

## Towards automatic field plant disease recognition

Penghui Gui<sup>a,1</sup>, Wenjie Dang<sup>b,1</sup>, Feiyu Zhu<sup>a,\*</sup>, Qijun Zhao<sup>a,b,\*</sup>

<sup>a</sup> College of Computer Science, Sichuan University, Chengdu, China

<sup>b</sup> National Key Laboratory of Fundamental Science on Synthetic Vision, Sichuan University, Chengdu, China



### ARTICLE INFO

#### Keywords:

Convolutional neural networks  
Plant disease recognition  
In-the-field conditions  
Data augmentation  
Discriminative feature learning

### ABSTRACT

Plant disease is a significant threat to food security and subsistence farmers. Despite the rapid development of automatic recognition of plant disease under controlled laboratory conditions since the employment of deep learning technology, it is still quite challenging to distinguish plant disease under uncontrolled field conditions. In this paper, based on a backbone convolutional neural network (CNN), we propose an improved CNN model towards field plant disease recognition (FPDR) by exploring the potential and generalization capabilities of the CNN model. To train the model, we propose background replacing to make the model more robust to background distraction, and leaf resizing to deal with inconsistent size and location of disease symptoms. Both background replacing and leaf resizing are used as data augmentation methods of the improved model. To further enhance the feature discriminativeness, we propose channel orthogonal constraint to improve the ability of feature to distinguish similar categories, and utilize species information as an auxiliary species classification task. In addition, we collect 665 plant disease images under field conditions, namely Field-PlantVillage (Field-PV) to remedy for lack of in-the-field images. The Field-PV is only used as an independent test set to evaluate the performance of the method applied to FPDR. Our improved CNN model improves the FPDR accuracy on Field-PV from 41.81% to 72.03%, though only the PlantVillage dataset is used for training. Experimental result on the PlantVillage achieves the state of the art performance (99.84%). Code and data are available at <https://github.com/PatrickGui/FPDR/tree/master>.

### 1. Introduction

Plant disease is not only related to global food safety, but also closely linked to subsistence farmers whose livelihoods depend on healthy plants (Abade et al., 2021). An average of 26% of the worldwide plant production is lost each year due to pre-harvest pests and pathogens (Oerke, 2006). The pests and pathogens that cause plant diseases include bacteria, fungi, oomycetes, viruses, nematodes, and insect (Moore et al., 2000). Symptoms of diseased plant leaves caused by different pathogens are different. Indeed, even the same pathogen causes various plant disease symptoms among different plant species. However, some disease symptoms are very similar to each other, and even nutritional deficiencies and pathogens can produce symptoms similar to those of some diseases (Barbedo, 2016; Bischoff et al., 2021; Moore et al., 2000). More importantly, most farmers have limited resources to deal with plant disease outbreaks because it is hard for them to immediately confirm the disease. Therefore, an automatic plant disease recognition system is

significant for the farmers to determine whether a plant is healthy or which disease the plant has.

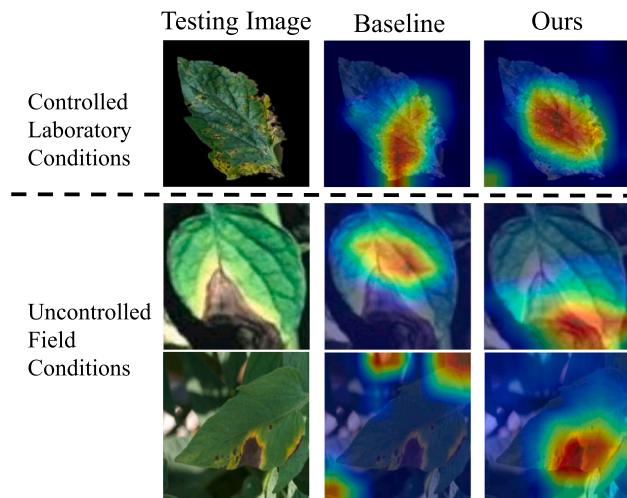
Many kinds of research have been carried out to diagnose a plant's disease through a digital leaf image (Abade et al., 2021; Barbedo, 2016; Bischoff et al., 2021). With the development of computer vision and artificial intelligence (AI), many deep learning methods, in particular Convolutional Neural Networks (CNNs), have found applications in plant disease recognition. A variety of deep or shallow CNNs have been devised to complete recognition tasks, and they outperform human beings on a public dataset, PlantVillage (Abade et al., 2019; Brahimi et al., 2018; Chen et al., 2020b; Li et al., 2020; Mohanty et al., 2016; Nazki et al., 2020; Too et al., 2019). It is worth noting that images in PlantVillage are taken under controlled laboratory conditions (Hughes et al., 2015). Consequently, despite the superior performance of CNNs on PlantVillage, it is still quite challenging to automatically recognize plant diseases under uncontrolled field conditions. As shown in Fig. 1, when applying a model that works well for images collected in the laboratory

\* Corresponding authors at: College of Computer Science, Sichuan University, Chengdu, China (Q. Zhao).

E-mail addresses: [penghui@stu.scu.edu.cn](mailto:penghui@stu.scu.edu.cn) (P. Gui), [feiyuz@scu.edu.cn](mailto:feiyuz@scu.edu.cn) (F. Zhu), [qjzhao@scu.edu.cn](mailto:qjzhao@scu.edu.cn) (Q. Zhao).

URL: <http://www.scubrl.org> (Q. Zhao).

<sup>1</sup> Penghui Gui and Wenjie Dang contribute equally to the article.



**Fig. 1.** Example diseased plant leaf images of the same category under different conditions (left) and the corresponding visualization results of diseased regions focused on by the baseline model and proposed improved model (right). Compared with laboratory conditions, the diseased leaf images in field conditions have not only more obvious symptom variation, but also cluttered background and unconstrained imaging conditions. The Class Activation Map (CAM) visualization results show where the model focuses on. When applying the baseline model that works well for images collected in the laboratory to images collected in the field, the visualization results show that the model is distracted. In contrast, our model can still effectively see the primary symptom regions. The pictures in uncontrolled field conditions are from the website of Wikimedia Commons (Commons, 2020).

to images collected in the field, the recognition performance degrades substantially because of symptom variation, cluttered background, and unconstrained imaging conditions (e.g., varying poses and illuminations).

Few researchers have discussed automatic plant disease recognition under field conditions in recent years (Boulet et al., 2019; Chen et al., 2020a; Ferentinos, 2018; Fuentes et al., 2020; Mohanty et al., 2016). Mohanty et al. (2016) first mentioned two limitations of their method that need to be addressed to accomplish field plant disease recognition. One is that plant images in the PlantVillage dataset are structured and simple, and thus can not reveal the challenges in FPDR. The other is that the method is currently constrained to the classification of single leaves, facing up, on a homogeneous background. Mohanty et al. divided the images in PlantVillage into a training set and a test set at a 1 : 4 ratio and achieved a high accuracy of 99.35%. However, the recognition accuracy was significantly reduced to 31.4% when they used images under uncontrolled conditions as the test set. Ferentinos (2018) used a dataset containing 87,848 images, of which 37.3% were acquired under field cultivation conditions, and the rest under PlantVillage-like controlled laboratory conditions. After training a CNN model with the dataset collected under controlled laboratory conditions, they tested the model on the dataset collected under uncontrolled conditions, resulting in an accuracy of 33.27%, which was lower than that under laboratory conditions. Barbedo (2016) discussed some intrinsic and extrinsic factors that affect the automatic recognition of plant diseases. Extrinsic factors include image background and acquisition conditions. Intrinsic factors include symptom segmentation, symptom variation, multiple simultaneous disorders, and different disorders with similar symptoms. These factors play an important role and have a significant impact on the effectiveness of FPDR.

