Fruit Classification based on ResNet and Attention Mechanism

Miaorun Lin *

Department of Telecommunications Engineering and Management, Beijing University of Posts and Telecommunications, Beijing, China

* Corresponding Author Email: 2019213241@bupt.edu.cn

Abstract. Numerous types of fruits are discovered, which can provide humans with nutrients and trace elements. There are 59 families and 694 species of fruits in China, of which more than 300 kinds of cultivated fruit trees are in abundance, including more than 10000 varieties. Therefore, the classification of fruit detection is very necessary. The traditional method of fruit classification mainly relies on manual methods, which has greatly reduced the cost effectiveness in recent years due to the increase of labor cost. This paper introduces the idea of transfer learning by comparing various automated methods for detecting and classifying fruits, and incorporates the attention mechanism module into the ResNet model to achieve recognition of image features that can effectively classify fruits. This paper seeks to provide a complete review on fruit classification concluding that the ResNet model with the introduction of the attention mechanism has good accuracy and verifying the efficiency of the model in the field of image recognition and classification.

Keywords: Fruit Classification, ResNet, Visual Attention, Residual Learning.

1. Introduction

In the process of agricultural production, the classification of products is a very important part. Manual methods of classification are labor-intensive and inefficient, and are not conducive to the standardization and systematic advancement of agricultural production [1]. Computer vision systems have been widely used in the agricultural industry and have proven to be a scientific and powerful tool, for example many machine vision algorithms are used in agricultural applications [2-4], for economical efficiency measurement evaluation and proper inspection. Image recognition technology, an important area of computer vision, has been rapidly developing in recent years to promote intelligent agricultural production. Image recognition techniques classify and recognize images by describing image features, and in the case of agricultural products differentiation by extracting fruit appearance features. Deep learning is a more efficient machine learning method that can perform large-scale and complex operations,[5] and its application in the field of image recognition can efficiently complete the processing of various images and effectively promote technological innovation in the field of computer vision. The image recognition technology based on residual learning can solve the degradation problem of deep learning, so as to achieve higher accuracy of image classification. Therefore, this paper proposes to use Resnet for residual learning and add an attention mechanism system, and use Fruit360 from kaggle as the dataset for training, so that fruit images can be classified accurately and efficiently.[6]

2. Method

2.1. Data processing

In the first step, Fruit360 was first divided into a training set (67,692 images) and a test set (22,688 images). In the second part, the dataset is observed and the fruits are classified using the features of mass, height, width and color. The classifications to be trained are copied from the original data folder, grouped numerically, shuffled, and the labels are converted into unique codes. Step 3, initialize the session and start training. Step 4, observe the results and test the accuracy.

2.2. ResNet

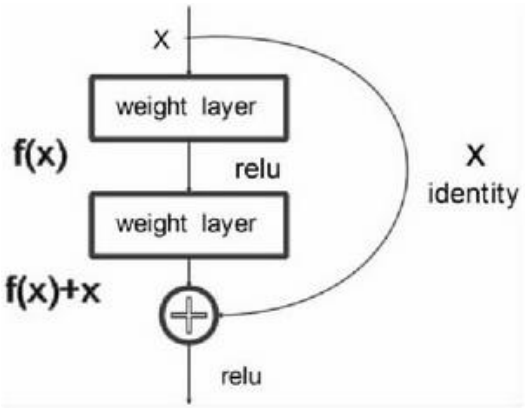


Figure 1. Framework of residual block from ResNet[7].

The traditional nonlinear representation of multilayer network structure is difficult to achieve the same output after stacking and stacking since before when deepening the number of layers, i.e., it is difficult to express constant mapping [7]. The problem of gradient explosion or disappearance occurs during the training process, while deep residual networks as a new network structure can reduce these some problems to some extent and perform well under some specific branching areas. The structure of residual learning is shown in Figure 1. In a stacked layer structure, the input is x, the learned feature is H(x), and the learned residual is F(x)=H(x)-x. When the residual is F(x)=0, the stacked layer can achieve a constant mapping. Although the residual will not actually be 0, this also allows the stacking layer to learn new features on top of the input, thus improving performance. The residual structure used in this study is shown in Figure 2. Such a residual structure can reduce the network parameters and operations, and is suitable for deep network construction.

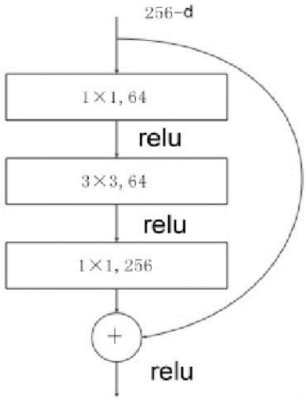


Figure 2. Framework of residual block used in our paper.

