



Comparing Convolution Neural Network Models for Leaf Recognition

Nurbaity Sabri^{1*}, Zalilah Abdul Aziz², Zaidah Ibrahim², Muhammad Akmal Rasydan Bin Mohd Rosni² and Abdul Hafiz bin Abd Ghapul²

¹Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Campus Jasin, Melaka, Malaysia

²Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia.

*Corresponding author E-mail: nurbaity_sabri@melaka.uitm.edu.my

Abstract

This research compares the recognition performance between pre-trained models, GoogLeNet and AlexNet, with basic Convolution Neural Network (CNN) for leaf recognition. Lately, CNN has gained a lot of interest in image processing applications. Numerous pre-trained models have been introduced and the most popular pre-trained models are GoogLeNet and AlexNet. Each model has its own layers of convolution and computational complexity. A great success has been achieved using these classification models in computer vision and this research investigates their performances for leaf recognition using MalayaKew (MK), an open access leaf dataset. GoogLeNet achieves a perfect 100% accuracy, outperforms both AlexNet and basic CNN. On the other hand, the processing time for GoogLeNet is longer compared to the other models due to the high number of layers in its architecture.

Keywords: AlexNet, CNN, GoogLeNet, Leaf recognition

1. Introduction

As human moving towards development, many trees and plants are deforestation to build buildings and houses for humans to live. Ironically, plant is very useful for human and some of these plants are at the risk of extinction. About 80% of Asian and African depend on herbal medicine for cure purposes and it is cheaper compared to modern medication. The importance of plants for human is slowly forgotten by the younger generations. Therefore, it is crucial to conserve and protect the information about these medicinal plants from extinction. With the current technology, researchers are able to develop and share databases to preserve this essential earth's ecology. This database must be able to store various information regarding medicinal plants including text and images, and allow for accurate and efficient recognition process and retrieval of information. The development of this database allows preserving the medicinal plant information for future uses. Furthermore, it also helps the botanists to keep and preserved the knowledge from extinction [1].

There are several ways to recognize a plant, like flower, root, leaf, vein and fruit. Unlike flowers or fruits, leaves can be easily captured throughout the year [2]. Capturing the root of a plant is not easily accessible as it may be harmful to the plant. Veins can only be seen clearly for recognition purposes with good lighting. Moreover, the leaf can provide enough discriminative features for plant recognition. In recent years computer vision methodologies and pattern recognition techniques have been applied towards automating the procedures of plant recognitions. Various researches on leaf recognition have been conducted but the accuracy obtained so far has not reached 100% accuracy [2][3][4][5][6]. This shows that there is still room for improvements to increase the accuracy result.

Convolution Neural Network (CNN) has been introduced in the past few years in computer vision with a remarkable accuracy compared to state-of-the-art approaches [7]. It is a multi-stage image processing and can extract high level hierarchical representations of data. It has been implemented in numerous pattern recognition researches such as palm oil [8] and fruit [9]. The feature detection layer of CNN is obtained during the training process; therefore it can avoid explicit feature extraction and implicitly learns from the training data. Currently many pre-trained models have been developed to increase the performance and recognition rates and two of the popular models are AlexNet [7] and GoogLeNet [10].

AlexNet was trained using ImageNet database that consist of 1.2 million images with 1000 classes of objects and manage to achieve a better performance compared to handcraft features [7]. Large Scale Visual Recognition Challenge (ILSVRC) is a competition using ImageNet dataset where it evaluates object classification on a large scale. In 2012, AlexNet model achieved a top-5 error of 16.4% for the classification of images into 1,000 possible categories [10]. In the last few years, various advances in deep CNN has been performed which manage to reduce the error rate to 3.57% [7]. While training large CNNs can be very time-consuming, the trained models can classify images very accurately.

