ISTP

RESEARCH ARTICLE

Fruit Tree Disease Recognition Based on Residual Neural Network and Attention Mechanism

Xiedong Song[1,2] and Vladimir Y. Mariano[1]

[1]College of Computing and Information Technologies, National University, Manila, Philippines
[2]School of Computer Science and Engineering, JiNing University, JiNing 273155, China

Abstract: Fruit growing has played a huge role in solving food supply issues in many countries. However, the yield and quality of fruits can be affected by various diseases, and thus timely and accurate identification of disease conditions is particularly important. Currently, using image recognition and object detection technology to diagnose fruit tree diseases has become a research hotspot in forestry informatization. Convolutional neural networks eliminate the preprocessing of manual feature selection and have high recognition performance. However, it is not easy to train due to the risk of gradient disappearance. In order to achieve better recognition effect, this research addresses the problem of applying small-scale data samples through data enhancement and transfer learning, and it optimizes the model by combining the two main attention mechanisms of SE and CBAM with ResNet50. Through experiments, it is found that the CBAM ResNet50 model has the best effect, improving the application performance of the studied model in actual scenarios.

Keywords: attention mechanism; disease recognition of fruit trees; ResNet; transfer learning

I. INTRODUCTION

Fruit has always been an important object to help solving the human food crisis and promote human vitamin intake. Apples are widely consumed by human beings and one of the important cash crops in China and many other countries. However, long-term cultivation has reduced the quality of the growth environment of fruit trees, and the occurrence of diseases in the planting process has become increasingly serious. It will cause a large reduction in apple production, causing incalculable economic losses to fruit farmers. At the same time, the massive use of pesticides has also brought hidden dangers to food safety [1]. Therefore, timely and correct identification of various diseases is crucial for maintaining fruit tree yield. With the continuous progress and application of new technologies related to artificial intelligence, traditional fruit tree protection methods are no longer suitable for modern forestry management. In the process of traditional forestry conservation, we mainly rely on experience to accumulate, and rely on technical personnel to identify and judge with the naked eye on the spot. This way of work has low identification efficiency, large error poor timeliness, and strong subjectivity. With the continuous exploration of forestry informatization construction, the examples of pattern recognition technology applied to fruit planting and production are gradually increasing, and forestry informatization and intelligence become the mainstream trend of development [2]. The application of image recognition technology to modern agricultural production can effectively identify and accurately operate pests and diseases and reduce the overuse of pesticides, this is of important significance for the modern management of forestry. Deep learning can eliminate subjective errors caused by manual operations during feature extraction and achieve automatic extraction. The target

detection is an important practical application of deep learning. With the deepening of research, a variety of object detection algorithms have been generated, which can locate and recognize single or multiple objects in the image, and classify the located objects and are continuously used in computer vision task exercises, such as pedestrian detection and video analysis [3]. Because early apple diseases can mostly be detected through leaf observation, this paper takes the leaves as the main detection object, uses ResNet50 network mode and combines the attention mechanism to make the model calculation more accurate than the traditional neural network model, and making the target detection algorithm more practical in the application of fruit tree disease and pest recognition [4]. The rest of the paper is structured as follows. Section II presents the relevant research work. Section III introduces the research methods used in this paper, including a comprehensive application of ResNet50 and data enhancement, transfer learning, attention mechanisms, loss function, etc. To better compare the results, this article selects CBAM and SE, two of the most popular attention mechanisms currently used, for application comparison. Section IV introduces the dataset used in this article. The Section V introduces the experimental process and analysis of the experimental results, and the Section VI provides a summary.

II. THE RELATED WORK

In recent years, intelligent recognition algorithms of plant disease have been constantly innovated. The existing algorithms are mainly based on traditional computer vision task and deep learning [5].

Most of the traditional methods of computer vision use image processing technology to preprocess the harmful image and then extract the features artificial selection according to personal experience. Common research methods include support vector machine,

---

Corresponding author: Xiedong Song (e-mail: sxdmail@163.com).

random forest, and morphology, such as Jia Qingjie proposed a multi threshold segmentation algorithm for apple diseased leaves, focusing on leaf spot disease, mosaic disease, and brown spot disease. First, the image of apple diseased leaves was filtered. Then, the fuzzy C-means were used to cluster the lesion images. Finally, a multithreshold algorithm was used to segment the apple disease leaf image, resulting in a lesion image with a segmentation accuracy of over 94% [6]. Wang Liwei uses the method of support vector machine to label the texture and shape information of the diseased area, with an accuracy rate of nearly 95% [7]. Hossain et al. conducted a variable statistical test based on the unique feature differences among three different diseases on tea and implemented disease identification using support vector machine methods, reducing computational time in the work and promoting an average acceleration of 300 milliseconds per leaf, with an overall accuracy of 93.3% [8]. Although the overall recognition effect of traditional computer image processing technology is relatively good, its feature extraction needs to be manually recognized by relevant experts before being sent to the selected model for training. This feature extraction process is too cumbersome, and the recognition accuracy is greatly affected by the extraction accuracy [9]. Therefore, the proposed methods are mostly targeted at a specific situation, and their transferability between different crop diseases and pests is not high, causing its weak universal applicability.

