Research on Crop Disease Classification Algorithm Based on Mixed Attention Mechanism

Zhen Chen¹, Maoyong Cao¹*, Peng Ji¹ and Fengying Ma¹

¹ School of Electrical Engineering and Automation, Qilu University of Technology (Shandong Academy of Sciences), Jinan, Shandong, 250000, PRC

*Corresponding author’s e-mail: cmy@qlu.edu.cn

Abstract. The prevention and control of crop diseases is an important measure to ensure the yield of crops, and the prerequisite for achieving this link is to improve the accuracy of crop disease identification. This paper constructs a new hybrid attention mechanism by realizing the serial connection of Spatial attention and Efficient Channel Attention (ECA) and proposes a new hybrid attention module Spatial_Efficient Channel Attention (S_ECA). This module is combined with the neural network to improve the disease classification and recognition ability of the network model, the network stability, and the overall performance of the network. In the classification of a part of the data of the crop disease data set in the 2018AI_Chllenger competition, a classification accuracy of 87.28% was achieved, which is an increase of 0.78% compared to the original network, thus verifying the effectiveness of the algorithm in this paper.

1. Introduction

Crop diseases are closely related to food security. The prevention and control of plant diseases are conducive to the growth of crops, thus improving the yield of crops and ensuring the quality of agricultural products, to greatly reduce economic losses[1]. The premise of improving the efficiency of pesticide control is to improve the accurate identification of plant diseases. Traditional crop-related identification tasks are usually completed manually by personnel with professional training or rich experience, but this kind of human behavior will more or less have certain errors or even wrong judgments. Therefore, a tool that can automatically identify plant diseases is particularly important[2].

With the development of computer vision and Internet and other emerging technologies, machine learning technology has been rapidly applied in the field of crop disease identification. Traditional machine learning methods are usually classified by spot extraction and edge feature extraction. However, this traditional machine learning method has many steps for image preprocessing, and the process is complex, so it can only be applied to a few diseases and insect pests detection of individual plants, and its portability is poor. However, with the development of deep learning technology, deep learning solves the problem of traditional machine learning requiring explicit extraction of disease and insect characteristics and poor mobility with its powerful learning ability and effectively improves the performance and accuracy of neural networks. With the characteristics of strong generalization, deep learning technology can identify and detect a wide variety of diseases and insect pests at the same time, which greatly improves the accuracy and control effect of crop disease classification. For example, by combining convolutional neural network with support vector machine (SVM), Qin Feng et al.[3] used the image of the affected area segmented by K-median clustering algorithm and linear discriminant analysis as the input training network model of the convolutional neural network to identify the Endospora leaf spot, rust, and small photobook leaf spot. Foreign Fujita et al.[4] built a model based on...
a convolutional neural network to recognize 7520 cucumber leaf images in the natural environment, and the accuracy rate reached 82.3%.

In this paper, the attention module is injected into the neural network. By realizing the serial connection between the Spatial attention mechanism and the Efficient Channel Attention (ECA), the hybrid attention mechanism Spatial _ Efficient Channel Attention (S_ECA) is constructed to reduce the feature extraction process. It can improve the overall performance of the network and improve the ability of the network to classify and recognize diseases.

2. The relevant knowledge

2.1. Attentional mechanism
In recent years, with the widespread use of deep learning, the research on image classification has achieved rapid development. However, the existing neural network models in image classification still have problems such as insufficient ability to represent the characteristics of the classified objects[5]. However, the attention model is widely used in various fields of deep learning because it can highlight more useful information in the process of feature extraction, suppress the interference of useless information, and effectively improve the classification performance of the network model.

According to different scopes, attention can be divided into spatial attention mechanisms, channel attention mechanisms, and mixed attention mechanisms. Spatial attention allows neural networks to pay more attention to the pixel regions that play a decisive role in the classification of images and ignore the irrelevant regions. Channel attention is an attention mechanism that considers the relationship between channels in a feature graph. Mixed domains combine these two types of attention.

2.2. ResNet neural network
ResNet is a residual network model proposed by He Kaiming et al. in 2016[6]. ResNet network uses residual units, which on the one hand reduces the number of parameters, on the other hand, direct connection channels are added to the network to improve the ability of CNN to learn features. Compared with the previous VGG network, it solves the problem of training degradation caused by the deepening of network depth, and greatly promotes the application of deep neural networks in the field of image recognition. The ResNet residual module is shown in figure 1.

