

Towards an Automatic Plant Identification System without Dedicated Dataset

Thi Thanh-Nhan Nguyen, Thi-Lan Le, Hai Vu, and Van-Sam Hoang

Abstract—A large image dataset plays a crucial role in building automatic vision recognition system. However, collecting and labeling data are tedious, laborious and time-consuming tasks. In some cases, it is chicken and egg problem: it is only possible to get application data after the system deployment. In our study, we are interested in building automatic plant identification systems from images. As plants distribution on the world is not uniform and may change in response to the availability of resources, the availability of species in different areas is different. That is why some species are very abundant in one region and non-existing in others regions. Even the distribution of plant species is diverse, plant species in the planet share common features. They all have organ types such as leaf, flower, etc. Taking into this observation, in this paper, we propose a new approach for building an image-based plant identification without an available image database based on the combination of deep learning, transfer learning, and crowd-sourcing. The proposed approach consists of four main steps: plant organ detection, plant image collection, data validation and plant identification. Plant organ detection aims to learn organ type characteristic from available image datasets of plants while the purpose of the data collection step is to crawl dataset from crowd-sourced sources. Then, plant organ detection will be used in data validation in order to remove the unwanted/invalid images while keeping the valid ones. Finally, plant identification method will be developed and evaluated from the new image dataset. We illustrate and demonstrate the use of the proposed approach for building a Vietnamese medicinal plant retrieval system.

Index Terms—Organ detection, plant identification, deep learning, convolutional neural network.

I. INTRODUCTION

Plant plays a very important role in human life, providing us with oxygen, food, fuel, medicine, wood, etc. Good knowledge of flora allows to improve agricultural productivity, biodiversity protection, ecosystem balance, planning and minimize the effects of change climate. Currently, the automatic plant identification technique is considered as a promising solution to help popularize

botanical knowledge among the public. According to the study of Bonnet *et al.* [1], automatic plant identification results are lower than the best experts but approximate experienced experts and far exceeds those of beginners or amateurs in plant taxonomy. Based on the impressive results on automatic plant identification, some applications have been deployed and widely used such as the PI@ntnet [2], Leafsnap [3], MOSIR [4], VnMed [5]. However, these systems are dedicated to species of specific regions. For instance, PI@ntnet [2] focuses on plants of France while Leafsnap [3] works with plants from the Northeastern United States and Canada. Moreover, the number of covered plant species (e.g., 10,000 in PlantClef [6]) is very small in comparison with the number of plant species on the earth (e.g., 400,000 [7]). Gaston *et al.* in the study [8] have proposed four suggestions for promoting the use of plant identification in reality. According to Gaston *et al.* [8], the research community should (1) overcome the production of a larger training dataset; (2) scale up; (3) reduce the error rate and (4) be able to detect novel species. Two first suggestions concern the data collection of plant species while two later ones relate to the performance of plant identification. Data collection is a crucial and unavoidable task in building an automatic classification system. Collecting images of a huge number of plant species using the first way is time-consuming and even infeasible. However, a majority of the current works focus on improving the performance of the identification while assuming that the working dataset is available [9]. In this paper, we propose a new approach for building an image-based plant identification without an available image database based on the combination of deep learning, transfer learning and crowd-sourcing. The idea comes from one observation that is even the distribution of plant species on the planet is diverse, plant species share common features: they all have organ types such as leaf, flower, etc. Therefore, we can employ the existing images dataset of species to build an automatic filter for data validation. The proposed approach consists of three main steps: plant organ detection, plant image collection, data validation. Plant organ detection aims to learn organ type characteristic from available image datasets of plants while the purpose of data collection step is to crawl dataset from crowd-sourced sources. Then, plant organ detection will be used in data validation in order to remove the unwanted/invalid images while keeping the valid ones. We demonstrate and evaluate the use of the proposed approach for building an automatic plant identification system without available dataset: Vietnamese medicinal plant retrieval.