Deep learning has become a research and application hotspot for many animal and plant disease recognition scholars [10]. Mideth Abisado and others use a combination of machine language and natural language to monitor the spread of diseases [11]. Dai Zehan built two kinds of recognizers using convolutional neural network and transfer learning. By comparison, they found that the eight types of recognizers based on MobileNet V1 structure performed best, with a recognition accuracy of 93.7% for all images [12]. Sladojevic and others will use data enhancement technology combined with migration CaffeNet parameter model to fine tune the model. The recognition result shows that the correct rate is 96.3% [13]. Brahimi M et al. used the AlexNet model based on transfer learning to realize tomato leaf disease recognition, and the experimental result reached 98.6% [14]. Mehment Metin Ozguven et al. realized the automatic detection of sugar beet leaf disease areas through the improved Fast R-CNN model, and the accuracy rate reached 95.48% [15]. Geetharmani G. et al. built a new plant leaf disease recognition model based on DCNN for plant leaf disease, The model was trained using a publicly available dataset containing 39 different types of plant leaves and their background images. After extensive simulation,

the classification accuracy of the model reached 96.46 [2]. The emergence of deep learning can solve the shortcomings of traditional computer vision technologies such as manually selecting features, and a large number of experiments have shown that it can achieve higher recognition accuracy than traditional computer vision [16]. Hiram Ponce et al. used a new CNN+AHN classifier to classify tomato diseases with an accuracy of 95% [17].

Therefore, this paper mainly adopts deep learning methods to conduct experiments and attempts to combine residual neural networks with attention mechanisms to find more efficient working modes among them.

III. THE RESEARCH METHOD

A. CONVOLUTIONAL NEURAL NETWORK (CNN)

Convolutional neural networks (CNN) are the most commonly used network models in various deep learning tasks. It consists of multilayer neural networks including convolutional layers, pooling layers, and fully connected layers [18]. The core idea of CNN is to quickly extract local feature information from input data through a series of convolution operations.

Convolutional layers are the most core part of neural networks, which use convolutional kernels to obtain important local feature information in input data, making the training process of the model simpler.

The pooling layer can achieve downsampling function, compress sample data, and it effectively reduces the redundant information represented by the feature map. There are usually two operating modes: average pooling and maximum pooling.

The fully connected layer is to convert feature maps into output results. In a fully connected layer, each neuron can be allowed to connect to all neurons in the previous layer, and the output data represent the probability of judging the input data as results of different categories [19].

The structure of a CNN can include multiple convolutional and pooling layers, and ultimately, the feature map is mapped to the output result through a fully connected layer [20]. It adjusts the deviation through a backpropagation algorithm to make the final output data result more accurate. At present, CNN has been proven to be an effective working method for object detection and image classification tasks [21]. The Fig. 1 shows us the workflow of CNN.

Based on the above characteristics, convolutional neural networks have become a common technical application in image classification tasks.
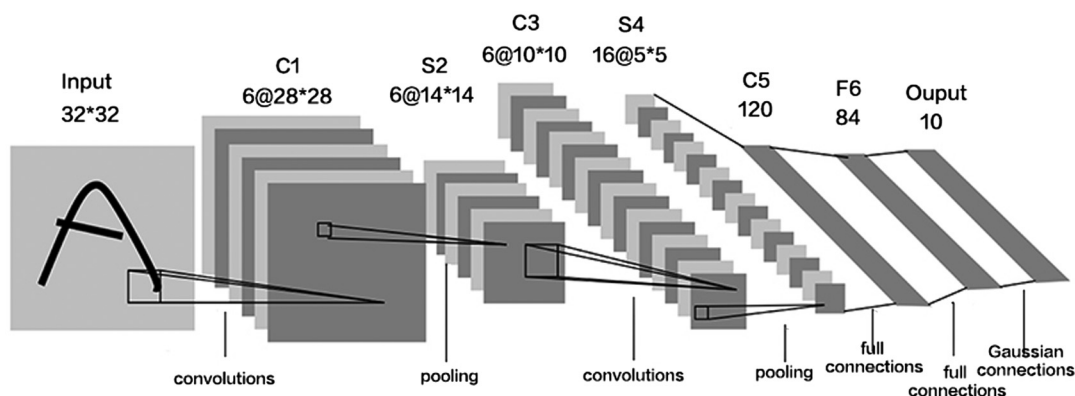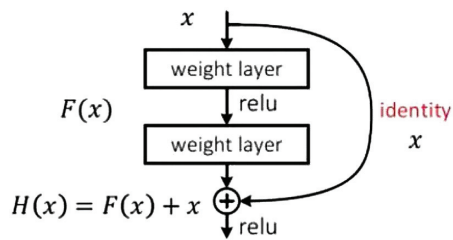

Fig. 1. CNN network workflow.

Fig. 2. ResNet network solution.

B. RESIDUAL NEURAL NETWORK (RESNET)

Residual neural network (ResNet) is an optimized neural network structure and a classic network model in the field of machine learning in recent years.

