AMERICAN JOURNAL OF BIOCHEMISTRY AND BIOTECHNOLOGY

RESEARCH ARTICLE

Crop Disease Detection via an Improved Residual Network

Guolin Chen[1], Weifeng Liu[2], Xinyu Zou[2], Xiaoxia Li[3]*

[1]Department of General Education, Gandong University, Fuzhou, Jiangxi 344000, China
[2]School of Information Engineering, Gandong University, Fuzhou, Jiangxi 344000, China
[3]Fuzhou Preschool Education College, Fuzhou, Jiangxi 344000, China

*Corresponding Author: xiaoling562@yeah.net

Abstract: Deep learning has shown substantial potential across many domains, and crop disease detection is no exception. This study investigates the application of deep learning to crop disease detection, systematically evaluating the performance of convolutional neural networks of different depths. Building on the classical ResNet-50 as the base model, we introduce a Squeeze-and-Excitation (SE) attention mechanism to enhance feature extraction. The results indicate that increasing network depth enables the extraction of more hierarchical and discriminative features, which is beneficial for distinguishing visually similar crop disease categories. Ablation experiments further show that adding the Squeeze-and-Excitation (SE) module to ResNet-50 increases validation accuracy from 80.7% to 85.2%, indicating that channel-wise attention helps emphasize disease-relevant feature channels while suppressing redundant information. Overall, deep learning provides a reliable technical foundation for crop disease detection and yields substantial performance gains.

Keywords: Deep Learning, Crop Disease Detection, SE-ResNet50, Squeeze-and-Excitation (SE) Attention Mechanism

Received: 24-11-2025 | **Revised:** 05-03-2026 | **Accepted:** 01-04-2026 | **DOI:** 10.3844/ajbbsp.2026.22.02.026

Introduction

Crop diseases remain a critical threat to global agriculture, leading to reduced yields, deteriorated crop quality, and substantial economic losses [1]. As agricultural production becomes more intensive and large-scale, the demand for timely and accurate disease detection has grown significantly. Traditional diagnostic approaches, such as visual inspection by experts, are constrained by subjectivity, labor intensity, and low efficiency, which makes them unsuitable for modern precision agriculture. Visible symptoms on plant leaves often appear at the early stages of infection [2, 3], yet the similarity of these manifestations across different diseases frequently hampers reliable recognition. Consequently, the development of automated and scalable methods for crop disease detection is of great importance to safeguard food security and support sustainable agricultural practices.

In recent years, advances in computer vision have provided promising tools for plant disease monitoring. Convolutional Neural Networks (CNNs), in particular, have shown strong capabilities in extracting multi-level image features and have been widely applied in various recognition tasks, including crop disease identification [4-10]. Early studies demonstrated the feasibility of using CNN architectures such as ResNet [11], VGG [12], and AlexNet [13] for disease classification, achieving encouraging results under controlled conditions. However, these approaches face notable challenges in practical applications.

DOI: 10.3844/ajbbsp.2026.22.02.026

Many models require large-scale and balanced datasets to achieve reliable performance, and their generalization often weakens in field scenarios where environmental variability and class imbalance are common [14-16]. Moreover, subtle visual differences between disease categories are difficult to capture with conventional CNN structures, limiting their ability to deliver consistent accuracy in diverse agricultural environments.

To address these limitations, recent research has explored enhanced network architectures and attention mechanisms to improve feature extraction and classification performance. For instance, a study integrated an optimized ResNet50 with Feature Pyramid Networks and an improved attention mechanism for orange leaf disease classification, enhancing the model's ability to capture lesion features under complex backgrounds while addressing class imbalance [17]. Similarly, the application of SE-ResNet50 to rice seed classification has demonstrated that channel-wise attention enables the model to capture subtle differential features more effectively than standard architectures [18]. Other work has developed a fine-tuned ResNet50 incorporating strategic layer unfreezing and regularization, which significantly improves feature reuse and generalization across diverse crop species [19]. Furthermore, a hybrid ML-DNN framework combining ResNet-based feature extraction with PCA and conventional classifiers has been proposed, achieving strong robustness to class imbalance and enhancing model interpretability [20]. More recently, researchers have advanced the field by integrating YOLOv8 for leaf region extraction, ResNet-50 for disease classification, and GPT-3.5 for generating treatment recommendations, highlighting the potential of multi-stage pipelines for comprehensive agricultural decision support [21].

Building upon these research advances, this study introduces an enhanced approach to crop disease detection by integrating a channel attention mechanism into the classical ResNet-50 architecture. Among existing attention mechanisms such as CBAM, ECA, and Squeeze-and-Excitation, the SE module is selected due to its ability to model channel-wise dependencies with low computational overhead, making it well-suited for capturing subtle disease-related features in crop images. By adaptively reweighting feature channels, the improved model enables the network to focus more effectively on disease-relevant patterns while suppressing redundant information. Comparative experiments against widely used CNN models demonstrate that the enhanced ResNet-50 achieves superior recognition accuracy and improved feature extraction capability. Ablation studies further confirm that the attention mechanism contributes significantly to these performance gains. By strengthening the representational power of residual networks, this research provides a practical and robust framework for automated crop disease detection, offering valuable support for intelligent agricultural systems.

Experimental Data

Datasets

The experimental dataset was derived from the publicly available Agricultural Disease dataset in the 2018 Global AI Challenge, which includes 27 disease categories affecting 10 common crops such as apples, corn, grapes, and citrus. Representative examples are shown in Fig. 1. Each category contains a set of leaf images that serve as samples for disease detection, with the total number of images exceeding 40,000. This corpus is instrumental for extracting discriminative disease features and provides a solid basis for developing accurate crop disease detection models.

