

Cassava Leaf Disease Classification Based On CNN

Boyang Ding*

School of Mathematics, Sun Yat-sen University, Guangzhou 510275, China

*Corresponding author email: dingby@mail2.sysu.edu.cn

Abstract. Cassava, a key food security crop in Africa, can provide a large amount of carbohydrates, which is widely grown in Africa. This kind of crop has been disturbed by viral diseases for a long time, which is the main reason for low production. The symptoms of cassava diseases can be reflected in their leaves but artificial visual diagnosis of diseases is inefficient. In order to get rid of this inefficient method, it is important to establish a model to automatically diagnose cassava by using the captured crop images to help farmers save their crops. Convolutional neural network has a good performance in image classification. Based on the research of Kaiming He et al. on residual convolution network, this paper trains the image dataset of cassava leaves from Kaggle by residual networks and then compares the classification effects by using different layers. Moreover, data augmentation and Focalloss function are used to optimize the model, which achieve better performance.

Keywords: Convolutional Neural Network, Resnet, Image Classification, Data Augmentation.

1. Introduction

F.Rosenblatt proposed the first artificial neural network that can be applied to practice in 1957-the perceptron[1], however, its single-layer structure determines the upper limit of its expression ability. Due to the limitation of technology, the development of neural network has been staying in the stage of one hidden layer neural network. Later, Back Propagation (BP) algorithm [2], one of the most important algorithms in the history of artificial intelligence proposed by D.E.Rumelhart in 1986, provided a practical path for training multi-layer networks, especially for training the weights of hidden layers, making it possible to build a deep network. Deep network can learn the characteristics of dataset, and has great development space in classification, visualization and other fields [3]. Convolution neural network (CNN) has a great effect on image processing, such as image classification, semantic segmentation and image captioning. The proposal of convolutional neural network benefited from the study of animal visual cortex. Scholars imitated the neural structure of visual cortex and proposed many neural network architectures for image recognition, and gradually evolved into the current convolutional network.

In 1998, LeCun proposed the LeNet-5[4] and introduced the concepts of convolution layer and pooling layer firstly, which is the most important architectural foundation in convolution neural networks. Later, AlexNet [5] was proposed in 2012, introducing the activate function of ReLU, which made the training faster and solved the gradient disappearance to a certain extent. By using data augmentation and a method called "dropout", it reduced overfitting. In 2014, GoogLeNet [6] introduced some new ideas such as Global average pooling (GAP) [7] and Inception Module [9], which reduced the amount of calculation and parameters and got better classification performance. In 2015, a milestone neural network called ResNet [8] is proposed by Kaming He et al. At the same year, it won the Imagenet competition, a famous dataset often used to test the accuracy of network. Beneficial from its idea of residual learning and structure, Resnet effectively solves the degradation problem and train the datasets in a high-speed, which made networks can be deeper than before.

2. Methodology

2.1. CNN

In the 1960s, Hubel et al. proposed the concept of receptive field since they found that some cells only responded to signals located in a small area in the visual cortex of cats [9]. They noticed that

some neurons respond to more complex local patterns, which can be viewed as combinations of lower modes. It is easy to imagine that these neurons receive signals from a range of neighboring lower-level neurons.

The structure of the convolution layer is inspired by the receptive field. Unlike the traditional networks that each layer of neurons does not fully connect to the lower layer of neurons, it receives only a small range of pixel signals. Each convolution layer has three important parameters: the height, width, and step of the field. The height and width can determine the size of the receptive field. The step size determines the degree of the overlap between the receptive fields of each neuron.

The purpose of the pooling layer is to down-sampling, which decreases the pixels of images and makes the parameters less. The structure of the pooling layer is similar to that of the convolution layer. Like the convolution layer, each neuron in the pooling layer also has a receptive field. The difference is that the pooling layer has no weight parameters. It simply processes the input matrix, such as taking the average value (average pooling layer) or the maximum value (maximum pooling layer). Convolution layer and pooling layer are arranged alternately, which extracts the feature maps and reduces the pixels of the images.

After flowing through convolution layer and pooling layer, the image data will enter a full connection layer at the end of the structure, which connects the neurons in previous layer and reduces the dimension of the data.

2.2. Residual Neural Network

Theoretically, the deeper the convolutional neural network is, the stronger the learning ability should be. The effect of the model should be better when it can learn more features. Most of the time, convolutional networks do benefit from layer deepening. However, in some scenarios, deepening the number of layers will not improve the accuracy of the model, but will lead to the poor effect of the model. Figure 1 shows this degradation phenomenon.

