Computers and Electronics in Agriculture 193 (2022) 106703

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Identification method of vegetable diseases based on transfer learning and attention mechanism

Xue Zhao [a], Kaiyu Li [a], Yunxia Li [a], Juncheng Ma [b], Lingxian Zhang [a,c,*]

[a] College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China
[b] Institute of Environment and Sustainable Development in Agriculture, Chinese Academy of Agricultural Sciences, Beijing 100081, China
[c] Key Laboratory of Agricultural Informationization Standardization, Ministry of Agriculture and Rural Affairs, Beijing 100083, China

ARTICLE INFO

Keywords:
Vegetable disease
Attention mechanism
Transfer learning
Convolutional neural network

ABSTRACT

Artificial Intelligence for disease identification is currently the focus of great research interest. Nonetheless, the approach has some problems, for example, identification takes a long time, has low accuracy, and is often limited to a single disease type. Here, we aimed to identify tomato powdery mildew, leaf mold and cucumber downy mildew against simple and complex backgrounds. We developed a vegetable disease identification model, DTL-SE-ResNet50, optimized by SENet and pre-trained by ImageNet to form a new model, SE-ResNet50. The SE-ResNet50 model was trained with the AI Challenger 2018 public database to obtain a new weight. The SE-ResNet50 model with the new weight was then trained by dual transfer learning with a self-built database to create the DTL-SE-ResNet50 model for the identification of vegetable diseases. The model was compared with convolutional neural networks EfficientNet, AlexNet, VGG19, and Inception V3. The experimental results showed that with the same experimental conditions, the identification precision of the new model reached 97.24%, and processing of a single image required 0.13 s. Compared with DTL-CBAM-ResNet50 and DTL-SA-ResNet50, three models has almost the same precision, but time consumption of DTL-SE-ResNet50 was 0.02 s higher than that of DTL-CBAM-ResNet50. Although the time consumption of DTL-SA-ResNet50 was 0.02 s higher than the proposed model, the precision was lower. At the same time, compared with the dual transfer learning model, the model's precision was 4.1% higher, and the processing of a single image was 0.06 s shorter. Compared with convolutional neural networks, the precision of DTL-SE-ResNet50 was 3.19% higher than the best result, the time consumption of a single image was 0.58 s shorter; Recall and F1 also increased. The method proposed in this paper has high identification precision and short identification time, and it meets the requirements for accurate and rapid identification of vegetable diseases.

1. Introduction

At present, agriculture is a major component of the world's economic development (Kulkarni, 2018), and the disease is an important factor that affects the yield and quality of vegetables. In the field of agricultural disease identification, image processing technology has been widely adopted. Artificial Intelligence (AI) has developed rapidly in recent years, gradually occupying an important position in the field of image identification and becoming an important trend in disease identification (Abade et al., 2021). AI includes machine learning, deep learning, transfer learning and attention mechanisms.

In traditional machine learning, disease features are extracted manually. This is time-consuming and laborious, and the coexistence of a variety of diseases increases the difficulty of feature extraction and leads to low identification accuracy. The emergence of deep learning solves these problems. Its self-learning ability does not require manual feature extraction (Brahimi et al., 2017; Gutierrez et al., 2021; Anagnostis et al., 2021), and it is therefore widely used for agricultural weed identification, disease diagnosis, and other tasks. Mohit et al. (2019) aimed at the problem of apple diseases identification and overlarge pre-training model structures such as VGG, instead proposing a method of apple diseases identification based on a convolutional neural network. The network consisted of three convolution layers, three maxpooling layers, and two fully connected layers. The experimental results showed

* Corresponding author at: College of Information and Electrical Engineering, China Agricultural University, No.17 Qinghua East Road, Haidian District, Beijing 100083, China.
E-mail address: zhanglx@cau.edu.cn (L. Zhang).

https://doi.org/10.1016/j.compag.2022.106703
Received 6 July 2021; Received in revised form 6 January 2022; Accepted 8 January 2022
Available online 20 January 2022

that its accuracy rate reached 99%, and its test time was 7.08 s; the model demonstrated improved identification accuracy and enhanced practicality. Rehman et al. (2021) combined ResNet101 with FPN to create Mask RCNN and combined it with transfer learning to achieve real-time detection of apple diseases. The experimental results showed that the average accuracy reached 96.6%. Rahman et al. (2020) and (Li et al., 2020a) constructed a new CNN network for diseases and insect pest detection, and compared it with the classical convolutional neural networks VGG16, Inception V3, ResNet50, and others. The experimental results demonstrated the effectiveness of the model and its improved identification accuracy. Zhang et al. (2018) took corn leaf diseases as the research object, improved the network structure of GoogLeNet and Cifar10, and compared the results with the classical machine learning model and the deep learning model. The final test results showed that the accuracy of GoogLeNet and Cifar10 models reached 98.9% and 98.8%, respectively. Liu et al. (2020) combined the structure and advantages of AlexNet, GoogLeNet V2 and DenseNet networks to create the Kiwi-ConvNet network model based on CNN. The experimental results showed that its accuracy rate reached 98.54%, and its convergence speed was faster.

Nonetheless, the training of deep learning requires a great deal of data support, and lack of data may cause problems, such as local optimal solutions or over-fitting. Researchers at home and abroad therefore use transfer learning to solve this problem. Transfer learning refers to transferring the weight trained on a certain database to a new model (Li et al., 2017; Wang et al., 2019; Ishengoma et al., 2021; Pan et al., 2010; Tian et al., 2019), thereby accelerating the model convergence and saving the training time. Lee et al. (2020) tested and compared different transfer learning mechanisms through VGG16, Inception V3, and GoogLeNetBN based on PlantVillage and common databases. The experimental results showed that the VGG16 model pre-trained by ImageNet has better generalization in adapting to new data. Espejo-Garcia et al. (2020) proposed a crop / weed identification system based on combining CNN network with a traditional machine learning classifier. The experimental results showed that the combination accuracy of DenseNet and SVM reached 99.29%, and the difference between the training and test database was very small, avoiding over-fitting and improving the model stability. Maeda-GutiéRrez et al. (2020) compared and analyzed the most advanced convolutional neural network architectures, such as AlexNet, GoogLeNet, Inception V3, ResNet18, and ResNet50, and then fine-tuned the structure based on them. Finally, an improved GoogLeNet model was used to identify tomato diseases. Trang et al. (2019) combined ResNet with transfer learning to build a mango disease identification model. The experimental results showed that the average accuracy of the proposed method reached 88.46%, and the difference between training and verification accuracy was very small. Compared with other models, this method had fewer over-fitting phenomena, and showed improved accuracy and data classification ability. Barbedo et al. (2018) used transfer learning and GoogLeNet to study the influence of database size and type on disease identification. The experimental results showed that the number of types, the feature similarity of images in the training and test databases, and the image background all had an impact on the disease identification result. Chen et al. (2020) combined the advantages of VGGNet and the Inception module, pre-trained VGGNet on ImageNet and the Inception module, and performed transfer learning. A self-built database in this paper was used to train a new neural network. The test results on a public database and a self-built database were better, and their accuracy was higher. Argueso et al. (2020) used the PlantVillage database, fine-tuned the Inception V3 network, learned the leaf characteristics for transfer learning, and compared the FSL network with classical transfer learning. The experimental results showed that the FSL network had high accuracy and could learn new diseases from a very small database, thereby reducing the data demand. (Wang and Wang, 2021)improved the SE-ResNeXt-101 model and combined it with transfer learning to develop the crop disease classification model TL-SE-ResNeXt-101. The experimental results showed that the accuracy of the proposed model reached 98%, higher than that of other models under the same experimental conditions. Under real-world conditions, the average accuracy reached 47.37%, and higher than the other models, demonstrating that the model had high identification accuracy and strong robustness. (Li et al., 2020b) for whether shallow CNN can be used for plant disease identification, two methods were proposed: SCNN-KSVM and SCNN-RF. VGG16 model was trained by transfer learning, then Kernel SVM and random forest processing embedding were selected. The experimental results showed that the combination of shallow CNN and machine learning algorithm also have good processing ability. The author said that if shallow CNN cannot achieve the needs, deep CNN network can be further considered. In order to solve the problem of plant leaf disease, a semi supervised less lens learning method for disease recognition was proposed by (Li and Chao, 2021) Transfer learning was used to fine-tuned the model, unmarked samples were sent to the fine-tuned model for prediction, multiple experiments were carried out to obtain the average accuracy. The experimental results showed that the proposed method can obtain better results with fewer samples, and have better performance than the classical less shot learning method based on transfer learning. Aiming at the problem of learning to automatically identify pests or plants from a small number of samples, (Li and Yang, 2021) proposed a task driven meta learning less shot classification method. The shallow CNN model was used for feature extraction, many task meta training models were prepared from the source set, and the average accuracy of a few shot classification was tested according to the configuration of the target set. Finally, the model parameters were frozen, the test accuracy of each task was recorded, and the average accuracy was finally output as the test index. The final results showed that this study can be used as a learning and comparison of a few lens learning tasks in the field of agriculture.

