

A CNN-based implementation of fruit recognition

Zixuan Huang *

McMaster University, 1280 Main St W, Hamilton, ON L8S 4L8, Canada

* Corresponding author: Huanz86@mcmaster.ca

Abstract. Image recognition technology is now widely used in various industries, and CNN has played an indispensable role in it over the past decade. The paper discusses the use of Convolutional Neural Networks (CNN) for fruit image recognition, aiming to verify the impact of different CNN designs on model training time, test accuracy, and test accuracy. The experiment uses data from the Kaggle Fruits 360 project and focuses on ten different categories of fruit. The CNN model is built using 3*3 convolutional kernels and features four combinations of convolutional and relu layers. The final test accuracy is recorded as 98.1714%. The paper also discusses potential reasons for lower-than-expected accuracy and attempts to address these issues, including overfitting, image resolution, and the simplicity of the training set. The impact of regularization and different image resolutions on model accuracy is observed. The paper concludes by highlighting the practicality of CNN in image recognition, but also acknowledges limitations such as training time, computational resources, and the abstract nature of extracted features. It also emphasizes the importance of choosing an appropriate training set for model accuracy and suggests that other AI models may offer solutions to the shortcomings of CNN. Overall, the paper provides insights and experiences for those working with CNN in image recognition and acknowledges the rapid development of artificial intelligence in recent years.

Keywords: CNN, Fruit classification, Recognition.

1. Introduction

Currently, information extraction from images is becoming more and more common, for example, car plate recognition, road recognition. There are different ways to achieve automation and liberating manpower, like edge detection algorithm, and template matching. One of the best ways for information feature extraction is using Artificial Intelligence. Artificial intelligence has undergone tremendous changes and achieved great development from its conception to the present. The incorporation of artificial intelligence (AI) into different industries such as healthcare, banking, and transportation has resulted in major improvements in efficiency, accuracy, and decision-making [1]. The performance of deep learning in image recognition has surpassed humans Compared with traditional AI, deep learning has the ability to automate the learning of feature sets for several tasks, enables learning and classification to be achieved in a single shot [2].

Convolutional neural network (CNN) is a special type of neural network in deep learning. It plays an important role in image recognition technology and has unique advantages. For example, in the medical field, (convolutional neural network) CNN has brought about breakthroughs in pattern recognition of images including medical images. In usual CAD algorithms, designing an image-feature extractor is important and difficult. On the other hand, a CAD algorithm by use of CNN does not necessarily require the image-feature extractor [3]. Since the concept of CNN was proposed 30 years ago, it continuously improved during use. At the beginning, the CNN is a three-layer continuous time neural network containing an input a state and an output layer. The densities of the layers are the same [4]. As different CNNs were implemented, it was discovered that the size and amounts of layers of CNNs can be designed differently according to different purposes and needs. In general, increasing the amount of convolution layers will improve training accuracy, but also increase the risk of overfitting which will decrease test accuracy. At the same time, time cost is also an issue that cannot be ignored.

In the above research background, this article aims to verify the impact of different CNN designs on model training time, test accuracy and test accuracy. With the development of social economy,

different types of fruits from all over the world have entered the human market. For most people, it is necessary to look at pictures to learn about the types of fruits, who is allergic to them, and how to eat them. Under this premise, fruit classification becomes increasingly important. By establishing and training a CNN model for fruit image recognition, the main idea of this article is to discover and try to solve the problems encountered in the recognition process. Through the research in this article, we will give some tips and experiences to those who try CNN later.

2. Data and Methods

2.1. Data and the Data Preprocessing

The entire experiment followed the Kaggle Fruits 360 project guide and used the data provided by the project [5]. The data are pictures of 131 different kinds of fruits, with a uniform background of pure white. The data is divided into training set and test set, which are used to train the model and test the accuracy of the model respectively. Using this data can well verify the influence of the data set for the actual model prediction accuracy.

The experiment didn't choose to use full training set provided by Fruit 360 Project, only pick 10 different categories of fruit: Apple Golden 2, Banana, Ginger Root, Mango, Onion Red, Peach, Potato Red, Tomato 1, Cherry 1, Watermelon. Compared with previous 131 different categories, it significantly reduces training time from 1 day to around 1 hour. In order to reduce potential problems such as poor model generality which is caused by the reduction of training samples, the experiment use image transform techniques to provide the model with more complex and diverse data. First add one Lambda layer to transform graph from RGB to HSV which can easily change parameters for saturation and hue. During transform process, the picture will randomly rotate 0 to 90 degrees, randomly change saturation from 0.9 to 1.2 and randomly change hue with range 0.02. All of those data augmentation will reduce the risk of overfitting, enhancing the model's generalization ability and increase test accuracy.

