*Article*

# Tomato Leaf Disease Diagnosis Based on Improved Convolution Neural Network by Attention Module

**Shengyi Zhao, Yun Peng, Jizhan Liu * and Shuo Wu**

Key Laboratory of Modern Agricultural Equipment and Technology, Ministry of Education, Jiangsu University, Zhenjiang 212013, China; 2111916017@stmail.ujs.edu.cn (S.Z.); 2111716004@stmail.ujs.edu.cn (Y.P.); 1000004697@ujs.edu.cn (S.W.)
**\*** Correspondence: 1000002048@ujs.edu.cn; Tel.: +86-511-88797338

**Abstract:** Crop disease diagnosis is of great significance to crop yield and agricultural production. Deep learning methods have become the main research direction to solve the diagnosis of crop diseases. This paper proposed a deep convolutional neural network that integrates an attention mechanism, which can better adapt to the diagnosis of a variety of tomato leaf diseases. The network structure mainly includes residual blocks and attention extraction modules. The model can accurately extract complex features of various diseases. Extensive comparative experiment results show that the proposed model achieves the average identification accuracy of 96.81% on the tomato leaf diseases dataset. It proves that the model has significant advantages in terms of network complexity and real-time performance compared with other models. Moreover, through the model comparison experiment on the grape leaf diseases public dataset, the proposed model also achieves better results, and the average identification accuracy of 99.24%. It is certified that add the attention module can more accurately extract the complex features of a variety of diseases and has fewer parameters. The proposed model provides a high-performance solution for crop diagnosis under the real agricultural environment.

**Keywords:** tomato leaf disease; deep learning; convolutional neural network (cnn); attention mechanism; classification

## 1. Introduction

Tomato is an important vegetable crop in the world, with a per capita consumption about 20 kg per year, accounting for around 15% of the total vegetable consumption [1]. The global annual output of fresh tomato exceeds 170 million tons, ranking first in vegetable crop production [2]. The United States, India, Turkey, Egypt and China are the main producers of tomatoes [3]. According to the survey data of the Food and Agriculture Organization of the United Nations, tomato disease is the main reason for the decrease in global tomato production, with an annual loss rate of as high as 8%–10% [4]. However, most tomato diseases start from the leaves and then spread to the entire plant [5]. Automatic identification of tomato leaf diseases accurately can help to improve the management of tomato production and provides a good growth environment.

Traditional expert diagnosis on tomato leaf disease has a high cost and has a subjective misjudgment risk. With the rapid development of computer technology, computer vision, machine learning, and deep learning are widely used in crop disease detection [6,7]. Traditional machine vision methods segment the RGB images of crop diseases by color, texture, or shape features. However, the characteristics of different diseases are similar, so it is difficult to judge the types of diseases, and the accuracy of disease recognition is poor in a complex natural environment. The Convolutional Neural Network (CNN) is a high-performance deep learning network; it abandons complex image preprocessing and feature extraction operations, and adopts an end-to-end structure, which greatly simplifies the recognition process compared to its learning [8–10]. Nowadays, CNN is widely used in

crop disease recognition for real agriculture environments [11–14]; the automatic detection of tomato leaf diseases combined with CNN is conducive to improving the accuracy of diagnosis and reducing labor costs.

Several studies have been carried out to use deep learning technology to improve the survival rate of vegetables, fruits, and field crops through early disease detection and subsequent disease management. Wang et al. [15] apply transfer learning to the original Alex Net network, and the average recognition rate of 10 categories of tomato leaves is better. Rangarajan et al. [16] use the original AlexNet, VGG16 network structure, combined with migration learning to obtain an accuracy of about 97% on the seven segmented tomato diseased leaves. The effects of weight, deviation, and learning rate on the accuracy and speed of disease detection are analyzed. Alcaro et al. [17] use cameras with different resolutions to capture images of 9 tomato diseases and insect pests, and use Faster R-CNN, R-FCN, and SSD for training. Long et al. [18] trained AlexNet and GoogleNet networks by using transfer learning technology for Camellia oleifera diseases identification. Kaur et al. [12] use a pre-trained ResNet network to classify 7 tomato diseases with an accuracy rate of 98.8%. Karthik et al. [19] proposed a deep detection model structure for tomato leaf diseases, optimized and improved the residual network, and used transfer learning to obtain important disease classification features. Although transfer learning can achieve better recognition results, the original AlexNet and VGG16 networks have complex structures and numerous parameters, which cannot meet the actual application and deployment of the model.

Based on the previous investigations, CNNs have more potential than the traditional feature extraction methods. Razavi et al. [20] used the improved CNN network to train the disease detection model for the open source disease dataset, and compared them with traditional classifiers such as SVM, LBP, and GIST, which proved that the model is higher than other classifications in terms of classification accuracy. Yang et al. [21] used the saliency analysis of the image to locate the pests in tea gardens, reduced the number of network layers and convolution kernels for AlexNet, and combined with the Dropout model optimization algorithm to improve the accuracy. The optimized model is effective against 23 pests in tea gardens. The average recognition accuracy rate reaches 88.1%. Sun et al. [22] improved the Alex Net network model by reducing the size of the convolution kernel, which improved the disease accuracy and reduced the parameters required for the model. Liu et al. [7] improved a CNN model based on AlexNet to identify four apple leaf diseases, and the model could achieve an average recognition accuracy of 97.62%. Grinblat et al. [23] developed a powerful neural network for the successful identification of three different legume species based on the morphological patterns of the leaves' veins. In the real agriculture environment, the above mentioned study provides a lot of reference for the diagnosis of tomato leaf diseases.

In recent years, due to the characteristic of extracting discriminative features of the area of interest, the attention network began to be widely used in machine translation, generative adversarial, and so on [24,25]. However, it is still in the exploratory stage in the field of agriculture diseases detection. Tang et al. [26] added the attention module into ShuffleNet, which improved the recognition rate of grape diseases in the PlantVillage dataset to 99.14%. Zhong et al. [27] based on the ResNet18 added a group attention module, and the pixel accuracy of semantic segmentation of cucumber diseased leaves in the natural environment reached 93.9%.

For tomato leaf disease diagnosis, the diseased area only occupies a part of the leaf image size. So, this study adds an attention module to the original CNN network model to automatically extract important disease feature information from a complex environment. The feature extraction is focused on the disease feature channel, and the invalid feature channel information is eliminated. In this paper, an improved CNN network model is proposed to diagnosis tomato multiple leaf disease accurately.

The main contributions of this paper are summarized as follows:

- In order to meet the diagnosis requirements of various tomato leaf diseases in the natural environment, this paper constructs a dataset of 9 tomato leaf diseases and healthy leaves. Furthermore, through data enhancement methods, the generalization ability and adaptability of the model in practical applications are improved.
- This paper proposes a multi-scale CNN network structure for the diagnosis of tomato leaf diseases. Based on the residual block, a multi-scale feature extraction module is added. The SE module is deeply integrated into the ResNet-50 network model.
- This paper established a multi-dimensional dependency relationship between the three dimensions (C, H, W) of the extracted tomato leaf disease feature map and used channel and spatial information with a small amount of calculation. In this way, effective features of lesions can be obtained in a complex background, and contextual information can be discriminated.