The rest of the paper is organized as follows. Section II introduces the challenges of building automatic plant

Manuscript received August 25, 2018; revised November 3, 2018. This research is funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 106.06-2018.23

Thi Thanh-Nhan Nguyen is with International Research Institute MICA, HUST-CNRS/UMI-2594-GRENOBLE INP, Hanoi, Vietnam and University of Information and Communication Technology, Thainguyen University, Thainguyen, Vietnam (e-mail: ntnhan@ictu.edu.vn).

Thi-Lan Le and Hai Vu are with International Research Institute MICA, HUST-CNRS/UMI-2594-GRENOBLE INP, Hanoi, Vietnam (e-mail: thi-lan.le@mica.edu.vn, hai.vu@mica.edu.vn).

Van-Sam Hoang is with Vietnam Forestry University, Hanoi, Vietnam (e-mail: hoangsam@vfu.edu.vn).

identification. The proposed framework is described in Section III. Section IV presents experimental results and Section V is a conclusion and future work.

II. CHALLENGES OF BUILDING AUTOMATIC PLANT IDENTIFICATION SYSTEM

Two crucial factors of an automatic object identification in general and plant identification in particular are database and identification methods. In this section, we will analyze and point out the challenges encountered in building dataset as well as identification methods.

Most studies focus only on plant identification methods. Recent works for plant identification either employs available benchmark images datasets such as Flavia, ImageClef or collect their own datasets. There are two main approaches for plant image collecting: manually image acquisition and crowdsourcing-based image collecting.

In the first approach, images of the plant of interest are manually collected. Due to the small number of people participated, the collected datasets are small and homogeneous and often cover a small number of species on a specific region. These datasets are then manually annotated. This type of dataset is suitable for laboratory studies, but not for practical application. Table I lists the most common plant image datasets collected by using the first approach.

TABLE I: PLANT IMAGE DATASETS USING THE CONVENTIONAL APPROACH

Dataset	Place	Organ	#species	#images
Swedish Leaf [10]	Sweden	leaf	15	1,125
Flavia [11]	China	leaf	32	1,907
Leafsnap [12]	America	leaf	184	30,866
ICL [13]	China	leaf	220	17,032
Malaya Kew Leaf [14]	England	leaf	44	2,288
Oxford flower 102 [15]	United Kingdom	flower	102	8,189

In the second approach, it is called crowdsourcing. Crowdsourcing is a technique that aims to take contributions from a large group of people, especially an online community where each person's contribution combines with those of others to achieve a cumulative result. That is a new trend today with many examples such as Oxford English Dictionary, Wikipedia, etc. There are two different ways of collecting crowdsourced data [2]. The first one is to automatically crawl huge amounts of images on the web then filter or validate them. ImageNet, the largest image database, has been built by this way [16]. The ImageNet is evaluated by humans. This is achieved by using the service of Amazon Mechanical Turk, an online platform on which one can put up tasks for users to complete and to get paid. ImageNet contains many plant images, but it does not organize itself into specific species that organized in form of Wordnet structure. Collecting image data on the Internet will contain a lot of noise, so it is important to remove these noise images. For example, ImageNet aims to eventually offer 500-1,000 clean images per synset, each synset has over 10,000 images on average. The second one is to collect data from users via tools developed for this specific purpose such as Pl@ntnet, Leafsnap, iNaturalist, iSpot, Flora finder, Plantifier. The

crowdsourcing systems can propose explicit or implicit participation to users. The main characteristic of implicit systems is that they collect information from users via indirect way. This method of gathering data will be less noise than a web-based collection. The validation of data is based on user communities, botanical experts and automated plant identification tools.

Table II shows some datasets which are collected by the second approach. These datasets are relatively large and diverse. Images usually have a complex background so they are very useful for evaluating plant identification in the real environment. However, this approach requires a lot of time for validation task. Even a number of plant datasets are available for research purpose, they cover the plants of some specific regions such as Europe, England, America, and China. Many regions in the world do not have public visual botanical dataset. For example, Vietnam has a very rich flora system, but Vietnam has no public plant images dataset. This is a challenge to build an automatic plant identification system for these areas.

