Cognitive Robotics 2 (2022) 155–163

Contents lists available at ScienceDirect

Cognitive Robotics

journal homepage: http://www.keaipublishing.com/en/journals/cognitive-robotics/

# Research on plant disease identification based on CNN

Xuewei Sun[a], Guohou Li[a,*], Peixin Qu[a], Xiwang Xie[b], Xipeng Pan[c], Weidong Zhang[a]

[a] School of Information Engineering, Henan Institute of Science and Technology, Xinxiang 453000, China
[b] Information Science and Technology College, Dalian Maritime University, Dalian 116026, China
[c] School of Computer Science and Information Security, Guilin University of Electronic Technology, Guilin 541000, China

ARTICLE INFO

*Keywords:*
Identification of plant diseases
Convolutional neural network
EfficientNet
Neural architecture search
Focal loss function

ABSTRACT

Traditional digital image processing methods extract disease features manually, which have low efficiency and low recognition accuracy. To solve this problem, In this paper, we propose a convolutional neural network architecture FL-EfficientNet (Focal loss EfficientNet), which is used for multi-category identification of plant disease images. Firstly, through the Neural Architecture Search technology, the network width, network depth, and image resolution are adaptively adjusted according to a group of composite coefficients, to improve the balance of network dimension and model stability; Secondly, the valuable features in the disease image are extracted by introducing the moving flip bottleneck convolution and attention mechanism; Finally, the Focal loss function is used to replace the traditional Cross-Entropy loss function, to improve the ability of the network model to focus on the samples that are not easy to identify. The experiment uses the public data set new plant diseases dataset (NPDD) and compares it with ResNet50, DenseNet169, and EfficientNet. The experimental results show that the accuracy of FL-EfficientNet in identifying 10 diseases of 5 kinds of crops is 99.72%, which is better than the above comparison network. At the same time, FL-EfficientNet has the fastest convergence speed, and the training time of 15 epochs is 4.7 h.

## 1. Introduction

China is a large traditional agricultural country. Agricultural production is related to the national economy and the people's livelihood. Ensuring agricultural production is the foundation of China. However, in the growth process of crops, affected by factors such as environment, climate, and soil, they have been threatened by diseases, which seriously affect grain yield and quality. Therefore, if we can identify plant diseases quickly, nondestructive and effectively, and apply drugs accurately, we can reduce economic losses to a great extent [1].

With the continuous development of artificial intelligence, deep learning has been widely used in the field of agriculture [2]. At present, the methods of plant disease identification are mainly divided into traditional methods and deep learning methods. Traditional methods are based on Support Vector Machine (SVM), K-Nearest Neighbor (KNN), K-means clustering algorithm, and other algorithms et al. Such as Thaiyalnayaki K [3] and others took cotton in the United States as the research object. Through Multi-Layer Perceptron (MLP), combined with morphological segmentation, pattern matching, tone matching, and other technologies, KNN and SVM were used as an overlapping classification to identify plant diseases such as soybean through SVM, with an accuracy of 94.1%. Prashar K [4] and others took cotton in the United States as the research object. Through multi-layer perceptron, combined with morphological segmentation, pattern matching, tone matching, and other technologies, KNN and SVM were used as an overlapping

[*] Corresponding author at: Henan Institute of Science and Technology, China.
*E-mail address:* liguohou_hnkjxy@163.com (G. Li).

https://doi.org/10.1016/j.cogr.2022.07.001
Received 17 May 2022; Received in revised form 1 July 2022; Accepted 4 July 2022
Available online 6 July 2022

**Table 1-1**
Advantages and disadvantages of single optimization strategy.

| Optimization strategy | advantage | shortcoming |
|---|---|---|
| Increase depth | The extracted features are richer and have good generalization ability. | If the network is too deep, the gradient will disappear and the training will be difficult. |
| Increase width | The training is easier and the fine-grained features extracted are higher. | If the network is shallow but the width is large, the network cannot learn deeper features. |
| Add resolution | The higher fine-grained feature template is obtained and deeper features can be learned. | A higher resolution is affected by the marginal effect of accuracy, which increases the amount of calculation. |

classification to identify plant diseases, and the recognition accuracy was 96%. Kumari C [5] and others took cotton and tomato as research objects, extracted the characteristics of contrast, correlation, mean, and variance, and used the K-means method for cluster analysis, with an average accuracy of 90%. Deep learning mainly uses convolutional neural networks (CNN), such as AlexNet, GoogLeNet, ResNet, and other network models. In recent years, some scholars have introduced the attention mechanism [6,7] into the image classification task, such as Zhao Y [8], who embedded the attention mechanism into the residual method based on neural structure search technology, and confirmed that higher accuracy can be achieved through oversampling, subsampling balance processing and grey transformation of training data. Sagar A [9] et al. Combined SE block with MobileNet and proposed SE-MobileNet. Combined with migration learning, the accuracy of SE-MobileNet on the plant Village dataset is 99.33%. By introducing migration learning and fine-tuning the network parameters of ResNet50, Chen J [10] et al. Achieved an accuracy of 98.2 on the plant Village dataset.

