthermore, SwinDA-DeepLabv3+ achieved a Recall of 87.01%, an improvement of 0.99 percentage points over DeepLabv3+, demonstrating stronger capability in detecting diseased areas. This enables coverage of more affected regions and effectively reduces the false negative rate.

SwinDA-DeepLabv3+ builds upon DeepLabv3+ by integrating the Dual Attention Expanded Dilated Convolution Module (DAEDCM) to enhance feature extraction. It further improves segmentation performance for rice leaf diseases through: - Multi-scale information capture via Extended Dilated Convolution (EDCB) - Spatial dependency enhancement with Position Self-Attention (PSB) - Feature selection reinforcement using Channel Self-Attention (CSB) Overall, SwinDA-DeepLabv3+ enables the network to better handle rice leaf disease segmentation tasks in complex backgrounds through in-depth modeling of multi-scale features and optimized attention mechanisms.

Table 5. Comparison of segmentation indexes of different network models

Network Model	mIoU/ %	mPA/ %	Precision /%	Recall/ %
UNet	74.86	83.37	86.28	83.37
DeepLabv3+	78.66	86.02	89.14	86.02
PSPNet	72.68	84.26	82.44	84.26
HRNet	71.89	81.63	83.98	81.63
SwinDA-DeepLabv3+	83.34	87.53	91.20	87.01

Figures 5 display the mIoU curves of different deep learning models over epochs during training. As shown, all models exhibit gradually increasing mIoU with epochs, indicating continuous improvement in segmentation performance during training. Among them, SwinDA-DeepLabv3+ demonstrates the fastest mIoU growth, rising rapidly in the early stages before stabilizing and ultimately achieving the highest mIoU, highlighting its superior segmentation capability. DeepLabv3+ also achieves a high mIoU, second only to SwinDA-DeepLabv3+, and converges toward stability in the later training stages. PSPNet and UNet exhibit similar mIoU performance, ultimately stabilizing around 70%. In contrast, HRNet demonstrates a relatively low mIoU, converges more slowly, and ultimately fails to reach the mIoU levels achieved by the other models.

Overall, SwinDA-DeepLabv3+ demonstrated the best performance throughout the training process, consistently achieving higher mIoU than other models. This indicates its superior suitability for datasets of rice leaf diseases in complex environments.

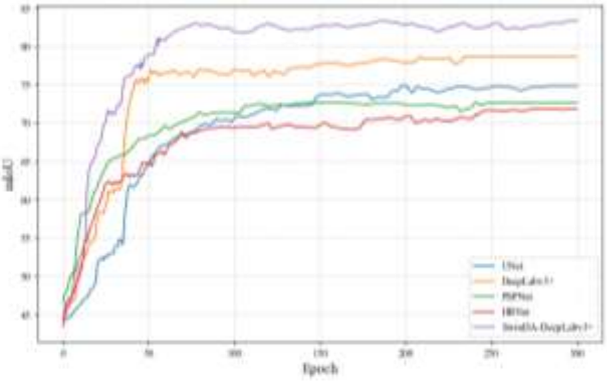


Figure. 5 Graph showing the variation of mIoU for different networks

4.3 Loss Function Analysis

Addressing the challenge of segmenting rice leaf diseases against complex backgrounds, the small lesion areas of rice blast and brown spot diseases, coupled with their minimal contrast against the background, caused severe class imbalance that significantly disrupted model learning. FocalLoss mitigated this imbalance by adjusting parameters, thereby enhancing the model's focus on distinguishing between difficult-to-separate samples.

Figure 6 below illustrates the loss decline trend during training for various semantic segmentation models: UNet, DeepLabv3+, PSPNet, HRNet, and SwinDA-DeepLabv3+. The horizontal axis represents training epochs, while the vertical axis shows the loss value of the hybrid loss function. Overall, the loss of all models gradually decreases as training iterations increase, indicating continuous learning and parameter optimization. The steepest decline occurs within the first 50 training epochs, after which the loss stabilizes—a pattern consistent with the training characteristics of deep learning models.

