Indonesian Journal of Electrical Engineering and Computer Science
Vol. 12, No. 2, November 2018, pp. 468~475
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v12.i2.pp468-475

Evaluation of CNN, Alexnet and GoogleNet for Fruit Recognition

Nur Azida Muhammad[1], Amelina Ab Nasir[2], Zaidah Ibrahim[3], Nurbaity Sabri[4]
[1,2,3]Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia
[4]Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Campus Jasin, Melaka, Malaysia

Article Info

Article history:

Received May 19, 2018
Revised Jun 30, 2018
Accepted Jul 11, 2018

Keywords:

Alexnet
CNN
Fruit Recognition
Googlenet
Neural Network

ABSTRACT

Fruit recognition is useful for automatic fruit harvesting. Fruit recognition application can reduce or minimize human intervention during fruit harvesting operation. However, in computer vision, fruit recognition is very challenging because of similar shapes, colors and textures among various fruits. Illuminations changes due to weather condition also lead to a challenging task for fruit recognition. Thus, this paper tends to investigate the performance of basic Convolutional Neural Network (CNN), Alexnet and Googlenet in recognizing nine different types of fruits from a publicly available dataset. The experimental results indicate that all these techniques produce excellent recognition accuracy, but basic CNN achieves the fastest recognition result compared with Alexnet and Googlenet.

Corresponding Author:

Nur Azida Muhammad,
Faculty of Computer and Mathematical Sciences,
Universiti Teknologi MARA,
Shah Alam, Selangor, Malaysia.
Email: azidamuhammad@gmail.com

1. INTRODUCTION

Fruit recognition is useful for automatic fruit harvesting. By having fruit recognition application, it can reduce or minimize human intervention in their fruit harvesting operations. Fruit recognition system will automatically detect and inspect the fruit for harvesting within the image. The implementation of fruit recognition application can also increase the value of products to the consumers [1]. In addition, it can reduce the operation time and harvesting cost. Fruit recognition application is also useful for fruit disease detection in the early stage. For classical approach, the detection and identification of fruit disease is based on human naked eyes which is time consuming and costly [2]. Through automatic fruit recognition process, it can facilitate the control of fruit diseases as the disease can be avoided by appropriate sprinkling of pesticides.

Various researches on fruit recognition based on images have been performed. Multiple feature based analysis that include color, shape and texture have been applied to recognize six different types of fruits that are read apple, banana, lychee, orange, pineapple and pomegranate [3]. The researchers have used the Log Gabor filter to recognize the texture of a fruit. The hue has been calculated for color and shape was being analyzed by counting the perimeter and area pixels. In addition, the Artificial Neural Network (ANN) was being used for the classification and it achieves about 90 % classification accuracy.

The use of deep learning has dramatically improves the performance of object detection, speech recognition, visual object recognition and many other domains like genomics and drug discovery [1]. Deep learning is a class of machine learning algorithms that uses multiple layers that contain nonlinear processing units. Convolutional Neural Networks (CNNs) are classified as a deep learning algorithm [2]. It provides successful results in areas of image recognition and classification. Besides that, Alexnet and Googlenet are pre-trained CNN models that have produced very good results for the past few past years [3]. Alexnet is the winner of ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in 2012 while Googlenet is the winner in 2014 [4]. These models show big impacts on image recognition and classification tasks as they produce outstanding performance. As a result, CNN models were widely used in the field of computer vision [5].

As the CNN model goes deeper in their convolution architecture, it can reach a lower identification error rate compared to the human's eyes. Thus, the CNN model was implemented for fruit and vegetables classification as it produces great results for other object recognition applications. However, in computer vision, the fruit classification gives challenges in image recognition because of the similar shapes, colors and textures among various fruits [6]. The changes in the location and eye-sight view of the fruits also lead to this issue. Besides that, in the supermarket, the staff still requires to weigh the selling fruit which effects the cost of labor, time and the efficiency is low [7]. Thus, the main objective of this research is to investigate the recognition accuracy performance of basic CNN, Alexnet and Googlenet in recognizing fruit images to see whether the results will achieve more that 90% accuracy or not.

