

PAPER • OPEN ACCESS

Experimental study on crop disease detection based on deep learning

To cite this article: Xiaoling Xia *et al* 2019 *IOP Conf. Ser.: Mater. Sci. Eng.* **569** 052034

View the [article online](#) for updates and enhancements.

Experimental study on crop disease detection based on deep learning

Xiaoling Xia¹, Yongbo Wu^{1*}, Qinyang Lu¹, Changqi Fan²

¹School of Computer Science and Technology, Donghua University, Shanghai 201620, China

²i-soft Co., Ltd.

2181826@mail.dhu.edu.cn

Abstract: At present, deep learning has been widely used in our daily life, and the recognition of crop disease degree based on deep learning has gradually entered the public's vision. The degree of crop diseases can usually be judged by the characteristics of leaf colour, shape and spot. According to the data set provided by AI Challenge Competition, the degree of crop diseases can be divided into many situations. According to the current open source data, the recognition accuracy rate on this data set has not been able to exceed 90%. In order to achieve better recognition results, on the basis of the convolutional neural network CNN, the transfer learning method is used to load the residual network ResNet pre-training model and SeNet + ResNet to carry out deeper network training to improve the efficiency and accuracy of recognition. This paper carries out many experiments under the condition of controlling variables. The experimental results show that Resnet can play a good role in this data set with training parameters and achieve 83% accuracy, while SeNet + ResNet can achieve 90% accuracy without using pre-training parameters, and 87% accuracy of test set recognition. It can be seen that the SeNet + ResNet method has better effects in image classification and recognition tasks in this data set.

1. Introduction:

Image classification and recognition is a very popular research direction. For example, face recognition, which has been studied for many years but is still a research hotspot at present, the scenes for image classification and recognition applications are very rich, and many of the network models whose performance is constantly refreshing records are based on ImageNet. The effect of identification and classification on this data set has become an important indicator of the performance of a network model, such as VGG16, AlexNet, GoogLeNet, ResNet and so on. Especially since the advent of ResNet, the research of deep learning has been raised to a new level again, which makes the efficiency and effect of recognition continuously improve.

This paper uses the data set of crop detection in last year's AI Challenger competition. This data set has 10 species, 27 disease types, 61 categories, a total of nearly 50,000 clear pictures, of which more than 30,000 training sets, moderate size. Simple model invocation can not achieve a satisfactory result for such data sets with large amount of data, numerous classifications and little discrimination, which has been verified in the experimental part of this paper. Especially for those pictures which are somewhat blurred and the features are not obvious enough, often make mistakes in recognition, and even people can't distinguish themselves from each other.



In this data set, there are also some dirty data. There are one or more copies of the same picture. These pictures are redundant. They will not help the results but also affect the accuracy of recognition. Therefore, the first step is to eliminate those redundant pictures to reduce the workload of training. Secondly, the image is preprocessed by some traditional image processing methods, which improves the recognition effect and efficiency. In addition, the pre-training model is added to the experiment process, and the model is adjusted to suit the needs of the experiment.

The model used in this paper is the combination of ResNet and SeNet. ResNet is the champion of the ImageNet Competition in 2015. It represents that it can achieve real deep learning and can easily reach hundreds of layers. It solves the problem of network degradation and over-fitting when the layers are deeper. The concept of identical mapping is added. The reason why the learning effect of deepening layers is worse in the past is that there are some layers that are useless for training, so it is necessary to automatically judge which layers are useless in the training process. The advantage of this is that even if you don't need a deep layer, you don't need to think about setting it up. You can let the network decide which layers you need based on the training situation. SeNet is a process to strengthen the training results of the model, mainly to strengthen the important features, so that these important parameters become more obvious, and improve the accuracy of the output later.

2. Relevant work:

Convolutional Neural Network[1]: CNN is an indispensable feedforward neural network in deep learning at present. For example, in image processing, because the colour and pixels in a local area of a picture are comparatively similar, if every pixel is added to the parameters, it will increase the calculation. But CNN extracts the pixels in the same area uniformly and convolutes them. The operation of the convolution kernel can reduce the size of the original larger image, such as a 5×5 area into a 3×3 convolution kernel operation will become a 3×3 area. Convolutional neural networks are usually used in conjunction with the pooling layer to achieve the best results. Of course, there are many excellent neural networks such as RNN[8], FCN[9] and R-CNN.

