Medicinal Plant identification in the wild by using CNN

1st Trung Nguyen Quoc
Faculty of Information Technology
Ho Chi Minh Open Universty, VietNam
trungnq.188i@ou.edu.vn

2nd Vinh Truong Hoang
Faculty of Information Technology
Ho Chi Minh Open Universty, VietNam
vinh.th@ou.edu.vn

***Abstract*—Plant identification based on deep learning received many attention and effort from the research community with many promising results. It becomes an active trend in the recent years. We apply Convolutional Neural Network (CNN) to recognize Vietnamese medicinal plant images in this paper. Different frameworks are evaluated such as: VGG16, Resnet50, InceptionV3, DenseNet121, Xception and MobileNet. The highest accuracy reached by Xception with 88.26%. We might think this approach will greatly contribute to the discovery and conservation of valuable medicinal plants.**

***Index Terms*—VietNam medicinal plant, Deep learning, VGG16, Resnet50, InceptionV3, DenseNet121, Xception, MobileNet, VNPlant-200.**

I. INTRODUCTION

Plants is a main source of food as well as a major source of medicine. We all consumed plants and their related products to support our daily nutrition. Most of the world's population relied on plant-based measures to meet the healthcare needs, such as their traditional medicine, or as complementary and alternative medicine. Human needs to have a good understanding of plant to distinguish and recognize different plant species. This helps to promote the pharmaceutical industry and protect the ecosystem. Thus, the sustainable development also allows to increase the agricultural productivity.

The advancement of computer vision methods has been introduced for plant recognition automatically. An industrial application on mobile devices have been highly appreciated by users such as LeafSnap [1] or Pl@ntNet [2] for identifying plants. Among them, deep learning method is gained a good achievement. Most researchers attempted to extract local features from leaves, flowers, and bark for plant classification by using characteristic variation of leaves. There are several datasets are released, for example, the Swedish leaves dataset [3] with 75 leaves per species, Flavia dataset [4] with 1,907 leaf images of 32 different species. However, these datasets consist of single leaf which were scanned on a simple background, so it gives very high results by using hand-crafted algorithms to extract features. Moreover, a single image of plant in a complex background composed with many leaves, flowers or trees is needed to further investigate. However, hand-crafted features are limited in this case and many preprocessing steps need to consider. The identification results are very poor by using SURF and SIFT descriptors [5] on the VNPlant-200 dataset. These images are naturally acquired a real environment.

In recent years, the CNN model has had tremendous success, it has been playing a principal role to understand the various features of images. We apply CNN models, such as Resnet50 and DenseNet to enhance the accuracy rate on the VNPlant-200 dataset [5]. This paper is organized as follows. Section 2 presents the related works. Next, section 3 and section 4 introduce the method and experimental results, respectively. Section 5 gives perspective and conclusion.

II. RELATED WORKS

Computer vision methods help people to recognize plant accurately and easily. We simply need a smartphone to obtain a plant's information from queried images. Plant identification applications are first based on many elements such as a plant's leaves, roots, flowers, and bark. The authors in [6] identified plant via its bark by proposing a new descriptor, namely Gradient Local Binary Patterns. Fekri-Ershad [7] improves Local Ternary Pattern and Multi layer network for bark texture recognition. Zhang et al. [8] suggested a method by fusing SVD with sparse representation (SR) for tree classification. The reliability is not high when identifying plant based on a single part such as flowers or leaves. Shanwen Zhang et al. [9] introduced a method of combining flowers and leaves for identification. The team used the Modified local discriminant Canonical Correlation analysis (MLDCCA) method to extract the characteristic from the two flower and leaf organs, resulting in a high accuracy of 92.73% (Euclidean distance, respectively). The leaves of tree contain valuable information for plant identification. In the work of two authors F. Mostajer Kheirkhah and H. Asghari [10], they applied GIST texture features to extract local features. The main characteristics will then be selected using the Principal Component Analysis method. In the classification step, the authors use three algorithms that are KNN, SVM and PatternNet network. The best results on the Flavia data set when using GIST in combination with PCA were 98.7%. Another study [11] that used texture of bark and its color to characterize textual information.

Deep learning based on CNNs provided a promising result for many related tasks in machine vision in the recent years [12]. The use of deep learning methods will give more accurate results than conventional methods in a natural environment.

