Journal of Electronic Research and Application, 2026, Volume 10, Issue 3
http://ojs.bbwpublisher.com/index.php/JERA
ISSN Online: 2208-3510
ISSN Print: 2208-3502

Enhancing Tea Leaf Disease Classification with Cross-Attention Fusion and Magnitude-Aware Linear Attention

Jiaxin Zhu*

Department of Electronic and Information Engineering, Liaoning Technical University, Huludao 125100, China

***Corresponding author:** *Jiaxin Zhu, 2306110233@stu.lntu.edu.cn*

Abstract: Accurate tea leaf disease classification in real-world scenarios is hindered by complex backgrounds and the loss of fine-grained lesion details during CNN down sampling. To address this, we propose ResNet50-Dual-Fusion. It integrates a Cross-Attention Feature Fusion module (CAmodule) to adaptively reconstruct tiny lesion edges via cross-spatial interaction between shallow and deep features. Furthermore, a Magnitude-Aware Linear Attention (MALA) module with 2D Rotary Position Embedding (RoPE) is introduced to rectify magnitude neglect, effectively suppressing background noise. Evaluated on a 5,276-image dataset, our model achieves 85.96% accuracy (+3.00% over the baseline), outperforming architectures like ViT and Swin-Tiny. Grad-CAM visualizations confirm its superior lesion localization, providing a robust paradigm for automated crop disease diagnosis.

Keywords: Tea leaf disease classification; Deep learning; Residual network (ResNet-50); Cross-attention mechanism; Magnitude-aware linear attention (MALA)

Online publication: April 10, 2026

1. Introduction

Automated diagnosis is crucial for protecting vulnerable tea yields [1]. While CNNs like ResNet and initial visual models pioneered disease classification, subsequent architectures utilizing RetinaNet and deeper networks further enhanced detection robustness [2–5]. However, complex agricultural backgrounds cause misclassifications, and continuous down sampling obscures fine-grained lesion details [6]. To resolve this, we propose ResNet50-Dual-Fusion. Inspired by cross-domain matching, our Cross-Attention Feature Fusion (CAmodule) adaptively reconstructs tiny lesions [7]. Additionally, we integrate a MALA module to rectify magnitude neglect, dynamically suppressing backgrounds to achieve superior classification accuracy [8].

2. Related work

2.1. Plant and tea leaf disease classification

Deep learning drives automated crop disease recognition. Initial DNNs proved effective for plant diseases [1]. For

tea leaves, researchers proposed improved CNNs for Grey Blight and TeaDiseaseNet for multi-scale lesions [6,9]. Detection models like YOLO and global-focused ViTs further advanced feature extraction [10,11]. However, existing models struggle to allocate dynamic attention amidst complex backgrounds and high inter-class similarities.

2.2. Attention mechanisms and cross-feature fusion

Standard self-attention models robust global contexts but incurs immense computational costs. Linear attention reduces overhead but degrades performance by neglecting query magnitude [8,12]. Fan *et al.* resolved this via MALA [8]. Additionally, FAMNet demonstrated the superiority of cross-feature fusion in overcoming domain interference [7]. Inspired by these, we integrate cross-attention fusion with MALA to precisely identify tiny, complex lesions in real-world scenarios.

3. Methodology

To address complex background noise and the loss of fine-grained lesion details in tea leaf disease classification, we propose a dual-branch fusion attention network based on ResNet-50. This framework innovatively integrates CAmodule for spatial detail reconstruction and MALA module for dynamic background suppression.

3.1. Base network: ResNet-50

ResNet-50 serves as our backbone, balancing computational efficiency and robust feature extraction. It mitigates vanishing gradients via shortcut connections (**Table 1**), with the basic residual unit formulated as:

$$y = \sigma(F(x, \{W_i\}) + W_s x) \tag{1}$$

where x and y are inputs and outputs, F is the residual mapping, σ denotes ReLU, and W_S matches dimensions.

Comprising four bottleneck stages, we remove its terminal pooling and fully connected layers. This allows direct feeding of high-dimensional feature maps into the CAmodule and MALA, maximally preserving fine-grained spatial lesion textures.

