Weakly Supervised Improved CBAM-ResNet18 Model for Recognition of Multiple Apple Leaf Diseases: Postprint

Authors: Zhang Wenjing, Jiang Zezhong, Qin Lifeng

Date: 2023-05-15T00:00:00+00:00

Abstract

To address the problem of low recognition accuracy for apple leaf disease images under weak supervision with only image-level annotations, an improved CBAM-ResNet algorithm is proposed for apple leaf disease recognition. Using ResNet18 as the base model, the Multilayer Perceptron (MLP) in the channel attention module of the lightweight Convolutional Block Attention Module (CBAM) attention mechanism is improved through dimensionality increase to amplify disease feature details; the improved CBAM is integrated into residual modules to enhance extraction of key detail features; AlphaDropout combined with SeLU (Scaled Exponential Linear Units) is integrated into the network to prevent overfitting and accelerate model convergence; finally, a one-cycle cosine annealing algorithm is employed to adjust the learning rate to obtain the disease recognition model. Training is conducted under weak supervision where sample images are only annotated at the image level, significantly reducing annotation costs. Through ablation experiments, the optimal dimensionality increase factor for the MLP in the improved CBAM is determined to be 2, which improves accuracy by 0.32% compared to the original CBAM, and reduces training time per epoch by 8 seconds despite a 17.59% increase in parameters. Experimental testing was conducted on a dataset of 6185 images covering five diseases including apple Alternaria blotch, brown spot, mosaic disease, grey spot, and rust. The results show that under weak supervision, the model achieves an average recognition accuracy of 98.44% for the five apple diseases, with the improved CBAM-ResNet18 achieving a 1.47% improvement over the original ResNet18, and outperforming the comparison models VGG16, DenseNet121, ResNet50, ResNeXt50, EfficientNet-B0, and Xception. In terms of learning efficiency, the improved CBAM-ResNet18 reduces training time per epoch by 6 seconds compared to ResNet18 despite a 24.9% increase in parameters, and completes model training at the fastest speed of 137 seconds per epoch among the comparison models VGG16, DenseNet121, ResNet50, ResNeXt50, EfficientNet-B0,

and Xception. Based on confusion matrix results, the model's average precision, average recall, and average F1-score reach 98.43%, 98.46%, and 0.9845, respectively. These results demonstrate that the improved CBAM-ResNet model can perform apple leaf disease recognition with favorable results, providing technical support for intelligent apple leaf disease recognition.

Full Text

Identifying Multiple Apple Leaf Diseases Based on the Improved CBAM-ResNet18 Model Under Weak Supervision

ZHANG Wenjing[1],[2], **JIANG Zezhong**[1], **QIN Lifeng**[1],[3]* [1]College of Mechanical and Electronic Engineering, Northwest A&F University, Yangling, Shaanxi, China [2]Key Laboratory of Agricultural Internet of Things, Ministry of Agriculture and Rural Affairs, Yangling, Shaanxi, China [3]Key Laboratory of Agricultural Information Perception and Intelligent Services, Yangling, Shaanxi, China

Abstract: To address the problem of low recognition accuracy for apple leaf disease images under weak supervision with only image-level annotations, this paper proposes an improved CBAM-ResNet algorithm for apple leaf disease identification. Using ResNet18 as the base model, the multilayer perceptron (MLP) in the channel attention module of the lightweight Convolutional Block Attention Module (CBAM) attention mechanism is improved through dimensionality expansion to amplify the feature details of apple leaf diseases. The improved CBAM is integrated into the residual modules to enhance key detailed features, while AlphaDropout and Scaled Exponential Linear Units (SeLU) are incorporated into the network to prevent overfitting and accelerate model convergence. Finally, a single-cycle cosine annealing algorithm is employed to adjust the learning rate, yielding the disease recognition model. Training is conducted under weak supervision where all sample images are only annotated with image-level labels, significantly reducing annotation costs. Through ablation experiments, the optimal dimensionality expansion ratio for the improved MLP is determined to be 2×. Compared with the original CBAM, this achieves a 0.32% accuracy improvement while reducing per-epoch training time by 8 seconds, despite a 17.59% increase in parameters. Tests were conducted on a dataset of 6,185 images containing five diseases: apple alternaria leaf spot, brown spot, mosaic, gray spot, and rust. The results demonstrate that under weakly supervised learning, the model achieves an average recognition accuracy of 98.44% for the five diseases. In terms of recognition accuracy, the model outperforms ResNet18 by 1.47% and surpasses VGG16, DenseNet121, ResNet50, ResNeXt50, EfficientNet-B0, and Xception. Regarding learning efficiency, the improved CBAM-ResNet18 reduces per-epoch training time by 6 seconds compared to the original ResNet18, despite a 24.9% parameter increase, and completes training faster than all control models at 137 seconds per epoch. Analysis of the confusion matrix reveals that the model's average precision, average recall, and average F1-score reach

98.43%, 98.46%, and 0.9845, respectively. These results indicate that the improved CBAM-ResNet model can effectively identify apple leaf diseases with high accuracy and provide technical support for intelligent apple leaf disease recognition.