It can be seen from the above literature that before the convolutional neural network was widely used, most studies used traditional machine learning methods. Machine learning has low efficiency and unsatisfactory recognition accuracy compared with deep learning. It needs to manually extract the texture, shape, color, and other features of crop leaves, and also needs to combine histogram or Principal Components Analysis (PCA)dimensionality reduction, most disease identification studies have gradually abandoned the traditional machine learning and adopted more efficient deep learning methods. At present, most network models are improved based on the existing excellent networks, and the recognition accuracy is also improving. However, most networks do not consider the impact of network model parameters and training time on the recognition efficiency. If these models are transplanted to the mobile terminal in the future, the model recognition efficiency will be low, can't meet the daily needs, and can't be put into actual production, to solve the above problems, combined with the EfficientNet-B0 network model, this paper introduces Focal loss function, FL-EfficientNet is proposed. The network model identifies 10 disease images of five types of crops and combines migration learning, DropConnect, parameter adjustment, and other strategies to greatly reduce the number of model parameters and training time of the network, which provides a certain reference value for subsequent deployment to the mobile terminal.

## 2. FL-efficientNet

### 2.1. EfficientNet

EfficientNet is a group of CNN-based backbone feature extraction networks released by Tan M [11] et al in 2019. The paper points out that the existing methods to improve the performance of CNN are based on increasing depth, width, and resolution. However, if only one parameter is optimized, it will lead to problems such as gradient loss, gradient explosion, a large amount of calculation, and so on. The impact of single optimization depth, width, and resolution is shown in Tab. 1-1.

### 2.2. Neural network architecture search technology

The neural Network Architecture Search (NAS) [12] technology is used to pre-limit the content and calculation amount. The optimal solution between depth, resolution, and channel width is searched through the grid, and a set of fixed scaling coefficients are found($\alpha$, $\beta$ and $\gamma$), By using the mixed scaling method, the depth, width and resolution are uniformly scaled by using the mixed factor Ø to balance the network dimension and improve the overall performance of the network model.

$$\begin{aligned}
&\text{depth}: d = \alpha^{\phi} \\
&\text{width}: w = \alpha^{\phi} \\
&\text{resolution}: r = \alpha^{\phi} \\
&s.t.\alpha \cdot \beta^2 \cdot \gamma^2 \approx 2 \\
&\alpha \ge 1, \beta \ge 1, \gamma \ge 1
\end{aligned} \tag{1}$$

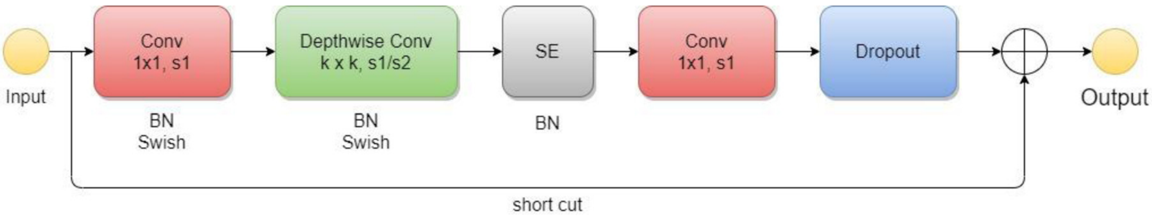


**Fig. 1-1.** MBConv module structure diagram.

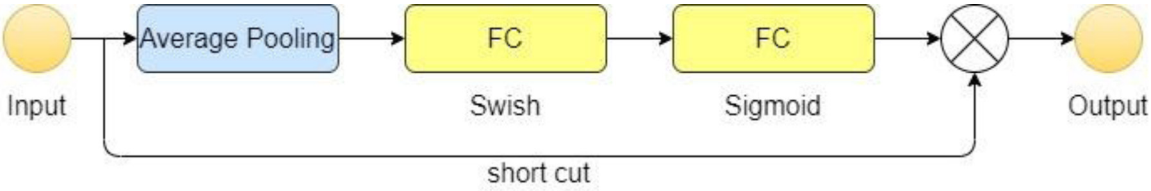


**Fig. 1-2.** SE block structure diagram.

In Eq. (1), The $\alpha$,$\beta$ and $\gamma$ are constants obtained by network search, and the mixing factor is intervened manually. The calculation amount of convolution operation is the same as $d$, $w^2$, $\gamma^2$.Under this condition, the calculation amount of the network is about $2^{\emptyset}$ times that of the previous one, and the total theoretical calculation amount (FLOPs) is about($\alpha \cdot \beta^2 \cdot \gamma^2$).When limited $\alpha \cdot \beta^2 \cdot \gamma^2 \approx 2$, FLOPs is increased by $2^{\emptyset}$ times for any $\emptyset$.

