IEEE Access
Multidisciplinary Rapid Review Open Access Journal

Received 27 July 2023, accepted 26 August 2023, date of publication 30 August 2023, date of current version 6 September 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3309925

RESEARCH ARTICLE

Bit-Plane and Correlation Spatial Attention Modules for Plant Disease Classification

XIAOTIAN WANG AND WEIQUN CAO
School of Information Science and Technology, Beijing Forestry University, Beijing 100083, China
Engineering Research Center for Forestry-Oriented Intelligent Information Processing, National Forestry and Grassland Administration, Beijing 100083, China
Corresponding author: Weiqun Cao (weiqun.cao@126.com)

ABSTRACT The effective and accurate classification of plant diseases is an important task for agricultural production. Therefore, some studies have utilized convolutional neural networks to identify categories of plant diseases, which can effectively reduce the reliance on crop experts. To further improve the accuracy of plant disease classification methods based on convolutional neural network, this paper proposes an attention model designed for plant disease classification tasks. The proposed attention model contains bit-plane attention and a correlation spatial attention. The bit-plane attention localizes disease areas by exploiting bit-plane information. The correlation spatial attention enhances the weight of important areas in the feature map by establishing the correlation between different areas. The accuracy of the proposed attention model inserted into ResNet101 on the AI Challenger 2018 and PlantVillage datasets is 87.11% and 99.82%, respectively. The performance is better than that of other methods studied on the public plant disease classification dataset. Experiments show that the proposed attention model outperforms the widely used universal attention models SE, CBAM, CA, ECA, BAM and GC. In addition, ablation experiments are conducted to verify the influences of different variants of the proposed attention model on the results.

INDEX TERMS Image classification, plant disease classification, convolutional neural network, attention mechanism, bit-plane attention.

I. INTRODUCTION

Food security is becoming increasingly important with the rapid growth of the global population. Plant diseases seriously affect plant yield and quality. Crop failures caused by plant diseases have triggered food crises in many countries, resulting in disastrous consequences. With the increase in the number of crop species, the expansion of planting areas and the diversity of cultivation methods, the classification of plant diseases has become a costly and time-consuming task. However, due to the limited number of agricultural experts, a method for automatically classifying crop diseases is needed. With the popularity of the internet and smartphones, farmers can take photos of crop diseases and utilize offline plant disease classification software to correctly classify crop disease types. This could reduce the dependence on crop experts and minimize food losses caused by crop diseases.

The associate editor coordinating the review of this manuscript and approving it for publication was Gangyi Jiang.

Plant disease recognition algorithms based on machine learning require much prior expert knowledge, and hand-designed feature extraction methods are used to extract discriminant features from images and send them to the classifier. Bashish et al. [1] used K-means clustering to segment images and input the texture features of objects into neural network to complete classification. Qin et al. [2] combined the K_median clustering algorithm and linear discriminant analysis to extract texture, color and shape features from plant images, and random forest, SVM and K-nearest neighbors classifier were used to complete disease classification. Machine learning methods require a large number of image preprocessing operations, which can bring additional overhead. There are many factors that affect the accuracy of the above machine learning methods, such as lighting in complex situations.

A convolutional neural network (CNN) makes full use of the end-to-end learning mode, has its own feature detection mechanism and is widely used in image detection and classification tasks. AlexNet [3] proposed by Alex Krizhevsky, was the champion of the ILSVRC competition. This model uses ReLU as the activation function of the convolutional neural network to solve the gradient dispersion problem of deep CNNs. VGGNet [4] uses small convolution kernel and designs a deep network structure to achieve improved classification accuracy. ResNet [5] is a landmark model in the field of convolutional neural networks. It addresses degradation in deep networks through residual connection. Deep learning methods are currently replacing machine learning methods because of their high accuracy.

At present, the classification methods of plant diseases mainly fall into two categories: deep learning and transfer learning. Deep learning methods need to design convolutional neural network structures according to the characteristics of plant disease images. Ferentinos et al. [6] introduced VGG network to plant disease identification tasks. Amara et al. [7] proposed a deep learning method for banana disease classification, which is based on the LeNet architecture and considers lighting, background, size and resolution in complex environments. Sladojevic et al. [8] proposed a convolutional neural network model to identify 13 different plant diseases, with an accuracy between 91% and 98%. Sun et al. [9] proposed a model for classifying 26 plant leaf diseases based on a convolutional neural network, which combines batch normalization and global pooling, and obtained a classification accuracy of 99.56%.

The accuracy of deep learning models depends on large dataset, and in the field of agriculture, the diversity of plant disease types cannot meet the modeling requirements of deep learning. The idea of plant disease classification based on transfer learning is to transfer knowledge from a source domain to a target domain by relaxing the assumption. Wang et al. [10] pretrained the convolutional neural network model in PlantVillage and adjusted the neural network parameters on their own dataset. The classification accuracy of crop disease images in a small dataset reached 90.84%. Long et al. [11] first pretrained the AlexNet model on the ImageNet dataset and then fine-tuned the camellia leaf disease image, achieving an accuracy of 96.53%. Yuan et al. [12] pretrained the VGGNet model on PlantVillage and fine-tuned a small dataset through training strategies. The experimental results showed that the accuracy reached 95.93%.

The above studies have made great progress, but to further improve the accuracy of plant disease classification, the convolutional neural network structure should be designed according to the characteristics of plant disease images to further improve the accuracy. However, exploring the optimal convolutional neural network structure is a time-consuming process. Therefore, some work [13], [14], [15], [16], [17] has introduced the attention mechanism into the model of plant disease classification based on convolutional neural networks. The introduction of an attention model can effectively improve the classification accuracy and avoid the time-consuming convolutional neural network design process.

However, the current models of plant disease classification based on attention mechanisms usually use variations of the universal attention model to enhance the performance of convolutional neural networks in plant disease classification tasks. To further improve the performance of the attention mechanism in the plant disease classification network, we propose an attention model specifically designed for plant disease classification to improve accuracy. We propose introducing bit-planes as auxiliary information to further improve the ability of the attention mechanism to locate disease areas. At the same time, we also design the correlation spatial attention by aggregating the information of different spatial areas. The proposed attention model can effectively improve the accuracy of plant disease classification by embedding the ResNet series network, which is widely used in the field of image classification. The proposed model is verified on two public plant disease classification datasets, and the experimental results show that the performance of the proposed attention module is superior to that of the widely studied attention models squeeze-and-excitation (SE) [18], convolutional block attention module (CBAM) [19], coordinate attention (CA) [20], efficient channel attention (ECA) [21], bottleneck attention module (BAM) (22), and global context (GC) [23]. By inserting the proposed attention model into the convolutional neural network, the accuracy of plant disease identification can be improved more efficiently, and the time-consuming convolutional neural network design process can be avoided. In addition, the accuracy of plant disease classification by inserting the proposed attention model into the ResNet101 network is better than that of previous work studied on public disease classification datasets.