2. RELATED WORK

For the past few years, many researchers have been working on developing fruit recognition and classification approaches. WC Seng and SH Mirisaee [11] developed a fruit recognition system that combine features likes color, size and shape based. They used the nearest neighbor classification. The result showed a good performance for single fruit recognition only but is not suitable to use for fruit recognition that are in a bunch. An efficient fusion of texture and color for fruit type recognition has been proposed. However, the result of the recognition rate is not very encouraging [1]. Lecun, Bengio and Hinton [8] proposed fruit recognition using CNN. It involved without feature extraction and the input images were directly entered into the network. The results showed that the recognition rate is improved and it is suitable to identify multiple types of fruits.

Due to the rising values of agricultural supplies such as agrochemicals, water irrigation and power has lead to the agriculture industry as one of the most cost-demanding areas. A fruit detection system by using deep neural networks is proposed in [9]. The purpose of their paper is to build an accurate, fast and reliable fruit detection system which is an important element of an autonomous agricultural robotic platform. They adapt the technique of Faster Region-based CNN (R-CNN) for the fruit detection by using imagery obtained from two modalities which is color (RBG) and Near-Infrared (NIR). They performed fine-tuning of VGG16 network based on pre-trained ImageNet model. The combination of RGB and NIR multi-modal is retrained to perform the detection of seven types of fruits. As a result, the accuracy is improved and it is faster to be deployed to recognize a new fruit type. It takes only four hours to annotate and train the new model per fruit.

2.1 CNN (Convolutional Neural Network)

The architecture of CNN is structured as a series of layers, that consists of three layers which are convolve layer, pooling layer and Rectified Linear unit (ReLu) [10]. Convolve layer extracts features of an image using filter and image patch that strides over the input image. ReLu layer replaces all negative pixel values in the feature map with zero while pooling layer allows the feature map to be down-sampled after ReLu layer to reduce the dimensionality. Max pooling computes the maximum local of feature map. Neighboring pooling takes input from feature maps that are shifted or stride by more than one rows or columns. Figure 1 shows the architecture of a CNN.

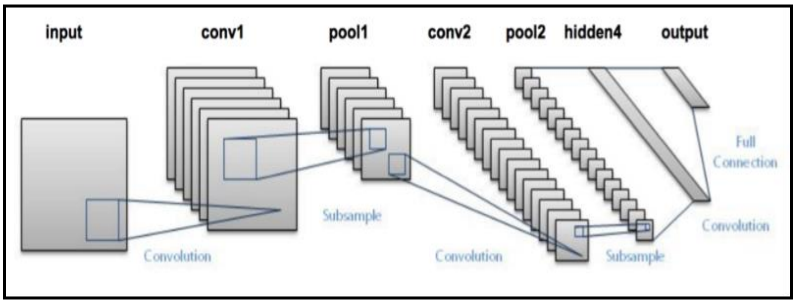

Figure 1. An illustration of CNN layers [10]

2.2 AlexNet

Alexnet is also known as transfer learning model where knowledge is learnt from training large amount of datasets. AlexNet won the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in 2012. It consists of 25 layers that combine a few stacks of convolutional layers and fully connected layers [13]. An illustration of the architecture of AlexNet is shown in Figure 2.

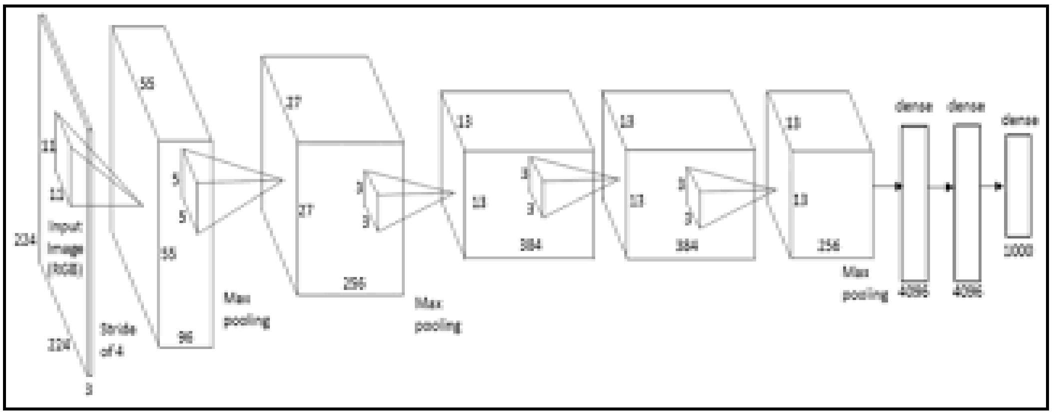

Figure 2. An illustration of AlexNet layers [14]