A common neural network structure consists of multiple layers and allows workers to determine the number of stacked layers themselves. Because each layer will transform the input from the previous layer into a higher level feature representation and output it. Therefore, in traditional neural networks, for the sake of pursue better computational performance, researchers attempt to add as many layers as possible to continuously map inputs to outputs. But as the number of layers in the structure increases, the gradient signal will gradually disappear, making the training process more difficult [22,23]. To address this deficiency, in 2015, He Kaiming et al. proposed an optimization method for ResNet, which mainly uses residual blocks to solve the gradient vanishing problem in deep neural networks.

ResNet utilizes residual blocks to introduce skip connections on the basis of common neural networks. When the calculation results are affected by the vanishing gradient, it allows the input to be directly connected to the output of subsequent layers, thereby preserving the original data information and making the gradient easier to propagate [24] as shown in Fig. 2.

The architecture of the ResNet model can be composed of multiple residual blocks, which are concatenated to form a new working model of the neural network. ResNet can also use algorithms such as random gradient descent (SGD) and cross entropy loss function to help optimize the training process [25,26]. Through multiple rounds of iterative training, the parameters of the model are adjusted to the optimal value, achieving higher recognition accuracy results in target detection or image classification tasks.

C. ATTENTION MECHANISM

(1) THE WORKING PRINCIPLE OF ATTENTION MECHANISM.

The attention mechanism originated from the study of neuroscience, based on the principle that the brain pre judges the importance of all information during work, prioritizing the processing of those important information. Later, attention mechanisms were introduced and applied in the field of artificial intelligence [27]. At present, it has been applied to all aspects of deep learning, including data analysis, natural language processing, image recognition, and reinforcement learning.

The core of attention mechanism is to assign weights to data information at different positions, marking and distinguishing its importance in training tasks through the size of the weights, thereby helping to extract effective features.

In the comprehensive application of various neural networks, attention mechanism adjusts the importance of data at various positions through dynamic weighting and then highlights the characteristics of important parts.

The application of attention mechanism in image classification and recognition tasks mainly involves introducing attention modules into selected network models, which are used to select important regions in the image for classification and help improve the classification accuracy of the model.

Usually, image classification models need to process image data containing a large number of pixels, and attention mechanisms can help the model better focus on important areas. Specifically, the attention mechanism can be achieved through the following steps:

a. Feature extraction: The working model first converts the input image data into a series of high-dimensional feature vectors for feature extraction.

b. Attention calculation: Then, the working model performs attention calculation on the feature vectors to obtain the weight of each feature vector.

c. Feature weighting: Next, the working model applies attention weight to the feature vector to obtain the weighted feature vector.

d. Feature fusion: Finally, the model fuses the weighted feature vectors into a global feature vector and transports it to the classifier for classification calculation.

In this way, the attention mechanism can help the model better focus on important areas in the image, thereby improving classification performance. At the same time, attention mechanism can also help the model learn more explanatory features, thereby improving the interpretability of the model.

Below, we will introduce the two main attention mechanisms used in this experiment.

(2) SE MODEL.
SE attention networks mainly add attention in the channel, and the two most important parts are squeeze link and excitation link.

The SE module can calculate the importance of all channels in the feature map through automatic learning, next,set different weight values to each feature based on this. Those feature of channels that are helpful for the task will be strengthened, while those that are not useful will be suppressed.

As shown in the following Fig. 3, before using the SE attention mechanism (Figure C on the left), the importance of each channel in the feature map was the same. After using SENet (Figure C on the right), different weights are represented by different colors, so that each feature channel has different importance for the measurement of the network model.

The work steps of the SE module are summarized as follows:

a. Input feature map: In SENet, the input feature map can be image data of any size.

b. Enter the squeeze module: This module is responsible for globally averaging and pooling the input feature maps, compressing them into global features for each channel, and then inputting this feature vector into the exit module.

c. Enter the exception module: This module is responsible for learning the global features of each channel and obtaining the weights of each channel. This module consists of two fully connected layers. The first fully connected layer converts the input feature vectors into smaller vectors, and the second fully connected layer converts the smaller vectors into the weights of each channel. The weights learned through this
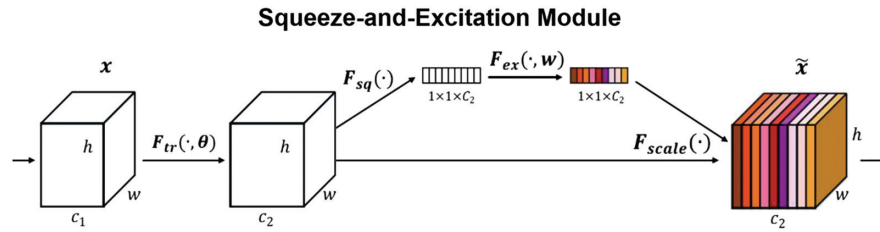
Squeeze-and-Excitation Module

Fig. 3. SE model.

module can adaptively reflect the importance of each channel for image classification.

d. Conduct channel attention weighting: Multiply the weight of each channel by the feature map of that channel to obtain the weighted feature map. Here, the role of weights is to highlight important channel features and suppress unimportant channel features, thereby enhancing the model's attention to important features in the image.

e. Perform feature extraction and classification for the next layer: The weighted feature map is used as the input for the next layer to continue feature extraction and classification. Due to the fact that the weighted feature map can more accurately reflect the features of the image, it can help the network better distinguish different categories and improve the accuracy of the model.