The main contributions of this paper are summarized as follows:

- This paper introduces the bit-plane technique into the attention model for the first time.
- An attention model specially designed for plant disease classification task is proposed, which can improve the classification accuracy by inserting Res-Net101. The performance of this attention model is superior to that of the current influence universal attention model in plant disease classification task.
- Compared with existing methods studied on public plant disease classification datasets, the proposed method shows more efficient performance.

The rest of the paper is organized as follows: Section II provides a brief overview of existing attention models and existing work on plant disease classification based on attention mechanism. Section III is devoted to the proposed attention method. Section IV discusses the experiment and results. Section V illustrates the conclusion and the work that can be carried out in the future.

II. RELATED WORK

A. ATTENTION MECHANISM

The attention mechanism originated from the study of the human brain. To make rational use of limited resources, people selectively pay attention to the most important information while ignoring other redundant information. When convolutional neural network features are input into the next layer, the network cannot suppress the redundant information that affects the accuracy of the network. A neural network based on an attention mechanism gives lower weight to unnecessary information in the feature map and increases the weight of important information, thus achieving improved performance. SENet uses a squeeze-and-excitation module for the adaptive reweighting of different channels. The squeeze operation specifically compresses the feature map size by global average pooling to emphasize the relationships between channels. The excitation operation adaptively calibrates the weights of the different channels by compressing and expanding the channels. The CBAM was proposed as an attention mechanism that integrates channel attention and spatial attention. The CA attention module aggregates different spatial axis features and then encodes the feature map into two attention maps, each of which contains the remote dependence of the two spatial axis features. ECA as a channel attention model uses 1D convolution instead of a full connection layer to compress channels, thus significantly reducing the number of parameters. BAM applies spatial and channel attentional mechanisms in parallel, and its spatial attentional mechanism employs dilated convolution to enlarge receptive field. The GC attention model captures global dependencies by aggregating global context information. LMFFNet [24] proposed the MAD module for combining multiple scale features to generate more accurate feature maps. D2Anet [25] proposed a difference occurrence module to calculate local correlations between multi-level changes.

B. ATTENTION-BASED PLANT DISEASE CLASSIFICATION METHODS

Exploring the optimal network structure is a time-consuming task, and the introduction of an attention mechanism can effectively improve the accuracy achieved on the image recognition task. The SACNN [13] constructs a backbone network to extract global features and designs a self-attention network to extract local disease features. DCPSNET [14] proposes a CPSA attention structure, which first detects the position of the target in the feature map by using a position self-attention mechanism, and a channel self-attention mechanism is then used to explore the interdependencies between channels. ECA_ResNet [15] is based on ResNet as the backbone network and uses the dual-branch channel attention mechanism proposed by the authors to establish the dependency relationships between channels. Dual-branch channel attention uses GAP and GMP to compress the feature map size, and a 1D convolution filter feature independently selects key features in the dual branches via adaptive parameters. Ric-Net [16] improves CBAM in its backbone to further improve its plant disease recognition accuracy. DTL-SE-ResNet50 [17] is based on the attention model SE and is trained on ImageNet and additional plant disease datasets to improve the accuracy of plant disease recognition. However, these methods only utilize the universal attention models, ignoring the design of attention model for plant disease tasks.

III. METHOD

A. OVERVIEW

Images with the same labels and semantics can be misidentified as belonging to other categories due to differences in the angles, lighting conditions, and positions of the object. Attention mechanism-based models address this problem by establishing relationships between spatial areas. Due to the local receptive fields of convolutional neural networks, it is difficult to establish dependence among different spatial area features. We propose a correlation spatial attention module, which further improves the performance of the spatial attention mechanism by establishing relationships between different areal features. In the attention module of existing works, only the features extracted from the given image are used to establish context information, but this approach cannot further improve the accuracy of recognition due to the smaller differences between some categories. In this paper, we propose a bit-plane attention module, in which the performance of the attention module is enhanced by supplementary information according to the bit-plane information of the image. Therefore, the proposed attention model includes correlation spatial attention and bit-plane attention.

The ResNet series network consists of one convolutional layer, several residual blocks and a full connection layer. The number of residual blocks is determined by the depth of the ResNet network. For example, the ResNet18 network consists of 8 residual blocks. In the proposed attention model, three bit-plane images need to be calculated based on the plant disease image first, and then bit-plane information is fused by a 1×1 convolution to input into each residual block, as shown in Fig.1(a). Fig.1(b) shows the residual block structure adopted by ResNet18. The residual block of ResNet50 and ResNet101 is composed of three convolutional layers, as shown in Figure Fig.1 (c). The calculation process of the bit-plane image is given in chapter *B*. As shown in Fig.1(b), the proposed attention model is inserted after the second convolution layer of the residual block of ResNet18. Due to the deep layers of ResNet50 and ResNet101, the residual blocks are composed of two 1×1 convolutions and one 3×3 convolution. The proposed attention model is inserted after the second 1×1 convolution layer, as shown in Fig.1(c). The attention module can be placed after the convolutional layer to reassign weights to the features extracted from the convolutional layer to emphasize important information.

As shown in Fig.2, the proposed attention model is different from the universal attention models SE and CA. The SE and CA attention modules only utilize image information,

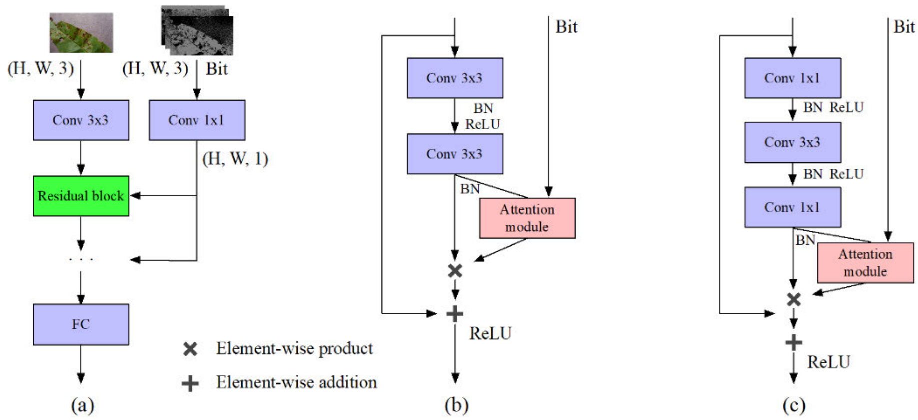

FIGURE 1. **The embedding position of the proposed attention module. H, W, C denotes the tensor shape height, width and depth. (a) Insert into the residual block of ResNet series Networks. (b) Inverted residual block in ResNet18. (c) Inverted residual block in ResNet50 and ResNet101.**

