



**HAL**  
open science

# A Pear Leaf Diseases Image Recognition Model Based on Capsule Network

Zhida Jia, Wenqian Mu, Junhua Gong, Yi Zong, Yongjie Liu

► **To cite this version:**

Zhida Jia, Wenqian Mu, Junhua Gong, Yi Zong, Yongjie Liu. A Pear Leaf Diseases Image Recognition Model Based on Capsule Network. 12th International Conference on Intelligent Information Processing (IIP), May 2022, Qingdao, China. pp.358-368, 10.1007/978-3-031-03948-5\_29 . hal-04178749

**HAL Id: hal-04178749**

**<https://inria.hal.science/hal-04178749v1>**

Submitted on 8 Aug 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



This document is the original author manuscript of a paper submitted to an IFIP conference proceedings or other IFIP publication by Springer Nature. As such, there may be some differences in the official published version of the paper. Such differences, if any, are usually due to reformatting during preparation for publication or minor corrections made by the author(s) during final proofreading of the publication manuscript.

# A pear leaf diseases image recognition model based on capsule network

Zhida Jia<sup>1</sup>, Wenqian Mu<sup>2</sup>, Junhua Gong<sup>3</sup>, Yi Zong<sup>4</sup>, Yongjie Liu<sup>\*</sup>

<sup>1</sup> College of Information Science and Engineering,  
Shandong Agricultural University, Taian, China, jzd0608@163.com

<sup>2</sup> College of Information Science and Engineering,  
Shandong Agricultural University, Taian, China

<sup>3</sup> ShiKeFeng Ecological Agriculture Engineering Co., LTD, Linyi, China

<sup>4</sup> Feicheng Bianyuan Town Agricultural Comprehensive Service Center, Taian, China

<sup>\*</sup> College of Plant Protection, Shandong Agricultural University, Taian, China  
lyj@sdau.edu.cn

**Abstract.** Image recognition of pear leaf diseases is an important task of plant protection. The lesion area of pear leaf diseases is not fixed in the whole leaf, which has the characteristics of randomness. The convolution neural network is used to identify the images of pear leaf diseases, due to its rotation and translation invariance, the generalization ability of the model is weak. The capsule network uses feature vectors to replace feature value, and uses dynamic routing to replace pooling to obtain spatial information between entities. However, the size of lesion area of pear leaf diseases is random, and the capsule network cannot fully extract features, resulting in a decrease in recognition rate. To solve the problem, a pear leaf diseases image recognition model based on capsule network was proposed, which uses conditional convolution to customize specific convolution kernels for each input to adapt to pear leaf diseases images of different sizes. The experimental results show that the recognition accuracy, precision, recall and F1score of the proposed algorithm are 91.33 %, 91.40 %, 91.33 % and 91.36 %, which are better than capsule network.

**Keywords:** Capsule Network, Conditional Convolution, Image Recognition, Images of Pear Leaf Diseases.

## 1 Introduction

Pear is the third largest fruit in China, accounting for about 80 % of the world ' s planting area [1]. However, due to the occurrence of pear leaf diseases, resulting in pear branches withered, fruit decay, less fruit, the yield and quality of pear have a serious impact. The prevention and control of pear leaf diseases is mainly divided into two steps. Firstly, the leaf diseases of pear are identified, and then the appropriate pesticides are selected according to the identification results. At present, the leaf diseases of pear are mainly identified manually by plant protection experts based on their experience, which are subjective and prone to errors. At the same time, plant protec-

tion experts and technical personnel are limited, which is difficult to timely diagnose the occurrence of early diseases, resulting in missing the best prevention period and causing serious economic losses to fruit farmers. Therefore, automatic identification of pear leaf diseases is an urgent problem to be solved in plant protection.