Keywords: disease identification; residual network; attention mechanism; cosine annealing learning rate; transfer learning; convolutional block attention module; multilayer perceptron

1 Introduction

Disease is a critical factor causing apple yield reduction, quality degradation, and decreased commercial value. Timely and accurate disease identification is essential for prompt prevention and control to minimize economic losses. Early disease identification primarily relied on manual field inspection, which is time-consuming, labor-intensive, and susceptible to subjective influences, resulting in low accuracy and efficiency. Digital image processing and machine learning-based disease recognition technologies have significantly advanced disease detection, identification, and diagnosis levels. However, traditional machine learning techniques face difficulties in image feature extraction and exhibit insufficient robustness and accuracy when dealing with complex backgrounds.

In recent years, deep learning technology has been extensively studied for crop disease identification and has achieved excellent results. Zhu et al. [**?**] utilized Inception V2 with Batch Normalization (BN) to provide multi-scale image features for a Region Proposal Network (RPN), achieving good performance in plant leaf recognition under complex background conditions. Ding et al. [**?**] constructed a convolutional capsule network based on the VGG-16 model, improving the noise resistance of lily disease diagnosis models. Li et al. [**?**] proposed a deep learning-based custom backbone for constructing a plant pest video detection system, highlighting information regions to enhance model recognition capability with good robustness. Zhou et al. [**?**] proposed a tomato leaf disease classification and recognition method based on improved MobileNetV3, enabling real-time non-destructive detection of tomato diseases. Although deep learning has achieved obvious advantages in recognition accuracy, it requires a large number of precisely annotated images as a foundation. In disease identification tasks, annotation costs are extremely high due to factors such as specialized domain knowledge requirements, complex backgrounds, and target diversity [**?**]. Unsupervised learning often struggles to achieve good results under complex backgrounds [**?**]. How to utilize the large number of imprecisely annotated samples that exist in practice for model learning while achieving sufficiently high recognition performance is an urgent problem to be solved [**?**]. In this context, weakly supervised learning [**?**, **?**] has attracted significant attention. In image recognition tasks, weakly supervised learning only requires image-level information annotation—i.e., labeling image categories without marking the specific

location of targets in images—to train models for recognition. Durand et al. [**?**] proposed a weakly supervised learning method for deep convolutional neural networks (WILDCAT), which learns multi-level locally enhanced features under spatial invariance constraints, achieving good results in three visual recognition tasks: image classification, weakly supervised object localization, and semantic segmentation.

The key challenge in weakly supervised learning is how to effectively focus on features in images without pixel-level annotations. Since attention mechanisms can automatically focus learning on important regions in images by assigning different weights to each channel, making model learning more flexible, they are widely used in weakly supervised learning. Choe and Shim [**?**] proposed an Attention-Based Dropout Layer (ADL) that utilizes self-attention mechanisms to process model feature maps, highlighting information regions to improve model recognition capability. Regarding deep features for determining localization, Zhou et al. [**?**] proposed adjustments to the global average pooling layer to enable convolutional neural networks to retain excellent localization properties.

For apple disease identification problems, although strongly supervised learning networks achieve good recognition results [**?**], they suffer from high annotation costs and limited disease samples. Existing weakly supervised learning methods still cannot quickly and effectively learn leaf disease detail features from apple disease images with complex backgrounds, resulting in low learning efficiency and unstable convergence. To address the lack of high-precision pixel-level annotated datasets and the inability of current weakly supervised methods to effectively focus on dense, small-target apple leaf diseases—leading to low model learning efficiency and unstable convergence—this paper proposes an improved CBAM-ResNet18 model. This model employs dimensionality-expanded multilayer perceptron (MLP) to improve the channel attention mechanism in the Convolutional Block Attention Module (CBAM) and integrates the improved CBAM into the ResNet18 model. AlphaDropout modules are introduced into the model along with the Scaled Exponential Linear Units (SeLU) activation function to prevent overfitting, and a single-cycle cosine annealing algorithm is used to adjust the learning rate. Trained under weak supervision using only image-level category labels, the model achieves accurate identification of apple leaf diseases.

