Plant Disease Recognition: A Large-Scale Benchmark Dataset and a Visual Region and Loss Reweighting Approach

Xinda Liu, Weiqing Min, *Member, IEEE*, Shuhuan Mei, Lili Wang, *Member, IEEE*, and Shuqiang Jiang, *Senior Member, IEEE*

***Abstract*— Plant disease diagnosis is very critical for agriculture due to its importance for increasing crop production. Recent advances in image processing offer us a new way to solve this issue via visual plant disease analysis. However, there are few works in this area, not to mention systematic researches. In this paper, we systematically investigate the problem of visual plant disease recognition for plant disease diagnosis. Compared with other types of images, plant disease images generally exhibit randomly distributed lesions, diverse symptoms and complex backgrounds, and thus are hard to capture discriminative information. To facilitate the plant disease recognition research, we construct a new large-scale plant disease dataset with 271 plant disease categories and 220,592 images. Based on this dataset, we tackle plant disease recognition via reweighting both visual regions and loss to emphasize diseased parts. We first compute the weights of all the divided patches from each image based on the cluster distribution of these patches to indicate the discriminative level of each patch. Then we allocate the weight to each loss for each patch-label pair during weakly-supervised training to enable discriminative disease part learning. We finally extract patch features from the network trained with loss reweighting, and utilize the LSTM network to encode the weighed patch feature sequence into a comprehensive feature representation. Extensive evaluations on this dataset and another public dataset demonstrate the advantage of the proposed method. We expect this research will further the agenda of plant disease recognition in the community of image processing.**

Manuscript received March 11, 2020; revised August 28, 2020 and October 10, 2020; accepted December 28, 2020. Date of publication January 14, 2021; date of current version January 22, 2021. This work was supported in part by the National Natural Science Foundation of China under Project 61932003 and Project 61772051; in part by the National Key Research and Development Plan under Grant 2019YFC1521102; in part by the Beijing Natural Science Foundation under Grant L182016; in part by the Beijing Program for International S&T Cooperation Project under Grant Z191100001619003; and in part by the Shenzhen Research Institute of Big Data (Shenzhen). The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Guo-Jun Qi. *(Corresponding author: Lili Wang.)*

Xinda Liu and Lili Wang are with the State Key Laboratory of Virtual Reality Technology and Systems, Beijing Advanced Innovation Center for Biomedical Engineering, Beihang University, Beijing 100191, China, and also with the Peng Cheng Laboratory, Shenzhen 518066, China (e-mail: liuxinda@buaa.edu.cn; wanglily@buaa.edu.cn).

Weiqing Min and Shuqiang Jiang are with the Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China, and also with the University of Chinese Academy of Sciences, Beijing 100049, China (e-mail: minweiqing@ict.ac.cn; sqjiang@ict.ac.cn).

Shuhuan Mei is with Beijing Puhui Sannong Technology Company Ltd., Beijing 100190, China (e-mail: long8622416@163.com).

Digital Object Identifier 10.1109/TIP.2021.3049334

***Index Terms*— Plant disease recognition, fine-grained visual classification, reweighting approach, feature aggregation.**

I. Introduction

PLANT diseases cause severe threats to global food security by reducing crop production all over the world. According to the statistics, about 20%-40% of all crop losses globally are due to plant diseases [1]. Therefore, plant disease diagnosis is critical to the prevention of spread of plant diseases and reduction of economic losses in agriculture. Most of the plant disease diagnosis methods heavily rely on either the molecular assay or plant protector's observation. However, the former is complicated and constrained to centralized labs while the latter is time-consuming and prone to errors. Currently, image-based technologies are being widely applied to various interdisciplinary tasks via deciphering visual content, e.g., medical imaging [2], food computing [3] and cellular image analysis [4]. Benefitting from recent advances in machine learning, especially deep learning [5], we assert that plant image analysis and recognition can also provide a new way for plant disease diagnosis. Meanwhile, the applications in visual plant disease diagnosis conversely promote the development of image processing technologies.

The research and exploration on plant image analysis in this field have begun to develop, such as aerial phenotyping [6] and fingerprinting of leaves [1]. However, these methods heavily rely on either expensive devices or complex molecular technology, and thus are not easily popularized. Recently, some works [7]–[12] adopt deep learning methods for plant disease recognition. However, most of them directly extract deep features from plant disease images without considering characteristics of the task. In addition, these works are restricted to small datasets with fewer categories and simple visual backgrounds.

According to our survey, there are mainly three distinctive characteristics for plant disease images taken in real-world scenarios. (1) **Randomly distributed lesions.** The foliar lesions probably randomly occur in the plant leaves. As shown in Fig. 1 (a), the cherry fungal shot hole disease is distributed in many different parts of the leaf, including the top, left and right positions. Because deep convolutional neural networks trained with image level labels only tend to focus on the

Fig. 1. Some plant disease samples with different characteristics: (a) Randomly distributed lesions. (b) Diverse symptoms. (c) Complex backgrounds. The diseased parts are annotated by red boxes.

most discriminative parts while missing other object parts, as claiming in [13], many lesions are easy to be neglected. (2) **Diverse symptoms.** Even for the same plant disease, there are probably various visual symptoms on the plant leaves at different time periods. As shown in Fig. 1 (b), due to different degrees of infection, the allium fistulosum black spot shows different symptoms in different leaves of the same plant. The middle leaf is severely infected while the others appear to be mild. The appearances vary considerably in different infected stages, leading to large intra-class variations. (3) **Complex backgrounds.** There usually exist various background clutters in real-world scenarios. As shown in Fig. 1 (c), there are dense leaves and any other irrelevant objects in the background. In contrast, the disease symptoms are not salient, making plant disease recognition more difficult.

To advance the plant disease recognition research in agricultural image processing, we collect a large-scale plant disease dataset Plant Disease Dataset 271 (PDD271) with 220,592 plant leaf images belonging to 271 plant disease categories. To the best of our knowledge, this dataset is the first large-scale plant disease dataset that is meaningful for image processing research in the agricultural field. Some image samples are shown in Fig. 2 while Fig. 3 shows the number of images per category, sorted in decreasing order in different categories. Pie chart in Fig. 3 indicates the overall balance among the three macro-classes, no matter in terms of category number or image number per category. All the images are taken in real-world scenarios with different conditions.

Taking the characteristics of the plant disease image into consideration, we tackle visual plant disease recognition via reweighting both visual regions and loss. Particularly, considering randomly distributed lesions, we explore the multi-scale strategy by dividing the plant disease images into non-overlapping patches, and compute the weights of these patches according to the cluster distribution of these patches in order to indicate the discriminative level of each patch. By setting different weights to different patches, we enhance the influence of patches with diseased symptoms and reduce the interference of irrelevant patches. We further allocate the weight to each loss for each patch-label pair during weakly-supervised training for diseased parts learning. Finally, we extract patch features from the network trained with loss reweighting and adopt a LSTM network to encode the weighted patch feature sequence into a comprehensive feature representation.

In summary, we make the following main contributions:

(1) We conduct for the first time the systematical investigation and analysis of the problem of plant disease recognition in the community of agricultural image processing.

(2) We propose a novel framework, which can explore a multi-scale strategy and reweight both visual regions and the loss to emphasize discriminative diseased parts for plant disease recognition.

(3) We collect the largest labeled plant disease dataset PDD271 with 271 plant disease categories and 220,592 images to date and conduct extensive evaluations on newly proposed PDD271, demonstrating the effectiveness of our method.

The rest of this paper is organized as follows. Section II reviews the related work. Section III introduces the construction of PDD271. Section IV elaborates the proposed plant disease recognition framework. Experimental results and analysis are reported in Section V. Finally, we conclude the paper and give the future work in Section VI.

