# Apple Leaf Disease Identification in Complex Background Based on BAM-Net

Yuxi Gao [1], Zhongzhu Cao [1], Weiwei Cai [2], Gufeng Gong [1],*, Guoxiong Zhou [1],* and Liujun Li [3]

1 College of Computer & Information Engineering, Central South University of Forestry and Technology, Changsha 410004, China
2 School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi 214122, China
3 Department of Soil and Water Systems, University of Idaho, Moscow, ID 83844, USA
* Correspondence: t20090523@csuft.edu.cn (G.G.); t20060599@csuft.edu.cn (G.Z.)

**Abstract:** Apples are susceptible to infection by various pathogens during growth, which induces various leaf diseases and thus affects apple quality and yield. The timely and accurate identification of apple leaf diseases is essential to ensure the high-quality development of the apple industry. In practical applications in orchards, the complex background in which apple leaves are located poses certain difficulties for the identification of leaf diseases. Therefore, this paper suggests a novel approach to identifying and classifying apple leaf diseases in complex backgrounds. First, we used a bilateral filter-based MSRCR algorithm (BF-MSRCR) to pre-process the images, aiming to highlight the color and texture features of leaves and to reduce the difficulty of extracting leaf disease features with subsequent networks. Then, BAM-Net, with ConvNext-T as the backbone network, was designed to achieve an accurate classification of apple leaf diseases. In this network, we used the aggregate coordinate attention mechanism (ACAM) to strengthen the network's attention to disease feature regions and to suppress the interference of redundant background information. Then, the multi-scale feature refinement module (MFRM) was used to further identify deeper disease features and to improve the network's ability to discriminate between similar disease features. In our self-made complex background apple leaf disease dataset, the proposed method achieved 95.64% accuracy, 95.62% precision, 95.89% recall, and a 95.25% F1-score. Compared with existing methods, BAM-Net has higher disease recognition accuracy and classification results. It is worth mentioning that BAM-Net still performs well when applied to the task of the leaf disease identification of other crops in the PlantVillage public dataset. This indicates that BAM-Net has good generalization ability. Therefore, the method proposed in this paper can be helpful for apple disease control in modern agriculture, and it also provides a new reference for the disease identification of other crops.

**Keywords:** digital agriculture; apple leaf; disease recognition; deep learning

## 1. Introduction

Apples are one of the world's most popular fruits [1]. The apple's cultivation area and production is increasing year after year due to its high economic and nutritional value [2]. Apples are susceptible to various pathogens during the growing process, which can lead to a variety of diseases. Lower yields, lower quality, and even serious economic losses can result from this. Apple disease symptoms are typically first seen on the foliage [3], and the type of disease that infects apples can be effectively determined by observing the foliage characteristics. Therefore, the timely and effective identification of apple leaf diseases and the determination of disease types are important for the accurate control of apple diseases and a reduction in economic losses.

Traditionally, plant leaf disease diagnosis has relied on manual observation by farmers or experts, which requires farmer experience and expert knowledge [4]. This is inefficient and vulnerable to subjective human factors. With the advancement of computer vision

technology [5], new opportunities for detecting leaf diseases have emerged. For example, Chakraborty et al. [6] used the Otsu thresholding algorithm to binarize leaf images; in addition, they also used the SVM algorithm to classify the binarized images of apple leaves into healthy and infected leaves. Chuan et al. [7] used the thresholding algorithm to segment leaf images and used the genetic algorithm (GA) for feature extraction to select important color, texture, and shape features to input into the SVM model for classification, with a 90% average accuracy. Zamani et al. [8] enhanced the leaf images with histogram equalization, then used the K-means algorithm to segment the images and extract the leaf features with principal component analysis. To achieve this, they used the random forests algorithm to classify the leaf disease images. Tan et al. [9] used the 2D-Renyi algorithm to segment the images, then fed the image features into the ABC-SVM classifier to determine the disease type, achieving an accuracy of 98.45%. To a certain extent, the above methods can classify plant diseases, but they rely too heavily on the manual extraction of shallow features [10], such as color and texture, while ignoring deeper features in the images. As a result, the method is less robust and incapable of adapting to changing natural environments.

Deep learning models have been widely used in a variety of livelihood areas [11], such as forest smoke detection [12,13], crack detection [14,15], and fruit species classification [16,17], due to their excellent algorithmic performance and their ability to obtain deep features directly from images. In recent years, certain researchers have chosen to use deep learning approaches for plant leaf disease identification and classification. Jiang et al. [18] proposed a deep CNN model for apple leaf disease detection based on a convolutional neural network (CNN) combined with the Inception module, which achieved a 78.80% mAP in the detection of five common apple leaf diseases. Bansal et al. [19] used image enhancement algorithms to preprocess the dataset, and then used multiple sets of pre-trained deep learning models (DenseNet121, Efficient-NetB7) to achieve a 96.25% disease classification accuracy. Suo et al. [20] introduced an attention mechanism with an asymmetric multiscale module based on CoAtNet to achieve a 95.95% accuracy in the classification of grape leaf disease. Liao et al. [21] achieved the highest recognition rate of 95.79% for strawberry leaf disease detection using a two-channel residual network with a multi-directional attention mechanism. Sun et al. [22] proposed using the mean SSD for real-time apple leaf disease detection, which achieved an 83.12% mAP at 12.53 FPS. Using multi-level feature fusion, Yang et al. [23] improved the EfficientNet network to achieve a 99.11% recognition accuracy. Bi et al. [24] introduced efficient channel attention (ECA) and dilated convolution on top of MobilenetV3 and achieved a 98.23% accuracy on their maize leaf disease identification task.

Although the abovementioned studies have achieved excellent results in plant leaf disease identification, they mostly focus on disease identification scenarios with a single background. This is not applicable to leaf disease recognition in a field environment with complex background interferences. Therefore, inspired by the above studies, this paper aims to achieve efficient apple leaf disease identification in the context of complex backgrounds with the use of deep learning methods. On this basis, this paper focuses on the following three problems: (1) in the actual detection of orchards, leaf images are easily disturbed by noise due to the filming equipment and filming environment, resulting in unclear leaf disease features in the images, thus influencing the subsequent disease recognition network; (2) interference from complex backgrounds makes it difficult for the network to detect small diseases on the leaves, resulting in inaccurate feature extraction; and (3) apple leaf diseases have inter-class similarities, meaning that different diseases may present similar features [25]. Considering only the shallow features of leaf diseases is not sufficient to correctly diagnose the disease.

To address the above three problems, we propose a new method for apple leaf disease identification in the context of complex backgrounds, the framework of which is shown in Figure 1. The primary contributions of this manuscript are as follows:

(1) An MSRCR algorithm based on bilateral filtering is used to perform preprocessing operations on the images. This method replaces the center-surround function of the conventional MSRCR algorithm with a bilateral filtering function, which enhances color while preserving texture features in the image, resulting in clearer images that facilitate the neural network's extraction of leaf features.

(2) To achieve the efficient identification and classification of apple leaf diseases in complex backgrounds, this paper proposes a BAM-Net, which is designed as follows:

   a. A method called the aggregate coordinate attention mechanism (ACAM) is proposed to address the issue of interference from background information in the context of feature extraction while against complex backgrounds. This method assigns feature weights in both the horizontal and vertical directions, then uses pointwise convolution to correct the weights, improving the network's focus on disease features and filtering out redundant interference information.

   b. A multi-scale feature refinement module (MFRM) is proposed to address the issue of misclassification caused by the inter-class similarity of leaf diseases. This module extracts feature information from multiple scales and refines deep features through cascaded channel information interactions, identifying disease feature information similarities and differences.

(3) The proposed method in this paper achieved a recognition accuracy of 95.64% and an F1-score of 95.25% in a self-made dataset of apple leaf diseases in the context of complex backgrounds. Compared with other methods, BAM-Net has a higher recognition efficiency, which provides a reference value for modern producers to detect and identify apple leaf diseases in a timely fashion. Additionally, it provides significant help for the early maintenance and production of agriculture.
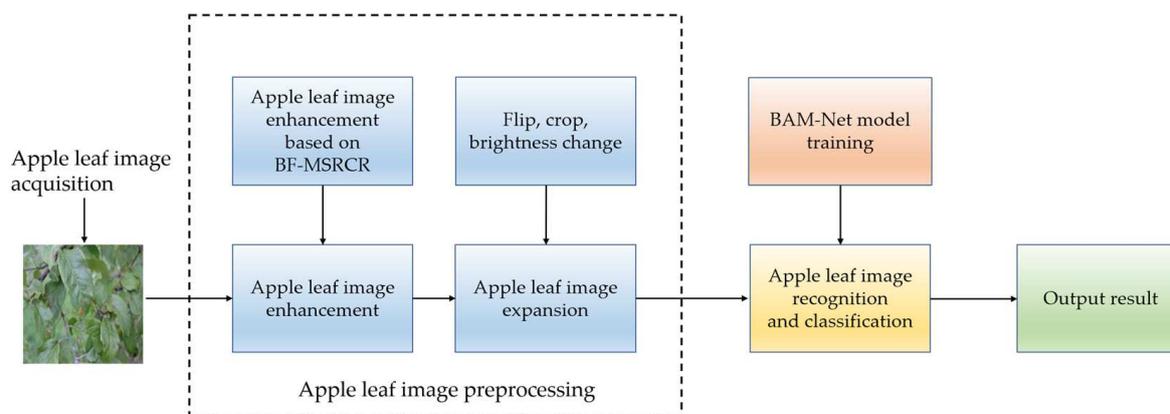


**Figure 1.** The overall structure of the method proposed in this paper.

## 2. Materials and Methods

### 2.1. Data Acquisition

In this study, the apple leaf disease dataset is the basis for studying the classification of apple leaf diseases. We selected 6 types of apple leaves for research, and the characteristics of different apple leaves are shown in Table 1 ("Before" and "After", respectively, represent before and after the dataset expansion. The specific concept of data expansion is shown in Section 2.2). The dataset employed in this study of apple leaves comprises 2 parts: the first part was collected in the southern apple base of Xiangtan City, Hunan Province. The researchers of this paper used a Sony A6300 camera with a resolution of 2420 × 3680 to take outdoor photos, totaling 2429 images of healthy and diseased apple leaves at different periods under natural light conditions. The second part was collected from the Kaggle open dataset Appleleaf9 (https://www.kaggle.com/datasets/jasonyangcode/appleleaf9 (accessed on 10 January 2023)) [23]. We compiled the collected apple leaf images, then reclassified the

apple leaf images in the dataset by consulting experts and scholars and by reviewing the literature; in addition, we selected images with relatively complex backgrounds. Finally, we obtained 3500 apple leaf images with complex backgrounds.