The rest of the paper is organized in the following manner. Section 2 introduces tomato leaf disease dataset augmentation and tomato leaf disease diagnosis model improvement. Section 3 conducted comparative experiments on the proposed model performance and verified the applicability of the model on other crop disease datasets. Section 4 compares the model results of this study with other researchers in detail. Finally, the conclusion is provided in Section 5.

## 2. Materials and Methods

### 2.1. Build the Dataset

The image data of tomato leaf health and disease in this paper comes from the PlantVillage open source database [22]. The database contains a large number of plant disease images and is the world's largest crop database. After initially acquiring the image of tomato leaves, the image data needed for the research is manually screened to avoid problems such as image duplication and classification errors in the dataset. Finally, a dataset contained 4585 tomato leaf images is obtained, and the size of each picture is fixed at $224 \times 224$. The dataset contains a total of 10 tomato leaf categories, such as bacterial spot, early blight, healthy, late blight, leaf mold, mosaic virus, septoria leaf spot, target spot, two-spotted spider mite, and yellow leaf curl virus. The images of tomato leaves in 10 categories are shown in Figure 1.
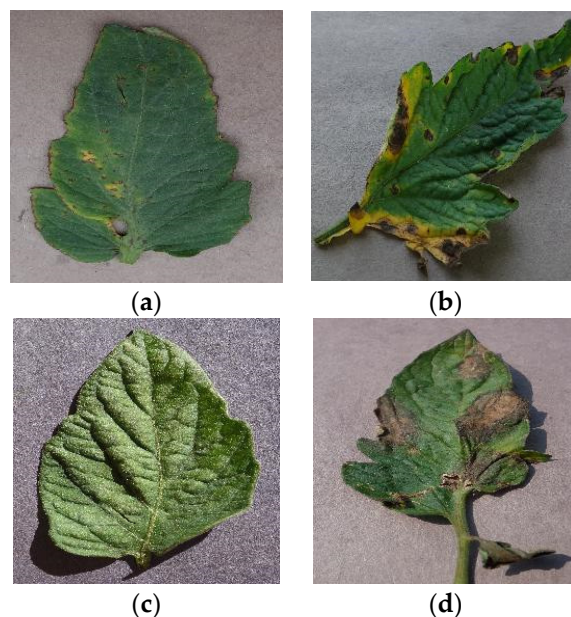


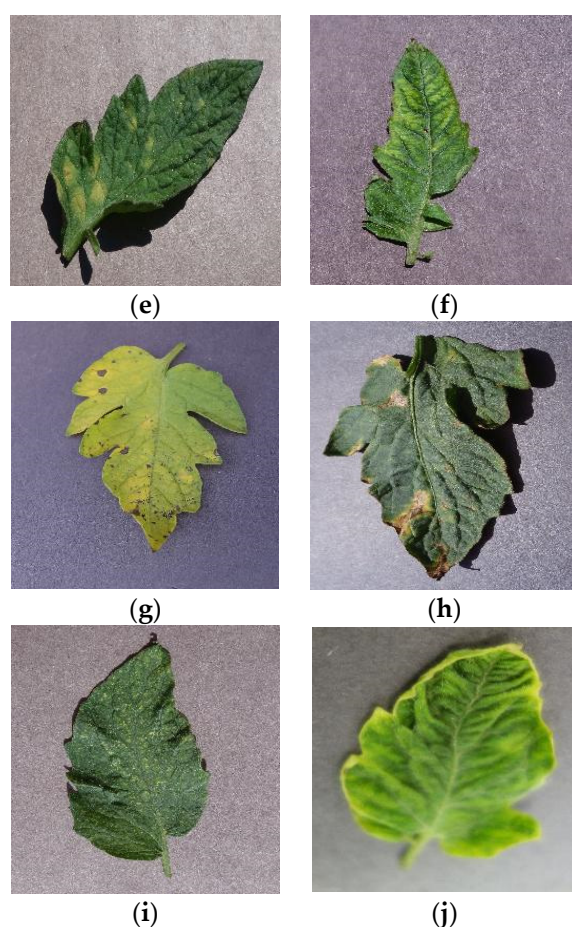| (a) | (b) |



| (c) | (d) |

**Figure 1.** *Cont.*

**Figure 1.** Tomato leaf dataset: (**a**) bacterial spot; (**b**) early blight; (**c**) healthy; (**d**) late blight; (**e**) leaf mold; (**f**) mosaic virus; (**g**) septoria leaf spot; (**h**) target spot; (**i**) two-spotted spider mite; (**j**) yellow leaf curl virus.

*2.2. Data Augmentation*

In deep learning, the diversity of the data set can enhance the generalization ability and robustness of the model [28]. Therefore, this paper used a variety of image enhancement techniques, and enhanced image data in combination with OpenCV under the Pytorch framework.

1.  Spin: Rotated the picture randomly by 0°, 90°, 180°, and 270° will not change the relative position of the diseased spot and the healthy part, simulated the randomness of the shooting angle under natural conditions.
2.  Zoom: Reduced an image according to a certain ratio helps to identify targets at multiple scales. For the zoomed image, the resolution of the image is expanded to 224 × 224 pixels by filling 0 pixels.
3.  Add noise: Added salt and pepper noise or gaussian noise to the image to simulate images with different definitions taken in the natural environment.
4.  Color jitter: Changed the brightness, saturation, and contrast of the image to simulate the image difference caused by the light intensity when shooting in the natural environment.

Using the above data enhancement method, the number of samples in each category was expanded by 4 times, and the enhanced tomato leaf disease data set contained 22,925 images. The trainset and validationset are randomly divided into 8:2. The detailed information of the dataset is shown in Table 1.

**Table 1.** Tomato leaf diseases dataset.

| Type | Class | Origin Images | Augmentation Images | Trainset |
|---|---|---|---|---|
| bacterial spot | 0 | 425 | 2125 | 1700 |
| early blight | 1 | 480 | 2400 | 1920 |
| healthy | 2 | 481 | 2405 | 1924 |
| late blight | 3 | 463 | 2315 | 1852 |
| leaf mold | 4 | 470 | 2350 | 1880 |
| mosaic virus | 5 | 448 | 2240 | 1792 |
| septoria leaf spot | 6 | 436 | 2180 | 1744 |
| target spot | 7 | 457 | 2285 | 1828 |
| two-spotted spider mite | 8 | 435 | 2175 | 1740 |
| yellow leaf curl virus | 9 | 490 | 2450 | 1960 |
| Total | | 4585 | 22,925 | 18,340 |

### 2.3. Deep Learning Models

#### 2.3.1. The Feature Extraction Network

Feature extraction is the key to deep learning, the different feature extraction networks have different parameters, speeds, and performance. Nowadays, some wide range of convolutional neural network models have been proposed, such as AlexNet [29], VGGNet [30], and GoogleNet [31]. However, these CNN models reduced the speed of training and detection due to the large numbers of parameters and computational operations [26]. He et al. [32] proposed a residual network with 101 layers to effectively solve the problem of gradient degradation, and won the 2015 ImageNet Large-scale Visual Recognition Challenge. Compared with AlexNet, VGGNet, and GoogLeNet, ResNet has less computation and higher performance. In this paper, ResNet-50 with less calculation and better performance is used as the feature extraction network.