2.2. Method

The CNN model extracts feature through a convolutional layer, followed by a relu layer. Reduce the size of the output of the convolutional layer through pooling layers to reduce training time. Finally, Fully Connected Layers integrate features and classify them. In the CNN model, each neuron only needs to process a small part of the data, which makes the CNN model have stronger spatial level learning capabilities.

3. Result

3.1. Building model

For the establishment of the model, at beginning, the experiment uses 5*5 convolutional kernel, because the smaller receptive field may not be suitable for images with larger features like fruits. Then to verify the idea, retrained the model using 3*3 convolutional kernel with the same training set and test set to conduct controlled experiments. Fig.1 shows the accuracy for 5*5 and 3*3 convolutional kernel, it proves after multiple times epoch there are only small different between test accuracy which can be ignored and choosing 3*3 convolutional kernels reduces training time. In this paper, when using the same data set, total parameters used for 3*3 convolutional Kernel are lesser than 5*5 convolutional kernel models, and the experiment use 3*3 convolutional kernels for the final model [6].

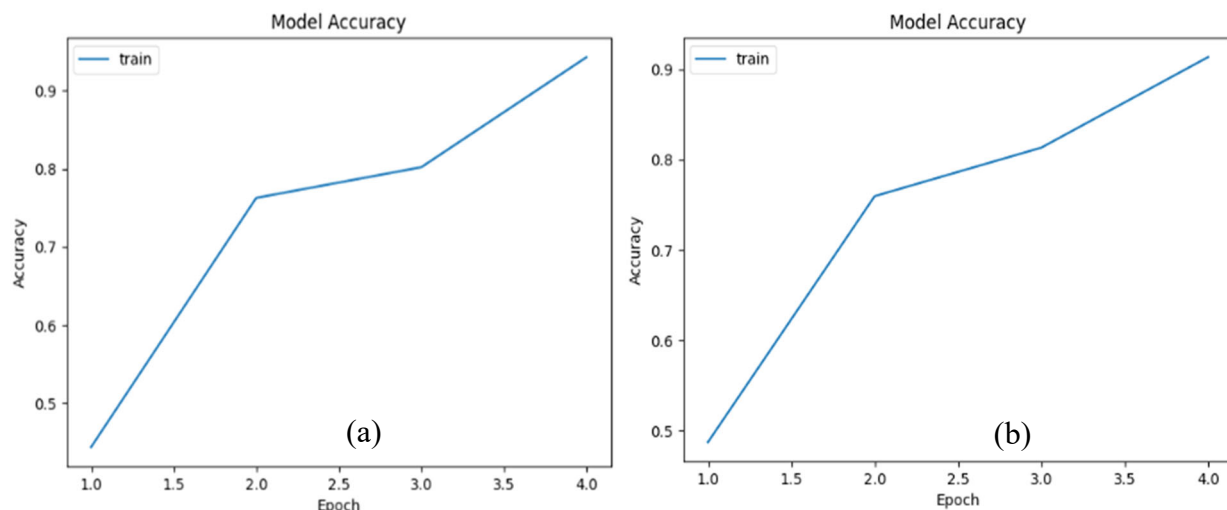


Figure 1. Accuracy of the different convolution kernel:(a)5*5 Convolutional Kernel; (b)3*3 Convolutional Kernel

The model uses four combinations of convolution layers and relu layers, The first convolutional layer uses 16 filters of size 3x3. The second convolutional layer uses 32 filters of size 3x3. The third convolutional layer uses 64 filters of size 3x3. The fourth convolutional layer uses 128 filters of size 3x3. Fig.2 shows the model accuracy for different epochs, it proves after epoch 10, model have enough high train accuracy. In order to prevent over fitting, choose epoch equal to 10 for final model, which can reduce training cost at same time.