**Table 1.** Types and numbers of complex apple leaf disease datasets.

| Category | Example | Characteristics | Number (Before/After) | Proportion (Before/After) |
|---|---|---|---|---|
| Healthy | | The leaf color is green, the texture is obvious, no disease spots. | 530/2650 | 15.1%/14.9% |
| Brown spot | | Brown, irregular spots on leaves progress to yellow and eventually become brown as the disease worsens. | 568/2840 | 16.2%/16.0% |
| Alternaria leaf spot | | Dark brown spots are present on the leaves. As the disease progresses, the color of the spots will change to black. | 588/2934 | 16.8%/16.5% |
| Mosaic | | Large mosaic-like spots appear on the leaves and yellowish irregular spots are present on the veins of the leaves. | 564/2826 | 16.1%/15.9% |
| Powdery mildew | | Large amounts of white mycelium are present on the leaves. As the disease progresses, the leaves will become distorted. | 608/3025 | 17.3%/17.1% |
| Rust | | Bright spots appear on the leaves. As the disease progresses, the spots will gradually enlarge and turn orange or red. | 632/3148 | 18.0%/17.8% |

*2.2. Data Expansion*

The training of convolutional neural networks requires the support of a large number of samples. We used the OpenCV package in Python to expand the original dataset to help the neural network model train better and to avoid issues such as overfitting. Figure 2 depicts the dataset expansion methods used in this paper, which include vertical flip, mirror flip, random crop, a brightness increase by 30%, and a brightness decrease by 30%. Table 1 shows the total number of expanded images (17,423 images).
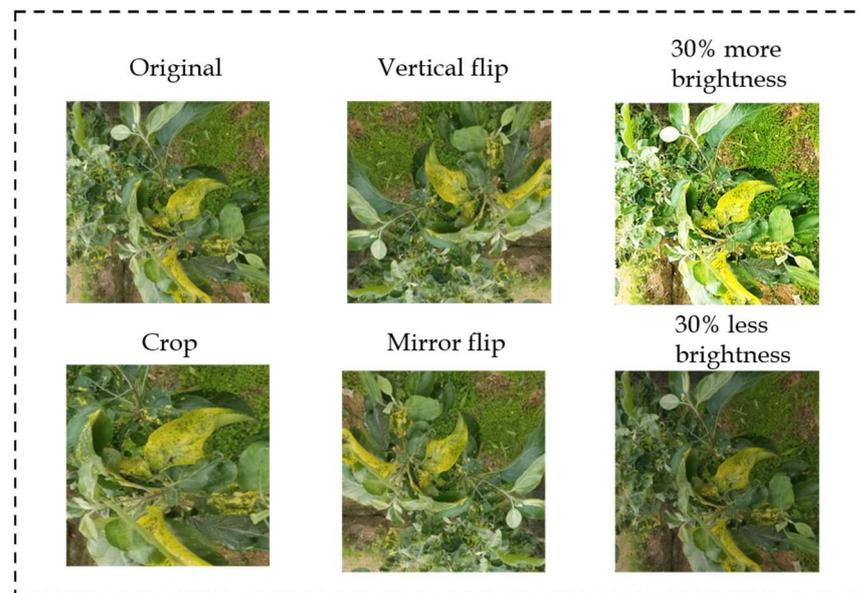
**Figure 2.** Data expansion methods.

### 2.3. BF-MSRCR

The process of acquisition, transmission, and processing is affected by the equipment and the shooting environment. Some of the apple leaf images in this dataset were of low resolution, and the color and texture characteristics of the leaves were obscured. This will have an impact on the subsequent neural network's extraction of leaf features. Therefore, it is necessary to pre-process the images for image enhancement before they are input into the neural network.

In previous studies, the proposed algorithm based on Retinex theory [26] has been widely used for the image enhancement of low-quality images. MSRCR [27] is an image enhancement algorithm derived from the Retinex algorithm; it can improve the overall color contrast of the image and alleviate the blurriness caused by uneven illumination. However, using MSRCR for image enhancement can lead to a certain degree of blurring of the texture edges in the image. When classifying leaf diseases, the texture edges of the leaves is an important factor that affects disease diagnosis. Therefore, it was necessary to pre-process the images for image enhancement before they were input into the neural network. MSRCR is a color restoration image enhancement algorithm extended from the Retinex algorithm, and it can improve the overall contrast in the color of an image and alleviate the unclear image caused by uneven illumination. However, using MSRCR for image enhancement may result in some blurriness along the texture edges. Therefore, this paper adopts a bilateral filter-based MSRCR algorithm (BF-MSRCR) [28] to address this issue. The algorithm uses a bilateral filter function as the center-surround function of the MSRCR algorithm to improve image clarity, highlight disease features, and retain as much of the edge texture information of the leaf images as possible.

BF-MSRCR consists of 2 parts: the multi-scale Retinex algorithm (MSR) and color recovery. We first used MSR for image clarification. Its expression is shown in Equation (1).

$$MSR_i(x,y) = \sum_{k=1}^{N} w_k \{\log[(I_i(x,y))] - \log[H_k(x,y)I_i(x,y)]\} \tag{1}$$

where $i$ represents the 3 color channels R, G, and B; $I_i(x,y)$ represents the original image corresponding to different color channels; $N$ represents the number of scales, which is empirically set to 3 and represents the large, medium, and small scales; $w_k$ represents the weighting factor of the $N$th scale; and $H_k(x,y)$ is the center-surround function of the Retinex

algorithm. In the MSR algorithm, the center-surround function is generally a low-pass Gaussian filter function. Its expression is shown in Equation (2).

$$F_k(x,y) = \frac{1}{2\pi\sigma^2} e^{\left(-\frac{(x-x_0)^2+(y-y_0)^2}{2\sigma^2}\right)} \tag{2}$$

where $(x_0, y_0)$ is the centroid of the Gaussian filter window. According to Equation (2), it is known that the closer the coordinates are to the centroid in the filtering window, the greater the weight obtained. However, when the pixel is in the edge region, its distance from the center point is farther and the pixel similarity is lower, which is easily lost in the filtering process. To reduce the loss of image edge texture information, the bilateral filter function is used as the center-surround function of MSRCR in this paper. The bilateral filter function is a nonlinear filter function that can consider the difference of neighboring pixel values for edge protection during image smoothing. Its expression is shown in Equation (3):

$$H_k(x,y) = \exp\left(-\frac{(x-x_0)^2+(y-y_0)^2}{2\sigma_s{}^2} - \frac{\|f(x,y)-f(x_0,y_0)\|^2}{2\sigma_r{}^2}\right) \tag{3}$$

where $f(x,y)$ is the image centroid coordinate; $f(x_0, y_0)$ is the image centroid gray value; $\sigma_s$ is the standard deviation of the null domain; and $\sigma_r$ is the standard deviation of the value domain.

Although BF-MSR can improve the image sharpness and preserve the edge texture features of the image well, it still causes local color distortion. Therefore, a color recovery factor needs to be added to perform color recovery processing. The specific expression of BF-MSRCR is shown in Equations (4) and (5).

$$BF - MSRCR_i(x,y) = C_i(x,y)BF - MSR_i(x,y) \tag{4}$$

$$C_i(x,y) = \beta \log[\alpha I_i'(x,y)] \tag{5}$$

where $C_i$ denotes the color recovery factor of the $i$th channel, which is used to adjust the color ratio of the 3 channels; $\beta$ denotes the gain constant; and $\alpha$ denotes the controlled nonlinear intensity. By adjusting the color balance ratio of the 3 original image color channels adjusted by the color restoration factor, it is possible to effectively enhance the information in the dark regions and to solve the problem of color distortion in BF-MSR.

Figure 3 displays the apple leaf images processed by different algorithms. MSR enhancement can improve image clarity, but it may cause color distortion. Although MSRCR solves the problem of color distortion, it loses the edge features of the image. In contrast, the BF-MSRCR algorithm effectively addresses these issues by enhancing image clarity and highlighting disease features while preserving the edge features of the leaf.