In Figure 2, the tomato leaf disease image is input into the ResNet-50 network structure, first passes through convolutional layer, BN layer, and activation layer, and then the obtained feature map is maximized pooling. The ResNet50 model mainly includes Stage 1–4, and each stage consists of 1 sampling module and multiple identity mapping modules. The output feature map undergoes an AVG pooling operation, and then passes through the Flatten layer to make the output multi-dimensional features one-dimensional, and finally output through the fully connected layer. When deepening the number of network layers, if the internal characteristics of the network have reached the optimal level in a certain layer, the subsequent superimposed network layers will not change the characteristics. The residual module (Figure 3) in ResNet-50 can effectively solve the problem of identity mapping [32] and can also reduce network parameters and calculations.
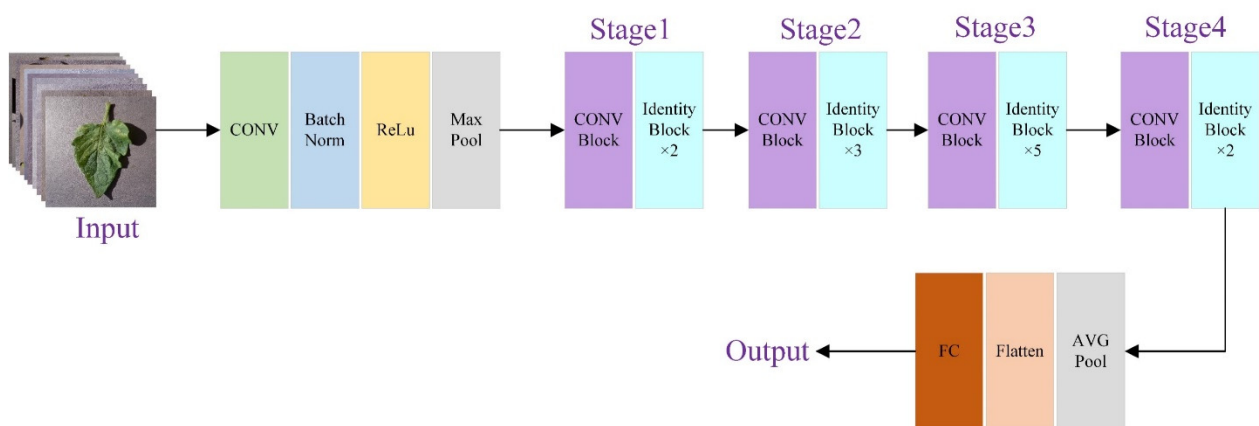


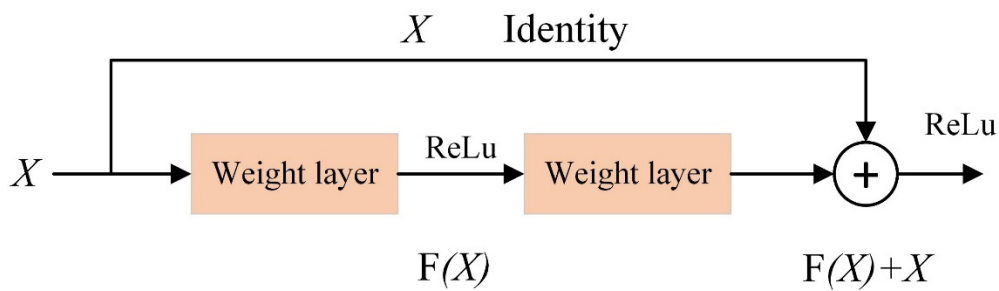**Figure 2.** The structure of the ResNet-50.

**Figure 3.** The residual block.

### 2.3.2. Attention Module

Multiple small disease lesions can occur on the tomato leaf, which are usually of various shapes. Using channel dependency is an important way to improve CNN model performance. To boost the performance of existing state-of-the-art models with slight computation cost. As shown in Figure 4, Hu et al. [33] mentioned the Squeeze-and-Excitation Networks in the CVPR 2017 ImageNet Workshop speech. The weights of different channels are trained through the cost function, and the weight coefficients of each feature channel are automatically obtained. Then, according to the size of the weight coefficient of each feature channel, the effective feature channel is enhanced, and the invalid feature channel is suppressed.
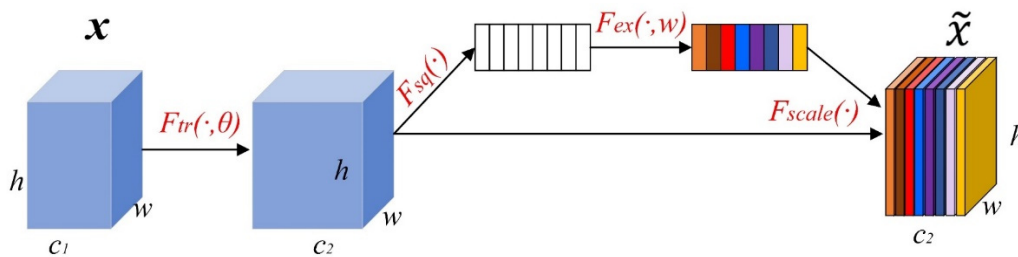


**Figure 4.** The Squeeze-and-Excitation Networks.

### 2.4. Tomato Diagnosis Model of ResNet Fused of the SE Module

Due to the flexibility of the SE Module, it can be directly applied to existing network architectures. In this paper, SENet is added to the original model structure of ResNet-50 to obtain the SE-ResNet50 mode. The network architectures of improved ResNet-50 are depicted in Figure 5. In the SE-ResNet50 network structure, SENet-block uses global average pooling to compress feature maps. Connect the two fully connected layers together to form a modular structure to express the correlation and dependence between the characteristic channels and keep the number of characteristic channels unchanged at the input and output of the two fully connected layer modules.
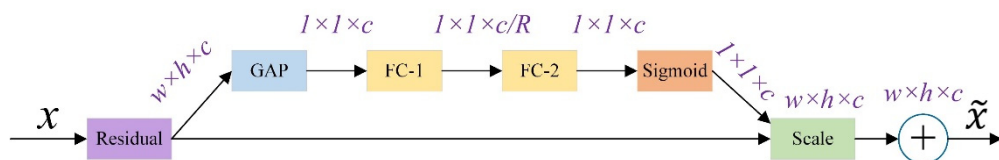


**Figure 5.** The SE-ResNet50 module.

The upper parts of the frameworks shown in Figure 5 are the SE module. When the input feature map reaches the first fully connected layer, the feature dimension is reduced to 1/R of the input. Then it is processed by the ReLu activation function, and then input to the second fully connected layer. At this time, the zoom ratio is R, that is, the number of feature channels is changed back to the input size ($1 \times 1 \times c$).

