Vietnamese Herbal Plant Recognition Using Deep Convolutional Features

Anh H. Vo, Hoa T. Dang, Bao T. Nguyen, and Van-Huy Pham

Abstract—**Herbal plant image identification is able to help users without specialized knowledge about botany and plan systematics to find out the information of herbal plans, thus it has become an interdisciplinary focus in both botanical taxonomy and computer vision. A computer vision aided herbal plan identification system has been developed to meet the demand of recognizing and identifying herbal plants rapidly. In this paper, the first herbal plant image dataset collected by mobile phone in natural scenes is presented, which contains 10,000 images of 10 herbal plant species in Vietnam. A VGG16-based deep learning model consisting of 5 residual building blocks is used to extract features from the images. A comparative evaluation of seven classification methods using the same deep convolutional feature extraction method is presented. Experiments on our collected dataset demonstrate that deep learning features worked well with LightGBM classification method for herbal plant recognition in the natural environment with a recognition rate of 93.6%.**

Index Terms—**Deep feature, deep learning, herbal plant, plant identification.**

I. Introduction

It is a desire to have an automated plant identification system that helps users without specialized knowledge and in-depth training in botany and plant systematics to find out the information of some herbal plants by taking pictures of the plants to feed into an automated plant recognition system. Computer vision aided plant identification systems have been developed to meet the demand of botanists to recognize and identify unknown herbal plants more rapidly. The core tasks of the systems are image recognition and retrieval, which have attracted much attention from researchers in the field of computer vision.

Studies on the identification of plants have been conducted by many authors and achieved certain results. In the early stages, the authors used low-level features such as shape, color, and texture of leaves to distinguish between species [1]-[5]. Kumar *et al*. [2] implemented the first mobile application for identifying plant species using automated visual recognition tools. This system, called Leafsnap, identifies plant species from photos of leaves. The key of this system is to extract features that represent the curvature of leaf border on multiple scales. The system achieves remarkable performance on the actual image. Cerutti *et al*. [5] presents a method for identifying plant species based on specialized algorithms using plant-inspired descriptors. Focusing on leaf analysis, identification of species is started from the image of a leaf in a complex natural background. A 2-step boundary segment algorithm based on the polygon leaf pattern is implemented to obtain the outline of the leaf. Extracted features are high level geometric descriptors that can be semantically deducted. In [3], Aakif *et al*. proposed a tree identification algorithm in three steps: preprocessing, extraction, and finally sorting. Different leaf characteristics, such as morphological characteristics, Fourier descriptions and a new characteristic are proposed in terms of shape. These characteristics become the input of artificial neural network (ANN). Classifier was trained with 817 leaf samples from 14 different fruit trees and gave an accuracy of over 96%.

It is quite clear that most of the studies mentioned above have focused on the recognition with hand-crafted image features, but there are two limitations in this approach. Firstly, most of these hand-crafted features are low-level image representation, which is easily affected by noise and background. Secondly, the input images should be very clean without any backgrounds, which makes it difficult to use in practical applications. Therefore, in order to be used in practical applications, it requires to design a high-level image representation with less affecting by environment and good for recognition and retrieval in real world plant images. This trend recently attracts more attention in literature [6]-[10]. Barre *et al*. [4] developed a deep learning method to learn distinctive features from leaf images along with a classification for plant species. Authors have demonstrated that learned features from a Convolutional Neural Network (CNN) can provide better features for leaf images than the hand-crafted features. In [11], Sun *et al*. have studied the use of CNN in the identification and query of herbal information. The authors used a CNN for Chinese herbal medicine images. For the recognition, the soft-max loss was used to optimize the recognition network; then for the retrieval problem, the recognition network was fine-tuned by adding a triplet loss to search for the most similar herbal medicine images.

Inspired by the recent progress of deep learning in computer vision, we realize that deep learning methods may provide robust herbal plants image representation. In this paper, we propose to use the Convolutional Neural Network (CNN) for Vietnamese herbal plant image feature extraction together with different classification methods. A comparative

Manuscript received October 20, 2018; revised April 27, 2019

Anh H. Vo and Hoa T. Dang are with the Faculty of Information Technology, Ton Duc Thang University, Ho Chi Minh City, Vietnam (e-mail: vohoanganh@ tdtu.edu.vn, dathaha86@gmail.com,).