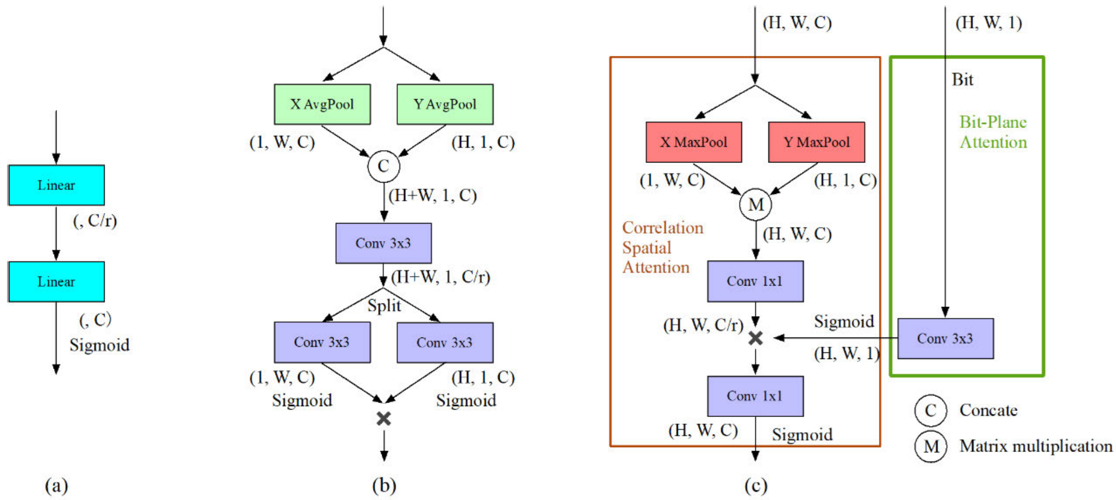

FIGURE 2. **Comparison of attention models. (a) The SE attention model. (b) The CA attention model. (c) The proposed attention model.**

while the proposed attention module utilizes both image information and bit-plane information. The SE attention mechanism is a classic channel attention model. The SE attention model uses two convolution layers to compress and expand channels to concentrate information and finally uses sigmoid to calculate the weight of different channels. CA, as a pioneering attention model in recent years, processes the information of the width axis and height axis, fuses the information through a 3×3 convolution layer, then processes the information of the width axis and height axis through two 3×3 convolution layers, and finally calculates the weight of different regions through sigmoid. The CA attention mechanism is a classic attention model that combines channel and spatial attention. The proposed model deals with the width and height axis information, respectively, but unlike the CA attention model, the proposed model uses matrix multiplication to recover the feature map size. The proposed attention model is shown in Fig.2(c), which consists of two parts: correlation spatial attention and bit-plane attention. The proposed attention model is described in detail in the following sections.

B. BIT-PLANE ATTENTION

We think that the bit-plane contains unique clues regarding the characteristics of the location of disease. Because most of the plant leaf disease areas are black, the gradient of the relevant area changes obviously, and the pixel value of the area of disease in the bit-plane image varies greatly between 0 and 1, so the bit-plane information can be used to locate the disease area. Therefore, we first propose the attention mechanism based on bit-plane.

Before using the bit-plane information, it is necessary to calculate the bit-plane image through the image. Let the gray level of the pixel at location (z, y) in an image with gray levels be represented as [26]:

$$f(x, y) = a_{K-1}2^{k-1} + a_{K-2}2^{k-2} + \ldots - + a_1 2^1 + a_0 2^0 \tag{1}$$

where $a_k, 0 \leq k \leq K-1$, is either 0 or 1. Let the kth-order bit-plane image be denoted by $b_k(x, y)$. For the case of an 8-bit image, the image is composed of eight-bit-planes $b_0(x, y) \sim b_7(x, y)$ ranging from plane 0 to plane 7. We only use the last three bit-planes to explore the relationships between the

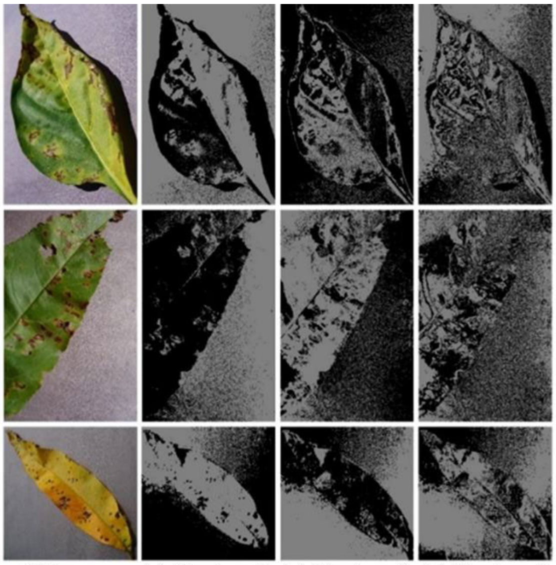

FIGURE 3. Visualization of bit-planes 5~7.

channels. Fig.3 shows the result of visualizing bit-planes 5~7. The higher-order bit-plane images contain more visually significant data. Among the different bit-planes, are areas in the image with high complexity. For example, if need convert an 8-bit RGB image in the pixel range 0 to 255 to a bit-plane image. First, an RGB image needs to be converted to a grayscale image, so K is defined as 8 in the above formula, $a_0 2^0$ is the pixel value of the bit-plane image at layer 0, and $a_{K-1} 2^{k-1}$ is the pixel value of the bit-plane image at layer 7. The range of the pixel value of the bit-plane image is 0 or 1.

As shown in Fig.3, plant disease-relevant areas tend to have large color variations, so pixel values vary significantly between 1 and 0 in bit-plane images. Therefore, the bit-plane image can be used as auxiliary information to generate the attention map, which makes the network focus on the areas of the image with large gradient changes. To feed the bit-plane image into the network, we need to aggregate the three channels of the bit-plane using a 1×1 convolution and then feed the aggregated feature map into the different residual blocks. In each residual block, a 3×3 convolution is used to generate the attention map to guide the correlation spatial attention focus to the disease area. This process can be expressed as:

$$BA = Sigmoid(C_{3\times3}(bit)) \tag{2}$$

where *bit* is the bit-plane feature map, $C_{3\times3}$ is the 3×3 convolution, *Sigmoid* is the sigmoid operation, and *BA* is the bit-plane attention map.

C. CORRELATION SPATIAL ATTENTION

Spatial attention focuses on areas that contain more important information. In existing works, pooling operations were applied to aggregate different channel information along the channel axis to highlight areas with important information. To focus on the areas in the given image that are more important to the classification task, we model the correlations among different spatial areas. Plant leaf disease areas often do not exist alone, and there are corresponding other disease areas in the leaves. When the spatial attention mechanism can establish the correlation between different disease areas, it can improve the classification accuracy. In the proposed spatial attention module, we use matrix multiplication to aggregate different spatial area features to establish correlations between different spatial areas.