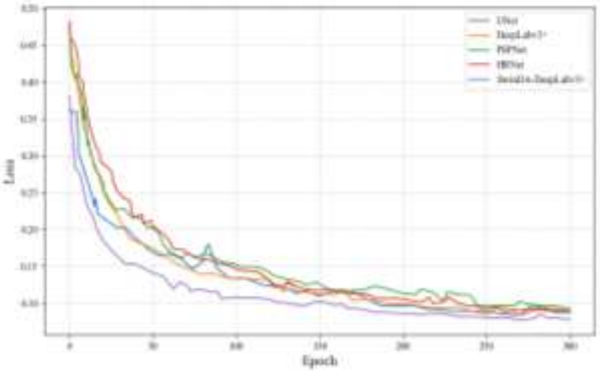


Figure. 6 Comparison of Training Loss Across Different Semantic Segmentation Models

Comparing model performance, SwinDA-DeepLabv3+ exhibits the lowest loss values throughout training and converges most rapidly, indicating its superior performance in rice disease segmentation. DeepLabv3+ and PSPNet also show rapid loss reduction, eventually stabilizing at lower levels. In contrast, HRNet demonstrates higher loss values and significant oscillations during training, attributed to its complex architecture leading to reduced training stability. UNet exhibited a high initial loss value. Although it showed a clear downward trend, its final loss value remained higher than that of other models.

In terms of training stability, SwinDA-DeepLabv3+ displayed the smoothest curve, indicating a more stable training process and stronger generalization capability. UNet and HRNet experienced some oscillation during training, which is related to learning rate settings, dataset complexity, and model architecture. Overall, SwinDA-DeepLabv3+ demonstrates superior performance in loss convergence speed, final stability, and training stability, making it the most promising model for this task.

4.4 Melting Experiment

This study employs DeepLabv3+ as the baseline network. Table 6 presents ablation experiments conducted to validate the feasibility of the proposed SwinDA-DeepLabv3+ model. Scheme 1: Original DeepLabv3+ baseline model; Scheme 2: Integrates the Swin Transformer architecture onto the original DeepLabv3+ foundation; Scheme 3: Replaces the original DeepLabv3+ feature pyramid module with the Extended Dilated Convolution Block (EDCB); Scheme 4: Combines Schemes 2 and 3, introducing the Channel Self-Attention Block (CSB) and Position Self-Attention Block (PSB); Scheme 5: Applies a hybrid loss function combining FocalLoss and DiceLoss for model training based on Scheme 4.

Among the results of the five ablation experiments, Scheme 2 achieved a 0.87% improvement over Scheme 1 in mPA and a 0.28% improvement in mIoU. This improvement indicates that the Swin Transformer architecture enhances the model's ability to capture global features, thereby improving its adaptability in complex environments. Comparing Scheme 1 and Scheme 3, the results show that Scheme 3 only achieved a 0.19%

improvement in mPA. The above experiments demonstrate that the EDCB architecture plays a role in enhancing local feature extraction and segmentation accuracy, particularly for complex rice leaf disease regions. To further analyze model optimization, we compared the performance between Approach 1 and Approach 4. Approach 4 achieved a 1.22% improvement in mPA and a 1.46% improvement in mIoU, indicating that the dual-attention structure enhances feature representation and increases the model's focus on diseased regions. Based on this, we further analyzed the performance difference between Approach 4 and Approach 5. Approach 5 employs additional FocalLoss and DiceLoss to address class imbalance. Experimental comparisons show that Scheme 5 achieves a 0.29% increase in mPA and a 4.68% improvement in mIoU. This demonstrates that the dual optimization of FocalLoss and DiceLoss enhances the model's learning of small target regions, mitigates misclassification of small targets caused by category imbalance, and further improves segmentation performance. In summary, after integrating all techniques in Approach 5, the mPA reached 87.53% and mIoU reached 83.34%, achieving the optimal segmentation results. This experiment provides practical references and theoretical basis for optimizing segmentation models for rice leaf diseases.