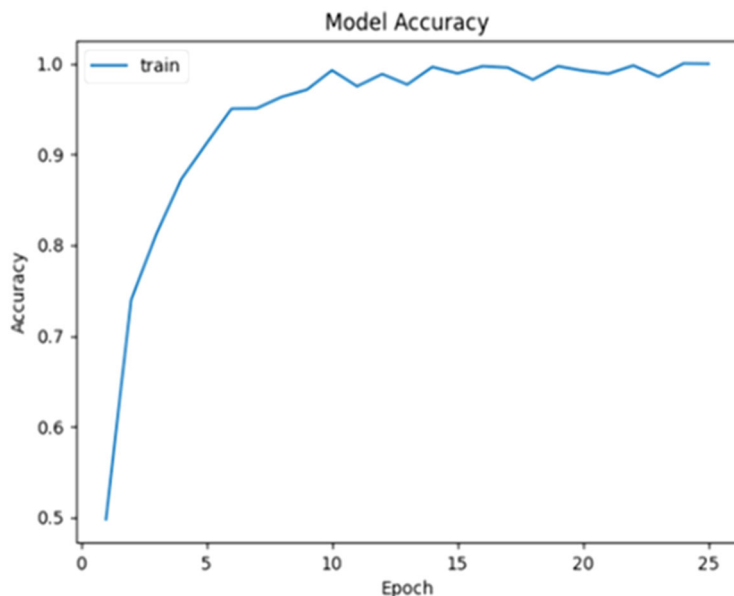


Figure 2. Model accuracy for different epochs

3.2. Based on the accuracy of the reason for speculation and try

By using TensorFlow and Keras to create and training mode, and training for 10 epoch, final test accuracy is 98.1714%. In actual more than 100 times tests, the accuracy of the model's prediction results was much lower than the theoretical value. This is reflected in the comments on the training and testing data. The paper will discuss and verify potential problems, solutions and Impact. The possible reasons are mainly reflected in these three aspects. Firstly, the model is overfitting. Secondly, training set is too simple, the model cannot handle various complex situations in reality. Finally, it is about the image size issue

To decrease overfitting, implement L2 regularization for building model. This regularization technique will update the cost function by adding a new term called the regularization term. Due to

the addition of the regularization term, the value of the weight matrix will decrease. Then the model raises the assumption that CNN with a smaller weight matrix means it is a simpler CNN model so that it can reduce overfitting [7]. Start with smaller lambda 0.001, the result didn't change compared with previous result, the actual prediction accuracy has no different. Retry training model with regularization lambda 0.01, From the accuracy result, the impact of the weight penalty become obvious, and the accuracy rate is rising slowly than before, the model converges slower. Although regularization decreases risk of overfitting, prediction accuracy lower than expected.

To reduce training time, use 100*100 low resolution images to train the model. For low resolution, there may be too much detail lost during training the model. This may cause actual prediction accuracy lower than expected. Change the resolution to its original size 224*224 for training the model, time cost become higher, after testing, the actual prediction accuracy become higher than 100*100, but still lower than expected.

Simple training set is another result which causes low prediction accuracy. From the given training set, all of them are too simple. Fig.3 is the training set for Apple Golden 2. For one category of fruit, it only uses one fruit for 360° shooting. The model trained by this training set will not be able to predict the fruit correctly. At the same time, the background of all fruits is pure white, which is very different from the complex background in reality.