As shown in Fig.2, the max pooling operation is applied to aggregate the information along the width and height axes to generate feature maps $Mw \in \mathbb{R}^{C\times W\times 1}$ and $Mh \in \mathbb{R}^{C\times 1\times H}$, respectively. Max pooling can better retain areas with large gradient changes without being affected by background information. To capture correlations across areas, matrix multiplication is applied to aggregate information from different spatial areas. The larger the dot product in the feature map is, the smaller the included angle between the vectors, which indicates that the correlation between the two areas is stronger. The feature map *Fmx* obtained after matrix multiplication can be expressed as:

$$Fmx = mm(GMP_w(x), GMP_h(x)) \tag{3}$$

where $GMP_w(x)$ and $GMP_h(x)$ denote global max pooling along the width and height axes, respectively, and *mm*() is the matrix multiplication operation. To enhance the weight of important channels, it is important to capture the relationship between channels. We utilize two 1×1 convolutional layers to compress and expand the feature map channels. The compression rate r is 32. In this process, we use the element-wise product operation to further guide the spatial attention to locate the disease location using the bit-plane attention map:

$$SSA = Sigmoid(C_2(C_1(Fmx) \times BA)) \tag{4}$$

where C_1 and C_2 are convolution layers with different numbers of convolution channels and *SSA* is the correlation spatial attention map.

IV. EXPERIMENTS

A. DATASETS AND IMPLEMENTATION DETAILS

The PlantVillage dataset [27] and AI Challenger 2018 dataset are utilized to train the proposed model. The PlantVillage dataset is an open source dataset with 54,306 images in 38 categories, which cover 24 types of diseases and 14 types of crops. The numbers of images in different categories are evenly distributed. The types of crops are grape, soybean, blueberry, cherry, orange, peach, bell pepper, potato, raspberry, squash, apple, strawberry, and tomato. The plant diseases include bacterial disease, mold disease, viral disease and mite disease. The AI Challenger 2018 dataset contains 10 crops and 27 diseases, with a total of 36,379 images divided into 61 categories. The images of the same class of diseases in the PlantVillage dataset are similar. Different from those in the PlantVillage dataset, the early disease images of crops in the AI Challenger 2018 dataset are very similar to the healthy images and the images of the same class in AI Challenger 2018 have larger differences, so they are more difficult to classify. We use all the classes in the dataset to train our model. The ratio of the training set to the test set is 8:2 for the datasets.

The proposed method is implemented with Python 3.8.8, PyTorch 1.10.2 and CUDA 10.2. All experiments are performed on an NVIDIA RTX2080TI GPU based on Ubuntu 18.04. During the training phase, Adam with weight decay 1e-4 is used to optimize the network parameters. The initial learning rate was set to 1e-3 and reduced to 1e-4 in epoch 5. The batch size is 32. We set the epoch to 50. The proposed model is pretrained on the ImageNet dataset.

B. EVALUATION METRICS

We evaluate the proposed model based on the following metrics: accuracy, precision, recall, and F1 score. Accuracy, as an important index of image classification, is defined as:

$$Accuracy = \frac{TN + TP}{TN + TP + FN + FP} \tag{5}$$

True negative (TN) represents the number of predicted results that are negative and actual class that are also negative. True positive (TP) represents the number where the predicted result is positive and the actual class is positive. False negative (FN) represents the number where the predicted result is negative and the actual class is positive. False positive (FP) represents the number of predicted results that are positive and the actual class that is negative.

Precision is used to calculate the number of objects classified as positive that are truly positive:

$$Precision = \frac{TP}{TP + FP} \tag{6}$$

Recall is used to calculate the proportion of positive samples that are correctly classified:

$$Recall = \frac{TP}{TP + FN} \tag{7}$$

The F1 score is used to calculate the weighted average value of the recall and precision metrics:

$$F1score = \frac{2 \times Recall \times Precision}{Recall + Precision} \tag{8}$$

C. EXPERIMENTAL RESULTS ON THE AI CHALLENGER 2018 DATASET

We compare the proposed attention model with the universal attention model widely used in recent years. Table 1 shows the results of different attention models on ResNet. As a channel attention mechanism, SE improves the accuracy by 0.29% in ResNet101. However, the SE attention model only enhances the important channel weights, so it cannot make the network accurately focus on the disease area. The CBAM attention model uses both channel and spatial attention mechanisms. The spatial attention mechanism makes the network focus on the disease area by assigning higher weights to the disease area in the feature map. The accuracy of CBAM is 0.78% higher than that of SE. As an advanced attention model in recent years, the CA attention model shows excellent performance on general tasks, and its performance in the field of plant disease recognition also exceeds that of the SE attention models. The ECA attention model is a channel attention model that performs better than the SE model and adds the fewest parameters among all the attention models. The BAM attention model combines channel and spatial attention, which shows excellent performance but also has the highest number of parameters. GC attention is modeled based on global context information, and the number of parameters is similar to CBAM, but the performance is better than CBAM. As an attention model specifically designed for the field of plant disease recognition, the BCSA model can better focus on disease areas with disease to enhance network performance. We also compared the attention models CAM [28] and CAE [29] that applied in other classification tasks. The CAM and CAE have much higher parameters than the universal attention model, while their accuracy has not significantly improved. Compared with other universal attention models, the BCSA attention model can improve the classification accuracy of plant diseases more effectively, and only a small number of parameters were added. The computational complexity of the proposed model is not significantly different from that of other universal attention models.

We also compared our model with several other plant disease classification models, and the results are shown in Table 2. All the comparison methods were studied on public plant disease classification datasets. The accuracy of the above models was obtained by training on the AI Challenger 2018 dataset. References [6], [30], and [31] was an early work for plant disease classification. They used a universal image classification model to classify plant disease images. Reference [32] improved the VGG19 model and enhanced the accuracy of the VGG model in plant disease recognition task. Reference [33] introduced the attention mechanism into the CNN model, which has only 0.7 M parameters and is suitable for running on mobile devices. Reference [15] proposed an improved CBAM attention model, which can significantly improve the accuracy of plant disease recognition after being inserted into ResNet variant models. Reference [16] used the CBAM attention model to enhance the accuracy of CNN classification, and satisfactory results were obtained. The model proposed by [34] and [35] achieves excellent performance with a small number of parameters. Due to the excellent image classification performance of ResNet101, an accuracy of 87.11% is achieved after the insertion of the proposed classification performance of ResNet101, an accuracy of 87.11% is achieved after the insertion of the proposed attention model, which is also the highest accuracy among these models.

The testing error during training is shown in Fig.4. As the number of epochs increases, the recognition accuracies of several models increase gradually on the test set. We can observe that ResNet101 with the SE, CBAM, CA, ECA, BAM, GC and BCSA attention models reduces the overall error on the test set with increasing epochs. For all models, the test error decreases rapidly in the first 10 epochs and becomes stable after 16 epochs. The test error of ResNet

TABLE 1. Performance comparison among different attention models on the AI challenger 2018 dataset.