### 2.4. BAM-Net

Firstly, in the context of practical applications in orchards, the identification of apple leaf diseases is easily disturbed by other factors in the image, such as branches, soil, and other complex backgrounds, resulting in the network not being able to correctly extract the disease characteristics of the leaves. Secondly, apple leaf diseases have the characteristics of inter-class mutation and inter-class similarity. As the degree of disease deepens, leaves with the same disease may show different texture and color characteristics. Leaves with different diseases may also have similar disease characteristics. This makes it difficult to correctly classify apple leaf diseases. To address the aforementioned problems, this paper proposes BAM-Net for apple leaf disease classification in complex backgrounds using ConvNext-T [29] as the backbone network.
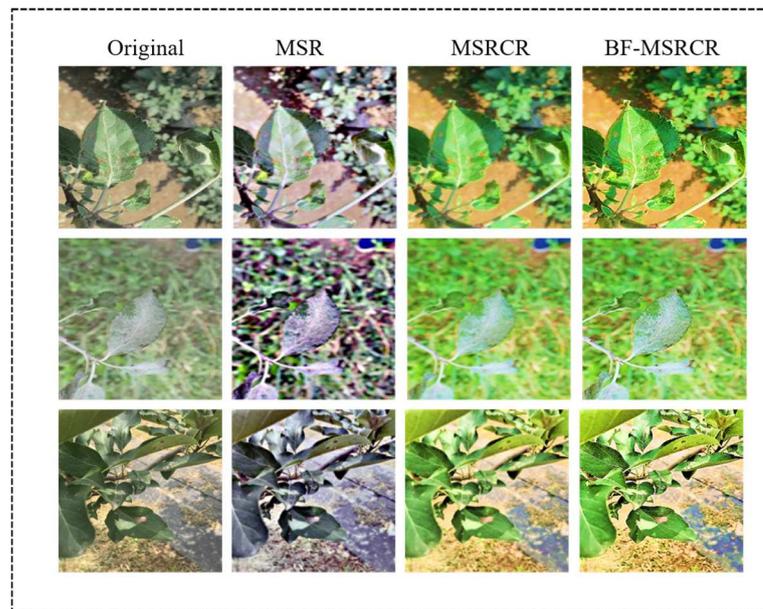
**Figure 3.** Enhancement of the original image using MSR, MSRCR, and BF-MSRCR, respectively.

The BAM-Net structure, as shown in Figure 4, can be divided into 3 parts:

(1) The first part is the feature extraction network of BAM-Net, which mainly consists of ConvNext-Stage and ACAM. ConvNext-Stage was used for the basic feature extraction of apple leaf images after BF-MSRCR processing. ACAM was used after each stage to help the network focus on the important feature information and to filter out the interference information.

(2) The second part is the feature refinement module, comprising several $1 \times 1$ convolutions and three $3 \times 3$ convolutions with varying expansion rates. MFRM divides the output of the feature from the first part into 4 branches. Then, feature extraction at different scales and channel information interaction operations are performed in the 4 branches to refine the leaf disease features.

(3) The third part is the classification output module, which includes global average pooling, layer normalization, and linear layers. Firstly, the network's extracted features are subjected to global pooling and normalization operations. Then, the fully connected layer and Softmax function transforms the output into a probability distribution, providing the classification results for apple leaf disease images.

This article provides detailed descriptions of the ConvNext-T backbone, ACAM, and MFRM in the subsequent content.

### 2.4.1. ConvNext-T Backbone

Convolutional neural networks have been used as the dominant model in computer vision for the past decade [30]. However, in recent years, transformer-like networks, such as Swin Transformer [31] (Swin-T), have achieved better performance than pure convolutional neural networks on certain tasks and have rapidly become a research hotspot. The complexity of the internal structure of transformer networks makes them inapplicable for tasks that require practical applications, such as apple leaf disease classification. To demonstrate that pure convolutional neural networks still have great room for improvement, Li et al. combined some of the design ideas of the transformer model and proposed the ConvNext model from ResNet [32].
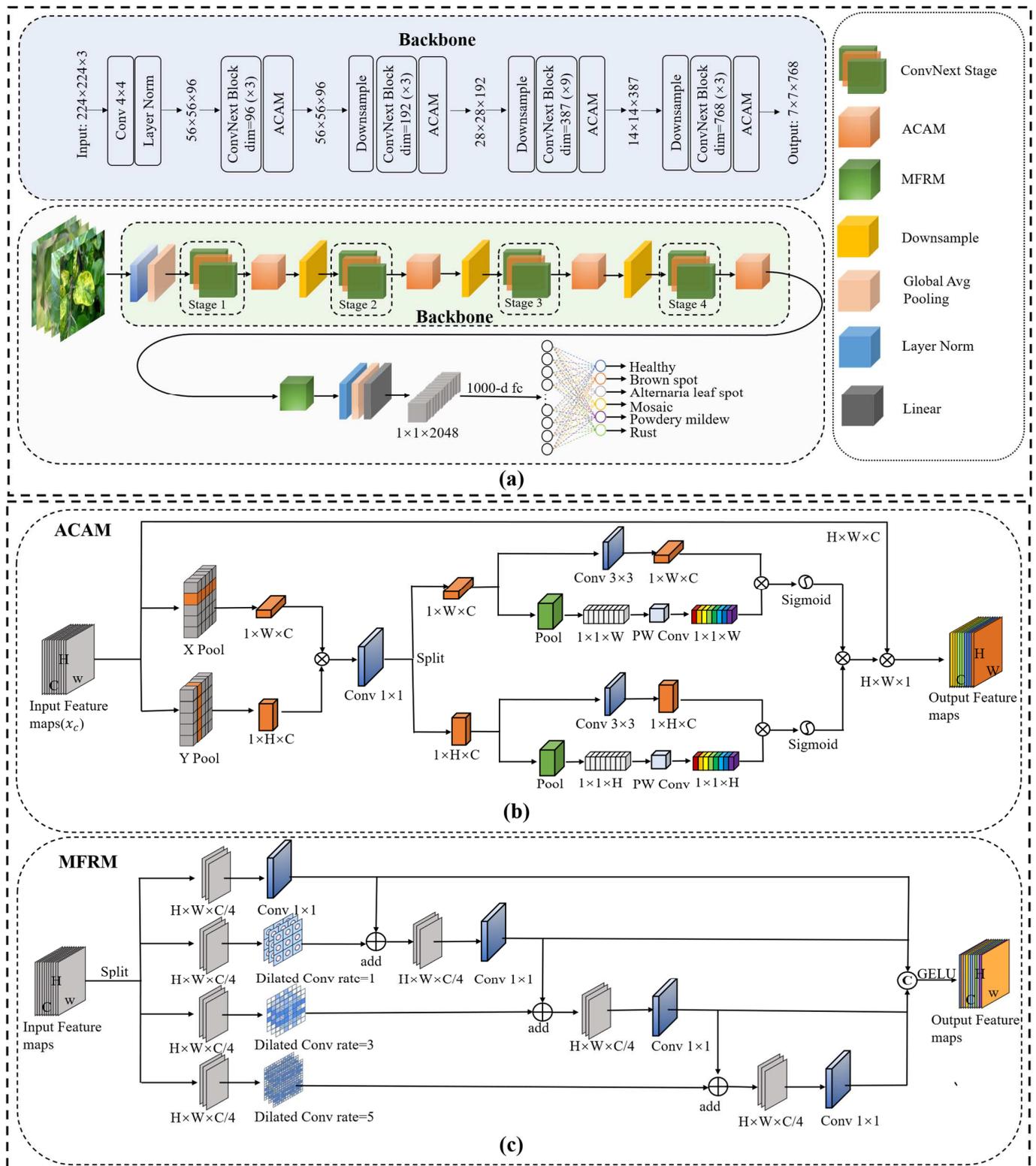
**Figure 4.** General network structure diagram. (**a**) BAM-Net structure diagram; (**b**) ACAM structure diagram; and (**c**) MFRM structure diagram.

The changes in ConvNext are as follows. (1) Adjust the model scale, i.e., the number of block stacking, to the same scale as Swin-T. (2) Patchify Stem is changed to a 4 × 4 convolution with a step size of 4. (3) Depthwise convolution is adopted, aiming to control the number of channels of the output feature matrix by the number of

convolution kernels. (4) The feature extraction block of ConvNext adopts an inverted bottleneck structure, which can reduce information loss during information extraction. Compared with ResNet and Swin-T, ConvNext has higher accuracy and faster computation speed.

The task of apple leaf disease classification requires high real-time performance. In addition, the irregular size and shape of leaf diseases make it difficult to accurately identify them based on size and shape features alone. The differences between leaf diseases and their surroundings are an important factor in achieving accurate disease identification. The $7 \times 7$ large convolution used by ConvNext can effectively capture the relationship between the extracted features and their contexts. Therefore, we chose ConvNext as the backbone network of BAM-Net. Training large-scale models often requires more computing resources. Given the limited computing resources available to our team in this research, which is a server equipped with an NVIDIA GeForce RTX 3090, we chose the ConvNext-T with the lowest number of parameters to explore the possibility of using deep learning methods to improve the efficiency of apple leaf disease classification.

### 2.4.2. ACAM

During the actual recognition process, the images of apple leaf diseases captured in real time do not simply contain the leaves to be inspected. Other irrelevant factors such as tree branches, fruits, and soil in the images will affect the extraction of the features of apple leaf diseases. Therefore, it is important to enhance the network's ability to focus on disease features and to suppress irrelevant interference factors. This will help the network better learn the disease features and improve the accuracy of apple leaf disease classification.

The attention mechanism is similar to human visual features [33], emphasizing the degree of importance for important areas through weight size and suppressing interference backgrounds. Traditional attention mechanisms usually only focus on single-dimensional feature weights [34], which may overlook important feature information. However, coordinate attention (CA) [35] is a new lightweight attention mechanism that can effectively capture long-distance dependency relationships by aggregating features along 2 spatial dimensions and retaining accurate positional information. Therefore, this paper proposes ACAM based on CA, combined with apple leaf disease features.

The ACAM structure is shown in Figure 4b and consists of 3 parts:

(1)　Bi-directional pooling

Traditional global average pooling compresses spatial information into channels, ignoring direction-related positional information. Therefore, we performed pooling on the input feature map $x_c$ in the X and Y directions separately to capture the precise positional information. After pooling, feature maps $z^w$ and $z^h$ with sizes of $1 \times W \times C$ and $1 \times H \times C$ were generated, respectively. Then, the generated feature maps were multiplied and fed into a $1 \times 1$ convolutional kernel to obtain feature map $x$, thus achieving information interaction between channels.

$$z^w(w) = \frac{1}{H} \sum_{0 \le j \le H} x_c(j, w) \tag{6}$$

$$z^h(h) = \frac{1}{W} \sum_{0 \le i \le W} x_c(h, i) \tag{7}$$

$$x = z^w \times z^h \tag{8}$$

(2)　Aggregate feature correction

After a $1 \times 1$ convolution, the feature map x is divided into 2 branches: the upper branch is the feature map $z_1^w$ ($1 \times W \times C$) and the lower branch is the feature map $z_1^h$ ($1 \times H \times C$). Since the 2 branched operations are the same, the following is a detailed description of the aggregate feature correction using the upper branch as an example.