2.1 Apple Leaf Disease Image Acquisition

Alternaria leaf spot, brown spot, mosaic, gray spot, and rust occur in large numbers and have wide distribution ranges on apple leaves. These diseases appear as small, densely distributed lesions that require high detail feature extraction, making them difficult to identify. Therefore, this study selected these five apple leaf diseases as recognition targets. The image data used in this study were partly collected from 2019 by Zhou Minmin [**?**] at three experimental stations: Baishui Apple Experimental Station (109°33 32 N,

35°12 45 E), Luochuan Apple Experimental Station (109°22 35 N, 35°47 25 E), and Qingcheng Apple Experimental Station (107°55 36 N, 36°0 30 E) of Northwest A&F University. A total of 3,217 images were collected under different weather conditions (sunny, cloudy, rainy) with a resolution of $512\$\times512 pixels. The other part was collected from July to October 2022 in an apple orchard in Qian County, Shaanx$ pixels. Both datasets contain the five common diseases—alternaria leaf spot, brown spot, mosaic, gray spot, and rust—under both simple and complex backgrounds, as shown in Figure 1 [Figure 1: see original paper].

2.2.1 Data Augmentation

To improve model generalization capability, anti-interference ability under complex backgrounds, and avoid training overfitting, image samples were augmented [**?**]. In the pre-training dataset collected by Zhou Minmin [**?**], 11 data augmentation methods were applied: two image rotations, one horizontal and vertical flip, two sharpness adjustments, two brightness adjustments, two contrast adjustments, and one Gaussian blur, resulting in 24,348 disease images in the final pre-training set. For images captured in the Qian County orchard, four augmentation types—rotation, shift, flip, and brightness variation—were used to expand the dataset, yielding 6,185 disease photos. In this study, only image-level category annotations were performed without pixel-level labeling, significantly reducing annotation costs.

3.1 ResNet Model

To address gradient vanishing and explosion problems in deep networks, He et al. [**?**] proposed the Residual Network (ResNet) based on skip connections and identity mapping in 2015, as shown in Figure 2 [Figure 2: see original paper]. Let x be the input signal, which undergoes two linear transformations to obtain $H(x)$. During model learning, the output signal $H(x)$ tends to stabilize, and the originally weighted x becomes an identity mapping, making the output signal $H(x)$ equal to the input signal x [**?**]. For apple leaf disease identification, since most scenarios require rapid and correct disease recognition on mobile devices, a residual network model with fewer parameters should be adopted. Therefore, this study employs ResNet18.

3.2 CBAM Attention Mechanism and Its Improvement

Traditional convolutional neural networks transmit features indiscriminately to the next layer, making it impossible to focus on effective information, especially under weak supervision conditions. To address this issue, this paper integrates an attention mechanism into the convolutional neural network. CBAM [**?**] consists of a Channel Attention Module (CAM) [**?**] combined with a Spatial Attention Module (SAM) [**?**]. It first applies average pooling and global pooling to the feature map U, compressing it to $1 \times 1 \times C$ format. The two compressed feature maps are then fed into an MLP with a two-layer neural network containing a

hidden layer in the middle, which reduces the dimensionality of the compressed $1 \times 1 \times C$ feature map to $1 \times 1 \times C/r$. The MLP outputs undergo element-wise addition, and finally, a Sigmoid function generates channel attention feature weights that act on the feature map U to produce a new feature map. The spatial attention mechanism complements channel attention by extracting key feature information in the image space.

Since apple leaf diseases are small in size and densely distributed, accurate and rapid recognition requires high detail feature extraction. To enable the shared neural network in the CBAM channel attention module to better capture disease details, this paper proposes an improved structure, as shown in Figure 3 [Figure 3: see original paper]. The original shared neural network first reduced the $1 \times 1 \times C$ feature map to $1 \times 1 \times C/r$ and then expanded it to $1 \times 1 \times C$. However, dimensionality reduction causes significant detail loss, affecting the final generated channel attention weights and failing to emphasize detailed features. The improved CBAM attention mechanism reverses this process by first expanding to $1 \times 1 \times rC$ and then reducing to $1 \times 1 \times C$, thereby enhancing CBAM's ability to distinguish apple leaf disease features.

