Multi-maxpooling Convolutional Neural Network for Medicinal Herb Leaf Recognition

Zaidah Ibrahim^{a,*}, Nurbaity Sabri^b and Dino Isa^c

^{a*}Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia.

^bFaculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 77300 Merlimau, Melaka, Malaysia.

^cConnect Initiative, Crops For the Future, Jalan Broga, 43500 Semenyih, Selangor, Malaysia.

*Corresponding Author: zaidah@tmsk.uitm.edu.my

Abstract

This paper evaluates the leaf recognition performance using basic Convolutional Neural Network (CNN) and two popular pre-trained CNN namely, AlexNet and GoogLeNet. CNN is known to be dependent highly on large amount of training data but the result achieved is tremendously accurate. This research tends to find an alternative to obtain excellent results but with low amount of training data. Publicly available leaf dataset that is, Leafsnap, is being used in this experiment and data augmentation has been applied to relatively increase the number of training data. The dataset consist of images in different sizes, orientations and shapes. Thus, a robust approach is utilized by adding multi-maxpooling layers before the fully-connected layers of the basic CNN. The experimental results indicate that the performance of basic CNN can be improved by applying multi-maxpooling layers where the accuracy results produced are nearly as good as AlexNet or GoogLeNet but at a faster rate.

Keywords: Alexnet, CNN, Googlenet, leaf recognition.

1. Introduction

Medicinal plants are available everywhere throughout the world including Malaysia but not many people recognize them and know about their users especially for medical purposes. Preservation of the knowledge of these medicinal plants is important since it enables the general public to gain useful knowledge which they can apply when necessary. Various researches have been conducted to recognize plants and the main part of the plant that is

popular to be used for plant recognition is the leaf⁽¹⁻⁵⁾. Leaf identification can be performed using two approaches, namely handcrafted approach and deep learning.

In handcrafted approach, suitable features and classifiers need to be determined prior to applying them for object recognition. From out previous research⁵, texture features such as Histogram of Oriented Gradient (HOG), Local Binary Pattern (LBP) and Speeded-Up Robust Feature (SURF) produce very promising results for leaf recognition. A combination of texture, color and shape features for leaf recognition also achieve very good accuracy results³. However, identifying the good feature and classifier for any object recognition may be time consuming since various experiments on the features need to be conducted in order to achieve the desired accuracy. Furthermore, prior knowledge on the images such as shape, colors, variation of shapes due to shape deformation or different categories, and variation in colors due to illumination or lighting changes also need to be considered so that they will not reduce the recognition results.

In recent years, Convolutional Neural Network (CNN) has gained popularity in computer vision applications including plant and leaf recognition^{1,6-7}. The application of deep layers in CNN is possible due to the availability of vast amount of data and powerful computer devices. Information on feature and classifier is not needed when applying CNN. On the other hand, large amount of data consumes a long training time. To overcome this limitation, utilizing pre-trained CNN models such as AlexNet¹¹ and GoogLeNet¹², reduce the training time and the huge amount of training data. Furthermore, they arrive to very accurate results that reach 0.99 accuracy. But they

still require relatively a long time for the training and validation processes.

A basic CNN model consists of three basic layers that are convolve layer, ReLu layer and pooling layer. These three layers are called a stack. Having more stacks in a basic CNN may increase the accuracy performance and processing time. Thus, to avoid having a deep CNN with a long processing time, multi-pooling approach has been proposed for multi-font printed Chinese character recognition before the classification layer since this approach is robust to spatial layout variations and deformation in the shape of the characters⁹. This paper tends to apply similar multi-pooling⁹ approach since the leaf images also consist of different spatial layout and shapes.

This paper is organized as follows. Some related work on leaf and plant recognition is explained in Section 2. Experimental results are being discussed in Section 3 while section 4 concludes this research.

2. Related Work

2.1 Convolutional Neural Network (CNN)

The architecture of a typical CNN is structured as a series of layers. A stack of CNN consists of three layers that are convolutional layer, Rectified Linear unit (ReLu) layer, pooling layer and followed by fully connected layer (classification layer)⁸. The convolutional layer extracts features of an image by using a filter that stride over the input image and produces a feature map. Different filter produces different feature map that act as feature detectors. Multiple convolutional layers can form different feature maps to ensure full extraction of various features. ReLu layer replaces all negative pixel values in the feature map with zero. A pooling layer down-sizes the feature map after ReLu layer to reduce the dimensionality. A typical pooling layer is max-pooling that computes the maximum of a local feature map. An average pooling layer takes the average of a local feature map. Neighboring pooling takes input from feature maps that that are shifted or stride by more than one row or columns. This operation reduces the dimension of the feature maps and act as invariance to distortion or small shifts. Fully connected layer performs the classification process. A simple illustration of the CNN layers is shown in Fig. 1.