The original feature map of the upper branch $z_1^w$ is pooled to obtain a $1 \times 1 \times$ W scalar. To mitigate the scale variation and to emphasize smaller objects, point-by-point convolution is used as a local up-down aggregator. It uses spatial locations for point-by-point channel interactions and yields a 1-D vector with weights. Subsequently, the feature map $z_1^w$ is multiplied by a $3 \times 3$ convolution with a 1-D vector for the purpose of feature map correction.

$$z_2^w = f_1(z_1^w) \times f_c[p(z_1^w)] \tag{9}$$

$$z_2^h = f_1(z_1^h) \times f_c[p(z_1^h)] \tag{10}$$

where $f_1$ represents $3 \times 3$ convolution; $p$ represents pooling; and $f_c$ represents pointwise convolution. In addition, $z_2^w$ and $z_2^h$ represent the re-calibration maps of the 2 branches.

(3)  Feature fusion output

The aggregation-corrected bi-directional spatial feature information $z_2^w$ and $z_2^h$ are multiplied by the activation function to fuse the feature coordinate information in both spatial directions. Then, the output feature map $x_c$ is obtained by multiplying the residual structure with the original feature map $x_c'$. The output feature map pays more attention to apple diseases and less attention to other noises. Therefore, it can attenuate the interference of irrelevant factors in apple leaf images.

$$x_c' = [\delta(z_2^w) \times \delta(z_2^h)] \times x_c \tag{11}$$

where $\delta$ represents the activation function sigmoid.

2.4.3. MFRM

As apple leaf diseases increase in severity, the same disease may take on different characteristics. For example, a brown spot is usually a small yellow-brown spot initially, but over time the spot will take on a dark, black color. In addition, different diseases may also have very similar characteristics. For example, the initial spots of the Alternaria leaf spot and rust are both brownish and blotchy. The intraclass variability and interclass similarity of apple leaf diseases increases the difficulty of classification. In order to fully exploit the differences and similarities between deep disease features, this paper proposes MFRM, which extracts features under multiple sensory fields and interacts with the information via the features of adjacent channels. The structure of MFRM is shown in Figure 4c.

First, we divided the input feature map into 4 branches with the same number of channels, and each branch used a different operation. The first and second branches were convolved using a $1 \times 1$ convolution and $3 \times 3$ convolution, respectively, aiming to obtain the salient features of the apple leaf diseases. To reflect the disease features more comprehensively, we performed operations in the third and fourth branches using dilated convolution [36] with a void ratio of r = 3 and r = 5, respectively, to obtain the feature information at different scales.

Then, to enrich the disease feature information, we performed the add operation on the feature maps between adjacent branches. The fused images passed through a $1 \times 1$ convolution. This was performed with the aim of achieving information interaction between the channels and to enhance the expression of the feature maps.

$$S_2' = f_1[add(S_1, S_2)] \tag{12}$$

$$S_3' = f_1[add(S_2', S_3)] \tag{13}$$

$$S_4' = f_1[add(S_3', S_4)] \tag{14}$$

where $f_1$ represents the $1 \times 1$ convolution kernel, and $S_1$, $S_2$, $S_3$, and $S_4$ represent the feature maps obtained after the first convolution operation for each of the 4 branches.

Finally, we used the concat method on the feature maps of these 4 branches and passed them through the activation function GELU. The use of the activation function GELU can effectively reduce the phenomenon of gradient explosion and gradient disappearance, thus helping the network to obtain stronger generalization ability.

$$S = \sigma[concat(S_1, S_2', S_3', S_4')] \tag{15}$$

where $\sigma$ represents the activation function GELU and $S$ represents the output feature map.

## 3. Experimental Results Analysis

In this section, we demonstrate through experiments the superiority of the method proposed in this paper for apple leaf disease recognition and classification tasks in complex backgrounds. Specifically, the section is divided into the following parts: (1) in Section 3.1, we present the experimental settings and parameters, including hardware and software environments, training methods, and parameter settings; (2) in Section 3.2, we show the performance evaluation indicators used in this paper; (3) in Section 3.3, we compare BAM-Net with other networks; (4) in Section 3.4, we assess the efficacy of each module in the proposed methodology; (5) in Section 3.5, we explore the influence of various modules on the performance of BAM-Net through ablation experiments; (6) in Section 3.6, we compare BAM-Net with baseline and state-of-the-art networks; (7) and in Section 3.7, we discuss the performance of BAM-Net on public datasets, verifying its generalization ability.

### 3.1. Experimental Environment and Parameter Setting

To ensure the fairness and validity of the experiments, all experiments in this paper were conducted in the same hardware and software environment. The hardware and software environments used in this paper are shown in Table 2. To reduce the computational burden, we set the input size of the images to be uniformly $224 \times 223 \times 3$. The original dataset was obtained after the data expansion operation in Section 2.2, and a total of 17,688 apple leaf images were obtained.

**Table 2.** Hardware and software environment.

| | | |
|---|---|---|
| Hardware Environment | CPU | Intel(R) Xeon(R) Platinum 8352M |
| | ARM | 80 GB |
| | Video Memory | 50 GB |
| | GPU | NVIDIA GeForce RTX 3090 |
| Software Environment | OS | Windows 11 |
| | PyTorch | 1.11.0 |
| | Python | 3.8 |
| | Cuda | 11.3 |
| | MATLAB | R2019a |

In this paper, we undertook a five-fold cross-validation for the training test and divided the images into training and test sets according to the ratio of 4:1. Moreover, we conducted five repetitions of the experiments. We chose the average of the results of these five experiments as the final experimental results. This strategy effectively avoids the chance of unreliable experimental results.

To improve the training effect of the model, AdamW was used as the training optimizer in this paper; by considering the actual performance of the hardware device, we set the initial learning rate to $1 \times 10^{-3}$, the batch size to 64, the epoch to 50, and the momentum parameter to $5 \times 10^{-4}$.

### 3.2. Evaluation Indicators

To assess the efficacy of the proposed model, this paper employed accuracy, precision, recall, and F1-score as the evaluation metrics for the purpose of classifying apple leaf diseases. Accuracy denotes the ratio of accurately classified samples to the total number of samples. Precision refers to the ratio of accurately classified positive samples to predicted positive samples. Recall represents the ratio of accurately classified positive samples to the actual positive samples. F1-score is the weighted average of precision and recall. The specific expression is shown in Equations (16)–(19):

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TN} + \text{FN} + \text{TP} + \text{FP}} \tag{16}$$

$$\text{Precision} = \frac{\text{TP}}{\text{FP} + \text{TP}} \tag{17}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{18}$$

$$\text{F1} - \text{Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{recall}} \tag{19}$$

Assume that the healthy apple leaves are positive samples. Then, TP represents the number of leaves predicted to be healthy and whether the actual situation is healthy; FP represents the number of leaves predicted to be healthy and whether the actual result is non-healthy; FN is the number of leaves predicted to be healthy and whether the actual result is non-healthy; and TN is the number of leaves predicted to be non-healthy and whether the actual result is non-healthy.

To evaluate the effectiveness of the image enhancement algorithm, this paper uses average gradient (AG), information entropy (IE), and standard deviation (Std) to evaluate the quality of the image after enhancement. The corresponding calculation formulas are as follows:

$$\text{AG} = \frac{1}{M \times N} \times \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} \sqrt{\frac{(F(i,j) - F(i+1,j))^2 + (F(i,j) - F(i,j+1))^2}{2}} \tag{20}$$

$$\text{IE} = -\sum_{m=0}^{l-1} p(m) \log_2 p(m) \tag{21}$$

$$Std = \sqrt{\frac{1}{M \times N} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} (F(i,j) - u)^2} \tag{22}$$

AG is the average gradient of the image, which is used to reflect the sharpness of the image, and Std is the standard deviation of the image, which is used to measure the contrast level of the image. Moreover, $M$ and $N$ are the length and width of the image, $F(i,j)$ is the pixel value of the image point, and $u$ is the pixel mean value; IE is the information entropy of the image, which is used to reflect the amount of information contained in the image; and $p(m)$ is the distribution density of the image gray level $m$, where $l$ is the image gray level and Std is the standard deviation of the image, which is used to measure the contrast level of the image.

### 3.3. Comparison with Classical Networks

In this section, we compare BAM-Net with certain conventional classification models: ConvNext-T [29], VGG-16 [37], ResNet-50 [32], ResNest-50 [38], and Densenet-121 [39]. First, we trained and tested all models in the same environment (the loss vs. accuracy curves are shown in Figure 5). It can be found that BAM-Net starts to converge at approximately

20 rounds, and the convergence speed is only lower than VGG-16 and Densenet-121; however, the average accuracy is significantly higher than the other classification models in this experiment.