### 2.3. Mobile inverted Bottleneck convolution and SE block

The Mobile inverted Bottleneck Convolution (MBConv) module [13] is used for reference in the MobileNet-V3 network. Meanwhile, an attention mechanism is introduced to highlight the more important features of disease characteristics. The attention mechanism is mainly realized by the Squeeze and Excitation (SE) block.

The MBConv module uses an Inverted Residual structure and is the core module of the network in this paper. The module contains a 1*1 boosted convolutional layer that uses a BN (Batch Normalization) strategy and a SiLU (Sigmoid Weighted Liner Unit) activation function. A Depthwise Separable Convolution module, also using BN and SiLU, is also included, followed by an SE block for implementing the attention mechanism, followed by a 1*1 convolutional layer, a downscaling of the feature map, and finally the addition of a Dropout layer to mitigate overfitting. The structure of the MBConv module is shown schematically in Fig. 1-1.

The SE block focuses on the relationship between different channels, which consists of a global average pooling and two fully connected layers [14]. Through the fully connected network, the feature weight is automatically learned according to the loss function, thus increasing the weight of the effective channel, and realizing the attention mechanism by learning the importance of different channel characteristics. The structure of the SE block is shown schematically in Fig. 1-2.

### 2.4. DropConnect

Because the EfficientNet model scales the depth, width, and resolution at the same time, the model is complex. Although the dropout layer is added to the model, the overfitting phenomenon will occur during the training process. Therefore, DropConnect is added to the model. When the feature map shape of the input MBConv structure has the same parameters as the output feature map shape, the shortcut connection is enabled; If you enable the shortcut connection, DropConnect is enabled at the same time. Unlike dropout, dropout randomly discards the output of hidden nodes during training, while DropConnect randomly discards the input of the hidden layer. By comparison, DropConnect is more effective and DropConnect’s drop rate is proportional to the parameter of the EfficientNet network model.

### 2.5. Focal loss function

Most of the data samples with the wrong classification are concentrated on the unbalanced samples. The number of samples used in this study varies from 2000 to 2500 from the disease category, but the number of samples varies greatly among different crop species, such as 11006 tomato samples and 4437 maize samples. Similar features will be extracted from the disease images of the same crop category in the process of feature extraction, which will lead to more accurate classification of crop categories with more samples, while the classification effect of the categories with fewer samples is not good. Therefore, this paper uses the Focal loss function [15] to replace the traditional Cross-Entropy (CE) loss function, which makes the network reduce the weight of the easily divided samples in the training process, and increase the weight of the hard classification samples so that the model is more focused on the difficult classification samples in the training process. In the feature extraction of neural networks, Focal loss can effectively solve the problem of unbalanced sample numbers and improve the average recognition rate of the network.

The Focal loss function is obtained by modifying the Cross-Entropy loss function. In the EfficientNet model, after regression through the softmax function, the function of the Cross-Entropy loss function is shown in formula (2).

$$\mathrm{CE}(x) = -\lg\left(\frac{e^{x_i}}{\sum_j e^{x_j}}\right) \tag{2}$$

In formula (2), $x$ is the eigenvalue, $i$ and $j$ represent the category number. The expression of the Focal loss function is shown in formula (3).

$$\mathrm{FL}\left(p_t\right) = -\left(1 - p_t\right)^{\gamma} \lg\left(p_t\right) \tag{3}$$

In formula (3), $p_t$ is the probability that the sample belongs to a certain category, $\gamma$ is the focusing parameter, where $0 \leq \gamma \leq 5$. $(1 - p_t)^{\gamma}$ Is the modulation coefficient. The value range is [0,1]. If the $p_t$ value is close to 1, the sample is easy to train, and if the $p_t$ value is close to 0, the sample is difficult to train. When $\gamma$=0, then $(1 - p_{\mathrm{t}})^{\gamma} = 1$. The Focal loss function is equivalent to the Cross-Entropy loss function. After many experiments, when $\gamma$=2, the training effect is the best. If the sample accuracy of prediction is very high (i.e. $p_t$ is very high), then the value of $(1 - p_{\mathrm{t}})^{\gamma}$ is very small, the value of the loss function will be very small, and the change rate of the loss function will be correspondingly small, and vice versa. The loss function of the FL-EfficientNet network is shown in formula (4).

$$\text{Focal Loss }(x) = -\left(1 - \frac{e^{x_i}}{\sum_j e^{x_j}}\right)^{\gamma} \mathrm{g}\left(\frac{e^{x_i}}{\sum_j e^{x_j}}\right) \tag{4}$$

### 2.6. Adam optimizer

Adam [16] optimizer is used to update the gradient. Adam optimizer combines the advantages of RMSProp optimizer and AdaGrand optimizer. Its updated formula is as follows:

$$w_t = w_{t-1} - \alpha * \frac{\widehat{m_t}}{\sqrt{\widehat{v_t}} + \epsilon} \tag{5}$$

In formula (5), $t$ is the number of updates, $\alpha$ Is the learning rate, $\epsilon$ Is a decimal number greater than 0, mainly used to prevent $\sqrt{\hat{v}_t} = 0$, $\hat{m}_t$ is the correction of $m_t$, $\hat{v}_t$ is the Correction of $v_t$.