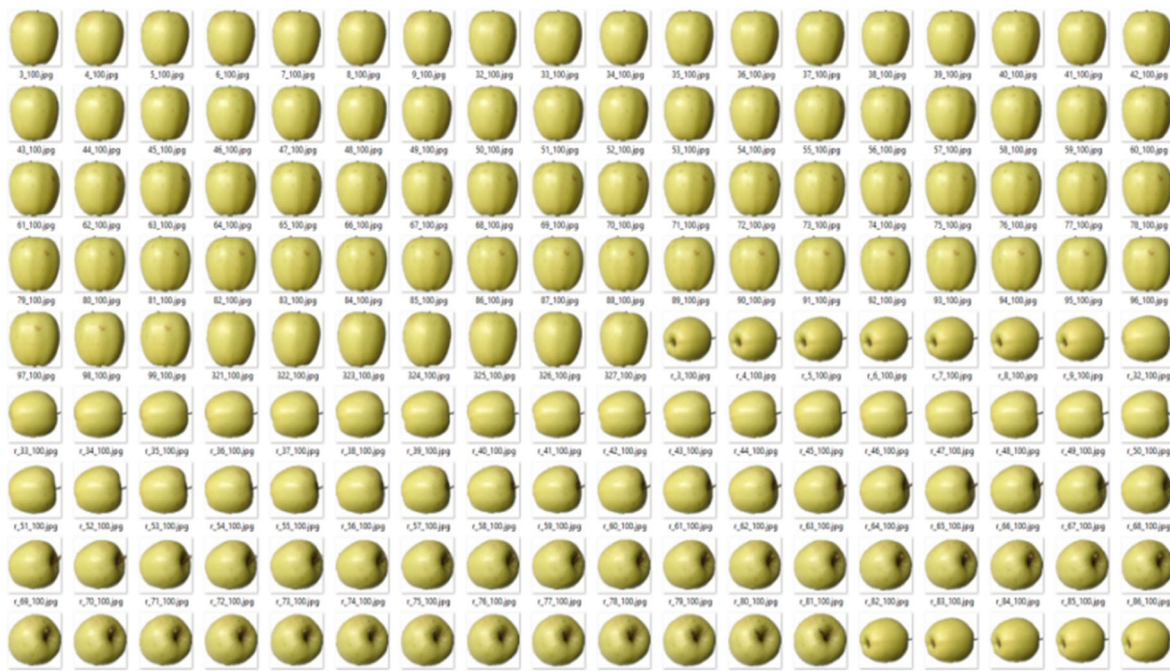


Figure 3. Training set of Apple Golden 2

4. Conclusion

The paper uses many attempts for rebuilding model with different ways to improve and observe influence for actual prediction accuracy. Regularization provides a positive impact on the accuracy of the model, effectively reducing the risk of overfitting. Using different image resolutions to train the model will affect the accuracy of the model. High resolution will allow the model to learn more subtle features, but it will increase the training time. Although there are many algorithms that can optimize the model and improve the accuracy of the model, the impact of the model accuracy problem caused by using an overly simple training set in this experiment cannot be repaired, for example, this experiment used randomly rotating images and adjusting grayscale to increase the complexity of the training set. Choosing an appropriate training set is crucial and directly determines the final accuracy in this experiment.

This experiment proved the practicality of CNN in image recognition, but at the same time, some shortcomings of CNN cannot be ignored. The prerequisite for using CNN is to train the model, which will take a lot of time and computer resources. Even after the model is trained, it still takes a long time to use the model and pictures to infer the type of fruit. In this experiment, on average, every five seconds the fruit video takes more than thirty seconds to predict the type of fruit, so the time delay is very high. The CNN model established in this experiment is not suitable for some industries that require immediate prediction results. In this experiment, CNN often has black box characteristics. The extracted features are highly abstract and difficult to be understood by people. This creates certain obstacles to data processing during debugging. The data used during training plays an outsized role in the performance of the model.

For artificial intelligence, there are still more excellent theories waiting to be discovered. In the past 10 years, artificial intelligence has developed rapidly. The disadvantages of the CNN model may be made up for by other models, such as the transformer model.

References

- [1] F. Amanov and A. Pradeep. The Significance of Artificial Intelligence in the Second Scientific Revolution - A Review. 2023 15th International Conference on Electronics, Computers and Artificial Intelligence (ECAI), Bucharest, Romania, 2023, 01 - 05. doi: 10.1109/ECAI58194. 2023. 10194056.
- [2] Alzubaidi, L., Zhang, J., Humaidi, A.J. et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. J Big Data 8, 53, 2021. <https://doi.org/10.1186/s40537-021-00444-8>.
- [3] S. Kido, Y. Hirano and N. Hashimoto. Detection and classification of lung abnormalities by use of convolutional neural network (CNN) and regions with CNN features (R-CNN). 2018 International Workshop on Advanced Image Technology (IWAIT), Chiang Mai, Thailand, 2018, 1 - 4, doi: 10.1109/IWAIT.2018.8369798.
- [4] Á. Zarándy, C. Rekeczky, P. Szolgay and L. O. Chua. Overview of CNN research: 25 years history and the current trends. 2015 IEEE International Symposium on Circuits and Systems (ISCAS), Lisbon, Portugal, 2015, 401 - 404, doi: 10.1109/ISCAS.2015.7168655.
- [5] Mihai, O. Fruits 360, A dataset with 90380 images of 131 fruits and vegetables. <https://www.kaggle.com/datasets/moltean/fruits>, 2020.05. 18. 0.
- [6] D. Chansong and S. Supratid. Impacts of Kernel Size on Different Resized Images in Object Recognition Based on Convolutional Neural Network. 2021 9th International Electrical Engineering Congress (iEECON), Pattaya, Thailand, 2021, 448 - 451, doi: 10.1109/iEECON51072. 2021. 9440284.
- [7] W. Swastika, R. B. Widodo, G. A. Balqis and R. Sitepu. The Effect of Regularization on Deep Learning Methods for Detection of Malaria Infection. 2021 International Conference on Converging Technology in Electrical and Information Engineering (ICCTEIE), Bandar Lampung, Indonesia, 2021, 87 - 90, doi: 10.1109/ICCTEIE54047. 2021. 9650646.