Models	Settings	Param	GFlops	Precision	Recall	F1 score	Accuracy
ResNet18	-	**11.20 M**	**35.70 G**	81.81%	79.91%	80.85%	84.82%
ResNet18	+SE	11.28 M	35.72 G	83.34%	81.20%	82.26%	85.04%
ResNet18	+CA	11.26 M	35.74 G	83.47%	81.80%	82.63%	85.24%
ResNet18	+CBAM	11.28 M	35.73 G	83.85%	82.02%	82.92%	85.52%
ResNet18	+ECA	11.20 M	35.71 G	83.89%	82.08%	82.98%	85.69%
ResNet18	+BAM	11.39 M	36 G	83.96%	82.34%	83.14%	85.76%
ResNet18	+GC	11.29 M	35.72 G	83.82%	82.03%	82.92%	85.59%
ResNet18	+CAM	11.98 M	35.86 G	83.78%	82.01%	82.89%	85.47%
ResNet18	+CAE	12.58 M	40.06 G	83.92%	82.26%	83.08%	85.72%
ResNet18	+BCSA	11.24 M	35.76 G	**84.73%**	**82.90%**	**83.81%**	**86.66%**
ResNet50	-	**23.63 M**	**83.94 G**	83.10%	80.18%	81.61%	84.95%
ResNet50	+SE	26.11 M	84.05 G	83.44%	81.70%	82.56%	85.21%
ResNet50	+CA	25.49 M	84.29 G	83.52%	81.95%	82.73%	85.38%
ResNet50	+CBAM	26.16 M	84.08 G	83.92%	82.33%	83.12%	85.72%
ResNet50	+ECA	23.63 M	84.01 G	84.02%	82.49%	83.25%	85.87%
ResNet50	+BAM	28.86 M	84.17 G	84.19%	82.51%	83.34%	86.03%
ResNet50	+GC	26.17 M	84.05 G	83.76%	82.25%	83.00%	85.62%
ResNet50	+CAM	46.34 M	85.11 G	83.84%	82.26%	83.04%	85.66%
ResNet50	+CAE	63.86 M	221.84 G	84.15%	82.49%	83.31%	85.98%
ResNet50	+BCSA	24.88 M	84.38 G	**84.91%**	**83.05%**	**83.97%**	**86.76%**
ResNet101	-	**42.62 M**	**161.94 G**	83.52%	81.85%	82.68%	85.27%
ResNet101	+SE	47.35 M	162.13 G	83.76%	82.15%	82.95%	85.53%
ResNet101	+CA	46.20 M	162.59 G	84.04%	82.37%	83.20%	85.77%
ResNet101	+CBAM	47.40 M	162.17 G	84.42%	82.62%	83.51%	86.31%
ResNet101	+ECA	42.63 M	162.08 G	84.62%	82.72%	83.66%	86.54%
ResNet101	+BAM	52.50 M	162.32 G	84.72%	82.91%	83.81%	86.65%
ResNet101	+GC	47.43 M	162.13 G	84.55%	82.65%	83.59%	86.42%
ResNet101	+CAM	85.55 M	163.85 G	84.33%	82.55%	83.43%	86.22%
ResNet101	+CAE	118.55 M	446 G	84.64%	82.78%	83.70%	86.44%
ResNet101	+BCSA	44.99 M	162.45 G	**85.23%**	**83.45%**	**84.33%**	**87.11%**

TABLE 2. Performance comparison among different work on the AI challenger 2018 dataset.

Study	Year	Network	Param	Accuracy
Ferentinos [6]	2018	VGG	138 M	82.57%
Too et al. [30]	2019	DenseNet	8 M	84.25%
Kamal et al. [31]	2019	MobileNet	0.5 M	83.74%
Chen et al. [32]	2020	VGG19 variant	41 M	83.61%
Ramamurthy et al. [33]	2020	CNN+attnetion	0.7 M	78.12%
Ronghua Gao et al. [15]	2021	ResNet18 variant +attention	51 M	86.09%
Zhao et al. [16]	2022	CNN+attention	59 M	84.91%
Li et al. [34]	2023	CNN	4 M	85.73%
Singh Thakur et al. [35]	2023	VGG variant	6 M	85.37%
Our	2023	ResNet101+attention	45 M	**87.11%**

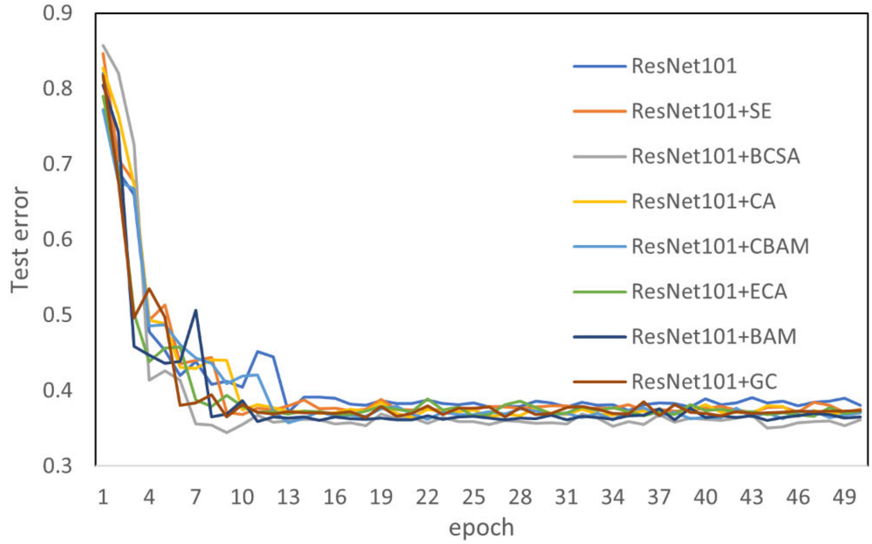

FIGURE 4. Loss curve of the attention model in the test set.

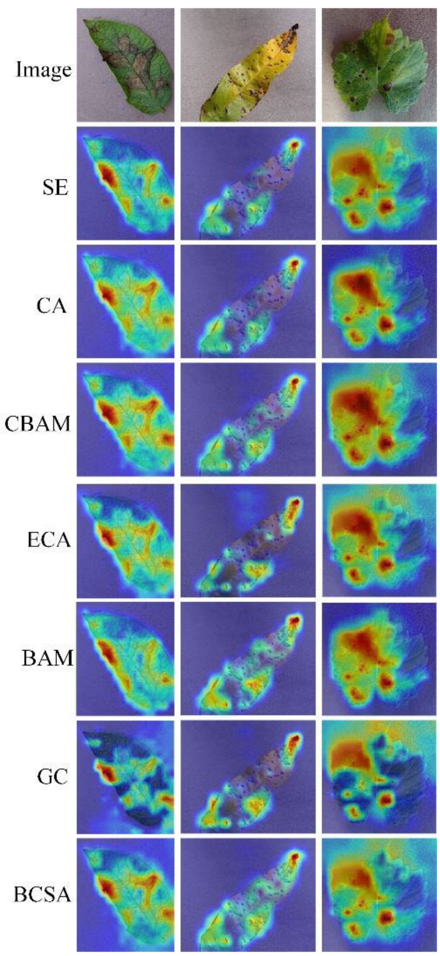

FIGURE 5. Visualization of the disease classification process.

without attention is higher than that of the attention model after epoch 13. Our model achieves the highest accuracy with almost no fluctuation.