AlexNet: AlexNet is the winner of the 2012 ImageNet Competition, which has been proposed as the pioneer of many deeper network models like VGG and GoogLeNet. AlexNet has eight layers, which are composed of five convolution layers and three full connection layers. After the convolution layer, it will be operated by norm layer and pooling layer, and of course, it will be corrected through ReLU immediately after the convolution layer. The last three layers are mainly for the classification, using the fully connected layer. Although the first two fully connected layers will use ReLU activation function as before, they are different from the previous convolution layers in that they will not go through the pooling layer operation, but will drop out to prevent over-fitting. Of course, the last layer is simply the output of the results, and this step is also predicted. [4]

VGG: VGGNet is one of the current mainstream network models with excellent performance and won the second place in the ILSVRC competition that year, it is named after a team of Oxford University, namely Visual Geometry Group. VGG is divided into two versions, named according to the depth of its. There are 16 layers of VGG 16 and 19 layers of VGG 19. The most well-known is the VGG16 network. VGG16 uses ReLU activation function after each convolution through stacked convolution layer. After two or three convolutions, it carries out a pooling layer operation. The final three fully connected layer is the same as the AlexNet network. VGG network has the characteristics of simple model structure and high requirement for training hardware because of its many parameters. But it is undeniable that it is still a very excellent network model. [5]

GoogLeNet: It and the VGG network mentioned before almost simultaneously appear in the public's vision, are the best in the ILSVRC competition in 2014, both of which were called "double heroes". GoogLeNet proposes a new neural network structure Inception, which is different from the previous VGG network in that although it has deeper network layers than VGG, its parameters are much less than the former and its performance is better than the former. All its convolutions are also corrected by ReLU. Moreover, it is no longer the use of fully connected layer for direct operation, but combined with the

characteristics of sparse connection and dense matrix. Inception currently has four versions, of which the most widely used version is Inception V3.[6][11][12][13]

3. Method:

3.1. ResNet + SeNet

In the first part of the experiment, without using the net only using ResNet. In order to make the loss of convergence faster, it is a good way to load the parameters of the pre-training model by using the method of fine-tuning. Therefore, this paper loaded the parameters that ResNet trains on ImageNet to fine-tune, and finds that the experiment is the first. After the round of training, it achieved very good results, which can be seen in the experimental part. Residual network determines which layers are identical mappings by fitting identical mapping functions, i.e. $H(x)=x$. $F(x)=H(x)-x$ is added here. It only needs to fit $F(x)$. When $F(x)=0$, it is identical mapping.[2]

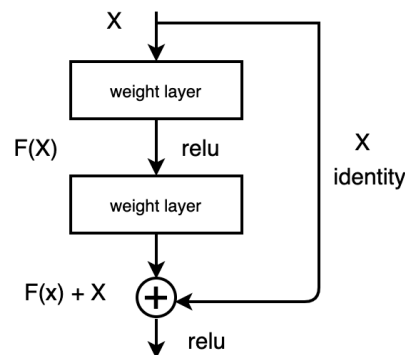


Figure 1. Residual learning: a building block

SeNet can actually be paired with any network model. It is connected behind the output layer of the model, mainly to adjust the output results, that is, the enhancement of important features and the weakening of non-critical features. SeNet is mainly divided into two steps. The output of the ResNet model training has two directions. It is not a direct output, but it is necessary to retrain these features through two layers of fully connected layers. The focus is on the above steps of sq and ex. First, the output features are compressed, and the features of $H \times W \times C$ in the original U are compressed into a feature sequence of $1 \times 1 \times C$, so that the original two-dimensional matrix in each dimension becomes an eigenvalue, which has a global receptive field and can represent the weight of the information of this dimension. These weighting factors are then activated by a second layer of the entire connection layer (Excitation), and finally these weights are appended to each of the U channels to obtain an output.[3]

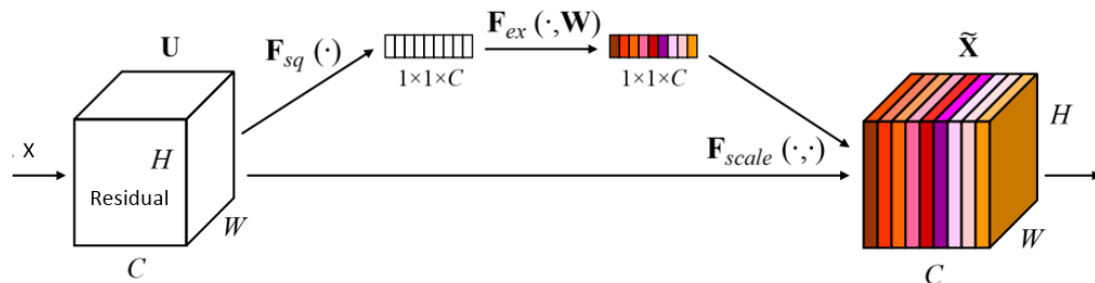


Figure 2. A ResNet + Squeeze-and-Excitation block

3.2. Loss Function

In the experiment, it is very important to choose an appropriate loss function, which is related to the convergence speed of the whole training model and the final experimental results. This paper uses categorical_crossentropy (cross-entropy loss function), which is defined as follows:

$$y = -[y \log \hat{y} + (1 - y) \log (1 - \hat{y})] \quad (1)$$

3.3. Optimize Strategy

Adam, which is one of the best algorithms at present. On the learning rate, we choose the dynamic change mode, which can play a very satisfactory effect with the Adam optimizer.[7] This is better than the gradient descent method. [10]

$$w = w - \alpha \frac{V_{dw}^{corrected}}{\sqrt{S_w^{corrected} + \varepsilon}}. (\varepsilon = 10^{-8}) \quad (2)$$

3.4. Image Processing

In the task of image classification, we often need to use some traditional image processing methods, such as image rotation and flip, image sharpening, histogram equalization, gray scale change, image enhancement and corrosion expansion in morphology. In this experiment, because the most needed image features are colour, shape and so on, only rotation, flip, brightness enhancement, sharpening can be performed.

3.5. Learning Rate

The setting of the learning rate is related to the convergence speed of the model. The final experimental results have been made. If the setting is not correct, the phenomenon of under-fitting or over-fitting may occur. This is fully considered in this paper, so in the experiment The dynamic learning rate method was chosen to make the learning rate gradually decrease with the training process. The LearningRateScheduler method was selected in the experiment and added to the callbacks function. Through a custom learning rate change law, the learning rate is reduced to 100 times the start after the training.

3.6. Parameter adjustment

As shown in Table 2, first perform a Squeeze operation on the feature map to prepare for the next Excitation operation, which is mainly to extract data. This process uses global pooling, which transforms the feature map into a $1 \times 1 \times C$ tensor to collect information for each channel. Then in the fully connected layer of Excitation, the number of channels can be controlled to achieve the effect of reducing the model parameters. In this paper, we first choose to reduce the number of channels by 16 times in the first layer FC (refer to $r=16$), and then restore the channel in the second layer FC. This can not only retain the necessary multi-channel features, but also play a role in dimensionality reduction. Then the paper sets $r=64$, and the experimental results show that the number of parameters is significantly reduced, reaching more than one million. At 23 epochs, the effect of $r=16$ has been achieved, and the training time is significantly reduced.

4. Experiment:

This experiment is divided into two parts, one using ResNet for training, the other using ResNet + SeNet for training. When choosing ResNet model for training, we use the method of transfer learning to train ResNet parameters trained on ImageNet, and only need fine-tuning. However, when using SeNet + ResNet, there are no pre-training parameters, so the training parameters need to start from scratch can not only be fine-tuned, but also need to be constantly tried in the experimental process.

Because the data set used has some dirty data, such as some duplicate files, the first thing is to filter out these duplicate files. After statistics, it is found that there are only 1-2 images labeled 44 and 45 in the whole data set. It is necessary to shuffle the JSON file after reading the information, in order to ensure randomness.[14]

The rule of rotation is to rotate according to the center of the picture, and of course to flip, and the direction of each rotation or flip is random. This can be controlled by generating a random number at a time. If it is 0, then rotating according to the x-axis means 180 degrees in the horizontal direction and 180 degrees in the vertical direction.[15]

Secondly, sharpening can make the image contour clearer. In this step, the image convolution operation in OpenCV is used. In addition, we need to enhance the brightness of the pictures. The method in OpenCV is still used here. The size of the picture is set at $224 * 224$, because most of the original size of the picture is basically medium size, so this setting will not lead to a poor quality of the picture. A previous experiment that doubled the size of a larger image ($1080 * 1920$) has reduced the quality of the image considerably, of course, the final result is not very satisfactory.