Plant leaf diseases image recognition mainly includes classical machine learning and deep learning. Image recognition algorithm based on classical machine learning mainly includes Bayesian classification, support vector machine (SVM) and other algorithms. Its advantage is that it can use a small amount of data samples to train, but requires manual extraction of features. Image recognition algorithm based on deep learning, especially convolution neural network, is widely used in the field of image recognition. The convolutional neural network has strong feature extraction ability, but it has rotation and translation invariance, and its training needs massive data. The capsule network [2] is a new attractive neural network structure. Its feature vector can represent the size and direction of the target, and solve the problem of rotation and translation invariance. However, the feature extraction ability of capsule network is weak, resulting in low recognition accuracy. To solve this problem, this paper proposes a capsule network based on conditional convolution [3], which uses the characteristics of conditional convolution to customize convolution kernel for each input sample to adapt to different sizes of diseases characteristics and improve the recognition accuracy.

## 2 Related Work

Automatic identification of plant leaf diseases images is an urgent problem to be solved for plant protection. Early, with the development of machine learning, many researchers used the combination of image processing and machine learning to achieve plant diseases image recognition. Almadhor et al. [4] took pomegranate leaf and fruit diseases images as the research object, used color difference to segment the diseases area, and extracted the RGB and HSV color histogram and LBP texture features of the diseases. Finally, the extracted features are classified using Bagged Tree. Padol et al. [5] uses K-means clustering to segment the diseases area, and then extracts color and texture features. Finally, SVM is used to classify grape leaf diseases images. Later, with the rapid development of deep learning, it has been widely used in agriculture. Hang et al. [6] proposed a model combining Inception module, squeeze-and-excitation module and global pooling layer to realize plant leaf diseases image recognition. The model has the advantages of less parameters and easy training. Wu et al. [7] proposed a dual-channel model based on ResNet50 and VGG16 to identify maize leaf diseases. Bansal et al. [8] proposed integrating DenseNet and EfficientNet to detect apple diseases images. The purpose of capsule network is to replace convolution neural network, which has been widely used in agriculture due to its less parameters and easy training. Verma et al. [9] used capsule network to achieve potato diseases image classification. Patrick et al. [10] proposed capsule network based on K-Means routing to realize tomato diseases image recognition. Kwabena et al. [11] proposed Gabor capsule network, which combines global Gabor filters and convolu-

tion kernels to extract features to identify blurred and deformed tomato and citrus diseases images. Wang et al. [12] proposed a capsule network based on attention mechanism. The attention module was added to the capsule to reduce the influence of noise and realize the fine-grained identification of crop pests.

### 3 Proposed Method

#### 3.1 Capsule Network

Capsule network [2] (CapsNet) is a new vector neural network, whose input and output are vectors. The direction of the vector can represent the spatial information of the target, such as attitude, position, size, direction and deformation. The length of the vector represents the probability of the existence of the target. The structure of capsule network mainly includes convolution layer, Primarycap layer, Digitcaps layer and Decoder layer. The convolution layer is used to extract features. Primarycap converts the extracted features into vector capsules, and then converts them into digital capsules through dynamic routing. Finally, the class of images is obtained according to  $\|L2\|$  of digital capsules.

The dynamic routing structure of the capsule network as shown in Fig. 1. The capsule network uses dynamic routing to combine the key features in the feature map. In the dynamic routing, high-level capsules and low-level capsules are fully connected, and the output  $V_j$  of high-level capsules is shown in (1).

$$v_j = \frac{\|s_j\|^2}{1 + \|s_j\|^2} \cdot \frac{s_j}{\|s_j\|} \quad (1)$$

Where  $V_j$  is the output of high-level capsule,  $S_j$  is the weighted sum of all low-level capsule prediction vectors, as shown in (2).

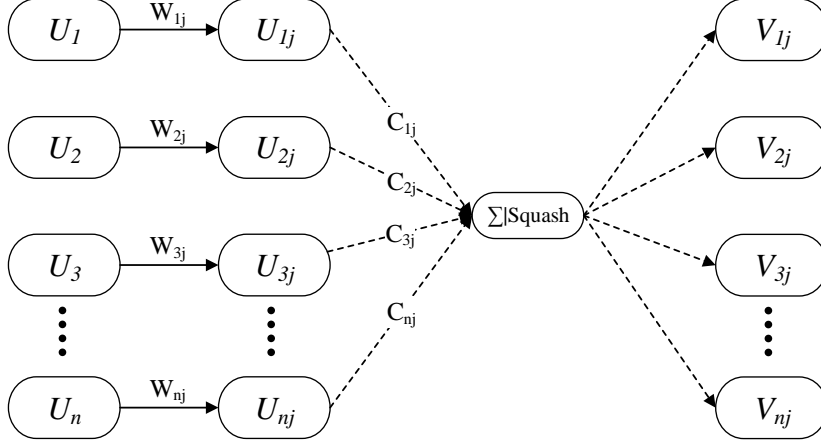
$$s_j = \sum_i c_{ij} \hat{u}_{ji} \quad (2)$$

