Influencing factors analysis in pear disease recognition using deep learning

Fang Yang¹ · Fuzhong Li¹ · Kai Zhang¹ · Wuping Zhang¹ · Shancang Li²

Received: 6 August 2020 / Accepted: 23 November 2020 / Published online: 11 December 2020
© The Author(s) 2020

Abstract

Influencing factors analysis plays an important role in plant disease identification. This paper explores the key influencing factors and severity recognition of pear diseases using deep learning based on our established pear disease database (PDD2018), which contains 4944 pieces of diseased leaves. Using the deep learning neural networks, including VGG16, Inception V3, ResNet50 and ResNet101, we developed a “DL network + resolution” scheme that can be used in influencing factors analysis and diseases recognition at six different levels. The experimental results demonstrated that the resolution is directly proportional to disease recognition accuracy and training time and the recognition accuracies for pear diseases are up to 99.44%, 98.43%, and 97.67% for *Septoria piricola* (SP), *Alternaria alternate* (AA), and *Gymnosporangium haracannum* (GYM), respectively. The results also shown that a forward suggestion on disease sample collection can significantly reduce the false recognition accuracy.

Keywords Pear diseases dataset · Disease recognition · Deep learning (DL) · Transfer learning · Resolution

1 Introduction

Pear cultivation has a history of more than 3000 years in China, and it is planting in a large area covering 80% of area in the world [1]. Pear diseases and their contagiousness can significantly affect the normal growth of pear trees. As a result, scientific diagnosis measures are of crucial important to avoid misuse of prescriptions, excessive application of pesticides, pesticide residues, can cause significant reduction of yield of pear etc. Also, it can increase the cost of disease prevention and control and reduction of economic benefits together with the enthusiasm of farmers. Even worse, it can cause food safety problems [1]. With the concerns of experts in agricultural technology promotion and the increasingly labor

costs, it is difficult and expensive for professionals to help farmers diagnose diseases and there are urgent demands for the disease automatic recognition technology that could detect, identify and possibly cure the diseases of pear trees.

Plant diseases recognition based on leaf lesions images is a promising method that has been widely studied and successfully applied for fruits, vegetables, and crops [2–4]. It is low-cost, simple, and convenient way compare to molecules, volatile organic compounds, and spectrum methods. At present, there are mainly two different technical routes for researching plant diseases: 1) traditional image recognition processing technology, which is based on a small dataset, and does not require large numbers of disease samples, but include image pre-processing, disease segmentation, feature extraction, classifier construction, model building, etc. have so many manual processing steps. This technology is of subjective, laborious and error-prone [5]; 2) Deep learning (DL) technology another promising technology that uses DL network models to automatically extract and identify disease features based on a large dataset. There are fewer manual processing steps, but great demand for the number of disease samples [6].

Deep learning is proven a promising way in object recognition [7], and a number of datasets have been established, including ILSVRC (ImageNet Large Scale Visual Recognition Challenge) [8]. The DL technology has made remarkable achievements in processing images, speech recognition, and text recognition, and has been widely used in many

This article is part of the Topical Collection: *Special Issue on P2P Computing for Deep Learning*
Guest Editors: Ying Li, R.K. Shyamasundar, Yuyu Yin, Mohammad S. Obaidat

✉ Fuzhong Li
lifuzhong@sxau.edu.cn

✉ Shancang Li
shancang.li@ieee.org

¹ School of Software, Shanxi Agricultural University, Jinzhong 030801, Shanxi, China

² Department of Computer Science, University of the West of England, Bristol BS16 1QY, UK

areas like medical, industrial and other fields. A strong advantage of deep learning is feature learning, i.e. the automatic feature extraction from raw data, with features from higher levels of the hierarchy being formed by the composition of lower level features [9, 10], the process is transparent to users. Therefore, the paper makes full use of the end-to-end characteristics of DL [11, 12], for reflecting Occam's razor law that is simple and effective, minimizes manual intervention, expresses image features in the original way, and explores the factors that affect DL technology to diagnose pear diseases.

The major contributions of this work are summarised as follows:

- 1) In this work, a pear disease database PDD2018 is established that contains more than 7000 pieces of diseased leaves (including front and back);
- 2) This work investigated the influence of disease image resolution specification variation on training time, cross-entropy loss, disease recognition accuracy and other indicators of different DL network models. It is the first time to put forward a simple and convenient "model + resolution" combination mode which is suitable for the recognition of pear disease;
- 3) Using ResNet50 and ResNet101 models, this work investigated the influence of the similarity between diseases classes, number of DL network layers, setting of epoch parameters on disease recognition accuracy.

The remainder of this paper is organized as follows: in Section 2, materials and methods are introduced. Experiments are presented in Section 3. Results and discussion are reported in Section 4. Finally, the conclusions are drawn in Section 5.

2 Related works

Dataset quality is critical in many machine learning (ML) or artificial intelligence (AI) researches. For plant diseases identification using DL, many research efforts focus more on network models or classification algorithms, rather than on optimisation of the extraction of artificial disease features [13, 14]. While ignoring the importance of plant disease dataset itself, such as precautions for dataset construction, and image resolution pre-processing specification, etc. which may affect the DL technology apply for disease recognition. This may be related to the difficulty and cost of disease samples collection, most researchers used public disease dataset such as *PlantVillage* [15], *LifeCLEF* [16], *MalayaKew* [17], *UC Merced* [18] and *Flavia* [19]. Barbedo explored number of dataset samples, difficulty of labeling dataset, manifestation of disease symptoms, image background, image capture

conditions, multiple simultaneous and other factors which may affect DL apply for disease recognition [20]. However, this work failed on conducting experiments to analyze which types of disease samples are easy to identify errors, or summarize the common points of identifying the wrong samples, and to study how to construct a high-quality disease dataset.