In general, there are three main difficulties when developing FPDR methods. 1) Most plant images available in existing datasets have simple backgrounds, but the background in field plant leaf images might be quite complex due to the different growth scenarios. 2) The size and location of symptoms of the same plant disease could vary largely across

different leaf images because of the regional divergence and different development stages (Raja et al., 2018). 3) The symptoms of some different plant diseases could be very similar on different leaf images, even on leaf images of different plant species (Barbedo, 2016).

Some attempts have been made to recognize plant diseases on in-the-field (in-the-wild) leaf images (Argüeso et al., 2020; Arsenovic et al., 2019; Boulet et al., 2019; Picon et al., 2019a; Picon et al., 2019b; Zhao et al., 2020). Boulet et al. (2019) surveyed 19 studies that relied on CNNs to automatically identify plant diseases and provided guidelines to “maximize the potential of CNNs deployed in real-world applications”. The guidelines described the best practices for field diseased plant image acquisition, dataset preparation (image preprocessing and augmentation), and training and evaluation phases. To enable the CNNs to handle the real-field complexities of plant disease recognition, an intuitive method is to add in-the-field leaf images to the training set. Some works collected the images under field conditions as training and test set and achieved success by using well-known CNNs architectures (Picon et al., 2019a; Picon et al., 2019b; Zhao et al., 2020). However, collecting many diseased field leaf images including different expressions of symptoms for a given category is very complicated in an agricultural context. Therefore, it is also important to improve the performance of FPDR from the perspective of improving the deep learning algorithms instead of enlarging the set of field images. Arsenovic et al. (2019) not only discussed the limitations of existing plant disease recognition models but also used algorithms of traditional data augmentation methods and generative adversarial networks (GANs) to overcome the obstacles of complex background and various conditions in FPDR (Goodfellow et al., 2014). But the improvement produced by their method is very slight, i.e., from 86.17% to 86.88%. Argüeso et al. (2020) used algorithms of distance metric few-shot learning approach based on the Triplet loss to identify new plant disease categories with very few annotated training images. Their experimental results on PlantVillage show the possibility of FPDR based on few leaf images taken in the field conditions.

This paper makes attempts towards automatic field plant disease recognition with an improved CNN model to address the three main difficulties in FPDR from the perspective of data augmentation and discriminative feature learning (Lin et al., 2015; Wang et al., 2020; Yun et al., 2019). Specifically, considering the complex background variations in FPDR, this paper proposes a data augmentation method of background replacing to make the CNN model learn a representation invariant to background changes. Considering the large variation in size and location of plant symptoms within the same class, we use a data augmentation method of leaf resizing to improve the diversity of symptom variation in the training data. To address the difficulty that symptoms are very similar between some different classes, we propose channel orthogonal constraint and species classification task to further enhance the feature discriminativeness. The channel orthogonal constraint improves the ability of the CNN model to distinguish between similar categories, and the auxiliary species classification task introduces species information to supervise the feature learning. To evaluate the effectiveness of proposed method, this paper uses some basic data augmentation methods and a backbone of ResNet50 as the baseline model (He et al., 2016). The basic data augmentation methods consist of the geometrical and color-space transformation of the original images (Shorten and Khoshgoftaar, 2019). What's more, we collect 665 plant disease images under field conditions to extend the PlantVillage dataset. We call this set of in-the-field plant images as Field-PlantVillage (Field-PV). To sum up, our contribution in this paper is as follows:

- We propose an improved CNN model without requiring any additional training data for FPDR via data augmentation and discriminative feature learning.
- We collect a set of in-the-field plant disease images under field conditions as an independent test set named Field-PlantVillage (Field-PV).

- Evaluation results on Field-PlantVillage show that our proposed improved CNN model can significantly improve the accuracy of FPDR.

## 2. Materials and methods

### 2.1. Materials

PlantVillage is an open and free dataset containing 54,306 images of 38 different categories of diseased and healthy plant leaves (Hughes et al., 2015). Every sample in this dataset has a corresponding segmented version with a black background and gray-scale version. We name these three versions of sample as Color, Segmented, and Grayscale, as shown in Fig. 2.

Note that the three versions of images from PlantVillage are obtained in the controlled laboratory, which is different from the real field. We use the Segmented images to synthesize a new dataset named Syn-PlantVillage (Syn-PV). The images in Syn-PV have a uniform complex background, which is synthesized by background replacing. The background used in Syn-PV does not appear in the training phase, and the images in Syn-PV are only used for testing.

Field-PlantVillage (Field-PV), a supplementary to PlantVillage, is a set of in-the-field plant images that can be used as an independent test set. These in-the-field images are obtained from trusted online sources, including Google Image Search and Bing Image Search. We first search the images corresponding to the 38 categories with different sizes or multiple leaves. Then we crop the images to ensure that one image only contains one leaf. After this operation, we finally acquire 665 images of 38 categories with at least 8 images in each category.

Sample images from these three datasets are displayed in Fig. 2. It can be clearly seen that Field-PV is more representative of the real field situation. The images in Field-PV have various complex backgrounds, variable size and location of symptoms, similar symptoms of different diseases, and differences due to illumination, view angles, and geography. Compared with the very challenging Field-PV, images in Syn-PV suffer from only one of the difficulties in FPDR, i.e., complex background. Field-PV and Syn-PV datasets are available at <https://drive.google.com/file/d/15aEoop-GT4OJsm94T91tHy-1SYsWv1Lp/view?usp=sharing>.

### 2.2. Method overview

The framework of the proposed method is shown in Fig. 3. There are four parts to the framework: backbone model, data augmentation, channel orthogonal constraint, and species classification task. It is worth noting that due to the scarcity of in-the-field images of diseased plants, our method only uses raw data under controlled laboratory conditions to train the model, which is then directly applied to FPDR. In the rest of this section, we introduce these four parts in detail.