A pre-trained CNN model like AlexNet, also called

transfer learning model, is where knowledge is learnt from training large amount of datasets. It won the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in 2012. AlexNet consists of 25 layers that combine a few stacks of convolutional layers and fully connected layers¹¹. An illustration of AlexNet layers is shown in Figure 2.

Fig. 1. An illustration of CNN layers.

Fig. 2. An illustration of the architecture of AlexNet¹¹.

GoogLeNet is the winner of ILSVRC 2014 competition. It achieved a top-5 error rate of 6.67%¹². This was very close to human level performance which the organizers of the challenge were forced to evaluate. It has been realized that it was actually a difficult task even for a human to perform the image recognition 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 but the number of parameters is reduced from 60 million (AlexNet) to 4 million (GoogLeNet). Figure 3 illustrates the different layers of GoogLeNet¹².

Fig. 3. Illustration of the layers of GoogleNet¹².

2.2 Multi-maxpooling Shallow CNN

The architecture of the CNN used in this study is shown in Fig. 4. It consists of two stacks of layers followed by additional maxpooling layers, fully-connected layer, softmax layer and classification layer. Each stack of layers consists of one convolutional layer, one ReLu layer and one maxpooling layer. The first convolutional layer in the first stack filters the 227x227 input images with 16 kernels of size 9x9 with a stride of 3 pixels. The second convolutional layer in the second stack takes the output of the first maxpooling layer and filters it with 10 kernels of size 9x9 with a stride of 2 pixels. Additional maxpooling layer is being added after the second stack and the accuracy performance is recorded.

Fig. 4. Illustration of CNN architecture used in this study.

2.3 Leaf Dataset

There are a few publicly available leaf datasets that have been tested by researchers such as Flavia¹⁰ and Leafsnap². Flavia dataset consists of 33 species and the leaf images were captured with white background. Fig. 5 illustrates some sample images from Flavia dataset. Fig. 6 illustrates some sample images from Leafsnap dataset where it consists of 52 species. This research tests the images of the leaves from Leafsnap dataset since they have more number of species and they are more challenges since there are variations of images even though they come from the same species

The Leafsnap dataset consists of images with different sizes and orientations, with some deformed images and shadows. Fig. 7 shows some sample images from Leafsnap dataset from two different species. By looking at Fig. 7, we can see that a robust approach is highly needed in order to perform the leaf recognition since the leaf images from the same species have different shapes. The images were captured with different background color as it represents the natural background that may be in various colors due to illumination changes. Some of the images also consist of slight shadows. Data augmentation is performed on each species where they are being transformed into various sizes and orientations to produce a total of 5200 images.

Fig. 5. Sample leaf images from Flavia dataset¹⁰.

Fig. 6. Sample leaf images from Leafsnap dataset².

Fig. 7. Some sample leaf images with different orientations and shapes from two different species from Leafsnap dataset².

3. Experimental Results and Discussions

The experiment is conducted using Matlab 2018a where eighty percent of the images were randomly selected for training and the remainder is for testing. The leaf images were resized to 227x227 pixels. Since AlexNet has less number of layers compared to GoogLeNet, it produces faster result compared to GoogleNet. But, GoogLeNet achieves slightly more accurate result compared to AlexNet. The results are shown in Table 1.

Table 1 also illustrates the experimental results for basic CNN with the addition of maxpooling layers. The numbers in the square bracket represent the parameter values of convolve layer and maxpooling layer. The additional values outside of the square brackets are the values of the extra maxpooling layers that are being added to the basic CNN. A few filter sizes and stride values have been tested but the values in Table 1 were the ones that produce significant results based on our experiments and the memory constraints of our computer. The accuracy is computed as the sum of the correct predicted images divided by the sum of the total tested images.

By looking at Table 1, we can see that the results for 1 stack show an increase in the accuracy as the number of maxpooling layer is added to the CNN but it begins to decrease when the maxpooling layer is 5. When another stack is being added along with the additional number of maxpooling layers, the values of the accuracy increases. The best accuracy results obtained is 0.9702 which is

relatively close to the results produced by AlexNet and GoogLeNet, and yet at a much faster rate. This shows that the additional maxpooling layers can still recognize the leaf images with various orientations and deformations as good as AlexNet and GoogLeNet. The results also illustrate that a small size of pooling layer is robust to various shapes and layout.

This research has also captured ten different images from 42 different species of the Malaysian medicinal herb plants. Then, data augmentation is performed where the images were transformed into various sizes and orientations to produce 4200 images. Fig. 8 illustrates some sample images of the Malaysian medicinal herb leaves from three species. AlexNet, GoogLeNet and basic CNN with multi-maxpooling layers have been applied to our own leaf dataset and the results produced were 0.91, 0.99 and 0.90 respectively. This shows that the additional multi-maxpooling layers are robust and can achieve accurate results as the pre-trained CNN models.

Table 1. Accuracy results for leaf recognition for Leafsnap dataset.