2.3 GoogleNet

Googlenet (a.k.a. Inception V1) is the winner of the ILSVRC 2014 competition from Google. It achieved a top-5 error rate of 6.67% [15]. This was very close to human level performance which the organizers of the challenge were forced to evaluate. As it turns out, this was actually rather hard to do and required some human training in order to perform the task. The human expert (Andrej Karpathy) was able to achieve a top-5 error rate of 5.1% (single model) and 3.6% (ensemble). The network used CNN inspired by LeNet but implemented a novel element which is dubbed an inception module. It used batch normalization, image distortions and RMSprop. This model is based on several very small convolutions in order to drastically reduce the number of parameters. Their architecture consisted of 22 layers of deep CNN but the number of parameters is reduced from 60 million (AlexNet) to 4 million (Googlenet). An illustration of the layers in GoogleNet is shown in Figure 3.

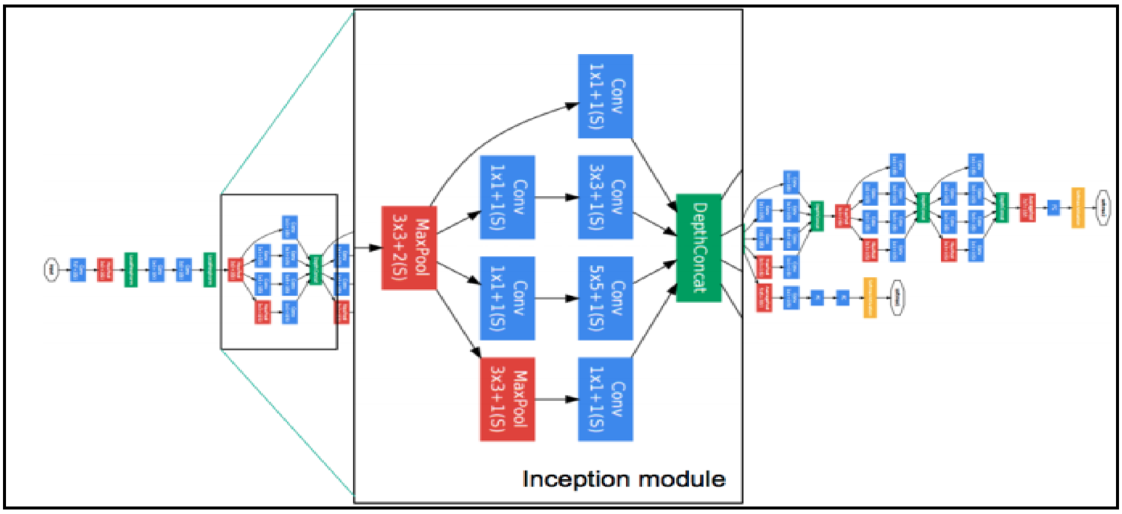

Figure 3. An illustration of the layers of GoogleNet [15]

3. RESEARCH METHOD

In this study, MATLAB 2018a is used to perform the experiments. In order to compare the performance of the three types of deep learning models, a set of fruit images are obtained from the GitHub [12] which is a freely available dataset. The dataset consists of 4900 training images and 1640 validation images. In addition, it is divided into 9 classes of fruit images which are kiwi, banana, strawberry, salak, pomegranate, pineapple, mandarins, dates, limes and carambula. The images consist of frames that were rotated by position. Table 1 shows the list of the number of instances of each class used for training as well as testing purposes. Figure 4 shows some of the specimen images for each class. The size of each image is 100 by 100 pixels.

Table 1. The number of images for training and validation [12]

No of class	Label	Number of Training Images	Number of Validation Images
1	Pomegranate	246	82
2	Salak	490	162
3	Banana	490	166
4	Pineapple	490	166
5	Mandarins	490	166
6	Dates	490	166
7	Limes	490	166
8	Carambula	490	166
9	Strawberry	492	164