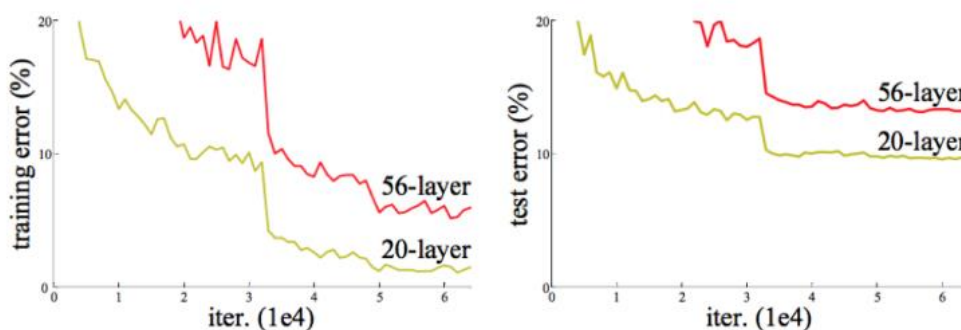


Figure 1. Degradation of deep networks [8]

There are many reasons may result in model degradation. One of most important reasons is gradient disappearance / explosion. Kaiming He et al. came up with a concept of Residual learning. The author pointed out that Deep network degradation is still a common phenomenon. The accuracy of deep network tends to be saturated but the final effect is not as good as that of common neural network.

The author put forward a solution: assuming that the target mapping of network learning is $F(x)$, and then the next step is not to let the network directly learn $F(x)$ itself, but to learn its residual $F(x) - x$, which is residual learning. Experiments showed that the problem of degradation in neural network has been greatly solved when using residual learning.

Kaiming He et al. also showed the structure of Resnet. Resnet is improved on the basis of VGG [10] and effectively solves the degradation problem. In 2017, Kaiming he et al. published an article, Aggregated Residential Transformations for Deep Neural Networks, which proposed Resnext, integrated the characteristics of Resnet and inception, and has achieved the very good results.

2.3. Optimization Strategies

2.3.1 Data augmentation

Data augmentation is a kind of method that generates a large number of new data from existing training examples and then increases the size of training sets.

As for image classification, we can use the original training image to randomly flip, rotate, adjust the brightness and contrast, and generate some new images to add to the training set. Figure 2 shows the augmentation process.



Figure 2. Data Augmentation

At first, data augmentation was only used to extend the dataset when there were insufficient training samples, but later it gradually used as a method to prevent over-fitting.

2.3.2 Under-sampling

An unbalance distribution in the number of samples exists in the classification task. There are too many samples belonging to one category, so the final effect of the model shows a deviation to the large category.

There are some methods to effectively solve the problem of unbalanced sample. The first method is under-sampling. We randomly remove some samples in the largest class to keep the balance of the number of samples. However, the disadvantage is that some important information might be ignored.

2.3.3 Focalloss

Generally, cross entropy is used as the loss function. Formula 1 is the crossentropy function.

$$L = -\frac{1}{N} \sum_{i=1}^N \log(p_t) \quad (1)$$

Where N represents the size of samples, p_t represents the probability when sample belonging to the correct category.

By using the reciprocal of the sample number proportion as the sample weight coefficient in the cross entropy formula, we can increase the weight of small categories in the loss function. Formula 2 is the function.

$$L = -\frac{1}{N} \sum_{i=1}^N \alpha_t \log(p_t) \quad (2)$$

Where α_t represents the reciprocal of the proportion when sample belonging to the correct category.

Based on cross entropy loss, Kaiming He et al. came up with another method called Focalloss [11], which reduces the weight of the well-classified samples to make the model concentrate more on hard samples while training. Formula 3 is the function.

$$L = -\frac{1}{N} \sum_{i=1}^N \alpha_t (1 - p_t)^\gamma \log(p_t) \quad (3)$$

Where τ is a super parameter, which determines the skew degree of the weights of difficult and easy samples.

3. Results

3.1. Data Description

The data of this study comes from Kaggle, a famous data analysis platform. This dataset contains more than 20000 plant pictures (800×600), and five categories of cassava leaves. We want to establish a model to identify these five categories. They are Cassava Green Mottle (CGM), Cassava Brown Streak Disease (CBSD), Cassava Mosaic Disease (CMD), Cassava Bacterial Blight (CBB) respectively and one health category. The label of each category is as follows. Figure 3 shows some samples of the images.