Where  $C_{ij}$  is the coupling coefficient of each low-level capsule  $i$  connected to the high-level capsule  $j$ , as shown in (3).  $\hat{u}_{ji}$  is the prediction vector of low-level capsules, as shown in (4).

$$c_{ij} = \frac{\exp(b_{ij})}{\sum_k \exp(b_{ik})} \quad (3)$$

$$\hat{u}_{ji} = W_{ij} u_i \quad (4)$$

Where  $W_{ij}$  is the weight matrix,  $u_i$  is the output vector of low-level capsules, and  $b_{ij}$  is the prior probability of coupling capsule  $i$  and capsule  $j$ .



**Fig. 1.** The structure of dynamic routing

Similar to convolution neural network, capsule network measures the gap between the predicted results of the model and the actual results through the loss function in the training process. However, the difference is that the capsule network allows multiple classes to exist at the same time, so the cross-entropy loss function cannot be simply used, but is composed of margin loss and reconstruction loss. For each class by a margin loss, the margin loss of the  $k$  class is shown in (5).

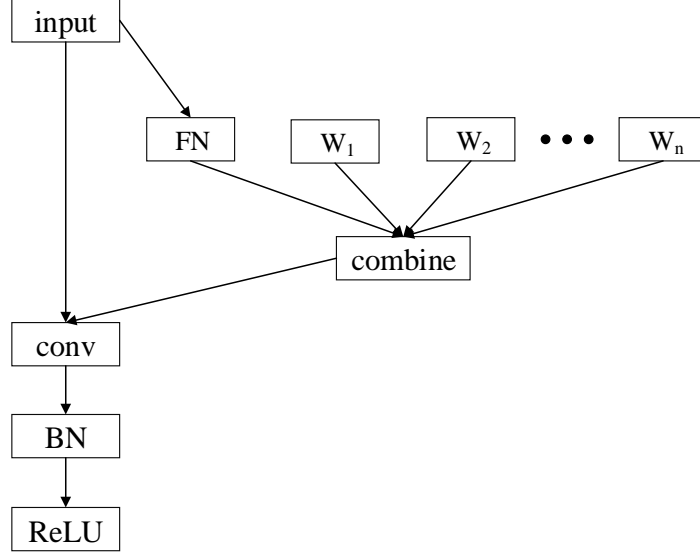
$$L_k = T_k \max(0, m^+ - \|v_k\|)^2 + \lambda(1 - T_k) \max(0, \|v_k\| - m^-)^2 \quad (5)$$

Where  $T_k$  denotes whether class  $k$  exists. If  $k$  class exists, the value is 1, else  $k$  class does not exist, the value is 0.  $m^+$  is the upper bound of 0.9 to punish false positives. When  $k$  class exists but not predicted, the loss value increases.  $m^-$  is the lower bound of 0.1 to penalize false negatives.  $\lambda$  is a proportional coefficient, which is used to balance the upper and lower bounds, and the value is 0.5. If class  $k$  exists,  $\|v_k\|$  is greater than 0.9, and if class  $k$  does not exist,  $\|v_k\|$  is less than 0.1.

The reconstruction loss is composed of  $32 \times 32$  pixels of the input image minus  $32 \times 32$  pixels of the reconstructed image, and the square is multiplied by the coefficient 0.005. The margin loss is the main component of the loss function.