Bao T. Nguyen is with the Faculty of Information Technology, University of Education and Technology, Ho Chi Minh City, Vietnam (e-mail: baont@hcmute.edu.vn).

Huy V. Pham is with the AI Lab, Faculty of Information Technology, Ton Duc Thang University, Ho Chi Minh City, Vietnam (e-mail: phamvanhuy@ tdtu.edu.vn).

doi: 10.18178/ijmlc.2019.9.3.811

evaluation of seven classification methods using the same deep convolutional feature extraction method is presented. Experiments on our collected dataset provide an effective solution to choose a classification method suitable for deep learned features in herbal plant image recognition systems.

II. Herbal Plant Recognition

A. Deep Convolutional Feature Extraction

Inspired by deep convolutional feature representation in [12], [13], we use the approach proposed in [12] to extract visual features from herb plant images as the inputs for the classifiers to predict the herb classes. In this approach, the fully connected layers are removed from the original VGG16 model and then the global average pooling operation is applied to each block inside the convolutional layers, as illustrated in Fig. 1. The final feature vector is obtained by concatenating from block 2 to block 5 into a single vector of 1408 dimensions, which is then fed into the classification stage. In this work, we make use the entire plant image instead of cropping into many sub-regions as done in previous works [12] to leverage the appearance characteristics of herb plants and to avoid time-consuming and overfitting of model.

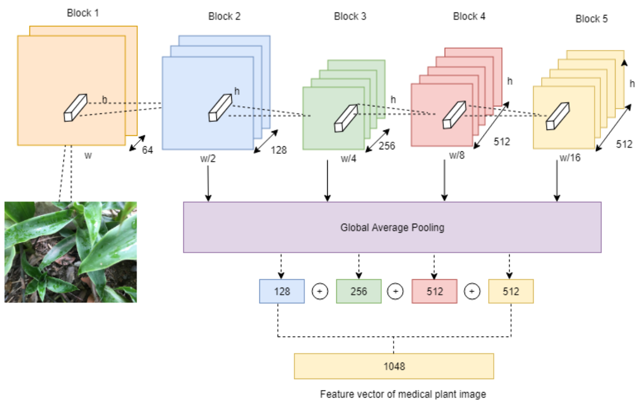

Fig. 1. Deep convolution features are extracted with a modified VGG16 network proposed in [12]. The fully connected layers are removed, and the global average pooling operation is applied to each block inside the convolutional layers.

B. Classification Methods

1) Random forests

Random forests method is one of the most well-known ensemble methods used in both classification and regression problem. A random forests classifier consists of a number of trees, each of which is grown using some form of random tree (e.g. decision tree) until it reaches a leaf node which is considered as the target class. All the posterior probabilities are then averaged, and the argmax is taken as the final prediction of input images. In addition, random forest prevents overfitting based on creating random subsets of features and building smaller trees from these subsets and handle missing feature values.

2) Support Vector Machine (SVM)

SVM is used to classify the herb classes of an input sample. Given a training set of labeled examples $\{(x_i, y_i), i=1,...,k\}$ where $x \in R^n$ and $y_i \in \{1,-1\}$, SVM classifies a new test sample x based on the following functions:

$$f(x) = \operatorname{sgn}\left(\sum_{i=1}^{l} \alpha_i \gamma_i K(x_i, x) + b\right)$$

where α_i are Lagrange multipliers of a dual optimization problem that describes the separating hyperplane; $K(.,.)$ is a kernel function; and b is a threshold parameter of the hyperplane. The training sample x_i (with $\alpha_i > 0$) is called support vectors, and SVM results in a hyperplane that maximizes the distance between the hyperplanes.