Data Augmentation

For disease categories with relatively few images, it is challenging to extract representative features from such limited data. Moreover, in real-world scenarios, images captured at different times, under varying environmental conditions, from diverse angles, or with different devices may exhibit variations in color tone, brightness, and other attributes. To enhance the generalization ability of the model, the original images were augmented through random rotation, color temperature adjustment, brightness and contrast modification, and cropping, as illustrated in Fig. 2.

Specifically, random rotation and cropping force the model to learn rotation-invariant features, simulating the varying orientations of leaves in field conditions. Color temperature and brightness adjustments mimic diverse lighting environments, thereby preventing the model from overfitting to specific illumination conditions present in the training set. After data augmentation, the dataset was divided into training, validation, and testing sets in a 7:2:1 ratio. This ratio is common in deep learning, ensuring the model has sufficient data for feature learning while retaining enough independent samples for hyperparameter tuning and unbiased performance evaluation. The final distribution of the dataset is presented in Table 1.

DOI: 10.3844/ajbbsp.2026.22.02.026

Table 1: Distribution of disease image quantities

Type	Training set	Validation set	Test set
Apple scab	786	224	112
Apple gray spot disease	1792	512	256
Apple rust	764	218	109
Cherry powdery mildew	1323	378	189
Corn gray leaf spot	1503	429	214
Corn rust	3519	1005	502
Corn leaf spot	2969	848	424
Maize dwarf mosaic virus	3423	978	489
Grape black rot	3540	1011	505
Grape ring spot	3939	1125	562
Grape brown spot	2902	829	414
Citrus greening disease	15582	4452	2226
Peach scab	6833	1952	976
Pepper scab	2788	796	398
Potato early blight	2994	855	427
Potato late blight	2927	836	418
Strawberry leaf blight	3255	930	465
Tomato powdery mildew	5397	1542	771
Tomato scab	1751	500	250
Tomato early blight	2910	831	415
Tomato late blight	8766	1647	823
Tomato leaf mold	2776	793	396
Tomato leaf spot	2024	578	289
Tomato target spot	5157	1493	736
Tomato red spider mite damage	3414	975	487
Tomato yellow leaf curl virus	16325	4664	2332
Tomato mosaic virus	1096	313	156