Attention mechanism has been widely used in recent years. In the past, the focus of deep learning model was the whole input image, while the focus of attention mechanism was the target region in the input image, which changes the focus from the whole to the local. In recent years, attention mechanisms have been widely used in the field of deep learning. At present, they are mainly used in natural language processing and image segmentation. (Li and Rai, 2020) used ResNet18, ResNet34, and VGG16 models to add residual blocks to identify and classify apple diseases. The experimental results showed that ResNet18 with few network layers had the best identification ability. Yu et al. (2020) proposed a point attention mechanism for leaf identification that combined FS-SubNet with SAC-SubNet. FS-SubNet used a new database for training and connected SubNet with SAC-SubNet to achieve apple leaf disease identification. The experimental results showed that the accuracy of disease identification reached 89.4%, and more discriminative features and semantic point information were extracted, improving the identification accuracy. (He, 2020) proposed an end-to-end deep learning image compressed sensing framework that combined a channel attention mechanism with multi-scale features to extract features. The experimental results showed that the final reconstruction rate reached 96.65%, further improving the reconstruction quality.

In order to solve the problems of low precision, long time-consuming and single identification type, and realize rapid and accurate identification of diseases, this study proposes a vegetable disease identification model based on dual transfer learning and attention mechanism. Based on ResNet50, the model is optimized through the use of the attention mechanism. The new model is trained on the AI Challenger 2018 database and a self-built database. The training results are evaluated based on the model running time, identification precision Recall, F1 Scor and the number of parameters as evaluation index providing a reference for the disease identification technology in vegetables.

The paper is structured as follows: Section 2 presents the collection of the image database. Section 3 introduces the method used to create the vegetable disease identification model along with related concepts and the proposed approach. Section 4 explains the experimental setup and results. In Section 5, conclusions are drawn and future prospects are discussed.

2. Data acquisition and preprocessing

The research in this paper used two databases: the AI Challenger 2018 database and a self-built database. The AI Challenger 2018 database contains 27,113 images, including 10 healthy samples and 27 disease samples from apple, cherry, strawberry, tomato, corn, grape, orange, peach, pepper, and potato. Images of tomato powdery mildew, tomato leaf mold, and healthy tomato leaves were selected from the database. Some low quality disease images were removed, and form a new database, which comprised 299 tomato leaf mold images, 28 healthy tomato leaf leaves images, and 14 tomato powdery mildew images. The self-built database included seven types images of tomato powdery mildew, tomato leaf mold, healthy tomato leaves, and cucumber downy mildew and healthy cucumber leaves with simple and complex backgrounds. The images were collected at Beijing Academy of Agricultural Sciences, Changping, Xiaotangshan and other research sites. Mobile phones were used to select different angles for image collection. The collection time was from 9 a.m. to 3 p.m. on sunny days, when the external light was suitable, and the image backgrounds were mainly soil or crops. Some images were collected with white paper as the background. All images were obtained in auto exposure mode, the total was 3226 and transmitted to a computer in JPG format. Some examples images from the public database are shown in Fig. 1.

Because the data quantity of two kinds of crop diseases is not balanced, it is easy to achieve disease training results with large amounts of data in the training process (Liu and Gao, 2021). However, training of a convolution neural network requires a great deal of data, and when the amount of data is small, over-fitting can easily occur. It is therefore necessary to expand the data. In this paper, image flipping, image rotation, and other methods were used to expand the data:

(1) Image flipping: Flip the input image horizontally and vertically.
(2) Image rotation: Randomly rotate the input image from 0° to 360°.
(3) Image clipping: Clip a part of the input image randomly, or clip an image containing multiple recognized objects into multiple images containing a single recognized object. The examples of image augmentation are shown in Fig. 2.

The results of data expansion are shown in Table 1. Through image augmentation, the limited data can produce the same value as more data without substantially increasing the data. Eighty percent of the experimental data were used for training and twenty percent for testing. Some examples images samples from the self-built database are shown in Fig. 3.

As shown in Table 1, there are 11,531 self-built databases and 4765 AI 2018 databases. The databases in this paper includes three kinds of diseases and four kinds of healthy leaves of tomato and cucumber: tomato powdery mildew, leaf mold and healthy leaves; cucumber downy mildew and healthy leaves under simple backgrounds and complex backgrounds.

3. Image identification model for vegetable diseases

Convolutional neural networks have become the most commonly used method of deep learning in computer vision, because they can automatically learn features from the input image. This approach can reduce or solve over-fitting by sharing weights and local connections to achieve a better network (Maeda-Gutierrez et al., 2020). A convolutional neural network model usually consists of more than one convolution layer, a pooling layer and a full connection layer. Convolution layer is the key to construct convolution neural network. Its main function is to extract the features of the input data, which contains multiple convolution kernel. The pooling layer is usually after the continuous convolution layer, which is mainly a down sampling operation to reduce the feature space of the feature map and the amount of parameters in the network structure. The FC layer is usually located at the back, and the main purpose is to classify. The overall working schematic diagram of convolutional neural network are shown in Fig. 4.

The calculation formula of convolution layer is as shown in Eq. (1):

$$x_j^l = f\left(\sum_{i \in M_j} x_i^{l-1} k_{ij}^l + b_j^l\right) \tag{1}$$

where, x_j^l is the output of the jth neuron in layer l, x_i^{l-1} is the output of the ith neuron in layer l-1, M_j represents the input feature mapping set, l is the layer serial number, k_{ij}^l is the convolution kernel, b_j^l is the offset term, $f(\bullet)$ is the nonlinear activation function.

The pooling process is usually as shown in Eq. (2):

$$x_j^l = f_{down}(x_i^{l-1}) \tag{2}$$

where, $f_{down}(\bullet)$ represents the down sampling function.

3.1. Convolutional neural network (CNN)

Convolutional neural network (CNN) can automatically learn features from the input image, reduce or solve the over fitting problem by sharing weights and local connections, so as to achieve better network performance, and develop rapidly in the field of agriculture. As a milestone of CNN, AlexNet uses Dropout to effectively solve the over fitting problem, and proposes a local response normalization layer to enhance the generalization ability of the model; VGG19 uses multiple small convolution cores to improve efficiency and reduce the amount of parameters, so as to improve the network performance; Inception V3 splits the large convolution core into multiple small convolution cores to

Fig. 1. Examples of vegetable disease leaf image samples from public database: tomato powdery mildew, tomato leaf mold, and healthy tomato leaves.

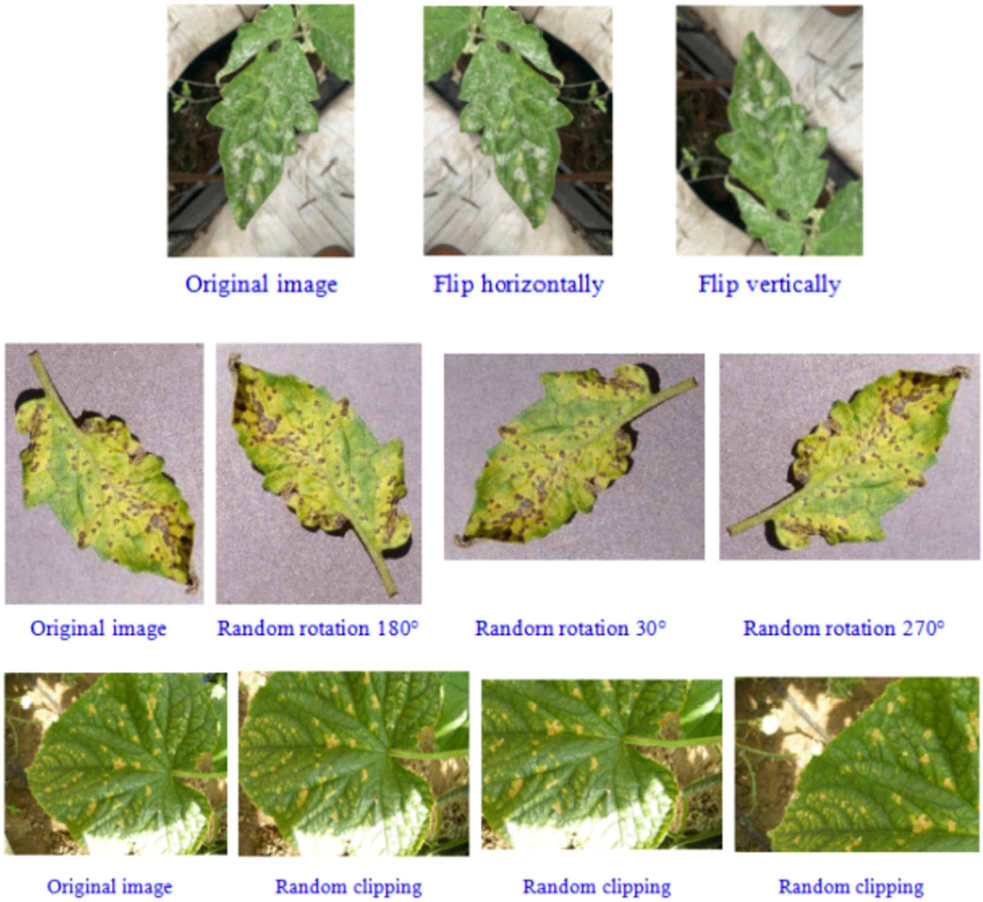


Fig. 2. Examples of image augmentation.