However, limited research efforts have been done on fruit tree diseases [21, 22], such as Park et al. focused on the source of disease dataset. Actually, in the past few years, a number of new tree disease datasets have been set up [14, 18, 23]. However, most disease datasets focus more on the front side of leaf, and ignore the back side, which may affect the promotion of DL technology, because observing the lesion on the back side of leaf is a conventional auxiliary means for diagnosing disease in practice.

Regardless of whether it is based on a public disease dataset or self-built, many studies specified the image resolution of disease dataset which inputted to DL network model as model default resolution. There is very little literature attempt to change the disease image resolution, that is, to pre-process the same dataset to different resolution specifications, and explore the resolution variation impact on performance of deep neural network (DNN) model (such as train time, cross-entropy loss, recognition accuracy, etc.). Kerkech [21] used the LeNet-5 network [24] to diagnose vine disease with a sliding window segmentation method, a 4608×3456 pixels image was divided into three resolution specifications (16×16 pixels, 32×32 pixels, 64×64 pixels) to form three different disease sample content datasets. It is worth noting that the three datasets with different contents are not different resolution specifications of the same dataset.

In [14], Lu et al. proposed an accurate wheat disease recognition scheme using VGG network by comparing the FCN (Fully Convolutional Network) and CNN (Convolutional Neural Network), in which the FCN method used 832×832 pixels resolution combined with MIL (Multiple instance learning), BBA (a bounding box approximation) for disease location and identification, while CNN method was based on 224×224 pixels resolution.

3 Proposed method

3.1 Pear disease database

As far as we know, the disease in early and middle stages of pear leaves is mainly a single disease and the leaves are susceptible to mixed diseases only in the later growing stages of the pear tree or when the farmers have abandoned management. In this work, we will focus on a single pear disease from the perspective of plant protection and timely treatment. Some disease datasets, such as [15, 25], omit lesion segmentation and lesion labeling, which can reduce the amount of image

pre-processing calculations and labor costs, and meet the experience needs of users in the Internet era for rapid system identification and response [26].

Since 2018, we formed a profession team in Plant Protection Institute of the Shanxi Academy of Agricultural Sciences and Plant Protection Station of the Shanxi Provincial Department of Agriculture, collected diseased and healthy pear leaves with front and back. After carefully identification and classification, we selected effective images and established a pear diseases database, namely PDD2018. Imaging devices involved include *NIKON D700*, *Canon PowerShot G9 X Mark II*, *Canon PowerShot G7 X Mark II*, *SONY ILCE6000* in both field and indoor scenarios. The distribution of disease collection points is as follows: Yuci (112.72, 37.68), Taigu (112.53, 37.42), Xi County (110.93, 36.7), Yanhu District (110.97, 35.03), Pinglu County (111.20, 34.12), Wanrong County (110.83, 35.42), Ruicheng County (110.68, 34.71). The collected pears varieties include representative with large planting areas such as crisp pear, Yulouxiang pear, Bartlett (*Pyrus communis*), Red Bartlett (*Pyrus communis* L.), and Xiaobai pear. We mainly focused on and collected three common diseases: *Septoria piricola* (SP), *Alternaria alternate* (AA), and *Gymnosporangium haracannum* (GYM). Only a few other diseases such as powdery mildew, dry rot, and coal pollution are rare, have been collected (no more than 10 photos taken per class).

In order to minimize the number of negative samples and establish a high-quality pear leaf disease dataset, we have invited 5 fruit trees and plant protection experts to conduct a preliminary review and classification at the collection site in first round, and a detailed review and classification in the laboratory in second round, this process eliminated about 2300 negative diseased leaf samples and remained 4944, which constituted the basic database of pear disease, we called Pear Disease Database 2018 (PDD2018) as shown in Table 1 and Fig. 1. It is estimated that the cost of each sample is about 30p (£0.3).

The PDD2018 has characteristics of diverse environment, strong regionality, rich variety, different growth stages, and attention to the leaf spots on the back side. These characteristics are significantly different from other plant disease datasets

Fig. 1 L-F with SP disease, L-B with SP disease, L-F with AA disease, L-B with AA disease

used by other researchers. To the best of our knowledge, no one has established such a pear disease dataset, which is one of the contribution points of this work.

3.2 Proposed DL methods

DL network models It is not easy to train a DL network model from scratch requires massive samples. Techniques, such as data augmentation and GAN network are widely used to enhance the dataset with massive samples. The data augmentation [27–31] is used to expand the number of samples in the dataset, such as *flipping*, *jittering*, *rotating*, *noise*, *changing image attributes*, etc., or GAN networks are used to generate pseudo samples to transform small dataset into large dataset [32, 33]. The other is *transfer learning*, utilizing the adjusted and optimized weight parameters of DL network models that have been trained on large dataset to retrain on another small similar dataset [18, 34–36]. The two techniques can solve the problem of insufficient samples and achieve good results.

Because the number of labeled pear disease samples we collected is not sufficient to support learning from scratch, in this work we use a feature-based supervised transfer learning, and selected DL network models which are similar to our classification task and have been trained on large dataset, such as *VGG16*, *Inception V3*, *ResNet50*, *ResNet101 models trained on ImageNet*, loading DL network models and their optimized network weight parameters of each layer, and fine-tuning their fully connected layers. Many studies indicate that transfer learning could get good results without requiring a large number of samples [18].

Table 1 Pear disease dataset 2018

	SP	AA	GYM	sum
indoor front	1070	863	435	2368
indoor back	707	738	431	1876
field front	90	189	154	433
field back	39	87	141	267
Sum	1906	1877	1161	4944

In order to solve the overfitting problem that may cause by a limited number of disease samples and improve the generalization ability of DNN network model, we expanded the number of disease samples using data augmentation techniques [8, 20, 21, 33, 37]. In the experiment, we used online data augmentation methods, which mainly include random cropping, center cropping, and random horizontal flipping [30].