Lee et al. [13] used deep learning method by using and combining two specific extraction methods, Convolutional Neural Network (CNN) based on conv-max pool-convaverage pool-fc-fc and Deconvolutional Network (DN) architectures. CNNs do not require to use hand-crafted features as traditional approach. Geetharamani el al. [14] proposed a deep convolutional neural network framework that achieved a good result for leaf plant disease recognition. Lee et al. [15] introduced a CNN method to taxon classification based on leaf images of 44 species.

Alex Olsen et al. [16] have had grass identification studies all over Australia. They used an automated robot with a camera to capture images of grass in parts of Australia by using Inception-v3 and ResNet-50 frameworks on the actual dataset of weed. Xiao el al. [17] studied on real-world plant species identification by using deep CNN on the BJFU100 dataset. All images of this database are captured by mobile devices at Beijing Forestry University campus.

III. Method

We apply the CNN models, advanced computer vision solutions and transfer learning method to classify images of Vietnamese medicinal plants.

A. Convolutional Neural Network

Deep Learning (DL) has grown extremely fast which is mainly based on a huge training data. The most used database for DL was ImageNet with 1.2M images with 1,000 different classes. Many of the DL models won the ImageNet LSVRC competitions as AlexNet [18], GoogLeNet [19], VGG [20]. Other alternatives and more efficient advanced architectures have been proposed, including DenseNet [21], FractalNet [22], Inception units [19], [23], and Residual Networks [24]. It can achieve a better performance than conventional methods. CNN's architecture included an input layer, multiple alternating convolution layer, aggregated or sub layer, and non-linear layers. The second layer included a small number of fully connected layers while the final layer is usually a softmax classification layers (as figure 1).

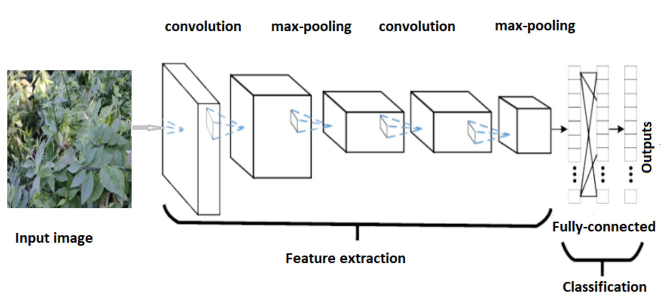

Fig. 1. An illustration of CNN.

B. State of the art models

VGG16: is a first deep neuron architecture after a success of Alexnet. VGG team did stack of many convolutional and full connected layers together and archived better performance by utilizing the smallest inception filter of 3×3 convolutional filters. VGG16 achieves the top-5 accuracy on the ImageNet [25] dataset which consists of more than 14 million images of 1,000 classes. Figure 2 presents the overall of VGG16 architecture.

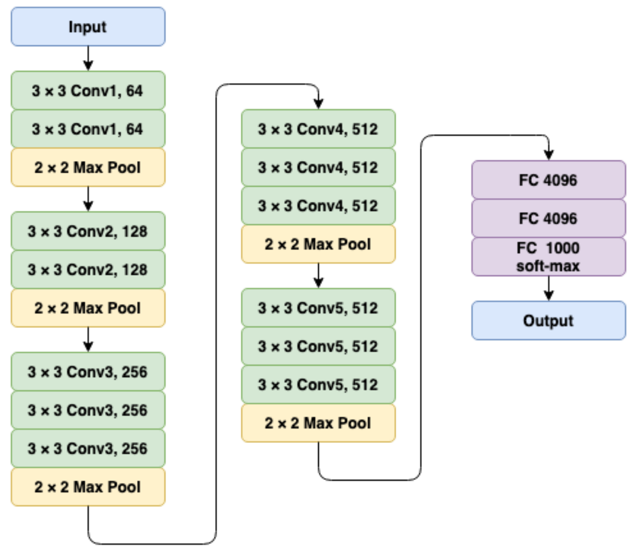

Fig. 2. The architecture of VGG16.