Table 1
Number of images from different crops and disease types in two databases.

Varieties	Original database					Expanded database			
	AI 2018 database		Self-built database			AI 2018 database		Self-built database	
	Disease	Number	Disease		Number	Disease	Number	Disease	Number
Tomato	Powdery Mildew	14	Powdery Mildew		993	Powdery Mildew	1554	Powdery Mildew	2015
	Leaf Mold	299	Leaf Mold		604	Leaf Mold	1592	Leaf Mold	1670
	Healthy leaves	28	Healthy leaves		310	Healthy leaves	1619	Healthy leaves	1556
Cucumber	Downy Mildew	/	Simple background	Downy Mildew	83	Downy Mildew	/	Downy Mildew	1530
	Healthy leaves	/		Healthy leaves	115	Healthy leaves	/	Healthy leaves	1554
	Downy Mildew	/	Complex background	Downy Mildew	736	Downy Mildew	/	Downy Mildew	1619
	Healthy leaves	/		Healthy leaves	385	Healthy leaves	/	Healthy leaves	1587

increase the network depth, and replaces the full connection layer with the average pooling layer.

(1) ResNet50

ResNet (Residual Network) solves the problems of gradient disappearance and gradient explosion with increasing network layers in the actual environment (Geng et al., 2018). The most frequently used ResNet networks are ResNet50, ResNet101, and others; among them, ResNet50 has better identification accuracy and real-time performance (He et al., 2016), and its structure is shown in Table 2.

(2) EfficientNet

Classical neural networks generally obtain better precision through increasing the network depth through residual structure neural network, the network width, and the resolution of the input image. Usually, the classical neural network has only one of the three ways, and EfficientNet combines the three ways to get a better network structure through adjustment. Efficientnet-B0 includes 16 MBConvs, two Convs, a global average pooling layer and one FC classification layer (refer to Tan et al., 2019; (Bao et al., 2021a)). The network structure of EfficientNet-B0 is shown in Fig. 5.

3.2. *Attention mechanism module*

In recent years, attention mechanisms have been widely used in deep learning (Wang et al., 2017). Great progress has been made in applying attention mechanisms in the fields of natural language processing and image segmentation (Chen et al., 2021). Attention mechanisms focuses attention on useful or interesting areas, suppresses unnecessary features, assign different weights to different features, and improves the

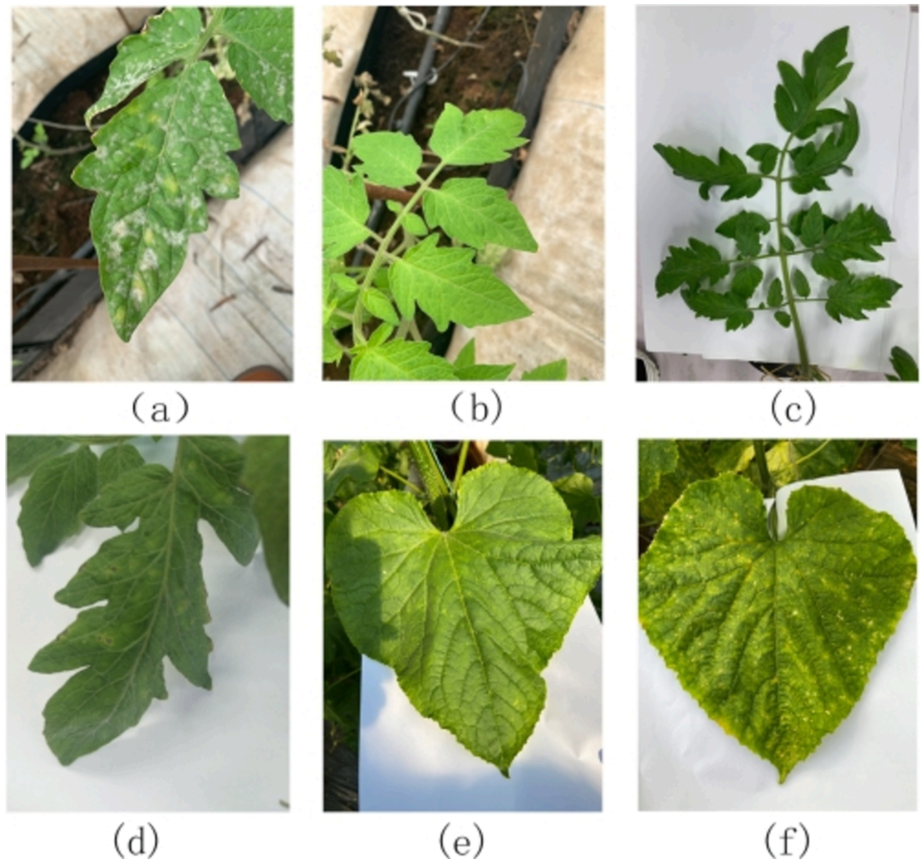


Fig. 3. Examples of vegetable disease leaf image samples from the self-built database: (a) tomato powdery mildew; (b) healthy tomato leaves in the field environment; (c) healthy tomato leaves in the laboratory; (d) tomato leaf mold; (e) healthy cucumber leaves; (f) cucumber downy mildew.

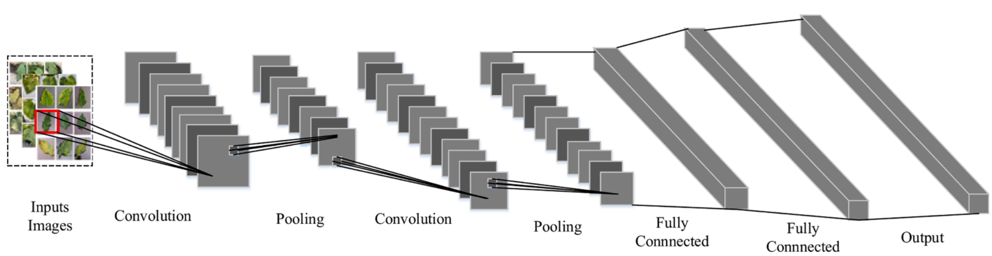


Fig. 4. The overall working schematic diagram of convolutional neural network.

efficiency and accuracy of the model (Zhu et al., 2019). SENet (SqueezeNet-and-Excitation Network) is a representative channel attention mechanism that has the advantages of simple structure, easy deployment, and no requirement for the introduction of new functions or layers (Hu et al., 2020). SENet is mainly composed of three basic parts (Chen et al., 2019; Qin et al., 2020):

(1) Squeeze operation, a compression step that uses global average pooling to pool H × W × C into specific data;
(2) Excitation operation, which generates different weights to act on each channel and constructs the dependencies between different channels;
(3) Reweight operation, which adds the weight obtained by the exception operation to the previous feature by multiplication.

The SE-ResNet50 model is built by embedding SENet into the ResNet50 model: the resulting structure is shown in Fig. 6.

CBAM implements the attention structure of channel to space, which is embedded into the CNN structure and trained with CNN (Bao et al., 2021b); Woo et al., 2018; Wang et al., 2021). The CBAM structure is shown in Fig. 7.

The overall attention generation process of CBAM is shown in the Eq. (3):

$$\begin{cases} F_1 = M_C(F) \otimes F \\ F_2 = M_S(F_1) \otimes F_1 \end{cases} \tag{3}$$

where, F is the feature map. $M_C(F)$ is the output weight of F through channel attention. $M_S(F_1)$ is the output weight of F through spatial attention. $\otimes$ represents the multiplication of corresponding elements.

3.3. Dual transfer learning

Transfer learning is a method used to solve similar unknown problems in different fields. The use of transfer learning can avoid the process of labeling a large number of data in a convolutional neural network, reduce data dependence, increase the calculation speed, and improve the model efficiency (Dai et al., 2020).

Table 2
ResNet50 network structure.