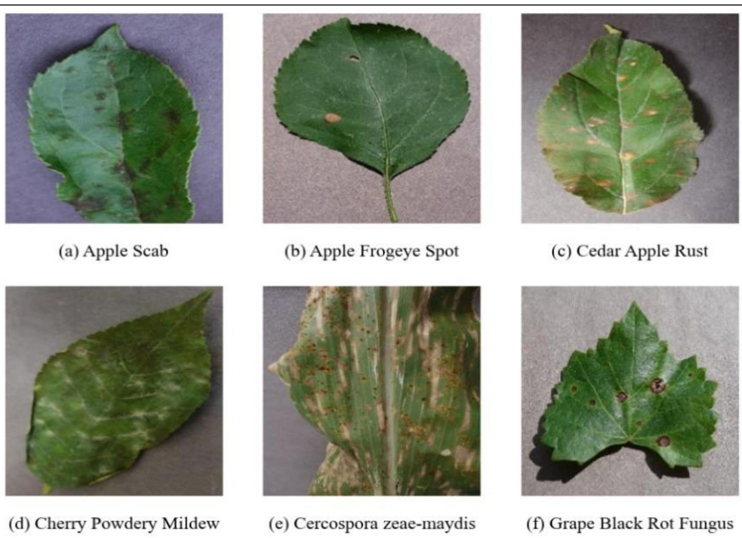

Fig. 1: Images of selected diseases

DOI: 10.3844/ajbbsp.2026.22.02.026

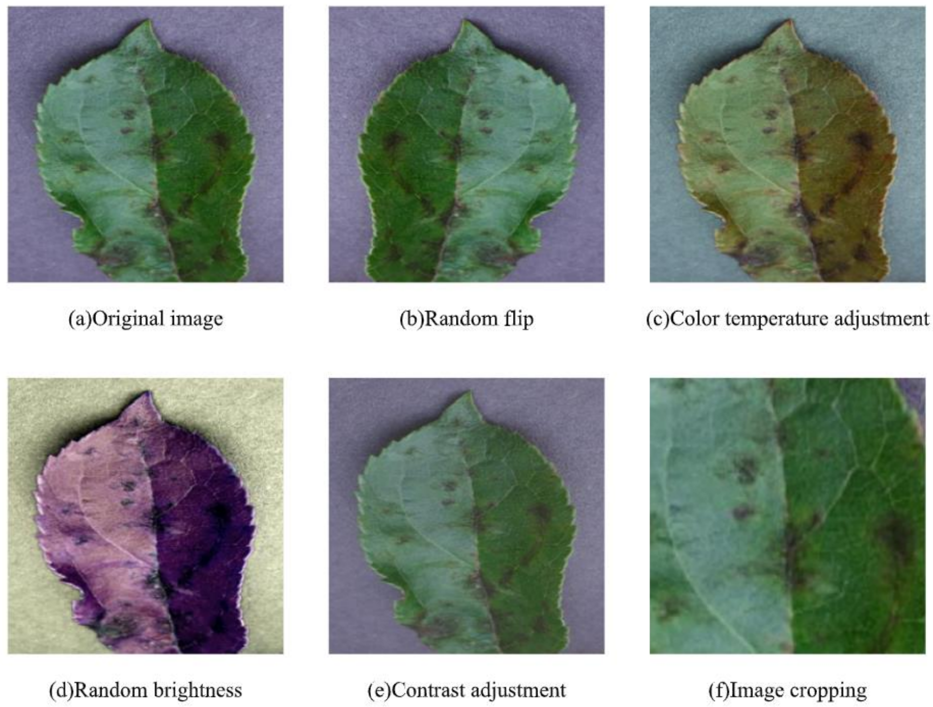

Fig. 2: Image after data augmentation

Materials and Methods

ResNet-50 Network

Early convolutional networks such as AlexNet, VGG-16, and GoogLeNet advanced image recognition primarily by scaling depth and width. However, plain deep stacks soon exhibited the degradation problem in which optimization becomes harder, and training error can increase as depth grows, even without severe vanishing gradients. To address this problem, [22] introduced the ResNet architecture, which effectively mitigates network degradation associated with increasing depth. In deep network models, residual blocks are incorporated to directly transmit input data to the output, thereby preserving the original features. The output processed by convolutional layers is then added to the original input, enabling the layers to fit a residual mapping. The residual mapping is defined in Eq. (1):

$$H(x) = x + F(x) \tag{1}$$

In Eq. (1), x denotes the input of the residual unit, $H(x)$ represents the weights of the residual unit, F and refers to the underlying mapping.

The ResNet-50 architecture is composed of convolutional layers, max-pooling layers, four types of residual block modules, a global average pooling layer, and a fully connected layer. The core component is the residual block, which consists of two 1×1 convolutional layers and one 3×3 convolutional layer. This model is designed to learn the residual function as defined in Eq. (2):

$$G(x) = F(x) - x \tag{2}$$

When $G(x)=0$ $F(x)=x$ it corresponds to the identity mapping. ResNet does not introduce additional computational complexity or parameters, and when the optimization target approaches an identity mapping, it is easier to learn slight perturbations around the identity than to fit an entirely new mapping function. Through identity connections, low-level feature information can be directly propagated to higher layers, enabling the network to learn diverse feature representations more effectively. The structure of the residual block that accomplishes this functionality is illustrated in Fig. 3.

DOI: 10.3844/ajbbsp.2026.22.02.026

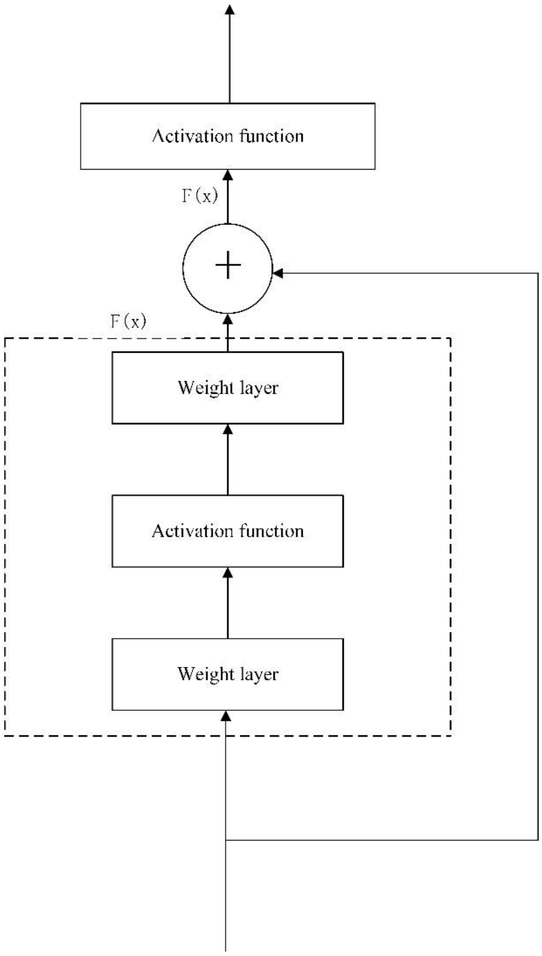

Fig. 3: Residual block module

SE Module

Convolutional neural networks are built upon convolutional operations, which extract informative features by integrating spatial and channel information within local receptive fields. To enhance the representational capacity of networks, [23] proposed the Squeeze-and-Excitation Networks (SENet). The SE module, introduced in SENet, is a channel attention mechanism that has been widely adopted in visual models due to its plug-and-play property. It strengthens channel-wise feature representations of the input feature maps while preserving their original spatial dimensions. In this module, S denotes the squeeze operation (global average pooling), as defined in Eq. (3).

$$z_c = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} x_c(i, j) \tag{3}$$

This operation compresses the spatial information of the input feature maps.

E denotes the excitation operation (channel attention weights), as defined in Eq. (4):

$$s = \sigma(W_2 \cdot \delta(W_1 \cdot z)) \tag{4}$$

Here, $W_1 \in R^{\frac{c}{r} \times C}, W_2 \in R^{C \times \frac{c}{r}}$ δ represents the ReLU activation function and σ denotes the Sigmoid function. This operation integrates the learned channel attention information with the input feature maps, thereby generating feature

DOI: 10.3844/ajbbsp.2026.22.02.026

maps with channel attention. The operation of the SE module is illustrated in Fig. 4. The overall workflow of the model is as follows:

First, the feature maps of crop diseases with dimension $H \times W \times C$ s are fed into the network. The spatial features of the input maps are then compressed, and global average pooling is applied along the spatial dimension to obtain feature maps of dimension $1 \times 1 \times C$. Subsequently, channel-specific feature learning is performed on the compressed maps through Fully Connected (FC) layers, producing channel-attentive feature maps that retain the same dimension $1 \times 1 \times C$. Finally, the channel-attentive feature maps are multiplied by the original input feature maps with channel-wise weighting, yielding the output feature maps with enhanced channel attention.