VGG16 is a version of VGG architecture [38] which is CNN model developed by the Visual Geometry Group from the University of Oxford, it consists of 13 convolution layers with the kernel size 3×3 , each layer followed by ReLU layer which increases the non-linearity, some of the convolution layers are followed by max-pooling to reduce the dimension. Additionally, two fully-connected (FC) layers each with 4096 nodes and a soft-max classifier.

Inception V3 is trained from the ImageNet Large Scale Visual Recognition Challenge (ILSVRC), It is the subsequent manifestations of GoogLeNet architecture [39], it extends the original GoogLeNet implementation and enhances the Inception Module to improve the accuracy by factorisation of convolutions and improved normalization [40]. Inception V3 factorizes the traditional 7×7 convolution into three 3×3 convolutions, grid reduction is applied to three traditional inception modules to reduce to a 17×17 grid with 768 filters, then grid reduction is applied again to five factorized inception modules to reduce to a $8 \times 8 \times 1280$ grid.

The ResNet architecture won the 1st place on with error rate of 3.57% in the ILSVRC 2015 challenge [41]. The most important difference between ResNet and a plain CNN is the addition of an identity connection to the underlying network element, this connection makes it possible to train hundreds or more layers while achieving enhanced performance [42]. And it is developed with many different numbers of layers including 50 and 101, ResNet50 and ResNet101 contain, respectively, 50 and 101 convolutional layers including one FC layer at the end of the network [43, 44].

The four models *VGG16*, *Inception V3*, *ResNet50*, *ResNet101* were created and loaded with pre-trained weights. Additionally, a fine-tuning method is proposed for the four models by truncating the original soft-max layer and replace it with our own according to different experimental objectives of the paper, i.e., the number of pear disease class.

Image Pre-processing The disease image resolution in PDD2018 mainly has three specifications: 3008×1688 pixels, 4256×2832 pixels, 5472×3648 pixels. The first step is offline image pre-processing, we used Python programming to subsample, reducing the disease images resolution to 1200×1200 pixels. The second step is online image pre-processing during model train, validation and test, we converted the input disease images (which would be fed into DL network model) resolution into different specifications

according to experimental goals. The last step is to normalize and standardize the disease images.

Experiment Setup In the experiment, all initial weight parameters of the four DL network models inherited parameters of their pre-trained models respectively. And the hyper-parameters used to train the DL network model were the following: Base Learning Rate is 0.001; Weight Decay is $1e-5$; Mini Batch Size is 16; and Number of Epochs is set as 100. Cross-entropy loss, recognition accuracy, and training time are mainly used as evaluation indicators. The optimal network model preservation strategy is to select the model with the highest recognition accuracy after the recognition accuracy curve is smooth.

4 Experiments

In order to investigate the impact of input disease images resolution variation on the performance of DL network model, we carried out model train, validation and test experiments under the same conditions, i.e., disease dataset, hardware and software resources, model train parameters and hyper-parameters. Firstly, we studied the influence of different resolution specification on the same DL network model. Then, we compared the performance of different DL network models under the same resolution specification. Thirdly, a combination mode of “DL network model+ resolution” which is more suitable for pear leaf disease recognition was proposed, it could provide valuable reference for similar research.

4.1 DL network models

This paper used SP, AA, GYM indoor disease samples as experimental objects. After removing obviously duplicate samples of PDD2018, the remaining 4226 constitute the PDIRE dataset, the proportion of train and validation sets is 8:2 based on experience [2, 6, 19, 20, 43], and were randomly selected from the PDIRE dataset. The details are shown in Table 2.

We used three classic DL network models such as VGG16, Inception V3 (In-V3), and ResNet50, and the experimental settings as shown above. The input disease images resolution has six types such as 224×224 pixels (except for Inception

Table 2 Develop PDIRE dataset

	SP	AA	GYM	sum
Train set	1421	1273	687	3381
Validation set	355	318	172	845
sum	1776	1591	859	4226

V3), 299×299 pixels (only for Inception V3), 448×448 pixels, 512×512 pixels, 600×600 pixels, 700×700 pixels. Unfortunately, all three models have memory overflow faults at 700×700 pixels resolution. We performed 12 experiments with three models under the above five resolution specifications. The experimental flow is shown in Fig. 2, in which the dataset PDIRE ($N = 3381$) was used for training and dataset PDIRE ($N = 845$) was used to validate the trained model.

The black dashed box on the left indicates train process, including online image pre-processing and data augmentation of the train disease dataset, and the disease images are input to the DL network model with modified resolution specifications for train. The green dashed box on the right indicates validation process, including online image preprocessing and data augmentation of the test disease dataset, and the disease image are input to the trained DL network model for validation.

The first step is down-sampling 3381 train samples of the PDIRE dataset to form a train set called PDIRE Dataset1 with a resolution of 1200×1200 pixels, as shown by the black arrow on the left in Fig. 2. The second step is performing online data augmentation and normalization on the PDIRE Dataset1 to form a train set called PDIRE Dataset2 with a preset resolution specification, as shown by the blue arrow on the left in Fig. 2. The third step is inputting the PDIRE Dataset2 into the DL network model with modified resolution specifications and starting first epoch train. The fourth step is performing similar operations to the first and second steps on the 845 validation samples of the PDIRE dataset to form a validation set called PDIRE Dataset2 and performing the first

epoch validation. The last step is repeating up to 100 epochs, and save the best DL network model.