Layer name	50-layer
Conv 1	7×7, 64, stride 2
	3×3 max pool, stride 2
Conv 2_x	$\begin{bmatrix}1\times1, 64\\3\times3, 64\\1\times1, 256\end{bmatrix}\times3$
Conv 3_x	$\begin{bmatrix}1\times1, 128\\3\times3, 128\\1\times1, 512\end{bmatrix}\times4$
Conv 4_x	$\begin{bmatrix}1\times1, 256\\3\times3, 256\\1\times1, 1024\end{bmatrix}\times6$
Conv 5_x	$\begin{bmatrix}1\times1, 512\\3\times3, 512\\1\times1, 2048\end{bmatrix}\times3$
	Average pool, 1000-d fc, softmax

Zhao et al. (2020) proposed a diagnostic method based on transfer learning for the real-time diagnosis of tomato leaf water stress process, fine-tuned the structure of DenseNet169, and combined it with transfer learning to classify the degree of leaf water stress. To establish an efficient and accurate detection system for grape leaf diseases, Fan et al. (2021) fine-tuned VGG16 model, combined it with transfer learning, and compared it with AlexNet, ResNet50, and Inception V3 models.

In this paper, the ResNet50 model was first pre-trained with the ImageNet database to obtain the weight; then the weight was transferred to the improved ResNet50 model, and the AI Challenger 2018 database was used for training to obtain the new weight. Finally, the self-built database was used to train the model with the new weight. Two transfer learning steps were performed, and the approach is therefore called dual transfer learning (DTL). The training process of dual transfer learning is shown in Fig. 8.

3.4. DTL-SE-ResNet50 model for vegetable diseases image identification

DTL-SE-ResNet50 vegetable disease identification model uses an attention mechanism to optimize ResNet50 and adopts a dual transfer learning method to train the model. A flow chart of the vegetable disease identification model is shown in Fig. 9.

The method for combining dual transfer learning with an attention mechanism involves first using SENet to optimize the ResNet50 model pre-trained by ImageNet to obtain SE-ResNet50, then using the AI Challenger 2018 database to train the new model, and obtain the new

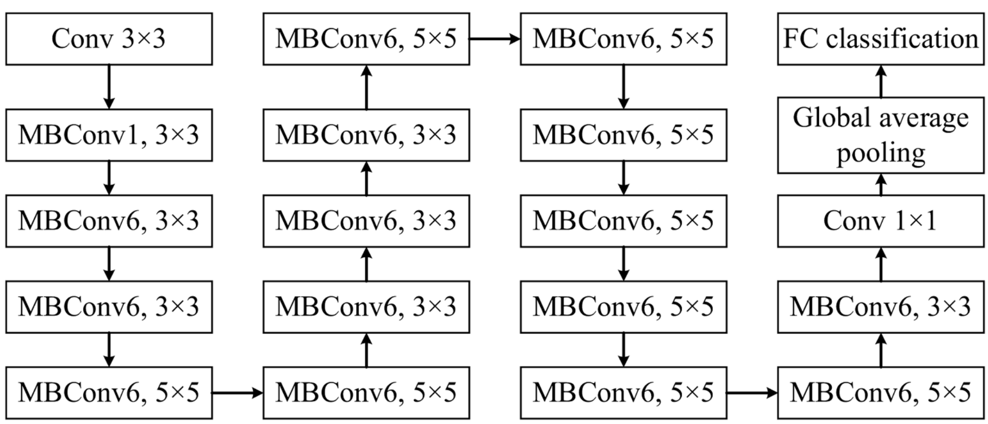


Fig. 5. Structure of EfficientNet model.

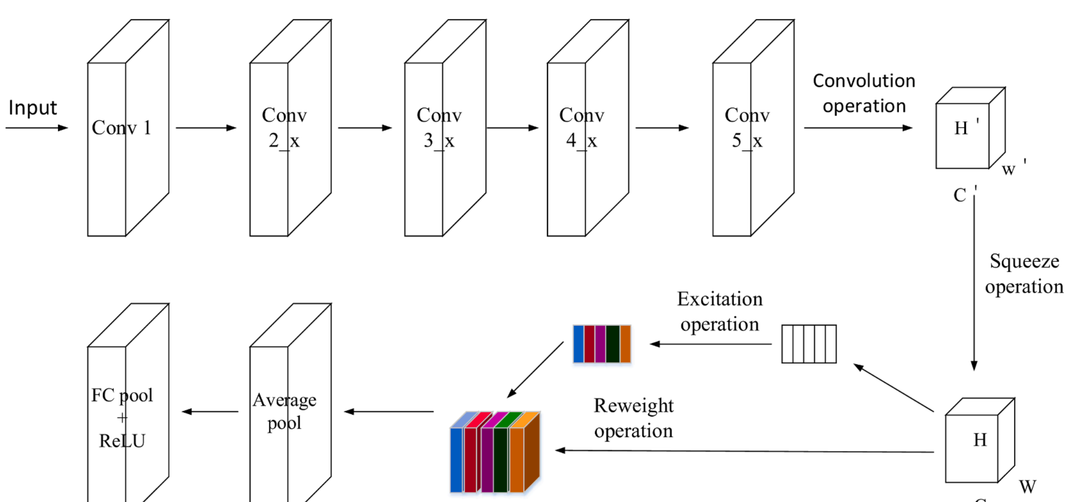


Fig. 6. Structure of SE-ResNet50 model.

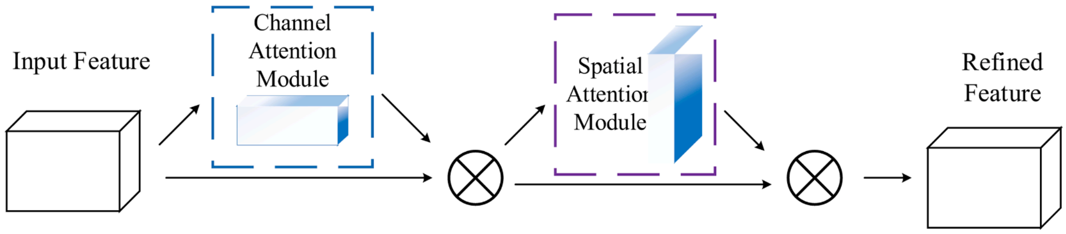


Fig. 7. Structure of CBAM module.

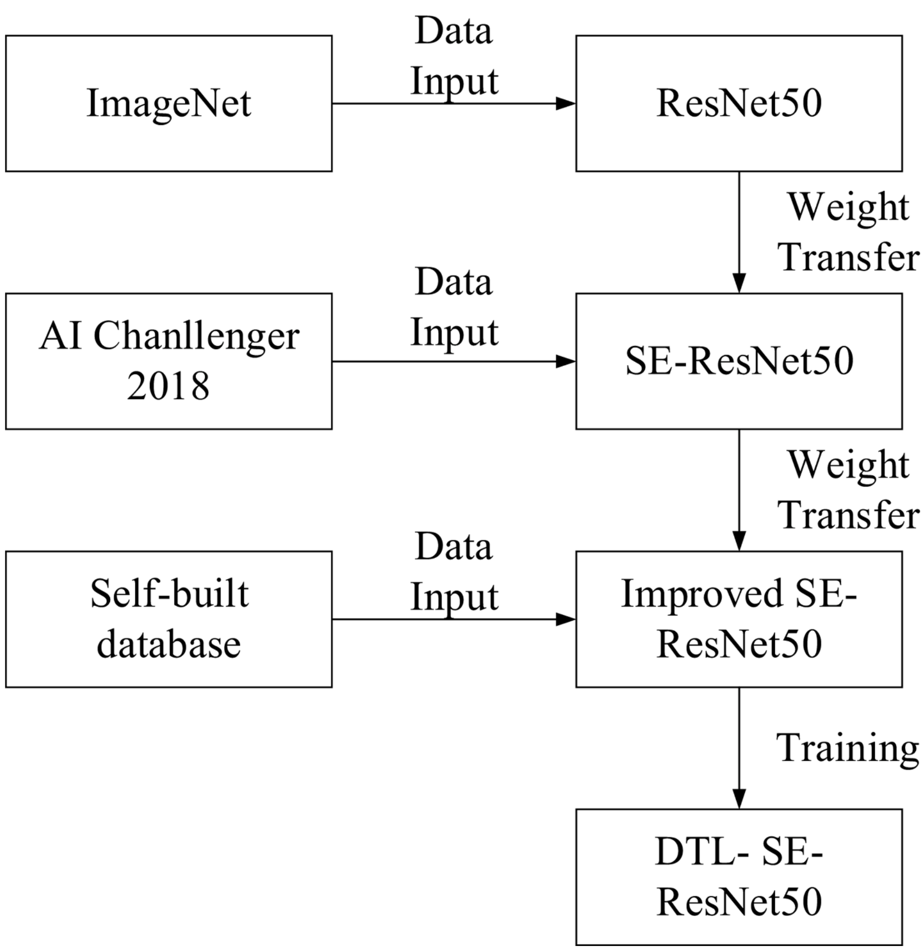


Fig. 8. Process of Dual Transfer Learning (DTL) training.

weight, and finally training the model with the new weight using the self-built database to identify vegetable diseases. At the same time, optimized ResNet50 with CBAM and spatial attention to form DTL-CBAM-ResNet50 and DTL-SA-ResNet50, the experimental results after training were compared with DTL-SE-ResNet50.