Stack of layers	Accuracy	Processing time (seconds)
Basic CNN : 1 stack		
[9x16, 3x3]	0.026	46
[9x16, 3x3], 2x2	0.0327	47
[9x16, 3x3], 2x2, 2x2	0.0875	47
[9x16, 3x3], 2x2, 2x2, 2x2	0.2106	52
[9x16, 3x3], 2x2, 2x2, 2x2, 2x2	0.7298	188
[9x16, 3x3], 2x2, 2x2, 2x2, 2x2, 2x2	0.6529	214
Basic CNN: 2 stacks		
[9x16, 3x3],[9x10, 2x2]	0.8481	274
[9x16, 3x3],[9x10, 2x2], 2x2, 2x2	0.9548	398
[9x16, 3x3],[9x10, 2x2], 2x2, 2x2, 2x2	0.9702	523
[9x16, 3x3],[9x10, 2x2], 2x2, 2x2, 2x2, 2x2	0.8317	543
AlexNet	0.9962	1526
GoogLeNet	0.9971	2092

Fig. 8. Some sample images of Malaysian medicinal herb leaves from three species.

4. Conclusions

In this paper, we propose additional multi-maxpooling layers before the fully-connected layer that makes the basic CNN robust to images with different orientations, shapes or formations and even with shadows. The results indicate that very good accuracy performance can still be achieved with less number of training images and shallow layers utilizing basic CNN with multi-maxpooling. Future work is to test with other datasets and apply such approach for mobile devices.

Acknowledgment

The authors would like to thank Institute of Research Management & Innovation (IRMI), Universiti Teknologi MARA (UiTM), Shah Alam, Selangor, Malaysia, for sponsoring this research under 600-IRMI/MyRA 5/3/LESTARI (060/2017) grant.

References

- (1) Wang-Su Jeon and Sang-Yong Rhee: “Plant Leaf Recognition using a Convolutional Neural Network”, *International Journal of Fuzzy Logic and Intelligent Systems*, Vol. 17, No. 1, pp. 26-34, March 2017.
- (2) Neeraj Kumar, Peter N. Belhumeur, Arijit Biswas, David W. Jacobs, W. John Kress, Ida Lopez, and Joao V. B. Soares: “Leafsnap: A Computer Vision System for Automatic Plant Species Identification”, *European Conference on Computer Vision (ECCV)*, pp. 502-516, 2012.
- (3) Nursuriati Jamil, Nuril Aslina Che Hussin, Sharifalillah Nordin and Khalil Awang: “Automatic Plant Identification: Is Shape the Key Feature?”, *IEEE International Symposium on Robotics and Intelligent Sensors (IRIS)*, *Procedia Computer Science* 76, pp. 436-442, 2015.
- (4) Milan Sulc and Jiri Matas: “Texture-based Leaf Identification”, *European Conference on Computer Vision (ECCV)*, pp. 185-200, 2014.
- (5) Zaidah Ibrahim, Nurbaity Sabri and Nur Nabilah Abu Mangshor, “Leaf Recognition Using Texture Features for Herbal Plant Identification”, *International Journal of Electrical Engineering and Computer Science (IJECS)*, Vol. 9, No. 1 2018, pp.152-156.
- (6) T L I Sugata and C K Yang: “Leaf App: Leaf recognition with deep convolutional neural networks”, *IOP Conf. Series: Materials, Science and Engineering* 273, pp. 1-6, 2017.
- (7) S. H. Lee, C. S. Chan, P. Wilkin and P. Remagnino, “Deep-plant: Plant identification with convolutional neural networks”, *IEEE International Conference on Image Processing (ICIP)*, 2015.
- (8) Alexandra Toma, Liviu Daniel Stefan and Bogdan Ionescu: “UPB HES SO@PlantCLEF 2017: Automatic Plant Image Identification using Transfer Learning via Convolutional Neural Network”, 2017.
- (9) Zhuoyao Zhong, Lianwen Jin and Ziyong Feng : “Multi-font Printed Chinese Character Recognition using Multi-pooling Convolutional Neural Network”, *13th International Conference on Document Analysis and Recognition (ICDAR)*, pp. 96-100, 2015.
- (10) Stephen Gang Wu, Forrest Sheng Bao, Eric You Xu, Yu-Xuan Wang, Yi-Fan Chang and Qiao-Liang Xiang: “A Leaf Recognition Algorithm for Plant Classification using Probabilistic Neural Network”, *IEEE 7th International Symposium on Signal Processing and Information Technology*, 2007.
- (11) A. Krizhevsky, I. Sutskever, G. E. Hinton: “Imagenet classification with deep convolutional neural networks”, in: F. Pereira, C. Burges, L. Bottou, K. Weinberger (Eds.), *Advances in Neural Information Processing Systems 25*, Curran Associates, Inc., pp. 1097-1105, 2012.
- (12) Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich: “Going Deeper with Convolutions”, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.