Evaluation of basic convolutional neural network and bag of features for leaf recognition

Nurul Fatihah Sahidan, Ahmad Khairi Juha, Zaidah Ibrahim
Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia

Article Info	ABSTRACT
	This paper presents the evaluation of basic Convolutional Neural Network (CNN) and Bag of Features (BoF) for Leaf Recognition. In this study, the performance of basic CNN and BoF for leaf recognition using a publicly available dataset called Folio dataset has been investigated. CNN has proven its powerful feature representation power in computer vision. The same goes with BoF where it has set new performance standards on popular image classification benchmarks and has achieved scalability breakthrough in image retrieval. The feature that is being utilized in the BoF is Speeded-Up Robust Feature (SURF) texture feature. The experimental results indicate that BoF achieves better accuracy compared to basic CNN.

Corresponding Author:

Zaidah Ibrahim,
Faculty of Computer and Mathematical Sciences,
UniversitiTeknologi MARA,
Shah Alam, Selangor, Malaysia.
Email: zaidah@tmsk.uitm.edu.my

1. INTRODUCTION

Several parts of a plant can be used by a botanist in order to recognize a plant. This includes flowers, leaves, and roots. However, leaves are the most widely used as it is more convenient to be used and the results are great [1]. The purpose of identifying plants is to categorize the plants for recording purposes. The process of identifying a plant using leaves is an easy task for botanists as they can simply recognize it using their senses [2]. On the contrary, for machines to achieve the same recognition results requires performing image-processing techniques to extract visual information and compare them to existing sets of data [3]. Structured learning or better known as deep learning, has been recognized as a new area in computer vision that has been reported to produce excellent results [4].

Deep learning, a class of machine-learning techniques used to extract characteristics of data, and CNN (Network Neural Convolutional), a series of artificial neural networks that have been expanded into space using shared weights, have been found to be suitable for computer vision tasks [5]. Within the past few years, deep learning algorithms particularly convolutional neural networks (CNN) have proven their much powerful feature representation power in computer vision [6].

Convolutional Neural Network has resulted in ground breaking decisions over the last decade in various fields related to pattern recognition; from image processing to voice recognition [7]. CNN's capabilities have become a known and used in voice analysis [8], image classification [9], scene classification [10], vehicle recognition [11], fruit classification and ripeness grading recognition [12], and food recognition [13].

Bag of Features (BoF) model is also widely implemented in image processing and classification tasks [14]. It is a machine learning model and an integration of Bag of Words (BoW) which makes it suitable for image classification [15], for example, in breast histopathology images recognition and Arabic

handwritten word recognition [16]. In this paper, a comparison between CNN and BoF is being conducted in order to analyze the accuracy performance of leaf recognition between the two models.

This paper presents evaluation of basic Convolutional Neural Network and Bag of Features for Leaf Recognition. In this study, comparison is made to determine which model is the most appropriate to recognize the leaf from Folio dataset. Research about leaf recognition has been conducted by several researchers using various techniques. One technique used is Support Vector Machine (SVM) with texture features and the result achieved is 99% accuracy [2]. With data augmentation, the accuracy of 99.04% using AlexNet and 99.42 using GoogleNet are obtained [16]. Besides that, shape features and colour histogram with k-nearest neighbour classifiers have been applied with 87.2% accuracy [3]. Since the results from using this dataset has been very positive, this dataset has been chosen to be experimented in this research.

2. RESEARCH METHOD

2.1. Bag of Features (BoF)

Bag-of-Features (BoF) represents the images by instances of local features extracted from the image. This framework can be perceived into a two-level framework. The first level is associated to the pixel intensity of the image and extraction of local features type. For the second level, it consists of two part which are encoding and pooling [16]. The encoding part converted local features into code books, in which the most representative visual vocabulary patterns are coded as visual or code words using codebook learning. Then, a histogram or feature vector is produced through an easy frequency analysis of each codeword inside that image in the pooling part [15]. Figure 1. shows the general structure of BoF framework.

In this project, Speeded up Robust Features (SURF) has been used in BoF because the performance of this feature is excellent and only require low computational cost [18]. It is an image detector and descriptor that is based on Hessian matrix measures. It uses a 2D Haar wavelet transform for a descriptor that uses only 64 dimensions leading to quick feature extraction [18].

