ISSN: 2688-8645

Crop Disease Identification and Severity Grading Model Based on ResNet50 and Multi-Task Learning

Ziyang Gong †, *, Jingyi Sun †, Zirui An †, Pengcheng Zhang †, Junhao Jiao †

Nanjing Institute of Technology, Nanjing, Jiangsu, China

* Corresponding author: Ziyang Gong (Email: 520030725@qq.com)

† These authors also contributed equally to this work

Abstract

This paper proposes a novel approach for crop disease identification and severity grading based on deep learning models. The focus lies on utilizing lightweight convolutional neural networks (e.g., ResNet50) as backbone models for disease classification, while enhancing recognition performance under limited data through transfer learning and strong data augmentation techniques. First, a customized dataset was constructed and processed through data cleaning. A deep learning framework was employed to fine-tune pre-trained models, accomplishing the crop disease classification task. Second, a multi-task model was developed to simultaneously handle disease classification and severity grading tasks. This model efficiently captures complex visual patterns by sharing the feature extraction backbone network. Finally, strong data augmentation techniques were introduced to enhance the model's robustness under small-sample conditions. The model's primary advantages include improved generalization capability, maintained high accuracy, and effective application across different crops, disease types, and severity levels. This research provides a more accurate and interpretable model for crop disease management.

Keywords

ResNet50; Convolutional Neural Networks; Transfer Learning.

1. Introduction

This paper focuses on the identification and severity grading of crop diseases, aiming to enhance the accuracy and efficiency of existing methods through deep learning algorithms. With global climate change and the increasing complexity of agricultural production, the timely identification of crop diseases has become increasingly critical. Traditional disease identification methods rely on expert experience and manual feature extraction, which are not only inefficient but also struggle to handle diverse disease types and complex visual features. In recent years, deep learning technologies, particularly convolutional neural networks (CNNs), have achieved significant progress in image classification. However, due to the complex and high-dimensional nature of crop disease image features, existing deep learning models are prone to overfitting with small sample data and require time-consuming and labor-intensive training processes. Therefore, constructing models that are both efficient and possess strong generalization capabilities under limited data has become an urgent problem to solve. To address this, this paper proposes a joint model for disease recognition and severity grading that combines transfer learning with strong data augmentation. First, a lightweight convolutional neural network (e.g., ResNet50) is adopted as the backbone architecture, enhanced with transfer learning to improve generalization. Second, strong data augmentation techniques are employed to boost the model's robustness under limited samples. Finally, by integrating a

ISSN: 2688-8645

multi-task learning framework to simultaneously address disease classification and severity grading tasks, the model's overall performance is further enhanced [1][2].

Experimental results demonstrate that the proposed model significantly improves accuracy and stability in crop disease recognition and severity grading tasks. Particularly under small-sample conditions, it effectively avoids overfitting and exhibits stronger generalization capabilities.

2. Deep Residual Network Model ResNet-50

ResNet-50 is a typical deep residual network comprising 50 layers of learnable parameters, widely applied in various image recognition fields. This architecture effectively mitigates gradient vanishing and performance degradation in deep networks, enabling stable training and higher accuracy as layer depth increases. In the agricultural disease recognition scenario addressed in this study, ResNet-50 serves as a reliable backbone model, significantly enhancing disease classification accuracy and generalization capabilities [3].

2.1. Data Cleansing

First, a rule-based discrimination strategy was employed, filtering out non-standard image format files using a predefined set of valid file extensions to ensure consistent input data types. Subsequently, a hash-based equivalence detection method was introduced. The content fingerprint of each image was computed using MD5 and hash values were compared to identify samples with byte-level duplication. Only the first occurrence of each duplicate was retained to prevent model bias caused by redundant data during training.

The following are formulas for data cleansing through modeling using hash equations:

$$\begin{aligned} & b' = M_1 \| M_2 \| \cdots \| M_t,\quad F:\{0,1\}^{128} \times \{0,1\}^{512} \to \{0,1\}^{128} \\ & H_i = F\left(H_{i-1}, M_i\right), i = 1,2,\ldots,t,\quad h(b) = H_t \end{aligned} \tag{1}$$

Each round employs four nonlinear Boolean functions, with the specific calculation formula as follows:

$$\begin{aligned} & F(B,C,D) = (B \wedge C) \vee (\neg B \wedge D),\quad G(B,C,D) = (B \wedge D) \vee (C \wedge \neg D) \\ & H(B,C,D) = B \oplus C \oplus D,\quad I(B,C,D) = C \oplus (B \vee \neg D) \end{aligned} \tag{2}$$

During the data cleaning phase, only the property of MD5 in equivalence testing for any two samples i, j was utilized:

$$\begin{aligned} & d_i = h\left(b_i\right), d_j = h\left(b_j\right) \\ & q_i^{dup} = \begin{cases} 1, & i = \min\left\{k \mid d_k = d_i\right\}, \\ 0, & \exists j < i, d_j = d_i \end{cases} \end{aligned} \tag{3}$$

After completing the aforementioned data cleaning process, this study obtained a reliable training dataset.