Figure 3. Some images of cassava leaves

Figure 4 shows the proportion of each category.

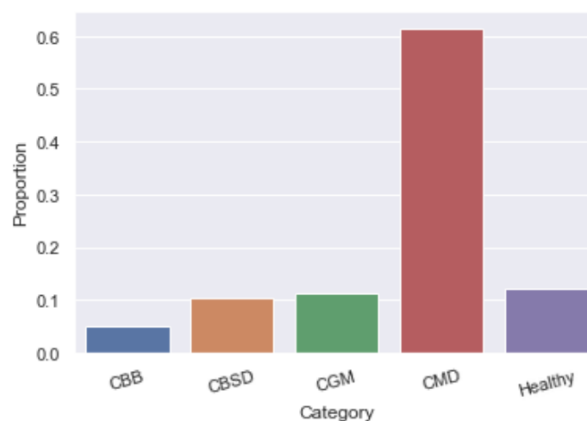


Figure 4. The proportion of each category

It can be seen that the disease category CMD accounts for about 60% of the sample, which means that it's difficult for the model to learn the characteristics from other small-size categories. We must consider the impact of large category deviation.

3.2. Analysis of experimental results

In this section, multiple groups of experiments will be conducted on the data set to verify the effects of each model and optimization strategy.

3.2.1 The Effect of Different Networks

First, Resnet34, Resnet50, Resnet101 and Resnext50 are set up respectively for training. In this part, we use Focalloss and data augmentation. The effect of these two optimization strategies will be compared in next part. The results are shown in Table 1.

Table.1. Comparison of different networks

Model	Train Error	Test Error	Time
Resnet34	39.8%	39.9%	39min
Resnet50	25.4%	27.3%	103min
Resnet101	28.8%	30.1%	132min
Resnext50	24.2%	24.3%	102min

Training process is shown in Figure 5.

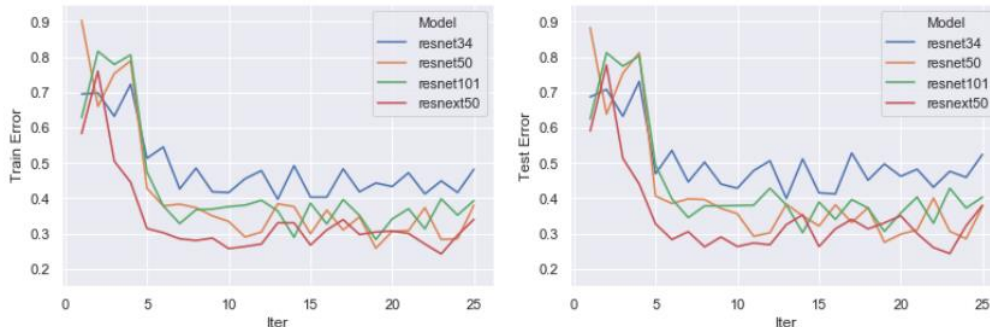


Figure 5. Train error(left) and Test error(right)

From the results, we can see that Resnet34 performs much worse than others. Resnet101 and Resnet50 showed some degradation, the accuracy in training set and test set even decreased by about 3% when time goes by. Resnext50 has the better performance.

3.2.2 The Effect of Different Strategies

The results based on Resnext50 are presented in Table 2 as follows.

Table.2. The effect of different strategies

Model	Strategy	Train Error	Test Error
Resnext50	under-sampling + Aug	35.8%	40.0%
Resnext50	FocalLoss + Aug	24.2%	24.3%
Resnext50	No solution + Aug	33.8%	32.3%
Resnext50	FocalLoss + no Aug	4.9%	33.7%

Training process is shown in Figure 6.