Table 6. Results of ablation experiment

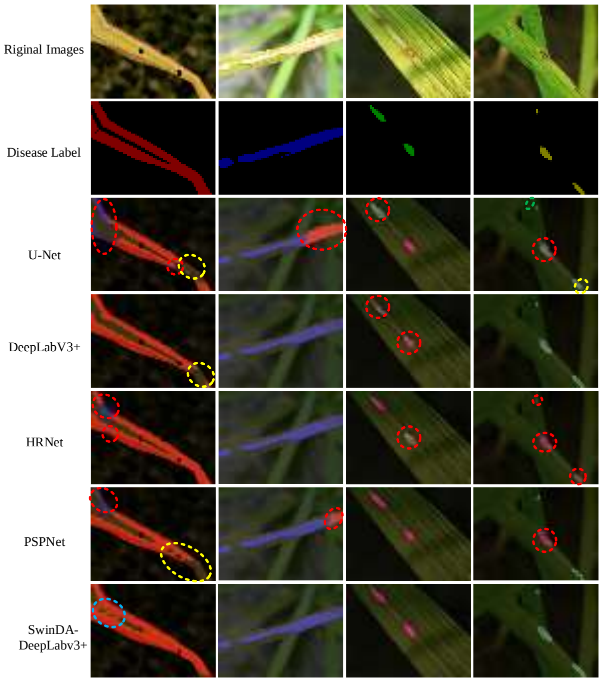
Model	mPA (%)	mIoU (%)	Params (MB)	FLOPS (G)	FPS
1	86.02	78.66	53.72	175.34	20.16
2	86.89	78.94	52.91	148.27	18.52
3	86.21	77.65	54.20	162.15	19.35
4	87.24	80.12	53.30	135.82	16.48
5	87.53	83.34	53.30	115.41	15.33

Among the results of the five ablation experiments, Scheme 2 achieved a 0.87% improvement over Scheme 1 in mPA and a 0.28% improvement in mIoU. This improvement indicates that the Swin Transformer architecture enhances the model's ability to capture global features, thereby improving its adaptability in complex environments. Comparing Scheme 1 and Scheme 3, the results show that Scheme 3 only achieved a 0.19% improvement in mPA. The above experiments demonstrate that the EDCB architecture plays a role in enhancing local feature extraction and segmentation accuracy, particularly for complex rice leaf disease regions. To further analyze model optimization, we compared the performance between Approach 1 and Approach 4. Approach 4 achieved a 1.22% improvement in mPA and a 1.46% improvement in mIoU, indicating that the dual-attention structure enhances feature representation and increases the model's focus on diseased regions. Based on this, we further analyzed the performance difference between Approach 4 and Approach 5. Approach 5 employs additional FocalLoss and DiceLoss to address class imbalance. Experimental comparisons show that Scheme 5 achieves a 0.29% increase in mPA and a 4.68% improvement in mIoU. This demonstrates that the dual optimization of FocalLoss and DiceLoss enhances the model's learning of small target regions, mitigates misclassification of small targets caused by category imbalance, and further improves segmentation performance. In summary, after integrating all techniques in Approach 5, the model achieves an mPA of 87.53% and an mIoU of 83.34%, delivering optimal segmentation results. This experiment provides practical guidance and theoretical basis for optimizing segmentation models for rice leaf diseases.

4.5 Visual Experiment

4.5.1 Comparison of Segmentation Results for Different Disease Lesions

To visually demonstrate segmentation performance on rice leaf diseases, this paper compares four semantic segmentation networks on rice bacterial leaf blight (dark red), rice sheath blight (purple), rice blast disease (magenta), and rice brown spot disease (grayish blue), as shown in Figures 7. Red circles mark incorrectly segmented disease areas, yellow circles indicate missed segments, and green circles highlight oversegmented regions.