Global average pooling is used in the compression stage, and the $H \times W$ spatial dimensions of the whole image are shrunk to $F \in R^c$. The compression process follows Equation (1):

$$F_{sq}(u_c) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} u_c(i,j) \tag{1}$$

After compression, a feature map of $1 \times 1 \times c_2$ is obtained, a parameter $W$ is introduced, and a weight is generated for each feature channel through $W$. The different parameters represent the different importance of the characteristic channel, which is the core of the entire SENet module. These weights are allocated to the input feature maps. This process is called feature recalibration, namely and gating mechanism. The excitation process follows Equation (2):

$$F_{ex}(z, W) = \sigma(g(z, W)) = \sigma(W_2 \delta(W_1 z)) \tag{2}$$

where $z$ is the result of the compression process; $\sigma$ is the ReLU function, $W_1 \in R^{\frac{C}{r} * C}$ and $W_2 \in R^{C * \frac{C}{r}}$: a dimensionality-reduction layer with parameters $W_1$ with reduction ratio $r$, a ReLU and then a dimensionality-increasing layer with parameters $W_2$. Reweight is a re-calibration process, which uses the output weight of excitation as the importance of each feature channel after feature selection. According to the degree of importance, the channel is added to the original feature through the Equation (3) and keeping the number of feature channels unchanged and not introducing new feature dimensions.

$$x = F_{scale}(u_c, s_c) = u_c \times s_c \tag{3}$$

where $\overline{x}$ and $F_{scale}$ refers to the channel wise multiplication between the scalar $s_c$ and the feature map $u_c \in R^{H \times W}$.

The above is the complete structure and operation process of SENet module. This sub-network structure is embedded in ResNet50. The combination of characteristic channel recalibration strategy and residual network can effectively improve the network performance, and thus does not need to increase the computational cost much. Through feature refinement, the learning ability of complex disease features is enhanced. The entire tomato leaf disease diagnosis network structure is shown in Figure 6.
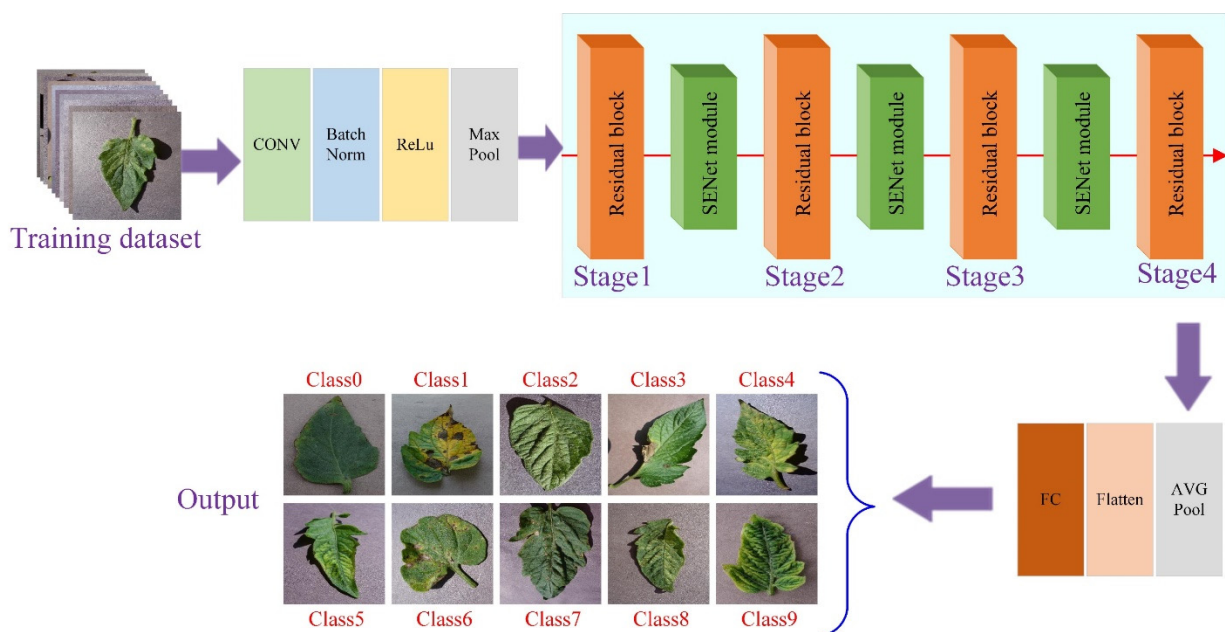


**Figure 6.** The structure of tomato leaf disease diagnosis.

### 2.5. Experiment Setup

The operating platform for this experiment is a Dell T7920 graphics workstation, the operating environment is Windows 10, the CPU is two Intel Xeon Gold 6248R, and the GPU is two NVIDIA Quadro RTX 5000, 64GRAM, 1T solid state drive. The training environment is created by Anaconda3, and the environment configuration is Python 3.6.13 and Pytorch 1.4.0, torchvision 0.5.0 artificial neural network library. At the same time CUDA 10.1 deep neural network acceleration library is used.

The weight value of the feature extraction network uses the parameters of the pretrained ImageNet classification model. This method can greatly reduce the model calculation cost and calculation time. After each training, the validationset is tested and the model is saved, and the model with the highest accuracy is selected.

### 2.6. The Evaluation Index

In order to evaluate the performance, the proposed network is compared with several famous CNN networks: VGG-19, Xception, ResNet-101, and GoogleNet. The average accuracy evaluation index recognized in the field of image classification is used to evaluate the classification results, including *Precision* (*PPV*), *Recall* (*TPR*), *F1 Score* (*F1*), and *Detection speed* ($T_A$).

$$PPV = \frac{T_P}{T_P + F_P} \tag{4}$$

$$TPR = \frac{T_P}{T_P + F_N} \tag{5}$$

$$F_1 = 2 \times \frac{PPV \times TPR}{PPV + TPR} \tag{6}$$

where $T_P$ (true positive) is the number of positive samples predicted as positive samples, $F_P$ (false positive) is the number of negative samples considered to be positive samples, and $F_N$ (false negative) is the number of negative samples considered to be negative samples.

$$T_A = \frac{T}{N} \tag{7}$$

where $T$ is the total detection time for validationset and $N$ is the total number for the validationset.

## 3. Results

### 3.1. Comparison of Various Convolution Neural Networks

The comparison of various CNN model test accuracy curves of different networks is shown in Figure 7. The training iteration epochs are plotted on the X axis and the corresponding training accuracy is plotted on the Y axis.
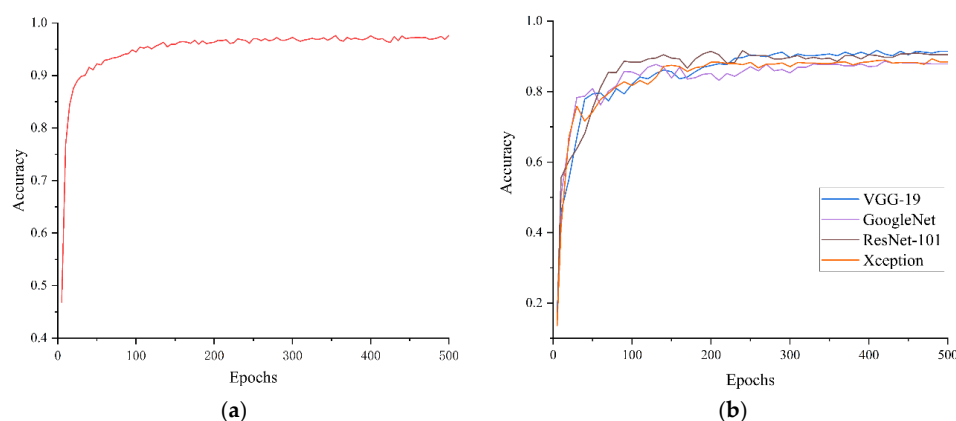


**Figure 7.** The training accuracy curves: (**a**) SE-ResNet50 accuracy curves; (**b**) Comparison model accuracy curves.