For BoF training, the strongest features from each category are set to 80 percent. Based on the result, the average accuracy is 0.85 which shows that by using this method, the accuracy is more than 80 percent. The clustering of the data has been completed on the 20th iteration in which it is about 4.39 seconds/iteration. Figure 2 shows the Visual Word graph for BoF based on the data used in this project.
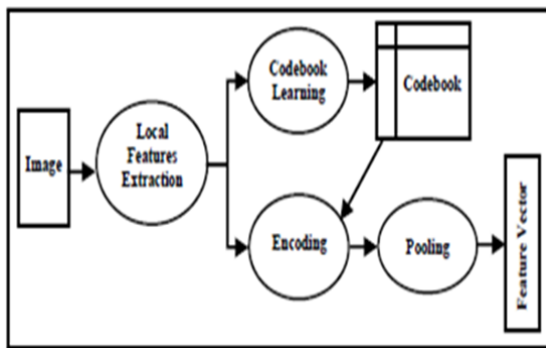

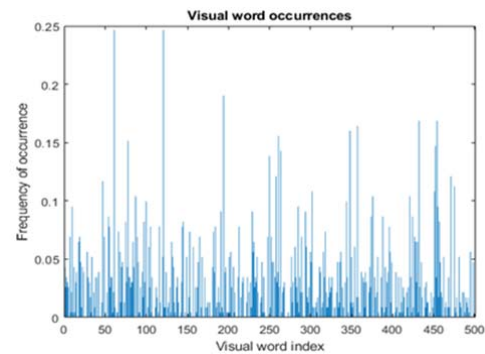Figure 1. General structure of BoF framework [16]

Figure 2. Visual Word graph based on the BoF in this project

Literature review that has been done author used in the chapter "Introduction" to explain the difference of the manuscript with other papers, that it is innovative, it are used in the chapter "Research Method" to describe the step of research and used in the chapter "Results and Discussion" to support the analysis of the results [2]. If the manuscript was written really have high originality, which proposed a new method or algorithm, the additional chapter after the "Introduction" chapter and before the "Research Method" chapter can be added to explain briefly the theory and/or the proposed method/algorithm [4].

2.2. Convolutional Neural Network (CNNs)

Convolutional Neural Network (CNNs) consists of four types of layers which are convolution layers, pooling layer, Rectified Linear unit (ReLu) layer and fully connected layers. Convolution layers

extract the input of an image by using convolution operation and produce a feature map [1]. Multiple convolutional layers can be applied for different feature maps as well. This method is to ensure complete extraction of various features. Next, pooling layer lower the size of the feature maps. This process makes the input robust against noise and distortion [6]. Neural networks and CNN particularly rely on the third layer which is the activation function. CNN may use specific functions such as ReLUs functions to efficiently implement non-liner triggering. All negative pixel values in the feature map are replaced with zero by using ReLu layer [3]. Fully connected layer which is the last layer, total the weightage of previous layer of features to determine the output.

Figure 3 shows the CNN architecture that extracts features by using convolution technique on the input image, resize the feature map during pooling layer and classifies it in the fully connected layer. The first convolution layer usually extracts the low-level features such as edges while the second convolution layer extracts the high-level features such as the shape.
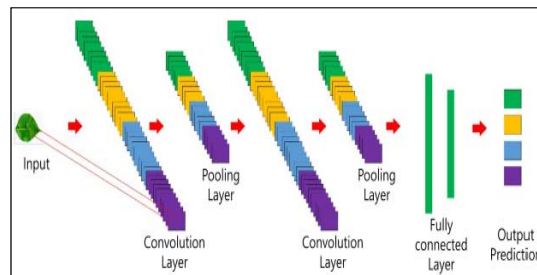

Figure 3. Structure of Basic CNN [19]

3. RESULTS AND ANALYSIS

The laptop used to run the CNN and BoF for this project was Lenovo with Windows 10 as the operating system, Intel Core i7 processor, and an 8.00 GB RAM while the software used was Matlab 2018a. The dataset used is Folio Leaf Data Set [19]. Leaves pictures are taken from plants on the farm of the University of Mauritius and nearby locations. There are 32 categories of plant and for each category 20 images of leaves are experimented. All the images are resized into 224 by 224 pixels to ensure the consistency of the data for each method. Figure 4 shows sample images of Folio Leaf dataset from all the 32 categories.

Figure 4. Sample images of Folio Leaf dataset

Experiments were conducted by changing the number of layers, the values of the parameters in the convolve layer, pooling layer and the learning rate. The purpose is to determine the best combination of values to produce the highest accuracy for leaf recognition from Folio dataset. The result of the experiments was recorded in Table 1. By referring to Table 1. the first column indicates the number of stacks of layers

where a stack consists of one convolve layer, one max-pooling layer and one ReLu layer. In column Convolve layer, the first number in the square bracket represents the size of the convolve filter while the second number represents the number of convolve filters. The third layer represents the size of max-pooling filter and the number of stride. The number of epoch and learning rate is shown in column 4. The number of epoch determines the number of repetitions of all the training data while the learning rate is the amount of adjustment that is being made to the weights during the training process. By looking at Table 1. the best accuracy is achieved when there are three stacks of layers, but the accuracy starts to decrease when the number of stacks is more than three.