II. Related Works

A. Plant Disease Recognition

Plant disease diagnosis is critical to the prevention of spread of crop diseases and reduction of economic losses in agriculture [1]. Most of traditional methods rely on the molecular technologies [14], [15] that are complicated, time-consuming and constrained to centralized laboratories. Therefore, some works adopt traditional computer vision methods for plant disease recognition, such as hyperspectral image analysis [16], artificial bee colony algorithm [17], and image segmentation [18]. Recently, there have been more attempts to utilize deep learning in plant disease recognition for its powerful capability of discriminative feature learning [7]–[10]. For example, Wang *et al.* [10] finetuned the VGG, ResNet50, and GoogleNet directly on the leaf disease set and Ferentinos *et al.* [9] finetuned the AlexNet and GoogleNet directly to identify 14 crop species and 26 diseases. However, most of them directly extract deep features without considering the characteristics of the plant disease image. Besides, most of the works conduct their evaluations on small-scale datasets. Table I summarizes the most common plant disease and crop pest datasets. We can see that PlantVillage Dataset [21] is the largest plant disease dataset, but only contains 38 plant disease categories. In addition, the images from this dataset are taken on the table, and not in the real-world scenarios. We show some samples of these leaf datasets in Fig. 4.

Different from these works, we systematically analyze the problem of plant disease recognition and propose a multi-scale method to reweight visual regions and the loss to emphasize discriminative diseased parts for plant disease recognition based on the characteristics of the plant disease image. Furthermore, we collect a large-scale plant disease dataset PDD271, which not only has the advantage in data volume and category coverage, but also is collected in real-world scenarios with complex background (as shown in Table I and Fig. 4). In particular, there is another agricultural dataset IP102 [24], which is relevant to crop pest. This dataset contains more

TABLE I
STATISTICS ON EXISTING PLANT DISEASE DATASETS

Dataset	Image Number	Class Number	Coverage
Leaflet Cassava Dataset [19]	1,896	6	Only Cassava
Kaggle Cassava Disease [20]	9,436	5	Only Cassava
PlantVillage Dataset [21]	54,309	38	Fruit, Crop
Leaf Disease Dataset [7]	4,483	15	Only Fruits
Apple Leaf Disease Dataset [22]	404	3	Only Apple
Crop Pests Dataset [23]	4,500	40	Crop Pest
IP102 [24]	75,222	102	Crop Pest
PDD271	**220,592**	**271**	**Fruit, Vegetable, Crop**

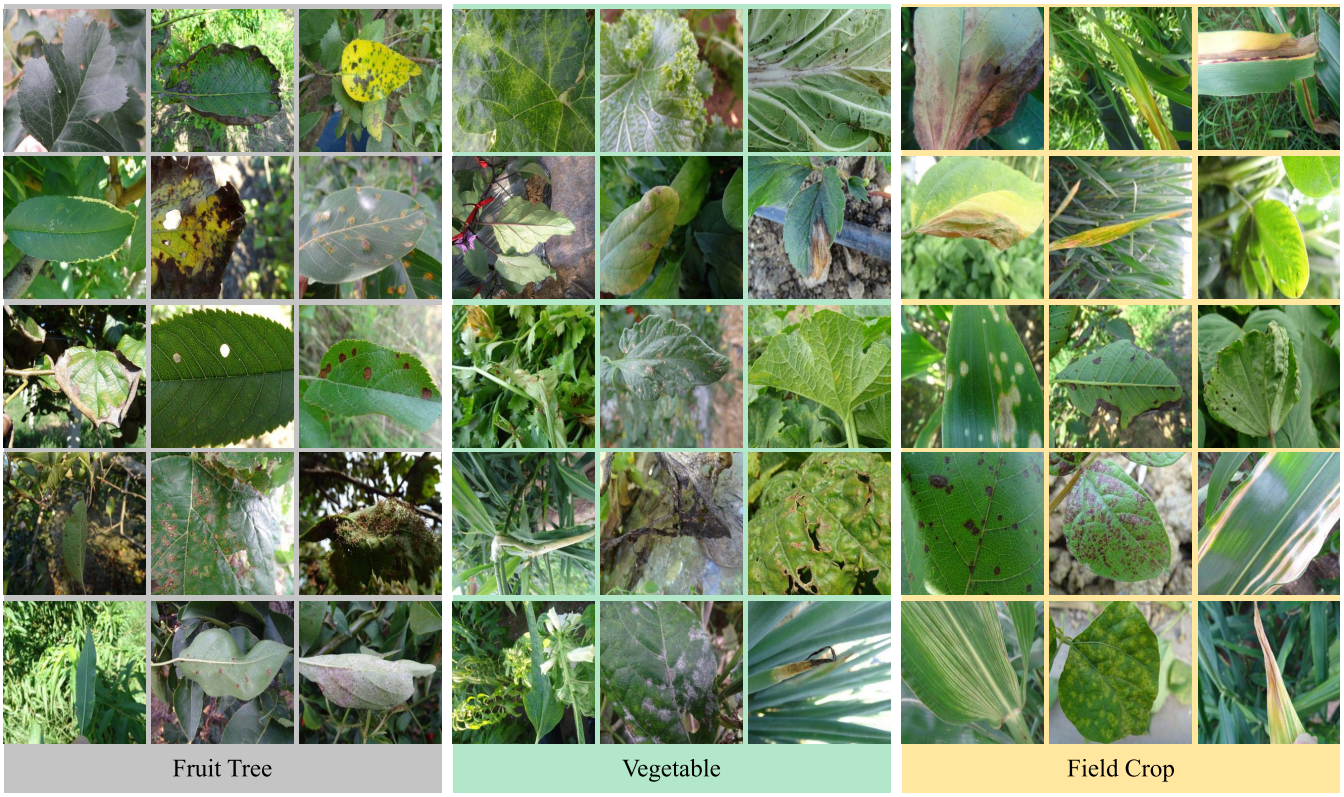


Fig. 2. Disease leaf image samples from various categories of PDD271 (one samples per category). The dataset contains three macro-classes: Fruit Tree, Vegetable, and Field Crops.

than 75, 000 images belonging to 102 categories for insect pest recognition. In contrast, PDD271 aims at advancing plant disease recognition. We believe that PDD271 and IP102 are very complementary and can jointly promote the development of intelligent agriculture analysis and understanding in the image processing and computer vision community.

B. Fine-Grained Visual Classification

Fine-grained image recognition aims to distinguish sub-ordinate categories, such as birds and food. In the early stage, researchers [25], [26] based on deep learning first used strong supervised mechanisms with part bounding box annotations to learn to attend on discriminative parts. Recent researches [3], [13], [27]–[32] focused on weakly-supervised recognition methods without high-cost object part locations or attribute annotations. For example, Yang *et al.* [32] initialized many anchors randomly and extracted their features as their informativeness using the RPN method, and finally chose the informative region to improve the classification performance. There are also several attention-based methods proposed for Fine-Grained Visual Classification. For example, Hu *et al.* [33] used attention maps to guide the data augmentation, Peng *et al.* [34] proposed the object-part attention model to select discriminative regions subjecting to the object-part spatial constraint, and SeNet154 [35] enhance the recognition performance with spatial-channel attention. However, attention-based methods probably focus on the most discriminative parts while missing other parts for the whole image.

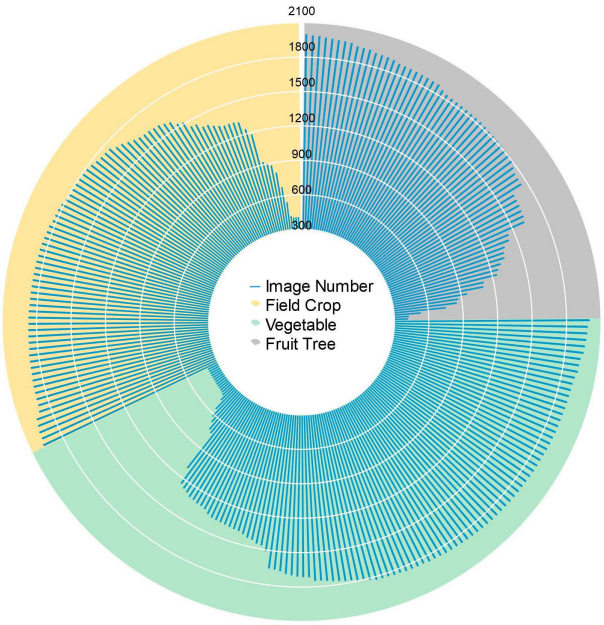


Fig. 3. Sorted distribution of image number per category in three macro-classes of the PDD271.

Fig. 4. Examples from different datasets.