The updated formulas of $\hat{m}_t$ and $\hat{v}_t$ are shown in Eqs. (6) and (7).

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t} \tag{6}$$

$$\widehat{v}_t = \frac{v_t}{1 - \beta_2^t} \tag{7}$$

In Eqs. (6) and (7), $\beta_1$ and $\beta_2$ is a constant to control the exponential decay, $m_t$ is the exponential moving mean of the gradient, which is obtained by the first moment of the gradient, $v_t$ is the square gradient, which is obtained by the second moment of the gradient.

Adam optimizer has a faster convergence speed, and fewer memory requirements and the scaling transformation of the gradient does not affect the parameter update. At present, it has become the mainstream gradient optimization algorithm and performs well in most classification tasks.

## 3. Model training and result analysis

### 3.1. Experimental data

This study uses the Kaggle official website to open the data set: the new plant diseases dataset. In this paper, 22427 pictures of crop leaves in JPG format are included in the data set, including 10 disease or health pictures of 5 kinds of crops, such as apple, cherry, and corn. The crop category and the corresponding picture quantity are shown in Table 2-1.

### 3.2. Pretreatment stage

Firstly, the repeated and invalid data in the data set are eliminated, and then the data set is randomly divided into the training set and test set according to the ratio of 8:2. According to the given label file, the data categories are divided, and the data sets of different diseases are stored in 10 folders respectively. The folder is named by the disease name, which is convenient for subsequent programming [17]. The RGB image of some pre-processed data sets is shown in Fig. 2-1.

To be better applied to the model, reduce training time and alleviate the convergence of overfitting merging [18,19], the data set needs to be preprocessed [20-23]. The preprocessing strategies used in this study include:

1) Normalization processing [24]: given coefficients ( [0.485, 0.456, 0.406], [0.229, 0.224, 0.225]), the normalized image has changed its pixel value range from the original [0, 255] to [0, 1]. The normalized image can accelerate the convergence of neural network and effectively alleviate the saturation of neuron output.

**Table 2-1**
Crop category and quantity.

| number | Disease category | Number of pictures |
|---|---|---|
| 1 | Corn Gray Spot | 2052 |
| 2 | Bacterial wilt of corn | 2385 |
| 3 | Tomato blight | 2181 |
| 4 | Tomato spot disease | 2284 |
| 5 | Bacterial spot | 2127 |
| 6 | Damage of tomato spider | 2176 |
| 7 | Tomato mosaic | 2238 |
| 8 | Apple cedar rust | 2200 |
| 9 | Grape Black Rot | 2360 |
| 10 | Potato late blight | 2424 |