The evaluation results of different approaches on the tomato leaf disease are obtained in Table 2. Under the same experimental conditions, the SE-ResNet50 model proposed in this paper has the highest average accuracy, with an accuracy of 96.81%. Compared with GoogleNet, ResNet-101, Xception, and VGG-19 models, the average accuracy is 9.54%, 6.68%, 8.65%, and 6.39% higher, respectively, significantly ahead of the 4 mainstream CNN networks. At the same time, it can be seen from Figure 7 that the SE-ResNet50 model starts to converge after 150 iteration epochs, the convergence rate is the fastest among all models. Also, the model tends to be stable after convergence, and the fluctuation range is smaller. Moreover, the SE-ResNet50 model proposed in this paper has the fastest average diagnosis time for a single disease image, which is only 31.68 ms. Compared with the second-ranked Xception model, the time is reduced by 1.23 ms, which meets the needs of real-time diagnosis of tomato leaf diseases. Synthesizing the above analysis, the proposed model achieves the best performance in terms of accuracy and convergence speed.

**Table 2.** The evaluation results.

| Model | Input | PPV | TPR | F1 | $T_A$ (ms) | Accuracy (%) |
|---|---|---|---|---|---|---|
| GoogleNet | 224 | 0.8716 | 0.8709 | 0.8712 | 33.56 | 87.27 |
| ResNet-101 | 224 | 0.8995 | 0.9013 | 0.9004 | 39.53 | 90.13 |
| Xception | 224 | 0.8825 | 0.8814 | 0.8819 | 32.91 | 88.16 |
| VGG-19 | 224 | 0.9039 | 0.9047 | 0.9243 | 40.26 | 90.42 |
| SE-ResNet50 | 224 | 0.9677 | 0.9681 | 0.9679 | 31.68 | 96.81 |

Figure 8 shows the confusion matrixes for 9 tomato leaf diseases and health leaf using our SE-ResNet50 model.
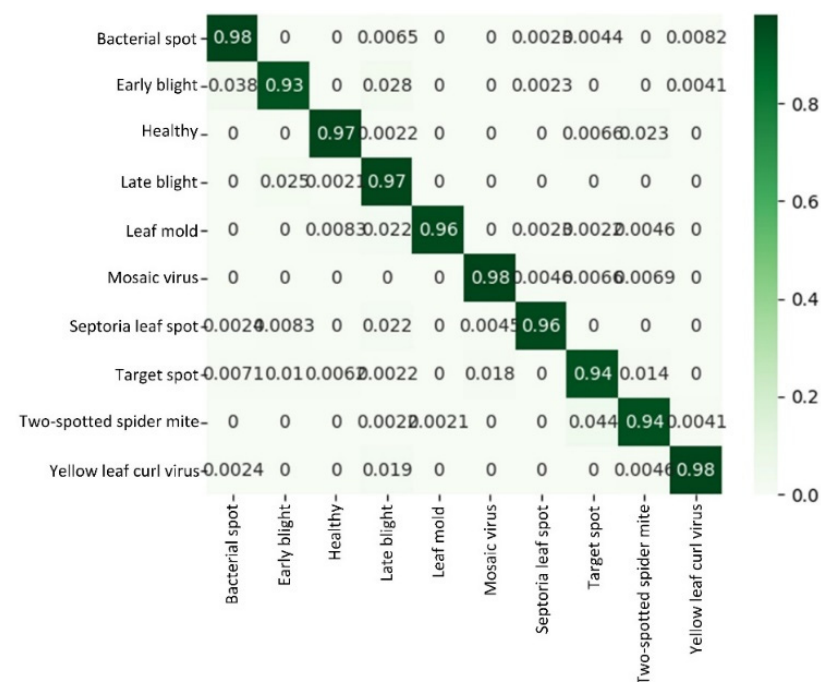


**Figure 8.** Confusion matrixes for tomato leaf diseases.

The SE-ResNet50 model proposed in this paper has an accuracy of over 97% for the diagnosis of healthy tomato leaves, and over 98% for the three diseases of bacterial spot, mosaic virus, and yellow leaf curl virus. The diagnosis accuracy of early blight, target spot and two-spotted spider mite is low, but they have reached 93%, 94%, and 94% respectively, which meets the accuracy requirements in actual diagnosis operations.

To better understand the learning capacity of the channel-wise mechanism, the visualizations with several tomato leaf disease feature maps of proposed SE-ResNet50 are

shown in Figure 9. The proposed model can retain more image details due to important feature reuse.
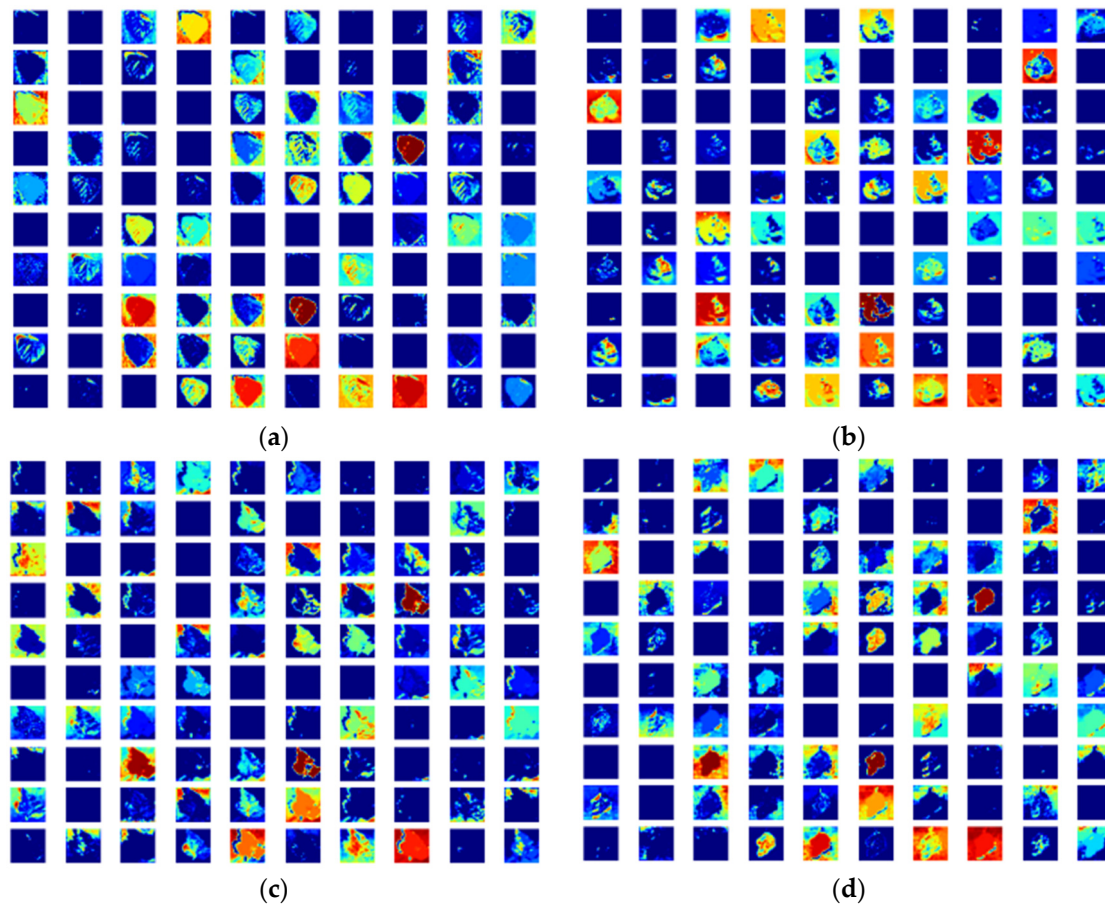


**Figure 9.** Visualization of tomato leaf network feature map: (**a**) healthy; (**b**) early blight; (**c**) target spot; (**d**) two-spotted spider mite.