Plant disease recognition belongs to fine-grained recognition. However, we can not simply and directly use existing fine-grained classification methods without considering the characteristics of plant disease images. For example, as shown in Fig. 1, in the plant disease recognition task, large intra-class variations are caused by not only different poses, scales and rotations, but also by different infected stages. In addition, plant disease symptoms are usually not very salient in the plant disease image. Hence, different from existing fine-grained recognition methods, taking characteristics of the plant disease image into consideration, we design a different fine-grained recognition method via reweighting both visual regions and the loss to emphasize the diseased parts for the plant disease recognition. Our method is heuristic with many possibilities to improve. For example, our method and attention-based methods can work together in one framework to further enhance the recognition performance, such as gate-attention deep networks [36]. We can also conveniently combine our framework with the spatial context through discriminative spatio-appearance kernels [37] to promote the performance.

III. Dataset Construction

Due to the complexity, diversity and variability of plant diseases, constructing a large-scale dataset with high-quality is difficult. First, building agricultural datasets should resort to many experts in different fields for annotation. For example, annotating diseases on apple fruit trees and juglans need different experts, which is demand-specific and time-consuming. Second, collecting plant disease images is extremely limited by the time and location. For example, the pear black spot occurs usually in May while the apple ring rot often occurs in the Bohai Sea area. We have to arrive at the place of disease in time. Otherwise all the work would be in vain.

In particular, the data construction is composed of the following three steps. (1) **Taxonomic System Establishment.** We establish a four-layer hierarchical taxonomic system for the PDD271 dataset. We invite several agricultural experts and discuss the common categories of plant diseases which exist in daily life. Each disease is assigned an upper-level class based on the plant that suffers from the disease. And each plant is assigned an upper-level class based on planting condition and plant morphology. For example, the apple brown spot spoils the apple, and the apple belongs to the fruit tree. Finally, we construct a structure with the dataset root, macro-classes, plant categories and plant diseases with 1, 3, 43 and 271 nodes in the first-layer, second- layer, third-layer and fourth-layer, respectively. Fig. 5 shows the results of plant disease hierarchy visualization. (2) **Dataset collection.** In order to collect large numbers of disease images, we organize ten teams. Each team consists of eight students from the agricultural university and four experts in relevant fields. Each team collects thirty kinds of diseases, where every disease contains over five hundred different images from different plants. Experts are responsible to guarantee the quality of disease images and their annotation. When capturing images, one standard protocol is that the distance between the camera and plant is in [20cm,30cm] to guarantee similar visual scope. Every plant disease category contains 500 images at least, and more than 200 plants are captured for one category. In addition, one plant can be captured from different angles. This is for the diversity of plant disease data and this diversity is good for gaining higher generalization power of networks. (3) **Dataset processing and expansion.** After image collection, each image is checked by three experts to make sure the label correctness. Then, experts remove blurry images and other noisy images to keep the dataset clean. For the categories with fewer images, we further collect more images to guarantee the image number of each category.

Fig. 5. Taxonomy of the PDD271 dataset.

The whole data construction takes about 2 years. The resulting PDD271 contains 220,592 images and 271 categories. As shown in Fig. 3, the minimum number of images per category is over 400 and the maximum one is 2000. The balanced distribution ensures the stability of model training. A reliable dataset plays an essential role in developing image processing technologies in a specific area. For example, HiEve [38] is vital to human-centric analysis, so as ATRW [39] to wildlife conservation. Likewise, the proposed dataset PDD271 offers a large coverage and diversity of plant diseases. It will further the plant disease recognition agenda and expand the image processing techniques into the agricultural area.

IV. Framework

In this section, we introduce the proposed framework which explores a multi-scale strategy and reweights both visual regions and the loss during the weakly-supervised learning to emphasize discriminative diseased parts for the purpose of the plant disease recognition. As shown in Fig. 6, this framework mainly consists of three stages, namely Cluster-based Region Reweighting (CRR), Training with Loss Reweighting (TLR) and Weighted Feature Integration (WFI). CRR takes all the divided patches from plant disease images as input and sets the weight of each patch according to the cluster distribution of the visual features of these patches. For each patch-label pair, TLR allocates the corresponding weight to each loss during weakly-supervised training in order to enable the discriminative disease part learning. Based on extracted patch features from TLR and corresponding weights from CRR, WFI utilizes the LSTM network to encode the weighed patch feature sequence into a comprehensive feature representation. Section IV-A details CRR, Section IV-B introduces TLR and Section IV-C presents WFI.

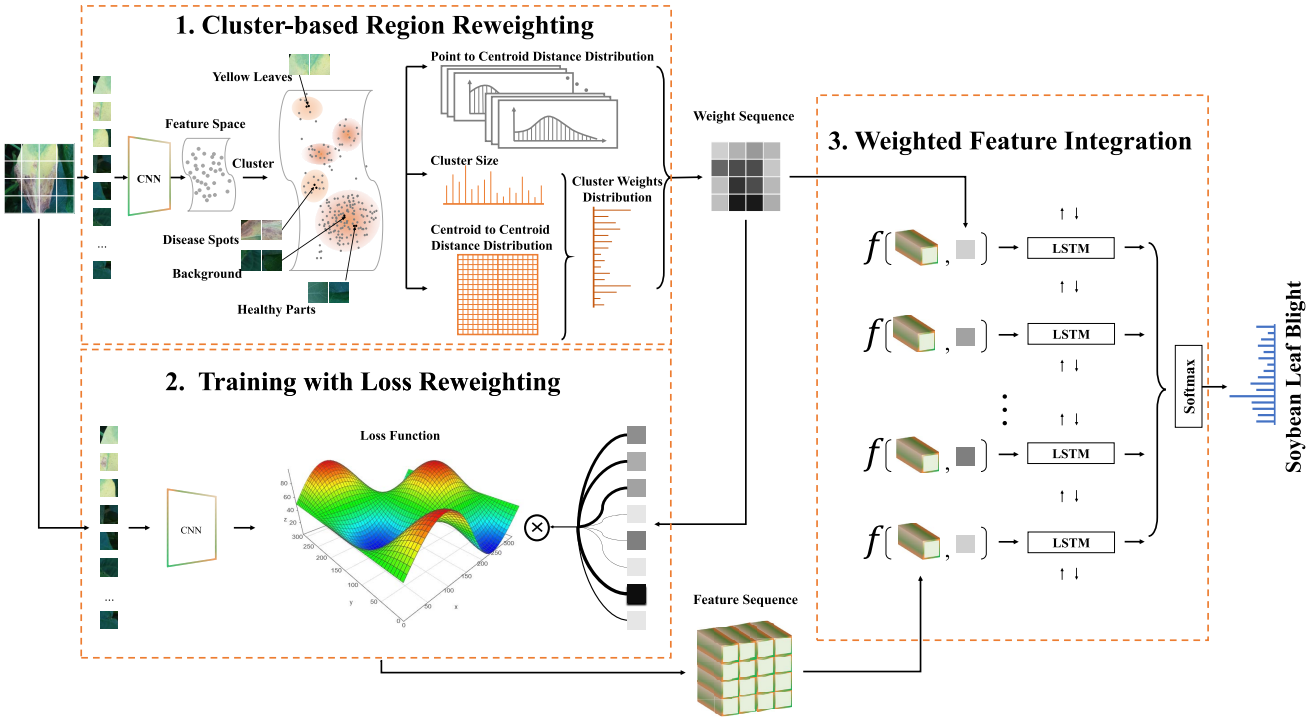


Fig. 6. The proposed plant disease recognition framework.

A. Cluster-Based Region Reweighting

Many diseases present small and scattered lesions, such as pumpkin mildew, pear frog-eye leaf spot and actinidia brown spot. The deep convolutional neural networks trained with image level labels often overlook these lesions while focusing on more salient parts. Considering these situations, we explore a multi-scale strategy by dividing the images into non-overlapping patches and enlarging every patch to avoid missing diseased patches. However, the disease-independent patches, such as the complex backgrounds and the healthy parts, are enhanced even more in the above process, which could lead to severe unbalance between the diseased patches and the irrelevant ones. To address this problem, we attempt to use the visual similarity among the same disease to cluster the patches of the same disease. Afterwards, we reweight the patches based on the clustering result and indicate the discriminative level of each patch.

