A Crop Leaf Disease Image Recognition Method Based on Bilinear Residual Networks

Yun He,1,2,3 Quan Gao,1 and Zifei Ma4

1School of Big Data, Yunnan Agricultural University, Kunming, Yunnan, China
2Key Laboratory for Crop Production and Intelligent Agriculture of Yunnan Province, Kunming, Yunnan, China
3State Key Laboratory for Conservation and Utilization of Bio-Resource, Yunnan, Kunming, Yunnan, China
4School of Water Conservancy, Yunnan Agricultural University, Kunming, Yunnan, China

Correspondence should be addressed to Zifei Ma; mzf@ynau.edu.cn

Received 16 February 2022; Revised 2 April 2022; Accepted 4 April 2022; Published 25 April 2022

Copyright © 2022 Yun He et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Deep learning models are widely used in crop leaf disease image recognition. These models can be divided into two categories: global model and local model. The global model directly takes the whole leaf disease images as input to training and recognition. It can achieve end-to-end training and recognition, which is very convenient to use. But this kind of model cannot very accurately and completely extract the features from the very small diseased spots in the image. Before training and recognizing, the local model needs to extract the diseased spots part from the image by image segmentation technology. Then the local model takes the disease spots part images as input to training and recognition. Features extracted by local model are more accurate and complete. But this kind of model cannot achieve end-to-end training and recognition, and the image segmentation will bring additional overhead. Considering the disadvantage of global model and local model, we proposed a crop leaf disease image recognition method based on bilinear residual networks (named DIR-BiRN). DIR-BiRN extracts features by two residual networks feature extractors and then integrates the features by a bilinear pooling function. By this way, it can extract features more accurately and completely while achieving end-to-end training and recognition. Experiments on the PlantVillage dataset show that, when compared with the standard ResNet-18 model, the DIR-BiRN improves on accuracy performance, recall performance, precision performance, and F1-measure performance by averages of 0.2918, 0.81641, 0.59185, and 0.52151 percentage points, respectively.

1. Introduction

Crop leaf disease image recognition is one of the most important agricultural engineering technologies. It is also one of the most important supporting technologies for crop disease recognition and control, which can provide effective guarantee for the safety of agricultural production [1]. How to quickly and accurately identify crop leaf diseases is one of the key research problems in the researches of crop leaf disease image recognition.

To address this problem, machine learning methods have long been applied in the crop leaf disease image recognition research, such as support vector machine [2], linear discriminant analysis [3], K-means [4], and Bayesian network [5]. These methods require manual feature selection, which will lead to high overhead. Moreover, the performance of these methods is limited. Usually, they can only recognize few diseases for a single crop [6]. In recent years, deep learning technology has developed rapidly and has become a popular technology in the research of crop leaf image disease recognition.

Different from the traditional machine learning methods, deep learning methods can be applied in an end-to-end way while extracting the features automatically and accurately [6]. The concept of deep learning comes from artificial neural network, but it constructs a deeper network model. As early as 2007, researchers [7] applied artificial neural network to the disease image recognition of Phalaenopsis seedling, but this method still needs to select features manually before applying. In 2012, Alex and others [8] applied the convolutional neural networks (CNN) to image recognition and achieved top-1 error rate of 15.3% on
the ImageNet dataset. Since then, many researchers have carried out studies to apply the CNN models to crop disease image recognition. Amara et al. [9] realized banana leaf disease image recognition based on LeNet and verified the effectiveness of deep learning model in the complex environment. Wu et al. [10] applied the GoogLeNet to classify five diseases of tomato and achieved an accuracy of 94.33%. Durmus et al. [11] trained an AlexNet to classify ten diseases of tomato leaves and achieved an accuracy of 95.65%. Srdjan and others [12] developed a deep convolutional neural network (DCNN) to detect different disease of 13 crops, with an accuracy of 91%–98%. Hu et al. [13] suggested a ResNet model to classify 59 diseases of different crops and achieved an accuracy of 85.22%. LeNet, ResNet, AlexNet, and GoogLeNet all belong to CNN models. These studies have verified the effectiveness of deep learning technology in the crop disease image recognition. The current research are no longer limited to detecting whether the crop is healthy or unhealthy; they also need to classify the crops that are unhealthy to a specific disease category [14]. This means we need more effective deep learning models, which can classify many different diseases for different crops. This kind of models is called fine-grained classification models. With more disease categories, the accuracy performance of the model is more difficult to achieve. This is because the features generally are hiding in the details of the image, such as tiny disease spots. This makes it very difficult for the models to extract the features from the image samples. So, some researchers will apply image segmentation technology to get the disease spots part image before feature extraction. But the image segmentation will lead to additional overhead and make the recognition model not able to be deployed in an end-to-end way.

To solve this problem, we propose a crop leaf disease image recognition method based on bilinear residual networks (named DIR-BiRN). DIR-BiRN incorporates two feature extractors by a bilinear pooling function, and the feature extractors are constructed by residual networks. The two feature extractors extract features, respectively, and then the features are combined by the bilinear pooling function. It can extract features with more accuracy and completeness compared to a single deep learning model in an end-to-end way. Figure 1 shows an example of crop leaf image recognition in which deep learning model is applied. It is essentially a classification problem, and each crop disease corresponds to a category label.