(a) Bacterial Leaf Blight (b) Sheath Blight (c) Blast Disease (d) Brown Spot Disease

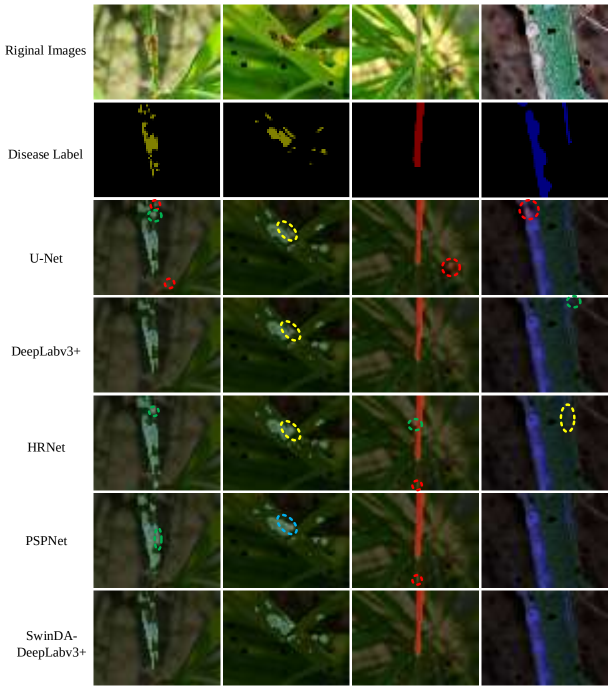
Figure. 7 Segmentation results of different diseases

In the rice bacterial leaf blight segmentation task, SwinDA-DeepLabv3+ delivered generally good results with only minor instances of lesion adhesion. U-Net, HRNet, and PSPNet all exhibited segmentation omissions, misclassifying rice bacterial leaf blight as rice sheath blight. Additionally, U-Net, DeepLabv3+, and HRNet exhibited segmentation omissions at rice leaf vein curling areas. In the rice sheath blight segmentation task, SwinDA-DeepLabv3+ maintained stable performance with excellent lesion segmentation. It accurately distinguished lesions against complex leaf texture backgrounds while avoiding misclassifications, yielding more precise segmentation results. U-Net and PSPNet, however, frequently misclassified sheath blight symptoms as bacterial leaf blight. This confusion arises from the high similarity in symptoms and coloration between the two diseases, coupled with insufficient discriminative power in their disease classification models. DeepLabv3+ and HRNet delivered generally good segmentation results, though oversegmentation persisted near leaf edges. For rice blast disease segmentation, SwinDA-DeepLabv3+ again performed well, demonstrating strong lesion segmentation and recognition capabilities with improved edge segmentation. This effectively minimized misclassifications, yielding more complete segmentation of diseased areas. PSPNet demonstrated relatively stable overall performance, accurately capturing lesion areas in segmentation results, though its edge segmentation quality was slightly inferior to SwinDA-DeepLabv3+. Regarding disease segmentation, U-Net, DeepLabv3+, and HRNet all exhibited segmentation errors in certain regions, incorrectly classifying portions of background as diseased areas. SwinDA-DeepLabv3+ and DeepLabv3+ demonstrated superior segmentation for rice brown spot disease, exhibiting high accuracy in lesion segmentation while maintaining good boundary continuity. They exhibited minimal mis-segmentation or omission of lesions, enhancing the integrity of the lesion regions. U-Net and PSPNet showed weaker lesion detection capabilities, leading to mis-segmentation of small lesions. HRNet exhibited severe misclassification, with some lesions incorrectly identified as rice blast disease, compromising segmentation quality.

Overall, SwinDA-DeepLabv3+ delivered the best segmentation performance for rice leaf diseases in complex backgrounds, with improved edge detail and significantly reduced misclassification and omission. This demonstrates the network's suitability for precise semantic segmentation diagnosis of rice leaf diseases.

4.5.2 Comparison of Segmentation Results Under Different Influencing Factors

Segmenting rice leaf diseases under natural light faces numerous challenges from field environmental factors. Noise from soil backgrounds, self-occlusion of leaves, overexposure or shadows caused by light variations, and interference from water droplet reflections all increase segmentation difficulty and reduce the generalization capability of segmentation models.