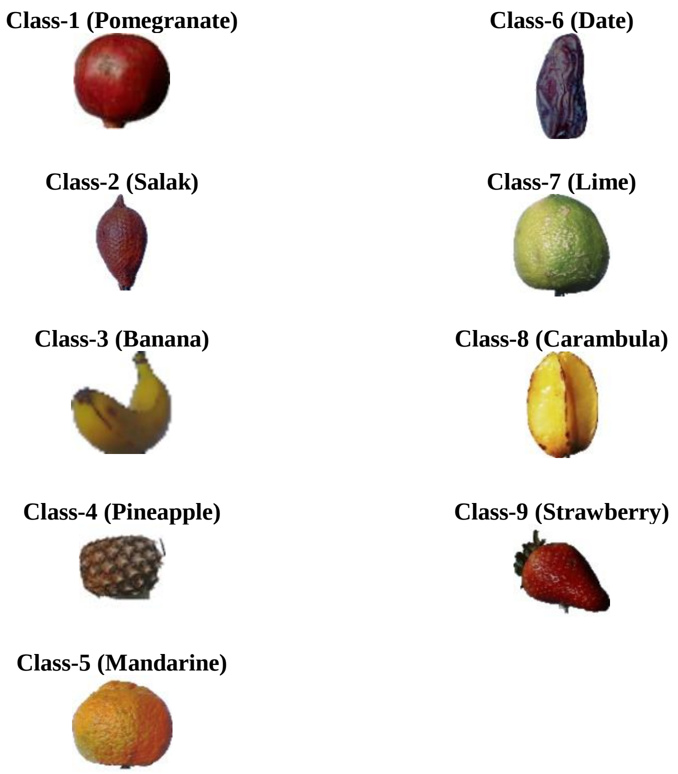

Figure 4. Specimen images for each type of fruits used in the experiment

4. RESULTS AND ANALYSIS

4.1 CNN (Convolutional Neural Network)

For the experiment using CNN, the size of the input image is set to 100 by 100 by 3 pixels due to the memory constraint of the computer used. The image only displays one dataset by identifying which categories it is. If it is true, then the text will be displayed in green color, but if it is false, the text is in blue color. In this case, the output is green (true), which is strawberry. Figure 5 shows the coding for the execution of CNN for an image of a strawberry.

```
labels = classify(net, imds_test);
ii = randi(490);
im = imread(imds_test.Files{ii});
imshow(im);
if labels(ii) == imds_test.Labels(ii)
colorText = 'g';
else
colorText = 'r';
end
title(char(labels(ii)),'Color',colorText);
```

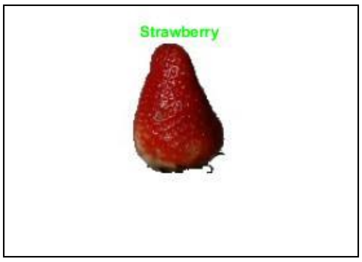

Figure 5. The sample coding and result for an image of strawberry

CNN takes the raw color image and the features are automatically extracted by the layers. A stack of CNN consist of convolve layer, pooling layer and ReLu layer while additional stack of layers can be added to compare the performance. The size in convolve layer and the value of stride in the pooling layer represent the number of column to be skipped for the sliding window that can change as these values can effect the result of the recognition performance. Besides that, the values of *maxepochs* represent the number of iteration for the training process and initial learning rate that represent the value of the weight to be adjusted during the training process, can be changed to view their effect to the recognition rate.

Next is the validation accuracy which is 100%, that makes the final accuracy is 1. The time to display the output image only takes 5 seconds. The training option need to be specified for CNN. An epoch is a full training cycle of the entire dataset. The maximum number of epochs for defined for CNN in this experiment is 10 with initial learning rate is 0.001. The frequency of CNN is 30 iterations**.** Figure 6 shows the training and validation progresses of CNN.