There are four main steps when we applying a deep learning model to crop leaf disease image recognition:

(1) Build training set. In this step, we need to collect leaf images which belong to the different diseases. The images need to be grouped into different disease labels. In order to distinguish healthy leaves from diseased ones, a healthy class must be added in the dataset. Then, we need to preprocess all the images in the dataset. The procedure of image preprocessing involved transforming all the images to uniform size and cropping of all the images, making the square around the leaves, in order to highlight the region of interest (plant leaves) [12]. If the training set is unbalanced, we generally will apply the data augmentation methods to prevent the model overfitting. The common data augmentation methods include image flipping, cropping, rotation, translation, and noise injection. Deep learning models can realize automatic complex feature extraction with sufficient training samples. So they can help us to avoid too many complex preprocessing operations.

(2) Build model. In this step, we need to construct the basic net structure of the recognition model and set the hyperparameter of the model. The model’s feature extraction performance has decisive influence on the image recognition method’s performance. So our main purpose in this step is to build a more effective model net structure as far as possible, which is suitable to extract the crop leaf disease images' fine-grained features.

(3) Training. In this step, we use the training set to iteratively train the model until the model is converged. Then we can apply the converged model to crop leaf disease image recognition.

(4) Recognizing. At last, we can input a new crop leaf disease image to the trained model. Then the model can tell us which kind of disease class should the image belong to.

Our main contributions in this paper are as follows:

(1) We propose the DIR-BiRN. In DIR-BiRN’s build model step, we built a novelty deep learning model net structure, which integrated two residual networks by a bilinear pooling. It can model local pairwise feature interactions in a translationally invariant manner which is very useful for crop leaf disease image recognition.

(2) DIR-BiRN can be deployed in an end-to-end way. It can extract fine-grained crop disease features with more accuracy and completeness without image segmentation.

(3) We experimented DIR-BiRN on the PlantVillage dataset. Compared with the single ResNet-18 model, DIR-BiRN has improvements on the performance of accuracy, recall, precision, and F1-measure.

2. Related Works

There are two kinds of deep learning models which are applied to crop leaf disease image recognition: global model and local model. In this section, we will respectively discuss the recognition methods which are based on the global model and local model.

Global model directly extracts features from the whole crop leaf image. The research based on global models tried to apply more advanced and appropriate model to get better recognition performance. Research [11] tested AlexNet and SqueezeNet on tomato leaf disease images and found that AlexNet had better recognition performance. Research [12] developed a deep convolutional neural network (DCNN) to detect different diseases of 13 crops, with an accuracy of
91%–98%. Research [13] suggested a ResNet model to classify 59 diseases of different crops and achieved an accuracy of 85.22%. Research [15] suggested a new deep convolution neural network (DCNN) for rice leaf disease recognition and achieved 95.48% accuracy. Research [16] proposed an optimized CNN model and achieved 78.8% accuracy when applying it to real-time recognition of 5 apple leaf diseases. Research [17] designed a framework for disease prediction in pearl millet and achieved 98.78% accuracy. It integrated IoT devices and deep learning model (Custom-Net), so it makes the model more suitable for automating disease detection. Research [18] introduced a pretrained CNN model (MobileNetV2) to cassava leaf disease identification, applied data augmentation techniques before training, and achieved 96.75% accuracy on low-quality test images. Similarly, many deep learning models have been applied to crop leaf disease image recognition, such as DenseNet-121 [19], VGG [20], and GoogLeNet [21]. We can find that most of the crop leaf disease research are based on global model. This is because the global model can be deployed in an end-to-end way which is very convenient to be used. It does not need complex preprocessing on the images before training and recognition. But the crop leaf disease image’s features are very different from other images’ features; they usually are hiding in the very fine-grained disease spots, and it is generally hard for the global models to extract this kind of fine-grained features. This reason limited the performance of the recognition methods which are based on the global models to a certain extent.

In order to make up for the shortcoming of the global model, some crop leaf disease image recognition research are based on local models. Local models need to extract the partial image of the disease spots from the whole image by image segmentation technology. Then the partial image is input to recognition model for feature extraction. Research [22] suggested an effective method to locate the disease spots. It proposed a spatial pyramid-oriented encoder-decoder cascade convolution neural network for crop disease leaf segmentation, which consists of a region disease detection network and a region disease segmentation network and can achieve higher disease spots segmentation accuracy. Research [23, 24] realized the tomato leaf disease recognition based on local model. Both of them located the disease spot locations before recognition and then applied a Yolo V3 CNN model and DNN model for diseases and pests recognition, respectively. Research [25] designed a framework that can realize real-time apple leaf disease identification and classification by local model. It configured MASK RCNN to detect the infected regions and then utilized a pretrained CNN model for features extraction and achieved a best accuracy of 96.6%. Research [26] designed a framework for recognition of guava plant diseases. It employed the ΔE color difference image segmentation to segregate the areas infected by the disease, and then color (RGB, HSV) histogram and textural features were applied to extract feature vectors. At last, it achieved an accuracy of 99% in recognizing four guava fruit diseases. Local models will locate the disease spots before training and recognition and can realize more accuracy and complete extraction for fine-grained disease features. Then we can achieve more accurate crop disease recognition. But the location step will lead to extra operations and overheads; this makes it very hard for the local models to be deployed in an end-to-end way.