3) Logistic regression

One of the most popular probabilistic classifiers is logistic regression whose probabilistic definition is presented as:

$$P(Y=y \mid X=x) = \frac{1}{1+\exp(y(\theta, x))},$$

where y is the class label vector and x is a CNN feature vector of a herb image, which is extracted in the previous step. We used the strategy of one-versus-all for the multiple classification. The maximum likelihood estimation and gradient descent are used as to estimate and optimize the parameters using the equation:

$$\begin{aligned}\theta_{MLE} &= \underset{\theta}{\operatorname{argmax}} \sum_{i=1}^{n} \log \frac{1}{1+\exp(y^{(i)}(\theta, x^{(i)}))}\\ &= \underset{\theta}{\operatorname{argmax}} \sum_{i=1}^{n} -\log(1+\exp(y^{(i)}(\theta, x^{(i)})))\\ &= \underset{\theta}{\operatorname{argmin}} \sum_{i=1}^{n} \log(1+\exp(y^{(i)}(\theta, x^{(i)})))\end{aligned}$$

The equation is solved to find a vector θ minimizing the above objective expression by using the method of gradient descent with the parameters θ_j, each of which is updated in consecutive steps until it becomes smaller than a threshold. α is the learning rate of the parameters as the gradient descent iteration increase.

$$\theta_i \leftarrow \theta_j - \alpha \frac{\delta \sum_{i=1}^{n} \log(1+\exp(y^{(i)}(\theta, x^{(i)})))}{\delta \theta_j}$$

4) Extreme gradient boosting

Extreme gradient boosting is known as XGBoost [14], which is a highly effective and widely used machine learning method. XGBoost algorithm is proposed by Chen and Guestrin [3] which is described as a scalable end-to-end tree-based boosting system. Given a training set $D=\{(x_1,y_1),...,(x_i,y_i)\}$, where $x_i \in R^m$ represents the i^{th} feature and $y_i \in L=\{0,\ldots,10\}$ indicates the class label of the herb plant. XGBoost makes use of a tree-based ensemble model with K additive functions to predict the target label using the following formula:

$$y_i = \sum_{k=1}^{K} f_k(x_i)$$

where $f_k \in F$ is the space of CART regression trees and $\hat{y}_i$ is the predicted label. The set of functions used in the model contributes to the regularized objective function as:

$$L(\phi) = \sum_i l(y_i, y_i) + \sum_k \Omega(f_k)$$

where $\Omega(f) = \gamma T + \frac{1}{2}\lambda|w|^2$

l is a differentiable convex loss function that measures the difference between the prediction $\hat{y}_i$ and the target label y_i.

5) Adaboost

Adaboost is a supervised algorithm based on boosting strategy which learns a strong classifier $H(x_i)$ by combining an ensemble of weak classifiers $h(x_i)$. The weights of training samples determine the probability of being selected for a feature and they are continuously updated in every iteration. If a training sample is accurately classified, then its chance of being used again in the next round is reduced. The weakly classified sample weights are increased while the weights of strongly classifier are decreased.

Input: Given $(x_1, y_1), \ldots, (x_m, y_m)$
$x_i \in X$, $y_i \in \{-1, 1\}$
Initialize weight weak classifier
$h_t: X \to \{-1, 1\}$ with minimum error
w.r.t distribution D_t; $D_1(i) = 1/m$
Output:
The strong classifier $H(x) = sign(\sum_{t=1}^{T} \alpha_t h_t(x))$
for t = 1,..., T
1. Choose $\alpha_t \in R$,
2. Update
$D_{t+1}(i) = \frac{D_t(i) exp(-\alpha_t y_i h_t(x_i))}{Z_t}$
where Z_t is a normalization factor chosen so that D_{t+1} is a distribution

6) K-nearest neighbors

The K-nearest neighbor classifier is a non-parametric classifier and has been widely used in the pattern classification and recoginition problems including natural image and medical image analysis applications.

Input: D is the set of feature vectors of training images, z is a feature vector of the test image, L is the set of class labels used to assign a label to z.
Output: $c_z \in L$, the class label of z
for each $y \in D$ **do**
Compute $d(z, y)$, distance of z and y;
end
Select $N \subseteq D$, the set of k closest training feature vectors from z;

$$c_z = \operatorname{argmax}_{v \in L} \sum_{y \in N} I(v = class(c_y));$$

where $I(.)$ is an indicator function that returns the value 1 if its argument is true and 0 otherwise.