4.2 Pear disease severity recognition

In order to realize the use of DL to identify the severity of pear diseases, and strengthen the guiding significance of this study on pear disease control production practice. We added 800 healthy pear leaves collected in 2019 on the basis of PDD2018. Because there are fewer early stage disease leaves collected, about 230, we randomly selected samples based on the minimum number of early stage disease leaves to form pear disease indoor severity recognition experiment (PDISRE) dataset, and balanced the number of train set, validation set and test set of each type to 180, 23 and 23 respectively, so that the proportion is close to 8:1:1. The labeled samples were graded from three perspectives, including early stage, middle and late stages, and health. So there are 7 types of *E_SP*, *ML SP*, *E_AA*, *ML AA*, *E_GYM*, *ML GYM* and *H* (*E*: Early stage, *ML*: Middle and late stages, *H*: Health). The details are shown in Table 3.

At the same time, we also carried out this experiment to study whether deepening the number of DL network layers is necessarily conducive to improving the disease recognition accuracy under such a condition, i.e., the number of PDISRE sets is less than 2000 and the number of seven types is absolutely balanced, i.e., comparing the characteristics of ResNet50 and ResNet101 for identification of diseases at various stages.

Fig. 2 The flow of DL network model training and validation

Table 3 Pear disease indoor severity recognition experiment (PDISRE) dataset

	E_SP	ML_SP	E_AA	ML_AA	E_GYM	ML_GYM	H	sum
Train set	180	180	180	180	180	180	180	1260
Validation set	23	23	23	23	23	23	23	161
Test set	23	23	23	23	23	23	23	161

The experimental settings are as follows: On one hand, we modified the final fully connected layer output number of ResNet50 and ResNet101 to seven, and retrained them on the PDISRE dataset. On the other hand, we unified the resolution of input disease images to 600×600 pixels according to the result of Experiment 3.1. The experimental process is as follows: Firstly, we used the similar process of Experiment 3.1 to train and validate ResNet50, ResNet101 respectively, and saved their respective optimal models. Secondly, we tested on PDISRE test set with ResNet50 and ResNet101 optimal models respectively.

5 Evaluation and discussion

5.1 Dataset images resolution factors

The highest recognition accuracy on validation set It can be clearly seen from Table 4 that the three models (VGG16, Inception V3, ResNet50) have a common feature, their highest recognition accuracy, recognition accuracy of 100th epoch, training time will rise with the input disease image resolution increasing. When the input disease image resolution is changed from 224×224 pixels (or 299×299 pixels), 448×448 pixels, 512×512 pixels to 600×600 pixels, the highest recognition accuracy of three models is improved as follows: VGG16 increased by 15.86%, 2.13%, 0.95%, Inception V3 increased by 13.37%, 1.9%, 1.54%, and Resnet50 increased by 22.37%, 1.18%, 1.18% respectively. This shows that under the condition of low resolution, greatly increasing the input disease image resolution can greatly improve the highest recognition accuracy. While after the resolution is increased to a certain level, such as 448×448 pixels,

and then continue to increase the resolution significantly, the highest recognition accuracy is only slightly increased.

Overfitting and large cross-entropy loss problems It can be clearly seen from Fig. 3 that overfitting occurs when the input disease image resolution is 224×224 pixels (or 299×299 pixels), e.g., the recognition accuracy of the three models on train set has remained above 99% since 20th epoch, such as the orange curve, while is less than 80% on validation set, such as the blue curve. In particular, the recognition accuracy of ResNet50 model on validation set is just over 60%.

It can be seen from Fig. 4 that when the input disease image resolution is increased to 448×448 pixels, the recognition accuracy of the three models on validation set is greatly improved, and the above-mentioned overfitting phenomenon disappears, such as 448×448 pixels orange curve. The recognition accuracy of the three models remains in the 90%–99% interval as a whole after the curve is smooth. Moreover, the input disease image resolution has a positive correlation with the recognition accuracy after training about 30th epoch. Intuitively, the red curve of 600×600 pixels is always above the green curve of 512×512 pixels, and the green curve of 512 pixels is always above the orange curve of 448 pixels. It shows that by appropriately increasing the resolution of the input disease image, the model overfitting problem can be solved, the recognition accuracy and the generalization ability of the model can be improved.

Similarly, the cross-entropy loss of the three models is large when the input disease image resolution is 224×224 pixels (or 299×299 pixels), the VGG16 loss value fluctuates in the interval [1, 3.1], the Inception V3 loss value fluctuates in the interval [1, 1.6], and the ResNet50 loss value fluctuates in the interval [1, 2.5]. While the loss value gradually returns to a

Table 4 Model and resolution combination comparison result on PDIRE dataset (train and validation).

Resolution (pixels)	The highest recognition accuracy (%)			Epochs of the highest recognition accuracy			The recognition accuracy of 100th epoch(%)			Training time (s)		
	VGG16	Inception V3	ResNet50	VGG16	Inception V3	ResNet50	VGG16	Inception V3	ResNet50	VGG16	Inception V3	ResNet50
224×224 (299×299)	78.46	81.3	74.32	9	24	25	64.02	76.21	59.05	3982	3651	2600
448×448	94.32	94.67	96.69	92	54	62,97	93.49	93.73	95.74	12,249	6913	7607
512×512	96.45	96.57	97.87	64	76	64,89	95.86	95.74	97.04	14,918	8815	9368
600×600	97.4	98	99.05	20,24,59,60,70,86,92,95	19,59,64,72,85,95,97	66	97.04	97.16	98.46	22,446	12,046	14,959

Fig. 3 The three models's disease recognition accuracy variation during train and validation under 224×224 pixels resolution (the input disease image)

reasonable value range as the resolution increases, and the larger resolution, the smaller loss value. Intuitively, the 600×600 pixels red loss curve of the three models remained stable below the other curves after 30th epoch.