The related works are summarized in Table 1. Global model methods are very easy to use, because they can be applied or deployed in an end-to-end way. They do not need additional preprocessing or overhead. But, generally, they cannot accurately and completely extract fine-grained disease spots features as local models. Local model methods can extract fine-grained disease features more accurately and
completely, but their image segmentation step will lead to additional operation and overhead. This makes local model methods not able to realize end-to-end deploying.

The most ideal method to realize the crop leaf disease image recognition is that which can accurately and completely extract the fine-grained disease features in an end-to-end way. For this purpose, we introduced a bilinear residual networks model. The concept of bilinear model is proposed by Lin et al. [27] in 2015. It can be applied in an end-to-end way and extract features accurately and completely by two extractors. Based on this concept, we proposed a bilinear residual networks model for crop leaf disease image recognition, which integrated the advantages of global model and local model.

3. Proposed Approach

3.1. Method Overview. In this section, we present our proposed crop leaf disease image recognition method based on bilinear residual networks (DIR-BiRN). The DIR-BiRN’s build training set, training, and recognizing steps are the same as those in other methods which are based on the global models. So we will not discuss these steps in this paper. Our main difference is in the build model step; we built a bilinear residual network for the DIR-BiRN. As shown in Figure 2, there are two residual networks feature extractors in the DIR-BiRN’s model. After an image is input to the model, these two extractors will extract features respectively. The features are integrated to a bilinear vector by a bilinear pooling function. Then the bilinear vector is input to the softmax function for disease classification. This method is related to the two pathways hypothesis of visual processing in the human brain [27]. The hypothesis indicates that human brain uses a pathway to locate object and another pathway to recognize the object [28]. The two extractors in the method can extract more different features from the images. The local features are integrated in a linear
way by a bilinear pooling function. So, this method can model local pairwise feature interactions in a translationally invariant manner [27].

DIR-BiRN is a quadruple, and it consists of the following:

$$\text{DIR} - \text{BiRN} = \{ f_{RNA}, f_{RNB}, f_{BP}, f_c \}.$$  \hspace{1cm} (1)

In this formula, $f_{RNA}$ and $f_{RNB}$ are feature extractors and they are constructed by residual networks. $f_{BP}$ is a bilinear pooling function, and $f_c$ is a classification function. Once image $I_t$ is input to DIR-BiRN, $f_{RNA}$ and $f_{RNB}$ will extract features in each location $L_s$, respectively, which can be represented with functions $f_{RNA}(I_t, L_s)$ and $f_{RNB}(I_t, L_s)$. The DIR-BiRN’s feature extraction for image $I_t$ in location $L'$ can be represented as

$$f_{\text{DIR-BiRN}}(I_t, L_s) = f_{RNA}(I_t, L_s)^T f_{RNB}(I_t, L_s).$$ \hspace{1cm} (2)

Then we can obtain an image representation vector $\text{vec}_t$ by the bilinear pooling function $f_{BP}$. It can be described as

$$\text{vec}_t = f_{BP} \left( \sum_{s \in t} f_{\text{BRM}}(I_t, L_s) \right).$$ \hspace{1cm} (3)

At last, we can get the classification result by inputting $\text{vec}_t$ to the classification function.

3.2. Base-Net. We construct the DIR-BiRN’s feature extractors by residual networks (ResNet). ResNet was proposed by He et al. [25] in 2015. ResNet is essentially a convolutional neural network model, but it introduced residual block in the networks. By the residual block, ResNet can avoid the degradation problem caused by the increase of networks depth. So, it can build more deep networks and realize more accurate feature representation with higher training efficiency [29]. In recent years, many researchers [30, 31] have applied ResNet to crop disease image recognition in their research and achieved better accuracy performance compared to some other models.

The Base-Net structure of the feature extractors in DIR-BiRN is shown in Figure 3. It has 18 layers: 17 convolutional layers and a max-pooling layer. The input images are resized to a size of 224×224 and 3 color channels. In Figure 3, “conv” means convolutional kernel, and the parameter in front of it shows its size. “s” means stride, and “p” means padding. The main characteristic of residual networks is as follows: a shortcut connection is added between each two convolution layers to form a residual block, and many residual blocks are constructed to a residual network. The shortcut connections are represented by red and blue arrows in Figure 3. They used different residual functions:

$$H(x) = f(x) + x.$$ \hspace{1cm} (4)

$$H(x) = f(x) + w(x).$$ \hspace{1cm} (5)

Blue arrows use formula (4), and red arrows use formula (5). In formula (5), $w(x)$ is a convolutional function, to downsampling and ascending dimension for $x$. Their connection structures are shown in Figure 4. The network structure in Figure 4(a) or Figure 4(b) is called a residual block.

Once the crop leaf disease image is input to the net, it will be transformed to a feature map with size of 512×7×7. There are a total of 7×7 = 49 locations in the feature map, and each location has a feature with 512 dimensions. There are two residual networks in DIR-BiRN, so we can obtain two different feature maps for each image. Then, we can input the two feature maps into bilinear pooling function $f_{BP}$ to get feature vector.