The SE mechanism can be introduced into ResNet for use, and its network structure is shown in Fig. 4.

Overall, SENet enhances the attention of neural networks to important channel features by introducing channel attention mechanisms, improves the accuracy and robustness of the model, and enables the network to better cope with various image classification problems.

(3) CBAM MODEL. Attention mechanism can be divided into two types according to the different goals of attention: channel attention mechanism (CAM) and space attention mechanism

(SAM). The channel focus mechanism pays more attention to the relevant information between channels in each feature map. Spatial attention mechanism is concerned about the spatial location of the key information on the feature map [28].

CAM and SAM have their own advantages and disadvantages. The advantage of channel focus mechanism is that it can strengthen the correlation between each channel and reasonably allocate the proportion of resources occupied by each channel. The disadvantage is that the information of feature positions is easily overlooked and compress it into a feature map of 1×1 channel information. The advantage of spatial attention mechanism is that it can realize fast location and recognition, but the disadvantage is that it overemphasizes the importance of location and ignores the information contained in the channel [29]. It causes all channels to be compressed into a single channel, which makes the information contained in the feature map relatively complex and difficult to identify key information.

Therefore, this paper combines the advantages of two attention mechanisms and adopts a hybrid attention mechanism: CBAM.

The CBAM flowchart is as follows. The input data of feature map first passes through the CAM, multiplies the channel weight and the input feature map, and then enters the SAM. The normalized spatial weight is multiplied by the input feature map of the SAM to obtain the final weighted feature map.

The structure of CBAM module is shown in Fig. 5.

As shown in the earlier figure, CBAM consists of an input link, a channel attention module link, a spatial attention module link, and an output link. Input features, and then perform a one-dimensional convolution operation in the CAM module. Multiply the result with
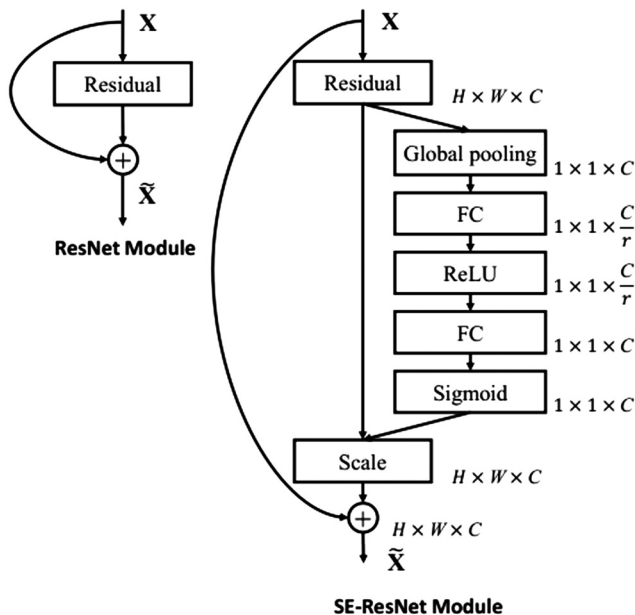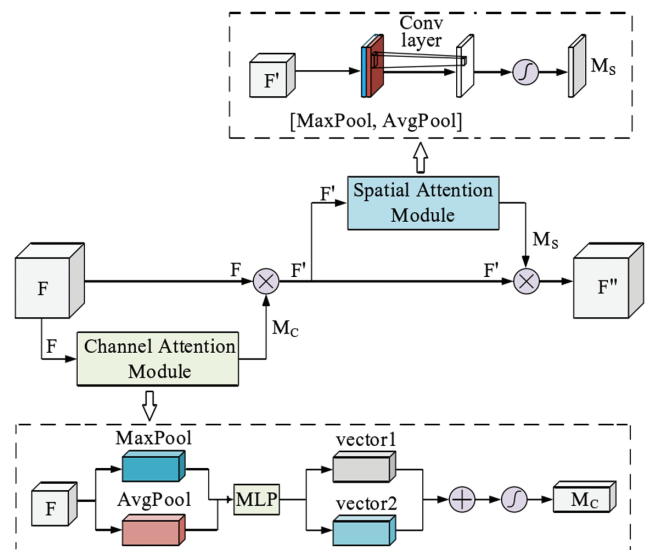

Fig. 4. SE-ResNet workflow.

Fig. 5. CBAM model.

the original image to obtain this value as the input. Perform a two-dimensional convolution operation through the SAM module, and multiply the output result with the original image again to obtain the desired result [30].

By adding CBAM modules to the ResNet50 model, the two can be combined to enhance their performance in graphic classification work.