We use class activation mapping (CAM) [36] to visualize the model disease recognition process. As shown in Fig.5, the red part represents the part concerned by the model, the blue part represents the part ignored by the model, and the yellow part represents the attention between the above two colors. The ability to accurately focus on disease areas in an image is the key to measuring the performance of an attention model. The SE attention can only enhance the weight of important information channels but not the weight of important spatial areas, so it is difficult to accurately focus on plant disease areas. The ECA and GC attention models are also susceptible to interference from unrelated regions. The CBAM, BAM and CA attention combine channel and spatial attention mechanisms, so they can focus more accurately on the area where the disease is located, but they are also disturbed by irrelevant areas. As an attention model specially designed for plant disease recognition, the proposed attention model can accurately focus on the disease area and is not easily disturbed by irrelevant areas.

D. EXPERIMENTAL RESULTS ON THE PLANTVILLAGE DATASET

The images of different classes of the PlantVillage dataset are more obvious. As shown in Table 3, the accuracy of ResNet in the PlantVillage dataset can achieve nearly 99%. The channel attention model SE can only improve the accuracy by 0.09% on ResNet101, while the CBAM, BAM and CA models combining channel and spatial attention improve the accuracy by 0.38%, 0.52% and 0.21%, respectively. The ECA attention model improved accuracy by 0.37% with almost no additional parameters. The accuracy of the global attention model GC was improved by 0.44% after it was plugged into the ResNet101 network. Although the accuracy of the ResNet series model reached 99% on the PlantVillage dataset, the accuracy of classification continued to be improved after the insertion of the attention model. The performance of

TABLE 3. Performance comparison among different attention models on the plantvillage dataset.

Models	Settings	Param	GFlops	Precision	Recall	F1 score	Accuracy
ResNet18	-	**11.20 M**	**35.70 G**	98.25%	98.14%	98.19%	98.62%
ResNet18	+SE	11.28 M	35.72 G	98.49%	98.32%	98.40%	98.77%
ResNet18	+CA	11.26 M	35.74 G	98.70%	98.59%	98.64%	98.91%
ResNet18	+CBAM	11.28 M	35.73 G	98.71%	98.64%	98.67%	99.12%
ResNet18	+ECA	11.20 M	35.71 G	98.82%	98.67%	98.84%	99.18%
ResNet18	+BAM	11.39 M	36 G	98.79%	98.69%	98.94%	99.21%
ResNet18	+GC	11.29 M	35.72 G	98.75%	98.62%	98.88%	99.16%
ResNet18	+CAM	11.98 M	35.86 G	98.67%	98.57%	98.62%	99.06%
ResNet18	+CAE	12.58 M	40.06 G	98.77%	98.72%	98.74%	99.18%
ResNet18	+BCSA	11.24 M	35.76 G	**99.12%**	**99.02%**	**99.07%**	**99.36%**
ResNet50	-	**23.63 M**	**83.94 G**	98.56%	98.40%	98.48%	98.84%
ResNet50	+SE	26.11 M	84.05 G	98.61%	98.56%	98.59%	98.92%
ResNet50	+CA	25.49 M	84.29 G	98.72%	98.61%	98.66%	98.98%
ResNet50	+CBAM	26.16 M	84.08 G	98.75%	98.57%	98.66%	99.02%
ResNet50	+ECA	23.63 M	84.01 G	98.88%	98.62%	98.72%	99.16%
ResNet50	+BAM	28.86 M	84.17 G	98.83%	98.67%	98.75%	99.15%
ResNet50	+GC	26.17 M	84.05 G	98.81%	98.68%	98.74%	99.08%
ResNet50	+CAM	46.32 M	85.11 G	98.77%	98.45%	98.61%	98.96%
ResNet50	+CAE	63.90 M	221.84 G	98.76%	98.64%	98.70%	99.04%
ResNet50	+BCSA	24.88 M	84.38 G	**99.24%**	**99.07%**	**99.15%**	**99.52%**
ResNet101	-	**42.62 M**	**161.94 G**	98.67%	98.52%	98.59%	98.95%
ResNet101	+SE	47.35 M	162.13 G	98.77%	98.53%	98.65%	99.04%
ResNet101	+CA	46.20 M	162.59 G	98.85%	98.64%	98.74%	99.17%
ResNet101	+CBAM	47.40 M	162.17 G	99.11%	98.98%	99.04%	99.33%
ResNet101	+ECA	42.63 M	162.08 G	99.05%	98.92%	98.98%	99.32%
ResNet101	+BAM	52.49 M	162.32 G	99.13%	98.97%	99.05%	99.47%
ResNet101	+GC	47.43 M	162.13 G	98.77%	98.83%	98.80%	99.29%
ResNet101	+CAM	85.55 M	163.85 G	99.08%	98.87%	98.97%	99.25%
ResNet101	+CAE	118.60 M	446.16 G	99.12%	98.92%	99.02%	99.42%
ResNet101	+BCSA	44.99 M	162.45 G	**99.44%**	**99.28%**	**99.36%**	**99.82%**

TABLE 4. Performance comparison among different work on the plantvillage dataset.

Study	Year	Network	Param	Accuracy
Ferentinos [6]	2018	VGG16	138 M	97.32%
Too et al. [30]	2019	DenseNet	8 M	98.16%
Kamal et al. [31]	2019	MobileNet	0.5 M	97.18%
Chen et al. [32]	2020	VGG19 variant	41 M	98.24%
Ramamurthy et al. [33]	2020	CNN+attnetion	0.7 M	97.86%
Ronghua Gao et al. [15]	2021	ResNet18 variant +attention	51 M	99.41%
Zhao et al. [16]	2022	CNN+attention	59 M	98.76%
Li et al. [34]	2023	CNN	4 M	98.49%
Singh Thakur et al. [35]	2023	VGG variant	6 M	98.72%
Our	2023	ResNet101+ attention	45 M	**99.82%**

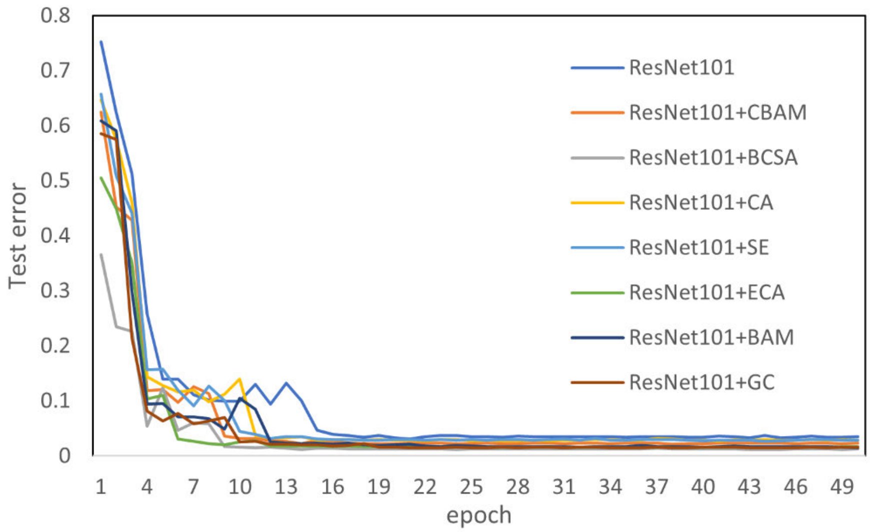

FIGURE 6. **Loss curve of the attention model in the test set.**

CAM attention model is lower than that of CBAM attention model. The CAE attention model has higher accuracy than the CAM attention model, but the number of parameters also significantly increases. By applying the proposed attention model, the accuracy of ResNet101 is improved by 0.87%, and the number of parameters is only increased by 2.4 M. Experimental results on the PlantVillage dataset also show the effectiveness of the proposed attention model in plant disease classification tasks.