7) Light gradient boosting machine

Light gradient boosting machine [15] is an algorithm based on gradient boosting while other algorithm grows trees horizontally meaning. LightGBM grows tree leaf-wise while other algorithms grow level-wise. In the same leaf, the leaf-wise algorithm can reduce more loss than a level-wise algorithm because it will choose the leaf with max delta loss to grow. Besides, light gradient boosting machine can take lower memory to run due to perform the large size of data. Gradient-based one-side sampling (GOSS) and Exclusive Feature Bundling (EFB) strategies are used in LightGBM. Since data samples with larger gradients play a more important role in the computation of information gain, in the case of a much smaller data size, the quite accurate estimation of the information gain can be obtained by using GOSS. Meanwhile, EFP bundle mutually exclusive features to reduce the number of features.

III. Experimental Results

A. Dataset Collection

A Vietnamese herbs dataset was collected from natural environment in Vietnam, which contains these images of herb species: Polyscias fruticosa (đinh lăng), Aloe vera (lô hội), Crinum latifolium (trinh nữ hoàng cung), Passiflora foetida (lạc tiên), Rhizoma belamcanda (xạ cần), Callisia fragrans (lược vàng), Perilla frutescens (tía tô), Coleus amboinicus (tần dày lá), Wedelia chinensis (sài đất), Achyranthes aspera L (ngưu tất) and 978 images of the unknown class. Besides, the herbs dataset was gained by crawling from website of Vietnamese herbal medicine, and then we manually cleaned the crawled data by removing duplicated images and the irrelevant images. The final dataset is composed of 10279 images of the 10 herbal plant species. In Fig. 2, we present the distribution of the herbal plant species in our collected dataset and some samples are shown in Fig. 3.

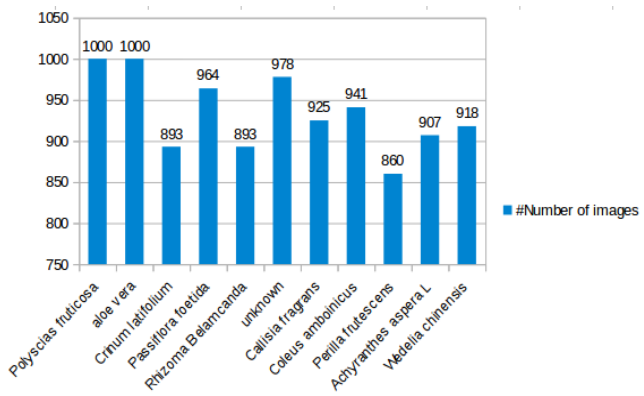

Fig. 2. Vietnamese herbal plant dataset: 10279 images of 10 classes and one unknown class. The number of images in each class ranges from 860 to 1000.

B. Experiments

The Vietnamese herbal plant dataset was split into 10 folds, and cross validation technique is used to train model. In each fold, the dataset is separated into the ratio of 0.8: 0.2 for the training and testing set. The training set contains a total of 8223 images and the validation set contains a total of 2055 images for 11 classes consisting of Polyscias fruticosa, Aloe vera, Crinum latifolium, Passiflora foetida, Rhizoma Belamcanda, Callisia fragrans, Perilla frutescens, Coleus amboinicus, Wedelia chinensis, Achyranthes aspera L and, an unknown class including the plant images of other classes. In each iteration, nine of ten subsets were used for training and one was used to test the trained model.

Our system was implemented in Keras using scikit-learn framework on a computer equiped with CPU Intel Core (™) i7 processor, 16GB RAM and GTX 1050 graphic card.

We aimed to evaluate the use of deep convolutional features with different classifiers. In average, the result in ten folds achieved 88% with the random forest based classifier, which is better than K-nearest neighbor based classifier achieved 76.5%. In experiments on SVM-based classifer, we concluded that the best kernel for SVM to train the model is the linear kernel by a hit rate of 90.8% which is outperformed the other kernels such as polynomial, RBF. Meanwhile, Adaboost and Logistic regression based classifiers obtained the rate of 91% and 92.6% respectively. Finally, we recognized that the LightGBM is the best classifier when it is used with deep convolutional features, and achieved at a rate of 93.6%, a little higher than XGBoost obtained at rate 93%. For evaluation, the results in Table I, Fig. 4, Fig. 5 show that LightGBM classifier outperformed than all other classifiers in all ten folds.