### 3.2 Conditional Convolution

Conditional convolution [3] (CondConv) breaks the limitation of using the same convolution kernel parameters for all input samples by traditional convolution, and realizes the customization of specific convolution kernel parameters for each sample input. In traditional convolutional neural networks, in order to extract sufficient features, the network is generally deepened, resulting in a sharp increase in model parameters and difficult training. By increasing the number of experts, conditional convolution achieves the purpose of extracting sufficient features with a small number of parameters, as shown in Fig. 2. By learning the input sample features, the convolution kernel is parameterized conditionally to achieve the effect of multiple convolutions.



**Fig. 2.** The structure of conditional convolution

The convolution kernel of conditional convolution is customized for each input sample, and the output is calculated using the customized convolution kernel, as shown in (6).

$$y = \sigma((\alpha_1 \cdot W_1 + \dots + \alpha_n \cdot W_n) * x) \quad (6)$$

Where  $y$  is the convolution kernel output,  $x$  is the convolution kernel input,  $W$  is the standard convolution kernel,  $\alpha$  is the scalar calculated by each expert according to the input, a total of  $n$  experts,  $\alpha = r(x)$ , as shown in (7).

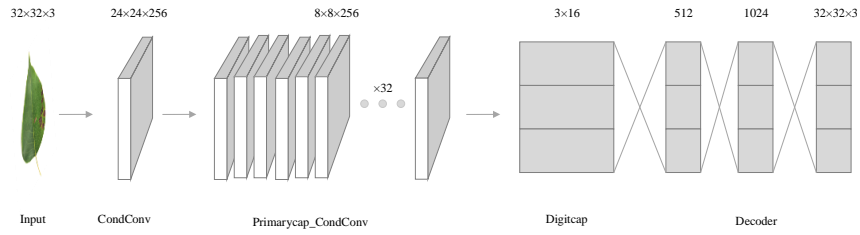
$$r(x) = \text{Sigmoid}(\text{GlobalAveragePool}(x)R) \quad (7)$$

Where  $R$  is a full connection layer with  $n$  neurons, and  $n$  expert weights are obtained. The values are constrained to  $[0,1]$  by Sigmoid as the weights of  $n$  convolution kernels.

### 3.3 CondConv CapsNet

Although CapsNet has the ability to capture the spatial information of features, such as the relative position relationship, relative size relationship and feature direction of features, which effectively solves the problem of rotation and translation invariance of CNN, its recognition accuracy still needs to be further improved. Due to the randomness of the lesion size of pear leaf diseases, the convolution kernel receptive field in CapsNet is fixed, and the diseases characteristics cannot be fully extracted, resulting in the decrease of recognition accuracy. Therefore, this paper introduces conditional convolution into the capsule network and proposes a capsule network model based on

conditional convolution, and its structure as shown in Fig. 3. The  $32 \times 32 \times 3$  pear leaf diseases images were input, and the diseases characteristics of different sizes were fully extracted by conditional convolution with 256 channels,  $9 \times 9$  size and 1 stride. Then, through the Primarycap layer, each capsule contains a conditional convolution with 8 channels,  $9 \times 9$  size and 2 stride, and a total of 32 capsules, of which low-level capsules and high-level capsules are connected by dynamic routing. Then, through the Digitcap layer, each class contains one 16D capsule, a total of three classes. Finally, through the Decoder layer, it is composed of three fully connected layers. Each layer contains 512, 1024 and  $32 \times 32 \times 3$  neurons, and finally the classification results are output.



**Fig. 3.** The structure of CondConv CapsNet

## 4 Experiment













### 4.1 Data Construction and Processing

In this paper, pear leaf diseases images were taken as the research object. There is no public pear leaf disease image dataset. So, the pear leaf diseases images were collected in Xiazhang pear orchard of Taian City, Shandong Province. A total of 900 images were collected, including 491 images of brown spot, 333 images of gray spot and 76 images of ring rot.

In order to improve the efficiency of model training and facilitate batch processing, the images size are unified to  $32 \times 32$ . In order to reduce the influence of complex background on recognition results, disease images under single background are collected. The constructed leaf diseases images of pear as shown in Table 1, where the size and location of lesion are random.