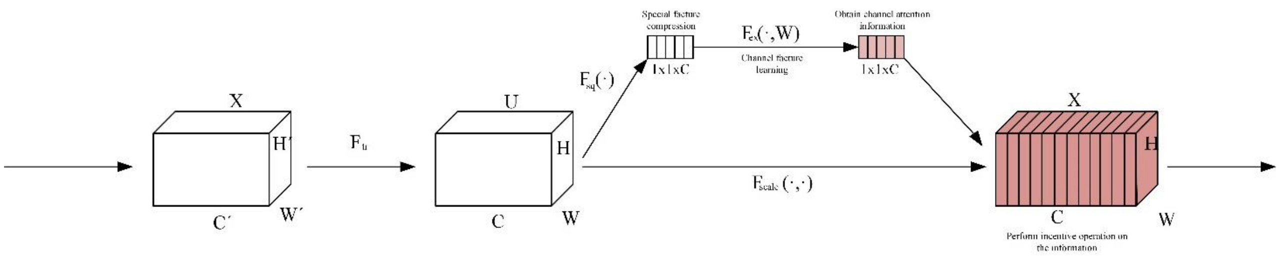

Fig. 4: SE module

ResNet-50 Optimization

Since the SE module can adaptively capture the importance of inter-channel relationships and dynamically adjust channel weights, it effectively reduces redundancy and noise while enhancing the accuracy of feature selection. When the SE module is embedded into the residual blocks of ResNet-50, each block is able to dynamically reweight feature channels, allowing the network to focus adaptively on more informative features during training.

At present, several strategies have been proposed to incorporate SE modules into residual structures. When inserted before the residual block, the module is referred to as an SE-PRE block, as illustrated in Fig. 5. When applied after the residual addition, it is called an SE-POST block, as shown in Fig. 6. Alternatively, the SE module can be placed immediately before the residual addition, which corresponds to the conventional SE block, illustrated in Fig. 7.

For the SE-PRE block, the SE module is positioned before the residual block. This insertion strategy enables the SE module to adjust channel attention of the input features before residual learning, which can benefit subsequent feature extraction. However, it may limit the ability to fully exploit the nonlinear transformations within the residual block to guide channel attention adjustment.

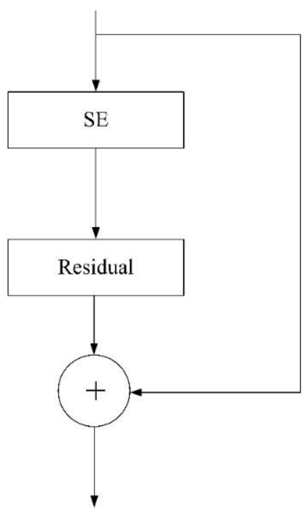

Fig. 5: SE-PRE module

DOI: 10.3844/ajbbsp.2026.22.02.026

For the SE-POST block, the SE module is positioned after the residual block. This placement allows it to capture features that have already undergone the nonlinear transformations within the block, thereby enabling a more refined adjustment of channel attention. However, the adjustment may also be influenced by the internal variations of the residual block, which can compromise its precision.

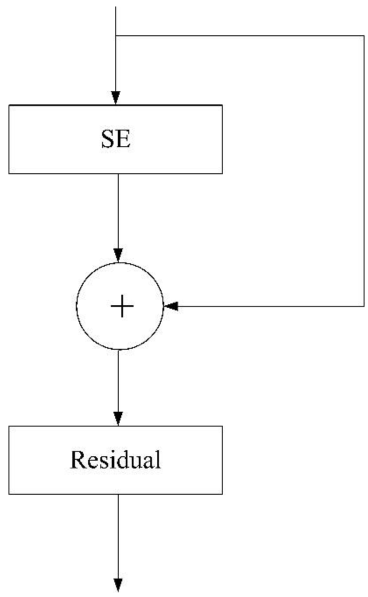

Fig. 6: SE-POST module

By comparison, applying the SE module after the residual has been formed enables the network to reweight features more precisely and focus on the most informative channels. We therefore adopt this configuration to refine the residual block and integrate it into the ResNet architecture, yielding the improved SE-ResNet module, as illustrated in Fig. 7.