Formally, all patches from all the original training images form a new training set. Let $X \in \mathbb{R}^{m \times N}$ denotes the visual features of these patches, where m is the dimension of the visual feature and N is the number of training patches. We then have these patches clustered into k cluster classes c with their centroids being $\{\mu_1, \mu_2, \ldots, \mu_k\} \in \mathbb{R}^m$. To compute the weight $w_x, x \in X$, the weights of the clusters $\mathbf{w_c}$ and the probability distribution $\mathbf{p_x}$ of x belonging to over all clusters are computed. Then, w_x is computed as

$$w_x = \mathbf{p_x} \cdot \mathbf{w_c}, \tag{1}$$

where $\mathbf{w_c} = [w_{c_1}, \ldots, w_{c_i}, \ldots, w_{c_k}]$ and w_{c_i} denotes the weight of the cluster c_i.

Normally, the patches containing similar visual symptoms are likely to be assigned to the same clusters. In case of small distance among clusters, the visual phenotypes of different diseases are similar and hard to distinguish by the deep model. Therefore, these clusters are given higher weights to enhance their influence in follow-up feature learning and integration. The size of cluster is also an important indicator. There is a highly skewed distribution of different disease patches. For example, the number of non-diseased patches containing complex backgrounds and foliar healthy parts is very large, but the number of patches containing cotton eye spot disease is small due to the concentrated symptom of this disease leading to the poor classification performance. Meanwhile, the distance between two clusters indicates their visual difference. If one cluster is far from the other clusters, we can easily obtain discriminative features for this cluster, thus assign a small weight to it. Hence we assign these clusters suitable weights to make their influences as balanced as possible.

Given all these, we assign the cluster weights according to the following rule: the larger size the cluster and the farther away from the others, the smaller its weight. We use a monotone decreasing function $F = e^{x/(x-1)}$ to model this change. According to the size of the cluster and the distance distribution among the cluster centroids, we compute the weights of the cluster c_i as follows,

$$w_{c_i} = F(N_{c_i}) \times F\Big(\sum_{j \neq i, j \in 1,\ldots,k} d(\mu_i, \mu_j)\Big), \tag{2}$$

where N_{c_i} is the number of patches in cluster c_i and $d(\mu_i, \mu_j)$ is the distance between the centroid μ_i and μ_j.

To compute the probability distribution $\mathbf{p_x}$, we use a soft assignment strategy based on the distances between a patch and the cluster centroids. The assignment probability distribution p_x is computed as following,

$$\mathbf{p_x} = F(d(x, \mu_i)), \quad i \in 1, \ldots, k. \tag{3}$$

The weight w_x is then computed using Equation 1.

For the patches from validation and testing datasets, we compute the probability distributions $\mathbf{p_x}$ based on the distance to the centroids learned from the training patches. The cluster weights have been computed and the patch weights are computed via Equation 1. The patches and their corresponding weights are used for the model training with loss reweighting.

B. Training With Loss Reweighting

To extract more discriminative regional features for the given patches, we train the network with a reweighted loss function. An observed input patch x shares the same label l with the original image. The model computes an observation o for this patch. The score can be interpreted as an estimation of a class posterior probability $p_\theta(o|x)$, where θ is the model parameters. Given labeled training data $\{(x_n, l_n) : n = 0, \ldots, N-1\}$, the original cross-entropy loss is defined as:

$$Loss_{ori}(x_n, l_n; \theta) = -\sum_{n=0}^{N-1}\sum_{l=1}^{M} y_{o_n,l_n} \log(p_\theta(o_n|x_n)) \tag{4}$$

where M denotes the number of plant disease classes, and y is one binary indicator defined as follows:

$$y_{o_n,l_n} = \begin{cases} 1 & o_n = l_n \\ 0 & o_n \neq l_n \end{cases} \tag{5}$$

However, this loss treats every patch equally. As a result, patches irrelevant to the disease symptoms distract the optimization of network. To solve this, we propose a new reweighted loss to enhance the influence of patches with discriminative diseased symptoms and to reduce the interference of irrelevant patches. For the observed input patch x, its weight w_x is precomputed via CRR. Given labeled training data $\{(x_n, w_{x_n}, l_n) : n = 0, \ldots, N-1\}$, we define the reweighted loss function as follows:

$$Loss_{rew}(x_n, w_{x_n}, l_n; \theta) = -\sum_{n=0}^{N-1}\sum_{l=1}^{M} w_{w_n} y_{o_n,l_n} \log(p_\theta(o_n|x_n)). \tag{6}$$

We allocate the weight to each loss for each patch-label pair. This loss forces the model to focus on the patches with discriminative diseased parts and to ignore the irrelevant patches as much as possible. This trained model can be used to extract visual features from all the patches. The patch features from the same image form a sequence as the input for the following weighted feature integration. For clarity, $\hat{x}$ denotes patch feature from the same image.

C. Weighted Feature Integration

The combination of diseased and healthy patches in plant images constitutes the complex and diverse visual patterns. We try to model the semantic correlation from the combination of local patches. Specifically, we propose a feature integration model with reweighting patch features as the inputs to induce the BiLSTM network to model the semantic correlativity among patches by end-to-end training.

Given a feature sequence $S = [\hat{x}_1, \ldots, \hat{x}_t]$ extracted from the network with TLR and its corresponding weight sequence $W = [\hat{w}_1, \ldots, \hat{w}_t]$ obtained from CRR, where t denotes the number of patches for one image, we combine the feature sequence with the weight via a following function $\boldsymbol{A}(S, W)$,

$$\boldsymbol{A}(S, W) = [f(\hat{x}_1, \hat{w}_1), \ldots, f(\hat{x}_t, \hat{w}_t)]. \tag{7}$$

Note that the function $\boldsymbol{A}(S, W)$ can be one of many aggregation methods, such as deep feedforward networks. Without loss of generality, the element-wise multiplication is adopted in our experiment.

For each image, a common two-layer stacked LSTM is adopted to fuse weighted patch feature sequences into the final representation. The hidden state of the first LSTM is fed into the second LSTM layer which follows the reversed order of the first one. The dimension of hidden states from both layers is 4,096. The output $o' = L(\boldsymbol{A}(S, W); \theta')$, where θ' is the model parameters. We use softmax to generate the class probability vector for each image $S_{i'}$, denoted as $\phi(L(\boldsymbol{A}(S_{i'}, W_{i'}); \theta')) \in \mathbb{R}^{M\times 1}$. The final loss function is defined as follows:

$$Loss_{lstm}(S_{i'}, W_{i'}; \theta') = -\sum_{i'=0}^{\frac{N}{t}-1}\sum_{l=1}^{M} y_{o'_{i'},l_{i'}} \log(\phi(L(\boldsymbol{A}(S_{i'}, W_{i'}); \theta')). \tag{8}$$

By optimizing this loss function, we obtain a weighted BiLSTM to encode the patch feature sequence into a comprehensive feature representation for plant disease recognition.

V. Experiment

A. Experimental Setting

1) Dataset Split and Evaluation Metric: The PDD271 contains 220,592 images belonging to 271 classes of diseases. We follow a roughly 7: 2: 1 split. The PDD271 is split into 154,701 training, 44,002 validation, and 21,889 testing images. Top-1 classification accuracy is adopted as the evaluation metric.

2) Hyperparameter Setting: All the images are resized to 224×224. Each image is divided into 4×4 patches. The initial learning rate is 0.001 and is divided by 10 after every 20 epochs with the standard SGD optimizer. Training converges after 100 epochs. The batch size is 128, and the momentum is 0.9. We adopt a random horizontal flip method for data argumentation in all the experiments. Our project will be made available at https://github.com/liuxindazz/PDD271https://github.com/liuxindazz/PDD271.

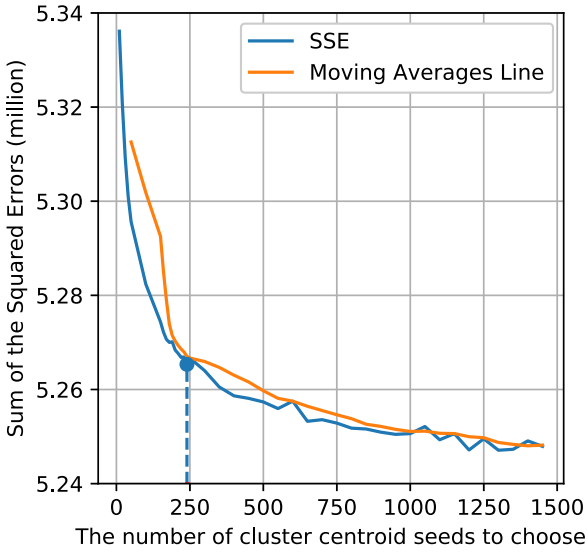


Fig. 7. The result of elbow method. The blue line shows that the SSE changes with the K, and the orange line is the MA line.