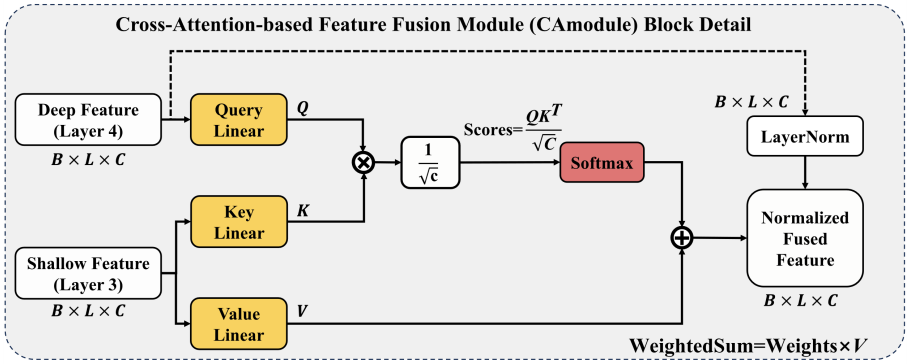
Table 1. Network configuration of the ResNet-18 backbone

Stage name	Output size	Layer configuration
Input stem	112×112	7×7,64, stride2 3×3 max pool, stride2
Stage 1	56×56	$\begin{bmatrix} 1\times1,64 \\ 3\times3,64 \\ 1\times1,256 \end{bmatrix}\times3$
Stage 2	28×28	$\begin{bmatrix} 1\times1,128 \\ 3\times3,128 \\ 1\times1,512 \end{bmatrix}\times4$
Stage 3	14×14	$\begin{bmatrix} 1\times1,256 \\ 3\times3,256 \\ 1\times1,1024 \end{bmatrix}\times6$
Stage 4	7×7	$\begin{bmatrix} 1\times1,512 \\ 3\times3,512 \\ 1\times1,2048 \end{bmatrix}\times3$
Custom routing	14×14 7×7	Output to CAmodule (Stage 3 & 4) Output to MALA (Stage 4)

3.2. Cross-attention feature fusion module

To address the loss of tiny lesion details in deep layers, we propose the CAmodule. Inspired by FAMNet, we adapt its cross-domain fusion mechanism for cross-layer interaction, utilizing high-resolution shallow details to dynamically calibrate deep semantics.

As illustrated in **Figure 1**, the module receives deep semantic features (F_{deep}) and shallow texture features ($F_{shallow}$). It projects F_{deep} into queries (Q), and $F_{shallow}$ into keys (K) and values (V). The adaptive cross-layer aggregation is formulated as:



$$F_{attn} = \text{Softmax}\left(\frac{QK^T}{\sqrt{C}}\right)V \tag{2}$$

Figure 1. Schematic diagram of the cross-attention feature fusion module.

To preserve primary semantic representations and stabilize training, a residual connection and Layer Normalization are applied:

$$F_{out} = \text{LayerNorm}\left(F_{deep} + F_{attn}\right) \tag{3}$$

This mechanism adaptively "recalls" fine-grained edges previously filtered by downsampling, establishing a robust foundation for accurate classification.

3.3. Magnitude-aware linear attention

To address query magnitude neglect in traditional linear attention, which causes imbalanced attention allocation, we introduce the MALA module. Maintaining complexity, MALA dynamically calibrates magnitudes to focus on discrete lesions.

As illustrated in **Figure 2**, input X is projected into Q, K, V, and gate O. LEPE extracts local details F_{leps} from V, while Q and K receive 2D Rotary Position Embedding (RoPE) to form $\widehat{Q}$ and $\widehat{K}$. Using a dynamic correction factor $Z = (\widehat{Q}\overline{K}^T)/\sqrt{d}$, MALA recalibrates global attention:

$$F_{attn} = \widehat{Q}\left(\widehat{K}^T V\right)\left(1 + \frac{1}{Z + \epsilon}\right) - Z\overline{V} \tag{4}$$

This effectively suppresses healthy backgrounds. Finally, features are fused via the output projection:

$$X_{out} = \text{Conv}_{1\times1}\left(\left(F_{attn} + F_{lepe}\right) \odot O\right) \tag{5}$$

This architecture significantly enhances global perception and precise localization of complex tea lesions.

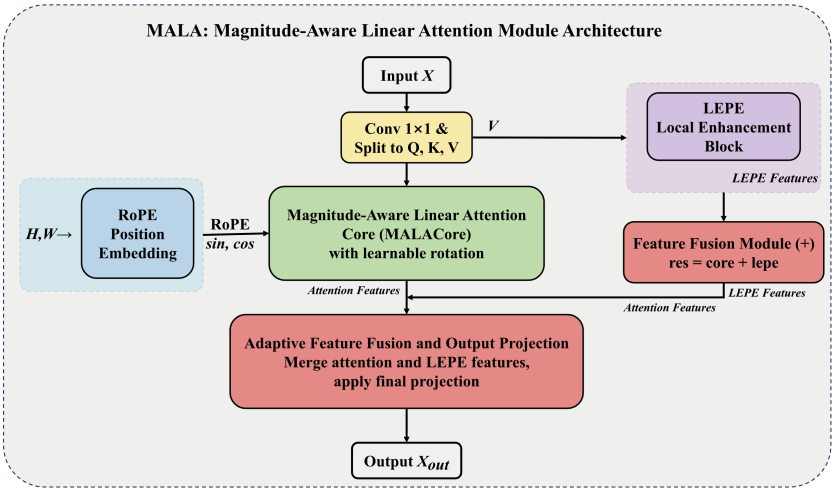


Figure 2. Schematic diagram of the magnitude-aware linear attention (MALA) module.