4. Results and analysis

4.1. Test platform and parameter setting

The running environment for the whole experiment is the Win10 (64 bit) operating system, with 16 GB of computer memory. The vegetable disease identification model based on transfer learning and an attention mechanism is programmed in MATLAB . A series of toolboxes such as image collection and image processing provided by MATLAB are used. Meanwhile, the MATLAB language is used for programming, the ResNet50 model structure is improved and combined it with transfer learning to create the vegetable disease identification model.

The ResNet50 is trained in two ways: dual transfer learning combined with an attention mechanism and dual transfer learning. Dual transfer learning refers to the two rounds of transfer learning of ResNet50 model trained by the AI Challenger 2018 database and the self-built database. Its experimental results were compared with those of the DTL-SE-ResNet50 model. In the original model, there were 1000 categories in the softmax classification layer. In this study, for four diseases and three health categories, the number of categories in the original classification layer was changed to seven, and the activation function used the rectified linear unit (ReLU) to accelerate the model convergence speed.

The feature extraction layer of ResNet50 is frozen, and the global average pooling layer and the fully connected layer of the original model are replaced by SENet, a new global average pooling layer, and a fully connected layer with ReLU activation function. Learning rate represents the update speed of model weight, which is an important parameter. The

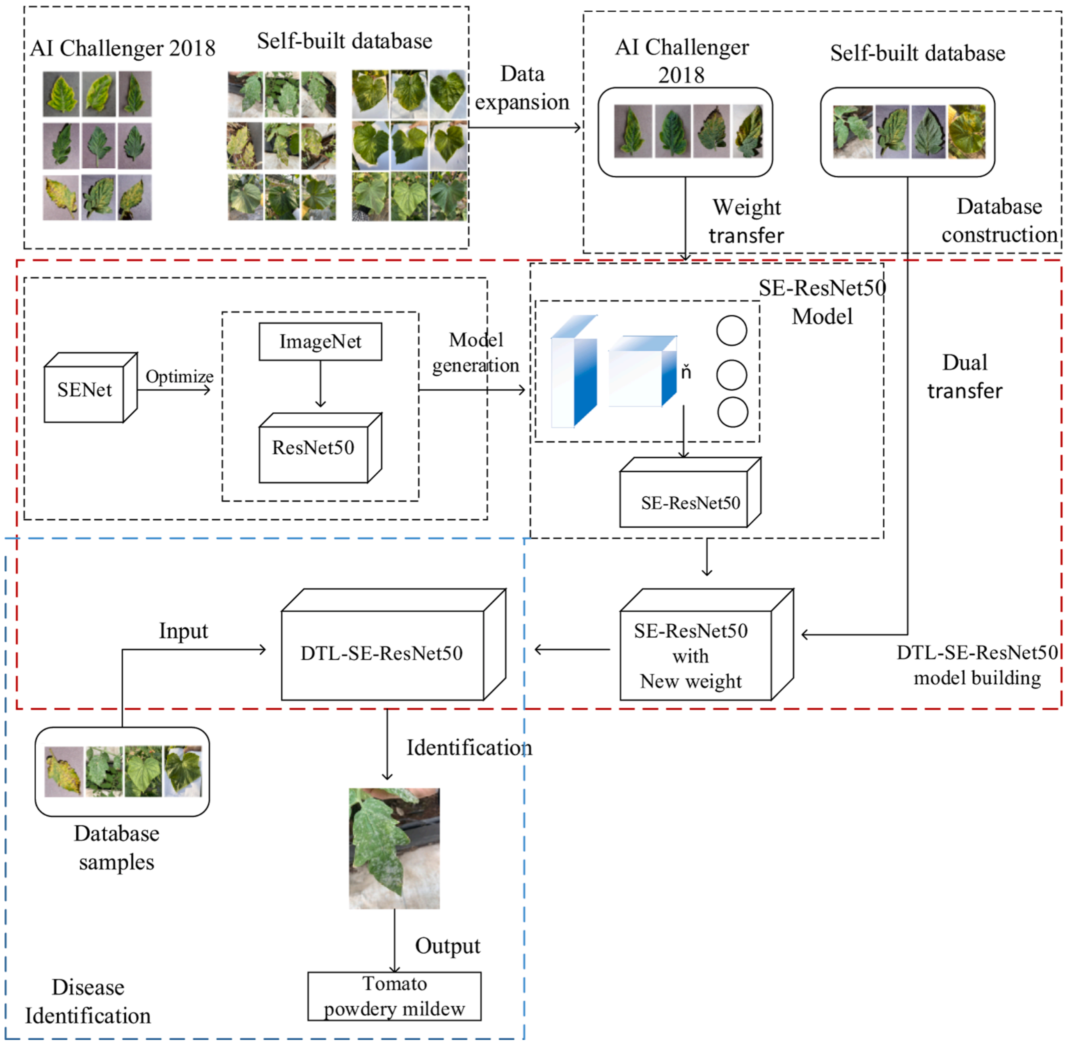


Fig. 9. Flow chart of the DTL-SE-ResNet50 vegetable disease identification model.

best learning rate is selected based on experimental comparison of learning rate. Learning rates of 0.1, 0.01, 0.001 and 0.0001 are tested respectively and the model performance was best when the learning rate is set to 0.001. Considering the hardware condition and training time of the computer, the batch size for testing and training is set to 20. The accuracy and loss training curve are shown in Fig. 10.

4.2. Model evaluation indexes

Many indexes can be used to evaluate the model performance. In this study, precision, the time consumption of a single image, Recall, F1 Score and the number of parameters were selected to evaluate the vegetable disease identification model.

(1) Precision (P)

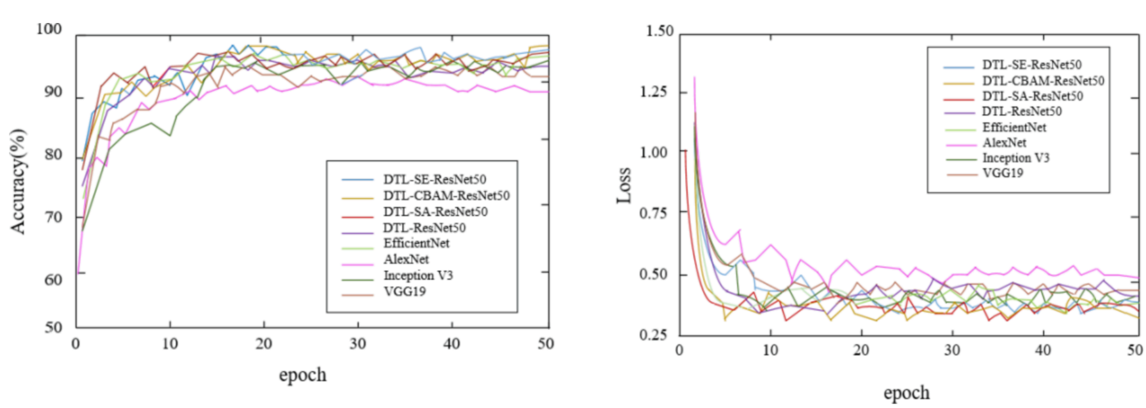


Fig. 10. Accuracy and loss curve in the training.

Precision (P) refers to the percentage of correctly predicted samples relative to all samples in the model, as shown in Eq. (4):

$$P = \frac{T_P}{T_P + F_P} \times 100\% \quad (4)$$

where, T_P is the number of samples predicted to be 1 and actually 1, and F_P is the number of samples predicted to be 1 and actually 0.

(2) Time consumption of a single image (T_S)

Time consumption of a single time (T_S) is a key index for model evaluation. Short model training times are conducive to parameter learning, and accelerate model training. The time required to analyze a single image can be calculated to facilitate model comparisons, as shown in Eq. (5):

$$T_S = \text{Total test time}/\text{Total number of test images} \quad (5)$$

(3) Recall (R)

Recall (R) refers to the number of correctly classified classes divided by the correct total number. Recall is shown in Eq. (6):

$$R = \frac{T_P}{T_P + F_N} \times 100\% \quad (6)$$

where, T_P is the number of samples predicted to be 1 and actually 1, and F_N is the number of samples predicted to be 0 and actually 1.

(4) F1 Score ($F1$)

F1 Score ($F1$) is the harmonic average of the precision and recall of the model, as shown in Eq. (7):

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} \quad (7)$$

(5) The number of parameters (M)

The number of parameters (M) is a comprehensive evaluation index of the overall performance of the model. The number of parameters affects the running speed and size of the model. Therefore, parameter quantity is an important index of model evaluation in this paper.

4.3. Effect of attention mechanism on model performance

To compare the effect of the attention mechanism on the model performance, and enable comparison with basic dual transfer learning, the network models proposed in this paper were trained under the same experimental conditions. The experimental results are shown in Table 3.