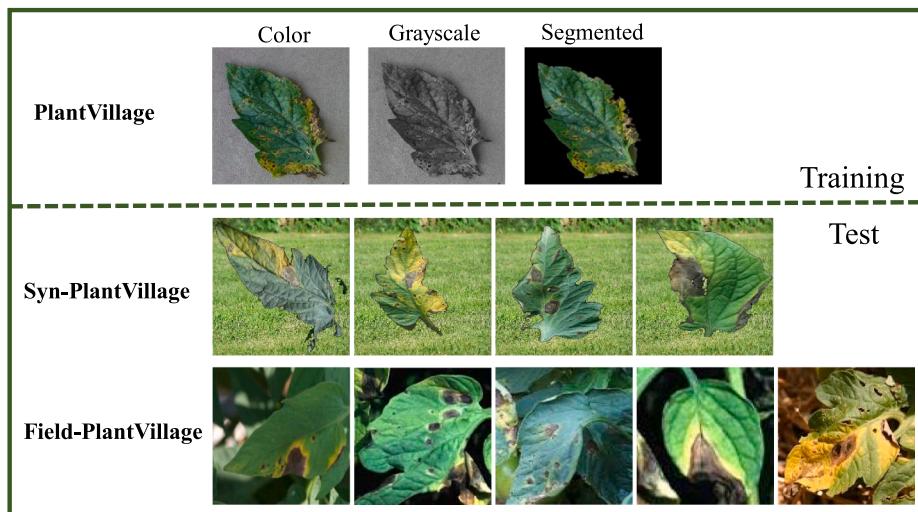
### 2.3. Backbone model

Most existing methods use deep CNN to solve the problem of plant disease recognition (Abade et al., 2019; Brahimi et al., 2018; Chen et al., 2020b; Mohanty et al., 2016). In this paper, we use a deep CNN of ResNet as the backbone model (He et al., 2016). The backbone will extract four convolutional features  $[M_1, M_2, M_3, M_4]$  from the input image after four residual learning blocks. Each convolutional features consists of  $C$  pieces of  $H \times W$  feature maps, i.e.,  $M_i \in R^{C \times H \times W}$ ,  $i = 1, 2, 3, 4$ . As the index increases, the size of the convolutional features  $M_i$  decreases smaller, the number of channels increases larger, and its semantic information becomes richer. So the feature map  $M_4$  has the richest information and is applied for plant disease classification.

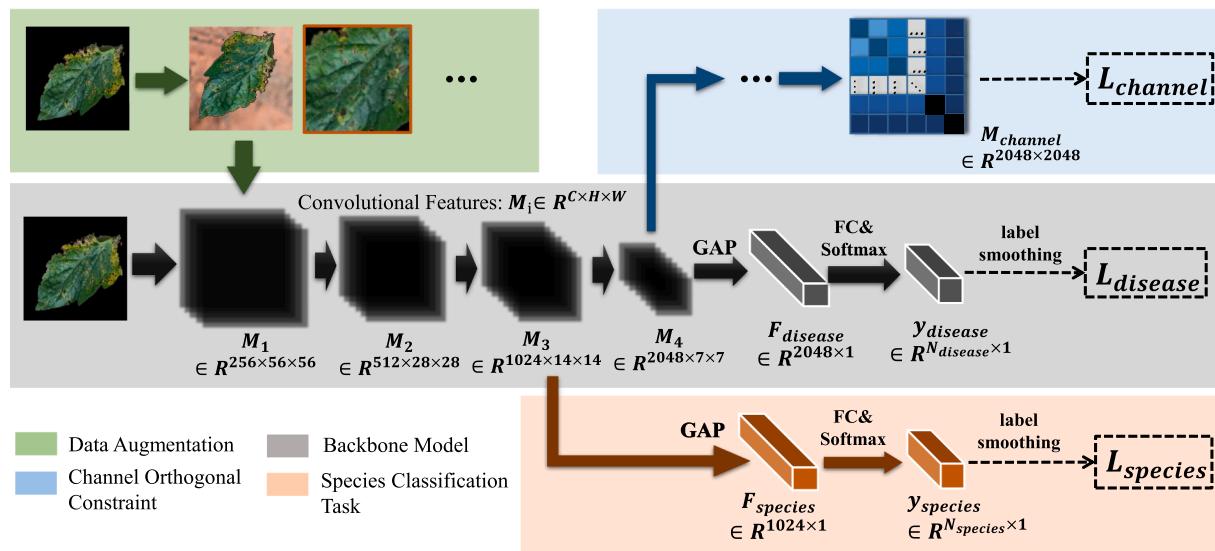
### 2.4. Data augmentation

#### 2.4.1. Background replacing

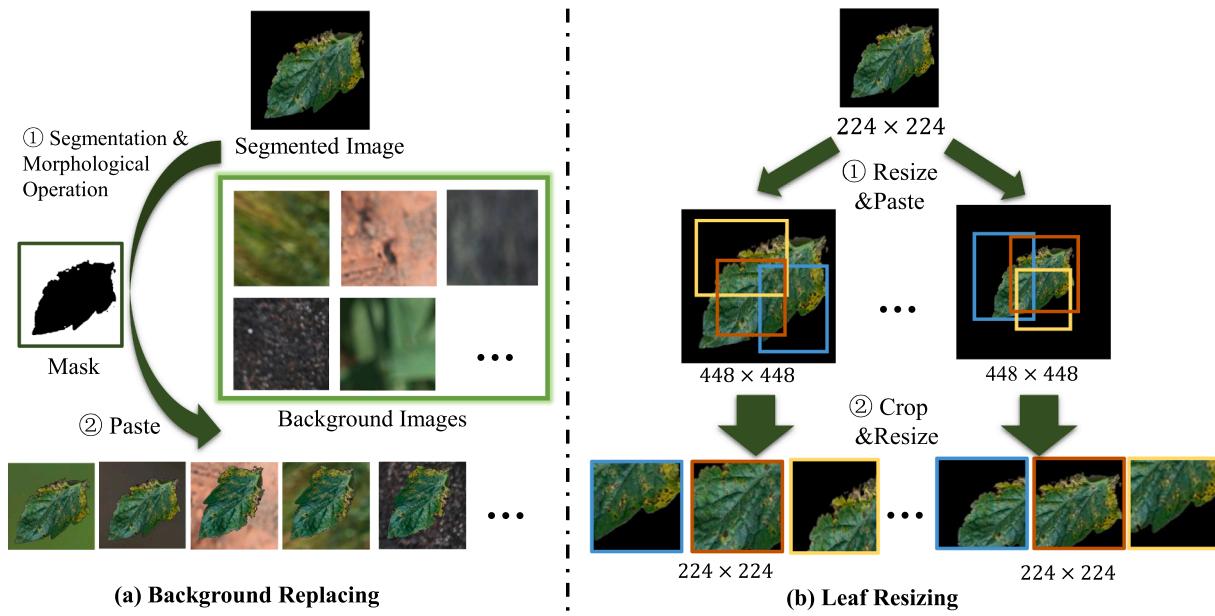
Images with various complex backgrounds in training data are useful for CNN models to learn a representation invariant to background changes. Being aware of this, we propose background replacing to synthesize diseased plant images with complex backgrounds. Fig. 4 depicts the steps in this method. First, we remove the black background in the Segmented image and get a binary mask of the leaf. Specifically, we obtain the binary mask (a value of 0 for black corresponds to the leaf, a value of 255 for white corresponds to the background) based on a pre-specified threshold (i.e., [180, 255, 23]) in the HSV color space. Note that some dark disease regions could be mistakenly treated as black background in segmentation, resulting in small white dots in the leaf mask. We thus apply a morphological open operation consisting of erosion and dilation in sequence to get the final mask. Finally, the leaf in the image is pasted onto different background images to obtain different mixed images with the help of the mask. We select three representative types of



**Fig. 2.** Sample images of tomato early blight disease in three datasets. Three versions of images in the PlantVillage dataset are taken in a controlled condition. Images in the Syn-PlantVillage are obtained by pasting the Segmented version of the images into a uniform complex background. Images in the Syn-PlantVillage and Field-PlantVillage are only used as test sets to evaluate the performance of FPDR.



**Fig. 3.** Overview of the proposed CNN model for FPDR. Data Augmentation, Channel Orthogonal Constraint, Species Classification Task, and three loss functions are employed for training the model.