The specific method is to selectively add CBAM modules to the residual blocks of the ResNet50 model. Between the input and output of each residual block, the CBAM module is used to increase the weighting process of channel and spatial attention. The CAM module can learn the correlation between channels by averaging and maximizing pooling for each channel and then use a fully connected layer to calculate channel attention. The SAM module can capture spatial structures in images through spatial average pooling and maximum pooling and use fully connected layers to calculate spatial attention. Finally, multiplying the channel and spatial attention will help enhance the feature representation ability of the image. Therefore, by adding the CBAM module to the ResNet50 model, the model can adaptively focus on important features and improve its feature extraction ability and classification accuracy.

In this experiment, we embedded CBAM into the bottleneck structure of ResNet50 to form a new structure to help suppress non critical feature channels in image data and enhance key feature channels.

Sixteen CBAM modules have been added to the original structure of ResNet50 to extract attention information. The ResNet work model is focused on key information to improve the recognition performance of the model. At the same time, the CBAM bottleneck structure involved 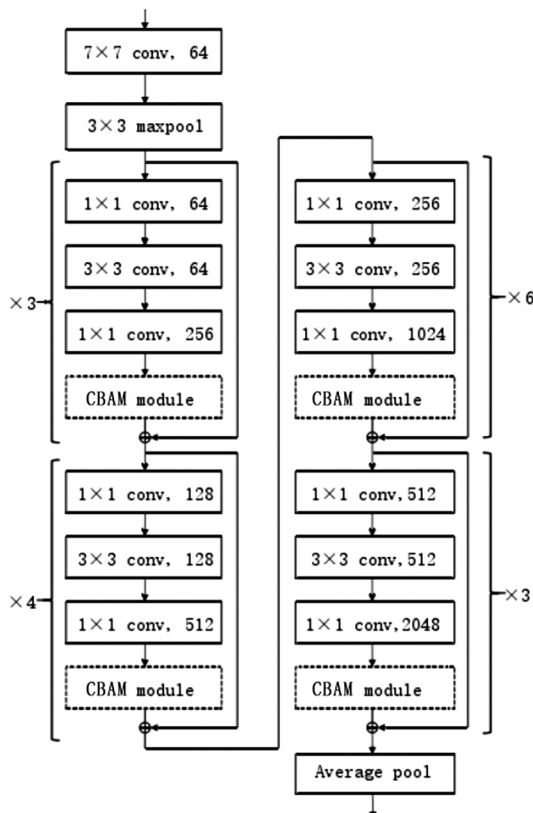in this paper is a modular structure, which can be copied to other residual network structures with different network layers, as shown in Fig. 6.

D. TRANSFER LEARNING

Transfer learning is an important data processing method that can be applied to deep learning training. It aims to apply the preacquired knowledge and experience to various working environments. It can borrow empirical data obtained from one task from another task under similar conditions. Transfer learning starts from this similarity between two tasks and applies the beneficial parameter transfer learned in the old model to the new model. In computer vision work, transfer learning is often applied to neural networks with convolution, which can extract and summarize the features of objects, thus obtaining some basic structures. And the basic structure of many neural networks is similar [31]. So we can use some mature deep learning networks to extract the network parameters of the features and then connect them to our own network. The resulting neural network can only extract more advanced features without extracting some basic structures [32]. The goal of training a well-performing network in a relatively short period of time can be achieved.

There are two common working methods of transfer learning: the method of fine-tuning the network and the method of taking the model as feature extraction. The difference between the two is that the former fine-tuning trains all network parameters after changing the network, while the latter freezes the previous layer and only trains the last layer, resulting in a fast training speed.

In the research of this paper, we first train the ImageNet dataset to find the appropriate feature quantity and parameter data, and then apply it to the self-collected data set to make up the shortage of small number of sample data sets in analysis work.

IV. THE DATASET

There are many types of fruit trees. In order to better detect diseases, we select in this experiment leaf image data of the four most common fruits in China as representatives, and we divide them into 10 types based on their health status. They are

(a) apple juniper rust, (b) apple health, (c) apple gray spot, (d) peach health, (e) peach bacterial plaque, (f) grape black rot, (g) grape health and (h) grape leaf blight (i) cherry powdery mildew and (j) cherry health. A total of 6023 image samples. As shown in Fig. 7.

Due to the limited number of images contained in this dataset, in order to facilitate model training, this experiment preprocessed and enhanced the collected data images, including resizing, cropping,
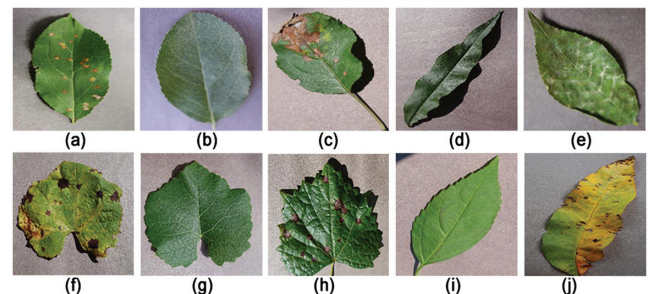

Fig. 6. CBAM-ResNet workflow.