### 3.2. Comparison of Diagnosis Performance with Attention Module

In order to prove the effect of adding attention mechanism on model accuracy, keeping the experimental conditions and parameters consistent, a comparison experiment of the performance of the SE-ResNet50 and ResNet50 models was carried out. The results of the comparative experiment of the proposed model and the ResNet50 model without attention module on tomato leaf disease are represented in Table 3.

**Table 3.** The results of SE-ResNet50 compared with ResNet50 without attention module.

| Model | PPV | TPR | F1 | $T_A$ (ms) | Accuracy (%) |
|---|---|---|---|---|---|
| ResNet-50 | 0.9251 | 0.9256 | 0.9253 | 33.85 | 92.56 |
| SE-ResNet50 | 0.9677 | 0.9681 | 0.9679 | 31.68 | 96.81 |

It can be seen from Table 3 that the results of the model are improved after adding the attention module. Adding the attention mechanism, the accuracy of the model is increased by 4.25%, the average detection time of a single disease is shortened by 2.17 ms, and the model parameters are only a little bit. It can be concluded that the proposed network is effective.

Based on the above results, it can be seen that the SE-ResNet50 model proposed in this paper can well complete the task of tomato leaf disease diagnosis and has high robustness and accuracy. This model can be a very useful detection tool in the field of crop diseases.

### 3.3. The SE-ResNet50 Effectiveness on Other Corp Disease Dataset

In order to verify the practical application performance of the SE-ResNet50 model proposed in this paper, we conducted experiments on the public dataset of grape leaf diseases. The public dataset contains 2750 grape leaf disease images, including black measles, black rot, brown spots, healthy, and leaf blight. Sample images are shown in Figure 10.
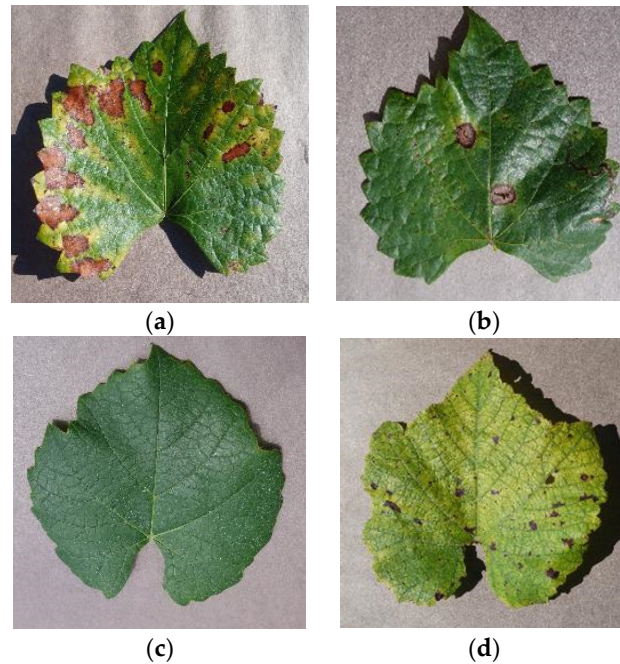


**Figure 10.** Grape leaf dataset: (**a**) black measles; (**b**) brown spots; (**c**) healthy; (**d**) leaf blight.

Under the same experimental conditions, this paper selected Googlenet, Resnet-50, and Xception for comparative experiments on grape leaf diseases. As shown in Figure 11, the convergence times of the four models are similar, but the final convergence accuracy of the SE-ResNet50 model is higher than that of the GoogleNet, ResNet-50, and Xception models. At the same time, the SE-ResNet50 model proposed in this paper has a small convergence accuracy fluctuation range. The convergence accuracy of GoogleNet, ResNet-50, and Xception models has a large fluctuation range.
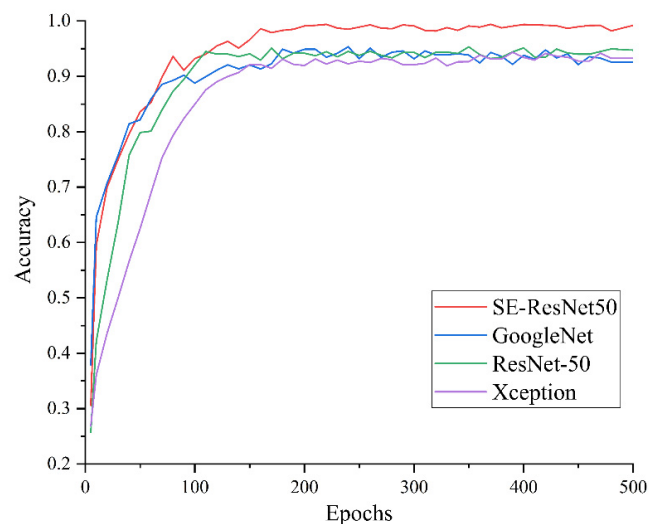


**Figure 11.** The training accuracy curves for grape leaf diseases.

The evaluation results of approaches on the grape leaf disease are obtained in Table 4. The SE-ResNet50 model proposed in this paper has an average diagnostic accuracy of 99.24% for the four classifications of grape leaves. Compared with tomato leaf diseases, the accuracy of diagnosis is increased by 2.43%, mainly due to the decrease of 6 kinds of diseases. Compared with ResNet-50, GoogleNet and Xception models, the average accuracy is 5.33%, 6.46% and 6.61% higher, respectively. Moreover, the SE-ResNet50 model proposed in this paper has the fastest average diagnosis time for a grape leaf image, which is only 31.42 ms. Compared with the second-ranked Xception model, the time is reduced by 0.55 ms. Synthesizing the above analysis, the proposed model for grape leaf diseases diagnosis also achieves the best performance in terms of accuracy and convergence speed.