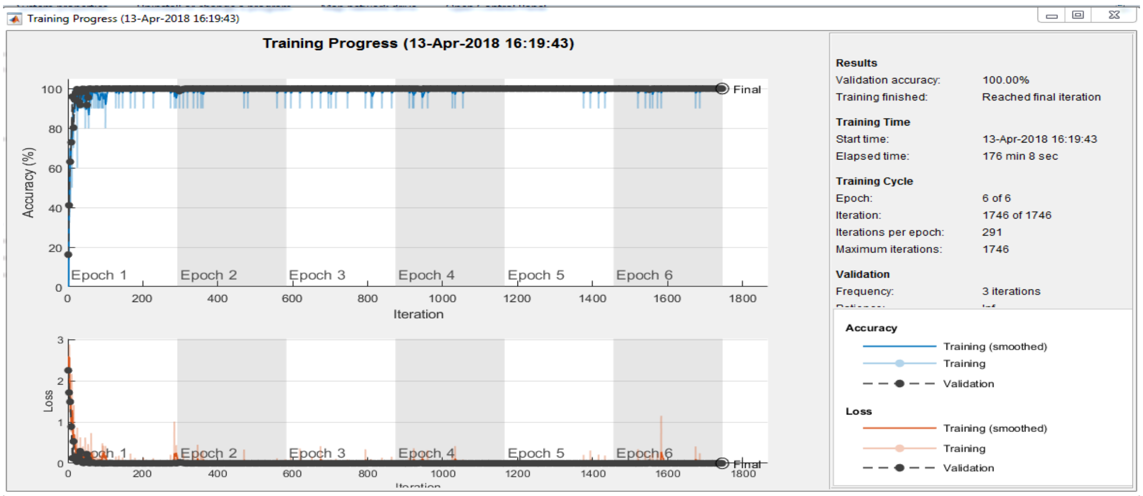

Figure 6. The results of CNN

4.2. AlexNet

AlexNet is also called as transfer learning model, which is the knowledge learnt from training large amount of database. For this experimnet, the size of each image is 227 by 227 pixels. The image displays four dataset of fruits with their predicted labels. Figure 7 shows some sample results produced by Alexnet which is strawberry, mandarine, dates and limes.

Alexnet consists of layers transfer with a fully connected layer, softmax layer and a classification output layer; by specifying the options of the new fully connected to the new data. By specifying the training options, transfer learning keeps the values of the parameters from the previous layers of the pretrained network. The initial learning rate is set to a small value to slow down the transfer layer. Besides that, the values of maximum epochs that represent the number of iteration for the training process and initial learning rate that represents the value of the weights to be adjusted during training process is set to 0.0001.

The training process took about 176 minutes and 8 seconds. In this experiment, the maximum number of epochs for Alexnet is 6 and the maximum number of iteration is 1746. Figure 8 shows the detail information of the results of Alexnet where it achieves 100% accuracy for fruit recognition.

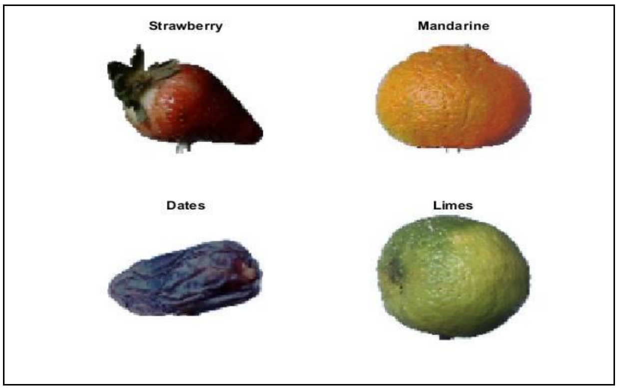

Figure 7. The results of Alexnet

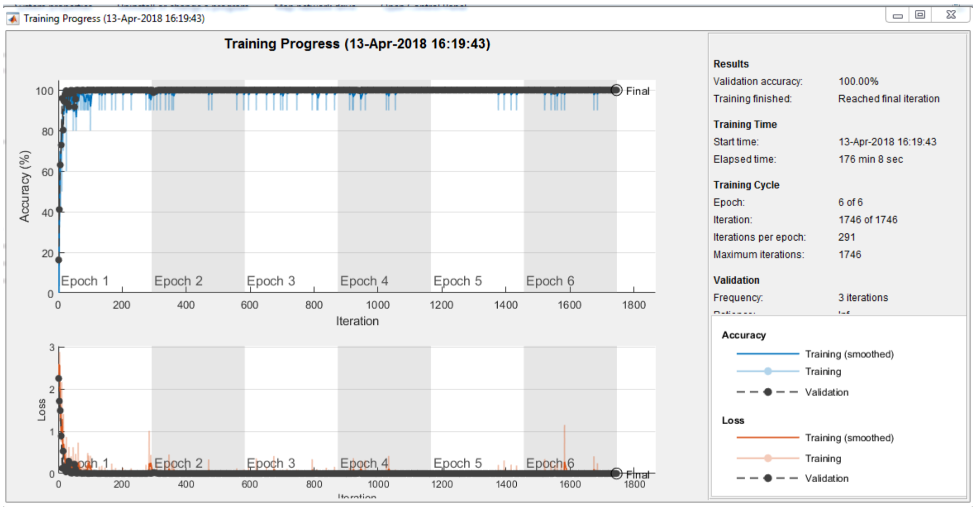

Figure 8. The detail results of Alexnet

4.3. GoogleNet

The size of an image in the input layer of Googlenet is 224 by 224. The result for the image is displayed with the predicted label (banana) and probabilities with the label which is 90.6%. Figure 9 shows the result produced by Googlenet where it displays the name of the fruit with the predicted probability.