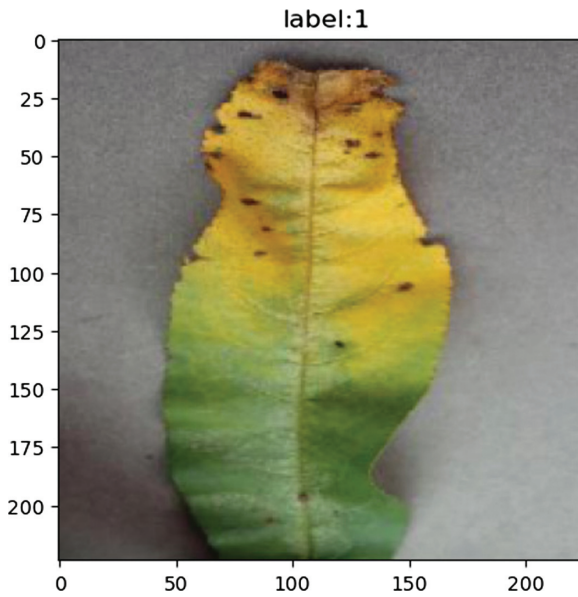

Fig. 7. Data set of fruit tree disease.
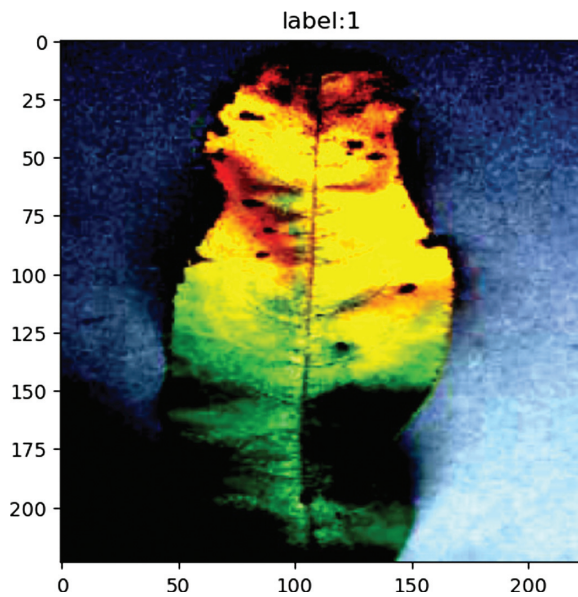
Fig. 8. Before image preprocessing.

Fig. 9. After image preprocessing.

tensor, and normalization operations. The comparison effect between the two is as follows Figs. 8 and 9.

In the experimental process used in this paper, the dataset was divided into two parts, with 80% of the data samples being divided into the training set and the other 20% being divided into the testing set.

V. EXPERIMENTAL DESIGN

A. EXPERIMENTAL ENVIRONMENT

In the experiment of this paper, Python 3.9 was used as the programming language, Anaconda3 as the development environment, Python as the learning library environment, and CUDA11.3 as the computing platform. The final training work is completed through cooperation between the local computer and the network server based on the computational complexity.

B. EVALUATION INDICATORS

To correctly evaluate the performance of this research in fruit tree disease recognition, combined with the objective situation that the sample category of the image dataset in this study is relatively small, I decided to use accuracy and recall as an index to evaluate the performance.

The accuracy is the number of predicted correct samples/all samples. The calculation formula is shown in formula (1):

$$Accuracy = (TP + TN)/(TP + TN + FP + FN) \qquad (1)$$

The recall rate refers to the ratio of the predicted correct quantity of a category to the actual total quantity of that category. The formula is shown in formula (2):

$$Recall = TP/(TP + FN) \qquad (2)$$

C. RESULT ANALYSIS

In this experiment, different network models such as normal ResNet50, SE-ResNet50 network, and CBAM-ResNet50 were tested for their effectiveness in identifying fruit diseases.

The test loss is shown in from Figs. 10 to 12:

According to the above graphical analysis, the testing losses of SE-ResNet50 and CBAM-ResNet50 in the later stage are superior to the traditional ResNet50 mode.

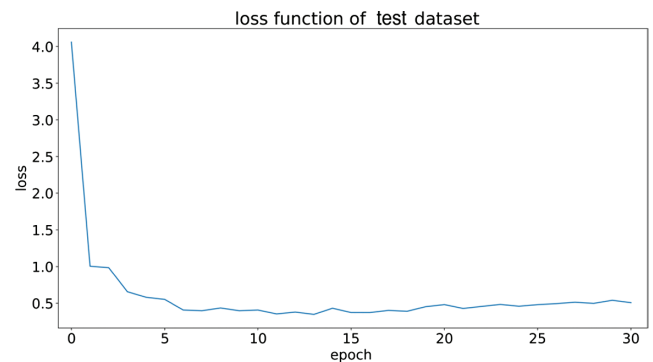The classification accuracy and recall rate are shown in Table I.