**Table 4.** The results of SE-ResNet50 compared with ResNet50 without the attention module.

| Grape | Input | PPV | TPR | F1 | $T_A$ (ms) | Accuracy (%) |
|---|---|---|---|---|---|---|
| SE-ResNet50 | 256 | 0.9918 | 0.9924 | 0.9921 | 31.42 | 99.24 |
| ResNet-50 | 256 | 0.9388 | 0.9392 | 0.9390 | 33.27 | 93.91 |
| GoogleNet | 256 | 0.9274 | 0.9269 | 0.9271 | 32.89 | 92.78 |
| Xception | 256 | 0.9261 | 0.9263 | 0.9262 | 31.97 | 92.63 |

The identification result is represented by confusion matrix in Figure 12; the diagnostic accuracy of black measles exceeds 98%, the diagnostic accuracy of brown spots and leaf blight exceeds 99%, and the diagnostic accuracy of healthy leaves is 100%. It is proven that the proposed method has a wide range of applicability and has better performance relative to deep based methods on other crop public datasets.
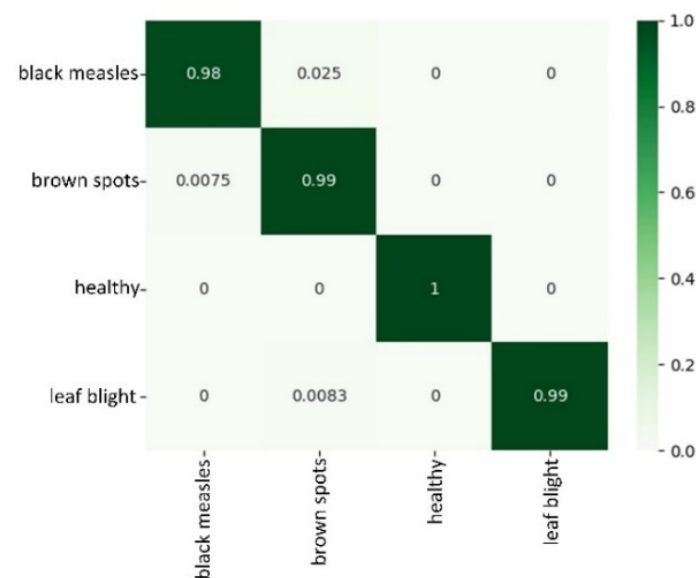


**Figure 12.** Confusion matrixes for grape leaf diseases.

## 4. Discussion

Crop diseases are a major threat to global vegetable supply security, and the latest technologies need to be applied to the agriculture field to control diseases. Due to the long-term continuous operations, ease of data acquiring, good robustness, and quick computing of deep-learning-based disease detection, it is widely investigated.

According to the complex characteristics of tomato leaf diseases, this study designed a diagnosis model for multi-scale extraction of disease features. In this study, the dataset was divided into 10-class classification (bacterial spot, early blight, healthy, late blight, leaf mold, mosaic virus, septoria leaf spot, target spot, two-spotted spider mite, and yellow leaf curl virus). The SE-ResNet50 model proposed in this paper reaches an average detection

accuracy of 96.81%, which is 4.25% higher than the original ResNet50 network accuracy. The diagnostic accuracy of this model for the four diseases exceeds 97%, and the detection accuracy for early blight is the worst, but it also exceeds 93%. The average diagnosis time of a single disease image is only 31.68 ms, and the diagnosis speed is faster to meet the needs of real-time operations.

The results in this study are compared with study results as summarized in Table 5. As shown in Table 5, Durmuş et al. [13], Wang et al. [15], Agarwal et al. [30], and Tm et al. [34] used the same dataset as this study. The accuracy of all these studies is lower than the model proposed in this paper. Even the accuracy of model designed by Guo et al. [29] for eight-class tomato leaves is 4.11% lower than our study. It is found that the proposed model accuracy of Kaur et al.[12], Rangarajan et al. [16], Karthik et al. [19], and Kaushik et al. [35] are higher than our results, which are attributed to the less classification classes of diseases (at most 7 classes). Overall, our model has good general performance and high diagnostic accuracy for tomato leaf diseases.

**Table 5.** Results in the paper compared with other state-of-the-art results.

| Paper | Dataset | Model | Classification | Precision | Recall | F1-Score | Accuracy |
|---|---|---|---|---|---|---|---|
| Durmuş et al. [13] | Plant Village | AlexNet | 10-class | — | — | — | 95.65% |
| Wang et al. [15] | Plant Village | AlexNet | 10-class | — | — | — | 95.62% |
| Agarwal et al. [30] | Plant Village | Vgg-16 | 10-class | 90% | 92% | 91% | 91.2% |
| Tm et al. [34] | Plant Village | LeNet | 10-class | 94.81% | 94.78% | 94.8% | 94% |
| Our model | Plant Village | ResNet-50+SeNet | 10-class | 96.77% | 96.81% | 96.79% | 96.81% |
| Kaur et al.[12] | Plant Village | ResNet-101 | 7-class | 98.8% | 98.8% | 98.8% | 98.8% |
| Rangarajan et al. [16] | Plant Village | AlexNet | 7-class | — | — | — | 97.49 |
| Karthik et al. [19] | Plant Village | ResNet + Dense | 4-class | — | — | — | 98% |
| Guo et al. [29] | Plant Village | AlexNet | 8-class | — | — | — | 92.7% |
| Kaushik et al. [35] | Plant Village | ResNet-50 | 6-class | — | — | — | 97.01% |

## 5. Conclusions

In this work, we successfully developed a multi-scale feature extraction model for tomato leaf diseases diagnosis. The model deeply integrates the residual block and the attention module and is trained for the detection of healthy and different unhealthy tomato leaf images. The obtained results show that our model outperforms some recent deep learning studies by using the most popular publicly available PlantVillage dataset.It was also found that the SE-ResNet50 was best suited for the diagnosis of tomato leaf diseases compared to other model. Besides, the performance of the SE-ResNet50 model generally further improved when trained with more different environment images. The trained models can be used in the early automatic diagnosis of tomato and other crop diseases. Thus, this work can be beneficial in early and automatic disease diagnosis of tomato crops enabled by the latest technologies such as smartphones, drone cameras, and robotic platforms.

In the next step, we will deploy the proposes model to the greenhouse inspection robot independently developed by the team to realize the automatic identification of tomato leaf diseases in the real agricultural environment. At the same time, we will also establish a dataset of tomato leaf diseases in a real agricultural environment to improve the diagnostic performance of the inspection robot. It can help farmers accurately identify diseases, carry out corresponding agricultural tasks such as applying pesticides and fertilizing according to the types of diseases, and successfully realize agricultural modernization and intelligence.

**Author Contributions:** Conceptualization, S.Z. and J.L.; methodology, S.Z. and J.L. and Y.P.; software, S.Z.; validation, S.Z. and J.L.; formal analysis, S.Z., J.L. and Y.P.; investigation, S.Z. and S.W.; resources, J.L.; data curation, S.Z. and S.W.; writing—original draft preparation, S.Z.; writing—review and editing, S.Z., J.L. and Y.P. All authors have read and agreed to the published version of the manuscript.