It can be seen from Table 3 that under the same experimental conditions, compared with the dual transfer learning model, the combination of an attention with dual transfer learning improved precision by 3.99% at least, and reduced the time consumption of a single image by 0.04 s at least; Recall increased by 3.24% at least, indicating that the number of correctly classified samples has increased; F1 comprehensively considers Precision and Recall, and its improvement can better reflect the superiority of the model's performance. By comparing the results of the models with different attentions, they showed little difference. The precision of DTL-CBAM-ResNet50 will be slightly higher than that of DTL-SE-ResNet50, but the time will be 0.02 s slower. This is because feature map goes through channel attention and spatial attention in turn, running time and parameters increase compared to SENet. Although time consumption of DTL-SA-ResNet50 is 0.02 s higher and the number of parameters is smaller than DTL-SE-ResNet50, the precision is slightly lower than the latter. Therefore, the effect of DTL-SE-ResNet50 is better. The addition of the attention mechanism improved the identification precision of the model, and reduced the time consumption for a single image. The precision of DTL-CBAM-ResNet50 and DTL-SA-ResNet50 are 97.83% and 97.13% respectively, which are almost the same, but the parameters of the two models are respectively 30.5×10^6 and 27.3×10^6. There is a large difference because DTL-CBAM-ResNet50 includes channel attention and spatial attention. The number of DTL-CBAM-ResNet50 model parameters is more than that of DTL-SA-ResNet50 with only spatial attention. A large number of parameters means that the memory occupied by the model is large. Therefore, DTL-CBAM-ResNet50 is suitable for PC when the hardware allows, while DTL-SA-ResNet50 is more suitable for mobile terminal because of less model parameters and less memory.

Table 3
Comparison of experimental results of disease identification models.

Model	P/%	T_S/s	R/%	F1	M
DTL-SE-ResNet50	97.24	0.13	92.58	94.85	30.2
DTL-CBAM-ResNet50	97.83	0.15	93.04	95.37	30.5
DTL-SA-ResNet50	97.13	0.11	91.97	94.48	27.3
DTL-ResNet50	93.14	0.19	88.73	90.88	26.7

4.4. Comparative analysis of model results

To verify the identification ability of DTL-SE-ResNet50, EfficientNet, AlexNet, VGG19 and Inception V3 were selected for dual transfer learning and compared under the same experimental conditions. The experimental results are shown in Table 4.

Compared with four convolutional neural networks, Table 4 shows that the proposed in this paper is superior in precision and training time. Its identification precision is 3.19% higher than the best result of the latter, and its time consumption of a single image is 0.58 s faster; At the same time, Recall and F1 Score have achieved good results, the number of parameters is the smallest. It shows that compared with the other models, the proposed research method has the largest number of correct samples, the performance is the best of these models. In the overall performance, the method proposed in this paper has more advantages. Compared with other three classical models, EfficientNet incorporates SENet for the highest precision and the shortest time, but the overall performance of the model is not better than the proposed model. AlexNet is not the largest of these four models, but its time consumption is the highest and its precision is the lowest. VGG19 has too many parameters, and consumes a lot of computational resources; most of its parameters are mainly concentrated in the first fully connected layer, making its running speed relatively slow, and increasing the time required to process a single image. The number of parameters and the model of Inception V3 are relatively small, so the single image processing time relatively fast.

The performance of the eight models on the test database was compared by using confusion matrix visualization, and the DTL-SE-ResNet50 model was analyzed further. A confusion matrix is an analysis table that summarizes the model classification and prediction results in machine learning. Each row represents the actual data for a category, and each column represents the predicted data for the category. Each value in the table is the probability that the data of the row category are predicted as the column category, and the value at the

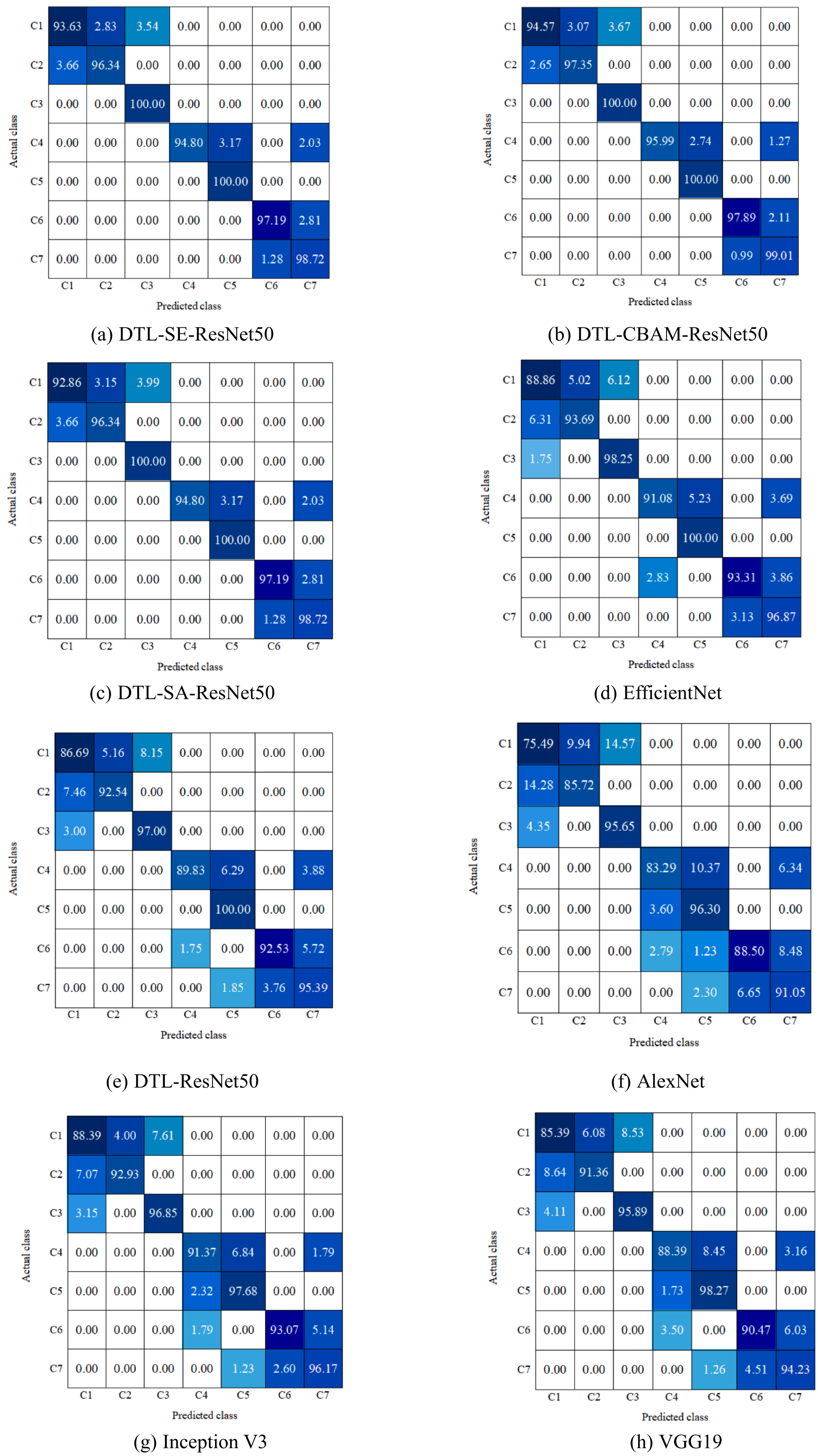
Table 4
Comparison of test results of disease identification models.

Model	P/%	T_S/s	R/%	F1	M
DTL-SE-ResNet50	97.24	0.13	92.58	94.85	30.2
EfficientNet	94.58	0.18	88.95	91.68	42.5
AlexNet	89.21	1.03	73.92	80.84	62.7
Inception V3	94.05	0.71	86.24	89.98	24.6
VGG19	92.00	0.89	83.95	87.79	145.2

diagonal position predicts the correct probability. Confusion matrices of the eight models are shown in Fig. 11.

Note: C1, C2, C3, C4, C5, C6 and C7 represent tomato powdery mildew, tomato leaf mold, healthy tomato leaves, cucumber downy mildew with the simple background, healthy cucumber leaves with the simple background, cucumber downy mildew with the complex background, and healthy cucumber leaves with the complex background.

It can be seen from Fig. 11 that the overall identification precision of



(a) DTL-SE-ResNet50 (b) DTL-CBAM-ResNet50 (c) DTL-SA-ResNet50 (d) EfficientNet (e) DTL-ResNet50 (f) AlexNet (g) Inception V3 (h) VGG19

Fig. 11. Confusion matrices of eight models.

DTL-SE-ResNet50 was higher than that of the other models. The identification precision for tomato leaf mold, healthy tomato leaves, healthy cucumber leaves with the simple background, healthy cucumber leaves with the complex background, and cucumber downy mildew with the complex background were higher in the proposed model, reaching 96.34%, 100%, 100%, 98.72% and 97.19% respectively. This was primarily because the attention mechanism reduced the learning of region features that were useless for disease identification, and directed more attention to the useful areas, enabling the extraction of more detailed disease features and improving the identification effect.