3.3. Bilinear Pooling Function. By the two residual networks, we can get two feature extraction functions $f_{RNA}(I_t, L_s)$ and $f_{RNB}(I_t, L_s)$ for each location $L_s$ in image $I_t$. Suppose that image $L_s$ can be represented by a matrix $M_{It,s}$; we can get $M_{It}$ by the following formula:

$$M_{It} = \sum_{s \in t} f_{\text{BRM}}(I_t, L_s) = \sum_{s \in t} f_{RNA}(I_t, L_s)^T f_{RNB}(I_t, L_s).$$ \hspace{1cm} (6)

$M_{It}$ can be reshaped to a vector:

$$\text{vec}_x = \text{reshape} \left( M_{It} \right).$$ \hspace{1cm} (7)

Then, we can get the feature vector of $L_s$ by normalizing $\text{vec}_x$. The normalization is performed by the following formulas:

$$\text{vec}_x = \sin(\text{vec}_x) \sqrt{||\text{vec}_x||},$$ \hspace{1cm} (8)

$$\text{vec}_{I_t} = \frac{\text{vec}_x}{||\text{vec}_x||_2}.$$ \hspace{1cm} (9)

3.4. Classification Function. At last, we can input $\text{vec}_{I_t}$ into classification function $f_c$ to achieve classification result. $f_c$ is defined as follows:

$$f_c(z_j) = \frac{e^{z_j}}{\sum_{i=1}^{n} e^{z_i}}.$$ \hspace{1cm} (9)

In formula (9), $z_j$ means the output value of node $j$, and $n$ means the total number of categories.

4. Case Study

4.1. Experimental Systems. Our approach is tested on the PlantVillage [32] dataset; it contains 54309 labelled images for 14 different crops. There are two research questions (RQ) in our experiments:

(1) RQ 1: Compared with traditional single residual networks model, does DIR-BiRN has performance improvement when classifying many different diseases for the same crop?

(2) RQ 2: Compared with traditional single residual networks model, does DIR-BiRN has performance
improvement when classifying many different diseases for different crops?

To answer RQ 1, we select the crops with more than 3 image categories to experiment. They are apple, corn, grape, potato, and tomato. Each of them has 3 more categories of images in the PlantVillage dataset (including a category of healthy). For RQ 2, we combine 5 kinds of crops’ datasets and then apply the combined dataset to experiment. The basic information of dataset used in the experiment is shown in Table 2, and the image samples in the dataset are shown in Figure 5. In the experiment, we use the standard single 18 layers’ residual networks model (ResNet-18) as baseline method.

4.2. Performance Metric. Many researchers have used accuracy, precision, recall, and F1-measure as metrics to evaluate classification performance. They are defined as follows:

\[
\text{Accuracy} = \frac{\sum_{i=1}^{n} (TP_i/A_i)}{n}.
\]

(10)

\[
\text{Precision} = \frac{\sum_{i=1}^{n} \text{Precision}_i}{n} = \frac{\sum_{i=1}^{n} TP_i/(TP_i + FP_i)}{n}.
\]

(11)

---

**Figure 3:** Base-Net in DIR-BiRN.

**Figure 4:** Residual block. (a) Using identity \( x \) as residual. (b) Using convolutional function as residual.
Recall equals \( \sum_{i=1}^{n} \text{Recall}_i \)

\[
\text{Recall} = \frac{\sum_{i=1}^{n} TP_i}{n} = \frac{\sum_{i=1}^{n} TP_i/(TP_i + FN_i)}{n}
\]

\[
F1 - \text{Measure} = \frac{\sum_{i=1}^{n} 2 \times \text{Precision}_i \times \text{Recall}_i/\text{Precision}_i + \text{Recall}_i}{n}
\]

In formula (10), \( n \) is the total number of categories, \( A_i \) is the total number of images of category \( i \), and \( TP_i \) means the number of images which are correctly classified as category \( i \). In formulas (11) and (12), Precision and \( \text{Recall}_i \) mean the precision and recall performance of category \( i \), respectively. \( FP_i \) is the number of images which are incorrectly classified as category \( i \), and \( FN_i \) means the number of images which should belong to category \( i \) but are incorrectly classified as other categories. However, precision and recall are reciprocal; they are unable to reflect the comprehensive performance of a disease image recognition method. Thus, we also employ the F1-measure to measure the method performance. It is the harmonic average of the precision and recall and is calculated in formula (13).

4.3. Experimental Parameter Settings. In the experiment, the batch size is set as 128, the learning rate is set as \( 1 \times 10^{-3} \), and the Adam optimizer is used. We used the hold-out method as the cross-validation method in the
Figure 6: Continued.
The dataset was randomly divided into training set and testing set, and the training set to testing set ratio was 1 to 4. In order to prevent overfitting, we used L2 regularization in the loss function. In the experiment, we found that the models’ loss function was basically converged after training 100 epochs. So, the epoch is set as 110 in the experiment.