Memory Requirements When the input disease image resolution is increased to 700×700 pixels, unfortunately, all three models suffer from memory overflow problems during the training process, which makes it impossible to continue train. It shows that increasing the input disease image resolution could improve the disease recognition accuracy, solve overfitting and reduce the cross-entropy value, but it should not be increased too much, otherwise it will cause memory overflow. At the same time, as far as we know, there are few literatures using resolutions above 600×600 pixels to study

plant diseases, only Lu [14] used the FCN basing on 832×832 pixels resolution, and Metin [45] used the Faster R-CNN basing on 600×600 pixels resolution.

Training Times As can be seen from Table 4, the training time of three models is directly proportional to the input disease image resolution specification. Additionally, when the input disease image resolution is 224×224 pixels (or 299×299 pixels), the three models have little difference in training time and descending order is VGG16, Inception V3, and ResNet50. In the other three resolutions, the three models have larger difference in training time and descending order is VGG16, ResNet50, and Inception V3. So, when the input disease image resolution is same, VGG16 is the most time-consuming of the three models. For example, at 600×600

Fig. 4 The model and resolution combination's disease recognition accuracy and cross-entropy loss variation during validation

pixels, VGG16 training takes about 6 h and 15 min, ResNet50 training takes about 4 h and 9 min, and Inception V3 training takes about 3 h and 21 min.

5.2 Dataset factors

Image Preprocess As can be seen from Table 4, Table 5, Fig. 3 and Fig. 4, the results show that DL network model trained on a disease dataset similar to the paper's experimental (such as taking photos with a solid color paper as background) is not particularly sensitive to the image background and does not need to remove them [20]. At the same time, from the perspective of simplifying network model train process and highlighting the “end-to-end” idea, the entire disease sample leaf can be inputted into DL network model directly without lesion segmentation, lesion labeling and other operations [15, 18, 25], this will reduce manual intervention, save manpower, and give full play to the advantages of DL network technology (Fig. 5).

Disease samples collection An interesting phenomenon was found according to the results of model and resolution combination experiment. Taking the ResNet50 model to identify GYM as an example, (a) had 4 recognition errors, and (b), (c) had 3 recognition errors. Taking the ResNet50 model to identify SP as an example, (d), (e) had 4 recognition errors, and (f) had 3 recognition errors. Taking the ResNet50 model to identify AA as an example, (g), (h) had 4 recognition errors, and (i), (j), (k) had 3 recognition errors.

5.3 Deep learning network model

Train Epoch As can be seen from Table 4, the recognition accuracy of 100th epoch for three models is lower than the highest recognition accuracy under the conditions of four input image resolutions. Moreover, the speed of getting the highest recognition accuracy of the 12 model resolution combination modes are different, some are quickly, such as “VGG16 + 224 × 224” in the 9th epoch, and some are slowly, such as “VGG16 + 448 × 448” in the 92nd epoch, more modes are in the interval of [24, 70] epochs. This indicating

that the DL network model recognition accuracy is not positively related to the number of train epochs. i.e., it is not that the larger the epoch value is, the higher the recognition accuracy of the model is. The highest recognition accuracy of different DL network models is different for the setting of train epoch parameters [37, 46–48].

The best mode for disease recognition As shown in Table 5, as with the model train process, the disease recognition accuracy for three models at 224 × 224 pixels (or 299 × 299 pixels) is extremely low, especially ResNet50 only reached 73.85% and the total number of recognition error samples reached 221. But with increase of resolution, three models achieved the highest disease recognition accuracy under the condition of 600 × 600 pixels respectively. The “ResNet50 + 600 × 600” combination mode achieved the highest disease recognition accuracy of 98.7%, which was 1.9% and 0.71% higher than “VGG16 + 600 × 600” and “Inception V3 (IcpV3) + 600 × 600” respectively, among them, SP, AA and GYM recognition accuracy reached 99.44%, 98.43%, and 97.67% respectively.

Also, by observing the performance of ResNet50 model during train, validation, and test, its disease recognition accuracy is lower than that of VGG16, Inception V3 at low resolution, such as 224 × 224 pixels (or 299 × 299 pixels), while significantly higher than VGG16, Inception V3 at high resolution, such as 448 × 448 pixels, 512 × 512 pixels and 600 × 600 pixels. It shows that the ResNet50 model is more suitable for higher resolution disease recognition scenarios. Therefore, on the premise of controllable disease data source, it is recommended to increase the disease images resolution appropriately to improve the ResNet50 model recognition accuracy.

ResNet50 vs ResNet101 As shown in Fig. 6, ResNet50 and ResNet101 have in common during the 100 epochs of train and validation: They reached the highest recognition accuracy on validation set respectively at about 27th epoch, and their recognition accuracy curve and loss value curve validation on train set are basically the same. However, there is a significant difference in the recognition accuracy curve trend between two models on validation set: The ResNet50 (green) curve fluctuated greatly before 50 epochs, and then remained

Table 5 Model and resolution combination comparison result on PDIRE dataset (test).