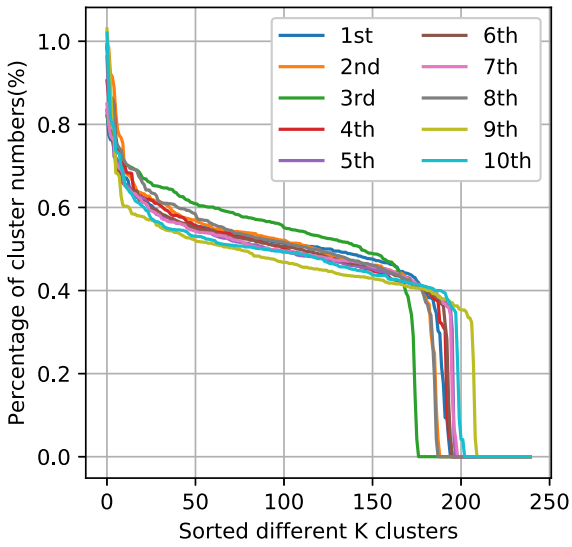


Fig. 8. The percentage distribution results from ten experiments.

B. Experiment on PDD271

1) The Choice of K in CRR: There are many notable cluster algorithms, such as Gaussian mixture model and K-means clustering. Considering the very large size of the dataset and the robustness of the algorithm, we chose the K-means++ as the cluster algorithm on all our experiments. We used the 'elbow' method to determine the value of K. As shown in Fig. 7, from the sum of the squared errors (SSE) and the trend line of SSE computed by the Moving Average (MA) method, we can obtain the obvious 'elbow' point in where K is 240.

We repeat the clustering procedure 10 times in Fig. 8, and observe that the distributions of these clustering results are similar and stable, quantificationally indicating that the clustering is converged in this dataset. Furthermore, Fig. 9 shows the t-SNE [40] map of clustering with patch samples in random five clusters to support the clustering results qualitatively.

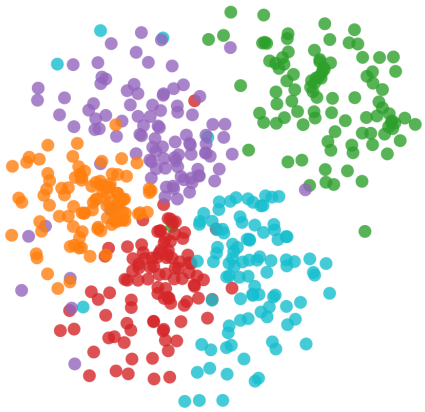


Fig. 9. The t-SNE visualizations of the result of K-means clustering(randomly choosing five cluters).

TABLE II
PERFORMANCE COMPARISON FOR DIFFERENT TRAINING METHODS

Method	Validate(%)	Test(%)
ResNet152	88.63	83.37
w/o WSL	81.98	77.14
w/o Weights	89.07	84.96
TLR	**89.44**	**85.40**

2) Evaluation of TLR: We used the following comparison for evaluating TLR.

- ResNet152. This method directly finetunes ResNet152 using the whole image.
- w/o WSL. This method uses the finetuned ResNet152 to directly extract the feature of each patch following the maxpooling and softmax layers. 'w/o WSL' means 'without weakly supervised learning'.
- w/o Weights. This method first uses the patches from all the images to finetune ResNet152, and then extract the feature of each patch following maxpooling and softmax layers. 'w/o Weights' means 'training on patches without weights'.

Table II shows the experimental results. We can see that (1) w/o WSL brings a drop of performance. The probable reason is that the information about backgrounds or healthy parts is harmful for predicting disease categories and blocking into patches exacerbates the influence of the information. (2) Training on patches without weights improves the recognition performance compared with ResNet152. It shows that weakly-supervised learning is effective for this task. (3) TLR gives further performance boost in both the validation set and the testing set, which demonstrates the advantage of loss reweighting in emphasizing diseased parts. (4) Due to the complexity, diversity, and variability of plant disease, it is hard to improve the recognition performance dramatically.

TABLE III
PERFORMANCE COMPARISON FOR INTEGRATION METHODS

Method	Validate(%)	Test(%)
LSTM	87.95	83.23
BiLSTM	89.70	85.50
sumBiLSTM	88.72	84.60
WFI	**89.91**	**85.54**

TABLE IV
ABLATION STUDY ON THE PDD271

Method	Validate(%)	Test(%)
w/o CRR	89.54	84.87
w/o TLR	89.23	84.98
w/o WFI	89.44	85.40
Full	**89.91**	**85.54**

However, the proposed TLR still gives a considerable improvement compared to the results of other competitive baselines.

3) Evaluation of WFI: We evaluated the effect of LSTM and its variants including BiLSTM and sumBiLSTM for integration of patch feature sequence without reweighting. In contrast, the input of WFI is the weighted patch feature sequence. As shown in Table III, the proposed WFI achieves the best recognition performance, and benefits significantly from the feature reweighting strategy.

4) Ablation Study: We further evaluated the effect of each component in our framework: CRR, TLR, and WFI. We designed different runs in the PDD271 datasets as follows.

- w/o CRR. In this baseline, we directly use the trained network to extract patch features, and then fuse these feature via BiLSTM.
- w/o TLR. In this baseline, we replace the reweighted loss with the original loss during the weakly-supervised training in our framework.
- w/o WFI. In this baseline, WFI is replaced with the Maxpooling layer in our framework.

As can be seen from the Table IV, (1) Any one of three components in isolation brings disease recognition performance gain; (2) Without TLR, the performance drops to 89.23%, showing that TLR is crucial in improving the performance. (3) Without CRR, the performance in testing set drops to 84.87%, indicating that CRR enhances the robustness of our framework. The ablation study validates our design is rational that it is necessary to jointly adopt three components in order to achieve the best performance.

5) Comparisons to the State-of-the-Art: For further verification for the proposed method, we compare our method against the state-of-the-art deep network architectures and fine-grained recognition models, as shown in Table V. The results show that (1) The performance of the SeNet154 is better than other single networks, since it can obtain the useful disease information with spatial-channel attention. (2) Our method can improve the performance of the Resnet152 by 1.28% without adding any attention modules. In addition, we also change the backbone network from Resnet152 to attention model SeNet154. The

TABLE V
COMPARISON WITH THE RESULTS FOR STATE-OF-THE-ART DEEP NETWORK ARCHITECTURES AND FINE-GRAINED RECOGNITION MODELS ON PDD271

Method	Validate(%)	Test(%)
VGG16 [42]	85.51	79.80
ResNet50 [43]	87.96	82.75
ResNet152 [43]	88.63	83.37
WRN [44]	86.15	78.95
DenseNet161 [45]	89.79	85.43
GoogleNet [46]	86.74	81.56
PolyNet [47]	87.42	82.54
PNASNet [48]	89.41	84.44
SeNet154 [35]	89.95	84.63
WS-DAN(ResNet152) [33]	89.36	84.25
NTS-NET(ResNet152) [32]	84.10	81.30
DCL(ResNet152) [41]	89.86	85.00
Our Method(ResNet152)	89.91	85.54
Our Method(SeNet154)	**90.01**	**85.58**

TABLE VI
PERFORMANCE COMPARISON FOR DIFFERENT PATCH SIZES.4×4 MEANS THE IMAGE IS DIVIDED INTO 4×4 PATCHES, AND THE SIZE OF EACH PATCH IS 56×56 PIXELS

Method	Validate(%)	Test(%)
2×2	89.95	85.53
3×3	89.99	85.55
4×4	**90.01**	**85.58**

experimental results also show that combining our method with SeNet154 improves performance by $\sim$1 percent in testing (from 84.63% to 85.58%) and achieves the state-of-the-art performance. This phenomenon further demonstrates that our method and attention-based methods can work together in one framework to futher enhance the recognition performance. (3) Overall, our method performs better than the state-of-the-art fine-grained methods, including the attention-based method WS-DAN [33] and the patch-based methods NTS-NET [32] and DCL [41]. This phenomenon shows that our method is more appropriate for plant disease recognition with considering the characteristics of plant disease image.