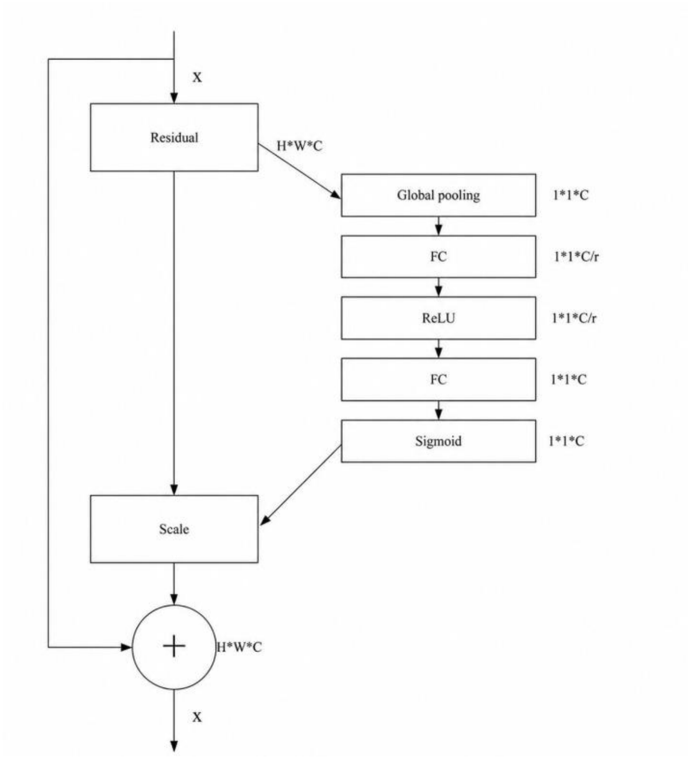

Fig. 7: SE-ResNet module

DOI: 10.3844/ajbbsp.2026.22.02.026

First, global average pooling is applied to aggregate spatial information within each channel, completing the squeeze operation. The resulting descriptor is then passed through two Fully Connected (FC) layers forming a bottleneck structure to model inter-channel dependencies and to output one weight per input channel. Specifically, the first FC layer reduces the feature dimension to c/r ; after a ReLU activation, a second FC layer restores it to the original dimensionality. Compared with using a single FC layer, this design introduces additional nonlinearity for more accurate modeling of complex channel correlations and, through dimensionality reduction, substantially lowers the number of parameters and computations. A Sigmoid function is subsequently used to obtain normalized weights, and a final scale operation applies these weights to the corresponding channels of the input feature maps.

Results and Analysis

Experimental Results

This study evaluates the feature extraction capability of the improved SE-ResNet-50 on crop disease recognition. To this end, three representative deep models VGG-16, ResNet-18, and ResNet-50 were trained under identical hyperparameter settings and compared against SE-ResNet-50.

VGG-16 is a classic deep convolutional network with 16 weight layers and approximately 138 million parameters. In our experiments, it achieved a validation accuracy of 70.6%. Its fixed and strictly sequential design, without residual or attention mechanisms, restricts the effective receptive field and reduces the degree of feature reuse. As a result, optimization becomes more challenging as the depth increases and the learned representations remain predominantly local. In the context of crop disease imagery, where lesions are often small, exhibit low contrast, and appear across multiple scales under varying illumination and complex backgrounds, the capacity of VGG-16 is insufficient. The model tends to rely heavily on shallow texture cues and shows weaker generalization under domain shift. Moreover, the balance between computational cost and accuracy is less favorable compared with more advanced architectures that incorporate skip connections, multi-scale feature integration, or attention mechanisms.

ResNet-18 and ResNet-50 are residual networks with different depths that employ identity skip connections to ease optimization and improve feature reuse. ResNet-18 comprises 18 weighted layers formed by basic residual blocks with 3×3 convolutions. Although it is relatively shallow, the residual pathway stabilizes gradient propagation and preserves informative cues from earlier stages, leading to a validation accuracy of 74.3% in our experiments. ResNet-50 increases representational capacity through bottleneck blocks arranged as 1×1-3×3-1×1 convolutions, which expand and then compress channel dimensions while controlling computation. This deeper configuration enhances sensitivity to fine-grained textures and cross-scale patterns that are common in crop disease imagery, and it achieved 80.7% validation accuracy.

Finally, we trained and evaluated the proposed SE-ResNet-50, which augments ResNet-50 with a Squeeze and Excitation attention module as illustrated in Fig. 8. The SE mechanism performs global average pooling to obtain channel descriptors and then applies a lightweight gating operation to reweight channels, thereby emphasizing disease-relevant responses and suppressing nuisance variations such as illumination changes and cluttered backgrounds. SE-ResNet-50 attained a validation accuracy of 85.2%, which is a clear improvement over ResNet-50. These results confirm the effectiveness of channel-wise recalibration and indicate that SE-ResNet-50 provides stronger feature extraction and better class separability for crop disease recognition. To further analyze training stability and convergence, Figure 9 presents the training curves of accuracy and loss values. Observation reveals that the proposed SE-ResNet-50 model converges faster than the baseline ResNet-50 model. The validation loss of this method decreases more significantly in the early epochs and stabilizes at a lower value; meanwhile, the accuracy curve consistently maintains an advantage over the baseline, ultimately reaching 85.2%.

For ease of comparison, the performance of different models on the validation set is summarized in Table 2. As shown in Table 2, accuracy increases as the network becomes deeper. In addition, the introduction of the SE attention module further strengthens feature extraction, and SE-ResNet-50 achieves the highest accuracy on the crop disease detection task [24].

Because the dataset is class-imbalanced, SE-ResNet-50 attains an overall validation accuracy of 85.2%. However, for disease categories with larger sample sizes, for example, peach scab, citrus huanglongbing, tomato late blight, and tomato yellow leaf curl virus, the model provides more accurate detection, as summarized in Table 3. These results indicate that the amount of training data strongly affects class-level accuracy. When sufficient samples are available, the model can learn disease-specific features more effectively.