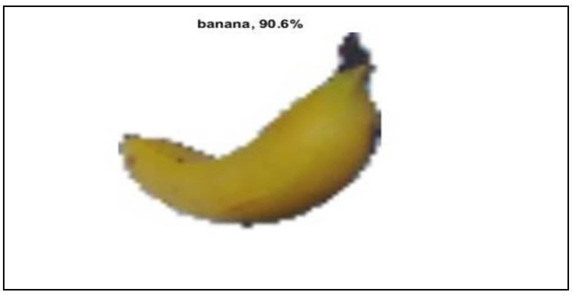

Figure 9. The result of Googlenet

Googlenet displays the top five predicted label and the probabilities. Figure 10 shows the top 5 predictions and probability of an image of a banana. The validation accuracy is 100%, that makes the final accuracy as 1. The training time took about 487 minutes and 49 seconds to complete the process. On the other hand, it needs a high power computer or laptop to complete the execution in a relatively short time. In this experiment, the maximum number of epochs for Googlenet is 6 and the maximum number of iteration is 1746. Figure 11 shows the detail results of GoogleNet.

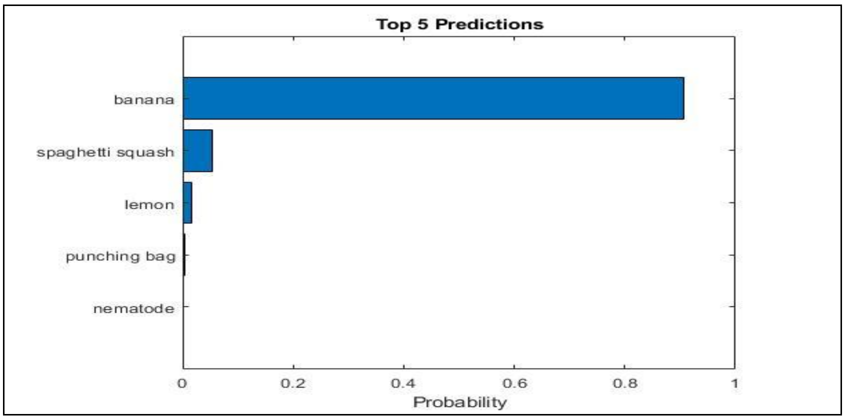

Figure 10. The top 5 prediction of probability of an image of a banana

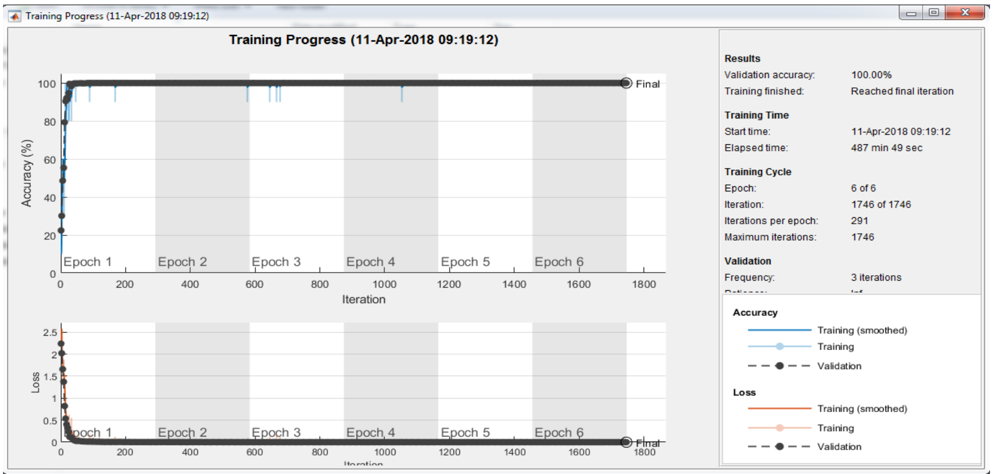

Figure 11. The detail results of Googlenet

Table 2 lists the overall results of fruit recognition using CNN, AlexNet and Googlenet. By looking at Table 2, we can see that all these three models produce the same perfect accuracy which is 1. But the runtime required by CNN is the lowest while Googlenet requires the longest time. This is due to the architecture of the models where CNN has the smallest number of layers while Googlenet has the largest number of layers.