(a) Loss curve
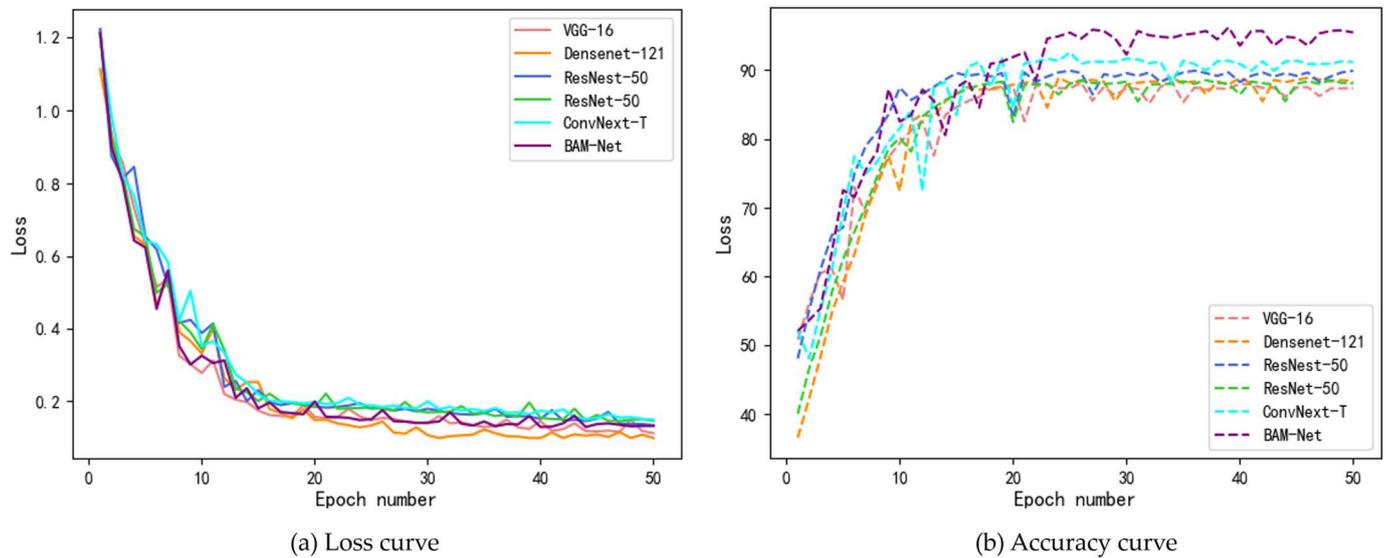


(b) Accuracy curve

**Figure 5.** Training loss and accuracy curves for different networks.

Table 3 shows the classification accuracies of these six classification models for the different categories of apple leaves. The BAM-Net proposed in this paper has a high accuracy in classifying apple leaves of all categories, including 96.84% and 95.61% for the healthy and brown spot groups, respectively. However, its accuracy in detecting Alternaria leaf spot is slightly lower than that of ConvNext-T. This is because the sample characteristics of Alternaria leaf spot samples are very similar to the sample characteristics of rust, which causes ConvNext-T to misclassify a large number of rust images as Alternaria leaf spot. This results in a false high precision for Alternaria leaf spot and a low precision for rust. On the other hand, due to the addition of ACAM and MFRM, the number of parameters of the BAM-Net model is slightly higher than that of ConvNext-T. This also leads to a higher training time for BAM-Net than ConvNext-T (+9 min 30 s).

**Table 3.** Comparison of classification accuracy across different networks.

| Network | Healthy (%) | Brown Spot (%) | Alternaria Leaf Spot (%) | Mosaic (%) | Powdery Mildew (%) | Rust (%) | Training Time |
|---|---|---|---|---|---|---|---|
| VGG-16 | 89.1 | 86.85 | 89.72 | 86.91 | 86.54 | 84.27 | 2 h 48 min 6 s |
| ResNet-50 | 92.67 | 84.19 | 93.46 | 85.91 | 89.05 | 86.34 | 2 h 45 min 28 s |
| Densenet-121 | 91.76 | 86.23 | 90.97 | 91.99 | 89.80 | 88.57 | 2 h 36 min 35 s |
| ResNest-50 | 89.01 | 90.83 | 91.93 | 88.75 | 90.71 | 85.83 | 3 h 4 min 55 s |
| ConvNext-T | 91.03 | 89.57 | 95.87 | 90.18 | 92.67 | 88.04 | 3 h 7 min 13 s |
| BAM-Net | 96.84 | 95.61 | 95.59 | 95.97 | 95.58 | 93.36 | 3 h 18 min 43 s |

In order to verify the ability of BAM-Net to recognize interspecific similarities and intraspecific variations in apple leaf diseases, we compared the disease recognition capabilities of ConvNext-T and BAM-Net on these images and the results were compared. As shown in Figure 6, ConvNext-T made recognition errors in some images, such as misidentifying rust as Alternaria leaf spot and mosaic as brown spot. This is because in the late stage of rust, the color of the lesion gradually becomes brown, which is similar to that of Alternaria leaf spot, while in the late stage of mosaic, the surface of the leaf is almost occupied by the yellow color that is similar to brown spot symptoms. However, BAM-Net

still performed extremely well in these special cases and accurately identified these leaf diseases. This indicates that BAM-Net has a high capability for identifying leaf diseases.
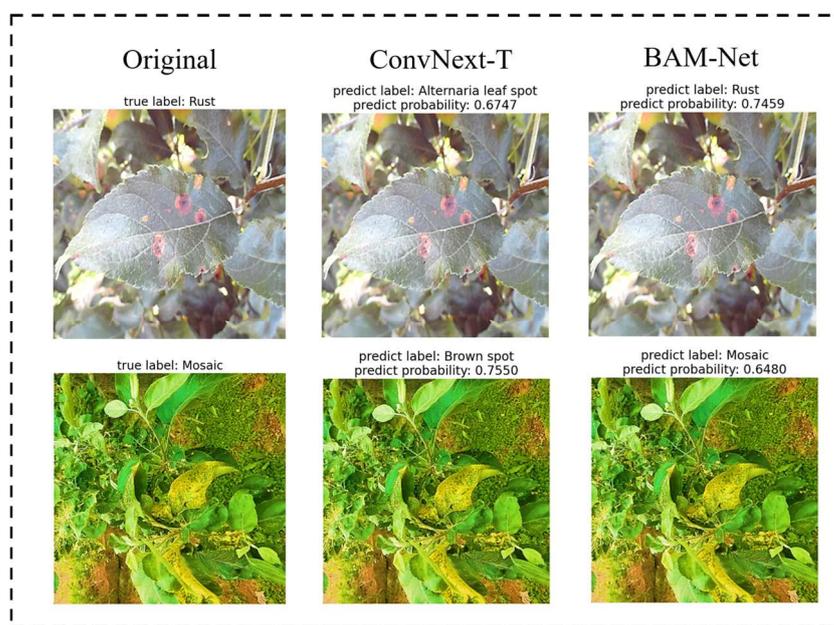


**Figure 6.** Identification results for the presence of interclass mutations as well as interclass similar leaf disease images.

### 3.4. Modules Effectiveness Analysis

In Section 3.3, we demonstrated that BAM-Net outperforms other conventional networks in the task of identifying and classifying apple leaf diseases in complex backgrounds. To explore the reasons for this result and to further validate the effectiveness of the proposed method, we conducted a module effectiveness analysis.

#### 3.4.1. Effectiveness of Image Pre-Processing

In Section 2.3, we visually demonstrated the effects of using MSR, MSRCR, and BF-MSRCR for image enhancement. Subjectively, BF-MSRCR can effectively highlight the color and texture features of leaves in the images. In this section, to evaluate the image-enhancement effects of different algorithms more objectively, we conducted comparative experiments, the results of which are shown in Table 4. Image quality evaluation after being enhanced by different algorithms). The images enhanced by BF-MSRCR have higher average gradients, information entropy, and standard deviation, indicating that BF-MSRCR also has better performance in improving image clarity and enhancing image details.

Next, we applied the same augmentation techniques, such as flipping, mirroring, cropping, and random brightness adjustments, to the original dataset, the MSRCR-enhanced dataset, and the BF-MSRCR-enhanced dataset in order to obtain the augmented versions of each. Then, we trained and tested ConvNext-T on these six datasets separately. Table 5 displays the obtained results. The accuracy of the ConvNext-T model was significantly improved after dataset expansion, which was due to the small number of samples in the original dataset. Meanwhile, the expanded dataset provided more learnable features for the model, which could better simulate apple leaf disease classification in the field environment and improved the robustness of the model. By comparing the original dataset with the image-enhanced dataset, it can be observed that the accuracy of the model was also improved after image enhancement. This is due to the fact that there are a certain number of blurred images in the dataset due to the influence of the shooting equipment and environment when collecting data and because the image enhancement can make the color features of leaves more obvious, as well as the fact that the model can extract

more accurate leaf disease features. In addition, the BF-MSRCR-processed dataset had a higher accuracy than the MSRCR-processed dataset because certain texture features may be lost when using MSRCR for image enhancement, while BF-MSRCR successfully solved this problem.

**Table 4.** Image quality evaluation after being enhanced by different algorithms.

| Methods | Image | Average Gradient | Information Entropy | Standard Deviation |
|---------|-------|------------------|---------------------|--------------------|
| Original |  | 8.73 | 7.62 | 49.62 |
| MSR |  | 4.80 | 7.81 | 61.04 |
| MSRCR |  | 8.61 | 7.81 | 61.55 |
| BF-MSRCR |  | 10.85 | 7.90 | 68.84 |

**Table 5.** The performance of the network before and after image pre-processing.

| Method | Accuracy (%) | Precision (%) | Recall (%) | F1-Score (%) |
|--------|--------------|---------------|------------|--------------|
| Original | 91.24 | 91.19 | 91.17 | 91.17 |
| MSRCR | 91.61 | 91.72 | 91.26 | 91.48 |
| BF-MSRCR | 91.94 | 91.84 | 91.85 | 91.84 |
| Original (Extended) | 91.82 | 91.05 | 91.56 | 91.30 |
| MSRCR (Extended) | 92.03 | 92.24 | 92.47 | 92.35 |
| BF-MSRCR (Extended) | 92.18 | 92.62 | 92.56 | 92.58 |

3.4.2. Effectiveness of ACAM

In this section, we introduced SE [40], CBAM [41], and CA [35] to the backbone network for comparative tests in order to investigate the classification effectiveness of BAM-Net for apple leaf diseases on the effects of different attention mechanisms. The test results are shown in Table 6. The results indicate that CBAM and SE did not perform well in the task of apple leaf disease classification. This is because they lack spatial interaction with positional information, which makes them less precise in locating disease features and irrelevant information. Although ACAM brings more parameters, it has a more outstanding performance. Compared with models without an attention mechanism, the F1-score increased by 1.03%, and accuracy increased by 2.67%. This is because the ACAM module can emphasize smaller objects, achieve channel interaction, and effectively guide the network to increase the attention to disease feature areas.