The identification error rates were higher for tomato powdery mildew and cucumber downy mildew against the simple background than the other categories. Analyzed the data images and found, the disease characteristics were not obvious when the disease degree of tomato powdery mildew or cucumber downy mildew was light, and infected leaves were similar to healthy leaves. However, the reason for the misidentification of tomato powdery mildew and leaf mold was that powdery mildew causes pale yellow, round lesions at the early stage of the disease, similar to leaf mold. Therefore, during the learning process, the model needs more detailed features for identification, which increases the difficulty of identification and leads to identification errors.

5. Conclusions

A DTL-SE-ResNet50 vegetable disease identification method was constructed based on an attention mechanism and transfer learning method. The ResNet50 model was optimized by SENet, spatial attention and CBAM, and experimental results were obtained from AI Challenger 2018 database and the self-built database using dual transfer learning training. The influence of the attention mechanism on the model performance was discussed, and the model was compared with convolutional neural networks EfficientNet, AlexNet, VGG19, and Inception V3. The conclusions are as follows:

(1) Compared with DTL-CBAM-ResNet50 and DTL-SA-ResNet50, the precision of DTL-SE-ResNet50 is almost the same with the others, but DTL-CBAM-ResNet50 is 0.02 s slower than DTL-SE-ResNet50 in time. Because CBAM module includes channel attention and spatial attention, and the feature map needs to pass through two attention modules. In terms of time consumption of a single image and model parameters, it is much more than DTL-SE-ResNet50. Although the time consumption of DTL-SA-ResNet50 is 0.02 s higher than the proposed model, the precision is lower, so the overall performance of SENet is better. And compared DTL-CBAM-ResNet50 with DTL-SA-ResNet50, the precision of the two models is almost the same, but the parameters of the two models are quite different. the parameters of DTL-CBAM-ResNet50 is larger than that of DTL-SA-ResNet50, so DTL-CBAM-ResNet50 is more suitable for PC, while DTL-SA-ResNet50 is more suitable for mobile terminal.

(2) Compared with DTL-ResNet50 model, the precision of the proposed model was 4.1% higher, and the time required to process a single image was 0.06 s shorter; Recall increased by 3.97%, and F1 Score has increased, taking into account the Precision and Recall, and the test model has been evaluated more comprehensively. Therefore, the attention mechanism improved the model precision, reduced the running time, and improved the identification effect of the model.

(3) Compared with convolutional neural networks after dual transfer learning, the precision of the new model was significantly improved, its identification precision was 3.19% higher than the best result of the other models, and the time consumption of a single image was reduced by at least 0.58 s; Recall and F1 Score have also increased. The proposed method effectively improves the performance of the model.

(4) Confusion matrices of eight models demonstrated that the overall identification precision of DTL-SE-ResNet50 was higher, and its effect was better. However, the method showed lower identification precision for tomato powdery mildew and cucumber downy mildew against the simple background. This result reflected the light degree of the disease caused healthy leaves similar to the other diseased leaves, resulting in low identification precision.

The DTL-SE-ResNet50 vegetable disease identification model based on transfer learning and an attention mechanism was trained with a large amount of data, maintained a good precision, and required a shorter running time. All databases in this paper were collected in practical application scenarios. In the future, we will continue to improve the disease identification model, increasing the identification of disease types and crop types, so that the disease identification model has a wider range of applications. At the same time, the model is simplified so that it can be deployed to mobile terminals to facilitate farmers' application.

Declaration of Competing Interest

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

Acknowledgements

Authors would like to acknowledge the financial support of the National Natural Science Foundation of China (NO. 62176261).

References

Kulkarni, O., 2018. Crop Disease Detection Using Deep Learning. In: In: Fourth International Conference on Computing, Communication Control and Automation (ICCUBEA), pp. 16–18. https://doi.org/10.1109/ICCUBEA.2018.8697390.

Abade, A., Ferreira, P.A., Vidal, F.D., 2021. Plant diseases identification on images using convolutional neural networks: A systematic review. Comput. Electron. Agric. 185, 1–31. https://doi.org/10.1016/j.compag.2021.106125.

Brahimi, M., Boukhalfa, K., Moussaou, A., 2017. Deep learning for tomato diseases: classification and symptoms visualization. Appl. Artif. Intell. 31 (4), 299–315. https://doi.org/10.1080/08839514.2017.1315516.

Gutierrez, S., Hernandez, I., Ceballos, S., Barrio, I., Diez-Navajas, A.M., Tardaguila, J., 2021. Deep learning for the differentiation of downy mildew and spider mite in grapevine under field conditions. Comput. Electron. Agric. 182, 1–9. https://doi.org/10.1016/j.compag.2021.105991.

Anagnostis, A., Tagarakis, A.C., Asiminari, G., Papageorgiou, E., Kateris, D., Moshou, D., Bochtis, D., 2021. A deep learning approach for anthracnose infected trees classification in walnut orchards. Comput. Electron. Agric. 182, 1–11. https://doi.org/10.1016/j.compag.2021.20598.

Mohit, A., Rohit, K., Gaurav, S., Suneet, K., 2019. FCNN-LDA: A Faster Convolution Neural Network model for Leaf Disease identification on Apple's leaf database. In: In: 12th International Conference on Information & Communication Technology and System (ICTS), pp. 246–251. https://doi.org/10.1109/ICTS.2019.8850964.

Rehman, Z.u., Khan, M.A., Ahmed, F., Damaševičius, R., Naqvi, S.R., Nisar, W., Javed, K., 2021. Recognizing apple leaf diseases using a novel parallel real-time processing framework based on MASK RCNN and transfer learning: An application for smart agriculture. IET Image Proc. 15 (10), 2157–2168. https://doi.org/10.1049/ipr2.12183.

Rahman, C.R., Arko, P.S., Ali, M.E., Iqbal Khan, M.A., Apon, S.H., Nowrin, F., Wasif, A., 2020. Identification and identification of rice diseases and pests using convolutional neural networks. Biosyst. Eng. 194, 112–120. https://doi.org/10.1016/j.biosystemseng.2020.03.020.

Li, D., Wang, R., Xie, C., Liu, L., Zhang, J., Li, R., Wang, F., Zhou, M., Liu, W., 2020. A identification Method for Rice Plant Diseases and Pests Video Detection Based on Deep Convolutional Neural Network. Sensors 2020, 20, 578, 1–21. DOI: 10.3390/s20030578.

Zhang, X., Qiao, Y., Meng, F., Fan, C., Zhang, M., 2018. Identification of maize leaf diseases using improved convolutional neural network. IEEE Access 6, 30370–30377. https://doi.org/10.1109/ACCESS.2018.2844405.

Liu, B., Ding, Z., Zhang, Y., He, D., He, J., 2020. Kiwifruit Leaf Disease Identification Using Improved Deep Convolutional Neural Networks. In: In: 2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC), pp. 1267–1272. https://doi.org/10.1109/COMPSAC48688.2020.00-82.

Li, X., Pang, T., Xiong,B., Liu, W., Liang, P., Wang, T., 2017. Convolutional Neural Networks Based Transfer Learning for Diabetic Retinopathy Fundus Image

Classification. In: 2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI 2017), 1-11. DOI: 10.1109/CISP-BMEI.2017.8301998.

Wang, S., Xie, S., Chen, X., Guttery, D.S., Tang, C., Sun, J., Zhang, Y., 2019. Alcoholism identification based on an AlexNet transfer learning model. Front. Psychiatry 10, 1–13. https://doi.org/10.3389/fpsyt.2019.002005.

Ishengoma, F.S., Rai, I.A., Said, R.N., 2021. Identification of maize leaves infected by fall armyworms using UAV-based imagery and convolutional neural networks. Comput. Electron. Agric. 184, 1–8. https://doi.org/10.1016/j.compag.2021.106124.

Pan, S., Yang, Q., 2010. A Survey on Transfer Learning. IEEE Trans. Knowl. Data Eng. 22 (10), 1345–1359. https://doi.org/10.1109/TKDE.2009.191.

Tian, J., Zhang, Y., Wang, Y., Wang, C., Zhang, S., Ren, T., 2019. A Method of Corn Disease Identification Based on Convolutional Neural Network. International Symposium on Computational Intelligence and Design (ISCID) 12, 245–248. https://doi.org/10.1109/ISCID.2019.00063.

Lee, S., Goeau, H., Bonnet, P., Joly, A., 2020. New perspectives on plant disease characterization based on deep learning. Comput. Electron. Agric. 170, 1–12. https://doi.org/10.1016/j.compag.2020.105220.

Espejo-Garcia, B., Mylonas, N., Athanasakos, L., Fountas, S., Vasilakoglou, I., 2020. Towards weeds identification assistance through transfer learning. Comput. Electron. Agric. 171, 1–10. https://doi.org/10.1016/j.compag.2020.105306.