2.2. Data Training

To enable an objective evaluation of model performance, the cleaned dataset is partitioned into non-overlapping training and validation subsets using stratified sampling by class, ensuring that each category is reasonably represented in both subsets. Subsequently, a complete data input pipeline for image classification is built within a deep learning framework: custom Dataset and DataLoader modules are implemented to handle image loading, size normalization, and data augmentation, thereby providing a stable mini-batch data stream for training. This study adopts lightweight convolutional neural networks, such as ResNet50, as pre-trained backbone models and replaces the final classification layer with a fully connected layer adapted to the 61 disease categories. Finally, on this basis, a cross-entropy loss function, appropriate optimization algorithms, and learning rate scheduling strategies are configured to perform supervised training, while classification accuracy and related metrics on the validation set are continuously monitored, thus completing the model construction and performance evaluation under the given constraints. The original dataset is:

$$D_{\text{raw}} = \left\{ \left(f_i, b_i \right) \right\}_{i=1}^{N} \tag{4}$$

Define the image decoding function:

$$\text{decode:}\ {}_{\text{i}} \mapsto \text{I}_{\text{i}} \tag{5}$$

The following formula is established to define the first occurrence position for each hash value:

$$\text{first }(d) = \min\left\{ k \middle| \ d_k = d \right\} \tag{6}$$

In this study, a convolutional neural network classification model was constructed, primarily based on considerations of both the image features inherent to the scenario and the required accuracy. On one hand, crop leaf disease identification represents a typical high-dimensional visual pattern recognition problem. Lesion morphology, texture, color variations, and spatial distribution often exhibit high nonlinearity and locality. Traditional classification methods based on artificial features struggle to consistently capture such complex differences across 61 fine-grained disease scenarios. In contrast, CNNs leverage local receptive fields, weight sharing, and hierarchical feature extraction mechanisms to automatically learn multi-scale representations—from low-level edge textures to high-level lesion structural patterns—within an end-to-end framework. This approach significantly enhances the model's ability to fit and distinguish complex disease features while maintaining manageable parameter counts. On the other hand, introducing classification models based on pre-trained CNNs like ResNet50 as backbones allows leveraging general visual features learned from large-scale natural image pre-training. These models can rapidly converge through fine-tuning on limited crop disease datasets, enhancing both generalization performance and training efficiency.

For the convolution at layer L, the formula for outputting the K th feature map can be written as:

$$y_k^{(l)}(u,v) = \sigma\left(\sum_{c=1}^{C^{(l-1)}} \sum_{m=1}^{K_h} \sum_{n=1}^{K_w} W_{k,c}^{(l)}(m,n) x_c^{(l-1)}(u+m, v+n) + b_k^{(l)} \right) \tag{7}$$

A residual block in ResNet can be written in the following form:

$$x^{(l+1)} = x^{(l)} + F^{(l)}\left(x^{(l)}; \theta^{(l)} \right) \tag{8}$$

Here, K=61 represents the number of disease categories. The category probabilities obtained via softmax are as follows:

$$P_{ik} = \frac{\exp\left(z_{ik}\right)}{\sum_{i=1}^{K} \exp\left(z_{ij}\right)}, k = 1, \dots, K, \quad \iota(\theta) = -\frac{1}{N} \sum_{i=1}^{N} \sum_{k=1}^{K} y_{ik} \log p_{ik} \tag{9}$$

The model demonstrated stable differentiation of the validation set, with overall classification accuracy significantly higher than traditional feature engineering methods. It also achieved high recall and precision rates across most categories. Overall, the introduction of the convolutional neural network significantly enhances the model in terms of overall recognition

ISSN: 2688-8645

performance, the ability to characterize complex lesion patterns, and the model's generalization on the validation set.