GoogLeNet is the winner of ILSVR 2014 with a top 5 error of 6.7%. It is different from AlexNet where it uses inception modules [10]. The model uses different size of convolution and concatenate filter for the next layer. AlexNet uses about 60 millions parameters and shared similar structures. Meanwhile, GoogLeNet used 12 times fewer parameters and deeper models compared to Alexnet [11]. GoogLeNet achieves a higher accuracy compared to AlexNet in breast cancer detection [12]. On the other hand, AlexNet

achieves a higher accuracy for palm oil grading recognition [13]. Therefore this paper aims to analyze the performance of basic CNN, GoogLeNet and AlexNet for leaf recognition using a publicly available dataset of the leaves.

The organization of this paper is as follows. The next section discusses about the works related to this research. After that, explanation about the architecture of CNN, AlexNet and GoogLeNet is produced. Next, the results and analysis from the experiment were discussed. Finally, this paper ends with conclusion and possible future work.

2. Related Work

In recent years, research in leaf recognition has attracted many researchers because the outcome can be useful in agriculture and heritage. Botanists can identify the name of plants by examining the leaves. Farmers can identify some plant diseases through the leaves. Knowledge about the use of medicinal plants is a heritage that needs to be preserved and this can be done by observing the leaves. Nutrients deficiency can also be identified through the leaves. Thus, the importance of research in automatic leaf recognition is important.

The process of automatic leaf recognition can be conducted in two approaches, that are handcrafted and deep learning. In handcrafted approach, suitable features need to be extracted before the recognition process. Gray-Level Co-occurrence Matrix and invariant features have been utilized for leaf recognition in [3] while Histogram of Oriented Gradient (HOG) has been applied in [6]. In fact, a fusion of texture, shape and color has also been used in [6].

CNN is a type of deep learning where the handcrafted feature extraction process is automatically being performed within it. This is possible due to the availability of tremendous amount of data for training purposes. The development of pre-trained CNN models such as AlexNet [16] and GoogLeNet [17] is also gaining its popularity in object recognition. Comparison of recognition performance among these CNN models has been conducted by a few researchers. The accuracy results of both Alex and GoogLeNet have reached 99% for SAR target recognition [19], [20]. Thus, this research investigates the performance of these pre-trained models for leaf recognition.

3. Research Method

3.1. Data Set

The MalayaKew (MK) dataset consist of about 2800 images from 44 different plant species were used in this experiment [14]. Each plant species has its own different leaf characteristic such as elongation, shape, size and colour. Each image is resized to 720 x 960 pixels using .png format. Fig. 1 shows some sample images of 10 species that were used in this research. Meanwhile, Fig. 2 contains some detail information of each plant such as its scientific name and the number of leaf specimens collected for each species. The plant species were numbered from 1 to 15 and from 22 to 36 that exhibit simple leaves, while species numbered from 16 to 21 and from 37 to 40 have complex leaves structures.

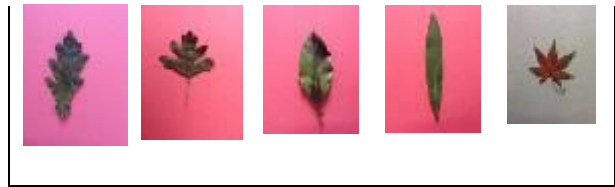


Fig. 1: Example of the leaf dataset [14].

C.	Scientific Name	#	C.	Scientific Name	#
1	<i>Quercus suber</i>	12	22	<i>Primula vulgaris</i>	12
2	<i>Salix atrocinerea</i>	10	23	<i>Erodium sp.</i>	11
3	<i>Populus nigra</i>	10	24	<i>Bougainvillea sp.</i>	13
4	<i>Alnus sp.</i>	8	25	<i>Arisarum vulgare</i>	9
5	<i>Quercus robur</i>	12	26	<i>Fannyinus japonicas</i>	12
6	<i>Cotoneaster monogyna</i>	8	27	<i>Ilex perado ssp.</i>	11
7	<i>Ilex aquifolium</i>	10	28	<i>Magnolia soulangeana</i>	12
8	<i>Nerium oleander</i>	11	29	<i>Buxus sempervirens</i>	12
9	<i>Betula pubescens</i>	14	30	<i>Urtica dioica</i>	12
10	<i>Tilia tomentosa</i>	13	31	<i>Podocarpus sp.</i>	11
11	<i>Acer palmatum</i>	16	32	<i>Acacia sellowiana</i>	11
12	<i>Celtis sp.</i>	12	33	<i>Hydrangea sp.</i>	11
13	<i>Corylus avellana</i>	13	34	<i>Pseudotsuga japonica</i>	11
14	<i>Castanea sativa</i>	12	35	<i>Magnolia grandiflora</i>	11
15	<i>Populus alba</i>	10	36	<i>Geranium sp.</i>	10
16	<i>Acer negundo</i>	10			

Fig. 2: Sample plant's scientific names.