3.4. Overall network architecture

The proposed ResNet50-Dual-Fusion architecture is illustrated in **Figure 3.** Using a modified ResNet-50 backbone, the network extracts Stage 3 (shallow) and Stage 4 (deep) features to balance local textures with global semantics. An adapter aligns the Stage 3 features with the deep branch.

In the interaction stage, Stage 4 features act as queries (Q), fusing with adapted Stage 3 features via the CAmodule to recall lesion details. This output is residually added to the original Stage 4 features. Subsequently, the MALA module, integrated with RoPE, recalibrates attention weights to suppress backgrounds. Finally, enhanced features are processed through Global Pooling and a Fully Connected head to output classification probabilities.

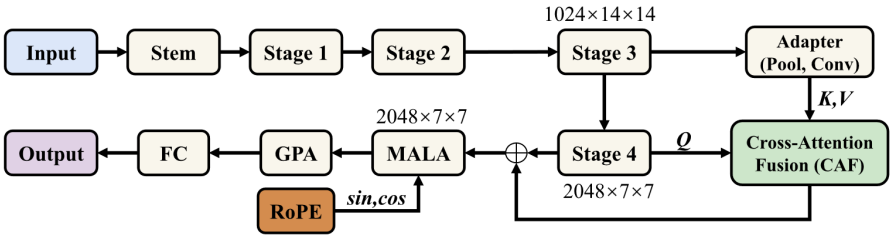


Figure 3. Schematic diagram of the overall ResNet50-dual-fusion network architecture.

4. Experimental results

4.1. Dataset and implementation details

This study utilizes a tea leaf disease dataset of 5,276 images across seven categories (e.g., Green mirid bug, Gray Blight), partitioned into training, validation, and test sets at an 8:1:1 ratio. Images are resized to and augmented via random flips and ImageNet normalization. Experiments are implemented using PyTorch on an NVIDIA RTX 3090 GPU. We employ the AdamW optimizer with a learning rate and weight decay of 0.0001, a batch size of 64, and Cross-Entropy Loss over 100 epochs. Model performance is evaluated using Accuracy, Precision, Recall, and F1-score to ensure robust assessment.

4.2. Model performance evaluation

Comparative experiments (**Table 2**) show that ResNet50 achieves the best performance with 82.96% accuracy, surpassing Swin-Tiny, ConvNeXt, and ViT. This superiority stems from its robust inductive bias and residual structure, which effectively capture fine-grained textures in small-scale datasets. Consequently, ResNet50 is selected as the base network for subsequent integration of the CAmodule and MALA.

Table 2. Performance comparison of different base models on the tea leaf disease dataset

Model	Acc (%)	Prec (%)	Rec (%)	F1 (%)
ResNet50	82.96	81.20	79.59	80.26
ViT-Base	72.47	69.43	67.49	68.15
Swin-Tiny	79.78	78.08	76.13	76.95
ConvNeXt-Tiny	70.79	68.06	66.36	67.01
PoolFormer-s12	73.22	70.90	69.65	70.16

4.3. Ablation studies

Ablation experiments (**Table 3**) validate the efficacy of CAmodule and MALA. Integrating CAmodule alone improves accuracy to 84.83% (+1.87%), demonstrating its ability to reconstruct fine-grained lesion details via cross-layer interaction. Meanwhile, MALA alone achieves 84.46% accuracy (+1.50%) by suppressing background noise through magnitude calibration. The full ResNet50-Dual-Fusion model achieves a peak accuracy of 85.96% (+3.00%). This highlights the synergy between "cross-layer detail reconstruction" and "intra-layer background suppression," significantly boosting overall classification performance.

Table 3. Ablation study results of different module combinations

Components		Performance metrics				
CrossAttn	MALA	Acc (%)	Prec (%)	Rec (%)	F1 (%)	Improvement
		82.96	81.20	79.59	80.26	-
	√	84.46	83.19	81.48	82.12	+ 1.50%
√		84.83	84.38	81.93	82.93	+ 1.87%
√	√	85.96	84.24	82.35	83.02	+ 3.00%

4.4. Model interpretability analysis

Grad-CAM visualizations (**Figure 4**) demonstrate our model's precise lesion localization across all categories. High-response regions strictly align with diseased areas, proving MALA's efficacy in suppressing background noise via magnitude calibration. Furthermore, CAF's cross-layer interaction enables the capture of fine-grained details, from large necrotic blights to discrete insect bites. These results confirm that ResNet50-Dual-Fusion learns discriminative pathological features rather than background biases, ensuring diagnostic reliability.