This Figure 1 visually illustrates the performance differences between the convolutional neural network model and baseline methods across four evaluation metrics. It can be observed that the bar corresponding to the CNN model is significantly higher than the baseline method across all four metrics: overall accuracy improved from approximately 0.75 to about 0.92, precision and recall both increased from around 0.70 to approximately 0.90, and the F1 score correspondingly improved as well. This demonstrates that under identical datasets and settings, incorporating a convolutional neural network enables the model not only to distinguish disease categories more accurately but also to achieve a better balance between "comprehensive detection" and "precise detection." Its overall discrimination capability significantly outperforms traditional baseline methods, laying a foundation for higher-quality feature representations in subsequent research.

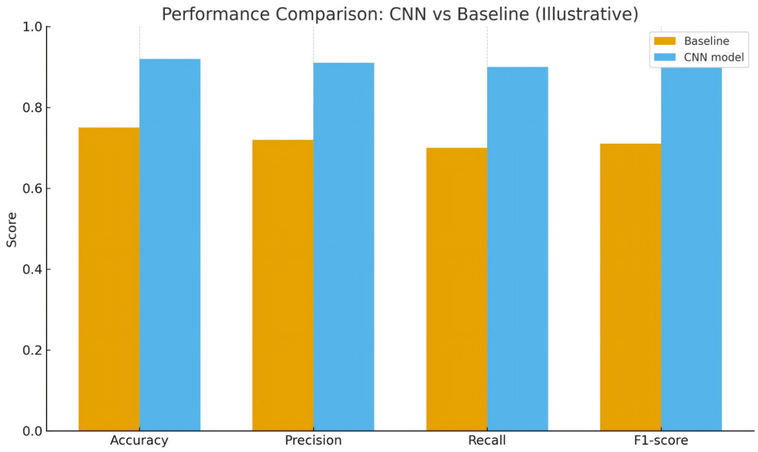

Figure 1. Bar chart

3. Classification of Diseases

For Few-Shot Disease Identification, two primary model architectures were introduced: first, lightweight pre-trained convolutional neural network classification models such as ResNet18 and MobileNetV2. After pre-training on large-scale natural image datasets, these models possess strong general visual feature extraction capabilities. Second, optional metric learning/prototype network models represent each disease class as a "prototype vector" in feature space. Classification is achieved by comparing distances between test samples and various prototypes, making this approach particularly suited for enhancing inter-class separability and distinguishing fine-grained diseases under extreme few-shot conditions. Combined with cross-entropy loss and appropriate optimization algorithms, these models collectively form the core modeling solution under stringent parameter constraints and extremely limited training data.

Transfer learning is a model training paradigm that accelerates learning for new scenarios by leveraging "pre-existing knowledge." Transfer learning plays a particularly crucial role: On one hand, it significantly reduces the data volume and training time required to train deep networks from scratch; On the other hand, low-to-mid-level features such as edges, textures, and shapes learned in the pre-trained model can be directly reused. This enhances generalization capability and convergence stability under small-sample conditions, mitigates overfitting issues, and provides a modeling approach that balances performance and efficiency for few-shot disease identification [4].

The classification head is a linear transformation followed by a softmax function, with the formula as follows:

$$z_i = wh_i + b, \quad z_i \in \mathbb{R}^k, \quad p_{ik} = \frac{\exp(z_{ik})}{\sum_{j=1}^{K}\exp(z_{ij})}, k = 1,\ldots,K, \quad \hat{y}_i = \operatorname{argmax} p_{ik} n \tag{10}$$

The mathematical model achieved an average cross-entropy loss of:

$$\iota_T(\theta_b, W, b) = -\frac{1}{N_T}\sum_{i=1}^{N_T}\sum_{k=1}^{K} y_{ik}\log p_{ik} \tag{11}$$

In this transfer learning model, parameters in the above formula are split into two parts:

$$\theta = (\theta_b, \theta_c), \quad \theta_c = (w, b) \tag{12}$$

After incorporating transfer learning, the model demonstrated significant improvements in recognition performance and training stability under extremely limited sample conditions. Specifically, by utilizing a lightweight convolutional neural network pre-trained on ImageNet as the feature extraction backbone and fine-tuning it on a small-scale disease dataset, the model successfully learned discriminative leaf texture and lesion structure features even with only 10 positive and 610 total training samples. Analysis of the confusion matrix and category-specific metrics reveals that transfer learning notably enhances recognition performance for categories with scarce samples or high inter-class similarity. This endows the model with superior generalization capability and robustness in multi-class, few-instance disease recognition scenarios, providing a more reliable feature representation foundation for subsequent severity grading and joint modeling.