Table 1. Experimental results on parameter tuning for basic CNN

No of Stack of Layers	Convolve Layer	Pooling layer and Stride	Epoch, Learning Rate	Accuracy (%)	Total Time/s
1	[3,16]	3	10, 0.001	71.92	6min 32s
	[5,20]	2	10, 0.0001	65.62	8min 23s
2	[3,16], [3,16]	3	10, 0.001	79.81	4min 27s
	[3,80], [3,64]	2	10, 0.001	73.82	15min 40s
3	[3,16], [3,16], [3, 32]	3	10, 0.001	76.66	4min 44s
	[5,20], [3, 20], [3, 16]	3	10, 0.001	82.03	3min 35s

Figure 5. shows training progress for convolve layer [5,20], [3,20], [3,16], pooling layer and stride 3, epoch 10 and learning rate 0.001. By using these parameter values the accuracy result reached at 82.03 with elapsed time of 3 minutes and 35 seconds which is faster compare to the other results. These layers and parameters is much more accurate compare to others.
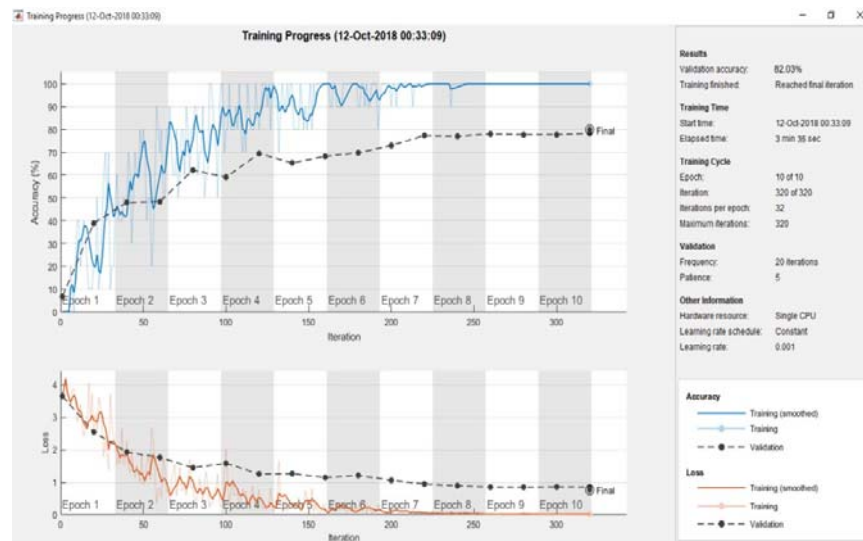

Figure 5. Training Progress

Table 2. Shows an overview of the accuracy performance of basic CNN compared to BoF based on Folio Leaf Dataset. By looking at Table 2, we can see that BoF is better than basic CNN but it took a longer time to achieve this result. This is because extracting of the SURF features is longer compare to the time to extract the low-level and middle level features by the basic CNN.

Table 2. The Performance overview for basic CNN and BoF for FolioLeaf Dataset

Model	Basic CNN	BoF
Validation Accuracy	0.82	0.85
Elapsed Time (s)	177	276

4. CONCLUSION

This paper has presented the evaluation of basic CNN and BoF for leaf recognition. In this paper, the accuracy performance for leaf recognition based on Folio leaf dataset is compared between basic CNN and BoF. The experimental results show that basic CNN achieves a lower accuracy rate compared to BoF since it requires a huge amount of data compared to BoF. Therefore, if the number of data is limited, BoF still provides a good result and preferable compared to CNN. For the future research, we plan to enhance the CNN architecture and increases the number of datasets to obtain a more accurate and faster results.

ACKNOWLEDGEMENTS

REFERENCES

[1] Z. Ibrahim, *et al.*, "Multi-maxpooling Convolutional Neural Network for Medicinal Herb Leaf Recognition," *Proceedings of the 6th IIAE International Conference on Intelligent Systems and Image Processing, Shimane, Japan*, pp. 327-331,September 2018.

[2] Z. Ibrahim,*et al.*, "Leaf Recognition using Texture Features for Herbal Plant Identification," *Indonesian Journal of Electrical Engineering and Computer Science*, Vol. 9, No. 1, pp. 152-156, January 2018.