**Fig. 4.** Illustration of Background Replacing and leaf resizing data augmentation methods. Both of them can be used together with other basic data augmentation methods during the training phase.

background images: yellow images simulating land background, green images simulating plant background, and black images simulating general background (see Fig. 4). In the training phase, the mixed images obtained by this method will greatly enrich the background diversity in training data. Models trained by images with both simple and various complex backgrounds can focus more on disease region, leaf shape, and texture instead of irrelevant features.

#### 2.4.2. Leaf resizing

To simulate varying symptom sizes and locations, we reduce or enlarge the original image to different sizes and then crop a partial area on the diseased leaf. As shown in Fig. 4, the original leaf images with the black background are randomly resized by 0.5~2.0 times and then pasted on a black background with a fixed size (e.g., 448 × 448). Next, we crop a partial area around the leaves. The location, length, and width

of the rectangular cropping boxes are randomly determined. Finally, the cropped images are resized to fit the CNN model. Like background replacing, this method is an online data augmentation methods that can be used together with basic data augmentation methods.

#### 2.4.3. Basic data augmentation

Considering that deep CNN models are easy to over-fit and ignore the strong color variations of disease symptoms, a bunch of basic data augmentation methods are used to increase the diversity of training data. We use some common geometric transformations and color space transformations as the basic data augmentation methods (Shorten and Khoshgoftaar, 2019). Geometric transformations, including horizontal flip, vertical flip, and rotation, enable the model to overcome positional and directional biases. Color space transformations include grayscale, brightness, contrast, saturation, and hue changes. Via grayscale

transformation, the model learns more knowledge about the shape of the diseased region versus the color of the region. Other color space transformations enhance the model to adapt to different color variations (intensity values changing and illumination variation).

### 2.5. Channel orthogonal constraint

As shown in Fig. 5, some plant disease categories have very similar symptoms to each other (e.g., Tomato Target Spot and Tomato Bacterial Spot). To distinguish similar symptoms among different plant diseases, more discriminative features which can capture subtle differences of these symptoms are needed. We obtain discriminative features by making the extracted features as distinguishable as possible. To this end, we propose a channel orthogonal constraint loss function to suppress the similarity between different channels of extracted features.

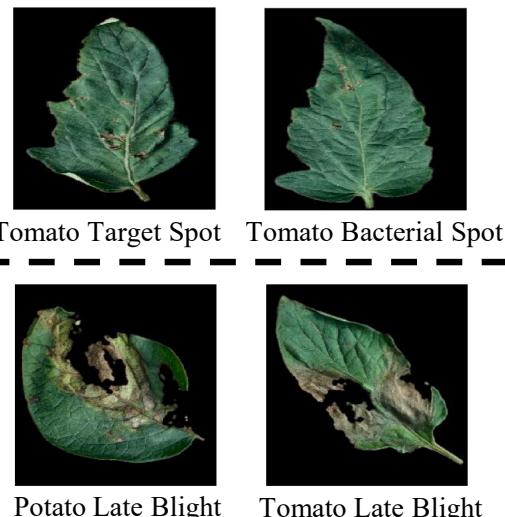
As shown in Fig. 6, given feature maps  $M_4$  extracted by the last convolutional layer in the model, we reshape the feature map of each channel into a feature vector to obtain the feature matrix  $X$ , and regard the feature matrix  $X$  as a set of feature vectors  $x_k \in R^{HW}$  ( $k$  indicates the index of the channels in  $M_4$ ,  $k = 1, 2, \dots, C$ , and  $C$  is the total number of channels). Then we multiply the matrix  $X$  by its transpose matrix to obtain a  $C \times C$  square matrix, where the element in the  $i^{\text{th}}$  and row  $j^{\text{th}}$  column represents the inner product of  $x_i$  and  $x_j$ . We further divide the element in the  $i^{\text{th}}$  row and  $j^{\text{th}}$  column by  $\frac{x_i \cdot x_j^T}{|x_i| \cdot |x_j|}$  such that the element measures the cosine similarity between  $x_i$  and  $x_j$ . The final square matrix obtained is named  $M_{\text{channel}}$ .

To obtain more discriminative features, we require that feature maps in different channels of  $M_4$  differ as much as possible. In other words, the similarity between  $x_i$  and  $x_j$  ( $i \neq j$ ) should be as small as possible. Therefore, we define the following channel orthogonal constraint loss function based on cosine similarity between feature maps.

$$L_{\text{channel}} = \max(M_{\text{channel}}^{ij}) = \max\left(\frac{x_i \cdot x_j^T}{|x_i| \cdot |x_j|}\right), \quad 1 \leq i < j \leq C \quad (1)$$

### 2.6. Species classification task

The categories of plant diseases caused by the same germ are different for different plant species. But their symptoms could be very similar. For example, as shown in Fig. 5, Potato Late Blight and Tomato Late Blight as different plant diseases are both caused by *Phytophthora infestans*, and consequently, they have quite similar symptoms. In



**Fig. 5.** Sample images of different disease categories with similar symptoms in PlantVillage.

response to this situation, we introduce the auxiliary species classification task to make the model learn the species-related features (e.g., leaf shapes and textures) rather than disease-related features and better recognize different plant diseases caused by the same germs.

The plant species classification module can be attached to any of the convolutional features (i.e.,  $M_1, M_2, M_3, M_4$ ). Given the feature  $M_3$ , the module first generates from it a feature vector  $F_{\text{species}} \in R^{1024 \times 1}$  via global average pooling (GAP), and then predicts the possibility of the input plant image belonging to different species according to  $y_{\text{species}} \in R^{N_{\text{species}} \times 1}$  via a fully connected (FC) layer followed by a dropout layer and a softmax function (Srivastava et al., 2014). The dimension  $N_{\text{species}}$  of  $y_{\text{species}}$  indicates the number of plant species categories. Based on the predicted species possibility  $y_{\text{species}}$ , we use cross-entropy as species classification loss.

### 2.7. Total loss function

The total loss function for training the CNN model is

$$L_{\text{total}} = \alpha \cdot L_{\text{channel}} + L_{\text{species}} + L_{\text{disease}} \quad (2)$$

where  $\alpha$  is the weight factor of  $L_{\text{channel}}$ .  $L_{\text{species}}$  is plant species classification loss function, which enforces the model to learn features that are related to plant species, based on which plant diseases can be better recognized.  $L_{\text{disease}}$  evaluates how accurately the features extracted by the model can distinguish between different plant diseases. Hence, we suppose that plant disease recognition is a finer-grained task than plant species recognition, and apply  $L_{\text{disease}}$  at the last layer of the network and  $L_{\text{species}}$  at an earlier layer. As for the  $L_{\text{channel}}$  loss, its purpose is to reduce the correlation between different feature maps and thus enforce the model to extract richer discriminative features for plant disease recognition. Therefore, we apply  $L_{\text{channel}}$  on the same layer as  $L_{\text{disease}}$ . For  $L_{\text{species}}$  and  $L_{\text{disease}}$ , they both use cross-entropy loss function with label smoothing. We use the label smoothing regularization to prevent the model from relying on over-predicted values (Szegedy et al., 2016). The cross-entropy (CE) loss function with label smoothing (LS) is defined as:

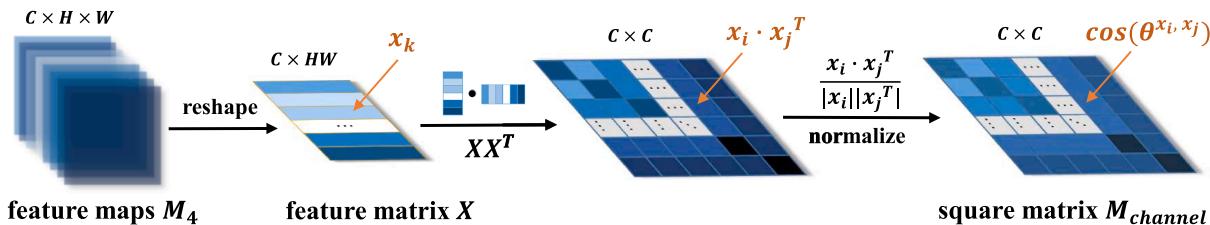
$$L_{\text{CE with LS}} = -\frac{1}{N} \sum_{i=1}^N ((1-\epsilon) \cdot \log(p_{\bar{y}_i}) + \frac{\epsilon}{K-1} \sum_{k \neq \bar{y}_i} \log(p_k)) \quad (3)$$

where  $N$  is the number of training samples,  $K$  is the total number of categories, and  $\epsilon$  is the smoothing factor.  $\bar{y}_i$  represents the ground truth label corresponding to the  $i^{\text{th}}$  sample. When the predicted category equals the ground truth category, i.e.,  $k = \bar{y}_i$ , the weight value of category  $k$  is  $1-\epsilon$ . When  $k \neq \bar{y}_i$ , the weight value is  $\frac{\epsilon}{K-1}$ . The probability value  $p_k$  of predicted category  $k$  in Eq. 3 is calculated by the softmax function:  $p_k = \frac{e^{z_k}}{\sum_{i=1}^K e^{z_i}}$ . The  $z_k$  represents the  $k^{\text{th}}$  value in the predicted value vector  $y$ . For  $L_{\text{disease}}$ , the predicted value  $y = y_{\text{disease}}$  and the number of disease categories is  $K = K_d$ . For  $L_{\text{species}}$ ,  $y = y_{\text{species}}$  and  $K = K_s$ .

## 3. Experiments and analysis

### 3.1. Experiment setup

The models in our experiments are pre-trained on the ImageNet dataset first and then fine-tuned on the PlantVillage dataset. We only input Segmented images and augment the training data online using background replacing, leaf resizing, and basic data augmentation methods during the training phase. In background replacing, we randomly select one background image from 9 different background images (three different background images for each color type) for mixing. The best probability of applying background replacing to each training image is 0.3 in our experiments. In leaf resizing, the input image size is  $224 \times 224$ , and the black background image size is  $448 \times 448$ . The best range of rectangular cropping box size is 0.3~1.0 times of input



**Fig. 6.** Illustration of Channel Orthogonal Constraint. The square matrix  $M_{channel}$  obtained from the last convolutional features  $M_4$  is used for loss function  $L_{channel}$ .

image size. In basic data augmentation, all transformations are randomly adopted to each image with default probability in training data during the training process. Then the augmented images are resized to a fixed size of  $224 \times 224$  and normalized by the mean and variance of the ImageNet (Deng et al., 2009).

During the training phase, the species classification loss  $L_{species}$  is applied on the third feature layer, and the channel orthogonal constraint loss  $L_{channel}$  using hyper-parameter  $\alpha = 0.1$ . The number of plant disease categories  $K_d$  is 38 and the number of plant species categories  $K_s$  is 14. The disease classification loss  $L_{disease}$  combines the above two losses to train the model in an end-to-end manner.

All experiments are implemented on the deep learning framework PyTorch and a PC with four NVIDIA GeForce GTX 1080 GPUs (Ketkar, 2017). For the baseline model and improved model, we fix the backbone network for the first three epochs and warm up the rest layers with a learning rate of  $lr = 0.001$ . We use Adam optimizer with basic learning rate  $lr = 0.0001$  and weight decay  $\lambda = 0.0004$  for rest epochs. The model's learning rate is set to  $0.1 \times lr$  when the test set's accuracy did not increase for three consecutive times. This learning rate schedule can prevent the model from falling into the local optimum. All the experiments use a batch size of 32, and the total number of epochs is 200. The dropout ratio is 0.5 in training. The smoothing factor  $\epsilon$  of label smoothing is set to 0.1.

We divide the PlantVillage data into a training set of 80% and a test set of 20% in all our experiments. Since the PlantVillage dataset has multiple images of the same leaf (taken from different orientations and environments), we make sure all the same leaf images go either in the training set or the test set during train-test split (Mohanty et al., 2016). Unlike PlantVillage, the Syn-PV and Field-PV are only used as test sets, and the results on these test sets are evaluated by the model trained with 80% of the Segmented version images of PlantVillage. The remaining 20% of the Segmented images are used to evaluate laboratory plant disease recognition performance under controlled laboratory conditions. The leaves in the Syn-PV are the same as the leaves in the test set of the PlantVillage. So the number of test images in the PlantVillage and Syn-PV are the same. The Field-PV test set has 665 images. We use Syn-PV and Field-PV to evaluate the performance of FPDR.

Further, for every experiment, we compute the overall accuracy as evaluation criteria. And we use it for the comparison of results across all the different experimental configurations.

### 3.2. Results on PlantVillage

**Table 1** shows the experimental results of different methods performed on the PlantVillage dataset. Among them, only Mohanty et al. evaluated their models on three versions of images in PlantVillage. For fair experimental comparison, we reimplemented the best model from the paper using PyTorch according to the experimental configuration in the paper. In this experiment, the training-test split ratio for all three versions is 80% for training and 20% for testing (Mohanty et al., 2016). To the best of our knowledge, Brahimi et al. (2018) and Too et al. (2019) achieved state-of-the-art performance on the color version of the PlantVillage. On all three version images of PlantVillage, the baseline achieves competitive results compared with the state-of-the-art methods.

**Table 1**

Experimental results of different plant disease recognition methods on the different versions of the public PlantVillage dataset. The baseline model is composed of basic data augmentation methods and a backbone of ResNet50. <sup>†</sup>Reimplemented method.

Method	CNN architecture	Transfer learning	PlantVillage		
			Color	Grayscale	Segmented
(Mohanty et al., 2016)	GoogleNet	Yes	0.9934	0.9800	0.9925
(Brahimi et al., 2018)	Inception	Yes	0.9976	-	-
(Too et al., 2019)	DenseNet	Yes	0.9975	-	-
<sup>†</sup> (Mohanty et al., 2016)	GoogleNet	Yes	0.9972	0.9826	0.9958
<sup>†</sup> (Brahimi et al., 2018)	Inception	Yes	0.9978	0.9920	0.9970
<sup>†</sup> (Too et al., 2019)	DenseNet	Yes	0.9979	0.9910	0.9972
Baseline	ResNet	Yes	<b>0.9984</b>	<b>0.9930</b>	<b>0.9972</b>

### 3.3. Results for FPDR

As can be seen from **Table 2**, although the existing methods have achieved high accuracy on PlantVillage, their performances significantly decreased when testing these methods on Syn-PV and Field-PV. The baseline model tested on the field-PV showed a worse result (from 99.72% to 41.81%) than that on the PlantVillage. Since the leaves in Syn-PV and leaves in the test set of PlantVillage are the same, the declining result (from 99.72% to 87.82%) can prove that the complex background is indeed a difficulty of FPDR.