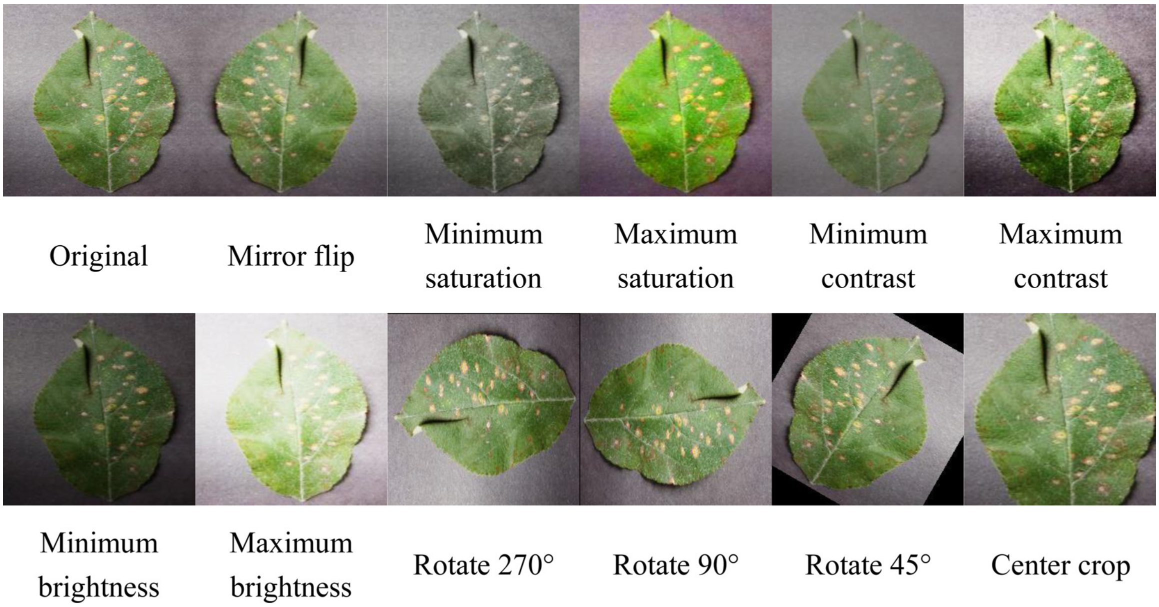


**Fig. 2-1.** The part of pre-processed image.

2) Random rotation: random rotation of random parts of data, such as 30 ° and 90 ° 180 °, etc.
3) Random image flip: randomly flip the original image according to the given probability(P=30%).
4) Brightness, saturation and contrast change: adjust the brightness and contrast of random data, so that the neural network has better generalization ability when facing images under different illumination backgrounds [25].
5) Center clipping: cut the centre of the random part of the picture, remove the edge background, keep only the middle leaves, and keep the picture size unchanged, to reduce the interference of the image background to the feature extraction process.
6) Add Gaussian noise: add Gaussian noise to the image to reduce the impact of noise on image recognition accuracy and enhance its generalization ability. The processed photos will have small white spots visible to the naked eye [26].

Finally, the pixels of all processed and unprocessed disease images are uniformly adjusted to 224*224 to adapt to the requirements of the neural network for image resolution. There are 12 pictures in Fig. 2-1. From left to right and from top to bottom, they are the original picture of Apple cedar rust, mirror flip, minimum saturation, maximum saturation, minimum contrast, maximum contrast, minimum brightness, maximum brightness, rotation 270°, rotation 90°, rotation 45°, and centre clipping. Limited to space, only one image of Apple cedar rust is preprocessed here. In addition, the above 12 images are added with Gaussian noise.

### 3.3. Experimental configuration and model parameters

The experimental hardware configuration is a computer equipped with AMD Ryzen 5 1600X six-core processor CPU and NVIDIA GeForce GTX 1650 GPU, equipped with windows10 64-bit operating system, deep learning framework, and version is PyTorch 1.9.1, using CUDA 11.1 and cuDNN 8.0.5 accelerate the training of network model by using GPU. The initial learning rate of the three network models is set to 0.001, and the learning rate attenuation strategy is introduced. The learning rate attenuation value (decay) after each parameter update is set to 0.001, the batch size is set to 16, and the training times (epoch) is set to 15; The performance of checkpoint is the best in the training process; Transfer learning is used to transfer the pre-learned knowledge of the three network

**Table 2-2**
Evaluating indicators of different network models.

| categorynetwork | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | Mean value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ResNet50 | Precision | 0.922 | 0.952 | 0.967 | 0.911 | 0.951 | 0.963 | 0.982 | 0.977 | 0.971 | 0.968 | 0.956 |
| | recall | 0.929 | 0.918 | 0.942 | 0.949 | 0.931 | 0.914 | 0.984 | 0.975 | 0.955 | 0.938 | 0.943 |
| DenseNet169 | Precision | **0.972** | 0.960 | 0.988 | 0.982 | 0.976 | 0.983 | 0.995 | **0.993** | 0.985 | 0.989 | 0.982 |
| | recall | 0.956 | **0.972** | 0.983 | **0.986** | **0.995** | 0.983 | 0.988 | 0.995 | 0.995 | 0.979 | 0.983 |
| EfficientNet | Precision | 0.959 | 0.979 | 0.983 | 0.976 | 0.983 | 0.984 | 0.997 | 0.982 | 0.992 | 0.989 | 0.982 |
| | recall | 0.971 | 0.957 | 0.975 | 0.970 | 0.987 | 0.993 | 0.993 | 0.982 | 0.993 | 0.977 | 0.979 |
| FL-EfficientNet | Precision | 0.959 | **0.980** | **0.995** | **0.982** | **0.985** | **0.986** | **0.997** | 0.990 | **0.995** | **0.993** | **0.986** |
| | recall | **0.975** | 0.966 | **0.997** | 0.984 | 0.988 | **0.997** | **0.995** | **0.997** | **0.997** | **0.995** | **0.989** |

**Table 2-3**
Accuracy and loss of different network models.

| model | Test set accuracy | Test set loss | Training duration |
|---|---|---|---|
| ResNet50 | 0.951 | 0.413 | 20049s (about 5.6 h) |
| DenseNet169 | 0.990 | 0.082 | 24741s (about 6.9 h) |
| EfficientNet | 0.984 | 0.105 | 17035S (about 4.7 h) |
| FL-EfficientNet | **0.997** | **0.053** | **16820s (about 4.7 h)** |

models in the ImageNet data set to this experiment. To compare the improvement effect, except that the FL-EfficientNet network uses the Focal loss function, other network models use the CE loss function.