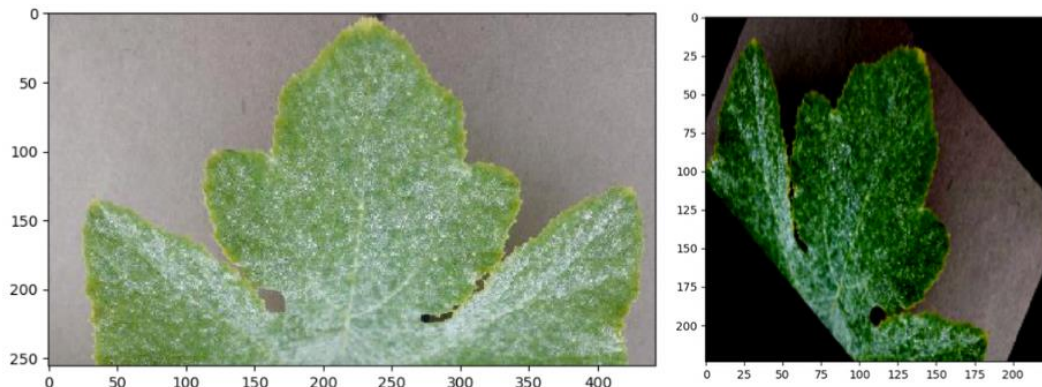


Figure 3. The left one is the original one and the right one is the preprocessing one

In terms of setting parameters, the setting of parameters directly affects the final training results. Improper parameter setting will cause the entire network model to fail completely. This has been verified in my first experiment, because the learning rate is not set to a larger 0.01, the optimizer chooses Momentum, BATCHSIZE is also set to a larger, resulting in serious over-fitting. So in this experiment, we choose the learning rate of 0.001, Adam optimizer, and adopt the method of dynamic learning rate. The learning rate decreases gradually according to certain rules. At the end of the training, the learning rate will be reduced to about 100 times of the original.

Table1. Model comparison table

Model	val_acc	val_loss	val_precision
ResNet without Image preprocessing	0.8246	0.4525	0.76
SeNet_ResNet without Image preprocessing	0.8132	0.4834	0.76
SeNet_ResNet with Image preprocessing	0.8748	0.3366	0.81

Table1 shows that SeNet and ResNet can achieve better results by comparing the data in the table. After 50 epoch of experiments, accuracy on the training set reaches 0.90, and accuracy on the verification set reaches 0.87. Although ResNet seems to perform better without image preprocessing, but SeNet + ResNet is without training parameters.

Table 2. Learning rate comparison table

Learning Rate;	val_loss	val_acc
0.001	0.38	0.8635
Dynamic learning rate	0.33	0.8748

Table 2 shows that the dynamic change of learning rate can converge better and achieve the best effect by using the above learning rate decreasing with the increase of the number of epoch.

5. Conclusion:

In this paper, we use the crop disease detection data set in the AI Challenge to detect crop diseases. From the data set, we can see that it is not an easy task because of the variety. Through the experiments in this paper, we can find a better method:

Use the ResNet + SeNet model for training and testing. From the experimental data, we can see that the effect of the whole task becomes more satisfactory after adding SeNet + ResNet model. At the same time, the image preprocessing process is also crucial to the improvement of the entire task. Therefore, we can conclude that SeNet has a very significant role in improving the classification task, and the preprocessed images are more suitable for the task.

References

- [1] Zeiler M D, Fergus R. Visualizing and understanding convolutional networks[C]//European conference on computer vision. Springer, Cham, 2014: 818-833.
- [2] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.
- [3] Hu J, Shen L, Albanie S, et al. Squeeze-and-Excitation Networks[J]. 2017.
- [4] Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton. ImageNet Classification with Deep Convolutional Neural Networks
- [5] Karen Simonyan, Andrew Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition
- [6] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 2818-2826.
- [7] Kingma D P, Ba J. Adam: A method for stochastic optimization[J]. arXiv preprint arXiv:1412.6980, 2014.
- [8] Zaremba W, Sutskever I, Vinyals O. Recurrent neural network regularization[J]. arXiv preprint arXiv:1409.1, 2014.
- [9] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 3431-3440.
- [10] Loshchilov I, Hutter F. Sgdr: Stochastic gradient descent with warm restarts[J]. arXiv preprint arXiv:1608.03983, 2016.
- [11] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 1-9.
- [12] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[J]. arXiv preprint arXiv:1502.03167, 2015.
- [13] Szegedy C, Ioffe S, Vanhoucke V, et al. Inception-v4, inception-resnet and the impact of residual connections on learning[C]//AAAI. 2017, 4: 12.
- [14] Zhang Xueqin, Chen Jiahao, Zhuge Jingjing, Yu Lijun. Rapid Plant Image Recognition Based on Deep Learning [J]. Journal of East China University of Technology (Natural Science Edition), 2018, 44 (06): 887-895.
- [15] Hu Qiuxia. Research on identification method of plant leaf diseases based on image analysis [D]. Northwest University of Agriculture and Forestry, 2013.