3.3 SeLU Activation Function

When models are large with numerous parameters, training is prone to overfitting. To address this, Dropout is introduced to achieve regularization effects. Dropout deactivates some neurons to improve model generalization, but the distribution of activation values may change after each Dropout operation. In response, Klambauer et al. [**?**] proposed AlphaDropout and a new activation function SeLU that can maintain consistent mean and standard deviation between input and output. Compared to the ReLU (Rectified Linear Activation Function) [**?**], SeLU has no dead zone, as shown in Equation (1), and its influence exists in a saturation region at negative infinity. After SeLU activation, the sample distribution is automatically normalized to zero mean and unit variance, ensuring that gradients do not explode or vanish during training.

$$\text{SeLU} = \lambda \begin{cases} x & \text{if } x > 0 \\ \alpha e^x - \alpha & \text{if } x \leq 0 \end{cases}$$

where λ and α are hyperparameters, with $\lambda \approx 1.05$ and $\alpha\lambda \approx 1.67$; x is the input value. This study combines AlphaDropout with SeLU to keep the feature map input distribution unchanged, enabling the model to better prevent overfitting and improve convergence speed and effectiveness.

3.4 Cosine Annealing Algorithm for Learning Rate Adjustment

This study employs a single-cycle cosine annealing algorithm [**?**] to decay the learning rate, as shown in Equation (2). Since this study uses a single cycle, i

is set to 1. $\eta_i^{\max}$ and $\eta_i^{\min}$ represent the maximum and minimum learning rates, respectively, defining the learning rate range. T_{cur} denotes the current epoch number, updated after each batch in every epoch, and T_i is the total number of epochs (100 for pre-training and 50 for the training stage in this study). The learning rate begins with a warmup phase at a very small value to prevent large learning rates from causing model instability. After the warmup phase, cosine annealing optimizes the learning rate.

$$\eta_t = \eta_i^{\min} + \frac{1}{2}(\eta_i^{\max} - \eta_i^{\min})\left(1 + \cos\left(\frac{T_{cur}}{T_i}\pi\right)\right)$$

where $\eta_i^{\min}$ is the minimum learning rate, $\eta_i^{\max}$ is the maximum learning rate, T_{cur} is the current epoch number, and T_i is the total number of epochs.

3.5 Improved ResNet18 Disease Recognition Model

Based on the above improvements, this study proposes an apple leaf disease recognition model called CBAM-ResNet18, with the structure shown in Figure 4 [Figure 4: see original paper]. The improved CBAM attention mechanism is integrated into each Res module of the network to better capture disease detail information and reduce complex background interference. ReLU is replaced with SeLU and combined with AlphaDropout to avoid neuron "death," prevent model overfitting, and maintain consistent input-output distributions. A single-cycle cosine annealing algorithm optimizes the learning rate to maintain stability during convergence and avoid fluctuations.

The input to the improved model is a 224$\times$224$pixelRGBimage.Afterinitialconvolutionalandpoolinglayersfo$ convolution. After Batch Normalization (BN), a CBAM module generates feature weights. In the improved CBAM channel attention mechanism, feature maps undergo channel expansion first, then reduction to $1 \times 1 \times 64$, $1 \times 1 \times 128$, $1 \times 1 \times 256$, and $1 \times 1 \times 512$ in the four Res modules, respectively, to obtain feature coefficients that are multiplied with the input feature map to produce new feature maps. All four Res modules use 3$\times$3$convolutionkernelswithquantitiesof$64, 128, 256, and512, $producingfeaturemapsizesof56\times$56, 28$\times$28, 14 pixels, respectively.

Global Average Pooling (GAP) and Fully Connected (FC) layers are followed by an AlphaDropout module to prevent overfitting. The Softmax function is used for classification, as shown in Equation (3), with the maximum probability value output as the result to complete apple leaf disease identification.

$$\text{Softmax}(x_i) = \frac{e^{x_i}}{\sum_j e^{x_j}}$$

where x_i is the input signal, e^{x_i} is the exponential function of input signal x_i, and $\sum_j e^{x_j}$ is the sum of exponential functions for all input signals x_i.

4.1 Model Training Environment and Parameter Settings

Model training was conducted on a CPU configured as Intel® CoreTM i7-11800H @ 2.30 GHz, with an RTX 3060 Laptop GPU (6 GB VRAM) and 16 GB RAM, based on the Windows 10 operating system. The deep learning algorithm training platform was built using the TensorFlow framework (Python version 3.9.7, TensorFlow version 2.8.0).