We also compared our method with other plant disease classification models on the PlantVillage dataset. The accuracy of most models is close to 99%, as shown in Table 4. Reference [6] used the VGG16 network to classify plant diseases. The accuracy of this model is 97.32% in the PlantVillage dataset. References [30] and [31] also used the universal image classification model to classify plant diseases, with accuracies of 98.16% and 97.18%, respectively. References [32] and [35] improved the VGG16 model, with improved accuracy compared to [6]. Reference [34] designed a CNN network for the characteristics of plant diseases. Because it has fewer model parameters, it is more suitable to run on mobile devices. Due to the performance of ResNet in image classification, both the proposed method and [15] achieve satisfactory accuracy. The number of parameters in the proposed model is lower than that in [15], and the accuracy increases by 0.41%.

Fig.6 illustrates the error on the test set during training. In the first four epochs, the error of all models decreases rapidly, and the fluctuation of all models from epoch 4 to epoch 16 is drastic. Since all models in the PlantVillage dataset can achieve high accuracy, the error stable value is reached from epoch 16 to epoch 50. The convergence performance of ResNet101 with attention is stronger than that of ResNet101, and the experimental results show that the attention model has better convergence performance.

E. ABLATION EXPERIMENTS FOR THE PROPOSED ATTENTION MODEL

The previous experiments have proven the effectiveness of the proposed attention model on plant disease classification tasks. We also designed several experiments to verify the effects of different components of the proposed attention model on the results.

To verify the effectiveness of the proposed bit-plane attention module, we design an ablation experiment to verify the performance improvement brought by the bit-plane module. As shown in Table 5, ResNet101 with the addition of the bit-plane attention module has 0.58% and 0.3% improvements on the AI Challenger 2018 and PlantVillage datasets, respectively. That is, after adding the bit-plane attention module, the attention can further locate the defect area according to the bit-plane information, thereby improving the accuracy of the model. Applying the bit-plane attention module will not add too many parameters to the model.

To demonstrate the performance of the proposed correlation spatial attention, we conduct a series of ablation experiments. As shown in Table 6, after adding correlation spatial attention, the ResNet101 network has 0.7% and 1.53% improvement on the PlantVillage and AI Challenger 2018 datasets, respectively.

Overall, there were multiple disease areas in the same leaf, and the correlation spatial attention can aggregate the different disease area information to establish correlations to improve the performance of the model. Experimental results show that when the correlation spatial attention module is inserted, the accuracy of plant disease recognition tasks can be improved.

We also tested the effect of the pooling operation in the correlation spatial attention module on the results. We can observe from Table 7 that the maximum performance can be achieved when the correlation spatial attention module aggregates different spatial axis information using max pooling.

TABLE 5. Evaluation results of variations in the bit-plane attention module on the plantvillage and AI challenger 2018 datasets. the backbone of the network is ResNet101.

Setting	Param	PlantVillage	AI Challenger 2018
w/o bit-plane attention	**44.99 M**	99.52%	86.53%
+ bit-plane attention	44.99 M	**99.82%**	**87.11%**

TABLE 6. Evaluation results of the correlation spatial attention module on the plantvillage and AI chal-lenger 2018 datasets. the backbone of the network is ResNet101.

Setting	Param	PlantVillage	AI Challenger 2018
w/o correlation spatial attention	**42.62 M**	99.12%	85.58%
+ correlation spatial attention	44.99 M	**99.82%**	**87.11%**

TABLE 7. Evaluation results of pooling in the correlation spatial attention module on the PlantVillage and AI Challenger 2018 datasets. The backbone of the network is ResNet101.

Setting	PlantVillage	AI Challenger 2018
max pooling	**99.82%**	**87.11%**
avg pooling	99.26%	86.69%

TABLE 8. Evaluation results of ratio r in the correlation spatial attention module on the PlantVillage and AI Challenger 2018 datasets. The backbone of the network is ResNet101.

Setting	Param	PlantVillage	AI Challenger 2018
4	61.59 M	99.12%	85.34%
8	52.10 M	99.59%	86.52%
16	47.36 M	99.49%	86.78%
32	44.99 M	**99.82%**	**87.11%**
64	**43.80 M**	99.45%	86.66%

This is because max pooling can retain the texture and color change information in the feature map and eliminate the interference of irrelevant information. Therefore, we utilize max pooling to aggregate the different spatial axis information to subsequently establish the correlation between different areas. Table 8 shows the influence of different ratios r in the correlation spatial attention module on the results. We set different ratios r to observe the performance change. The larger the ratio r value is, the fewer the model parameters. When the r value decreases, the model parameters increase accordingly. The best performance can be achieved by setting an appropriate value of r. We can observe that the best performance is achieved when the value of r is 32.

V. CONCLUSION

In this paper, an attention model designed specifically for the task of disease classification is proposed to enhance the accuracy of the plant disease classification task and reduce the time-consuming convolutional neural network design process. A bit-plane attention mechanism has been proposed to use bit-plane information to locate disease locations. This is also the first time that bit-plane technology has been introduced into the attention mechanism. The correlation spatial attention enhances the attention performance by matrix multiplication to establish the correlation of diseases in different areas. After the proposed attention model is inserted into the ResNet101 network, the accuracy of the test set on the AI Challenger 2018 and PlantVillage datasets is 87.11% and 99.82%, respectively. The performance of our method is better than that of the model that studies on the public plant disease classification dataset. Experimental results on two public plant disease datasets show that the proposed attention model inserted into the ResNet can enhance the accuracy of plant disease classification and that the proposed attention model outperforms the widely used SE, CBAM, CA, ECA, BAM, and GC attention models. In addition, we also design ablation experiments to verify the influence of different variants of the proposed attention model on the results. In future work, researchers can continue to improve the attention model proposed in this paper and insert it into their own CNN model to improve the accuracy of plant disease classification. In addition, researchers can further study the bit-plane attention mechanism to improve its performance and extend it to other image processing tasks.

REFERENCES

[1] D. Al Bashish, M. Braik, and S. Bani-Ahmad, "Detection and classification of leaf diseases using K-means-based segmentation and neural-networks-based classification," *Inf. Technol. J.*, vol. 10, no. 2, pp. 267–275, Jan. 2011.

[2] F. Qin, D. Liu, B. Sun, L. Ruan, Z. Ma, and H. Wang, "Identification of alfalfa leaf diseases using image recognition technology," *PLoS ONE*, vol. 11, no. 12, Dec. 2016, Art. no. e0168274.

[3] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017.