Inception module's idea is to concatenate many optimal local structures with high correlation analyzed from previous layer. It used a various size of convolutional operator such as 1×1, 3×3, 5×5 and fusing techniques. They are likely to be types of multi-scale presentation in pyramid scheme. For the Inception with reduction design, it allows increasing several nodes at each layer without effecting to next computation layer. This network is totally extended to 22 layers with pre-trained parameters (see figure 3).

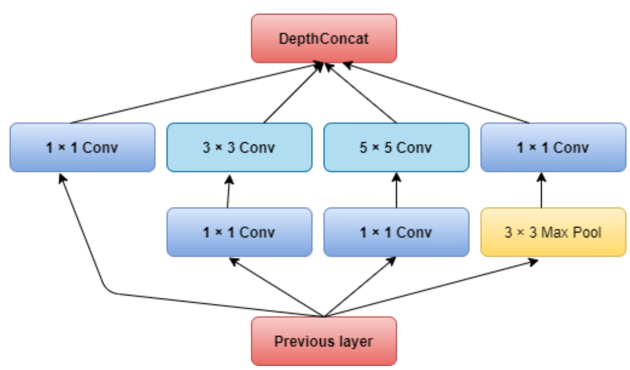

Fig. 3. Inception block.

Resnet: declared that originally $H(x)$ is predicted mapping function which learns a mapping from input to output. Alternatively, let define another mapping $F(x) = H(x) - x$ and so again $H(x) = F(x) + x$. Now $H(x)$ – residual function - is easier to optimize with reference to the layer input. This formula is also a type of shortcut connection which borrows in Long Sort Term Memory network by bringing a flow of memory from an input to output layer. For the ResNet-50 model, simply replace each two-layer residual blocks with a three-layer bottleneck block which uses 1×1 convolutions. This allows to reduce and subsequently restore the channel depth and computational load when calculating the 3×3 convolution (see figure 4).

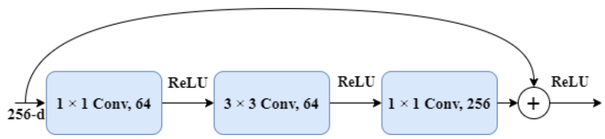

Fig. 4. Resnet 50 block.

DenseNet: is a recent proposed network for visual object identification. It is like Resnet by composing of dense blocks and transition layers. Stacked dense block- transition layers- dense block- transition layers. With traditional CNN, if we have L layer, we have L connection, and in densenet, we have $L(L+1) = 2$ connection. The Fig. 5 shows the DenseNet architecture.

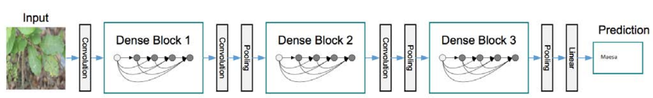

Fig. 5. The architecture of DenseNet.

Xception [26] is proposed to replace Inception module by using depth wise separable convolutions. This network began separating the two slightly by using 1×1 convolution to project the original input into smaller input spaces and using different type of filter to transform those smaller 3D blocks of data.

MobileNet: in recent years, AI built-in mobile device is extremely interesting to research community. Although achieving a great computation and storage capability for smart phone, it cannot compare with computer. Light neuron network with accepted accuracy is necessary. Howard at al. [27] introduced the MobileNet for mobile and embedded device which allowed to adjust hyper-parameter to be suitable for specified applications.

IV. Experimental Results

A. Data Preparation

The VNPlant-200 dataset contains total of 20,000 images of medicinal plant which belongs to 200 categorizes. All labeled images from this dataset were partitioned into $50 : 10 : 40$ corresponding to the training, validation, and testing sets. The images size of 256×256 pixels are considered randomly augmented for each training and validation subset. Next, the images are flipped horizontally and then cropped to the images size of 224×224 pixels. The Table 1 shows several classes of the VNPlant-200 dataset.

TABLE I
Several selected species of VNPlant-200 dataset.