**Table 1.** Image samples of pear leaf diseases.

Classes	Sample 1	Sample 2	Sample 3	Sample 4
Brown Spot				
Gray Spot				
Ring Rot				

In order to improve the generalization ability of capsule network and avoid overfitting during training, the images data of pear leaf diseases were enhanced. In order to enhance the adaptability of the model to different shooting angles, the original image is rotated and flipped. In order to improve the adaptability of the model to different illumination conditions, the brightness of the original image is adjusted. In order to improve the adaptability of the model to diseases at different locations, the original image is translated. In order to improve the anti-noise ability of the model, salt and pepper noise is added to the original image. At the same time, in order to reduce the impact of data imbalance on the performance of the model, each kind of diseases images are expanded to 5000, a total of 15000. The image enhancement method example is shown in Table 2. The enhanced data are divided into training set, verification set and test set according to the ratio of 6 : 2 : 2. Among them, 9000 training sets are used to train model parameters, 3000 verification sets are used to optimize model parameters, and 3000 test sets are used to evaluate model performance.

**Table 2.** Image enhancement methods.

Artwork	Flip	Rotated	Translation	Brightness	Noise
					

## 4.2 Evaluation Standard

In order to quantify the performance of the model, Accuracy, Precision, Recall and F1score were used as evaluation criteria in this experiment, as shown in (8)-(11).

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

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

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

$$F1score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (11)$$

Where TP is the number of positive samples predicted as positive samples, TN is the number of negative samples predicted as negative samples, FP is the number of negative samples predicted as positive samples, and FN is the number of positive samples predicted as negative samples.

## 4.3 Environment and Parameter

The hardware environment of this experiment is Intel Core i7-8700K CPU @ 3.70GHz, 16GB memory, NVIDIA GeForce GTX 1080Ti GPU 12G memory. The software environment is Windows 7 system, Python 3.6, TensorFlow 2.3.0. The hyperparameters of model training as shown in Table 3.

**Table 3.** Hyper-parameter setting.

Model	Learn rate	Batch size	Epoch	Lr_decay	Experts
CapsNet	0.0002	20	100	0.9	4

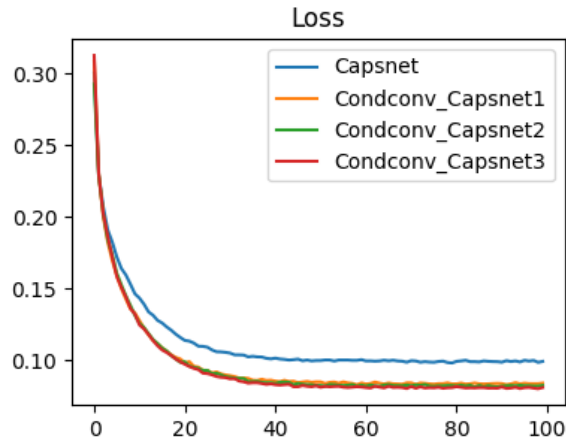
## 4.4 Results and Analysis

In order to verify the proposed method, the experimental scheme is designed as shown in Table 4, and the model in the experimental scheme is used to train the pear leaf diseases images. Where the structure of CapsNet is convolution layer, Primarycap layer based on convolution, Digitcap layer and Decoder layer. CondConv\_CapsNet1 replaces the convolution layer with the conditional convolution layer, and the other structures are the same as CapsNet. CondConv\_CapsNet2 replaces the Primarycap layer based on convolution with the Primarycap layer based on conditional convolution, and the other structures are the same as CapsNet. CondConv\_CapsNet3 replaces the convolution layer with the conditional convolution layer, and the Primarycap layer based on convolution with the Primarycap layer based on conditional convolution, and the other structures are the same as CapsNet.