### 3.4. Experimental results and analysis

In this paper, ResNet50, DenseNet169, and EfficientNet were used as the experimental NPDD data set. The data set has 22427 pictures of the leaves of crops, including 10 disease categories of 5 crops, and the proportion of training set and test set is 8:2. In addition, precision and recall are used as the evaluation indexes of the model to evaluate the performance of the experimental network. The larger the precision and recall values, the higher the model quality, and the formula is shown in formula (8)~(10).

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (8)$$

$$\text{Precision} = \frac{TN}{FP+TN} \quad (9)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (10)$$

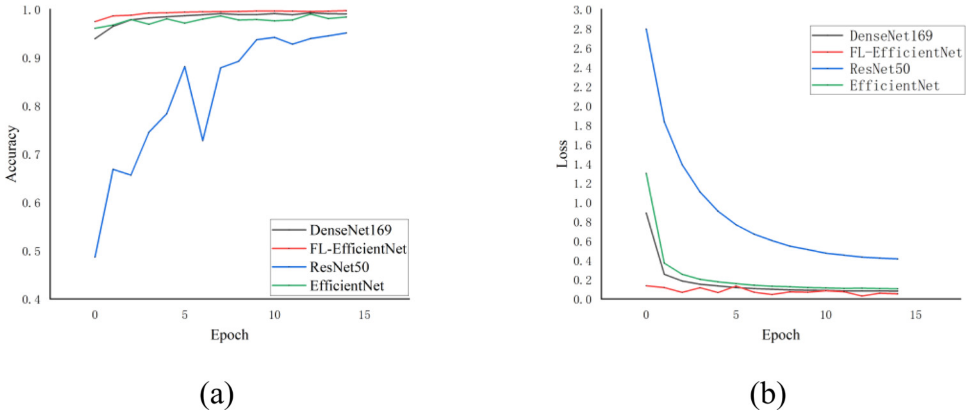
In formula (8)~(10), TP is the true positive, TN is the true negative, FP is the false positive, and FN is the false negative. The accuracy and recall rate of each network model is shown in Table 2-2.

In Table 2-2, FL-EfficientNet has higher precision and recall rate than the ResNet50 network. ResNet50 is a classic CNN structure. It uses residual structure to solve the gradient disappearance phenomenon when the network depth is too deep, but it does not consider the influence of network width and image resolution on classification results, and it can't extract fine-grained features of images, so the classification effect is far less than FL-EfficientNet and DenseNet169. There is little difference between FL EfficientNet and Densenet169 evaluation index values, but in general, FL-EfficientNet is the best. Only individual indicators of individual categories are not as good as DenseNet169, such as the accuracy rate of corn ash spot (No.1), recall rate of corn bacterial wilt (No.2), recall rate of tomato spot disease (No.4), recall rate of tomato scab disease (No.5) Accuracy of Apple cedar rust (No.8). Since DenseNet169 emphasizes the integration of the characteristics of each channel and the enhancement of the relationship between the middle layer and the layer of the network, especially the reuse of features through the branch structure, FL-EfficientNet uses the composite coefficient to uniformly scale the network width, network depth, and input image resolution, and emphasizes the balance between the three parameters, and has a slight shortage in feature reuse, This leads to the accuracy of individual categories slightly higher than EfficientNet.

The total quantitative evaluation indexes of ResNet50, DenseNet169, EfficientNet, and FL-EfficientNet in the NPDD data set are shown in Table 2-3 and Fig. 2-2.

Combined with Fig. 2-2, it is found that ResNet50 is close to convergence at the 13th epoch in the training process, and the convergence speed is far lower than DenseNet169 and FL-EfficientNet, which is mainly determined by the overall performance of the network. ResNet50 has the largest number of parameters and does not consider the impact of network width and core input image resolution on the network. The fluctuation of accuracy rate during ResNet50 training is because the hyper-parameter falls into the local optimal solution and jumps out of the local optimal solution after continuous iteration. DenseNet169, EfficientNet, and FL-EfficientNet almost converge at the fourth epoch, and the convergence speed is very fast, which is mainly due to the high performance of the network itself. In addition, migration learning is used to load the learned features into new tasks to accelerate the convergence of the network model. Although ResNet50 also uses transfer learning, the effect is not obvious because of its large number of parameters.