Following transfer learning principles, model training in this study is divided into pre-training and training stages. The pre-training stage learns on Zhou Minmin's [**?**] dataset and saves the best model, which is then fine-tuned using the Qian County dataset. In the pre-training stage, the input image resolution is 224$×$224 pixels, batch size is 16, and one iteration through all training samples constitutes one epoch with 100 total epochs. The Adaptive Moment Estimation (ADAM) algorithm optimizes the model with cross-entropy loss function. To maintain stable training while accelerating convergence, the initial learning rate is set to 10^{-4} and rises to 0.001 after the warmup phase, then adjusted using cosine annealing and finally decayed to 0. The training stage sets epochs to 50 and warmup epochs to 10, with other parameters identical to pre-training.

4.2 Ablation Experiments

To investigate the effects of the improved CBAM, AlphaDropout, and single-cycle cosine annealing optimization algorithm under weakly supervised learning conditions, three groups of ablation experiments were conducted.

4.2.1 Impact of Improved CBAM

To explore the impact of improved CBAM on the ResNet18 network and determine the optimal dimensionality expansion ratio for the channel attention mechanism's shared neural network, four different MLP settings were tested on the original ResNet18 network: (1) original ResNet18 with original CBAM, (2)-(4) improved CBAM with shared neural networks expanded to $1 \times 1 \times 2C$, $1 \times 1 \times 3C$, and $1 \times 1 \times 5C$ before reduction to $1 \times 1 \times C$. Training and testing were performed on the same dataset, with results shown in Table 2 .

Table 2. Model Performance Under Different Dimension-Up Strategies of the Shared Neural Network in CBAM

Strategy	Feature Map Size	Accuracy (%)	Parameters	Training Time (s/epoch)
Original CBAM	$1 \times 1 \times C/2$	97.02%	11,889,885	145
$1 \times 1 \times 2C$	$1 \times 1 \times 2C$	**98.44%**	13,981,725	**137**
$1 \times 1 \times 3C$	$1 \times 1 \times 3C$	97.14%	15,376,285	142
$1 \times 1 \times 5C$	$1 \times 1 \times 5C$	97.99%	18,165,405	140
No CBAM	-	96.70%	11,189,893	143

As shown in Table 2, expanding the feature map to $1 \times 1 \times 2C$ yields the best performance, with accuracy 1.3% and 0.45% higher than the other two expansion methods, respectively, and 0.32% higher than the original CBAM attention mechanism. Moreover, the $1 \times 1 \times 2C$ expansion has 9.97% fewer parameters than group (3) and 29.92% fewer than group (4), while being 17.59% higher than the original CBAM. It also achieves the fastest learning speed, reducing training time by 8 s, 5 s, and 3 s compared to groups (1), (3), and (4), respectively. These results indicate that expanding instead of compressing feature maps can amplify disease details and effectively extract image features, but excessive expansion also amplifies complex background noise, affecting recognition. Therefore, considering accuracy, parameters, and learning efficiency, this study selects the improved attention mechanism with $1 \times 1 \times 2C$ expansion.

4.2.2 Impact of AlphaDropout Combined with SeLU

To verify the impact of AlphaDropout combined with SeLU on ResNet18 for apple leaf disease recognition under weak supervision, the original ResNet18 network was tested with AlphaDropout+ReLU, AlphaDropout+SeLU, and the original ReLU. The performance of the three models under the same conditions is shown in Table 3 .

Table 3. Model Performances with Different Activation Functions

Activation Function	Accuracy (%)	Training Time (s/epoch)
ReLU	97.21%	139
AlphaDropout+ReLU	97.21%	139
AlphaDropout+SeLU	**97.34%**	**137**

Compared with ReLU, AlphaDropout+SeLU achieves the highest accuracy of 97.34%, an improvement of 0.13% over AlphaDropout+ReLU, while reducing per-epoch training time by 2 seconds. This is because SeLU effectively avoids the "dead neuron" problem, enhances model expression capability, and its self-normalizing property combined with AlphaDropout ensures zero-mean and unit-standard-deviation outputs, accelerating model convergence.

4.2.3 Impact of Single-Cycle Cosine Annealing Optimization Algorithm

To investigate the impact of the single-cycle cosine annealing optimization algorithm, it was compared with dynamic exponential decay learning rate strategies.

As shown in Table 4 , the cosine annealing optimization algorithm stabilizes the model during the initial warmup phase and later adjusts the learning rate to enable better learning of disease features, achieving 0.09% higher accuracy than dynamic decay. Figure 5 [Figure 5: see original paper] shows the validation accuracy curves using different learning rate strategies. The cosine annealing

optimized learning rate decays reasonably in the later stages, enabling rapid convergence while maintaining high accuracy. In contrast, exponential decay learning rate still exhibits large fluctuations after the same training epochs, resulting in poor convergence.