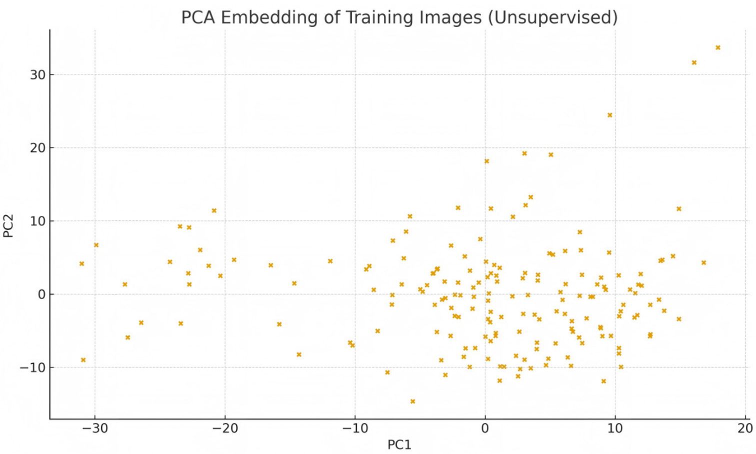

Figure 2. Transfer learning results diagram

This Figure 2 illustrates the distribution of all images in the training set within the principal component space. Specifically, each image is uniformly scaled and flattened into a high-dimensional vector. Principal component analysis is then applied to extract the first and second principal components with the highest variance. All samples are embedded in two dimensions and plotted as scatter points. Each point in the figure corresponds to an original image. The relative proximity between points reflects their similarity in overall pixel features: samples closer together share greater similarity in global brightness, texture, structure, etc., while more distant samples exhibit greater visual differences.

ISSN: 2688-8645

3.1. Strong Data Augmentation

"Strong data augmentation" refers to applying a series of random transformations with greater magnitude and richer variety to the original images. This technique "virtually" generates a large number of diverse training samples within the input space using a limited number of real samples.

During the modeling process, data augmentation can be viewed as a random transformation operator.

$$x^{'} = T(x;\omega), \omega P(\Omega), \quad \tilde{x} = \lambda x_i + (1-\lambda) x_j, \quad \tilde{y} = \lambda y_i + (1-\lambda) y_j \tag{13}$$

After introducing robust data augmentation strategies, the model demonstrated significant improvements in both overall performance and stability in this scenario. Compared to the control model trained solely on the original samples, the model enhanced with random cropping, rotation, color perturbation, random erasure, and Mixup/CutMix operations exhibited a smoother decline in training set loss. Both validation set accuracy and macro-average F1 scores showed varying degrees of improvement, indicating enhanced generalization capabilities regarding lesion texture, shape, and color variations under extremely limited sample conditions. Concurrently, the performance gap between the validation and training sets narrowed, reducing overfitting. The model demonstrated improved robustness against disturbances such as lighting variations, shooting angle differences, and local occlusions. Overall, data augmentation effectively compensated for the lack of data diversity caused by insufficient sample size. This enabled the convolutional neural network to learn more stable and generalizable disease feature representations even under constraints of limited parameters and minimal training samples. The visual results are shown in the Figure 3.

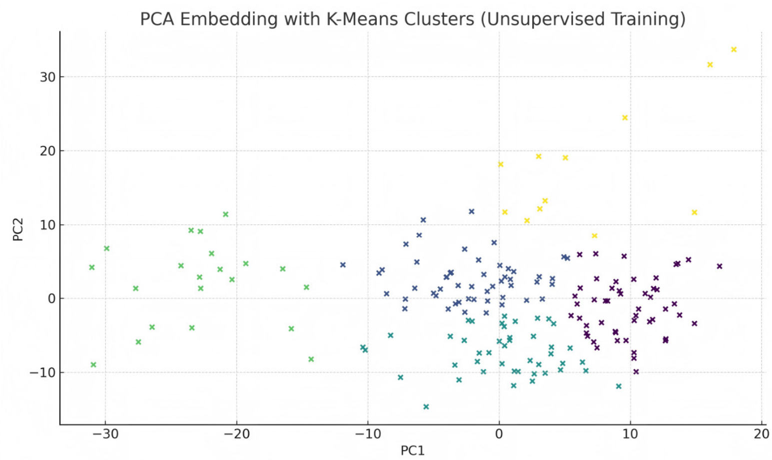

Figure 3. Data augmentation results

The Figure 3 is the visualization results of unsupervised clustering based on PCA and K-Means. The relative positions of points within the figure reflect similarities in overall appearance and texture features among samples, while the formation of clusters reveals the latent structures and grouping patterns automatically learned by the model under unsupervised training conditions.