[3] Munisami. T, *et al.*, "Plant leaf recognition using shape features and colour histogram with k-nearest neighbour classifiers," *Procedia Computer Science (Elsevier) Journal*. 58, pp. 740-747, 2015.

[4] L. Deng and D. Yu, "Deep Learning: Methods and Applications,"*Foundations and Trends in Signal Processing*, vol. 7, no. 3–4, pp. 197–387, 2013.

[5] E. Nishani And B. Cico, "Computer Vision Approaches Based On Deep Learning And Neural Networks," *2017 6th Mediterranean Conference On Embedded Computing (Meco), Bar, Montenegro*, 11-15 June 2017.

[6] M. Afiq, *et al.*, "Deep Learning for Roman Handwritten Character Recognition," *Indonesian Journal of Electrical Engineering and Computer Science*. Vol. 12, no. 2, pp. 455-460, November 2018.

[7] S. Albawi, et al., "Understanding of a convolutional neural network," *2017 International Conference on Engineering and Technology (ICET), Antalya*, pp. 1-6,2017.

[8] R. Durratun, *et al.*,"Performance Evaluation for Vision-Based Vehicle Classiffication Using Convolutional Neural Network," *International Journal of Engineering & Technology*, 7 (3.15) 86-90, 2018.

[9] T. Guo, *et al.*, "Simple convolutional neural network on image classification," *2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA), Beijing*, pp. 721-724,2017.

[10] G. Cheng, *et al.*, "Scene classification of high resolution remote sensing images using convolutional neural networks," *IEEE International Conference in Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 767-770, 2016.

[11] J. Sochor,*et al.*, "BoxCars: 3D Boxes as CNN Input for Improved Fine-Grained Vehicle Recognition," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV*, pp. 3006-3015, 2016.

[12] Z. Ibrahim, *et al.*, "Palm Oil Fresh Fruit Bunch Ripeness Grading Recognition Using Convolutional Neural Network," *Journal of Telecommunication, Electronic & Computer Engineering,* vol. 9, no. 3-2, pp. 109-113, 2018.

[13] C. Chin, *et al.*, "Smartphone-based food category and nutrition quantity recognition in food image with deep learning algorithm," *2016 International Conference on Fuzzy Theory and Its Applications (iFuzzy), Taichung*, pp. 1-1, 2016.

[14] R.A. Chen, et al., "Enhanced Bag-of-Features model for image classification," *2014 IEEE Workshop on Advanced Research and Technology in Industry Applications (WARTIA), Ottawa, ON*, pp. 1195-1198, 2014.

[15] S. H. Bhandari, "A bag-of-features approach for malignancy detection in breast histopathology images," *2015 IEEE International Conference on Image Processing (ICIP), Quebec City, QC*, pp. 4932-4936, 2015.

[16] M. O. Assayony and S. A. Mahmoud, "Integration of Gabor Features with Bag-of-Features Framework for Arabic Handwritten Word Recognition," *2017 9th IEEE-GCC Conference and Exhibition (GCCCE), Manama*, pp. 1-4, 2017.

[17] P. Pawara, et al., "Data Augmentation for Plant Classification," *Advanced Concepts for Intelligent Vision Systems: 18th International Conference, ACIVS 2017, Antwerp, Belgium*, Proceedings, September 18-21, 2017.

[18] A. K. Khan, et al., "Evaluation of SIFT and SURF using Bag of Words Model on a Very Large Dataset," *Sindh Univ. Res. Jour. (Sci. Ser.),* Vol.45 (3), pp. 492-495, 2013.

[19] J.W. Su and R.S. Yong, "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.

[20] Folio Data Set. Rerieved from http://archive.ics.uci.edu/ml/datasets/Folio.

BIOGRAPHIES OF AUTHORS

Nurul Fatihah Binti Sahidan is a Master's student in Computer Science (Web Technology) at UniversitiTeknologi MARA, Shah Alam, Selangor, Malaysia. Her area of interests is image processing, data science and big data, multimedia computing and web-based technology.

Ahmad Khairi Bin Juha is a Master's student in Computer Science (Web Technology) at UniversitiTeknologi MARA, Shah Alam, Selangor, Malaysia. His area of interests is image processing, data science and big data, business computing and web-based technology.

Zaidah Ibrahim is an Associate Professor at the Faculty of Computer and Mathematical Sciences, UniversitiTeknologi MARA, Shah Alam, Selangor, Malaysia. She has been teaching courses related to Artificial Intelligence for over ten years. She is actively involved in research and publications under Digital Image, Audio and Speech Technology (DIAST) research interest group that include text and object recognition