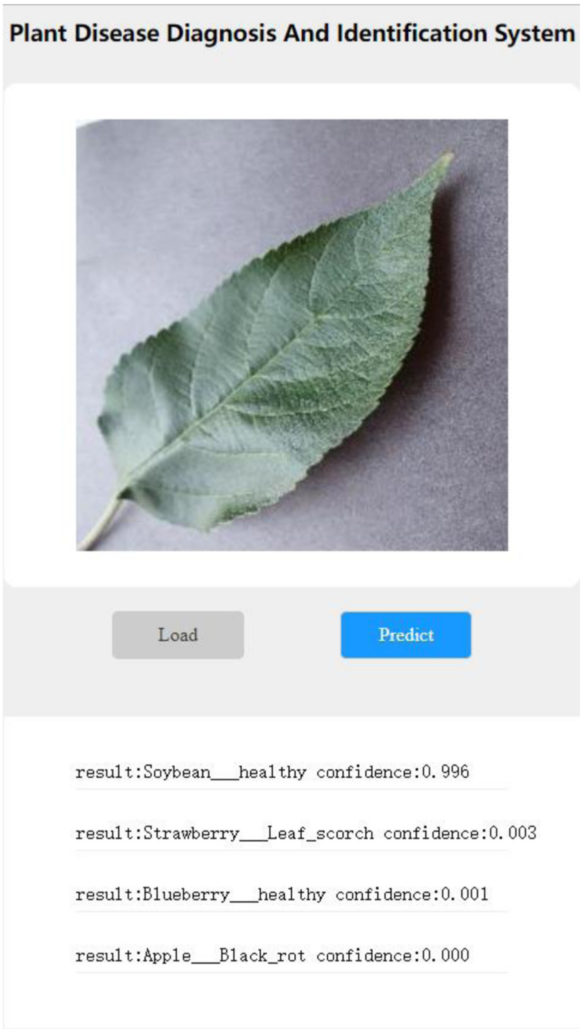
Combined with Table 2-3, it is found that FL-EfficientNet has the highest accuracy, which is due to the high classification accuracy of the network itself. By introducing Focal loss, the network focuses more on the categories with fewer samples. At the same time,



(a)

(b)

**Fig. 2-2.** Loss and accuracy curves in the training process of different network models.



Plant Disease Diagnosis And Identification System

```
result:Soybean___healthy confidence:0.996

result:Strawberry___Leaf_scorch confidence:0.003

result:Blueberry___healthy confidence:0.001

result:Apple___Black_rot confidence:0.000
```

**Fig. 2-3.** Image results of the diagnosis.

the DropConnect strategy is introduced to prevent overfitting to a certain extent. Moreover, the network uses NAS technology to dynamically adjust the network, and the number of network layers will not be too deep or too shallow, the network height will not be too high or too low, and the image resolution will not be too large or too small, which minimizes the number of network parameters. ResNet50 can't guarantee the recognition accuracy well on the premise of short training time, DenseNet169 can't consider the training time on the premise of ensuring the recognition accuracy, and FL-EfficientNet has slightly better recognition accuracy and training time than the original EfficientNet on the premise of considering to the training time and recognition accuracy, which not only has the highest recognition accuracy but also has the shortest training time.

To verify the recognition effect, we used PyTorch and Flask to build a simple web service, as shown in Fig. 2-3. After opening the web service, open the browser, enter the loopback address and the port set in advance, follow the prompts, click "Load" to select the disease picture to be predicted, and click "Predict" to display the prediction result. In Fig. 2-3, it is obvious that the result is "Soybean-healthy", and its prediction confidence is "99.6%".

## 4. Conclusion

Plant disease image recognition is a deep combination of computer vision and agriculture, which has a good application prospect. The plant disease identification and classification method based on the FL-EfficientNet network proposed in this paper solves the problem of imbalance in the number of samples of different kinds of plant diseases by introducing the Focal loss function in the task of multi-class plant disease classification, and effectively improves the accuracy of the network model in identifying disease types with a small number of samples, DropConnect is used to effectively reduce the occurrence of overfitting. Finally, the FL-EfficientNet network model is better than the comparison network in terms of convergence speed and recognition accuracy, with an average accuracy of 99.72%. According to the evaluation indicators of accuracy and recall rate, it can be seen that the FL-EfficientNet model is effective. The comparison with traditional algorithms shows that the model in this paper has faster convergence speed, accuracy, and timeliness, which lays a foundation for subsequent transplantation to the mobile terminal.