4.4. Experimental Results. We trained our models for a total of 110 epochs on each dataset and recorded the model parameters, training loss, accuracy, precision, recall, and F1-measure of each epoch. Researchers usually use the model parameters which have the optimal accuracy as the final model parameters. In our experiment, we also use this kind of way to obtain the final parameters.

4.4.1. Discussion of RQ 1. The DIR-BiRN model and ResNet-18 model are trained on the apple dataset, corn dataset, grape dataset, potato dataset, and tomato dataset, respectively. The training loss and accuracy in the training process are shown in Figure 6.

The training loss and accuracy in the 110 epochs on the different crop datasets are shown in Figure 6. In these figures, the red line represents training loss, and the black line represents accuracy. From the figures, we can find that the DIR-BiRN model and ResNet-18 model have similar training processes. Both of them have higher training loss and they are lower in the beginning epochs. With the increasing of epochs, the training loss begins to decrease, while accuracy increases. At last, the training loss reached minimum, and accuracy reached maximum. Then we recorded the model parameters with optimal accuracy as the final model parameters. The performance of the model with the final model parameters is called the model's optimal performance. To compare the performance of our approach with that of the single ResNet-18 model, we recorded their worst performance and optimal performance on the different crop datasets. The worst performance means the performance which is obtained by the model parameters with the lowest accuracy. The performance results on different crops are shown in Tables 3–7, respectively.

The worst performance is obtained in the beginning epochs. We can find that DIR-BiRN's worst performance is not as good as the ResNet-18 model from the tables. But, with the increasing of epochs, both models’ performances are improved. After reaching the maximum performance, DIR-BiRN obtained a better optimal performance compared to the ResNet-18 model. Generally, researchers will apply and deploy the recognition models by using the model parameters which obtained optimal performance. In this perspective, we can suggest that DIR-BiRN has better performance than ResNet-18. DIR-BiRN improves by as much as 0.4732 (apple), 0.6494 (corn), 0.0764 (grape), 0.2325 (potato), and 0.0275 (tomato) percentage points in accuracy when compared to ResNet-18 in each case. On the apple dataset, DIR-BiRN achieved 0.36155 percentage points precision improvement, 0.61825 percentage points recall improvement, and 0.48875 percentage points F1-measure improvement compared to ResNet-18. On the corn dataset, DIR-BiRN achieved 1.8168 percentage points precision improvement, 0.7528 percentage points recall improvement, and 1.0917 percentage points F1-measure improvement compared to ResNet-18. On the grape dataset, DIR-BiRN achieved 1.3952 percentage points precision improvement, 0.3603 percentage points recall improvement, and 0.1488
On the potato dataset, DIR-BiRN achieved 0.077 percentage points precision improvement, 0.2106 percentage points recall improvement, and 0.164 percentage points F1-measure improvement compared to ResNet-18. On the tomato dataset, DIR-BiRN achieved 0.4315 percentage points precision improvement, 1.0173 percentage points recall improvement, and 0.7143 percentage points F1-measure improvement compared to ResNet-18.

A confusion matrix is also one of the evaluation indicators of the recognition model. The confusion matrix of each crop on different datasets is shown in Figure 7. The column label of the confusion matrix represents the predicted category, and the row labels of the confusion matrix represent the true category of the predicted image. The value at the diagonal line shows correctly predicted tags. The darker the diagonal line indicates the better the model’s effect [33].

Through Figure 7, we can find that the DIR-BiRN recognized more correct images on 9 kinds of crop diseases’ recognition (2 kinds of apple diseases, 2 kinds of corn diseases, 3 kinds of grape diseases, 1 kind of potato disease, and 1 kind of tomato disease). ResNet-18 recognized more correct images on 3 kinds of crop diseases’ recognition (1 kind of corn disease, 1 kind of grape disease, and 1 kind of potato disease). From this perspective, we can consider that the DIR-BiRN model shows a better recognition performance compared to the ResNet-18 model.

By this, we can answer research question 1: compared with traditional single residual networks model, DIR-BiRN has an improvement on the optimal performance when classifying many different diseases for the same crop. Because researchers usually apply and deploy the recognition models by using the model parameters which obtained optimal performance, we can suggest that DIR-BiRN has a performance improvement.

### 4.4.2. Discussion of RQ 2

For RQ 2, we combined the apple, corn, grape, potato, and tomato datasets in a combined dataset. There are a total of 25 image categories in the combined dataset. Then we ran the DIR-BiRN model and the

| Crop   | Accuracy (%) | Recall (%) | Precision (%) | F1-measure (%) |
|--------|--------------|------------|---------------|----------------|
| Apple  | ResNet-18    | DIR-BiRN   | ResNet-18     | DIR-BiRN       |
| Worst  | 81.7035      | 70.1893    | 89.01825      | 63.2663        |
| Optimal| 97.3186      | 97.7918    | 97.53775      | 98.1560        |

| Crop   | Accuracy (%) | Recall (%) | Precision (%) | F1-measure (%) |
|--------|--------------|------------|---------------|----------------|
| Corn   | ResNet-18    | DIR-BiRN   | ResNet-18     | DIR-BiRN       |
| Worst  | 70.7792      | 65.2078    | 61.3640       | 35.9643        |
| Optimal| 93.8961      | 94.5455    | 90.7075       | 91.4603        |