DOI: 10.3844/ajbbsp.2026.22.02.026

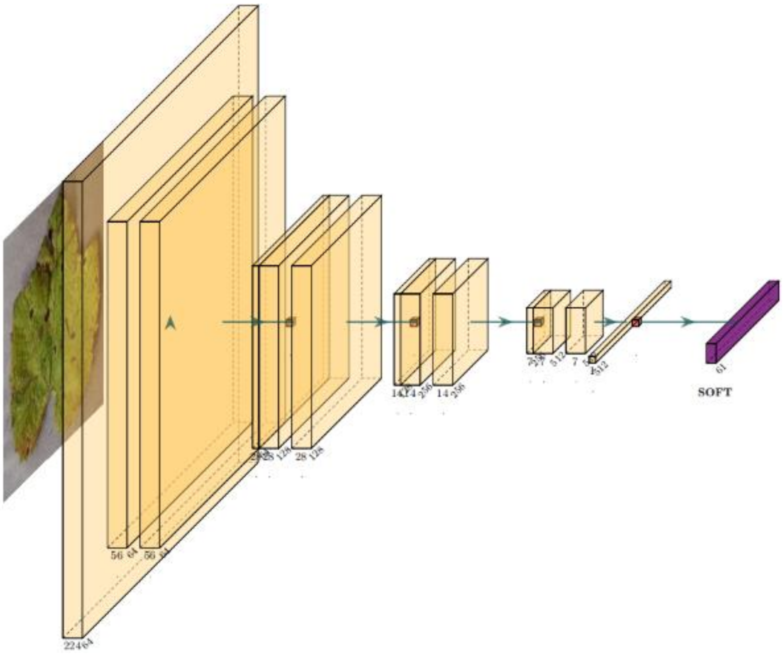

Fig. 8: Architecture of the SE-ResNet-50 model

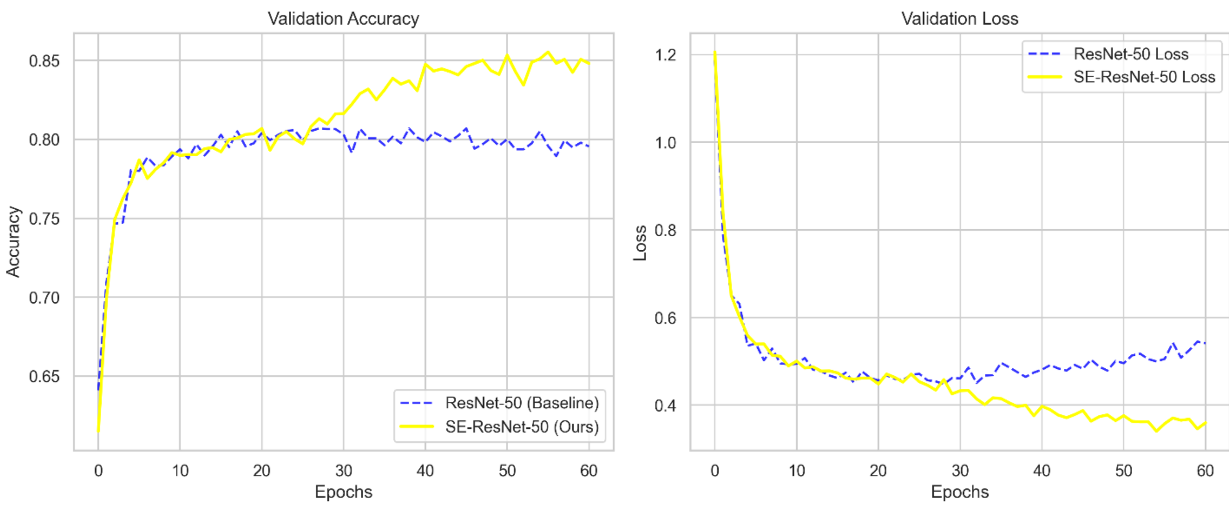

Fig. 9: Training accuracy and loss curves

Table 2: Performance comparison of different models on the validation set

Model	Accuracy	Precision	Recall	F1-Score
VGG-16	70.6%	69.1%	70.1%	69.6%
ConvNeXt	70.4%	69.5%	69.4%	69.5%
AlexNet	74.8%	70.8%	70.5%	70.6%
GoogleNet	75.9%	73.5%	73.5%	73.5%
ResNet18	74.3%	72.8%	73.8%	73.3%
ResNet50	80.7%	79.2%	80.2%	79.7%
SE-ResNet50	85.2%	83.7%	84.7%	84.2%

Table 3: Accuracy of disease detection

Type	Accuracy
Apple scab	83.36%
Apple gray spot disease	98.92%
Apple rust	93.64%
Cherry powdery mildew	70.26%
Corn gray leaf spot	86.96%
Corn rust	79.43%
Corn leaf spot	86.55%
Maize dwarf mosaic virus	97.33%
Grape black rot	81.48%
Grape ring spot	76.06%
Grape brown spot	93.43%
Citrus greening disease	89.58%
Peach scab	88.30%
Pepper scab	77.09%
Potato early blight	90.39%
Potato late blight	84.81%
Strawberry leaf blight	77.36%
Tomato powdery mildew	91.3%
Tomato scab	83.3%
Tomato early blight	73.02%
Tomato late blight	94.30%
Tomato leaf mold	77.08%
Tomato leaf spot	70.32%
Tomato target spot	81.31%
Tomato red spider mite damage	84.86%
Tomato yellow leaf curl virus	91.89%
Tomato mosaic virus	84.25%

Ablation Experiment

To demonstrate the effectiveness of adding the SE module to ResNet-50, we conducted an ablation study using ResNet-50 as the baseline to isolate the contribution of the Squeeze-and-Excitation (SE) module within the proposed pest-and-disease detection pipeline and to quantify its impact on model performance. The SE module was inserted into the baseline model, and performance was evaluated on the validation set. All other training settings, including data preprocessing, optimizer, learning rate, and training epochs, were kept identical to ensure a fair comparison. As reported in Table 4, the Squeeze-and-Excitation (SE) module effectively enhances feature discrimination by recalibrating channel-wise responses.