[4] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[5] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[6] K. P. Ferentinos, "Deep learning models for plant disease detection and diagnosis," *Comput. Electron. Agricult.*, vol. 145, pp. 311–318, Feb. 2018.

[7] J. Amara, B. Bouaziz, and A. Algergawy, "A deep learning-based approach for banana leaf diseases classification," in *Proc. Datenbanksysteme Bus., Technol. Web (BTW)-Workshopband*, 2017, pp. 79–88.

[8] S. Sladojevic, M. Arsenovic, A. Anderla, D. Culibrk, and D. Stefanovic, "Deep neural networks based recognition of plant diseases by leaf image classification," *Comput. Intell. Neurosci.*, vol. 2016, pp. 1–11, May 2016.

[9] J. Sun, W. Tan, H. Mao, X. Wu, Y. Chen, and L. Wang, "Recognition of multiple plant leaf diseases based on improved convolutional neural network," *Trans. Chin. Soc. Agricult. Eng.*, vol. 33, no. 19, pp. 209–215, 2017.

[10] J. Wang, L. Chen, J. Zhang, Y. Yuan, M. Li, and W. Zeng, "CNN transfer learning for automatic image-based classification of crop disease," in *Image and Graphics Technologies and Applications*. Beijing, China: Springer, Apr. 2018.

[11] M. Long, C. Ouyang, H. Liu, and Q. Fu, "Image recognition of *Camellia oleifera* diseases based on convolutional neural network & transfer learning," *Trans. CSAE*, vol. 34, no. 18, pp. 194–201, 2018.

[12] Y. Yuan, S. Fang, and L. Chen, "Crop disease image classification based on transfer learning with DCNNs," *Pattern Recognition and Computer Vision*. Guangzhou, China: Springer, Nov. 2018.

[13] W. Zeng and M. Li, "Crop leaf disease recognition based on self-attention convolutional neural network," *Comput. Electron. Agricult.*, vol. 172, May 2020, Art. no. 105341.

[14] Y. Zhang and Y. P. Liu, "Identification of navel orange diseases and pests based on the fusion of DenseNet and self-attention mechanism," *Comput. Intell. Neurosci.*, vol. 2023, Jun. 2023, Art. no. 5436729.

[15] R. Gao, R. Wang, L. Feng, Q. Li, and H. Wu, "Dual-branch, efficient, channel attention-based crop disease identification," *Comput. Electron. Agricult.*, vol. 190, Nov. 2021, Art. no. 106410.

[16] Y. Zhao, C. Sun, X. Xu, and J. Chen, "RIC-Net: A plant disease classification model based on the fusion of inception and residual structure and embedded attention mechanism," *Comput. Electron. Agricult.*, vol. 193, Feb. 2022, Art. no. 106644.

[17] X. Zhao, K. Li, Y. Li, J. Ma, and L. Zhang, "Identification method of vegetable diseases based on transfer learning and attention mechanism," *Comput. Electron. Agricult.*, vol. 193, Feb. 2022, Art. no. 106703.

[18] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.

[19] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 3–19.

[20] Q. Hou, D. Zhou, and J. Feng, "Coordinate attention for efficient mobile network design," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 13713–13722.

[21] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-Net: Efficient channel attention for deep convolutional neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 11534–11542.

[22] J. Park, S. Woo, J.-Y. Lee, and I. So Kweon, "BAM: Bottleneck attention module," 2018, *arXiv:1807.06514*.

[23] Y. Cao, J. Xu, S. Lin, F. Wei, and H. Hu, "GCNet: Non-local networks meet squeeze-excitation networks and beyond," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 1971–1980.

[24] M. Shi, J. Shen, Q. Yi, J. Weng, Z. Huang, A. Luo, and Y. Zhou, "LMFFNet: A well-balanced lightweight network for fast and accurate semantic segmentation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 6, pp. 3205–3219, Jun. 2023.

[25] J. Mei, Y.-B. Zheng, and M.-M. Cheng, "D2ANet: Difference-aware attention network for multi-level change detection from satellite imagery," *Comput. Vis. Media*, vol. 9, no. 3, pp. 563–579, Sep. 2023.

[26] J. W. Schwartz and R. C. Barker, "Bit-plane encoding: A technique for source encoding," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-2, no. 4, pp. 385–392, Jul. 1966.

[27] D. P. Hughes and M. Salathe, "An open access repository of images on plant health to enable the development of mobile disease diagnostics," 2015, *arXiv:1511.08060*.

[28] L. Xue, X. Zeng, and A. Jin, "A novel deep-learning method with channel attention mechanism for underwater target recognition," *Sensors*, vol. 22, no. 15, p. 5492, Jul. 2022.

[29] G. Cheng, P. Lai, D. Gao, and J. Han, "Class attention network for image recognition," *Sci. China Inf. Sci.*, vol. 66, no. 3, Mar. 2023, Art. no. 132105.

[30] E. C. Too, L. Yujian, S. Njuki, and L. Yingchun, "A comparative study of fine-tuning deep learning models for plant disease identification," *Comput. Electron. Agricult.*, vol. 161, pp. 272–279, Jun. 2019.

[31] K. Kc, Z. Yin, M. Wu, and Z. Wu, "Depthwise separable convolution architectures for plant disease classification," *Comput. Electron. Agricult.*, vol. 165, Oct. 2019, Art. no. 104948.

[32] J. Chen, J. Chen, D. Zhang, Y. Sun, and Y. A. Nanehkaran, "Using deep transfer learning for image-based plant disease identification," *Comput. Electron. Agricult.*, vol. 173, Jun. 2020, Art. no. 105393.

[33] R. Karthik, M. Hariharan, S. Anand, P. Mathikshara, A. Johnson, and R. Menaka, "Attention embedded residual CNN for disease detection in tomato leaves," *Appl. Soft Comput.*, vol. 86, Jan. 2020, Art. no. 105933.

[34] M. Li, G. Zhou, A. Chen, L. Li, and Y. Hu, "Identification of tomato leaf diseases based on LMBRNet," *Eng. Appl. Artif. Intell.*, vol. 123, Aug. 2023, Art. no. 106195.

[35] P. S. Thakur, T. Sheorey, and A. Ojha, "VGG-ICNN: A lightweight CNN model for crop disease identification," *Multimedia Tools Appl.*, vol. 82, no. 1, pp. 497–520, Jan. 2023.

[36] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," *Int. J. Comput. Vis.*, vol. 128, no. 2, pp. 336–359, Feb. 2020.

XIAOTIAN WANG is currently pursuing the Ph.D. degree with the Department of Information Science and Technology, Beijing Forestry University, Beijing, China. His research interests include artificial intelligence and digital image processing.

WEIQUN CAO received the Ph.D. degree from the State Key Laboratory of CAD and CG, Zhejiang University, China, in March 2000. From June 2000 to June 2004, she was a Research Scientist with the Institute of Media Communication, German National Research Center of Information Technology (GMD-IMK, later merged as Fraunhofer Institute of Media Communication). She is a Professor with the School of Information Science and Technology, Beijing Forestry University, China. Her current research interests include digital image processing and computer graphics.