| Crop   | Accuracy (%) | Recall (%) | Precision (%) | F1-measure (%) |
|--------|--------------|------------|---------------|----------------|
| Grape  | ResNet-18    | DIR-BiRN   | ResNet-18     | DIR-BiRN       |
| Worst  | 79.6798      | 71.3054    | 80.0173       | 73.7798        |
| Optimal| 94.9507      | 95.0271    | 95.5995       | 95.9598        |

| Crop   | Accuracy (%) | Recall (%) | Precision (%) | F1-measure (%) |
|--------|--------------|------------|---------------|----------------|
| Potato | ResNet-18    | DIR-BiRN   | ResNet-18     | DIR-BiRN       |
| Worst  | 61.6279      | 63.0233    | 44.2653       | 45.3407        |
| Optimal| 95.8140      | 96.0465    | 90.2647       | 90.4753        |

| Crop   | Accuracy (%) | Recall (%) | Precision (%) | F1-measure (%) |
|--------|--------------|------------|---------------|----------------|
| Tomato | ResNet-18    | DIR-BiRN   | ResNet-18     | DIR-BiRN       |
| Worst  | 78.5518      | 74.3392    | 70.6608       | 69.0021        |
| Optimal| 96.8337      | 96.8612    | 95.0058       | 96.0231        |

| Crop   | Accuracy (%) | Recall (%) | Precision (%) | F1-measure (%) |
|--------|--------------|------------|---------------|----------------|
| Apple  | ResNet-18    | DIR-BiRN   | ResNet-18     | DIR-BiRN       |
| Worst  | 81.7035      | 70.1893    | 89.01825      | 63.2663        |
| Optimal| 97.3186      | 97.7918    | 97.53775      | 98.1560        |

| Crop   | Accuracy (%) | Recall (%) | Precision (%) | F1-measure (%) |
|--------|--------------|------------|---------------|----------------|
| Corn   | ResNet-18    | DIR-BiRN   | ResNet-18     | DIR-BiRN       |
| Worst  | 70.7792      | 65.2078    | 61.3640       | 35.9643        |
| Optimal| 93.8961      | 94.5455    | 90.7075       | 91.4603        |

| Crop   | Accuracy (%) | Recall (%) | Precision (%) | F1-measure (%) |
|--------|--------------|------------|---------------|----------------|
| Grape  | ResNet-18    | DIR-BiRN   | ResNet-18     | DIR-BiRN       |
| Worst  | 79.6798      | 71.3054    | 80.0173       | 73.7798        |
| Optimal| 94.9507      | 95.0271    | 95.5995       | 95.9598        |

| Crop   | Accuracy (%) | Recall (%) | Precision (%) | F1-measure (%) |
|--------|--------------|------------|---------------|----------------|
| Potato | ResNet-18    | DIR-BiRN   | ResNet-18     | DIR-BiRN       |
| Worst  | 61.6279      | 63.0233    | 44.2653       | 45.3407        |
| Optimal| 95.8140      | 96.0465    | 90.2647       | 90.4753        |

| Crop   | Accuracy (%) | Recall (%) | Precision (%) | F1-measure (%) |
|--------|--------------|------------|---------------|----------------|
| Tomato | ResNet-18    | DIR-BiRN   | ResNet-18     | DIR-BiRN       |
| Worst  | 78.5518      | 74.3392    | 70.6608       | 69.0021        |
| Optimal| 96.8337      | 96.8612    | 95.0058       | 96.0231        |
### Figure 7: Continued.
The ResNet-18 model on the combined dataset. The training loss and accuracy in 110 epochs are shown in Figure 8.

The training process is similar to that in RQ 1. Both models have higher training loss and lower accuracy. With the increasing of epochs, the training loss decreased to a minimum, and the accuracy increased to a maximum. The worst and optimal performances of DIR-BiRN and ResNet-18 are shown in Table 8.

Both models’ worst performance is obtained in the first epoch, and the optimal performance is obtained in the last few epochs. Compared with ResNet-18, DIR-BiRN also has a better optimal performance on the combined dataset. It achieved 0.0319 percentage points accuracy improvement, 1.0534 percentage points precision improvement, 1.6804 percentage points recall improvement, and 1.3804 percentage points F1-measure improvement compared to ResNet-18. Because researchers usually apply and deploy the recognition models by using the model parameters which obtained optimal performance, we can suggest that, compared with traditional single residual networks model, DIR-BiRN can achieve performance improvement when classifying many different diseases for different crops.
4.5. Computational Complexity Evaluation. We realized the models by PyTorch and performed them with Intel i7-10850 CPU, NVIDIA Quadro T1000 GPU and 8 GB RAM. In the experiment, we record the average time consumption of training each epoch and recognition of each image and we also record the models' size on each dataset. The training time, recognition time, and model size on different datasets are shown in Table 9. We can find that DIR-BiRN needs to spend more time in training and recognition when compared to ResNet-18 in each dataset. The training time of DIR-BiRN is averagely increased by 3.24 percent compared to ResNet-18, and the recognition time is averagely increased by 9.85 percent. DIR-BiRN also has bigger model size; its model size is averagely increased by 19.49 percent compared to ResNet-18. From this perspective, we can think that DIR-BiRN has higher time and storage consumption than the ResNet-18 model.