Compared with the baseline and existing methods, we can obtain comparable results by proposed improved model on Syn-PV and Field-PV. The improvement (from 87.82% to 98.50%) of our improved model on Syn-PV shows the power of background replacing. From **Table 2**, we can see that the accuracy of the best counterpart method is 41.20% on Field-PV, while our method achieves an accuracy of 72.03%,

**Table 2**

Results of different methods on Syn-PV and Field-PV. The models of these methods are trained only by the data of PlantVillage (Segmented). The improved model is composed of baseline, background replacing (BR), leaf resizing (LR), channel orthogonal constraint (COC), and species classification task (SCT). <sup>†</sup>Reimplemented method.

Method	Backbone	Syn-PV	Field-PV
<sup>†</sup> Mohanty et al. (2016)	GoogLeNet	0.6112	0.3097
<sup>†</sup> Brahimi et al. (2018)	Inception-v3	0.6639	0.4045
<sup>†</sup> Too et al. (2019)	DenseNet-121	0.6881	0.4120
Baseline (BS)	ResNet-50	0.8782	0.4181
BS + BR + LR + COC + SCT	ResNet-50	<b>0.9850</b>	<b>0.7203</b>

which is a great improvement of about 74.8%. This may owe to the following factors: 1) background replacing can introduce different backgrounds into the training data so that the improved CNN model will reduce the attention on complex backgrounds in Syn-PV and Field-PV. 2) leaf resizing can increase the number of plant disease images with different sizes and locations of symptoms so that the improved model can identify the correct disease categories of hugely different plant disease images. 3) For the features extracted by the baseline model, channel orthogonal constraint and species classification task encourage the improved model to learn more discriminative features.

In order to show the improvement of the improved CNN model more clearly, we tested all plant disease categories in Field-PV, as shown in [Table 3](#). For the accuracy on Field-PV tested with baseline model, most of them are very low or even 0%, such as “Apple Cedar apple rust” and “Strawberry healthy”. For these categories, the recognition accuracy can be significantly increased by the improved model. The result on “Corn Northern Leaf Blight” has a decline with the improved model. From the [Fig. 7](#), we can see that some corn leaves in PlantVillage have not only big poke-shaped spots but also many small spots. We guess that the improved model with stronger learning ability may make these small spots as discriminative features for this class, which leads to performance degradation on Field-PV. Although the improved model has

**Table 3**

Different plant disease categories' different accuracy results were evaluated on the Field-PV dataset with the baseline model and the improved model.

Category Names	Number	Baseline Model	Improved Model
01Orange Haunglongbing (Citrus greening)	33	0.3333	0.8485
02Apple scab	32	0.6250	0.6563
03Apple Black rot	19	0.2631	0.7368
04Apple Cedar apple rust	12	0.0000	0.5833
05Apple healthy	9	0.4444	0.8888
06Blueberry healthy	12	0.2500	0.7500
07Cherry (including sour) healthy	10	0.0000	0.3000
08Cherry (including sour) Powdery mildew	10	0.2000	0.5000
09Corn (maize) Cercospora leaf spotGray leaf spot	22	0.8636	0.9545
10Corn (maize) Common rust	15	0.4000	0.6667
11Corn (maize) healthy	28	1.0000	1.0000
12Corn (maize) Northern Leaf Blight	17	0.6470	0.5882
13Grape Black rot	9	0.3333	0.3333
14Grape Esca (Black Measles)	11	0.0909	0.0909
15Grape healthy	20	0.3500	0.8000
16Grape Leaf blight ( <i>Isariopsis Leaf Spot</i> )	12	0.5000	0.5000
17Peach Bacterial spot	20	0.6500	0.8000
18Peach healthy	11	0.0000	0.0000
19Pepper bell Bacterial spot	13	0.7692	1.0000
20Pepper bell healthy	9	0.1111	0.8889
21Potato Early blight	11	0.2727	0.6364
22Potato healthy	10	0.0000	0.0000
23Potato Late blight	15	0.3333	0.8000
24Raspberry healthy	8	0.0000	0.5000
25Soybean healthy	23	0.3043	0.8261
26Squash Powdery mildew	25	0.9200	0.9200
27Strawberry healthy	32	0.0000	0.7500
28Strawberry Leaf scorch	48	0.9375	0.9375
29Tomato Bacterial spot	10	0.0000	0.6000
30Tomato Early blight	19	0.3157	0.8421
31Tomato healthy	11	0.0000	0.2727
32Tomato Late blight	28	0.5714	0.8929
33Tomato Leaf Mold	17	0.2941	0.8235
34Tomato Septoria leaf spot	17	0.2352	1.0000
35Tomato Spider mites	11	0.0000	0.0909
36Tomato Target Spot	11	0.0000	0.0000
37Tomato mosaic virus	8	0.0000	0.1250
38Tomato Yellow Leaf Curl Virus	37	0.3783	0.9469
All Categories	665	0.4181	0.7203

improved significantly on Field-PV, the total accuracy of 72% shows that there are still many challenges in FPDR. Multiple diseases in one plant leaves and covariate shift between training data domain and test data domain ([Sugiyama et al., 2008](#)). The plant leaves in the real field have different shapes because of its non-rigid property and different appearance because of the different growing environments.

### 3.4. Ablation study

We test the improved model by enabling or disabling the four novel designs on three test sets to demonstrate their effectiveness and necessity. The results are summarized in [Table 4](#). Obviously, the best accuracy on the Field-PV dataset is obtained when all the novel designs are enabled. It should be noted that leaf resizing plays a significant role in FPDR according to the results of the second row evaluated on the Field-PV dataset, which improves from 41.81% to 56.09%. The accuracy of the Syn-PV test set has noticeable improvement when background replacing is enabled and declines when it is disabled according to the third row in [Table 4](#). The channel orthogonal constraint and species classification task are both effective on PlantVillage and Field-PV datasets, which improve the generalization ability of the improved model in FPDR.

### 3.5. Model analysis

#### 3.5.1. Impact of background replacing's parameters

There are two parameters about the background replacing that needs to be discussed. One is the number of mixing background images. We tried three, six, and nine. As [Fig. 8](#) shows, the accuracy will increase as the number increases, although the result on nine images is only a little better than six images. This demonstrates that the improved model will have more adaptability on complex background in the field conditions if mixing background images could have more diversity. The other is the probability of this operation applying to each training image. In our experiments, the best results were obtained when the probability was 0.3, and we guess that higher probabilities may bring more redundant background information.

#### 3.5.2. Impact of leaf resizing's ratio ranges

As mentioned above, field plant disease' symptom sizes and locations are inconsistent, so the range of the rectangular cropping box has a considerable impact. In this experiment, we take four ratio ranges to control the size of the rectangular cropping box. [Table 5](#) gives the results, according to which we observe that best performance on Field-PV can be obtained by a ratio range of 0.3~1.0. It means that the length and width of the rectangular cropping box can be determined at a range of 0.3~1.0 times of input image size (i.e., 67~224 pixels). We set the ratio range value as 0.3~1.0 in the other experiments in this paper. We guess that the result with a ratio range of 0.1~1.0 is not optimal because too small a cropping box would crop to the healthy area of the diseased leaf, which would change the label of this diseased leaf to a healthy category.

#### 3.5.3. Effectiveness of channel orthogonal constraint

In order to succinctly and effectively prove that the channel orthogonal constraint is helpful for the difficulty of similar symptoms in different plant disease categories, we evaluate this method on some disease categories belonging to the same species. Since the tomato species has the most plant disease categories in the Field-PV, we test 10 different disease categories of tomato species. The results in [Fig. 9](#) can demonstrate the effectiveness of the channel orthogonal constraint, in which this method has a good improvement in most disease categories of tomato species.