## 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.

## Acknowledgement

This research was supported in part by National Natural Science Foundation of China (Grant No. 62002082), Guangxi Natural Science Foundation (Grant No. 2020GXNSFBA238014), Guangxi University Young and Middle-aged Teachers' Research Ability Improvement Project (Grant No. 2020KY05034), Key specialized research and development program of science and technology in Henan province (Grant No. 202102210388)

## References

[1] J Chen, J Chen, D Zhang, et al., Using deep transfer learning for image-based plant disease identification [J], Comput. Electron. Agricult. 173 (2020) 105393.
[2] H Lu, Y Li, M Chen, et al., Brain intelligence: go beyond artificial intelligence [J], Mobile Networks Appl. 23 (2) (2018) 368–375.
[3] K. T J Christeena, Classification of plant disease using SVM and deep learning [J], Mater. Today 47 (P1) (2021) 468–470.
[4] K Prashar, R Talwar, C. Kant, CNN based on overlapping pooling method and multi-layered learning with SVM & KNN for American cotton leaf disease recognition [C], in: 2019 International Conference on Automation, Computational and Technology Management (ICACTM), IEEE, 2019, pp. 330–333.
[5] C U Kumari, S J Prasad, G. Mounika, Leaf disease detection: feature extraction with K-means clustering and classification with ANN [C], in: 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC), IEEE, 2019, pp. 1095–1098.
[6] H Wang, G Xu, X Pan, et al., Multi-task generative adversarial learning for nuclei segmentation with dual attention and recurrent convolution [J], Biomed. Signal Processing Control 75 (2022) 103558.
[7] H Wang, G Xu, X Pan, et al., Attention-inception-based U-Net for retinal vessel segmentation with advanced residual [J], Comput. Electrical Eng. 98 (2022) 107670.
[8] Y Zhao, J Chen, X Xu, et al., SEV-Net: residual network embedded with attention mechanism for plant disease severity detection [J], Concurr. Comput. 33 (10) (2021) e6161.
[9] A Sagar, D. Jacob, On using transfer learning for plant disease detection [J], bioRxiv (2021) 2020–2025.
[10] J Chen, D Zhang, M Suzauddola, et al., Identification of plant disease images via a squeeze and excitation MobileNet model and twice transfer learning [J], IET Image Processing 15 (5) (2021) 1115–1127.
[11] M Tan, Q Le, Efficientnet: Rethinking model scaling for convolutional neural networks [C], in: International Conference On Machine Learning, PMLR, 2019, pp. 6105–6114.
[12] Zoph B, Le Q V. Neural architecture search with reinforcement learning [J]. arXiv preprint arXiv:1611.01578, 2016.
[13] H Lu, Y Zhang, Y Li, et al., User-oriented virtual mobile network resource management for vehicle communications [J], IEEE Trans. Intelligent Transport. Syst. 22 (6) (2020) 3521–3532.
[14] H Lu, Y Tang, Y.DRRS-BC Sun, Decentralized routing registration system based on blockchain [J], IEEE/CAA J. Automatica Sinica 8 (12) (2021) 1868–1876.
[15] T Y Lin, P Goyal, R Girshick, et al., Focal loss for dense object detection [J], IEEE Trans. Pattern Anal. Mach. Intell. 42 (2) (2020) 318–327.
[16] U M Khaire, R. Dhanalakshmi, High-dimensional microarray dataset classification using an improved adam optimizer (iAdam) [J], J. Ambient Intell. Humanized Comput. 11 (11) (2020) 5187–5204.
[17] W Zhang, L Dong, W. Xu, Retinex-inspired color correction and detail preserved fusion for underwater image enhancement [J], Comput. Electron. Agricult. 192 (2022) 106585.
[18] Cong R, Yang N, Li C, et al. Global-and-Local Collaborative Learning for Co-Salient Object Detection [J]. arXiv preprint arXiv:2204.08917, 2022.
[19] C Li, C Guo, L H Han, et al., Low-light image and video enhancement using deep learning: a survey [J], IEEE Trans. Pattern Anal. Machine Intell. (01) (2021) 1 -1.

[20] X Pan, L Li, H Yang, et al., Accurate segmentation of nuclei in pathological images via sparse reconstruction and deep convolutional networks [J], Neurocomputing 229 (2017) 88–99.

[21] W Zhang, L Dong, X Pan, et al., A survey of restoration and enhancement for underwater images [J], IEEE Access 7 (2019) 182259–182279.

[22] W Zhang, L Dong, T Zhang, et al., Enhancing underwater image via color correction and bi-interval contrast enhancement [J], Signal Processing: Image Communication 90 (2021) 116030.

[23] W Zhang, X Pan, X Xie, et al., Color correction and adaptive contrast enhancement for underwater image enhancement [J], Comput. Electrical Eng. 91 (2021) 106981.

[24] H Lu, M Zhang, X Xu, et al., Deep fuzzy hashing network for efficient image retrieval [J], IEEE Trans. Fuzzy Syst. 29 (1) (2020) 166–176.

[25] R Lan, L Sun, Z Liu, et al., MADNet: a fast and lightweight network for single-image super resolution [J], IEEE Trans. Cybernetics 51 (3) (2020) 1443–1453.

[26] R Lan, L Sun, Z Liu, et al., Cascading and enhanced residual networks for accurate single-image super-resolution [J], IEEE Trans. Cybernetics 51 (1) (2020) 115–125.