Table 4: Results of the ablation experiment

Model	Accuracy
ResNet50	80.7%
SE-ResNet50	85.2%

To provide a more intuitive demonstration of the effect of the SE module, Grad-CAM was employed to generate heatmaps. Grad-CAM, proposed by [25], is designed to interpret the decision-making process of deep learning models in image classification. Gradient-weighted Class Activation Mapping (Grad-CAM) is a gradient-based class activation technique, and its mathematical formulation is given in Eq. (5).

$$L^c_{Grad-CAM} = \mathrm{ReLU}\left(\sum\nolimits_k \alpha_k^c A^k\right) \tag{5}$$

Where $\alpha_k^c = \frac{1}{Z}\sum_{i,j}\frac{\partial y^c}{\partial A^k_{i,j}}$ A^k denotes the feature map at the k-th layer, and y^c is the score for class c. By computing the gradient of the class score with respect to c, Grad-CAM produces a visual heatmap that highlights the image regions most influential to the model's prediction. To elucidate the contribution of the Squeeze-and-Excitation (SE) module to feature extraction, we applied Grad-CAM to compare the attention maps of ResNet-50 and SE-ResNet-50. As shown in Figs. 10 and 11, SE-ResNet-50 assigns greater focus to disease-specific regions than the baseline ResNet-50. This visual evidence demonstrates that the Squeeze-and-Excitation (SE) module enhances the network's ability to suppress background noise and highlight key disease-related features, thereby improving the robustness and interpretability of the pest and disease detection process.

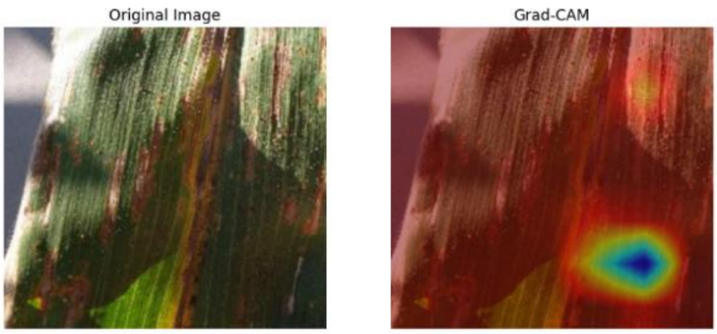

Fig. 10: ResNet50 heatmap

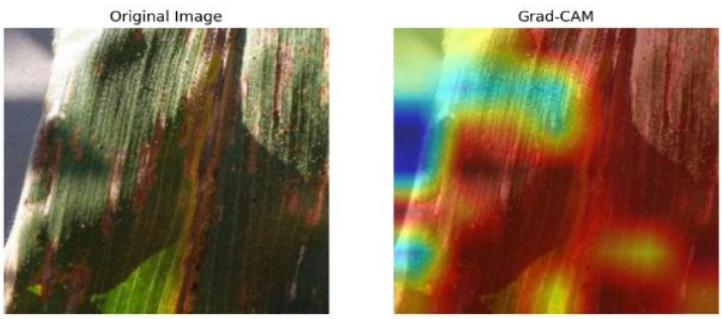

Fig. 11: SE-ResNet50 heatmap

Conclusion

This study explored the application of deep learning to crop disease detection. By augmenting the dataset and training VGG-16, ResNet-18, and other convolutional neural networks, we confirmed the feasibility of CNNs for this task. Building on this foundation, an SE (Squeeze-and-Excitation) attention module was integrated into the classical ResNet-50 architecture, which successfully improved the accuracy of crop disease detection. Experimental results demonstrated the superior feature extraction capability of the SE-ResNet-50 model, and ablation experiments further verified the effectiveness of the SE module.

Although SE-ResNet-50 achieved strong performance in this study, there remains room for further optimization. CNNs generally rely on large-scale training datasets, yet the dataset used here exhibits distributional heterogeneity. Considering that disease characteristics are influenced by geographic and climatic conditions, the current model still requires systematic validation in cross-regional transfer scenarios. Future work should enhance the model's ability to learn discriminative features under diverse regional and climatic settings. In addition, adaptive inference mechanisms that integrate environmental factors could be developed to comprehensively evaluate generalization performance across heterogeneous conditions, ultimately leading to a more robust and widely applicable crop disease recognition framework.

Acknowledgment

Thank you to the publisher for their support in the publication of this research article. We are grateful for the resources and platform provided by the publisher, which have enabled us to share our findings with a wider audience. We appreciate the efforts of the editorial team in reviewing and editing our work, and we are thankful for the opportunity to contribute to the field of research through this publication.

Funding Information

This work was supported by Jiangxi Provincial College Student Innovation Training Project (S202513432011).

Author's Contributions

Guolin Chen: Designed and performed the experiments.
Weifeng Liu and Xinyu Zou: Participated in collecting the materials related to the experiment.
Xiaoxia Li: Designed the experiments and revised the manuscript.