6) Influence of Different Patch Size: The size of patches is a important factor to avoid missing diseases, therefore we consider evaluating the influence of different patch size. We divide each image into 2×2, 3×3, 4×4 patches, respectively. Considering the computational complexity and the receptive field, we do not verify the smaller patch size. The result is shown in Table VI.

7) Influence of Different Patch Order: To evaluate the influence of different patch order, we compare the different orders of the patch feature squence in Table VII. We observe that the performances for the different fixed orders are almost the same. The certain different order of the patch list does not matter for the final prediction. The most surprising aspect of the data is that the random unfixed order is much better than others. A possible explanation for this might be that the plant

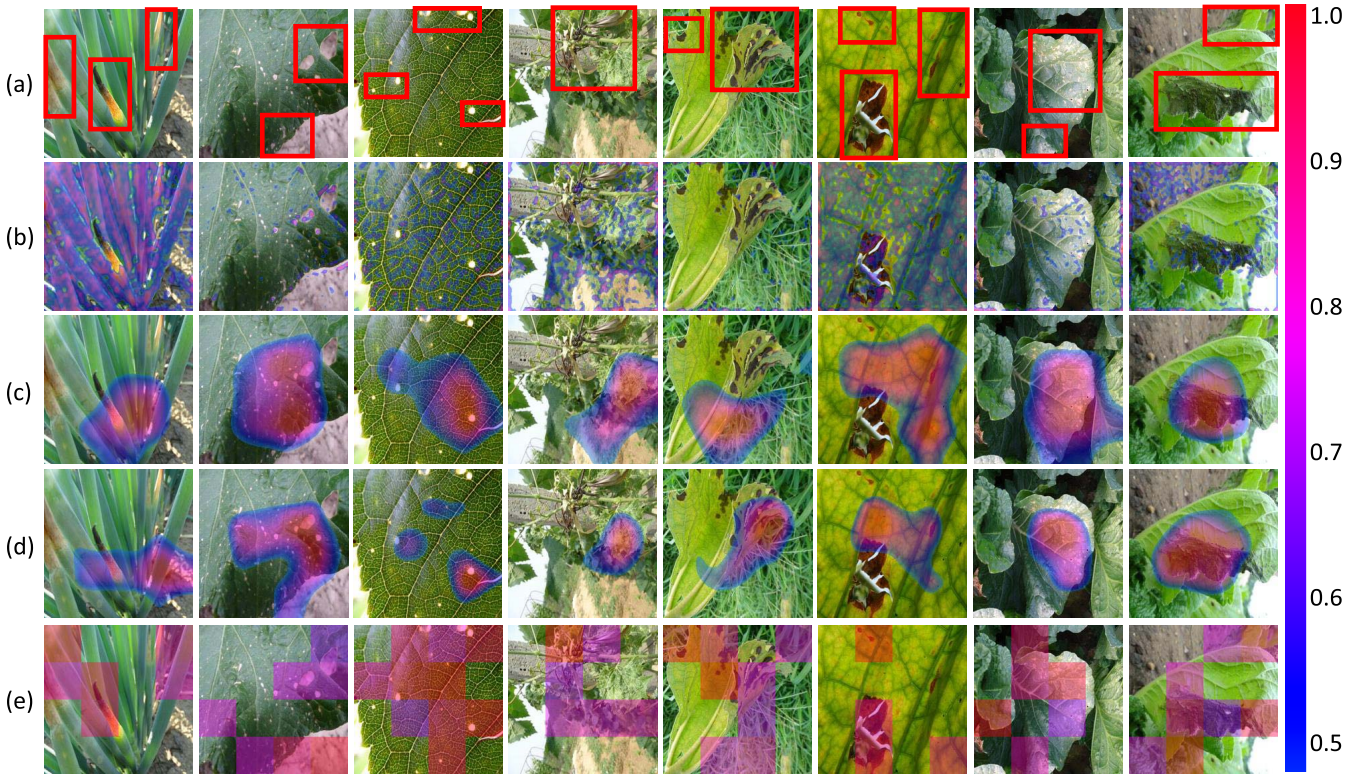


Fig. 10. Qualitative results. From top to bottom, (a) the original image with annotating diseased parts by red boxes, (b) the feature map from the last convolution layer of VGG16, (c) the feature map from the last convolution layer of ResNet152, (d) the feature map from the last convolution layer of SeNet154, (e) visualisation of the proposed CRR weights for each patch. The red means high weights and the blue means relatively low weights. For the best view, we only visualize the weights which are bigger than 0.75. CRR can consider more regions and obtain more characteristics.

TABLE VII

IMPACT OF ORDERS. THE 'T', 'B', 'L', AND 'R' DENOTE THE TOP, BOTTOM, LEFT AND RIGHT, RESPECTIVELY. THE 'T2B,L2R' MEANS THE PATCHES FROM EACH IMAGE ARE ORDERED FROM TOP TO BOTTOM AND LEFT TO RIGHT. THE 'RD' DENOTES THE RANDOM ORDER. THE 'FIXED' DENOTES THAT THE ORDER OF THE PATCH LIST FOR EACH IMAGE IS FIXED, AND THE 'UNFIXED' DENOTES THAT THE ORDER OF THE PATCH LIST FOR EACH IMAGE IS UNFIXED

Method	Validate(%)	Test(%)
t2b,l2r,fixed	90.01	85.58
l2r,t2b,fixed	90.01	85.48
t2b,r2l,fixed	90.01	85.62
r2l,t2b,fixed	90.01	85.66
rd,fixed	90.02	85.62
rd,unfixed	**90.10**	**85.70**

is diseased no matter where the lesions appear in. Another possible explanation is that the uncertain order is likely to enhance the power of networks.

8) Visualization: We visualize different emphasized parts in different methods via gradient-weighted class activation heatmap [49]. Fig. 10 shows the visualization results of some typical deep architectures, such as VGG16 and ResNet152. The reweighted maps of the proposed cluster-based region reweighting strategy are shown in Fig. 10 (d), where we only visualize the weight of the patch x when $w_x \geq 0.75$.

Compared with feature maps from typical deep networks, we can find that the proposed reweighted maps can cover more discriminative regions. The VGG16 and ResNet152 probably focus on disease-irrelevant regions, and meanwhile ignores some useful information. Our approach can pay attention to multiple scattered regions, which is more appropriate for plant disease recognition. The visualization results of the PDD271 further demonstrate the effectiveness of the proposed cluster-based reweighting strategy.

In addition, we further show the confusion matrix of our method on the PDD271 in Fig. 11, where the vertical axis shows the ground-truth classes and the horizontal axis shows the predicted classes. Yellower colors indicate better performance. We can see that our method still does not provide perfect performance for some plant disease categories. We enlarge specific regions to highlight the misclassified results and show some samples from confused categories. We can see that these plant disease categories are very similar in visual appearance and texture. Even the humans do not easily distinguish among these disease categories. The probable solution is to design more fine-grained visual feature learning methods or use multi-source information from different sensors to classify these plant disease categories.

C. Experiment on PlantVillage Dataset

Besides the PDD271, we also conduct the evaluation on another publicly available benchmark datasets, the

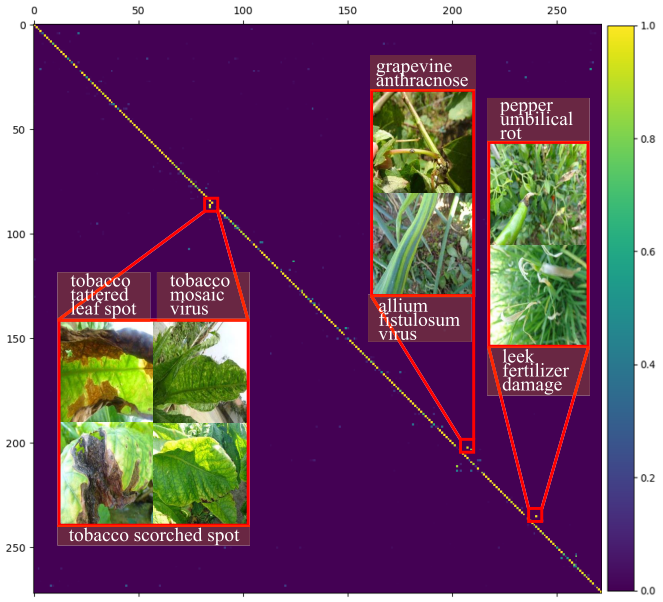


Fig. 11. Confusion matrix of our method on the PDD271.