**Table 6.** The impact of different attention networks on BAM-Net.

| Method | Accuracy (%) | F1-Score (%) | Param |
|---|---|---|---|
| Without attention | 92.18 | 92.56 | 27.80 M |
| +SE | 92.64 | 92.56 | 27.90 M |
| +CBAM | 92.71 | 92.87 | 27.90 M |
| +CA | 93.52 | 93.38 | 27.88 M |
| +ACAM | 94.33 | 94.41 | 28.14 M |

We selected five apple leaf disease images with complex backgrounds to visualize and validate the effects of different attention mechanisms. According to Figure 7 (Visual validation of different attention mechanisms), adding SE and CBAM can help the network focus on the disease feature region compared to not adding attention mechanisms, but the improvement is limited. Particularly with SE, interference regions similar to the disease feature still have a significant impact on the network's ability to extract disease features. With the addition of CA, the network's attention is mostly focused on the disease area, but some non-disease areas that are extremely similar to the disease area can still be misclassified as disease, affecting disease recognition accuracy. In contrast, the proposed ACAM in this paper can guide the network to focus on the leaf and disease regions of apple leaf disease images under different complex backgrounds, effectively suppressing the interference of complex background information and improving the extraction of disease feature information and recognition accuracy.
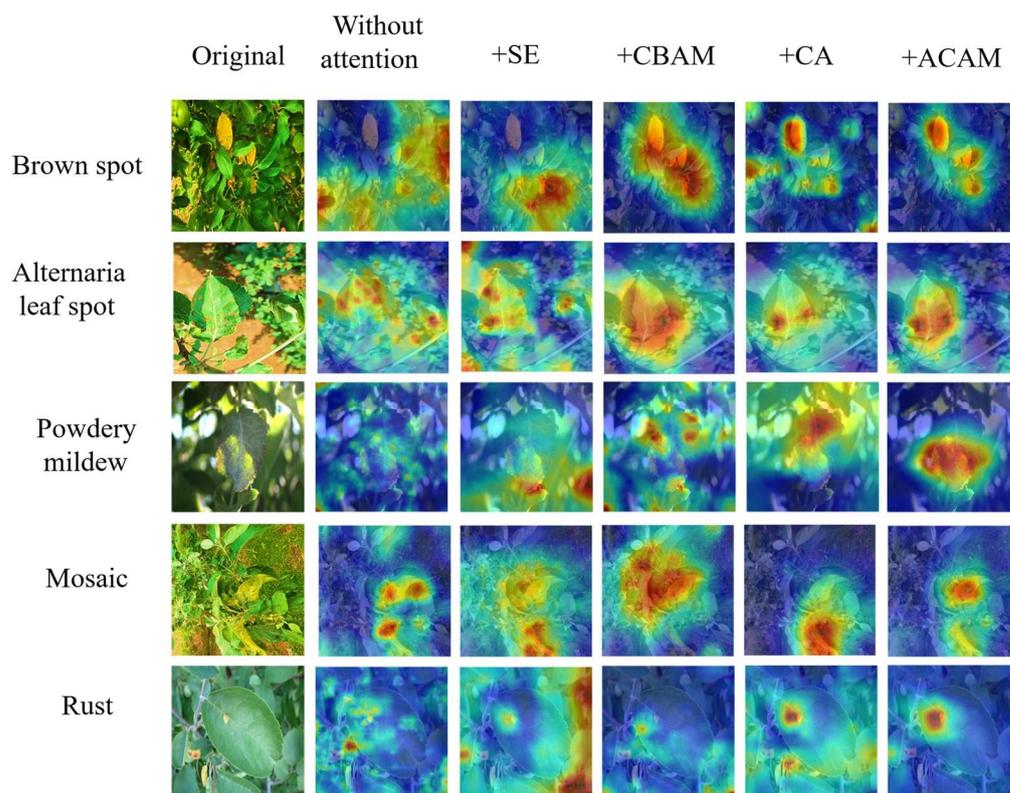


**Figure 7.** Visual validation of different attention mechanisms.

3.4.3. Effectiveness of MRFM

Different combinations of dilation rates can affect the performance of MFRM. To investigate the most suitable dilation rates for apple leaf disease classification, we conducted a comparative experiment. We set the dilation rate combinations of MFRM to A (1:1:1), B (1:2:3), C (1:3:5), D (1:5:8), and E (1:8:15), and then added them to the connection between the backbone network and the classifier. The results are depicted in Figure 8. The effect of

different expansion rates of MFRM on the performance of BAM-Net. The expansion ratios for each group are A (1:1:1), B (1:2:3), C (1:3:5), D (1:5:8), and E (1:8:15), respectively.
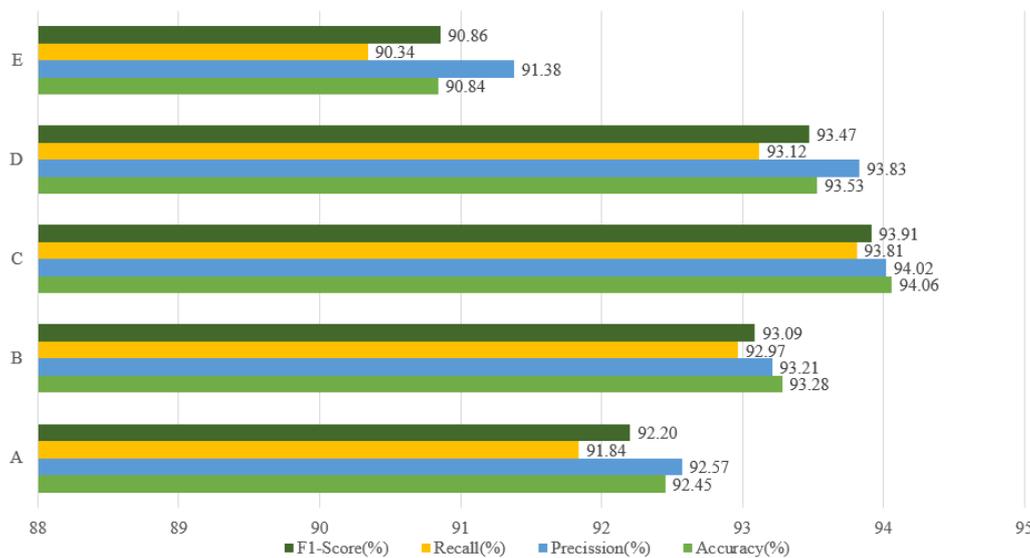


**Figure 8.** The effect of different expansion rates of MFRM on the performance of BAM-Net. The expansion ratios for each group are A (1:1:1), B (1:2:3), C (1:3:5), D (1:5:8), and E (1:8:15), respectively.

It can be seen that the model performs best in all performance indicators when the expansion ratio in MFRM is 1:3:5. This is due to the fact that when the expansion ratio is too large, the network cannot capture the minor disease features. Additionally, when the expansion ratio is too small, the network cannot measure the difference between the disease and its surrounding well. Additionally, when the expansion ratio in MFRM is 1:3:5, the model is able to learn from more detailed apple leaf disease features at multiple scales. The ability of the model to discriminate between similar disease features was improved.

### 3.5. Ablation Experiment

To evaluate the performance of BAM-Net, we conducted ablation experiments on our self-made dataset of apple leaf images with complex backgrounds. We gradually added BF-MSRCR, ACAM, and MFRM to the ConvNext-T backbone network and analyzed and evaluated the performance of each module by comparing the changes in parameter quantity and detection accuracy. The ablation experiment outcomes are illustrated in Table 7.

**Table 7.** Ablation experimental results.

| Group | Method | Accuracy (%) | F1-Score (%) | Param | FPS |
|---|---|---|---|---|---|
| 1 | Convnext-T | 91.24 | 91.17 | 27.80 M | 95.30 |
| 2 | Convnext-T+BF-MSRCR | 92.18 | 92.56 | 27.80 M | 95.30 |
| 3 | Convnex-Tt+MFRM | 93.59 | 93.17 | 28.84 M | 77.31 |
| 4 | Convnext-T+ACAM | 93.91 | 93.41 | 28.14 M | 91.64 |
| 5 | Convnext-T+ACAM+MFRM | 94.79 | 93.79 | 29.18 M | 74.97 |
| 6 | Convnext-T+BF-MSRCR+ACAM | 94.33 | 94.14 | 28.14 M | 91.64 |
| 7 | Convnext-T+BF-MSRCR+MFRM | 94.06 | 93.91 | 28.84 M | 77.31 |
| 8 | BAM-Net | 95.64 | 95.25 | 29.18 M | 74.97 |

Comparing groups 6–8 with groups 3–5, we observed an improvement in the model accuracy after image enhancement when using BF-MSRCR. This is because image enhancement makes the color and texture features of the leaves more prominent, making

it easier for the model to extract features. When comparing Group 2 with Group 6, we found that adding ACAM improved the model's ability to filter interference information during feature extraction, resulting in a significant improvement in accuracy (+2.15%). However, the speed of the model's disease recognition was not greatly affected, as the FPS only dropped by 3.66, going from 95.30 to 91.64. When comparing Group 2 with Group 7, we observed an improvement of 1.88% in model accuracy, which is due to the enhanced discriminative ability of the model to identify leaf diseases when using MFRM, which was achieved through further strengthening the feature information extracted from the backbone network. When comparing Group 2, Group 6, and Group 8, we can see that MFRM is the main reason for the decrease in model recognition speed as its use of a multi-branch structure increases the model's computational consumption to some extent. When comparing Group 1 with Group 8, we observed that the accuracy of the network increased by 4.40%, and the F1-score improved by 4.11% when we employed BF-MSRCR, ACAM, and MFRM, concurrently. Meanwhile, the model's FPS decreased from 95.30 to 74.97. Nevertheless, the drop in FPS is acceptable compared to the increase in leaf disease recognition accuracy. The eight experiments fully demonstrate the effects of BF-MSRCR, ACAM, and MFRM on the model performance.