Resolution (pixels)	The overall recognition accuracy (%)			The recognition accuracy of SP,AA,GYM (%)			The mistaken sample of SP,AA,GYM		
	VGG16	Inception V3	ResNet50	VGG16	Inception V3	ResNet50	VGG16	Inception V3	ResNet50
224×224 (299×299)	78.34	80.23	73.85	89.58,73.90,63.37	87.04,71.70,81.98	84.51,62.26,73.26	183(37,83,63)	167(46,90,31)	221(55,120,46)
448×448	93.61	93.61	96.09	96.34,91.19,92.44	95.77,89.94,95.93	98.31,92.77,97.67	54(13,28,13)	54(15,32,7)	33(6,23,4)
512×512	95.98	95.86	97.63	96.06,95.91,95.93	97.18,94.34,95.93	98.60,96.54,97.67	34(14,13,7)	35(10,18,7)	20(5,11,4)
600×600	96.80	97.99	98.70	97.17,96.22,97.10	98.03,98.43,97.10	99.44,98.43,97.67	27(10,12,5)	17(7,5,5)	11(2,5,4)

Fig. 5 The mistaken samples of GYM, SP, AA for ResNet50

Fig. 6 Disease recognition accuracy and cross-entropy loss variation on PDISRE dataset (train and validation) for ResNet50 and ResNet101

Table 6 Pear disease severity recognition result on PDISRE dataset (train and validation).

Model	The highest recognition accuracy (%)	Epochs of the highest recognition accuracy	The recognition accuracy of 100th epoch(%)	Training time (s)
ResNet50	81.99	26,32,47	80.12	5155
ResNet101	73.30	27	68.94	8060

basically stable at about 80%, and always above the ResNet101 (purple) curve which is basically stable at about 68%. In addition, the loss curve trends of two models on validation set are significantly different, the ResNet50 (green) curve fluctuated greatly before 30 epochs, and then basically hovered around 0.8, the ResNet101 (purple) curve achieved a maximum loss of 1.69 in 30th epoch, after which the loss values were all greater than 1. Table 5 describes the model and resolution combination comparison results on PDIRE dataset, in which we use 224^2 to denote 224×224 , similar for other size

It can be seen from Table 6 that the training time of ResNet50 model is 56.4% less than that of ResNet101, the highest recognition accuracy is 8.69% higher than that of ResNet101, and the recognition accuracy of 100th epoch is 11.18% higher than that of ResNet101.

It can be seen from Table 7 and Fig. 7 that the prediction effect of ResNet50 on disease severity was significantly improved than that of ResNet101, and its overall recognition accuracy was 14.29% higher than that of ResNet101. For ResNet50, the category with highest recognition accuracy was 95.65% of E GY M, and the category with lowest recognition accuracy was 69.57% of E AA. For ResNet101, the category with highest recognition accuracy was 82.61% of ML GY M, the category with lowest recognition accuracy was 47.83% of H. It shows that when the number of disease sets is less than 2000 and the number of various types is absolutely balanced, the recognition accuracy of ResNet50 is higher than that of ResNet101. It also shows that only deepening the number of DL network layers does not necessarily improve disease recognition accuracy. As He [44] compared ResNet1202 and ResNet110 based on CIFAR-10 dataset, the result is that the recognition error rate of ResNet1202 is 1.5% higher than that of ResNet110.

Table 7 Pear disease severity recognition result on PDISRE dataset (test).

Model	The overall recognition accuracy (%)	The recognition accuracy (%) of E_SP, ML_SP, E_AA, ML_AA	The mistaken sample of E_GYM, ML_GYM, H
ResNet50	83.23	86.96,78.26,69.57,82.61,95.65,86.96,82.61	27(3,5,7,4,1,3,4)
ResNet101	68.94	78.26,78.26,52.17,78.26,65.22,82.61,47.83	50(5,5,11,5,8,4,12)

5.4 Recognition characteristics of different diseases

As shown in Table 5, the recognition accuracy of SP is often higher than that of AA and GYM under the same DL network model and input image resolution. It may be because the SP number of samples is more than AA and GYM, which is conducive to model learning and extracting more disease characteristics. The recognition accuracy of GYM is often higher than that of AA, although the number of GYM samples is less than that of AA, this may indicate that GYM is easier to extract features and recognize than AA.

As shown in Table 7 and Fig. 7, We can conclude that SP and GYM are easier to recognize than AA, and E_SP disease is easier to recognize than ML_SP. Furthermore, there are three common features of ResNet50 and ResNet101 for disease severity prediction: Firstly, the recognition accuracy of the middle and late stage disease is relatively high. Secondly, the recognition accuracy of the early disease and healthy leaves is relatively low. Lastly, the recognition accuracy of E_AA disease is low, ResNet50 is 69.57% (lowest among 7 categories) and ResNet101 is 52.17% (except healthy leaves, the lowest among 6 category diseases).

5.5 Discussion

The results of pear disease severity recognition experiment (2.4.2) were worse compared with the results of model and resolution combination experiment (2.4.1), no matter on validation set or test set, especially the disease recognition accuracy of "ResNet50 + 600×600 " mode has decreased, maybe mainly for four reasons: Firstly, the number of target classifications expanded from 3 to 7. Secondly, the number of disease datasets reduced from 4226 to 1582, there are only 180 disease samples in each category for training. Thirdly, the classification target has the problem of high similarity or indistinguishable features between classes, such as healthy leaf, early disease leaf, and middle disease leaf, it may cause DL network model to extract features more difficultly, which reduced the recognition accuracy [20]. Finally, the existence of negative samples may lead to a decrease in recognition accuracy, which is related to the selection of early, middle and late disease leaves by naked eyes. So, we will study and establish quantitative standards to eliminate negative samples, e.g., the ratio of diseased area to total leaf area in the next step.

Fig. 7 The disease recognition accuracy confusion matrix for ResNet50 and ResNet101. (a) ResNet50, (b) ResNet101

6 Conclusion and discussion

Under the condition of low resolution (such as 224×224 pixels or 299×299 pixels), DNN network model will produce overfitting problem and large cross-entropy loss value, which will reduce the disease recognition accuracy. Whereas under higher resolution specification (such as 448×448 pixels, 512×512 pixels, 600×600 pixels), the disease recognition accuracy would be improved and descending order is ResNet50, Inception V3, and VGG16. So, the problem can be solved by increasing the resolution, that is, there is a positive correlation between resolution and disease recognition accuracy. However, when the resolution is increased to a certain level, if it continues to increase, the effect of improving the disease recognition accuracy will be very weak, especially under the condition of 700×700 pixels, the training model will have memory overflow. At the same time, there is a positive correlation between resolution and training time, VGG16 is the most time-consuming of the three models.