Maeda-Gutiérrez, V., Galván-Tejada, C.E., Zanella-Calzada, L.A., Celaya-Padilla, J.M., Galvan-Tejada, J.I., Gamboa-Rosales, H., Luna-Garcia, H., Magallanes-Quintanar, R., Mendez, C.A.G., Olvera-Olvera, C.A., 2020. Comparison of convolutional neural network architectures for classification of tomato plant diseases. Appl. Sci. Basel 10 (4), 1245. https://doi.org/10.3390/app10041245.

Trang, K., TonThat, L., Gia Minh Thao, N., Tran Ta Thi, N., 2019. Mango Diseases Identification by a Deep Residual Network with Contrast Enhancement and Transfer Learning. In: 2019 IEEE Conference on Sustainable Utilization and Development in Engineering and Technologies (CSUDET), 138-142. DOI: 10.1109/CSUDET47057.2019.9214620.

Barbedo, J.G.A., 2018. Impact of database size and variety on the effectiveness of deep learning and transfer learning for plant disease classification. Comput. Electron. Agric. 153, 46–53. https://doi.org/10.1016/j.compag.2018.08.013.

Chen, J., Chen, J., Zhang, D., Sun, Y., Nanehkaran, Y.A., 2020. Using deep transfer learning for image-based plant disease identification. Comput. Electron. Agric. 173, 1–11. https://doi.org/10.1016/j.compag.2020.105393.

Argueso, D., Picon, A., Irusta, U., Medela, A., San-Emeterio, M.G., Bereciartua, A., Alvarez-Gila, A., 2020. Few-Shot Learning approach for plant disease classification using images taken in the field. Comput. Electron. Agric. 175, 1–8. https://doi.org/10.1016/j.compag.2020.105542.

Wang, D., Wang, J., 2021. Crop disease classification with transfer learning and residue networks. Trans. Chinese Soc. Agric. Eng. (Transactions of the CSAE), 37(4), 199-207.(in Chinese with English abstract) . DOI:10.11975/j.issn.1002-6819.2021.4.024.

Wang, F., Jiang, M., Qian, C., Yang, S., Li, C., Zhang, H., Wang, X., Tang, X., 2017. Residual Attention Network for Image Classification. In: In: 30th IEEE Conference on Computer Vision and Pattern identification, pp. 6450–6458. https://doi.org/10.1109/CVPR.2017.683.

Li, Y., Nie, J., Chao, X., 2020b. Do we really need deep CNN for plant diseases identification? Comput. Electron. Agric. 178, 1–7. https://doi.org/10.1016/j.compag.2020.105803.

Li, Y., Chao, X., 2021. Semi-supervised few-shot learning approach for plant diseases recognition. Plant Methods 17 (1), 1–10. https://doi.org/10.1186/s13007-021-00770-1.

Li, Y., Yang, J., 2021. Meta-learning baselines and database for few-shot classification in agriculture. Comput. Electron. Agric. 182, 1–9. https://doi.org/10.1016/j.compag.2021.106055.

Li, X., Rai, L., 2020. Apple Leaf Disease Identification and Classification using ResNet Models. In: In: IEEE 3rd International Conference on Electronic Information and Communication Technology (ICEICT), pp. 738–742. https://doi.org/10.1109/ICEICT51264.2020.9334214.

Yu, H., Son, C., 2020. Leaf spot attention network for apple leaf disease identification. In: In: IEEE Computer Society Conference on Computer Vision and Pattern identification Workshops (CSPRW), pp. 229–237. https://doi.org/10.1109/CVPRW50498.2020.00034.

He, C., 2020. Image Compressive Sensing via Multi-scale Feature Extraction and Attention Mechanism. In: In: 2020 International Conference on Intelligent Computing, Automation and Systems (ICICAS), pp. 266–270. https://doi.org/10.1109/ICICAS51530.2020.00061.

Liu, Y., Gao, G., 2021. Identification of multiple leaf diseases using improved SqueezeNet model. Transactions of the Chinese Society of Agricultural Engineering (Transcations of the CSAE), 37(2), 187-195. (in Chinese with English abstract). DOI:10.11975/j.issn.1002-6819.2021.2.022.

Maeda-Gutierrez, V., Galvan-Tejada, C.E., Zanella-Calzada, L.A., Celaya-Padilla, J.M., Galvan-Tejada, J.I., Gamaboa-Rosales, H., Luna-Garica, H., Magallanes-Quintanar, R., Mendez, C.A.G., Olvera-Olvera, C.A., 2020. Comparison of Convolutional Neural Network Architectures for Classification of Tomato Plant Diseases. Appl. Sci. Basel 10 (4), 1–15. https://doi.org/10.3390/app10041245.

Geng, L., Xu, W., Zhang, F., Xiao, Z., Liu, Y., 2018. Dried Jujube Classification Based on a Double Branch Deep Fusion Convolution Neural Network. Food Sci. Technol. Res. 24 (6), 1007–1015. https://doi.org/10.3136/fstr.24.1007.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep Residual Learning for Image identification. In: In: 2016 IEEE Conference on Computer Vision for Image identification, pp. 770–778. https://doi.org/10.1109/CVPR.2019.90.

Tan, X., Le, Q., 2019. EfficientNet: Rethinking model scaling for convolutional neural networks. In: In: 36th International Conference Machine Learning(ICML), pp. 10691–10700.

Bao, Y., Liu, W., Niu, C., Li, R., Zhang, H., 2021. Scene Classification of Optical Remote Sensing Images Joint Ensemble Learning and EfficientNet. In: Computer Engineering, 47(10), 226-235. DOI: 10.19678/j. issn. 1000-3428.0059128.

Chen, J., Zhang, D., Zeb, A., Nanehkaran, Y.a., 2021. Identification of rice plant diseases using lightweight attention networks. Expert Syst. Appl. 169, 1–13. https://doi.org/10.1016/j.eswa.2020.114514.

Zhu, X., Cheng, D., Zhang, Z., Lin, S., Dai, J., 2019. An Empirical Study of Spatial Attention Mechanisms in Deep Networks. In: In: Proceedings of the IEEE International Conference on Computer Vision, pp. 6687–6696. https://doi.org/10.1109/ICCV.2019.00679.

Hu, J., Shen, L.i., Albanie, S., Sun, G., Wu, E., 2020. Squeeze-and-Excitation Networks. IEEE Trans. Pattern Anal. Mach. Intell. 42 (8), 2011–2023. https://doi.org/10.1109/TPAMI.2019.2913372.

Chen, Q., Liu, L., Han, R., Qian, J., Qi, D., 2019. Image identification method on high speed railway contact network based on YOLO v3 and SENet. In: In: 2019 Chinese Control Conference (CCC), pp. 8772–8777.

Qin, R., Fu, X., Lang, P., 2020. PolSAR Image Classification Based on Low-Frequency and Contour Subbands-Driven Polarimetric SENet. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 13, 4760–4773. https://doi.org/10.1109/JSTARS.2020.3015520.

Bao, W., Yang, X., Liang, D., Hu, G., Yang, X., 2021b. Lightweight convolutional neural network model for field wheat ear disease identification. Comput. Electron. Agric. 189, 1–15. https://doi.org/10.1016/j.compag.2021.106367.

Woo, S., Park, J., Lee, J., Kweon, I., 2018. CBAM: Convolutional Block Attention Module. In: Computer Vision-ECCV2018 11211, 3–19. https://doi.org/10.1007/978-3-030-01234-2_1.

Wang, M., Wu, Z., Zhou, Z., 2021. Fine-grained Identification Research of Crop Pests and Diseases Based on Improved CBAM via Attention. Trans. Chinese Soc. Agric. Mach. 4, 239–247. https://doi.org/10.6041/j.issn.1000-1298.2021.04.025.

Dai, Q., Cheng, X.i., Qiao, Y., Zhang, Y., 2020. Crop Leaf Disease Image Super-Resolution and Identification with Dual Attention and Topology Fusion Generative Adversarial Network. IEEE Access 8, 55724–55735. https://doi.org/10.1109/ACCESS.2020.2982055.

Zhao, Q., Li, L., Zhang, M., Lan, T., Sigrimis, N.A., 2020. Water Stress Diagnosis Algorithm of Greenhouse Tomato Based on Fine-tuning Learning. Trans. Chinese Soc. Agric. Mach. 51, 340–347. (in Chinese with English abstract). DOI:10.6041/J.ISSN.1000-1298.2020.s1.040.

Fan, X., Xu, Y., Zhou, J., Li, Z., Peng, X., Wang, X., 2021. Detection system for grape leaf diseases based on transfer learning and updated CNN. Trans. Chinese Soc. Agric. Eng. (Transactions of the CSAE) 37 (6), 151–159. https://doi.org/10.11975/j.issn.1002-6819.2021.06.019.