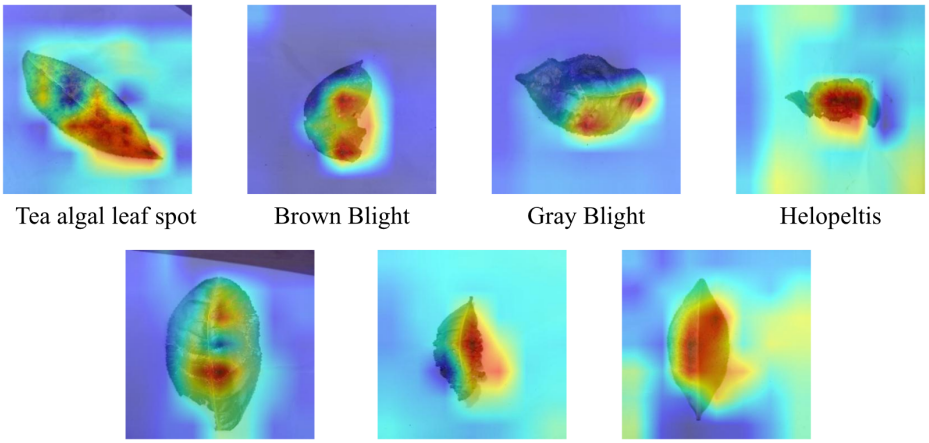


Figure 4. Grad-CAM visualization results of the proposed model on the tea leaf disease dataset.

5. Conclusion

This paper presents ResNet50-Dual-Fusion, integrating CAmodule and MALA to tackle complex backgrounds and fine-grained detail loss in tea disease classification. Experiments confirm that CAmodule enables cross-layer detail reconstruction, while MALA effectively suppresses environmental noise. Our method achieves 85.96% accuracy (+3.00% improvement), outperforming mainstream CNNs and Transformers. Grad-CAM results further validate its interpretability. Future work will focus on model pruning and knowledge distillation to facilitate lightweight deployment on edge-side agricultural intelligent devices.

Disclosure statement

The author declares no conflict of interest.

References

[1] Sladojevic S, Arsenovic M, Anderla A, et al., 2016, Deep Neural Networks based Recognition of Plant Diseases by Leaf Image Classification. Computational Intelligence and Neuroscience, 2016: 1–11.

[2] He K, Zhang X, Ren S, et al., 2016, Deep Residual Learning for Image Recognition, Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 770–778.

[3] Chen J, Liu Q, Gao L, 2019, Visual Tea Leaf Disease Recognition using a Convolutional Neural Network Model. Symmetry, 11(3): 343.

[4] Bao W, Fan T, Hu G, et al., 2022, Detection and Identification of Tea Leaf Diseases based on AX-RetinaNet. Scientific Reports, 12(1): 1–16.

[5] Datta S, Gupta N, 2023, A Novel Approach for the Detection of Tea Leaf Disease using Deep Neural Network. Procedia Computer Science, 2023(218): 2273–2286.

[6] Pandian J, Nisha S, Kanchanadevi K, et al., 2023, Grey Blight Disease Detection on Tea Leaves using Improved Deep Convolutional Neural Network. Computational Intelligence and Neuroscience, 2023: 1–11.

[7] Bo Y, Zhu Y, Li L, et al., 2025, FAMNet: Frequency-Aware Matching Network for Cross-Domain Few-Shot Medical Image Segmentation. Proc. AAAI Conf. Artificial Intelligence, 39(2).

[8] Fan Q, Huang H, Ai Y, et al., 2025, Rectifying Magnitude Neglect in Linear Attention, Proc. IEEE/CVF Int. Conf. Computer Vision (ICCV), 2025.

[9] Sun Y, Wu F, Guo H, 2023, TeaDiseaseNet: Multi-Scale Self-Attentive Tea Disease Detection. Frontiers in Plant Science, 2023(14): 1257212.

[10] Mathew M, Mahesh T, 2022, Leaf-based Disease Detection in Bell Pepper Plant using YOLO v5. Signal, Image and Video Processing, 16(3): 841–847.

[11] Dosovitskiy A, Beyer L, Kolesnikov A, et al., 2021, An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale, Proc. Int. Conf. Learning Representations (ICLR), 2021.

[12] Vaswani A, Shazeer N, Parmar N, et al., 2017, Attention is all you Need. Advances in Neural Information Processing Systems (NIPS), 2017(30): 5998–6008.