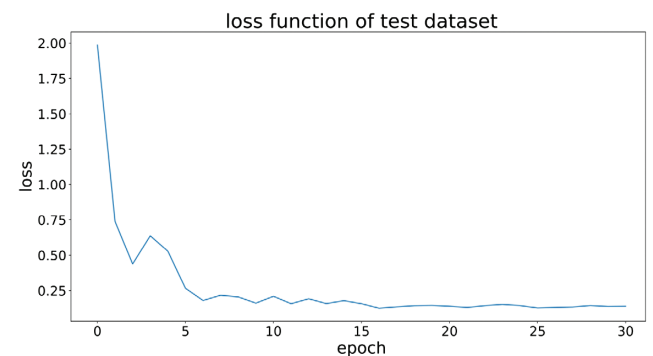**Fig. 10.** Test set loss for traditional ResNet50 work mode.
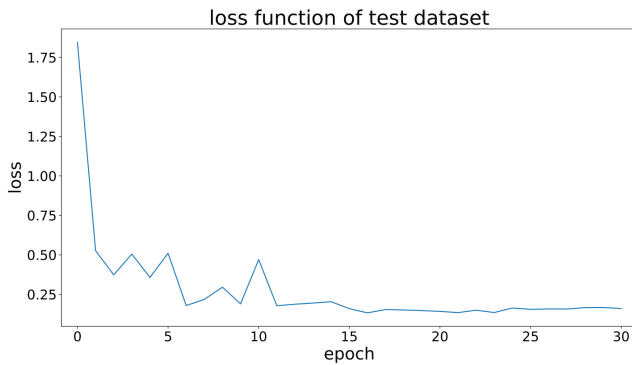

Fig. 11. Test set loss for SE-ResNet50 work mode.

Fig. 12. Test set loss in CBAM-ResNet50 work mode.

Table I. Resnet50 experimental

Model	Accuracy	Recall
ResNet	91.7611	88.8290
SE-ResNet	96.7388	95.8301
CBAM-ResNet	96.9447	96.1036

As shown in the table, the CBAM-ResNet50 model has the highest performance compared to the original ResNet50 model and SE-ResNet. In terms of accuracy, it has improved by 5.18% compared to the ResNet50 network model and 0.2% compared to the SE-ResNet model. In terms of recall rate, it has increased by 7.27% compared to the ResNet50 network model and 0.27% compared to the SE-ResNet model. This experimental result proves that using the CBAM-ResNet50 model can effectively improve the performance of disease classification.

VI. CONCLUSIONS

Rapid identification of plant disease is of great significance to modern fruit cultivation. This paper studies the use of small sample data through data preprocessing and transfer learning. We use ResNet50 as the main working model, and try to combine ResNet, data enhancement, transfer learning, attention mechanism, loss function, and other technologies related to machine learning for comprehensive application. Compared with the traditional ResNet model, the method adopted in this article can achieve better classification results. To explore the optimal attention selection, this experiment optimized ResNet by using two of the most popular attention mechanisms, CBAM and SE, respectively. It was found that in this experiment, the combination effect of CBAM-ResNet50 was slightly better than SE-ResNet50, with an accuracy of 96.94%. Therefore, this experiment demonstrates that the combination of convolutional neural networks and CBAM attention mechanism can effectively improve the accuracy of the model and achieve better performance in fruit tree disease recognition.

CONFLICT OF INTEREST STATEMENT

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

REFERENCES

[1] Y. M. Oo and N. C. Htun, "Plant leaf disease detection and classification using image processing," *Int. J. Res. Eng.*, vol. 5, no. 9, pp. 516–523, 2018.

[2] G. Geetharamani and A. Pandian, "Identification of plant leaf diseases using a nine-layer deep convolutional neural network," *Comput. Electric. Eng.*, vol. 76, pp. 323–338, 2019.

[3] Y. Liu and Y. Ji, "Target recognition of sport athletes based on deep learning and convolutional neural network," *J. Intell. Fuzzy Syst.*, vol. 40, no. 2, pp. 2253–2263, 2021.

[4] Y. Lu, G. Dan, S. He, L. Xin, and H. Zhengjun, "Research on rice leaf disease identification method based on deep learning," *Inf. Record. Mater.*, vol. 19, no. 12, p. 3, 2018.

[5] X. Song and V. Y. Mariano, "Image-based apple disease detection based on residual neural network and transfer learning," in *2023 IEEE 3rd Int. Conf. Power, Electron. Comput. Appl. (ICPECA)*, IEEE, Shenyang, China, Jan. 2023, pp. 365–369.

[6] Xu G., Z. Wang, and C. Li, "A tomato leaf disease recognition model based on deep learning," *Microprocessors*, vol. 41, no. 3, p. 7, 2020.

[7] Jia Q., G. Qi, and X. Hu, "A multi threshold apple disease image segmentation method based on fuzzy c-means clustering," *Intell. Comput. Appl.*, vol. 9, no. 2, p. 5, 2019.

[8] L. Wang, X. Xu, Y. Su, T. Song, Y. Deng, and Z. Fang, "Research on grape leaf disease identification based on computer vision Jiangsu," *Agric. Sci.*, vol. 45, no. 23, p. 4, 2017.

[9] S. Hossain, R. M. Mou, M. M. Hasan, S. Chakraborty, and M. A. Razzak, "Recognition and detection of tea leaf's diseases using support vector machine," in *2018 IEEE 14th Int. Colloquium Signal Process. Its Appl. (CSPA)*, IEEE, Penang, Malaysia, Mar. 2018, pp. 150–154.