Table 4. Model Performances with Different Learning Rate Adjustment Methods

Learning Rate Method	Accuracy (%)	Training Time (s/epoch)
Single-cycle cosine annealing	**98.44%**	137
Dynamic exponential decay	98.35%	139
Fixed learning rate	96.70%	143

4.3 Pre-training Process Comparison Analysis

After determining the optimal parameters for the improved model, the performance of the improved CBAM-ResNet18 during pre-training was evaluated. Under identical conditions, ResNet18, ResNet50, DenseNet121, VGG16, Xception, ResNeXt50, EfficientNet-B0, and the improved CBAM-ResNet18 were pre-trained. The accuracy and loss values of each model on the validation set after pre-training are shown in Figure 6 [Figure 6: see original paper].

As seen in Figure 6, except for VGG16, the other seven models show an overall upward trend in validation accuracy and decreasing loss values to a relatively stable range. VGG16 accuracy remains around 21% and cannot converge. Among converged models, EfficientNet-B0 exhibits large fluctuations with accuracy around 60% and poor convergence. ResNet18 and ResNet50 oscillate significantly around 96% accuracy, with ResNet18 showing more pronounced oscillations even dropping to 60-80%, while loss values fluctuate around 0.2. Multi-scale structures Xception and ResNeXt50 converge more stably with accuracy around 96% and minor fluctuations, with loss values stable around 0.1. DenseNet121 achieves stable accuracy and loss values around 96% and 0.1, respectively, performing better than ResNet50. This is because under weak supervision without pixel-level annotations, models have insufficient learning capability for effective disease features, resulting in unstable accuracy and loss values with large fluctuations. In contrast, the improved CBAM-ResNet18 model demonstrates significantly more stable performance.

4.4 Disease Identification Experiments and Results Analysis

To further compare model effectiveness for apple leaf disease identification, the pre-trained models were saved and transferred to the training set for fine-tuning to obtain target models. Disease identification experiments were then conducted on the test set to evaluate classification performance. Model performance was

evaluated using accuracy, parameter count, and training time, with accuracy calculated as shown in Equation (4).

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}}$$

where TP is the number of positive samples correctly predicted as positive, TN is the number of negative samples correctly predicted as negative, FP is the number of negative samples incorrectly predicted as positive, and FN is the number of positive samples incorrectly predicted as negative.

Experimental results are shown in Table 5 .

Table 5. Performance Comparison of Different Classic Networks for Apple Leaf Disease Recognition

Model	Validation Accuracy (%)	Test Accuracy (%)	Parameters	Training Time (s/epoch)
VGG16	21.00	20.51	134,281,029	143
DenseNet121	96.00	98.12	7,042,629	156
ResNet50	96.00	98.18	23,597,957	156
ResNeXt50	95.00	98.29	23,084,933	442
ResNet18	96.00	96.97	11,189,893	143
EfficientNet-B0	90.00	90.12	4,055,969	139
Xception	96.00	97.93	20,778,725	149
CBAM-ResNet18	**98.50**	**98.44**	**13,981,725**	**137**

As shown in Table 5, the improved CBAM-ResNet18 proposed in this study achieves test set accuracies 77.93%, 0.32%, 0.26%, 1.15%, 1.47%, 8.32%, and 0.51% higher than VGG16, DenseNet121, ResNet50, ResNeXt50, ResNet18, EfficientNet-B0, and Xception, respectively. ResNet50 utilizes residual structures to ensure strong feature learning capability, achieving 98.18% accuracy. DenseNet121's fundamental Dense Block structure enhances feature propagation and encourages feature reuse, enabling it to achieve similar accuracy to ResNet50 with only one-third the parameters. ResNeXt50's training time reaches 442 s/epoch due to numerous parallel branches that significantly reduce computational efficiency. Similar to ResNeXt50, Xception's multi-scale structure leads to cross-computation among model parameters, resulting in high computational time. In contrast, the improved CBAM-ResNet18 maintains ResNet18's low parameter count while reducing per-epoch training time by 19 seconds compared to ResNet50 and achieving 0.26% higher accuracy, making it suitable for deployment on hardware terminals with limited storage and computational performance.

Relative to ResNet18, despite a 24.9% parameter increase, the improved model not only avoids the negative impact of increased parameters on learning efficiency but actually improves both learning efficiency and speed, reducing per-epoch training time by 6 seconds. This is because the improved attention mechanism enables convolutional layers to more quickly focus on disease features in feature maps, enhancing feature learning rate and training efficiency while reducing training time despite increased parameters.