3.2. Convolution Neural Network (CNN)

CNN is a deep learning method along with autoencoders, Boltzman machines and sparse coding. It is similar to feedforward neural network where they are made up of neurons that stores learnable weights and biases. The basic CNN or ConvNet is a structure of layers where it consists of input layer, output layer and many hidden layers between the input and output layers. There are four main layers namely convolution, pooling, rectified linear unit (ReLU) and fully-connected layer or the classification layer. In a convolution layer, convolution operation is performed where features from the input images are extracted over a set of convolution filters and produces feature maps. The pooling layer reduces the number of parameters of the feature maps where a nonlinear down sampling is performed to simplify the output. Rectified linear unit (ReLU) changes the negative values in the feature map into 0 and remain the positive values. The fully-connected layer computes the category scores where the output neuron with the highest score is the classified result [15].

CNN trains data similar to the conventional feedforward neural network. Each image is submitted through the layers until a loss function is obtained at the top layer. Further, the loss is back-propagated by adjusting the weights through the layers based on Stochastic Gradient Descent (SGD) method. The process is repeated until a pre-defined number of training epochs is obtained.

Fig. 3 illustrated the CNN architecture for leaf recognition in this research.

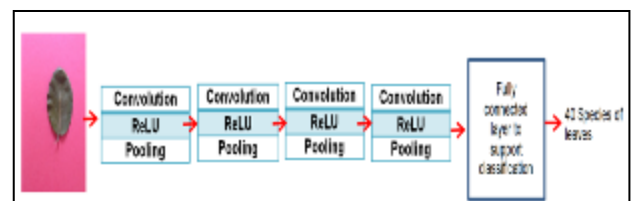


Fig. 3: The basic CNN layers for leaf recognition.

2.3 AlexNet

Two popular pre-trained CNN models investigated in this paper are AlexNet [16] and GoogLeNet [17], winner of the ImageNet Large Scale Visual Recognition Challenge (ILSVRC). AlexNet is designed with deeper architecture where it has more filter layers with stacked convolution layers compared to the basic CNN architecture. AlexNet is a combination of convolution and max pooling on the first three layers followed by two single convolution layer also known as dropout layer. The fifth layer of this single convolution layer ended with max pooling layer. The other three layers are fully connected layers. The output from fully connected layers are sent to softmax layer where it produces 1000 probabilities values for 1000 class labels [18]. For this research, the fully-connected layers are fine-tuned to classify 40 different categories since the dataset consists of 40 different leaf species.

Fig. 4 shows AlexNet architecture where it is designed to classify 1000 possible categories of images.

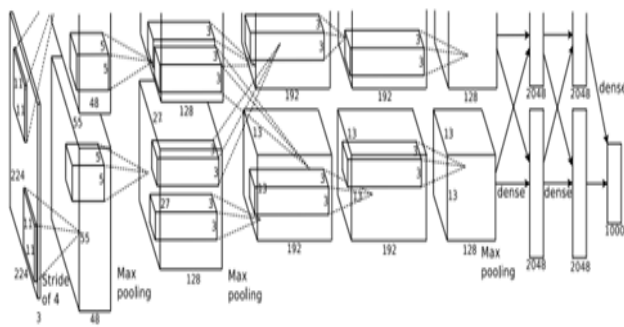


Fig. 4: AlexNet architecture [15].