References

1. Elangovan, M., Nigam, R., & Srivastava, G. (2024). *Plant diseases: diagnosis, management, and control*.
2. Rahaman, J., Paul, P., Chowdhury, A., & Quamruzzaman, M. (2026). A customized MobileNetV3Large-based deep learning framework for plant disease detection. *Discover Artificial Intelligence*, *6*(1), 1-8. https://doi.org/10.1007/s44163-025-00733-8
3. Bhargava, A., Shukla, A., Goswami, O. P., Alsharif, M. H., Uthansakul, P., & Uthansakul, M. (2024). Plant Leaf Disease Detection, Classification, and Diagnosis Using Computer Vision and Artificial Intelligence: A Review. *IEEE Access*, *12*, 37443-37469. https://doi.org/10.1109/access.2024.3373001
4. Fan, X., Luo, P., Mu, Y., Zhou, R., Tjahjadi, T., & Ren, Y. (2022). Leaf image based plant disease identification using transfer learning and feature fusion. *Computers and Electronics in Agriculture*, *196*, 106892. https://doi.org/10.1016/j.compag.2022.106892
5. Wang, S., Xu, D., Liang, H., Bai, Y., Li, X., Zhou, J., Su, C., & Wei, W. (2025). Advances in Deep Learning Applications for Plant Disease and Pest Detection: A Review. *Remote Sensing*, *17*(4), 698. https://doi.org/10.3390/rs17040698
6. Gómez-Camperos, J., Jaramillo, H., & Guerrero-Gómez, G. (2021). Digital image processing techniques for detection of pests and diseases in crops: a review. *Ingeniería Y Competitividad*, *24*(1), e20110973. https://doi.org/10.25100/iyc.24i1.10973
7. Nagaraju, M., Chawla, P., & Tiwari, R. (2025). Case Study of Plant Disease Detection and Safe Transportation Using Convolutional Neural Networks: A Systematic Review and Open Challenges. *Springer Nature Link*, 95-118. https://doi.org/10.1007/978-981-97-3222-7_5
8. Nazir, A., Hussain, A., & Assad, A. (2024). CNN in Food Industry: Current Practices and Future Trends. *Taylor and Francis Group Logo*, 329-354. https://doi.org/10.1201/9781032633602-17
9. Shoaib, M., Sadeghi-Niaraki, A., Ali, F., Hussain, I., & Khalid, S. (2025). Leveraging deep learning for plant disease and pest detection: a comprehensive review and future directions. *Frontiers in Plant Science*, *16*, 1-25. https://doi.org/10.3389/fpls.2025.1538163
10. Kong, J., Wang, H., Yang, C., Jin, X., Zuo, M., & Zhang, X. (2022). A Spatial Feature-Enhanced Attention Neural Network with High-Order Pooling Representation for Application in Pest and Disease Recognition. *Agriculture*, *12*(4), 500. https://doi.org/10.3390/agriculture12040500
11. Chen, S.-Y., Kong, C., Feng, F., Sun, B., & Wang, Z.-J. (2024). Apple leaf disease recognition based on an improved ResNet18 neural network. *Shandong Agricultural Sciences*, *56*(10), 174-180.
12. Zhang, J.-H., Kong, F.-T., Wu, J.-Z., Zai, Z.-F., Han, S.-Q., & Cao, S.-S. (2018). Cotton disease identification model based on improved VGG convolution neural network. *Journal of China Agricultural University*, *23*(11), 161-171. https://doi.org/10.11841/j.issn.1007-4333.
13. Long, M.-S., Ouyang, C.-J., Liu, H., & Fu, Q. (2018). Camellia oleifera disease image recognition based on convolutional neural network and transfer learning. *Transactions of the Chinese Society of Agricultural Engineering*, *34*(18), 194-201.
14. Salehi, A. W., Khan, S., Gupta, G., Alabduallah, B. I., Almjally, A., Alsolai, H., Siddiqui, T., & Mellit, A. (2023). A Study of CNN and Transfer Learning in Medical Imaging: Advantages, Challenges, Future Scope. *Sustainability*, *15*(7), 5930. https://doi.org/10.3390/su15075930
15. Shen, L., Sun, Y., Yu, Z., Ding, L., Tian, X., & Tao, D. (2025). On Efficient Training of Large-Scale Deep Learning Models. *ACM Computing Surveys*, *57*(3), 1-36. https://doi.org/10.1145/3700439
16. Zhang, C., Zhang, M., Zhang, S., Jin, D., Zhou, Q., Cai, Z., Zhao, H., Liu, X., & Liu, Z. (2022). Delving Deep into the Generalization of Vision Transformers under Distribution Shifts. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 7267-7276. https://doi.org/10.1109/cvpr52688.2022.00713
17. Zhou, Z. X., & Xia Shang, J. (2024). DF-SE-ResNet: Orange Leaf Disease Classification Method Based on Improved Attention Mechanism and FPN Fusion. *2024 International Conference on Image Processing, Computer Vision and Machine Learning (ICICML)*, 561-566. https://doi.org/10.1109/icicml63543.2024.10958027

18. Ma, Z., wang, S., Su, H., Li, J., Li, y., Bi, Z., & Li, X. (2025). Rice Seed Classification Based On Se-Resnet50. *INMATEH Agricultural Engineering*, *76*(12), 131-141. https://doi.org/10.35633/inmateh-76-12
19. Sagnika, S., Malhotra, M., Deol, I. K., Roy, S., & Kumar, S. (2025). PlantDiseaseNet-RT50: A Fine-tuned ResNet50 Architecture for High-Accuracy Plant Disease Detection Beyond Standard CNNs. *2025 IEEE International Conference on Advances in Computing Research On Science Engineering and Technology (ACROSET)*, 1-6. https://doi.org/10.1109/acroset66531.2025.11281259
20. Alsubai, S., Almadhor, A., Al Hejaili, A., & Gadekallu, T. R. (2025). Intelligent smart sensing with ResNet-PCA and hybrid ML-DNN for sustainable and accurate plant disease detection. *Frontiers in Plant Science*, *16*, 1691415. https://doi.org/10.3389/fpls.2025.1691415
21. Korade, N. B., Salunke, M. B., Bhosle, A. A., Asalkar, G. G., Joshi, D. M., Patil, A. S., & Sangve, S. M. (2025). Tomato Leaf Disease Detection with YOLOV8 Leaf Extraction, Resnet-50 Classification, and Gpt-3.5 for Treatment Recommendations. *International Research Journal of Multidisciplinary Scope*, *06*(01), 879-891. https://doi.org/10.47857/irjms.2025.v06i01.02864
22. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778). https://doi.org/10.1109/cvpr.2016.90
23. Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-Excitation Networks. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7132-7141. https://doi.org/10.1109/cvpr.2018.00745
24. Shi, R.-X. (2024). *Research on Crop Disease Image Recognition Based on Deep Learning*.
25. Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. *Proceedings of the IEEE International Conference on Computer Vision*, 618-626. https://doi.org/10.1109/iccv.2017.74