2.3. Implementation details

A 50-layer resnet model is constructed with the structure (The network model consists of a convolutional layer (Conv), a batch normalization layer (batch norm), a pooling layer (pool), a fully connected layer (FC), and finally a Softmax classifier to predict the classification results.): CONV2D -> BATCHNORM -> RELU -> MAXPOOL -> CONVBLOCK -> IDBLOCK*2 -> CONVBLOCK -> IDBLOCK*3-> CONVBLOCK -> IDBLOCK*5 -> CONVBLOCK -> IDBLOCK*2 -> AVGPOOL -> TOPLAYER

Because the dataset is huge, Relu is chosen as the activation function. It solves the problems of large computation, gradient disappearance, etc. [8] and increases the sparsity of the network.[9] The reLu activation is used for the last part of the main path and the result of the summation of the ri tensor of the cnn layer.

(1) Loss during training: the difference between the predicted value and the true value of the model.

(2) Accuracy: the accuracy of the test after each iteration in the training process.

(3) batch_size: the number of images used in one iteration of batch gradient descent.

(4) Learning rate(lr): the speed of updating neurons during backpropagation.

(5) Epoch: the process of passing a complete data through the neural network once and returning it once.

In the experiments, Epoch is set to 25. batch size is set to 8, learning rate (Ir) is set to 0.001, and the number of classifications is set to 3. The training starts from the pre-trained model. The loss and accuracy of the model training process are important indicators to judge the performance of the model, and this study judges the performance of the model by observing the loss and accuracy of the training process. The experiments were conducted using two different methods, Model A and Model B, to train them respectively. Model A is to add the attention mechanism module to the basic ResNet model; Model B is the ResNet base model.

Model A combines the convolutional neural model of ResNet50 with the attention mechanism for fruit classification. The feature extraction layer of ResNet50 is used as the backbone network, and the attention mechanism is added later to obtain the proposed feature extraction network model. This model is shown in Figure 3. The residual network structure can achieve fruit feature extraction by stacking multiple convolutional layers. The self-attention mechanism is used here, as shown in Figure 4. Since convolutional neural networks can only learn closer dependencies, and are less capable of learning long-range dependencies. However, in many real-life situations we need to construct long-range dependencies. The introduction of the self-attentive mechanism can introduce long-range dependencies, which can better solve this problem. The attention mechanism is used to improve the sensitivity to the key features, so that the network focuses on the useful key information and suppresses the useless information when performing feature extraction [10]. The addition of the attention mechanism module has a beneficial effect on the performance of classification tasks with small differences between fruit classes.

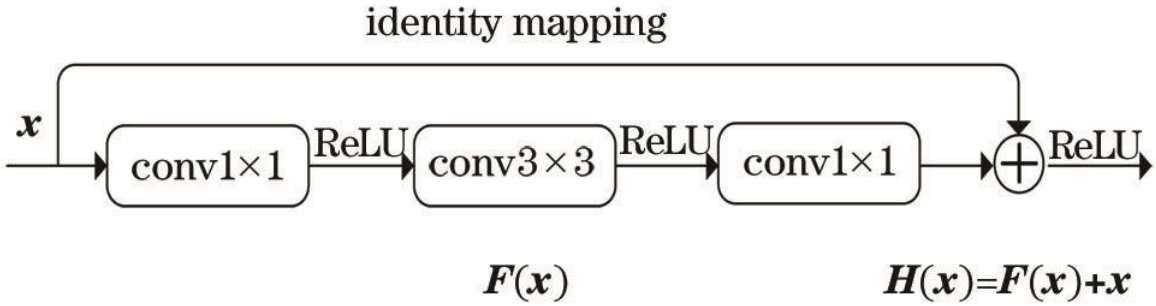


Figure 3. Residual block structure in ResNet50 Network

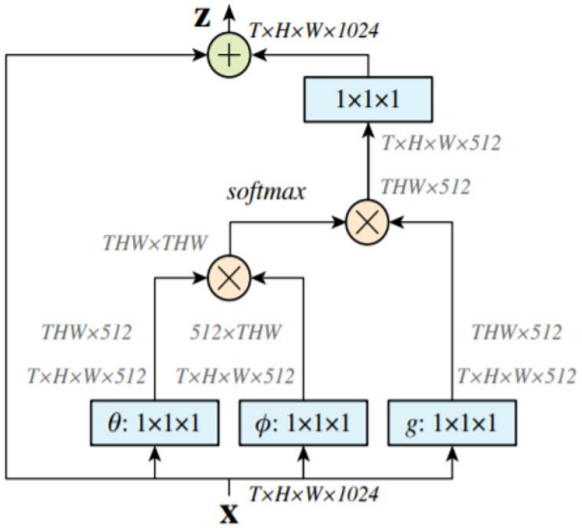


Figure 4. Non-local attention network.