![Figure 1. ResNet residuals module](image)

3. The algorithm in this paper

3.1. ECA module
In recent years, people generally choose to inject channel attention module into the convolution block of the neural network, so as to improve network performance. SENet is a typical network that is based on channel attention. The network generates the attention mechanism based on the channel domain by
obtaining the correlation degree between each channel and the important information. For the channel attention module in SENet, the SE module independently adopts global average pooling for each channel after given input characteristics, and then uses two fully-connected (FC) layers and nonlinear Sigmoid function to generate channel weights[7]. The module structure diagram is shown in figure 2.

Among them, the FC layer is used to capture the nonlinear cross-channel interaction. In this process, the complexity of the model is reduced by reducing the dimensionality. However, after dimensionality reduction, it will have an adverse effect on channel attention prediction, and the dependence between each channel will be reduced. At the same time, due to compression and excitation in SENet, the model will become more complicated and the calculation difficult.

In order to solve this problem, the Efficient Channel Attention(ECA) module is introduced here. Compared with the SE module, the Efficient Channel Attention(ECA) module learns the channel attention mechanism in a more effective way. Its model is not complicated and the computational difficulty is not high. At the same time, the reduction of the dimensionality in the channel interaction is avoided, and the efficiency of the channel interaction is improved. The structure is shown in figure 3. Since ECA performs channel-by-channel global average pooling without reducing the dimensionality, each channel and its k neighbors effectively capture the interaction between different channels, ensuring the effectiveness of channel attention[8]. Different from the SE module, ECA generates channel weights by performing fast one-dimensional convolution of size K, where K is adaptively determined by a function of channel dimension C.

$$K = \varphi(C)$$
3.2. Build the mixed domain attention S_ECA

The mixed domain attention mechanism is a combination of channel attention mechanism and spatial attention mechanism. There are many attention models in the mixed domain. The main difference lies in two aspects: first, there are differences in spatial domain attention mechanism and the computing mode of channel and attention; Second, some attention is processed in parallel in spatial domain channel domain, while some attention is processed in serial in the spatial domain and channel domain[9].

The algorithm in this paper uses the Efficient Channel Attention (ECA) module to replace the traditional channel attention module, and connects it in series with the Spatial attention mechanism to realize the attention structure from space to channel, namely Spatial Efficient Channel Attention (S_ECA) Module. This module uses spatial attention to identify relevant areas in the image that are decisive for classification and then passes them to the ECA module to achieve effective cross-channel interaction. This serial connection ensures the effect of the attention module and also improves the recognition accuracy and network stability in classification tasks.

The mixed attention module S_ECA was constructed by first input feature graph F, weighted the result $F_1$ by the Spatial attention mechanism, and then weighted by ECA to get the output feature graph $F_2$. The formula is shown below:

$$F_1 = M_S(F) \otimes F$$

$$F_2 = M_{EC}(F_1) \otimes F_1$$

In the formula, $M_S$ is the output weight of F after spatial attention; $M_{EC}$ is the output weight of $F_1$ after ECA.

The structure of the constructed mixed attention module S_ECA is shown in figure 4.

![Figure 4. S_ECA structure diagram of the mixed attention module](image)

4. Experimental comparison and analysis

4.1. Experimental data set and it is preprocessing

The data used in this article comes from a part of the crop disease data set in 2018AI_Chllenger, including ten crops such as apples, cherries, and strawberries, with a total of 27 diseases and 61 disease categories. Through the analysis of the data, since there is only one picture in the 44th and 45th categories, it is decided to delete these two categories to conduct the experiment. The processed data is divided into 59 categories according to the degree of disease, which is healthy—normal—severe. Among them, there are 31,493 pictures in the training set and 4,527 pictures in the validation set. Some images of the data set are shown in figure 5.

Due to raw data, centralized data distribution is extremely uneven, one of the most potato health categories a total of 1493 images, and cherries, powdery mildew, and severe general category only 119 images respectively, in order to alleviate class imbalance problems affect the performance of model training, this paper adopted the rotate, flip and normalization method to preprocess the original data set. The pre-processed image is shown in figure 6.

The horizontal flip uses the vertical axis of the image as the inversion line to flip the image left and right, the image obtained after horizontal flip is shown in (a); the vertical flip uses the horizontal axis of the image as the inversion line to flip the image up and down, the image after flipping, up and down, is
shown in (b); the image centered on the center of the image is rotated clockwise and then rotated a certain angle, and then the rotated image is shown in (c).