TABLE I: COMPARISON OF DIFFERENT SEVEN CLASSIFFICATION METHODS USING THE SAME DEEP CONVOLUTIONAL VIUSAL FEATURE REPRESENTATION EXTRACTED FROM A MODIFIED VGG16 MODEL

Method	Fold01	Fold02	Fold03	Fold04	Fold05	Fold06	Fold07	Fold08	Fold09	Fold10	Average
Random forest	88.0	86.7	88.8	89.5	87.5	87.4	87.8	88.4	88.1	87.9	88. 0 ±0.76
KNN	77.2	75.7	78.3	78.5	75.1	76.8	77.3	74.8	76.8	74.3	76.5±1.38
SVM	91.4	90.2	92.5	91.1	90.8	89.5	91.0	90.3	91.7	89.8	90.8±0.87
AdaBoost	90.8	90.2	92.2	91.9	91.2	90.8	90.7	91.7	91.5	90.2	91. 0 ±0.69
Logistic regression	93.1	92.0	93.1	94.6	93.3	91.7	93.0	92.3	92.0	91.2	92.6±0.93
XGBoost	93.0	91.1	94.6	94.6	93.8	92.6	92.3	92.9	93.0	91.8	93. 0 ±1.07
LightGBM	**93.8**	**91.1**	**94.7**	**95.0**	**94.2**	**93.8**	**93.0**	**93.4**	**94.0**	**92.8**	**93.6±0.99**

Fig. 3. Some samples from the collected Vietnamese herbal plant dataset: The leftmost column is the class name and the next columns are three sample images taken in real life environment.

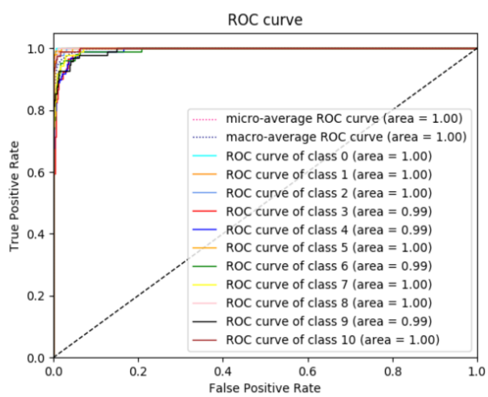

Fig. 4. The ROC curve of the proposed model for 11 different class when using deep convolutional feature representation with the light gradient boosting machine (LightGBM) classifier.

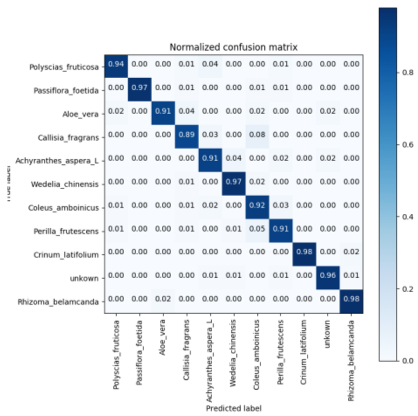

Fig. 5. The confusion matrix when applying the deep convolutional feature representation with LightGBM classifier.

IV. Conclusion

In this work, we evaluated the performance of deep convolutional feature to be used with the different classifiers. Through experiments, we conclude that LightGBM is outperform than the other classifiers including bagging algorithms, Adaboost, SVM, logistic regression and the state-of-the-art XGBoost in the performance. Besides, Deep convolutional feature and LightGBM classifier help herbal plant recognition system applying in real world because it reduces the feature dimension not only in feature extraction processing but also in classifier processing.

In the future, we will use the deep convolutional features extracted from the other architectures rather than VGG16 to improve the performance of the model by a better image representation. Moreover, it is also needed to compare between LightGBM classification technique with other state-of-the-art deep learning classifiers.