(a)- Bischofia trifoliata
(b)- Costus speciosus
(c)- Jatropha gossypiifolia
(d)- Dioscorea persimilis
(e)-Euphorbia pulcherrima
(f)- Pandanus tectorius
(g)- Phyllanthus Reticulatus Poir
(h)- Pluchea pteropoda Hemsl
(i)- Tacca subflabellata
(j)- Psychotria reevesii Wall

B. Experimental setup

Six deep learning models, such as VGG16, VGG19, Resnet50, InceptionV3, Densenet121, Xception and MobileNetV2 are applied to train on VNPlant-200 dataset. All available models in Keras and Tensorflow 2.2 have initialized as pre-trained model to learn new leaf features. Their original ImageNet trained architectures have been slightly modified to classify 200 species. The experiment and analysis are carried out on the PC cadenced at i7 processor, 16GB RAM and Gefore RTX 2060 GPU memory. Due to CNN models needed

TABLE II
CLASSIFICATION PERFORMANCE (ACCURACY) BY DEEP LEARNING APPROACHES THE VNPLANT-200 DATASET.

Models	Accuracy
VGG16	76.00
InceptionV3	82.50
MobileNetV2	87.92
Resnet50	88.00
Densenet121	88.00
Xception	**88.26**

TABLE III
CLASSIFICATION PERFORMANCE (ACCURACY) BY USING LOCAL IMAGE DESCRIPTOR ON THE VNPLANT-200 DATASET IN [5].

	SURF	SIFT
Accuracy	21.00	28.00

to be trained by a large data, we have adopted Keras's open source data enhancement techniques. This technique helped to rotate, zoom, and resize the current images into $224 \times 24 \times 3$ dimension. This technique is developed to reduce overfitting for deep learning models.

C. Results

The goal of this research is to propose a CNN-based method for large scale plant classification. The experimental models are VGG16, Resnet50, InceptionV3, Xception, DenseNet121 and MobileNetV2. The pre-trained model on ImageNet dataset are applied with the same configured weights by 200-neuron fully connected layer. The Adam optimization [28] and Early Stopping [29] techniques are applied with initial learning-rate was set to $lr = 0.0001$. During training stage of six deep CNNs, we have required 100 epochs of training to achieve a satisfying validation accuracy. The results are presented in Table 2, which shows that the Xception model reached 88,26% and outperforms other models.

The obtained results are compared with those in table 3. We see that the deep learning-based method significantly outperforms the traditional approach by using hand-crafted features.

V. CONCLUSION

Plant identification plays a major part in the research of medicinal plants and botany. We used deep learning CNN models (VGG16, Resnet50, Inceptionv3, DenseNet121, Xception and MobileNetV2) and the VNPlant-200 are trained on a pre-trained model. We classified 200 different leaf layers and significantly improved the classification performance. Although performance of the system is good enough, we believe that the performance could be enhanced by adding more images and adding more layers.

REFERENCES

[1] Neeraj Kumar, Peter N. Belhumeur, Arijit Biswas, David W. Jacobs, W. John Kress, Ida C. Lopez, and João V. B. Soares. Leafsnap: A Computer Vision System for Automatic Plant Species Identification. In Andrew Fitzgibbon, Svetlana Lazebnik, Pietro Perona, Yoichi Sato, and Cordelia Schmid, editors, *Computer Vision – ECCV 2012*, volume 7573, pages 502–516. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012. Series Title: Lecture Notes in Computer Science.

[2] Alexis Joly, Hervé Goëau, Pierre Bonnet, Vera Bakić, Julien Barbe, Souheil Selmi, Itheri Yahiaoui, Jennifer Carré, Elise Mouysset, Jean-François Molino, Nozha Boujemaa, and Daniel Barthélémy. Interactive plant identification based on social image data. *Ecological Informatics*, 23:22–34, September 2014.

[3] *O. Soderkvist, Computer Vision Classification of Leaves from Swedish Trees, Dissertation, 2001.*

[4] 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. In *2007 IEEE International Symposium on Signal Processing and Information Technology*, pages 11–16, Giza, Egypt, December 2007. IEEE.

[5] Trung Nguyen Quoc and Vinh Truong Hoang. Vnplant-200–a public and large-scale of vietnamese medicinal plant images dataset. In *International Conference on Integrated Science*, pages 406–411. Springer, 2020.

[6] Tuan Le-Viet and Vinh Truong Hoang. Local binary pattern based on image gradient for bark image classification. In Kezhi Mao and Xudong Jiang, editors, *Tenth International Conference on Signal Processing Systems*, page 39, Singapore, Singapore, April 2019. SPIE.