5. Discussion

This paper proposed a crop leaf disease image recognition method based on bilinear residual networks (named DIR-BiRN). According to our experiments, DIR-BiRN has better recognition performance (accuracy, recall, precision, and F1-measure) than the traditional ResNet-18 model. This is because DIR-BiRN's bilinear model can extract more fine-grained features in the crop leaf disease images, and these fine-grained features are very useful to crop disease recognition.

There are still some limitations in our methodology. At first, the DIR-BiRN's bilinear model is more complex than the single model, so it has higher time and storage consumption than the single-model method. Second, our experiments are only performed on the leaf disease images...
which have simple background. If our method is applied to the leaf disease images which have complex background, it is very possible to lead to recognition performance degradation. At last, we only tested our method on five kinds of crops; the recognition performance of DIR-BiRN on other crops still needs to be verified.

6. Conclusions

To address the fine-grained classification problems in the crop leaf disease image recognition, we proposed a method based on bilinear residual networks (named DIR-BiRN). It integrated two 18-layer residual networks feature extractors by a bilinear way. It can extract features more accurately and completely than the single residual networks model, while deploying and applying the model in an end-to-end way. So it has the advantages of both global model and local model. We tested DIR-BiRN on the PlantVillage dataset. In our experiment, DIR-BiRN showed a better performance (accuracy, recall, precision, and F1-measure) than the single residual networks model. From the confusion matrix results, we can also find that DIR-BiRN shows better recognition performance on some crop diseases which have very small disease spots (apple scab, apple black rot, grape black rot, etc.). This experimental result approved that our bilinear residual networks can extract more fine-grained crop disease features in the images, making our method able to realize more accurate disease recognition.

In our future works, we will try to integrate more different feature extractors to get a better recognition performance and also test our method on more datasets.

Data Availability

All data included in this study are available upon request by contact with the corresponding author.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant no. 32101611, the Youth Project of Basic Research Program of Yunnan Province under Grant no. 202101AU007002, the Scientific research fund project of Yunnan Provincial Department of Education under Grant no. 2020J0239, the Major Project of Science and Technology of Yunnan Province under Grants nos. 202002AE090010 and 202002AD080002, the Open Research Program of State Key Laboratory for Conservation and Utilization of Bio-Resource in Yunnan under Grant no. GZKF2021009, and the Open Fund Project of Yunnan Province Software Engineering Key Laboratory under Grant no. 2020SE501.

References

[1] M. G. Du and S. W. Zhang, “Crop disease leaf image segmentation based on genetic algorithm and maximum entropy,” *Applied Mechanics and Materials*, vol. 713-715, pp. 1670–1674, 2015.
[2] J. Mandal, S. Satapathy, M. Kumar Sanyal, P. Sarkar, and A. Mukhopadhyay, “Identifying two of tomatoes leaf viruses using support vector machine,” *Advances in Intelligent Systems and Computing*, vol. 339, pp. 771–782, 2015.
[3] S. Sindhuja, M. Ashish, M. Joe Mari, and R. Ehsan, “Visible-near infrared spectroscopy for detection of Huanglongbing in citrus orchards,” *Computers and Electronics in Agriculture*, vol. 77, no. 2, pp. 127–134, 2011.
[4] M. Wahabzada, A. K. Mahlein, C. Bauckhage, U. Steiner, E. C. Oerke, and K. Kersting, “Metro maps of plant disease dynamics—automated mining of differences using hyperspectral images,” *PLoS ONE*, vol. 10, no. 1, pp. 1–20, 2015.
[5] L. H. R. Deny, R. Q. Fernando, and G. J. Julian, “Integrating SOMs and a bayesian classifier for segmenting diseased plants in uncontrolled environments,” *The Scientific World Journal*, vol. 2014, Article ID 214674, 13 pages, November, 2014.
[6] S. P. Mohanty, D. P. Hughes, and M. Salathé, “Using deep learning for image-based plant disease detection,” *Frontiers of Plant Science*, vol. 7, p. 1419, September, 2016.
[7] K.-Y. Huang, “Application of artificial neural network for detecting Phalaenopsis seedling diseases using color and texture features,” *Computers and Electronics in Agriculture*, vol. 57, no. 1, pp. 3–11, 2007.
[8] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” *Advances in Neural Information Processing Systems*, vol. 25, pp. 1097–1105, 2012.
[9] J. Amara, B. Bouaziz, and A. Algergawy, “A deep learning-based approach for banana leaf diseases classification,” in *Proceedings of the 2007 Datenbanksysteme für Business, Technologie und Web Workshop*, pp. 6–10, Stuttgart, Germany, March 2017.
[10] Q. Wu, Y. Chen, and J. Meng, “DCGAN-based data augmentation for tomato leaf disease identification,” *IEEE Access*, vol. 8, pp. 98716–98728, 2020.
[11] H. Durmus, E. O. Gunes, and M. Kirici, “Disease detection on the leaves of the tomato plants by using deep learning,” in *Proceedings of the 6th International Conference on Agro-Geoinformatics*, pp. 1–5, Fairfax, VA, USA, August 2017.
[12] S. Srdjan, A. Marko, A. Andras, C. Dubravko, and D. Stefanovic, “Deep neural networks based recognition of plant diseases by leaf image classification,” *Computational Intelligence and Neuroscience*, vol. 2016, Article ID 3289801, 11 pages, June, 2016.
[13] W.-J. Hu, J. Fan, Y.-X. Du, B.-S. Li, N. Xiong, and E. Bekkering, “MDPC-ResNet: an agricultural IoT system to accurately recognize crop diseases–ResNet: an agricultural IoT system to accurately recognize crop diseases,” *IEEE Access*, vol. 8, pp. 115287–115298, 2020.
[14] M. V. Appalanaidu and G. Kumaravelan, “A review of machine learning approaches in plant leaf disease detection and classification,” in *Proceedings of the Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks*, pp. 716–724, Tirunelveli, India, February 2021.
[15] Y. Lu, S. Yi, and N. Zeng, “Identification of rice diseases using deep convolutional neural networks,” *Neurocomputing*, vol. 267, pp. 378–384, 2017.
[16] P. Jiang, Y. Chen, and B. Liu, “Real-time detection of apple leaf diseases using deep learning approach based on improved convolutional neural networks,” IEEE Access, vol. 7, pp. 59069–59080, 2019.