3.3. GoogLeNet/Inception

GoogLeNet architecture uses dubbed and inception model where it involves three processes which are the batch normalization, image distortions and Root Mean Square Propagation (RMSprop). To drastically reduce the number of parameters, very small convolution is used in this model. It consists of 22 layers of deep CNN but the number of parameters are reduced from 60 millions (AlexNet) to 4 millions. Inception model is the key features of GoogLeNet which means “network in network”. This inception model uses parallel convolution of 1×1, 3×3, 5×5 filter together with max pooling layer in parallel form. Therefore, this operation is able to capture numerous types of features in the same form. Filter concatenation layer simply concatenates the outputs of all these parallel layers. GoogLeNet uses nine inception models combined with the traditional convolutions and softmax for classification output [18].

Fig. 5 illustrates GoogLeNet architecture used in this research.

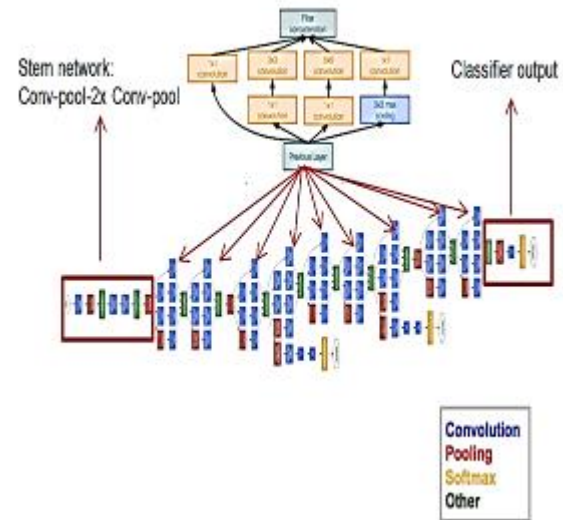


Fig. 5: GoogLeNet architecture [18].

4. Result and Discussion

A comparative study between basic CNN and pre-trained CNN models, namely AlexNet and GoogLeNet, was demonstrated. This experiment uses 44 species of leaves from MK Leaf dataset [14]. Five execution operations were conducted using the same parameter settings for all models where eighty percent of the data is used for training while the other twenty percent is used for validation, and an average accuracy is recorded. The learning rate is set to 0.0001, the maximum epoch is set to 20 and the minimum batch size is set to 40.

By referring to Table 1, we can see that the experimental results show that GoogLeNet achieved the highest average validation accuracy which is 100%. Meanwhile, AlexNet achieved 96% accuracy and basic CNN arrived to 99% accuracy. However, GoogLeNet shows the highest average elapsed time which is 41 minutes and 5 seconds, followed by AlexNet with 2 minutes and 1 second and the lowest elapsed time is basic CNN with 1 minute and 6 seconds. From the results obtained, it shows that GoogLeNet, the winner of ILSVRC in 2017, outperformed AlexNet and basic CNN. As mentioned earlier, GoogLeNet has more layers compared to AlexNet and basic CNN where the accuracy gets better when the number of layers increases. However, this theory is not applicable for AlexNet where it has more layers compared to basic CNN, but is less accurate. This possibly happen due to the increase number of noises in the leaf dataset used in this experiment compared to ImageNet where the data has less noises.

Table 1: Performance of average accuracy and elapsed time for GoogLeNet, AlexNet and basic CNN

Model	Average accuracy	Average Elapsed Time
GoogLeNet	100%	41 minutes and 5 seconds
AlexNet	96%	2 minutes and 1 seconds
Basic CNN	99%	1 minute and 6 seconds

5. Conclusion

In this paper, comparison between basic CNN, AlexNet and GoogLeNet has been performed with respect to accuracy and elapsed time. The experiment results show that GoogLeNet achieves a higher accuracy rate compared to the other two models which is 100%. However, GoogLeNet has the highest elapsed time due to the number of layers and more deeper learning process. The use of GoogLeNet is recommended for classification and recognition when processing time is not an issue. More models and more publicly available leaf datasets will be tested to achieve

a perfect accuracy with much lesser processing time for future work.

Acknowledgement

This work is supported by ARAS Grant of Universiti Teknologi MARA (UiTM), Malaysia.