Generally, among the three kinds of diseases, the recognition accuracy of SP is the highest, and that of AA is the lowest. For the recognition of six kinds of disease severity, the recognition accuracy of E_AA is the lowest, and the recognition accuracy of middle and late disease is generally higher than that of early disease, but E_SP is easier to identify than ML_SP. At the same time, the high similarity between disease classes is one of the factors that reduce the recognition accuracy. In addition, there is not a positive correlation between the setting of epoch parameters and disease recognition accuracy, and the setting of epoch for different models to achieve the optimal classification performance is also different. Meanwhile, there is not a positive correlation between DL network layer numbers and disease recognition accuracy, e.g., ResNet50 is better than ResNet101 in the recognition of disease severity.

With regard to improve the quality of disease datasets: Firstly, try not to collect diseased leaves that are difficult to learn and extract features for DL networks, such as those that

are curled or broken, those with insignificant disease or few lesions, and those with many surface pollutants; Secondly, try to eliminate the existence of negative samples, we will design quantitative standards (such as the proportion of diseased spots) to strictly divide the early disease and the middle and late disease; Thirdly, try to make the leaves occupy more than 50% of the space when taking pictures.

Taking comprehensive consideration of disease recognition accuracy, training time, cross-entropy loss and other factors, the “ResNet50 + 600 x 600” combination mode is the optimal matching method, which can achieve the highest disease recognition accuracy 98.7%. Among them, the disease recognition accuracy of SP, AA, and GYM is 99.44%, 98.43% and 97.67%, respectively.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Yearbook CS et al. (2012) National bureau of statistics of China, China Statistical Yearbook
2. DeChant C, Wiesner-Hanks T, Chen S, Stewart EL, Yosinski J, Gore MA, Nelson RJ, Lipson H (2017) Automated identification of northern leaf blight-infected maize plants from field imagery using deep learning. *Phytopathology* 107(11):1426–1432
3. Bai X, Li X, Fu Z, Lv X, Zhang L (2017) A fuzzy clustering segmentation method based on neighborhood grayscale information for defining cucumber leaf spot disease images. *Comput Electron Agric* 136:157–165