TABLE VIII
PERFORMANCE COMPARISON OF METHODS ON PLANTVILLAGE

Method	Test(%)
AlexNet [5]	99.24
DenseNet169 [45]	98.89
GoogleNet [46]	99.76
ResNet34 [43]	99.67
ResNet152 [43]	99.69
VGG13 [42]	99.49
Squeezenet [50]	99.20
Our Method(ResNet152)	**99.78**

PlantVillage dataset, to further verify the effectiveness of our method. The PlantVillage dataset contains 38 plant disease categories and a total of 54,309 images. It is split following the setup in [12], 80% of the dataset is used for training and 20% for validation. All the methods have good performance as shown in Table VIII, because all the images are shot on the table.

VI. CONCLUSION AND FUTURE WORK

Plant disease recognition is an interesting and practical topic. However, this problem has not been sufficiently explored due to the lack of systematical investigation and large-scale dataset. The most challenging step in constructing such a dataset is providing a reasonable structure from both the agriculture and image processing perspective.

In this paper, we systematically investigate the problem of plant disease recognition in the community of image processing. With the help of agriculture experts, we construct the first large-scale plant disease dataset with 271 plant disease categories and 220,592 images. Furthermore, we present a plant disease oriented framework for plant disease recognition based on their distinctive characteristics. We design a strategy to compute patch weights based on the cluster distribution of patch features and then use learned weights to reweight both patch features to highlight diseased patches and the loss to guide the model optimization. Qualitative and quantitative evaluations on the PDD271 and PlantVillage datasets demonstrate the effectiveness of the proposed method. Nevertheless, a limitation of our method is that the proposed method is a little slow due to adding the clustering process before training. We will try to accelerate our method in the future. Another interesting work is analyzing the impact of the random unfixed order of patches. The random unfixed order enhances the performance, which may seem counterintuitive at first glance. Additionally, we can further consider a more advanced variant of LSTM as the alternative to the conventional LSTM, such as SFMRNN [51] and H-LSTCM [52]. Besides, a further study of the imbalanced problem between disease and healthy classes of image patches could assess the long-term effects. Hard sample mining [53] and hard triplet generating [54] could be more efficient and accurate.

The study on the visual plant disease recognition is still at the initial stage. How to discover discriminative diseased regions more efficiently and accurately remains an open question for further investigation.

REFERENCES

[1] Z. Li *et al.*, "Non-invasive plant disease diagnostics enabled by smartphone-based fingerprinting of leaf volatiles," *Nature Plants*, vol. 5, no. 8, pp. 856–866, Aug. 2019.
[2] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017.
[3] W. Min, S. Jiang, L. Liu, Y. Rui, and R. Jain, "A survey on food computing," *ACM Comput. Surv.*, vol. 52, no. 5, pp. 92:1–92:36, 2019.
[4] E. Moen, D. Bannon, T. Kudo, W. Graf, M. Covert, and D. Van Valen, "Deep learning for cellular image analysis," *Nature Methods*, vol. 16, no. 12, pp. 1233–1246, 2019.
[5] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Neural Inf. Process. Syst.*, 2012, pp. 1106–1114.
[6] A. Bauer *et al.*, "Combining computer vision and deep learning to enable ultra-scale aerial phenotyping and precision agriculture: A case study of lettuce production," *Horticulture Res.*, vol. 6, no. 1, pp. 1–12, Dec. 2019.
[7] S. Sladojevic, M. Arsenovic, A. Anderla, D. Culibrk, and D. Stefanovic, "Deep neural networks based recognition of plant diseases by leaf image classification," *Comput. Intell. Neurosci.*, vol. 2016, pp. 1–11, May 2016.
[8] J. Wang, L. Chen, J. Zhang, Y. Yuan, M. Li, and W. Zeng, "CNN transfer learning for automatic image-based classification of crop disease," in *Image and Graphics Technologies and Applications*. Beijing, China: Springer, 2018, pp. 319–329.
[9] K. P. Ferentinos, "Deep learning models for plant disease detection and diagnosis," *Comput. Electron. Agricult.*, vol. 145, pp. 311–318, Feb. 2018.
[10] G. Wang, Y. Sun, and J. Wang, "Automatic image-based plant disease severity estimation using deep learning," *Comput. Intell. Neurosci.*, vol. 2017, pp. 1–8, Jul. 2017.
[11] M. RuBwurm and M. Korner, "Temporal vegetation modelling using long short-term memory networks for crop identification from medium-resolution multi-spectral satellite images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 11–19.
[12] M. Brahimi, M. Arsenovic, S. Laraba, S. Sladojevic, K. Boukhalfa, and A. Moussaoui, "Deep learning for plant diseases: Detection and saliency map visualisation," in *Human and Machine Learning*. Cham, Switzerland: Springer, 2018, pp. 93–117.
[13] W. Ge, X. Lin, and Y. Yu, "Weakly supervised complementary parts models for fine-grained image classification from the bottom up," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 3034–3043.
[14] A. J. Wakeham, G. Keane, and R. Kennedy, "Field evaluation of a competitive lateral-flow assay for detection of alternaria brassicae in vegetable Brassica crops," *Plant Disease*, vol. 100, no. 9, pp. 1831–1839, Sep. 2016.

[15] A. K. Lees, L. Sullivan, J. S. Lynott, and D. W. Cullen, "Development of a quantitative real-time PCR assay for phytophthora infestans and its applicability to leaf, tuber and soil samples," *Plant Pathol.*, vol. 61, no. 5, pp. 867–876, Oct. 2012.

[16] C. H. Bock, G. H. Poole, P. E. Parker, and T. R. Gottwald, "Plant disease severity estimated visually, by digital photography and image analysis, and by hyperspectral imaging," *Crit. Rev. Plant Sci.*, vol. 29, no. 2, pp. 59–107, Mar. 2010.

[17] F. Ahmad and A. Airuddin, "Leaf lesion detection method using artificial bee colony algorithm," in *Advances in Computer Science and its Applications*, vol. 279. Beijing, China: Springer, 2014, pp. 989–995.

[18] S. Prasad, P. Kumar, and A. Jain, "Detection of disease using block-based unsupervised natural plant leaf color image segmentation," in *Swarm, Evolutionary, and Memetic Computing*. Beijing, China: Springer, 2011, pp. 399–406.

[19] A. Ramcharan, K. Baranowski, P. Mcclowsky, B. Ahmed, and D. P. Hughes, "Using transfer learning for image-based cassava disease detection," *Frontiers Plant Sci.*, vol. 8, p. 1852, Oct. 2017.

[20] E. Mwebaze, T. Gebru, A. Frome, S. Nsumba, and J. Tusubira, "iCassava 2019 fine-grained visual categorization challenge," 2019, *arXiv:1908.02900*. [Online]. Available: https://arxiv.org/abs/1908.02900

[21] D. P. Hughes and M. Salathé, "An open access repository of images on plant health to enable the development of mobile disease diagnostics through machine learning and crowdsourcing," 2015, *arXiv:1511.08060*. [Online]. Available: https://arxiv.org/abs/1511.08060

[22] H. Yu and C. Son, "Apple leaf disease identification through region-of-interest-aware deep convolutional neural network," 2019, *arXiv:1903.10356*. [Online]. Available: https://arxiv.org/abs/1903.10356

[23] C. Xie *et al.*, "Multi-level learning features for automatic classification of field crop pests," *Comput. Electron. Agricult.*, vol. 152, pp. 233–241, Sep. 2018.

[24] X. Wu, C. Zhan, Y.-K. Lai, M.-M. Cheng, and J. Yang, "IP102: A large-scale benchmark dataset for insect pest recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 8787–8796.

[25] J. Donahue *et al.*, "Decaf: A deep convolutional activation feature for generic visual recognition," in *Proc. Int. Conf. Mach. Learn.*, vol. 2014, pp. 647–655.

[26] N. Zhang, J. Donahue, R. Girshick, and T. Darrell, "Part-based R-CNNs for fine-grained category detection," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 834–849.