![Image](attachment:image1.png)

(a) Apple healthy (b) Apple_Scab general (c) Apple_Scab serious

![Image](attachment:image2.png)

(d) Corn healthy (e)Corn rust is average (f) Corn rust is serious

Figure 5. Part of the crop disease data set image example

![Image](attachment:image3.png)

(a) Horizontal flip (b) Flip vertical (c) Rotating graph

Figure 6. Data preprocessing

4.2. Experimental environment
The experiments were all carried out by the deep learning framework PyTorch. The computer operating system was Linux, and the CPU was Intel(R) Xeon Gold 5118 CPU @2.30GHz. The GPU model is GeForce RTX 2080 Ti 11GB.

4.3. Experimental design and result analysis
In the experiment, the Spatial attention mechanism, ECA, and S_ECA proposed in this paper will be used to compare the accuracy of the ResNet50 convolutional neural network model on the dataset of crop diseases and insect pests in order to prove the effectiveness and superiority of the proposed E_SCA. Secondly, we can set up an experiment that first passes ECA and then connects spatial attention in series to compare with S_ECA. In this case, because the S_ECA module first extracts useful information for classification through spatial attention, it avoids The interference of useless information on the channel interaction can more accurately locate and select the diseased area in the blade, making the network performance after adopting the S_ECA module more stable and higher accuracy, thus verifying the rationality of the module setting.

In this experiment, a total of 50 epochs were set, the adam optimizer was used, the initial learning rate was set as 0.0001, and step attenuation was adopted. After every ten epochs, the learning rate was reduced to 1/10 of the original one, and the batch size was set as 32. The accuracy rate is used to evaluate the performance of the model, and the accuracy index is defined as follows:
where:

\[ N_T \] — the number of validation samples that are predicted to be correct

\[ N \] — the total number of validation samples

The comparison of accuracy after each attention module was added into ResNet50 is shown in figure 7. It can be seen from the figure that when different attention modules are injected into the neural network ResNet50, the network accuracy is improved by 0.78% compared with the original network after the S_ECA module provided in this paper is used. Its network stability and classification accuracy are also better than other attention modules. The experimental results are shown in table 1.

![Figure 7. Accuracy for embedding each attention module in ResNet50](image)

**Table 1. Results of the attention module experiments in ResNet50**

| Model                  | Acc     |
|------------------------|---------|
| ResNet                 | 86.50%  |
| ResNet + Spital        | 86.65%  |
| ResNet + ECA           | 86.98%  |
| ResNet + ECA + Spital  | 86.83%  |
| ResNet + S_ECA         | 87.28%  |

5. **Conclusion**

In this paper, we construct a hybrid attention mechanism that connects spatial attention mechanism and ECA in series and combines the model of ResNet50 convolutional neural network with it to make the network have good feature extraction ability. In the process of classification of pests and diseases, good experimental results have also been achieved. Compared with the original network of ResNet50, the accuracy has been improved by 0.78%, which also proves the feasibility of the ResNet50 model based
on the improved mixed attention mechanism. However, since the experiment requires a long time to train on a GPU device, the network model and training method can be improved in the future, so as to improve the training efficiency and performance of the network.

References
[1] Chen C, et al. (2019) Review on Development of Convolutional Network and Application in Computer Vision. Journal of Computer Science, 46: 63-73.
[2] Cao Lp. (2015) The Research Progress on Machine Recognition of Plant Diseases and Insect Pests. Chinese Agricultural Science Bulletin, 31:244-249.
[3] Qin F, et al. (2017) Image recognition of four different alfalfa leaf diseases based on deep learning and support vector machine. Journal of China Agricultural University, 22:123-133.
[4] Fujita E, et al. (2016) 2016 15th IEEE International Conference on Machine Learning and Applications. Frontiers in Plant Science, ICMLA 2016, p989-992.
[5] Jia Sp, Gao Hj, et al. (2019) Research Progress on Image Recognition Technology of Crop Pests and Diseases Based on Deep Learning. Transactions of the Chinese Society for Agricultural Machinery, 50: 313-317.
[6] He Km, et al. (2016) Deep Residual Learning for Image Recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. Las Vegas. 770-778.
[7] Xian Z, et al. (2019) Squeeze-and-Excitation Wide Residual Networks in Image Classification. In: 2019 IEEE International Conference on Image Processing (ICIP). Wuhan. pp. 395-399.
[8] Wang Qi, et al. (2020) ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 11531-11539.
[9] Wang Mh, Wu Zx, et al. (2021) Research on Fine-grained Identification of Crop Pests and Diseases Based on Improved CBAM via Attention. Journal of Transactions of the Chinese Society for Agricultural Machinery, 52:239-247.