(a) Complex Soil (b) Leaf Occlusion (c) Complex Illumination(d) Water Droplet Influence

Figure. 8 The results were divided under different influencing factors

This study analyzes the impact of soil background, leaf occlusion, lighting effects, and water droplet interference on segmentation using rice leaf disease images captured under natural light. Different models are compared, with experimental results shown in Figures 8. Red circles mark incorrectly segmented disease areas, yellow circles indicate missed disease segments, blue circles highlight disease overlap regions, and green circles denote over-segmented disease areas. Results demonstrate that SwinDA-DeepLabv3+ exhibits stable segmentation capabilities in complex environments, particularly showing strong resistance to interference from soil backgrounds, leaf occlusions, lighting variations, and water droplet interference.

In complex backgrounds, this model effectively suppresses misclassification of non-diseased areas, with superior edge handling compared to DeepLabV3+ and PSPNet, and clearer segmentation boundaries than U-Net. Under leaf-obscured conditions, SwinDA-DeepLabv3+ accurately identified small lesions while effectively reducing missed segmentation and lesion boundary coalescence. However, DeepLabV3+, U-Net, and HRNet exhibited slight deficiencies in handling minute lesion boundaries on rice leaves, resulting in missed segments. PSPNet's segmentation results showed coalescence of diseased regions. Under varying lighting conditions, SwinDA-DeepLabv3+ demonstrated outstanding interference resistance, precisely distinguishing between disease and bright areas while significantly reducing misclassifications. In contrast, DeepLabV3+ showed slightly lower stability, U-Net misclassified bright areas as disease, and both HRNet and PSPNet performed less effectively than SwinDA-DeepLabv3+ under light interference. When confronted with water droplet interference, this model avoided misclassifications caused by reflections. DeepLabV3+ exhibited minor misclassifications in areas where water droplets met with dead leaves. U-Net's visualization results showed segmentation errors, while HRNet's segmentation results exhibited omissions. PSPNet demonstrated greater stability than U-Net but slightly inferior performance in handling disease edges compared to SwinDA-DeepLabv3+. However, SwinDA-DeepLabv3+ incorrectly identified water droplets along rice leaf margins as sheath blight lesions in its segmentation results.

5. Discussion

Addressing the core challenge of distinguishing diseased areas in rice leaf disease segmentation tasks within complex farmland environments, this chapter proposes a semantic segmentation network, SwinDA-DeepLabv3+, achieving high-precision rice leaf disease segmentation through multi-scale feature fusion, global context modeling, and the introduction of a hybrid loss function. The backbone network employs the Swin Transformer architecture to enhance the model's ability to capture long-range dependencies. Within the dual-attention expanded dilated convolution module, background noise interference is suppressed through the introduction of the Expanded Dilated Convolution Block (EDCB), Position Self-Attention Block (PSB), and Channel Self-Attention Block (CSB), enhancing the model's adaptability to complex backgrounds. Furthermore, the loss function optimization strategy combines the Dice-Focal hybrid loss function to mitigate the imbalance in rice leaf disease samples, thereby improving segmentation accuracy. Experimental results demonstrate that the SwinDA-DeepLabv3+ network achieves outstanding performance across all metrics. achieving mIoU, mPA, Precision, and Recall of 83.34%, 87.53%, 91.20%, and 87.01%, respectively. Compared to other network architectures, SwinDA-DeepLabv3+ demonstrates superior performance in rice leaf disease segmentation tasks, validating its potential for semantic segmentation of rice leaf diseases in complex environments.