4. Azadbakht M, Ashourloo D, Aghighi H, Radiom S, Alimohammadi A (2019) Wheat leaf rust detection at canopy scale under different lai levels using machine learning techniques. *Comput Electron Agric* 156:119–128
5. Dubey SR, Jalal AS (2012) Adapted approach for fruit disease identification using images. *International Journal of computer vision and image processing (IJCVIP)* 2(3):44–58
6. Ma J, Du K, Zheng F, Zhang L, Gong Z, Sun Z (2018) A recognition method for cucumber diseases using leaf symptom images based on deep convolutional neural network. *Comput Electron Agric* 154:18–24
7. Hinton GE, Osindero S, Teh Y-W (2006) A fast learning algorithm for deep belief nets. *Neural Comput* 18(7):1527–1554
8. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks, in *Advances in neural information processing systems*, pp. 1097–1105
9. Meng T, Wolter K, Wu H, Wang Q (2018) A secure and cost-efficient offloading policy for mobile cloud computing against timing attacks,” *Pervasive and Mobile Computing*, vol. 45, pp. 4–18, [Online]. Available:
10. Lu Y, Yi S, Zeng N, Liu Y, Zhang Y (2017) Identification of rice diseases using deep convolutional neural networks. *Neurocomputing* 267:378–384
11. Tajbakhsh N, Suzuki K (2017) Comparing two classes of end-to-end machine-learning models in lung nodule detection and classification: Mtnns vs. cnns. *Pattern Recogn* 63:476–486
12. Zhang X, Lin T, Xu J, Luo X, Ying Y (2019) Deepspectra: an end-to-end deep learning approach for quantitative spectral analysis. *Anal Chim Acta* 1058:48–57
13. Johannes A, Picon A, Alvarez-Gila A, Echazarra J, RodriguezVaamonde S, Navajas AD, Ortiz-Barredo A (2017) Automatic plant disease diagnosis using mobile capture devices, applied on a wheat use case. *Comput Electron Agric* 138:200–209
14. Lu J, Hu J, Zhao G, Mei F, Zhang C (2017) An in-field automatic wheat disease diagnosis system. *Comput Electron Agric* 142:369–379
15. Mohanty SP, Hughes DP, Salathe M (2016) Using deep learning’ for image-based plant disease detection. *Front Plant Sci* 7:1419
16. Brahimi M, Boukhalfa K, Moussaoui A (2017) Deep learning for tomato diseases: classification and symptoms visualization. *Appl Artif Intell* 31(4):299–315
17. Yamamoto K, Togami T, Yamaguchi N (2017) Super-resolution of plant disease images for the acceleration of image-based phenotyping and vigor diagnosis in agriculture. *Sensors* 17(11):2557
18. Ramcharan A, Baranowski K, McCloskey P, Ahmed B, Legg J, Hughes DP (2017) Deep learning for image-based cassava disease detection. *Front Plant Sci* 8:1852
19. Ma X, Gao H, Xu H, Bian M (2019) An IoT-based task scheduling optimization scheme considering the deadline and cost-aware scientific workflow for cloud computing. *EURASIP J Wirel Commun Netw*, 2019 (249). DOI:<https://doi.org/10.1186/s13638-019-1557-3>
20. Barbedo JGA (2018) Impact of dataset size and variety on the effectiveness of deep learning and transfer learning for plant disease classification. *Comput Electron Agric* 153:46–53
21. Kerkech M, Hafiane A, Canals R (2018) Deep learning approach with colorimetric spaces and vegetation indices for vine diseases detection in uav images. *Comput Electron Agric* 155:237–243
22. Park K, Hong YK, Kim GH, Lee J (2018) Classification of apple leaf conditions in hyper-spectral images for diagnosis of marssonina blotch using mmr and deep neural network. *Comput Electron Agric* 148:179–187
23. Picon A, Alvarez-Gila A, Seitz M, Ortiz-Barredo A, Echazarra J, Johannes A (2019) Deep convolutional neural networks for mobile capture device-based crop disease classification in the wild. *Comput Electron Agric* 161:280–290
24. Gao H, Liu C, Li Y, Yang X (2020) V2VR: Reliable Hybrid-Network-Oriented V2V Data Transmission and Routing Considering RSUs and Connectivity Probability. *IEEE Transactions on Intelligent Transportation Systems(T-ITS)*, DOI: <https://doi.org/10.1109/TITS.2020.2983835>
25. Ferentinos KP (2018) Deep learning models for plant disease detection and diagnosis. *Comput Electron Agric* 145:311–318
26. Zhao S, Li S, Qi L, Xu LD (2020) Computational intelligence enabled cybersecurity for the internet of things. *IEEE Transactions on Emerging Topics in Computational Intelligence* 4(5):666–674
27. Dymann M, Karstoft H, Midtby HS (2016) Plant species classification using deep convolutional neural network. *Biosyst Eng* 151: 72–80
28. Wang S-H, Lv Y-D, Sui Y, Liu S, Wang S-J, Zhang Y-D (2018) Alcoholism detection by data augmentation and convolutional neural network with stochastic pooling. *Journal of medical systems* 42(1):2
29. Rahmehoonfar M, Sheppard C (2017) Deep count: fruit counting based on deep simulated learning. *Sensors* 17(4):905
30. Yang Z, Luo T, Wang D, Hu Z, Gao J, Wang L (2018) Learning to navigate for fine-grained classification,” in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 420–435
31. Kamilaris A, Prenafeta-Boldu FX (2018) Deep learning in agriculture: a survey. *Comput Electron Agric* 147:70–90
32. Frid-Adar M, Diamant I, Klang E, Amitai M, Goldberger J, Greenspan H (2018) Gan-based synthetic medical image augmentation for increased cnn performance in liver lesion classification. *Neurocomputing* 321:321–331
33. Sladojevic S, Arsenovic M, Anderla A, Culibrk D, Stefanovic D (2016) Deep neural networks-based recognition of plant diseases by leaf image classification, *Computational intelligence and neuroscience*, vol 2016
34. da Silva LA, Bressan PO, Goncalves DN, Freitas DM, Machado BB, Goncalves WN (2019) Estimating soybean leaf defoliation using convolutional neural networks and synthetic images. *Comput Electron Agric* 156:360–368
35. Ghazi MM, Yanikoglu B, Aptoula E (2017) Plant identification using deep neural networks via optimization of transfer learning parameters. *Neurocomputing* 235:228–235
36. Kaya A, Keceli AS, Catal C, Yalic HY, Temucin H, Tekinerdogan B (2019) Analysis of transfer learning for deep neural network based plant classification models. *Comput Electron Agric* 158:20–29
37. Chen SW, Shivakumar SS, Dcunha S, Das J, Okon E, Qu C, Taylor CJ, Kumar V (2017) Counting apples and oranges with deep learning: a data-driven approach. *IEEE Robotics and Automation Letters* 2(2):781–788
38. Gao H, Kuang L, Yin Y, Guo B, Dou K (2020) Mining consuming behaviors with temporal evolution for personalized recommendation in Mobile marketing apps. *ACM/Springer Mobile Networks and Applications (MONET)* 25(4):1233–1248
39. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A (2015) Going deeper with convolutions,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1–9
40. 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
41. ImageNet. Large scale visual recognition challenge 2015 (ilsvrc2015). [Online]. Available: <http://image-net.org/challenges/LSVRC/2015/index>
42. He K, Zhang X, Ren S, Sun J (2016) Identity mappings in deep residual networks,” in *European conference on computer vision*. Springer, pp. 630–645

43. Dose H, Møller JS, Iversen HK, Puthusserypady S (2018) An end-to-end deep learning approach to mi-ecg signal classification for bcis. *Expert Syst Appl* 114:532–542
44. 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
45. Ozguven MM, Adem K (2019) Automatic detection and classification of leaf spot disease in sugar beet using deep learning algorithms. *Physica A: Statistical Mechanics and its Applications* 535: 122537
46. Suh HK, Ijsselmuiden J, Hofstee JW, van Henten EJ (2018) Transfer learning for the classification of sugar beet and volunteer potato under field conditions. *Biosyst Eng* 174:50–65
47. Zhang S, Zhang S, Zhang C, Wang X, Shi Y (2019) Cucumber leaf disease identification with global pooling dilated convolutional neural network. *Comput Electron Agric* 162:422–430
48. Gao H, Huang W, Duan Y (2020) The Cloud-Edge Based Dynamic Reconfiguration to Service Workflow for Mobile Ecommerce Environments: A QoS Prediction Perspective *ACM Transactions on Internet Technology*, DOI: <https://doi.org/10.1145/3391198>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Yang Fang is an associate professor in Shanxi Agricultural University, male, born in 1982, graduated from Taiyuan University of Technology with a major in computer application technology, his main research direction is intelligent agricultural systems engineering.