### 3.6. Comparison with the Latest Network Model

In this section, we compared BAM-Net with other advanced classification models (i.e., Dual-Task Gabor CNN [42], Swin Transformer V2 [43], and FC-SNDPN [44]). The results are shown in Table 8. On the task of apple leaf disease identification and classification in complex backgrounds, all of the performance metrics showed an outperformance of the other methods.

**Table 8.** Comparison of BAM-Net with the latest networks.

| Network | Accuracy (%) | Precision (%) | Recall (%) | F1-Score (%) | Training Time |
|---|---|---|---|---|---|
| Dual-Task Gabor CNN | 93.23 | 93.14 | 92.93 | 93.03 | 2 h 26 min 26 s |
| Swin Transformer V2 | 94.77 | 93.84 | 94.63 | 93.72 | 3 h 46 min 48 s |
| FC-SNDPN | 94.28 | 94.79 | 94.57 | 94.67 | 3 h 24 min 25 s |
| BAM-Net | 95.64 | 95.62 | 95.89 | 95.25 | 3 h 18 min 43 s |

To visualize the classification network's ability to identify apple leaf diseases, we compared the confusion matrices of BAM-Net and the three other networks. In the confusion matrix, the numbers in the diagonal cells indicate the number of correctly predicted samples, and the numbers in the non-diagonal cells indicate the number of incorrectly predicted samples. As is shown in Figure 9, among the 3483 apple leaf images (test dataset), Dual-Task Gabor CNN correctly identified and classified 3247; Swin-Transformer V2 correctly identified and classified 3301; FC-SNDPN correctly identified and classified 3284; while BAM-Net correctly identified and classified 3331. Therefore, the BAM-Net model proposed in this paper had the highest apple leaf recognition and classification efficiency when compared with the latest networks.

### 3.7. Generalizability Experiment

The BAM-Net proposed in this paper was mainly used for apple leaf disease classification in complex backgrounds. Additionally, in this paper, we achieved outstanding performance on the self-made apple leaves in complex backgrounds dataset. In this section, to verify the generalization ability of BAM-Net, we re-trained and tested BAM-Net on the public PlantVillage dataset [45] for three types of plants: apple leaves, corn leaves, and grape leaves. A total of 3171 apple leaves were divided into four categories; a total of 3852 corn leaves were divided into four categories; and a total of 4062 grape leaves were divided into four categories. The test results are shown in Table 9. When we utilized BAM-Net to classify apple leaf diseases on the PlantVillage dataset, the accuracy that was attained was 99.41%. This result was much higher than the result on the homemade dataset

of this paper (+3.77%). This is because the backgrounds of the leaf images in the PlantVillage dataset are all very simple, and there is no interference from background information that is not related to the leaves, such as branches, fruits, and soil, thus allowing BAM-Net to extract disease features more accurately. BAM-Net also achieved good results when used on maize and grape leaves, with classification accuracies of 98.19% and 98.52%, respectively. This is due to the existence of certain similarities among the leaves of different plants, which means that leaf disease features are mostly reflected in color, texture, and shape. This proves that BAM-Net has a strong generalization ability, which is not only applicable in the identification and classification of apple leaf diseases in complex backgrounds but can also be extended to other plants.



(a) Dual-Task Gabor CNN

(b) Swin-Transformer V2
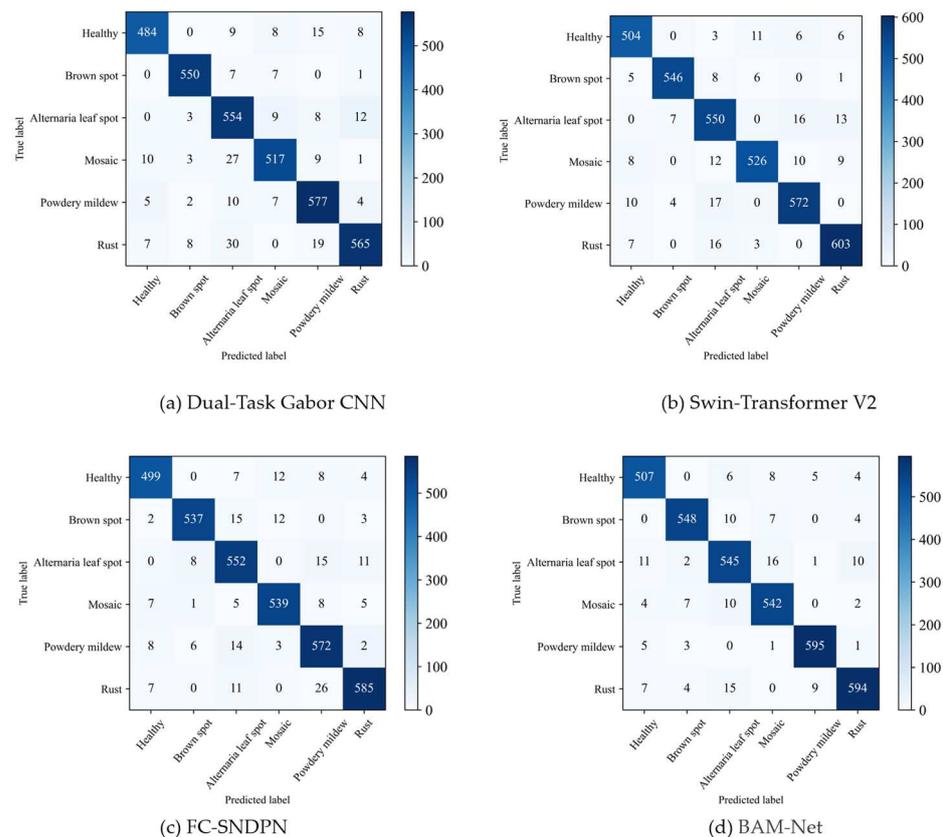
(c) FC-SNDPN

(d) BAM-Net

**Figure 9.** BAM-Net and the latest networks' confusion matrix for apple leaf disease classification (**a**) Dual-Task Gabor CNN; (**b**) Swin-Transformer V2; (**c**) FC-SNDPN; and (**d**) BAM-Net.

**Table 9.** Results of generalization experiments.

| Plant | Categories | Precision (%) | Recall (%) | F1-Score (%) | Accuracy (%) |
|---|---|---|---|---|---|
| Apple leaf | All categories | 99.54 | 99.39 | 99.46 | 99.41 |
| | Healthy | 100 | 99.66 | | |
| | Scab | 99.17 | 99.17 | | |
| | Rust | 99.01 | 97.18 | | |
| | Black rot | 100 | 99.55 | | |

**Table 9.** *Cont.*

| Plant | Categories | Precision (%) | Recall (%) | F1-Score (%) | Accuracy (%) |
|---|---|---|---|---|---|
| Corn leaf | All categories | 98.31 | 98.13 | 98.21 | 98.19 |
| | Healthy | 98.17 | 97.93 | | |
| | Rust | 98.54 | 98.25 | | |
| | Spot | 97.64 | 97.65 | | |
| | Leaf blight | 98.90 | 98.71 | | |
| Grape leaf | All categories | 98.47 | 98.45 | 98.45 | 98.52 |
| | Healthy | 98.59 | 98.22 | | |
| | Black measles | 98.37 | 98.93 | | |
| | Leaf blight | 98.64 | 98.75 | | |
| | Black rot | 98.28 | 97.91 | | |

## 4. Conclusions

In this paper, we proposed a network called BAM-Net, which aims to achieve the efficient identification of apple leaf diseases in complex backgrounds. We constructed a dataset of apple leaf images, totaling 17,688 images, in a complex background containing healthy leaves and five categories representing different diseases. To help the network train better, we used the BF-MSRCR algorithm to perform pre-processing operations on the leaf images in the original dataset to highlight the leaf color and texture features in the images. This approach effectively improved the accuracy of the subsequent network in identifying apple leaf diseases (+1.35%). With BAM-Net, we used ConvNext-T as the backbone network to extract the basic leaf disease feature information. To help the network more effectively locate leaves in complex backgrounds, we used ACAM to filter the complex interference information in images and to improve the network's focus on important features. Meanwhile, the MFRM module designed in this paper was used with the aim to refine the deep feature information and also to improve the network's ability to distinguish similar disease features.

In the experimental part of this study, we used a five-fold cross-validation method to validate the performance of BAM-Net. The results showed that BAM-Net achieved excellent performance in the classification of six apple leaves with an accuracy of 95.64% and an F1-score of 95.25%. In particular, for apple leaves suffering from brown spot and mosaic, the classification accuracy reached 95.61% and 95.97%, respectively. A comparison with the base network and the latest networks further validated the effectiveness of BAM-Net, which improved the accuracy of the model while maintaining the standard detection speed. This provides a reference value for modern agricultural producers to detect and identify apple leaf diseases in a timely manner and provides significant help for early agricultural conservation and production.

The proposed method can efficiently identify and classify apple leaf diseases in a complex context, but there are still the following shortcomings: (1) The dataset contains only six types of apple leaves, which is not comprehensive enough for practical applications. Therefore, subsequent studies will help to enrich the apple leaf disease categories in the dataset to help fruit farmers achieve more comprehensive detection and classification. (2) Although the number of parameters in BAM-Net was only 28.14M, it is still far from being truly lightweight. In the future, we will continue to explore more lightweight strategies and deploy the model on mobile applications. This will enable farmers to identify apple leaf diseases in real time using their smartphones anytime, anywhere. We believe that this will help farmers identify and treat apple leaf diseases more quickly, increasing their productivity and profitability. In addition, we plan to add a cloud-based database to the mobile app that will allow farmers to access data on known apple leaf

diseases and their treatments. This will further enhance the app's functionality and enable farmers to make more informed decisions about their disease management practices. Overall, we are committed to improving our technology to better meet the needs of the agricultural industry.