4.5 Multi-Class Recognition Confusion Matrix

The confusion matrix is commonly used for multi-class performance evaluation. This study involves five apple leaf diseases. The improved CBAM-ResNet18 model was used to classify the test set data, yielding the confusion matrix shown in Figure 7 [Figure 7: see original paper].

Figure 7. Confusion Matrix of Apple Leaf Disease Test Set

The confusion matrix evaluation metrics typically include Precision, Recall, and F1-score, calculated as shown in Equations (5)-(7).

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100\%$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100\%$$

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

where TP and FN represent the numbers of disease samples correctly and incorrectly identified as a certain class, respectively; FP and TN represent the numbers of disease samples incorrectly and correctly identified as not belonging to a certain class, respectively. Precision is the proportion of correctly predicted samples among all samples predicted as a certain class (%). Recall is the proportion of correctly predicted samples among all samples of the true class (%). F1-score is the harmonic mean of precision and recall, providing a comprehensive evaluation.

Based on the confusion matrix in Figure 7, the experimental results for the five apple leaf disease types are shown in Table 6 .

Table 6. Experimental Results of Precision, Recall, and F1 Scores for 5 Types of Apple Leaf Diseases

Disease Class	Precision (%)	Recall (%)	F1-Score
Alternaria Leaf Spot	98.41	98.00	0.9820
Brown Spot	98.51	99.00	0.9925

Disease Class	Precision (%)	Recall (%)	F1-Score
Mosaic	98.51	98.00	0.9825
Gray Spot	97.00	97.00	0.9700
Rust	98.51	99.00	0.9925
Average	**98.43**	**98.46**	**0.9845**

As shown in Table 6, the improved CBAM-ResNet model achieves good results in precision, recall, and F1-score. From Figure 7, the most misclassifications occur between gray spot and alternaria leaf spot. Ten alternaria leaf spot images were misidentified as gray spot, and seven gray spot images were misidentified as alternaria leaf spot. This is because gray spot appears as yellow-brown circular dots on apple leaves while alternaria leaf spot appears as brown circular dots, with very similar color, shape, and size distribution, making them easily confused. The other three diseases achieve precision, recall, and F1-scores close to 100% with good recognition performance. Four rust images were misidentified as mosaic and one mosaic image as rust due to their similar appearance when densely distributed. Additionally, two gray spot images were misidentified as small-area brown spot because brown spot leaves have green surfaces with brown spots. The overall number of misidentifications is small, indicating that the improved CBAM-ResNet model performs well and can be applied to real-world apple leaf disease identification.

5 Conclusion

To address the problems of low recognition accuracy and learning efficiency for apple leaf disease images under weak supervision with only image-level labels, this study proposes an apple leaf disease recognition model. Based on ResNet18 with inserted CBAM, the MLP in CBAM is improved by first expanding dimensions to better highlight apple leaf disease feature details before reduction, enabling the model to more effectively learn disease features and details. Additionally, improvements to the activation function and learning rate enhance training efficiency with greater focus on disease detail learning. The following conclusions are drawn:

1) Using ResNet18 as the base model, CBAM is inserted into each residual block. The MLP in CBAM is improved by reversing the dimensionality operation from "reduce-then-expand" to "expand-then-reduce," which effectively amplifies disease feature details in apple images and enhances convolutional layer efficiency for disease feature extraction.

2) Under weakly supervised learning conditions, the improved CBAM-ResNet18 network model achieves an average recognition accuracy of 98.44% for five apple leaf diseases. This not only surpasses control models (VGG16, DenseNet121, ResNet50, ResNeXt50, ResNet18, EfficientNet-B0, Xception) in accuracy but also, despite a 24.9% parameter increase

over the original ResNet18, avoids the negative impact of parameter increase on learning efficiency and training time seen in traditional convolutional neural networks. Instead, both learning efficiency and speed are improved, reducing per-epoch training time by 6 seconds compared to the original ResNet18 and outperforming all control models.

3) Through confusion matrix analysis of class-wise recognition results, the average precision, average recall, and average F1-score for the five diseases reach 98.43%, 98.46%, and 0.9845, respectively. Mosaic, brown spot, and rust achieve precision, recall, and F1-scores very close to 100%, verifying the superior performance of the improved CBAM-ResNet18 model in effectively distinguishing apple diseases.

4) Although the weakly supervised learning condition can effectively focus on apple leaf disease feature details to achieve excellent accuracy and learning efficiency, there remains room for improvement in model size and parameter count. Future work will consider establishing a cloud-based apple leaf disease recognition and diagnosis service system and developing a mobile App to meet industrial application requirements.