6. Conclusion

Deep learning enhances crop disease identification and improves monitoring efficiency. Rice leaf disease segmentation faces challenges due to diverse leaf morphologies, large variations in lesion sizes, and background interference, all limiting accuracy.This study proposes the SwinDA-DeepLabv3+ network: a Swin Transformer backbone captures global information, a DADCM module strengthens lesion attention, and Dice-Focal loss addresses sample imbalance. Experiments validate its segmentation effectiveness, adaptability to complex backgrounds, and efficient training.The model's limitation lies in lower segmentation accuracy for structurally complex lesions or those with minimal background contrast, as some lesions share similar leaf shapes and textures with the background.Future work may incorporate multimodal data such as hyperspectral and thermal infrared imagery to augment feature dimensions and enhance recognition accuracy.

7. References

[1] Reddy, C. A., Oraon, S., Bharti, S. D., et al. 2024. Advancing disease management in agriculture: A review of plant pathology techniques. Plant Science Archives, 9, 16-18.

[2] Lü, X. K. 2024. Identification and control techniques of major rice diseases and pests. Agricultural Science & Technology and Development, 3(06), 58-60.

[3] Shoaib, M., Shah, B., Ei-Sappagh, S., et al. 2023. An advanced deep learning models-based plant disease detection: A review of recent research. Frontiers in Plant Science, 14, 1158933.

[4] Shafik, W., Tufail, A., Namoun, A., et al. 2023. A systematic literature review on plant disease detection: Motivations, classification techniques, datasets, challenges, and future trends. IEEE Access, 11, 59174-59203.

[5] Li, W., Zhu, L., Liu, J. 2024. PL-DINO: An Improved Transformer-Based Method for Plant Leaf Disease Detection. Agriculture, 14(5), 691.

[6] Correspondent. 2024. Ministry of Agriculture and Rural Affairs issues "National Smart Agriculture Action Plan (2024-2028)". Jiangsu Agricultural Mechanization, (06), 6-9.

[7] Chen, C., Wang, X., Heidari, A. A., et al. 2021. Multi-threshold image segmentation of maize diseases based on elite comprehensive particle swarm optimization and otsu. Frontiers in Plant Science, 12, 789911.

[8] Rahmawati, A., Yulianti, I., Nurajizah, S. 2023. Image Segmentation Analysis Using Otsu Thresholding and Mean Denoising for the Identification Coffee Plant Diseases. Jurnal Riset Informatika, 6(1), 7-14.

[9] Yang, T., Zhou, S., Xu, A., et al. 2023. An approach for plant leaf image segmentation based on YOLOV8 and the improved DEEPLABV3+. Plants, 12(19), 3438.

[10] Shoaib, M., Hussain, T., Shah, B., et al. 2022. Deep learning-based segmentation and classification of leaf images for detection of tomato plant disease. Frontiers in Plant Science, 13, 1031748.

[11] Yang, T., Wang, Y., Lian, J. 2024. Plant Diseased Lesion Image Segmentation and Recognition Based on Improved Multi-Scale Attention Net. Applied Sciences, 14(5), 1716.

[12] Elmessery, W. M., Maklakov, D. V., El-Messery, T. M., et al. 2024. Semantic segmentation of microbial alterations based on SegFormer. Frontiers in Plant Science, 15, 1352935.

[13] Zhao, J. F. 2023. Research on rice leaf disease recognition algorithms for complex field environments. Doctoral Thesis. Guangdong Polytechnic Normal University.

[14] Fujii, Y., Uchida, D., Sato, R., et al. 2024. Effectiveness of data-augmentation on deep learning in evaluating rapid on-site cytopathology at endoscopic ultrasound-guided fine needle aspiration. Scientific Reports, 14(1), 22441.

[15] Kumar, A., Yadav, S. P., Kumar, A. 2025. An improved feature extraction algorithm for robust Swin Transformer model in high-dimensional medical image analysis. Computers in Biology and Medicine, 188, 109822.

[16] Zheng, Y., Tian, B., Yu, S., et al. 2025. Adaptive boundary-enhanced Dice loss for image segmentation. Biomedical Signal Processing and Control, 106, 107741.

[17] Zhong, R., Wang, C., Song, Y. F., et al. 2024. Few-shot face recognition method based on sample-balanced distillation. Computer Engineering and Design, 45(11), 3457-3462.