## References

1. Pardo, A.; Borges, P.A. Worldwide importance of insect pollination in apple orchards: A review. *Agric. Ecosyst. Environ.* **2020**, *293*, 106839. [CrossRef]
2. Zhong, Y.; Zhao, M. Research on deep learning in apple leaf disease recognition. *Comput. Electron. Agric.* **2019**, *168*, 105146. [CrossRef]
3. Khalil, A.J.; Barhoom, A.M.; Musleh, M.M.; Abu-Naser, S.S. Apple Trees Knowledge Based System. *Int. J. Acad. Eng. Res.* **2019**, *3*, 1–7.
4. Dong, Y.; Fu, Z.; Stankovski, S.; Peng, Y.; Li, X. A Cotton Disease Diagnosis Method Using a Combined Algorithm of Case-Based Reasoning and Fuzzy Logic. *Comput. J.* **2020**, *64*, 155–168. [CrossRef]
5. Tian, H.; Wang, T.; Liu, Y.; Qiao, X.; Li, Y. Computer vision technology in agricultural automation—A review. *Inf. Process. Agric.* **2019**, *7*, 1–19. [CrossRef]
6. Chakraborty, S.; Paul, S.; Rahat-uz-Zaman, M. Prediction of apple leaf diseases using multiclass support vector machine. In Proceedings of the 2021 2nd International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST), Dhaka, Bangladesh, 5–7 January 2021; pp. 147–151.
7. Chuanlei, Z.; Shanwen, Z.; Jucheng, Y.; Yancui, S.; Jia, C. Apple leaf disease identification using genetic algorithm and correlation based feature selection method. *Int. J. Agric. Biol. Eng.* **2017**, *10*, 74–83.
8. Zamani, A.S.; Anand, L.; Rane, K.P.; Prabhu, P.; Buttar, A.M.; Pallathadka, H.; Raghuvanshi, A.; Dugbakie, B.N. Performance of machine learning and image processing in plant leaf disease detection. *J. Food Qual.* **2022**, *2022*, 1598796. [CrossRef]
9. Tan, A.; Zhou, G.; He, M. Surface defect identification of Citrus based on KF-2D-Renyi and ABC-SVM. *Multimed. Tools Appl.* **2020**, *80*, 9109–9913. [CrossRef]
10. Jiang, H.; Li, X.; Shao, H.; Zhao, K. Intelligent fault diagnosis of rolling bearings using an improved deep recurrent neural network. *Meas. Sci. Technol.* **2018**, *29*, 065107. [CrossRef]
11. Deng, L.; Yu, D. Deep learning: Methods and applications. *Found. Trends®Signal Process.* **2014**, *7*, 197–387. [CrossRef]
12. Li, J.; Zhou, G.; Chen, A.; Lu, C.; Li, L. BCMNet: Cross-Layer Extraction Structure and Multiscale Downsampling Network With Bidirectional Transpose FPN for Fast Detection of Wildfire Smoke. *IEEE Syst. J.* **2022**, *17*, 1235–1246. [CrossRef]
13. Hu, Y.; Zhan, J.; Zhou, G.; Chen, A.; Cai, W.; Guo, K.; Hu, Y.; Li, L. Fast forest fire smoke detection using MVMNet. *Knowledg. Based Syst.* **2022**, *241*, 108219. [CrossRef]
14. Ali, R.; Chuah, J.H.; Abu Talip, M.S.; Mokhtar, N.; Shoaib, M.A. Structural crack detection using deep convolutional neural networks. *Autom. Constr.* **2022**, *133*, 103989. [CrossRef]
15. Kheradmandi, N.; Mehranfar, V. A critical review and comparative study on image segmentation-based techniques for pavement crack detection. *Constr. Build. Mater.* **2022**, *321*, 126162. [CrossRef]
16. Ibrahim, N.M.; Gabr, D.G.I.; Rahman, A.-U.; Dash, S.; Nayyar, A. A deep learning approach to intelligent fruit identification and family classification. *Multimed. Tools Appl.* **2022**, *81*, 27783–27798. [CrossRef]
17. Chen, X.; Zhou, G.; Chen, A.; Pu, L.; Chen, W. The fruit classification algorithm based on the multi-optimization convolutional neural network. *Multimed. Tools Appl.* **2021**, *80*, 11313–11330. [CrossRef]
18. Jiang, P.; Chen, Y.; Liu, B.; He, D.; Liang, C. Real-Time Detection of Apple Leaf Diseases Using Deep Learning Approach Based on Improved Convolutional Neural Networks. *IEEE Access* **2019**, *7*, 59069–59080. [CrossRef]
19. Bansal, P.; Kumar, R.; Kumar, S. Disease Detection in Apple Leaves Using Deep Convolutional Neural Network. *Agriculture* **2021**, *11*, 617. [CrossRef]

20. Suo, J.; Zhan, J.; Zhou, G.; Chen, A.; Hu, Y.; Huang, W.; Cai, W.; Hu, Y.; Li, L. CASM-AMFMNet: A Network Based on Coordinate Attention Shuffle Mechanism and Asymmetric Multi-Scale Fusion Module for Classification of Grape Leaf Diseases. *Front. Plant Sci.* **2022**, *13*, 846767. [CrossRef]

21. Liao, T.; Yang, R.; Zhao, P.; Zhou, W.; He, M.; Li, L. MDAM-DRNet: Dual Channel Residual Network with Multi-Directional Attention Mechanism in Strawberry Leaf Diseases Detection. *Front. Plant Sci.* **2022**, *13*, 869524. [CrossRef]

22. Sun, H.; Xu, H.; Liu, B.; He, D.; He, J.; Zhang, H.; Geng, N. MEAN-SSD: A novel real-time detector for apple leaf diseases using improved light-weight convolutional neural networks. *Comput. Electron. Agric.* **2021**, *189*, 106379. [CrossRef]

23. Yang, Q.; Duan, S.; Wang, L. Efficient Identification of Apple Leaf Diseases in the Wild Using Convolutional Neural Networks. *Agronomy* **2022**, *12*, 2784. [CrossRef]

24. Bi, C.; Xu, S.; Hu, N.; Zhang, S.; Zhu, Z.; Yu, H. Identification Method of Corn Leaf Disease Based on Improved Mobilenetv3 Model. *Agronomy* **2023**, *13*, 300. [CrossRef]

25. Barbedo, J.G.A. Impact of dataset size and variety on the effectiveness of deep learning and transfer learning for plant disease classification. *Comput. Electron. Agric.* **2018**, *153*, 46–53. [CrossRef]

26. Jobson, D.J.; Rahman, Z.-U.; Woodell, G.A. Retinex processing for automatic image enhancement. *J. Electron. Imaging* **2004**, *13*, 100–110. [CrossRef]

27. Wang, J.; Lu, K.; Xue, J.; He, N.; Shao, L. Single Image Dehazing Based on the Physical Model and MSRCR Algorithm. *IEEE Trans. Circuits Syst. Video Technol.* **2017**, *28*, 2190–2199. [CrossRef]

28. Yang, Y.n.; Jiang, Z.; Yang, C.; Xia, Z.; Liu, F. Improved retinex image enhancement algorithm based on bilateral filtering. In Proceedings of the 2015 4th International Conference on Mechatronics, Materials, Chemistry and Computer Engineering, Xi'an, China, 12–13 December 2015; pp. 1363–1369.

29. Liu, Z.; Mao, H.; Wu, C.-Y.; Feichtenhofer, C.; Darrell, T.; Xie, S. A convnet for the 2020s. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 11976–11986.

30. Zhao, Y.; Wang, G.; Tang, C.; Luo, C.; Zeng, W.; Zha, Z.J. A battle of network structures: An empirical study of cnn, transformer, and mlp. *arXiv* **2021**, arXiv:2108.13002.

31. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 10012–10022.

32. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.

33. Burt, P.J. Attention mechanisms for vision in a dynamic world. In Proceedings of the 9th International Conference on Pattern Recognition, Rome, Italy, 14 May–17 November 1988.

34. Brauwers, G.; Frasincar, F. A general survey on attention mechanisms in deep learning. *IEEE Trans. Knowl. Data Eng.* **2021**, *35*, 3279–3298. [CrossRef]

35. Hou, Q.; Zhou, D.; Feng, J. Coordinate attention for efficient mobile network design. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 13713–13722.

36. Yu, F.; Koltun, V. Multi-scale context aggregation by dilated convolutions. *arXiv* **2015**, arXiv:1511.07122.

37. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.

38. Zhang, H.; Wu, C.; Zhang, Z.; Zhu, Y.; Lin, H.; Zhang, Z.; Sun, Y.; He, T.; Mueller, J.; Manmatha, R. Resnest: Split-attention networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 2736–2746.

39. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.

40. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.

41. Woo, S.; Park, J.; Lee, J.-Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the European CONFERENCE on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.

42. Li, M.; Zhou, G.; Li, Z. Fast recognition system forTree images based on dual-task Gabor convolutional neural network. *Multimed. Tools Appl.* **2022**, *81*, 28607–28631. [CrossRef]

43. Liu, Z.; Hu, H.; Lin, Y.; Yao, Z.; Xie, Z.; Wei, Y.; Ning, J.; Cao, Y.; Zhang, Z.; Dong, L. Swin transformer v2: Scaling up capacity and resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 12009–12019.

44. Huang, X.; Chen, A.; Zhou, G.; Zhang, X.; Wang, J.; Peng, N.; Yan, N.; Jiang, C. Tomato leaf disease detection system based on FC-SNDPN. *Multimed. Tools Appl.* **2023**, *82*, 2121–2144. [CrossRef]

45. Mohanty, S.P.; Hughes, D.P.; Salathé, M. Using Deep Learning for Image-Based Plant Disease Detection. *Front. Plant Sci.* **2016**, *7*, 1419. [CrossRef] [PubMed]