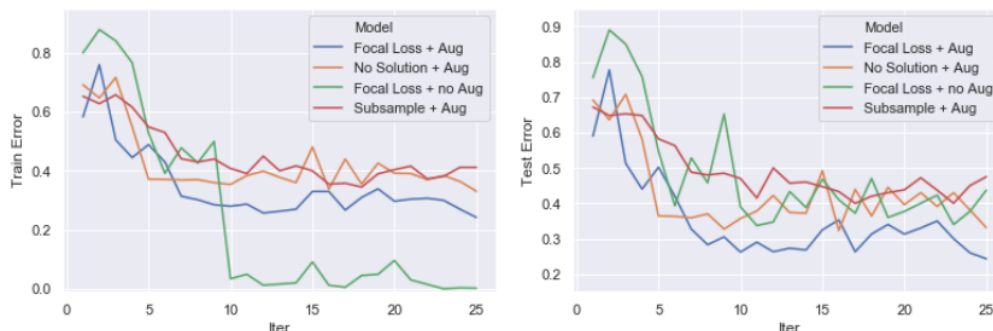


Figure 6. Train error(left) and Test error(right)

As we can see from the line chart, in the group without augmentation, the training error decreased to a very low level, but the test error did not decrease with it. This is a typical over-fitting phenomenon, and it does not appear in other groups. Therefore, it is concluded that data augmentation can prevent over-fitting effectively. Moreover, the model using Focalloss is significantly better than the model

using under-sampling. This may be because under-sampling needs to discard samples, and Focalloss uses more information. Therefore, Focalloss can effectively solve the problem of unbalanced sample.

4. Improvement

To increase the accuracy of classification, we can use other improved models. EfficientNets [12] balances network depth, width, and resolution, which can achieve much better accuracy and efficiency than the previous CNN. DenseNets [13] can reuse the feature and strengthen the propagation of feature, which makes the parameters less. V. V. Srinidhi et al. has used EfficientNetB7 and DenseNet on apple leaves disease detection and the accuracy reached at 99.8% and 99.75% respectively [14].

5. Conclusions

This study lists the development of artificial neural network and convolutional neural network. The deep residual network proposed by Kaiming He effectively solves the degradation problem. In order to get rid of over-fitting phenomenon and get better performance, some strategies which can optimize the image sample datasets are used, such as under-sampling, data augmentation, and Focalloss function. In the part of the experiment, by evaluating the performance of CNN model with different layers, it is shown that Resnext50 model has the highest accuracy of 75.8% among them. Moreover, by doing control experiment, both data augmentation and Focalloss can increase the accuracy of training and test set, which also indicated that the problem of over-fitting can be effectively solved by doing so. The setting of super parameters, such as batch size and learning rate, which are crucial important during the training, might be the reason for why the accuracy is not high. In fact, there was someone who use Resnext50 to classify the cassava leaves had reached the accuracy of 87%.

References

- [1] Rosenblatt F. The perception: a probabilistic model for information storage and organization in the brain[J]. Psychological review, 1958, 65(6): 386-408.
- [2] Rumelhart D E, Hinton G E, Williams R J. Learning representations by back-propagating errors[J]. Nature, 1986, 323: 533-536
- [3] HINTON G E, SALAKHUTDINOV R. Reducing the Dimensionality of Data with Neural Networks[J]. Science, 2006, 313: 504-507.
- [4] Lecun Y, Bottou L. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11):227-2324.
- [5] Krizhevsky A, Sutskever I, Hinton G. ImageNet Classification with Deep Convolutional Neural Networks[J]. Communications of the ACM, 2012, 60: 84-90.
- [6] Szegedy C, Liu W, Jia Y, et al. Going Deeper with Convolutions[J]. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015: 1-9.
- [7] Zhou B, Khosla A, Lapedriza A, et al. Learning Deep Features for Discriminative Localization[J]. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016: 2921-2929.
- [8] He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016: 770-778.
- [9] A Géron. Hands-on machine learning with Scikit-Learn and TensorFlow: concepts, tools, and techniques to build intelligent systems [C]. O'Reilly Media, 2017.
- [10] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition[J]. ArXiv, 2015, abs/1409.1556.
- [11] Lin T Y, Goyal P, Girshick R, et al. Focal Loss for Dense Object Detection[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2020, 42:318-327.

- [12] Tan M, Le Q V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks[J]. ArXiv, 2019, abs/1905.11946.
- [13] Huang G, Liu Z, Laurens V, et al. Densely Connected Convolutional Networks[C].2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 2261-2269.
- [14] Srinidhi V V, Sahay A, Deeba K. Plant Pathology Disease Detection in Apple Leaves Using Deep Convolutional Neural Networks: Apple Leaves Disease Detection using EfficientNet and DenseNet[C]. 2021 5th International Conference on Computing Methodologies and Communication (ICCMC). 2021,pp. 1119-1127.