[17] N. Kundu, G. Rani, V. S. Dhaka et al., “IoT and interpretable machine learning based framework for disease prediction in pearl millet,” Sensors, vol. 21, no. 16, p. 5386, August. 2021.

[18] O. O. Abayomi-Alli, R. Damaševičius, S. Misra, and R. Maskeliunas, “Cassava disease recognition from low-quality images using enhanced data augmentation model and deep learning,” Expert Systems, vol. 38, no. 7, Article ID 12746, June. 2021.

[19] Y. Zhong and M. Zhao, “Research on deep learning in apple leaf disease recognition,” Computers and Electronics in Agriculture, vol. 168, Article ID 1051469, January. 2019.

[20] S. Coulibaly, B. Kamsu-Foguem, and D. Kamissoko, “Deep neural networks with transfer learning in millet crop images,” Computers in Industry, vol. 108, pp. 115–120, 2019.

[21] X. Zhang, Y. Qiao, F. Meng, C. Fan, and M. Zhang, “Identification of maize leaf diseases using improved deep convolutional neural networks,” IEEE Access, vol. 6, pp. 30370–30377, 2018.

[22] Y. Yuan, Z. Xu, and G. Lu, “SPEDCCNN: spatial pyramid-oriented encoder-decoder cascade convolution neural network for crop disease leaf segmentation,” IEEE Access, vol. 9, pp. 14849–14866, 2021.

[23] J. Liu and X. Wang, “Tomato diseases and pests detection based on improved Yolo V3 convolutional neural network,” Frontiers of Plant Science, vol. 11, no. Jun, p. 898, 2020.

[24] A. F. Fuentes, S. Yoon, J. Lee, and D. S Park, “High-performance deep neural network-based tomato plant diseases and pests diagnosis system with refinement filter bank,” Frontiers of Plant Science, vol. 9, p. 1162, August. 2018.

[25] Z. Rehman, M. A. Khan, F. Ahmed et al., “Recognizing apple leaf diseases using a novel parallel real-time processing framework based on MASK RCNN and transfer learning: an application for smart agriculture,” IET Image Processing, vol. 15, no. 10, pp. 2157–2168, 2021.

[26] A. Almadhor, H. T. Rauf, M. I. U. Lali, R. Damaševičius, B. Alouffi, and A. Alharbi, “AI-driven framework for recognition of guava plant diseases through machine learning from DSLR camera sensor based high resolution imagery,” Sensors, vol. 21, no. 11, p. 3830, June. 2021.

[27] T. Y. Lin, A. Roychowdhury, and S. Maji, “Bilinear CNN models for fine-grained visual recognition,” in Proceedings of the 2015 IEEE International Conference on Computer Vision, pp. 1449–1457, Santiago, Chile, December 2015.

[28] M. A. Goodale and A. D. Milner, “Separate visual pathways for perception and action,” Trends in Neurosciences, vol. 15, no. 1, pp. 20–25, 1992.

[29] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778, Las Vegas, NV, USA, June 2016.

[30] T. Fang, P. Chen, J. Zhang, and B. Wang, “Crop leaf disease grade identification based on an improved convolutional neural network,” Journal of Electronic Imaging, vol. 29, no. 1, Article ID 013004, January. 2020.

[31] M. Prabhakar, R. Purushothaman, and D. P. Awasthi, “Deep learning based assessment of disease severity for early blight in tomato crop,” Multimedia Tools and Applications, vol. 79, no. 39-40, pp. 28773–28784, 2020.

[32] D. P. Hughes and M. Salathe, “An open access repository of images on plant health to enable the development of mobile disease diagnostics,” 2016, https://arxiv.org/abs/1511.08060.

[33] X. Wenchao and Y. Zhi, “Research on strawberry disease diagnosis based on improved residual network recognition model,” Mathematical Problems in Engineering, vol. 2022, pp. 1–13, Article ID 6431942, 2022.