References

- [1] Wu, S. G., Bao, F. S., Xu, E. Y., Wang, Y. X., Chang, Y. F., & Xiang, Q. L. (2007). A leaf recognition algorithm for plant classification using probabilistic neural network. In *Signal Processing and Information Technology, 2007 IEEE International Symposium on* (pp. 11-16).
- [2] Wu, H., Wang, L., Zhang, F., & Wen, Z. (2015). Automatic leaf recognition from a big hierarchical image database. *International Journal of Intelligent Systems*, 30(8), 871-886.
- [3] Chaki, J., Parekh, R., & Bhattacharya, S. (2015). Plant leaf recognition using texture and shape features with neural classifiers. *Pattern Recognition Letters*, 58, 61-68.
- [4] Yasar, A., Saritas, I., Sahman, M. A., & Dundar, A. O. (2015). Classification of Leaf Type Using Artificial Neural Networks. *International Journal of Intelligent Systems and Applications in Engineering*, 3(4), 136-139.
- [5] Jamil, N., Hussin, N. A. C., Nordin, S., & Awang, K. (2015). Automatic Plant Identification: Is Shape the Key Feature?. *Procedia Computer Science*, 76, 436-442.
- [6] Ibrahim, Z., Sabri, N., & Mangshor, N. N. A. (2018). Leaf Recognition using Texture Features for Herbal Plant Identification, *IJECS*, vol. 9, no. 1, 152-156.
- [7] Liang, M., & Hu, X. (2015). Recurrent convolutional neural network for object recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 3367-3375).
- [8] M. A. Hedhazi, I. Kourbane and Y. Genc, "On Identifying leaves: A Comparison of CNN with classical ML methods", *IEEE 25th Signal Processing and Communications Applications Conference*, 2017
- [9] Y. D. Zhang, Z. D. Dong, X. Chen, W. Jia, S. Du, K. Muhammad and S. H. Wang, "Image based fruit category classification by 13-layer deep convolutional neural network and data augmentation", *Journal of Multimedia Tools and Applications*, Springer, pp. 1-20, 2017.
- [10] Ballester, P., & de Araújo, R. M. (2016, February). On the Performance of GoogLeNet and AlexNet Applied to Sketches. In *AAAI* (pp. 1124-1128)
- [11] Tian, Y., Luo, P., Wang, X., & Tang, X. (2015). Deep learning strong parts for pedestrian detection. In *Proceedings of the IEEE international conference on computer vision* (pp. 1904-1912)
- [12] Lévy, D., & Jain, A. (2016). Breast mass classification from mammograms using deep convolutional neural networks. *arXiv preprint arXiv:1612.00542*
- [13] Silva, P. F., Marcal, A. R., & da Silva, R. M. A. (2013, June). Evaluation of features for leaf discrimination. In *International Conference Image Analysis and Recognition* (pp. 197-204). Springer, Berlin, Heidelberg.
- [14] Lee, S. H., Chan, C. S., Wilkin, P., & P. (2015, September). Deep-plant: Plant identification with convolutional neural networks. In *Image Processing (ICIP), 2015 IEEE International Conference on* (pp. 452-456).
- [15] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097-1105).
- [16] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. *arXiv*, 2014.
- [17] Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). Using deep learning for image-based plant disease detection. *Frontiers in plant science*, 7, 1419.
- [18] Zhong, Z., Jin, L., & Xie, Z. (2015, August). High performance of-line handwritten Chinese character recognition using GoogLeNet and directional feature maps. In *Document Analysis and Recognition (ICDAR), IEEE 13th International Conference on* (pp. 846-850).
- [19] Jiaqi, S., Changwen, Q. & Jianwei, L. (2017). A Performance analysis of convolutional neural network models in SAR target recognition. *IEEE SAR in Big Data Era: Models, Methods and Applications*, (pp. 1-6).
- [20] Maha A. M., Esra A. H. Bilal, T. & Naoufel, W. (2018). Automatic Target Automation in SAR Images. *IEEE International Conference on Artificial Intelligence and BigData*, (160-164).