[7] Shervan Fekri-Ershad. Bark texture classification using improved local ternary patterns and multilayer neural network. *Expert Systems with Applications*, 158:113509, November 2020.

[8] Shanwen Zhang, Chuanlei Zhang, Zhen Wang, and Weiwei Kong. Combining sparse representation and singular value decomposition for plant recognition. *Applied Soft Computing*, 67:164–171, June 2018.

[9] Shanwen Zhang, Chuanlei Zhang, and Wenzhun Huang. Integrating leaf and flower by local discriminant CCA for plant species recognition. *Computers and Electronics in Agriculture*, 155:150–156, December 2018.

[10] Fateme Mostajer Kheirkhah and Habibollah Asghari. Plant leaf classification using GIST texture features. *IET Computer Vision*, 13(4):369–375, June 2019.

[11] Sarah Bertrand, Rihab Ben Ameur, Guillaume Cerutti, Didier Coquin, Lionel Valet, and Laure Tougne. Bark and leaf fusion systems to improve automatic tree species recognition. *Ecological Informatics*, 46:57–73, July 2018.

[12] Pierre Barré, Ben C. Stöver, Kai F. Müller, and Volker Steinhage. LeafNet: A computer vision system for automatic plant species identification. *Ecological Informatics*, 40:50–56, July 2017.

[13] Sue Han Lee, Chee Seng Chan, Simon Joseph Mayo, and Paolo Remagnino. How deep learning extracts and learns leaf features for plant classification. *Pattern Recognition*, 71:1–13, November 2017.

[14] Geetharamani G. and Arun Pandian J. Identification of plant leaf diseases using a nine-layer deep convolutional neural network. *Computers & Electrical Engineering*, 76:323–338, June 2019.

[15] S. H. Lee, C. S. Chan, P. Wilkin, and P. Remagnino. Deep-plant: Plant identification with convolutional neural networks. In *2015 IEEE International Conference on Image Processing (ICIP)*, pages 452–456, 2015.

[16] Alex Olsen, Dmitry A. Konovalov, Bronson Philippa, Peter Ridd, Jake C. Wood, Jamie Johns, Wesley Banks, Benjamin Girgenti, Owen Kenny, James Whinney, Brendan Calvert, Mostafa Rahimi Azghadi, and Ronald D. White. DeepWeeds: A Multiclass Weed Species Image Dataset for Deep Learning. *Scientific Reports*, 9(1):2058, December 2019.

[17] Qingguo Xiao, Guangyao Li, Li Xie, and Qiaochuan Chen. Real-world plant species identification based on deep convolutional neural networks and visual attention. *Ecological Informatics*, 48:117–124, November 2018.

[18] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90, May 2017.

[19] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew

Rabinovich. Going deeper with convolutions. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–9, Boston, MA, USA, June 2015. IEEE.

[20] Karen Simonyan and Andrew Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv:1409.1556 [cs]*, April 2015. arXiv: 1409.1556.

[21] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q. Weinberger. Densely Connected Convolutional Networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2261–2269, Honolulu, HI, July 2017. IEEE.

[22] Gustav Larsson, Michael Maire, and Gregory Shakhnarovich. FractalNet: Ultra-Deep Neural Networks without Residuals. *arXiv:1605.07648 [cs]*, May 2017. arXiv: 1605.07648.

[23] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alexander A. Alemi. Inception-v4, inception-resnet and the impact of residual connections on learning. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, AAAI'17, page 4278–4284. AAAI Press, 2017.

[24] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, Las Vegas, NV, USA, June 2016. IEEE.

[25] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, 115(3):211–252, December 2015.

[26] Francois Chollet. Xception: Deep Learning with Depthwise Separable Convolutions. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1800–1807, Honolulu, HI, July 2017. IEEE.

[27] Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *arXiv:1704.04861 [cs]*, April 2017. arXiv: 1704.04861.

[28] Diederik P. Kingma and Jimmy Ba. Adam: A Method for Stochastic Optimization. *arXiv:1412.6980 [cs]*, January 2017. arXiv: 1412.6980.

[29] Yuan Yao, Lorenzo Rosasco, and Andrea Caponnetto. On Early Stopping in Gradient Descent Learning. *Constructive Approximation*, 26(2):289–315, August 2007.