3. Results

Table 2 described the learning curve of both two model. Compared with model B, model A with attention module has good fit ability with high training accuracy and high convergence speed. However, it spends more time for inference due to the increasing computation cost. Then, a comparison results on Fruit360 dataset are shown in Table 3. The author compares our models with other machine learning method, including KNN, gaussian naïve Bayes, Logistic regression, Decision tree, Linear discriminant analysis and SVM. Our model A achieves the best performance of 98% accuracy, which validate the effectiveness of ResNet with visual attention.

Table 1. Training statistics of both models.

	A Model		B Model	
Time	29min 7sec		14min 39sec	
Epoch	Training Accuracy	Test Accuracy	Training Accuracy	Test Accuracy
0	0.8634	0.9902	0.8315	0.9641
1	0.9295	0.9935	0.9218	0.9967
2	0.9416	0.9967	0.9339	0.9967
3	0.9703	1.0000	0.9548	0.9902
4	0.9515	1.0000	0.9438	0.9967
5	0.9714	1.0000	0.9515	0.9935
6	0.9846	1.0000	0.9581	0.9967
7	0.9725	1.0000	0.9659	0.9967
8	0.9857	1.0000	0.9626	0.9967
9	0.9824	1.0000	0.9581	0.9967
10	0.9879	1.0000	0.9526	0.9967
11	0.9890	1.0000	0.9526	0.9967
12	0.9934	1.0000	0.9581	0.9935
13	0.9857	1.0000	0.9637	0.9967
14	0.9857	1.0000	0.9581	0.9967
15	0.9890	1.0000	0.9471	0.9967
16	0,9890	1.0000	0.9659	0.9967
17	0.9846	1.0000	0.9559	0.9935
18	0.9835	1.0000	0.9526	0.9967
19	0.9857	1.0000	0.9504	0.9967
20	0.9934	1.0000	0.9416	0.9967
21	0.9857	1.0000	0.9681	0.9967
22	0.9879	1.0000	0.9626	0.9967
23	0.9879	1.0000	0.9659	0.9967
24	0.9835	1.0000	0.9504	0.9967

Table 2. Comparison results on Fruit360 dataset.

Method	Accuracy on training set	Accuracy on test
KNN	0.92	0.96
Gaussian naive Bayes	0.86	0.67
Logistic regression	0.70	0.41
Decision-making tree	0.98	0.73
Linear discriminant analysis	0.86	0.67
SVM	0.61	0.33
ResNet(A)	0.98	1.00
ResNet(B)	0.95	0.99

4. Conclusions

The KNN algorithm was the most accurate model tested among the six previously studied models. However, but in these datasets only a smaller dataset was used. As the network became deeper, the experiments revealed a degradation problem in the deep network, where the network accuracy saturated and even decreased as the network depth increased. So the author needs a better model to

solve this problem - Resnet. and achieved the best result :100% correct prediction rate. In this study, the fruit classification recognition method based on Resnet model, the pre-trained Resnet model is migrated to the experimental dataset for training, keeping the convolutional layer structure unchanged, replacing and training the fully connected layer, and adding the attention mechanism in between. The fruit recognition accuracy of the Resnet50 model using migration learning is significantly improved compared with other models. However, it can be seen from the experimental results that the training time to perform migration learning must be more time to perform deep learning training and the accuracy is higher. However, there is a problem that more time is spent and the accuracy improvement is not significant because of the sample limitation in the dataset, and subsequent research can be conducted by increasing the variety of fruits to train the model.

References

[1] He Zichao. Multi-task image classification based on deep learning and migration learning [D]. Guangzhou:South China University of Technology, 20177

[2] Gomes et al., “Applications of computer vision techniques in the agriculture and food industry: a review”, Eur. Food Res. Technol., 235 (6), 989–1000. 2012.

[3] Zhang et al., "Application of computer vision technology in agricultural field", Applied Mechanics and Materials, vol. 462. Trans Tech Publ, pp. 72–76, 2014.

[4] Vibhute et al., "Applications of image processing in agriculture: a survey", Int. J. Comput. Appl., 52 (2), 34– 40, 2012.

[5] Guo Hua. Deep learning and its significance [J]. Curriculum. Teaching materials. Teaching Methodology,2016(11):25-32.

[6] https://www.kaggle.com/datasets/moltean/fruits

[7] He K,Zhang XRen Set alDeep residual learningfor imagerecognition[C] Proceedings of the IEEE Conference on Computer Vision and PatternRecognition2016:770-778 Amjady N. Short-term hourly load forecasting using time series modeling with peak load estimation capability [J]. IEEE Transactions on Power Systems, 2001, 16(4): 798-805.

[8] Guo ZY,Shu X,Liu CHANGYAN,et al. Flower recognition algorithm based on convolutional neural network with ReLU function[J]. Computer Technology and Development,2018(5):154-157,163.

[9] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.

[10] Wang WG, Shen JB, Jia YD. A review of visual attention detection[J]. Journal of Software,2019(2):416-439.