Conflict of Interest Statement: This study has no conflicts of interest among researchers or with publicly disclosed research results.

References

[1] ZHU X L, ZHU M, REN H E. Method of plant leaf recognition based on improved deep convolutional neural network[J]. Cognitive systems research, 2018, 52: 223-233.

[2] DING Y J, ZHANG J J, LI M Z. Disease detection of lily based on convolutional capsule network[J]. Transactions of the Chinese society for agricultural machinery, 2020, 51(12): 246-251, 331.

[3] LI D S, WANG R J, XIE C J, et al. A recognition method for rice plant diseases and pests video detection based on deep convolutional neural network[J]. Sensors (basel, Switzerland), 2020, 20(3): ID 578.

[4] ZHOU Q L, MA L, CAO L Y, et al. Identification of tomato leaf diseases based on improved lightweight convolutional neural networks MobileNetV3[J]. Smart agriculture, 2022, 4(1): 47-56.

[5] REN D W, WANG Q L, WEI Y C, et al. Progress in weakly supervised learning for visual understanding[J]. Journal of image and graphics, 2022, 27(6): 1768-1798.

[6] MEI S A, YANG H A, YIN Z P. An unsupervised-learning-based approach for automated defect inspection on textured surfaces[J]. IEEE transactions on instrumentation and measurement, 2018, 67(6): 1266-1277.

[7] SUN M J, LYU C Z, HAN Y H, et al. Weakly supervised surface defect

detection based on attention mechanism[J]. Journal of computer-aided design & computer graphics, 2021, 33(6): 920-928.

[8] DESELAERS T, ALEXE B, FERRARI V. Weakly supervised localization and learning with generic knowledge[J]. International journal of computer vision, 2012, 100(3): 275-293.

[9] RUSSAKOVSKY O, LIN Y Q, YU K, et al. Object-centric spatial pooling for image classification[C]// European conference on computer vision. Berlin, Heidelberg, Germany: Springer, 2012: 1-15.

[10] DURAND T, MORDAN T, THOME N, et al. WILDCAT: weakly supervised learning of deep ConvNets for image classification, pointwise localization and segmentation[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2017: 5957-5966.

[11] CHOE J, SHIM H. Attention-based dropout layer for weakly supervised object localization[C]// 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2020: 2214-2223.

[12] ZHOU B L, KHOSLA A, LAPEDRIZA A, et al. Learning deep features for discriminative localization[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2016: 2921-2929.

[13] WANG Y L, WU J F, LAN P, et al. Apple disease identification using improved Faster R-CNN[J]. Journal of forestry engineering, 2022, 7(1): 153-159.

[14] ZHOU M M. Apple foliage diseases recognition in android system with transfer learning-based[D]. Yangling: Northwest A&F University, 2019.

[15] XIE Q J, WU M R, BAO J, et al. Individual pig face recognition combined with attention mechanism[J]. Transactions of the Chinese society of agricultural engineering, 2022, 38(7): 180-188.

[16] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2016: 770-778.

[17] TU X Y, LIU S J, QIAN C. Study on the identification methods of typical cultured fish based on ResNet[J]. Fishery modernization, 2022, 49(3): 81-88.

[18] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional block attention module[C]// Computer Vision-ECCV 2018. Berlin, Heidelberg, Germany: Springer International Publishing, 2018: 3-19.

[19] FU J L, ZHENG H L, MEI T. Look closer to see better: Recurrent attention convolutional neural network for fine-grained image recognition[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2017: 4476-4484.

[20] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2018: 7132-7141.

[21] KLAMBAUER G, UNTERTHINER T, MAYR A, et al. Self-normalizing neural networks[J/OL]. arXiv: 1706.02515, 2017.

[22] GLOROT X, BORDES A, BENGIO Y. Deep sparse rectifier neural networks[C]// Artificial Intelligence and Statistics Conference. Cambridge, US: MIT Press, 2011: 315-323.

[23] LOSHCHILOV I, HUTTER F. SGDR: Stochastic gradient descent with warm restarts[J/OL]. arXiv: 1608.03983, 2016.

ZHANG Wenjing, JIANG Zezhong, QIN Lifeng. Identifying multiple apple leaf diseases based on the improved CBAM-ResNet18 model under weak supervision[J]. Smart Agriculture, 2023, 5(1): 111-121.

(Visit www.smartag.net.cn for free access to the full electronic version)

Note: Figure translations are in progress. See original paper for figures.