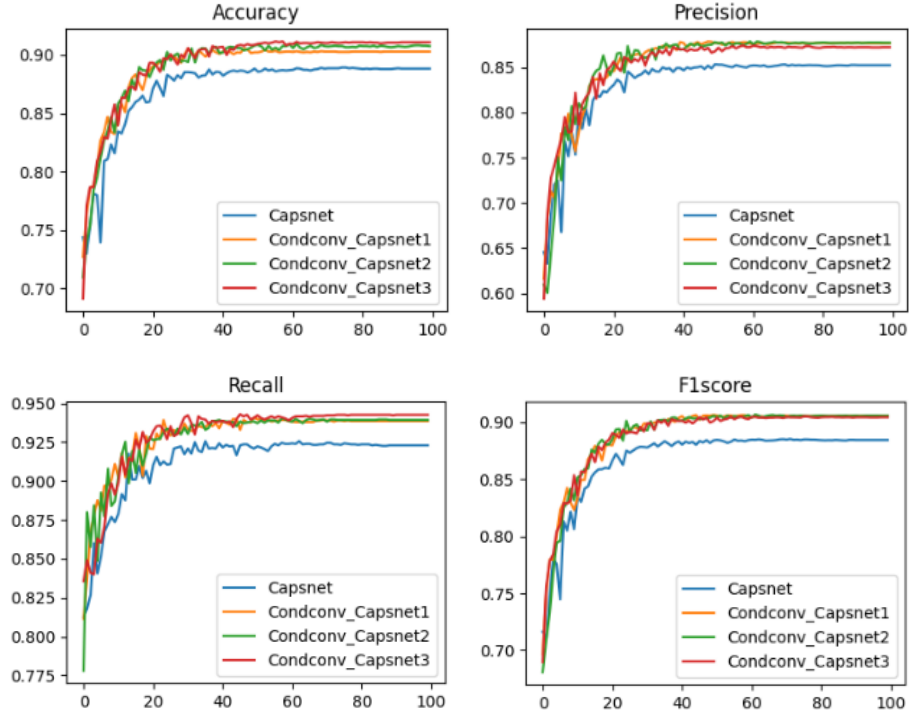
**Table 4.** Experimental schemes.

Model	Layer			
CapsNet	Conv	Primarycap_conv	Digitcap	Decoder
CondConv_CapsNet1	CondConv	Primarycap_conv	Digitcap	Decoder
CondConv_CapsNet2	Conv	Primarycap_condconv	Digitcap	Decoder
CondConv_CapsNet3	CondConv	Primarycap_condconv	Digitcap	Decoder

Four groups of models are trained, and the loss changes in the training process as shown in Fig. 4. CapsNet and the improved CapsNet converge after 40 rounds of training. The loss after convergence of CapsNet is 0.1, and the loss after convergence of the three groups of improved CapsNet is 0.084, 0.082 and 0.080, respectively, which is significantly lower than that of CapsNet, where the loss value of CondConv\_CapsNet3 is the lowest, indicating that the prediction results of CondConv\_CapsNet3 are closer to the real results.

**Fig. 4.** The loss variation curve

Four groups of models are trained, and the accuracy, precision, recall and F1score change curve of the four groups of models on the verification set are shown in Fig. 5. It can be seen from the four groups of change curves that the recognition effect of the three groups of improved CapsNet is significantly better than that of CapsNet, and Condconv\_CapsNet3 has the highest accuracy, recall and F1score.



**Fig. 5.** The accuracy and precision and recall and F1score variation curve

The four groups of convergent models are tested on the test set respectively. The test results as shown in Table 5. The three groups of improved CapsNet are higher than CapsNet in four evaluation criteria. Among them, CondConv\_CapsNet3 has the best recognition effect, which is 2.4 %, 2.35 %, 2.4 % and 2.37 % higher than CapsNet in accuracy, precision, recall and F1score, respectively.

**Table 5.** Test results of different models.

Model	Accuracy/%	Precision/%	Recall/%	F1score/%
CapsNet	88.93	89.05	88.93	88.99
CondConv_CapsNet1	90.40	90.55	90.40	90.47
CondConv_CapsNet2	91.00	91.09	91.00	91.04
CondConv_CapsNet3	91.33	91.40	91.33	91.36

The identification accuracy of the four models for different diseases is shown in Table 6. The four models have the highest recognition accuracy for Gray Spot. Because of its obvious characteristics and the sufficient original data. The recognition accuracy of Brown Spot and Ring Rot is low, because the color and texture features of the two diseases are similar. The recognition accuracy of Ring Rot is the lowest,

because the original data are less. The improved model has improved the recognition accuracy of the three diseases, and the model proposed in this paper has the highest recognition accuracy. Compared with the capsule network, the recognition accuracy of Brown Spot, Gray Spot and Ring Rot is increased by 1.9 %, 2.2 % and 3.1 %, respectively, indicating that the proposed model has strong generalization ability.