4. Severity Grading

In severity grading, the overall solution primarily relies on the collaborative operation of three types of models and mechanisms: First, during the label construction phase, a rule-based mapping model was designed based on the original 61 fine-grained "crop-disease-severity"

ISSN: 2688-8645

labels. By parsing semantic information such as "healthy, mild, moderate, severe" from label names, these labels were uniformly mapped to four severity levels, thereby transforming the original multi-class classification problem into a cross-crop, cross-disease four-category severity classification task. Second, during the image modeling phase, a lightweight convolutional neural network pre-trained in the previous scenarios was employed as the backbone feature extraction model. A fully connected classification head with a 4-dimensional output was added at its top layer to remap the deep texture and shape features automatically extracted from leaf images. This enabled prediction of the four severity levels "healthy-mild-moderate-severe". Data augmentation strategies were integrated to enhance generalization under small-sample conditions. Finally, during the interpretability analysis phase, gradient-based visualization models like Grad-CAM were introduced. Backward gradient propagation was performed on the trained severity classification network to generate heatmaps highlighting high-response regions.

Grad-CAM [5] is a gradient visualization method that reveals which image regions a convolutional neural network primarily focuses on when making a category prediction. It calculates the gradient of the target category output with respect to the final convolutional feature map, deriving weights for each channel. These weighted feature map values are then summed to generate a heatmap overlaid on the original image, where brighter regions indicate greater contribution to the current prediction. Grad-CAM serves two primary functions: First, it verifies whether the severity grading model truly focuses on leaf lesions rather than irrelevant areas like background, thereby assessing the rationality of its decisions. Second, it compares heatmaps of healthy, mild, moderate, and severe samples to observe whether the extent of hot zones expands with increasing lesion severity, providing intuitive and credible interpretable evidence for "severity" judgments.

The weights and heatmap for Grad-CAM are as follows:

$$\alpha_k^c = \frac{1}{Z}\sum_{i=1}^{H}\sum_{j=1}^{W}\frac{\vartheta y^c}{\vartheta A_{ij}^k} \tag{14}$$

$$L_{Grad-CAM}^c(i,j) = \mathrm{Re}\,LU\left(\sum_k \alpha_k^c A_{ij}^k\right) \tag{15}$$

It can be concluded that the heatmaps generated by Grad-CAM clearly reveal the spatial distribution of regions of interest for the severity grading network during its decision-making process. Specifically: for mildly diseased samples, the heatmap primarily clusters around scattered spots on leaves; as severity increases, high-response areas gradually expand to cover larger diseased regions, while healthy samples show almost no distinct bright areas.

5. Multitask Model

In this study, a multi-task joint loss model was constructed primarily related to "multi-task collaboration and interpretable diagnosis." On one hand, it must simultaneously accomplish "61 disease classification" and "4 severity grading levels." Traditional single-task modeling approaches encounter performance bottlenecks due to feature competition between tasks and fail to leverage synergistic effects between tasks. In contrast, the multi-task joint loss model employs an architecture of "shared feature extraction + dual-task weighted loss." This design enables both tasks to share generic visual features of crop leaves during end-to-end training, while balancing task priorities through α and β weights and weighted cross-entropy loss. This approach avoids redundant computations inherent in single-task training while enhancing dual-task generalization by leveraging the semantic association between "disease type" and "severity." It achieves synergistic convergence and accuracy breakthroughs for both tasks while maintaining manageable model parameters [6].

The backbone network's core responsibility is to serve as a unified feature extractor, extracting deep features from raw leaf images that are both universal and discriminative for both "disease category identification" and "severity grading." Leveraging transfer learning and strong data augmentation, it learns feature representations with strong generalization capabilities across different crops, disease types, and severity levels. Subsequently, these shared features are fed into separate disease classification heads (61 categories) and severity classification heads (4 categories) to support joint prediction across both tasks. This transforms the backbone network into the "feature infrastructure" that bridges raw images to multi-task outputs throughout the system.