#### 3.5.4. Impact of species classification task's attachment layers

We test the feature layer from  $M_1$  to  $M_4$  in this experiment to explore which layer is the best for species classification task. In order to be more

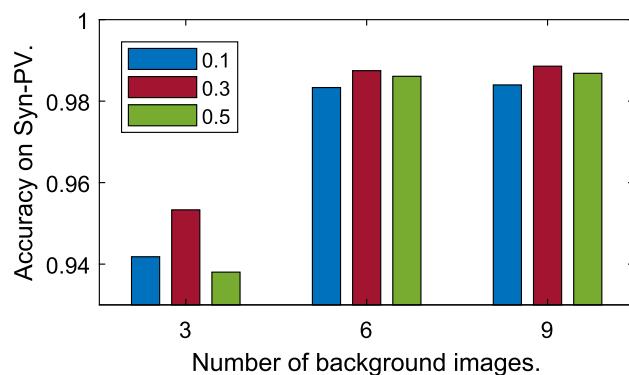


**Fig. 7.** Sample images of the Corn (maize) Northern Leaf Blight category in PlantVillage and Field-PV.

**Table 4**

Performance of the proposed improved model on three test sets when the four novel designs are enabled or disabled. The baseline is the backbone model without the new designs.

Leaf Resizing	Background Replacing	Species Classification Task	Channel Orthogonal Constraint	PlantVillage (Segmented)	Syn-PV	Field-PV
✗	✗	✗	✗	0.9972	0.8782	0.4181
✓	✗	✗	✗	0.9955	0.7683	0.5609
✗	✓	✗	✗	0.9968	<b>0.9886</b>	0.4662
✗	✗	✓	✗	<b>0.9986</b>	0.7783	0.4812
✗	✗	✗	✓	0.9979	0.8901	0.5008
✓	✓	✓	✓	0.9970	0.9850	<b>0.7203</b>



**Fig. 8.** Impact of background replacing with different probabilities and numbers of background images on the Syn-PV dataset with the baseline model trained on the PlantVillage. The different colors represent different probabilities.

**Table 5**

Impact of leaf resizing with different ratio ranges of rectangular cropping box on the Field-PV dataset with the baseline model. The length and width of the rectangular box are in the ratio range times of input image size.

Method	Ratio Range	Field-PlantVillage
Baseline + Leaf Resizing	0.1~1.0	0.5310
	0.3~1.0	<b>0.5610</b>
	0.5~1.0	0.5474
	0.7~1.0	0.5323

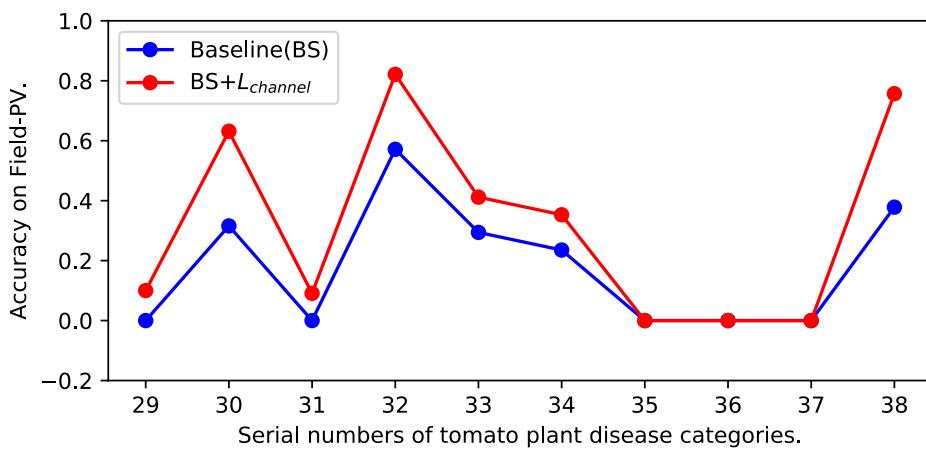
concise and clear, we only show the results of 14 disease categories, in which each category represents one species even though some species may have multiple disease categories. The category numbers in Fig. 10

are the same as the numbers of category name in Table 3. As we can see from Fig. 10, the best results of different categories are mostly obtained on  $M_3$ , on which the features are sufficiently informative for species classification. Since the number of disease categories under each species category varies from small to large, for the case of a large number, the last layer of residual blocks (i.e.,  $M_4$ ) is still needed to learn the feature for disease classification.

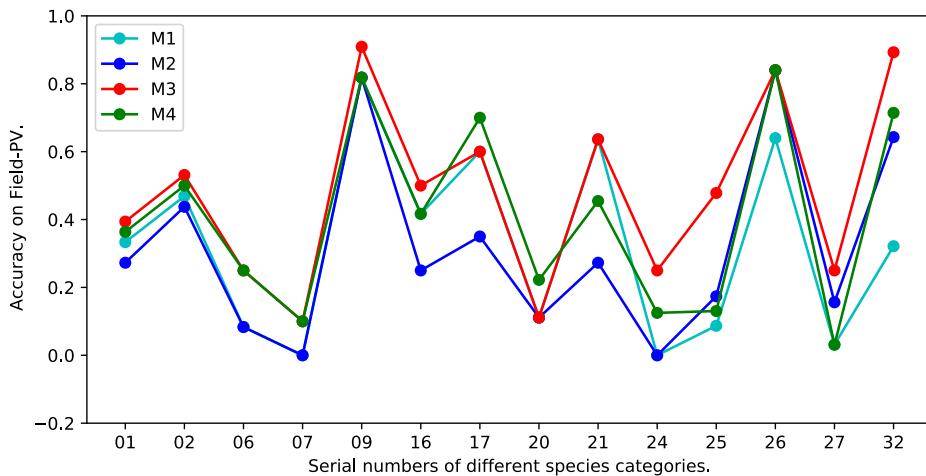
#### 4. Conclusion

Recent advances in deep learning provide solutions with a high recognition accuracy of plant diseases on the public PlantVillage dataset. However, existing methods fail to attain effective performance on FPDR. The experimental results of our reimplemented methods on the in-the-field dataset also show that automatic plant disease recognition in field conditions is a challenging task. After analyzing the differences in images of diseased plant leaves in the field and the laboratory, we found three difficulties: complex backgrounds, varying size and location of disease symptoms, and similar symptoms between different disease categories. To solve these difficulties, this paper proposes an improved CNN model for FPDR based on a CNN backbone network with the addition of data enhancement and discriminative feature learning design. To evaluate the method's performance on the FPDR task, we collect an in-the-field dataset Field-PV as an independent test set. The performance improvement (from 39.38% to 72.03%) of the improved CNN model on the Field-PV demonstrates that the design of data augmentation and discriminative feature learning is effective. The improved CNN model does not require additional training data of field conditions, which substantially improves the performance for FPDR from an algorithmic perspective. It avoids the dependence on new data and greatly improves the widespread and general applicability of practical applications.

Although significant improvement has been brought by our method, there are still many challenges for CNNs to be deployed in real-world



**Fig. 9.** Results of channel orthogonal constraint on different tomato plant disease categories of Field-PV.



**Fig. 10.** Impact of species classification task on different feature layers on 14 different species in Field-PV. The legend indicates the layer of convolutional features on which the method is used.

applications. In the future, we will try to deal with the following problems: multiple leaves in one image, multiple diseases in one plant leaf, and covariate shift between training data domain and test data domain (Sugiyama et al., 2008).