[27] T. Xiao, Y. Xu, K. Yang, J. Zhang, Y. Peng, and Z. Zhang, "The application of two-level attention models in deep convolutional neural network for fine-grained image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 842–850.

[28] M. Lam, B. Mahasseni, and S. Todorovic, "Fine-grained recognition as HSnet search for informative image parts," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6497–6506.

[29] T.-Y. Lin, A. RoyChowdhury, and S. Maji, "Bilinear convolutional neural networks for fine-grained visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 6, pp. 1309–1322, Jun. 2018.

[30] S. Jiang, W. Min, L. Liu, and Z. Luo, "Multi-scale multi-view deep feature aggregation for food recognition," *IEEE Trans. Image Process.*, vol. 29, pp. 265–276, 2020.

[31] W. Min, L. Liu, Z. Luo, and S. Jiang, "Ingredient-guided cascaded multi-attention network for food recognition," in *Proc. 27th ACM Int. Conf. Multimedia*, Oct. 2019, pp. 99–107.

[32] Z. Yang, T. Luo, D. Wang, Z. Hu, J. Gao, and L. Wang, "Learning to navigate for fine-grained classification," in *Proc. Eur. Conf. Comput. Vis.*, Sep. 2018, pp. 438–454.

[33] T. Hu and H. Qi, "See better before looking closer: Weakly supervised data augmentation network for fine-grained visual classification," *CoRR*, vol. abs/1901.09891, 2019.

[34] Y. Peng, X. He, and J. Zhao, "Object-part attention model for fine-grained image classification," *IEEE Trans. Image Process.*, vol. 27, no. 3, pp. 1487–1500, Mar. 2018.

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

[36] G.-J. Qi, "Hierarchically gated deep networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2267–2275.

[37] G.-J. Qi, X.-S. Hua, Y. Rui, J. Tang, and H.-J. Zhang, "Image classification with kernelized spatial-context," *IEEE Trans. Multimedia*, vol. 12, no. 4, pp. 278–287, Jun. 2010.

[38] W. Lin *et al.*, "Human in events: A large-scale benchmark for human-centric video analysis in complex events," 2020, *arXiv:2005.04490*. [Online]. Available: https://arxiv.org/abs/2005.04490

[39] S. Li, J. Li, H. Tang, R. Qian, and W. Lin, "ATRW: A benchmark for Amur tiger re-identification in the wild," in *Proc. 28th ACM Int. Conf. Multimedia*, Oct. 2020, pp. 2590–2598.

[40] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.

[41] Y. Chen, Y. Bai, W. Zhang, and T. Mei, "Destruction and construction learning for fine-grained image recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 5157–5166.

[42] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Represent.*, 2015, pp. 1–14.

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

[44] S. Zagoruyko and N. Komodakis, "Wide residual networks," in *Proc. Brit. Mach. Vis. Conf.*, 2016, p. 87.

[45] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2261–2269.

[46] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2818–2826.

[47] X. Zhang, Z. Li, C. C. Loy, and D. Lin, "PolyNet: A pursuit of structural diversity in very deep networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3900–3908.

[48] C. Liu *et al.*, "Progressive neural architecture search," in *Proc. Eur. Conf. Comput. Vis.*, Sep. 2018, pp. 19–35.

[49] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 618–626.

[50] F. N. Iandola, M. W. Moskewicz, K. Ashraf, S. Han, W. J. Dally, and K. Keutzer, "Squeezenet: Alexnet-level accuracy with 50x fewer parameters and 1mb model size," 2016, *arXiv:1602.07360*. [Online]. Available: https://arxiv.org/abs/1602.07360

[51] H. Hu and G. Qi, "State-frequency memory recurrent neural networks," in *Proc. Int. Conf. Mach. Learn.*, vol. 70, 2017, pp. 1568–1577.

[52] X. Shu, J. Tang, G. Qi, W. Liu, and J. Yang, "Hierarchical long short-term concurrent memory for human interaction recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, Sep. 17, 2018, doi: 10.1109/TPAMI.2019.2942030.

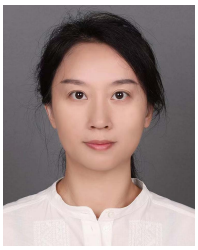
[53] K. Chen, Y. Chen, C. Han, N. Sang, and C. Gao, "Hard sample mining makes person re-identification more efficient and accurate," *Neurocomputing*, vol. 382, pp. 259–267, Mar. 2020.

[54] Y. Zhao, Z. Jin, G. Qi, H. Lu, and X. Hua, "An adversarial approach to hard triplet generation," in *Proc. Eur. Conf. Comput. Vis.*, vol. 11213, Sep. 2018, pp. 508–524.

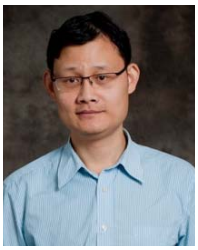
Xinda Liu received the B.E. degree from the China University of Mining and Technology, Beijing, China, in 2013, and the M.E. degree from Ningxia University, Yinchuan, China, in 2016. He is currently pursuing the Ph.D. degree with the State Key Laboratory of Virtual Reality Technology and Systems, School of Computer Science and Engineering, Beihang University, Beijing. He is also with the Peng Cheng Laboratory. His research interests include machine learning and image processing.

Weiqing Min (Member, IEEE) received the B.E. degree from Shandong Normal University, Jinan, China, in 2008, the M.E. degree from Wuhan University, Wuhan, China, in 2010, and the Ph.D. degree from the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, in 2015. He is currently an Associate Professor with the Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences. His current research interests include multimedia content analysis, understanding and applications, food computing, and geo-multimedia computing. He has authored or coauthored more than 40 peer-referenced papers in relevant journals and conferences, including *ACM Computing Surveys*, IEEE TRANSACTIONS ON IMAGE PROCESSING, IEEE TRANSACTIONS ON MULTIMEDIA, ACM TOMM, *IEEE Multimedia Magazine*, ACM Multimedia, AAAI, and IJCAI. He organized several special issues on international journals, such as *IEEE Multimedia Magazine*, *Multimedia Tools and Applications*, as a Guest Editor. He served as a TPC member of many academic conferences, including ACM MM, AAAI, and IJCAI. He was a recipient of the 2016 ACM TOMM Nicolas D. Georganas Best Paper Award and the 2017 *IEEE Multimedia Magazine* Best Paper Award.

Shuhuan Mei received the M.E. degree from the Shandong University of Science and Technology, China. His current research interests include multimedia content analysis, understanding, and applications.



Lili Wang (Member, IEEE) received the Ph.D. degree from Beihang University, Beijing, China. She is currently a Professor with the School of Computer Science and Engineering, Beihang University, where she is also a Researcher with the State Key Laboratory of Virtual Reality Technology and Systems. She is also with the Beijing Advanced Innovation Center for Biomedical Engineering. Her research interests include virtual reality, augmented reality, mixed reality, real-time, and realistic rendering.



Shuqiang Jiang (Senior Member, IEEE) is currently a Professor with the Institute of Computing Technology, Chinese Academy of Sciences (CAS), Beijing, and a Professor with University of Chinese Academy of Sciences. He is also with the Key Laboratory of Intelligent Information Processing, CAS. His research interests include multimedia processing and semantic understanding, pattern recognition, and computer vision. He has authored or coauthored more than 150 articles in the related research topics. He was supported by the New-Star Program of Science and Technology of Beijing Metropolis in 2008, the NSFC Excellent Young Scientists Fund in 2013, and the Young top-notch talent of Ten Thousand Talent Program in 2014. He has also served as a TPC member for more than 20 well-known conferences, including ACM Multimedia, CVPR, ICCV, IJCAI, AAAI, ICME, ICIP, and PCM. He won the Lu Jiaxi Young Talent Award from Chinese Academy of Sciences in 2012, and the CCF Award of Science and Technology in 2012. He is the Senior Member of CCF, a member of ACM, an Associate Editor of ACM TOMM, IEEE MULTIMEDIA, and *Multimedia Tools and Applications*. He is the Vice Chair of IEEE CASS Beijing Chapter and ACM SIGMM China chapter. He is the General Chair of ICIMCS 2015, the Program Chair of ACM Multimedia Asia2019 and PCM2017.