[10] B. Li and D. Lima, "Facial expression recognition via ResNet-50," *Int. J. Cognit. Comput. Eng.*, vol. 2, pp. 57–64, 2021.

[11] M. Abisado, A. Trillanes, A. Lacasandile, and A. De La Cruz, "Using low-resourced language in social media platforms towards disease surveillance for public health monitoring using artificial intelligence," in *2022 11th Int. Conf. Software Inf. Eng.*, Macau, China, Oct. 2022, pp. 77–85.

[12] Z. Dai et al., "Symptom recognition of citrus yellow dragon disease based on deep convolutional neural networks," *J. South China Agric. Univ.*, vol. 41, no. 4, p. 9, 2020.

[13] S. Sladojevic, M. Arsenovic, A. Anderla, D. Culibrk, and D. Stefanovic, "Deep neural networks based recognition of plant diseases by leaf image classification," *Comput. Intell. Neurosci.*, vol. 2016, pp. 210–221, 2016.

[14] M. Brahimi, K. Boukhalfa, and A. Moussaoui, "Deep learning for tomato diseases: classification and symptoms visualization," *Appl. Artif. Intell.*, vol. 31, no. 4, pp. 299–315, 2017.

[15] M. M. Ozguven and K. Adem, "Automatic detection and classification of leaf spot disease in sugar beet using deep learning algorithms," *Phys. A: Stat. Mech. Appl.*, vol. 535, p. 122537, 2019.

[16] M. S. Khan, T. D. Nath, M. M. Hossain, A. Mukherjee, H. B. Hasnath, T. M. Meem, and U. Khan, "Comparison of multiclass classification techniques using dry bean dataset," *Int. J. Cognit. Comput. Eng.*, vol. 4, pp. 6–20, 2023.

[17] H. Ponce, C. Cevallos, R. Espinosa, and S. Gutiérrez, "Estimation of low nutrients in tomato crops through the analysis of leaf images using machine learning," *J. Artif. Intell. Technol.*, vol. 1, no. 2, pp. 131–137, 2021.

[18] P. Dixit and S. Silakari, "Deep learning algorithms for cybersecurity applications: a technological and status review," *Comput. Sci. Rev.*, vol. 39, p. 100317, 2021.

[19] A. Ramcharan, K. Baranowski, P. McCloskey, B. Ahmed, J. Legg, and D. P. Hughes, "Deep learning for image-based cassava disease detection," *Front Plant Sci.*, vol. 8, p. 1852, 2017.

[20] A. A. Alfarisy, Q. Chen, and M. Guo, "Deep learning based classification for paddy pests & diseases recognition," in *Proc. 2018 Int. Conf. Math. Artif. Intell.*, Apr. 2018, pp. 21–25.

[21] W. Li, P. Chen, B. Wang, and C. Xie, "Automatic localization and count of agricultural crop pests based on an improved deep learning pipeline," *Sci. Rep.*, vol. 9, no. 1, p. 7024, 2019.

[22] J. Gu et al. "Recent advances in convolutional neural networks," *Pattern Recognit.*, vol. 77, 354–377, 2018.

[23] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," in *2017 Int. Conf. Eng. Technol. (ICET)*, IEEE, Antalya, Turkey, Aug. 2017, pp. 1–6.

[24] J. Wu, "Introduction to convolutional neural networks," *Natl. Key Lab Novel Softw. Technol.. Nanjing Univ. China*, vol. 5, no. 23, p. 495, 2017.

[25] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Y. Fu, and A. C. Berg, "Ssd: single shot multibox detector," in *Comput. Vis.–ECCV 2016: 14th Eur. Conf.*, Amsterdam, The Netherlands, Springer International Publishing, October 11–14, 2016, Proceedings, Part I 14, 2016, pp. 21–37.

[26] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Las Vegas, USA, 2016, pp. 770–778.

[27] G. Liu and J. Guo, "Bidirectional LSTM with attention mechanism and convolutional layer for text classification," *Neurocomputing*, vol. 337, pp. 325–338, 2019.

[28] Z. Niu, G. Zhong, and H. Yu, "A review on the attention mechanism of deep learning," *Neurocomputing*, vol. 452, pp. 48–62, 2021.

[29] W. Zhang, H. Li, Y. Li, H. Liu, Y. Chen, and X. Ding, "Application of deep learning algorithms in geotechnical engineering: a short critical review," *Artif. Intell. Rev.*, vol. 54, pp. 1–41, 2021.

[30] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "Cbam: convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Munich, Germany, 2018, pp. 3–19.

[31] Y. Zhao, L. Liu, C. Xie, R. Wang, F. Wang, Y. Bu, and S. Zhang, "An effective automatic system deployed in agricultural Internet of Things using Multi-Context Fusion Network towards crop disease recognition in the wild," *Appl. Soft Comput.*, vol. 89, pp. 106–128, 2020.

[32] R. Ribani and M. Marengoni, "A survey of transfer learning for convolutional neural networks," in *2019 32nd SIBGRAPI Conf. Graph., Patterns Images Tutorials (SIBGRAPI-T)*, IEEE, Alagoas, Brazil, Oct. 2019, pp. 47–57.