#### CRediT authorship contribution statement

**Penghui Gui:** Conceptualization, Methodology, Supervision, Visualization, Writing – original draft. **Wenjie Dang:** Investigation, Methodology, Funding acquisition, Software. **Feiyu Zhu:** Conceptualization, Writing – review & editing. **Qijun Zhao:** Supervision, Validation.

#### Funding

This work was supported in part by the National Natural Science Foundation of China under Grant 61971005.

#### Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### References

- Abade, A., Ferreira, P.A., de Barros Vidal, F., 2021. Plant diseases recognition on images using convolutional neural networks: A systematic review. *Comput. Electron. Agric.* 185, 106125.
- Abade, A., Vidal, F., Almeida, A., 2019. Plant diseases recognition from digital images using multichannel convolutional neural networks. *VISIGRAPP 5: VISAPP*, 450–458.
- Argüeso, D., Picon, A., Irusta, U., Medella, A., San-Emeterio, M.G., Bereciatua, A., Alvarez-Gila, A., 2020. Few-shot learning approach for plant disease classification using images taken in the field. *Comput. Electron. Agric.* 175, 105542.
- Arsenovic, M., Karanovic, M., Sladojevic, S., Anderla, A., Stefanovic, D., 2019. Solving current limitations of deep learning based approaches for plant disease detection. *Symmetry* 11 (7), 939.
- Barbedo, J.G.A., 2016. A review on the main challenges in automatic plant disease identification based on visible range images. *Biosyst. Eng.* 144, 52–60.
- Bischoff, V., Farias, K., Menzen, J.P., Pessin, G., 2021. Technological support for detection and prediction of plant diseases: A systematic mapping study. *Comput. Electron. Agric.* 181, 105922.
- Boulent, J., Foucher, S., Théau, J., St-Charles, P.-L., 2019. Convolutional neural networks for the automatic identification of plant diseases. *Front. Plant Sci.* 10, 941.
- Brahimi, M., Arsenovic, M., Laraba, S., Sladojevic, S., Boukhalfa, K., Moussaoui, A., 2018. Deep learning for plant diseases: detection and saliency map visualisation. In: *Human and Machine Learning*. Springer, pp. 93–117.
- Chen, J., Chen, J., Zhang, D., Sun, Y., Nanehkaran, Y.A., 2020a. Using deep transfer learning for image-based plant disease identification. *Comput. Electron. Agric.* 173, 105393.
- Chen, J., Zhang, D., Nanehkaran, Y., 2020b. Identifying plant diseases using deep transfer learning and enhanced lightweight network. *Multimedia Tools Appl.* 79 (41), 31497–31515.
- Commons, W., 2020. File:early blight on tomato leaves. [Online; accessed 14-May-2021]. [https://commons.wikimedia.org/w/index.php?title=File:Early\\_blight\\_on\\_tomato\\_leaves\\_\(7871930010\).jpg&oldid=464093452](https://commons.wikimedia.org/w/index.php?title=File:Early_blight_on_tomato_leaves_(7871930010).jpg&oldid=464093452).

- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L., 2009. Imagenet: A large-scale hierarchical image database. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255.
- Ferentinos, K.P., 2018. Deep learning models for plant disease detection and diagnosis. *Comput. Electron. Agric.* 145, 311–318.
- Fuentes, A., Yoon, S., Park, D.S., 2020. Deep learning-based techniques for plant diseases recognition in real-field scenarios. In: International Conference on Advanced Concepts for Intelligent Vision Systems. Springer, pp. 3–14.
- Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial networks. arXiv preprint arXiv:1406.2661.
- He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778.
- Hughes, D., Salathé, M., et al., 2015. An open access repository of images on plant health to enable the development of mobile disease diagnostics arXiv preprint arXiv: 1511.08060.
- Ketkar, N., 2017. Introduction to pytorch. In: Deep learning with python. Springer, pp. 195–208.
- Li, Y., Nie, J., Chao, X., 2020. Do we really need deep cnn for plant diseases identification? *Comput. Electron. Agric.* 178, 105803.
- Lin, T.-Y., RoyChowdhury, A., Maji, S., 2015. Bilinear cnn models for fine-grained visual recognition. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1449–1457.
- Mohanty, S.P., Hughes, D.P., Salathé, M., 2016. Using deep learning for image-based plant disease detection. *Front. Plant Sci.* 7, 1419.
- Moore, D., Robson, G.D., Trinci, A.P.J., 2000. 21st Century Guidebook to Fungi. Cambridge University Press. <https://doi.org/10.1017/CBO9780511977022>.
- Nazki, H., Yoon, S., Fuentes, A., Park, D.S., 2020. Unsupervised image translation using adversarial networks for improved plant disease recognition. *Comput. Electron. Agric.* 168, 105117.
- Oerke, E.-C., 2006. Crop losses to pests. *J. Agric. Sci.* 144 (1), 31–43.
- Picon, A., Alvarez-Gila, A., Seitz, M., Ortiz-Barredo, A., Echazarra, J., Johannes, A., 2019a. Deep convolutional neural networks for mobile capture device-based crop disease classification in the wild. *Comput. Electron. Agric.* 161, 280–290.
- Picon, A., Seitz, M., Alvarez-Gila, A., Mohnke, P., Ortiz-Barredo, A., Echazarra, J., 2019b. Crop conditional convolutional neural networks for massive multi-crop plant disease classification over cell phone acquired images taken on real field conditions. *Comput. Electron. Agric.* 167, 105093.
- Raja, M.U., Mukhtar, T., Shaheen, F.A., Bodlah, I., Jamal, A., Fatima, B., Ismail, M., Shah, I., 2018. Climate change and its impact on plant health: a pakistan's prospective. *Plant Protection* 2 (2), 51–56.
- Shorten, C., Khoshgoftaar, T.M., 2019. A survey on image data augmentation for deep learning. *J. Big Data* 6 (1), 1–48.
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. *J. Machine Learn. Res.* 15 (1), 1929–1958.
- Sugiyama, M., Nakajima, S., Kashima, H., Buenau, P.V., Kawanabe, M., 2008. Direct importance estimation with model selection and its application to covariate shift adaptation. In: Advances in Neural Information Processing Systems, pp. 1433–1440.
- Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z., 2016. Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2818–2826.
- Too, E.C., Yuji, L., Njuki, S., Yingchun, L., 2019. A comparative study of fine-tuning deep learning models for plant disease identification. *Comput. Electron. Agric.* 161, 272–279.
- Wang, Q., Xie, J., Zuo, W., Zhang, L., Li, P., 2020. Deep cnns meet global covariance pooling: better representation and generalization. *IEEE Trans. Pattern Anal. Mach. Intell.* 1.
- Yun, S., Han, D., Oh, S.J., Chun, S., Choe, J., Yoo, Y., 2019. Cutmix: Regularization strategy to train strong classifiers with localizable features. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 6023–6032.
- Zhao, Y., Liu, L., Xie, C., Wang, R., Wang, F., Bu, Y., Zhang, S., 2020. An effective automatic system deployed in agricultural internet of things using multi-context fusion network towards crop disease recognition in the wild. *Appl. Soft Comput.* 89, 106128.