**Table 6.** The accuracy of different diseases.

Model	Brown Spot/%	Gray Spot /%	Ring Rot /%
CapsNet	90.80	92.60	83.40
CondConv_CapsNet1	92.70	93.80	84.70
CondConv_CapsNet2	92.80	93.80	86.40
CondConv_CapsNet3	92.70	94.80	86.50

## 5 Conclusion

In this paper, the advantages and disadvantages of capsule network were analysed, and the characteristics of capsule network adapting to the rotation and translation of diseases spots were used to solve the problem of randomness in the location of diseases incidence area of pear leaves. At the same time, in view of the problem that the receptive field of the capsule network is fixed, and the size of the diseased area of the pear leaf is random, and the capsule network cannot fully extract the characteristics, a capsule network based on conditional convolution is proposed to realize the image recognition of pear leaf diseases. The convolution layer of the first layer is replaced by conditional convolution, and the conditional convolution is used to customize the specific convolution kernel for each input to fully extract the characteristics of different lesions. At the same time, the convolution of Primarycap is replaced by conditional convolution to enhance the feature expression ability of the whole network. The experimental results show that the proposed model has higher accuracy on the same test set, which provides a solution for automatic identification of pear leaf diseases images.

## Acknowledgements

This study was funded by College Students' Innovative Entrepreneurial Training Plan Program—the research on pear leaf diseases image recognition and fruit counting method based on deep learning and the project of introducing urgently needed talents in key support areas of Shandong Province in 2021—the key technology research and application of intelligent water and fertilizer integration based on big data.

## References

1. Bagheri N, Mohamadi-Monavar H, Azizi A, et al. Detection of Fire Blight disease in pear trees by hyperspectral data[J]. *European Journal of Remote Sensing*, 2018, 51(1): 1-10.
2. Sabour S, Frosst N, Hinton G E. Dynamic routing between capsules[J]. *arXiv preprint arXiv:1710.09829*, 2017.
3. Yang B, Bender G, Le Q V, et al. Condconv: Conditionally parameterized convolutions for efficient inference[J]. *arXiv preprint arXiv:1904.04971*, 2019.
4. Almadhor A, Rauf H T, Lali M I U, et al. AI-Driven Framework for Recognition of Guava Plant Diseases through Machine Learning from DSLR Camera Sensor Based High Resolution Imagery[J]. *Sensors*, 2021, 21(11): 3830.
5. Padol P B, Yadav A A. SVM classifier based grape leaf disease detection[C]//2016 Conference on advances in signal processing (CASP). IEEE, 2016: 175-179.
6. Hang J, Zhang D, Chen P, et al. Classification of plant leaf diseases based on improved convolutional neural network[J]. *Sensors*, 2019, 19(19): 4161.
7. Wu Y. Identification of Maize Leaf Diseases based on Convolutional Neural Network[C]//*Journal of Physics: Conference Series*. IOP Publishing, 2021, 1748(3): 032004.
8. Bansal P, Kumar R, Kumar S. Disease Detection in Apple Leaves Using Deep Convolutional Neural Network[J]. *Agriculture*, 2021, 11(7): 617.
9. Verma S, Chug A, Singh A P. Exploring capsule networks for disease classification in plants[J]. *Journal of Statistics and Management Systems*, 2020, 23(2): 307-315.
10. Patrick M K, Weyori B A, Mighty A A. Capsule network with K-Means routing for plant disease recognition[J]. *Journal of Intelligent & Fuzzy Systems*, 2021 (Preprint): 1-12.
11. Kwabena P M, Weyori B A, Mighty A A. Gabor Capsule Network for Plant Disease Detection[J]. *(IJACSA) International Journal of Advanced Computer Science and Applications*, 2020, 11(10).
12. Wang X, Wang X, Huang W, et al. Fine-Grained Recognition of Crop Pests Based on Capsule Network with Attention Mechanism[C]//*International Conference on Intelligent Computing*. Springer, Cham, 2021: 465-474.