References

[1] C. Zhao, S. S. F. Chan, W.-K. Cham, and L. M. Chu, "Plant identification using leaf shapes — A pattern counting approach," *Pattern Recognition*, vol. 48, no. 10, pp. 3203–3215, Oct. 2015.

[2] N. Kumar *et al.*, "Leafsnap: A Computer vision system for automatic plant species identification," in *Proc. Computer Vision – ECCV 2012*, Springer, Berlin, Heidelberg, 2012, pp. 502–516.

[3] A. Aakif and M. F. Khan, "Automatic classification of plants based on their leaves," *Biosystems Engineering*, vol. 139, pp. 66–75, Nov. 2015.

[4] P. Barré, B. C. Stöver, K. F. Müller, and V. Steinhage, "LeafNet: A computer vision system for automatic plant species identification," *Ecological Informatics*, vol. 40, pp. 50–56, Jul. 2017.

[5] G. Cerutti, L. Tougne, J. Mille, A. Vacavant, and D. Coquin, "Understanding leaves in natural images – A model-based approach for tree species identification," *Computer Vision and Image Understanding*, vol. 117, no. 10, pp. 1482–1501, Oct. 2013.

[6] S. H. Lee, C. S. Chan, P. Wilkin, and P. Remagnino, *Deep-Plant: Plant Identification with Convolutional Neural Networks*, 2015, pp. 452–456.

[7] G. L. Grinblat, L. C. Uzal, M. G. Larese, and P. M. Granitto, "Deep learning for plant identification using vein morphological patterns," *Computers and Electronics in Agriculture*, vol. 127, pp. 418–424, Sep. 2016.

[8] Y. Fu and C. Aldrich, "Froth image analysis by use of transfer learning and convolutional neural networks," *Minerals Engineering*, vol. 115, pp. 68–78, Jan. 2018.

[9] Y. Sun, Y. Liu, G. Wang, and H. Zhang, "Deep learning for plant identification in natural environment," *Comput Intell Neurosci*, vol. 2017, 2017.

[10] A. Kamilaris and F. X. Prenafeta-Boldú, "Deep learning in agriculture: A survey," *Computers and Electronics in Agriculture*, vol. 147, pp. 70–90, Apr. 2018.

[11] X. Sun and H. Qian, "Chinese herbal medicine image recognition and retrieval by convolutional neural network," *PLOS ONE*, vol. 11, no. 6, p. e0156327, Jun. 2016.

[12] A. Rakhlin, A. Shvets, V. Iglovikov, and A. A. Kalinin, "Deep convolutional neural networks for breast cancer histology image analysis," *arXiv:1802.00752 [cs]*, vol. 10882, pp. 737–744, 2018.

[13] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew, "Deep learning for visual understanding: A review," *Neurocomputing*, vol. 187, pp. 27–48, Apr. 2016.

[14] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16*, 2016, pp. 785–794.

[15] G. Ke *et al.*, "LightGBM: A highly efficient gradient boosting decision tree," in *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds. Curran Associates, Inc., 2017, pp. 3146–3154.

Anh H. Vo received the M.S. degree in computer science from University of Sciences, Ho Chi Minh City, Vietnam in 2015, and is currently a PhD candidate. Since 2012, she has been a lecturer and researcher at Information Technology Faculty, Ton Duc Thang University, Vietnam. Her main research interests include Image processing, Pattern recognition, Computer vision, Data mining.

Bao T. Nguyen currently serves as a lecturer at University of Technology and Education, HCM Vietnam. Before that, he worked at Laboratory for mathematics in imaging at Harvard University. He received his PhD from the Trento-FBK ICT program at Trento University, Italy. His research interests are image processing, computer vision, medical imaging and neuroinformatics.

Huy V. Pham received the Ph.D in computer science from Ulsan University, South Korea, in 2015, and M.S. degree in computer science from University of Sciences, Ho Chi Minh City, Vietnam in 2007. Since 2015, he has been a lecturer and researcher at Information Technology Faculty, Ton Duc Thang University, Vietnam. His main research interests include artificial intelligence, image processing, computer vision.