## References

1. Chowdhury, M.E.; Rahman, T.; Khandakar, A.; Ayari, M.A.; Khan, A.U.; Khan, M.S.; Al-Emadi, N.; Reaz, M.B.I.; Islam, M.T.; Ali, S.H.M. Automatic and Reliable Leaf Disease Detection Using Deep Learning Techniques. *AgriEngineering* **2021**, *3*, 294–312. [CrossRef]
2. Valenzuela, M.E.M.; Restović, F. Valorization of Tomato Waste for Energy Production. In *Tomato Chemistry, Industrial Processing and Product Development*; Royal Society of Chemistry: London, UK, 2019; pp. 245–258.
3. Elnaggar, S.; Mohamed, A.M.; Bakeer, A.; Osman, T.A. Current status of bacterial wilt (*Ralstonia solanacearum*) disease in major tomato (*Solanum lycopersicum* L.) growing areas in Egypt. *Arch. Agric. Environ. Sci.* **2018**, *3*, 399–406. [CrossRef]
4. Zhu, X.K. *Research on Tomato Disease Identification Based on Convolutional Neural Network*; Beijing University of Technology: Beijing, China, 2020.
5. Sardogan, M.; Tuncer, A.; Ozen, Y. Plant leaf disease detection and classification based on CNN with LVQ algorithm. In Proceedings of the 2018 3rd International Conference on Computer Science and Engineering (UBMK), Sarajevo, Herzegovina, 20–23 September 2018; pp. 382–385.
6. Pooja, V.; Das, R.; Kanchana, V. Identification of plant leaf diseases using image processing techniques. In Proceedings of the 2017 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR), Chennai, India, 7–8 April 2017; pp. 130–133.
7. Liu, B.; Zhang, Y.; He, D.; Li, Y. Identification of apple leaf diseases based on deep convolutional neural networks. *Symmetry* **2018**, *10*, 11. [CrossRef]
8. Boulent, J.; Foucher, S.; Théau, J.; St-Charles, P.-L. Convolutional neural networks for the automatic identification of plant diseases. *Front. Plant Sci.* **2019**, *10*, 941. [CrossRef] [PubMed]
9. Ma, J.; Zheng, F.; Zhang, L.; Sun, Z. Disease recognition system for greenhouse cucumbers based on deep convolutional neural network. *Trans. Chin. Soc. Agric. Eng.* **2018**, *34*, 186–192.
10. Chao, X.; Sun, G.; Zhao, H.; Li, M.; He, D. Identification of apple tree leaf diseases based on deep learning models. *Symmetry* **2020**, *12*, 1065. [CrossRef]
11. Arya, S.; Singh, R. A Comparative Study of CNN and AlexNet for Detection of Disease in Potato and Mango leaf. In Proceedings of the 2019 International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT), Ghaziabad, India, 27–28 September 2019; pp. 1–6.
12. Kaur, M.; Bhatia, R. Development of an improved tomato leaf disease detection and classification method. In Proceedings of the 2019 IEEE Conference on Information and Communication Technology, Baghdad, Iraq, 15–16 April 2019; pp. 1–5.
13. Durmuş, H.; Güneş, E.O.; Kırcı, M. Disease detection on the leaves of the tomato plants by using deep learning. In Proceedings of the 2017 6th International Conference on Agro-Geoinformatics, Fairfax, VA, USA, 7–10 August 2017; pp. 1–5.
14. Liu, Y.; Feng, Q.; Wang, S. Plant disease identification method based on lightweight CNN and its mobile application. *J. Agric. Eng.* **2019**, *35*, 194–204.
15. Wang, Y.; Zhang, H.; Liu, Q.; Zhang, Y. Image classification of tomato leaf diseases based on transfer learning. *J. China Agric. Univ.* **2019**, *24*, 124–130.
16. Rangarajan, A.K.; Purushothaman, R.; Ramesh, A. Tomato crop disease classification using pre-trained deep learning algorithm. *Procedia Comput. Sci.* **2018**, *133*, 1040–1047. [CrossRef]
17. Fuentes, A.; Yoon, S.; Kim, S.C.; Park, D.S. A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition. *Sensors* **2017**, *17*, 2022. [CrossRef] [PubMed]
18. Long, M.; Ouyang, C.; Liu, H.; Fu, Q. Image recognition of *Camellia oleifera* diseases based on convolutional neural network & transfer learning. *Trans. Chin. Soc. Agricult. Eng.* **2018**, *34*, 194–201.
19. Karthik, R.; Hariharan, M.; Anand, S.; Mathikshara, P.; Johnson, A.; Menaka, R. Attention embedded residual CNN for disease detection in tomato leaves. *Appl. Soft Comput.* **2020**, *86*, 105933.
20. Razavi, S.; Yalcin, H. Using convolutional neural networks for plant classification. In Proceedings of the 2017 25th Signal Processing and Communications Applications Conference (SIU), Antalya, Turkey, 15–18 May 2017; pp. 1–4.
21. Yang, G.; Bao, Y.; Liu, Z. Localization and recognition of pests in tea plantation based on image saliency analysis and convolutional neural network. *Trans. Chin. Soc. Agric. Eng.* **2017**, *33*, 156–162.

22. Sun, J.; Tan, W.; Mao, H.; Wu, X.; Chen, Y.; Wang, L. Recognition of multiple plant leaf diseases based on improved convolutional neural network. *Trans. Chin. Soc. Agric. Eng.* **2017**, *33*, 209–215.
23. Grinblat, G.L.; Uzal, L.C.; Larese, M.G.; Granitto, P.M. Deep learning for plant identification using vein morphological patterns. *Comput. Electron. Agric.* **2016**, *127*, 418–424. [CrossRef]
24. Luong, M.-T.; Pham, H.; Manning, C.D. Effective approaches to attention-based neural machine translation. *arXiv* **2015**, arXiv:1508.04025.
25. Zhang, H.; Goodfellow, I.; Metaxas, D.; Odena, A. Self-attention generative adversarial networks. In Proceedings of the International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019; pp. 7354–7363.
26. Tang, Z.; Yang, J.; Li, Z.; Qi, F. Grape disease image classification based on lightweight convolution neural networks and channelwise attention. *Comput. Electron. Agric.* **2020**, *178*, 105735. [CrossRef]
27. Zhong, C.; Hu, Z.; Yang, X.; Li, H.; Liu, F.; Li, M. Triple Stream Segmentation Network for Plant Disease Segmentation. In Proceedings of the 2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), Chongqing, China, 13–14 March 2021; pp. 496–501.
28. Pulgar, F.J.; Rivera, A.J.; Charte, F.; del Jesus, M.J. On the impact of imbalanced data in convolutional neural networks performance. In Proceedings of the International Conference on Hybrid Artificial Intelligence Systems, La Rioja, Spain, 21–23 June 2017; pp. 220–232.
29. Guo, X.; Fan, T.; Shu, X. Tomato leaf diseases recognition based on improved multi-scale AlexNet. *Trans. Chin. Soc. Agricult. Eng.* **2019**, *35*, 162–169.
30. Agarwal, M.; Gupta, S.K.; Biswas, K. Development of Efficient CNN model for Tomato crop disease identification. *Sustain. Comput. Inform. Syst.* **2020**, *28*, 100407. [CrossRef]
31. Shin, J.; Chang, Y.K.; Heung, B.; Nguyen-Quang, T.; Price, G.W.; Al-Mallahi, A. A deep learning approach for RGB image-based powdery mildew disease detection on strawberry leaves. *Comput. Electron. Agric.* **2021**, *183*, 106042. [CrossRef]
32. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
33. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.
34. Tm, P.; Pranathi, A.; SaiAshritha, K.; Chittaragi, N.B.; Koolagudi, S.G. Tomato leaf disease detection using convolutional neural networks. In Proceedings of the 2018 Eleventh International Conference on Contemporary Computing (IC3), Noida, India, 2–4 August 2018; pp. 1–5.
35. Kaushik, M.; Prakash, P.; Ajay, R.; Veni, S. Tomato Leaf Disease Detection using Convolutional Neural Network with Data Augmentation. In Proceedings of the 2020 5th International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 10–12 June 2020; pp. 1125–1132.