Table 2. The performance comparison between CNN, Alexnet and Googlenet

	CNN	Alexnet	Googlenet
Input size	100 100 3	227 227 3	224 224 3
Image display	1	4	1
Extra features	No	No	Display top prediction
Accuracy	1	1	1
Runtime	5 second	176 min 8 seconds	487 min 49 seconds
Epoch	10	6	6
Frequency	30 iteration	3 iteration	3 iteration

5. CONCLUSION

In this paper, we evaluate the recognition performance of CNN, Alexnet and Googlenet for nine different types of fruits. The experimental results show that the three models produce a perfect 100% recognition accuracy but with different range of run time. CNN model seems to be the best choice for this experiment since it is very accurate and fast. For future work, we will investigate other fruit datasets with more fruit types and involve fruits in a bunch.

ACKNOWLEDGEMENT

The authors would like to thank Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Shah Alam, Selangor, for sponsoring this research.

REFERENCES

[1] Lecun, Y. Bengio, and G. Hinton. (2015). “Deep learning”, Nature, vol. 521, no. 7553, 436–444.
[2] Stuart J. Russell, Peter Norvig. (1995). Artificial Intelligence - A Modern Approach. New Jersey, Simon and Schuster Company. ArticialIntelligence - A Modern Approach.
[3] G. Zheng (2017). *Fruit and Vegetables Classification System Using Image Saliency and Convolutional Neural Network*. IEEE 3rd. Information Technology and Mechatronics Engineering Conference. pp. 613–617.
[4] Ballester, P., &Araujo, R. M. (2016). On the Performance of GoogLeNet and AlexNet Applied to Sketches. Proceedings of the 30th Conference on Artificial Intelligence (AAAI 2016), 1124–1128.
[5] Jana, S., Basak, S. & Parekh, R. (2017). *Automatic Fruit Recognition from Natural Images using Color and Texture Features*. Conference on Devices for Integrated Circuit. 23–24.
[6] Sabri, N., Ibrahim, Z., Syahlan, S., Jamil, N., & Mangshor, N. N. A. (2017). *Palm Oil Fresh Fruit Bunch Ripeness Grading Identification Using Color Features*. Journal of Fundamental and Applied Sciences, 9(4S), 563-579.
[7] Ibrahim, Z., Sabri, N., & Mangshor, N. N. A. (2018). Leaf Recognition using Texture Features for Herbal Plant Identification.
[8] Ibrahim, Z., Kasiran, Z., Isa, D., & Sabri, N. (2016). *Multi-script Text Detection and Classification from Natural Scenes*. In International Conference on Soft Computing in Data Science (pp. 200-210). Springer, Singapore.
[9] Shukla, D., & Desai, A. (2017). *Recognition of fruits using hybrid features and machine learning*. International Conference on Computing, Analytics and Security Trends, CAST 2016, 572–577. https://doi.org/10.1109/CAST.2016.7915033
[10] Hou, L., Wu, Q., Sun, Q., Yang, H., & Li, P. (2016). *Fruit recognition based on convolution neural network*. 2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery, ICNC-FSKD 2016, 18–22. https://doi.org/10.1109/FSKD.2016.7603144
[11] Naskar, S. (2015). A Fruit Recognition Technique using Multiple Features and Artificial Neural Network, International Journal of Computer Applications. 116(20), 23–28.
[12] Li, P., Lee, S. H., & Hsu, H. Y. (2011). Review on fruit harvesting method for potential use of automatic fruit harvesting systems. Procedia Engineering, 23, 351-366.https://doi.org/10.1016/j.proeng.2011.11.2514
[13] Dubey, S. R., & Jalal, A. S. (2012). Detection and Classification of Apple Fruit Diseases
[14] Using Complete Local Binary Patterns. Third International Conference on Computer and Communication Technology, 346–351.
[15] Sa, I., Ge, Z., Dayoub, F., Upcroft, B., Perez, T., & McCool, C. (2016). Deepfruits: A fruit detection system using deep neural networks. Sensors (Switzerland), 16(8).
[16] A. Krizhevsky, I. Sutskever and G. E. Hinton. (2012). ImageNet Classification with Deep Convolutional Networks. Advances in Neural Information Processing Systems 25, 2012.