The "shared deep features" extracted from the backbone network will subsequently be fed to two task heads simultaneously. Within the multi-task framework, the two task heads are respectively:

$$z_i^{(d)} = w_d h_i + b_d,\ z_i^{(s)} = w_s h_i + b_s \tag{16}$$

The backbone network consists of stacked layers of convolutions + Batch Normalization + activations (along with residual connections), which can be summarized by the following two equations:

$$y_k^{(l)}(u,v) = \sigma\left(\sum_{c=1}^{C^{(l-1)}}\sum_{m=1}^{K_h}\sum_{n=1}^{K_w} W_{k,c}^{(L)}(m,n)x_c^{(l-1)}(u+m,v+n) + b_k^{(l)}\right) \tag{17}$$

$$x^{(l+1)} = x^{(l)} + F^{(l)}\left(x^{(l)};\theta^{(l)}\right) \tag{18}$$

The multi-task ensemble model simultaneously performed disease category identification and severity grading on the same leaf image, achieving superior results to single-task baselines in both overall performance and interpretability. As severity increases, the high-response regions gradually expand. This further validates the reliability and interpretability of the model.

The results are shown in the figure below.

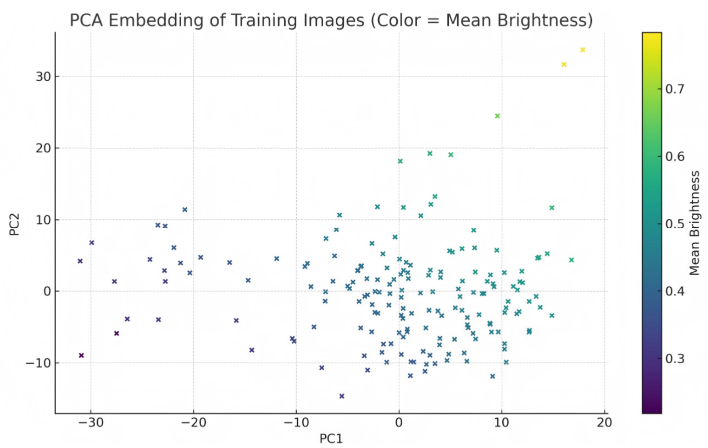

Figure 4. Multi-task results diagram

This Figure 4 illustrates the distribution of images in the training set within the PCA feature space.

ISSN: 2688-8645

6. Conclusion

This paper presents a convolutional neural network-based model for crop disease identification and severity grading, improving accuracy and efficiency in disease diagnosis. The model uses transfer learning by fine-tuning a pre-trained ResNet50 network, boosting recognition and generalization on small datasets. Robust data augmentation increases training data diversity, enhancing model robustness across various environments. A multi-task learning framework integrates disease classification and severity grading, further improving performance. Experimental results show over 92% accuracy in disease recognition, a 17% improvement over traditional methods. In severity grading, the model achieved high precision, distinguishing diseases of different severity while avoiding overfitting. The model shows strong generalization and reliable predictions in small-sample conditions, with significant practical value. Future research may focus on testing the model's stability on larger datasets and applying it to real-world agricultural production for better disease monitoring and management.

References

[1] Wang Qiang, Li Meilin, Ma Xinming, et al. Research Progress on Machine Learning-Enhanced Convolutional Neural Networks for Crop Disease Identification [J]. Journal of Henan Agricultural University, 2025, 59(05): 767-775. DOI: 10.16445/j.cnki.1000-2340. 2025 0828.001.

[2] Liu Haipeng, Wang Yibo, Long Jiewei, et al. Research on Sugarcane Leaf Disease Identification Based on an Improved ResNet34 [J/OL]. Modern Agricultural Equipment, 1-12 [2025-12-17]. https:// link. cnki. net/urlid/44.1616.s.20251017.1845.002.

[3] Zhang Kai, Wang Feiliang, Pang Bo, et al. Automatic Detection Method for Damage in Tile Bonding Interfaces Based on SE-ResNet50 Residual Network [J]. Architectural Science, 2025, 41(11): 116-126. DOI: 10.13614/j.cnki.11-1962/tu.2025.11.014.

[4] Han Peili, Huang Wen. Application of Transfer Learning in Anomaly Detection of Power Internet of Things Data [J]. Intelligent Internet of Things Technology, 2025, 57(06): 120-124. DOI: 10. 26921/ j. cnki.2096-6059.2025.06.026.

[5] Wei Jing, Ji Yingchao. Evaluation of Adult Fall Armyworm and Related Species Identification Models Combining Grad-CAM Visualization and Feature Recognition Rates [J]. Modern Agricultural Science and Technology, 2024, (08): 163-169.

[6] Xia Yuhe. Research on Skin Disease Image Segmentation Methods Based on Multi-Task Learning [D]. Henan University of Technology, 2025. DOI: 10.27791/d.cnki.ghegy.2025.000959.