TABLE II: PLANT IMAGE DATASETS ARE BUILT BY CROWDSOURCING DATA COLLECTION TOOLS

Dataset	Place	#species	#images
PlantClef 2015 [17]	Europe	1,000	113,205
PlantClef 2017 [6]	Around the world	10,000	1.8M
Social image collection [2]	France	2,258	44,810

To validate the plant image dataset collected from various sources, this task to do is to determine whether an input image is a plant or non-plant. If it is a plant image, the type of plant organ should be determined. This task has many challenges with plant images may contain text and other objects. Fig. 1 shows some examples where plant and non-plant classification is ambiguous.



Fig. 1. Some challenges in plant and non-plant classification.

Concerning the plant organ detection, this task also has to deal with different challenges: an image may consist of different organs; the shape of some organs of certain plants is quite similar. Fig. 2 illustrates some difficult cases for plant organ detection. The boldface word below is name of organ that is assigned by the user in the PlantClef 2015 dataset. Each image is assigned to one sole organ.



Fig. 2. Illustration of difficult cases for plant organ detection.

Currently, plant image datasets are assumed that the type of organ is manually predetermined. This task is very time-consuming and depends on the users. Plant retrieval system Pl@ntnet is the first application based on multiple

organ images [2]. However, when a user submits a query image, the system requires users must determine the name of an organ.

Some studies are proposed to detect automatically plant organs [3], [18], these studies usually utilize hand-crafted features (e.g. GIST features) and traditional classifiers such as K-Nearest Neighbor (kNN), Support Vector Machine (SVM). These methods only determine automatically an image is leaf or non-leaf. The authors evaluated on the Flavia and Leafsnap datasets (see Table I), images used in these datasets contain a single leaf on a uniform background and the datasets are small. To the best of our knowledge, it lacks studying for automatic organs detection on complex and diverse datasets.

In recent years, deep learning methods have developed rapidly because they able to achieve significant success in many problems. In computer vision, Convolutional neural networks (CNNs) have been introduced to learn automatically features of the input data through multiple layers of transforms. CNNs have proved effective in detecting and classifying objects [19]-[23]. The potential of CNNs has been exploited by many researchers to outperform conventional methods based on hand-crafted features [20], [24]. Some famous networks such as AlexNet [20], GoogLeNet [25], VGG [24] have also been applied for plant identification, especially in PlantClef competition from 2014 to 2017 and have obtained higher results compared to traditional methods based on hand-designed features [6], [14], [17], [26], [27]. However, these works based on an assumption that the working dataset is available. Moreover, none of the studies has applied CNNs for organ plant detection.

III. PROPOSED METHOD

A. Overall

In this paper, we proposed a new approach based on deep learning for building an automatic plant identification from images without available database following Fig. 3. Our framework includes 3 steps. The first step aims at collecting images from different sources. In order to collect as many as possible images for each plant, we employ both approaches for data collection: manual image acquisitions and crowdsourcing based. Concerning crowdsourcing approach, as plants in different regions may have different names while they have an identical scientific name, we employ JSoup in Java library to collect the plant images from Internet by using their scientific names and common names. Fig. 4 illustrates some collected images. There are some non-plant images. Then, these images are needed to go through the data validation step where only valid images are kept for the further processing. In the context of our work, valid images have to contain one of the five most common organs that are leaf, flower, stem, fruit and branch. To do this task, we build a convolutional neural network named OrganNet. Finally, a convolutional neural network named VnPlantNet is built for plant identification. Both OrganNet and VnPlantNet are pre-trained on available and diverse datasets that are PlantClef 2015 and ImageNet. In the following section, we

will describe the architectures of OrganNet and VnPlantNet.

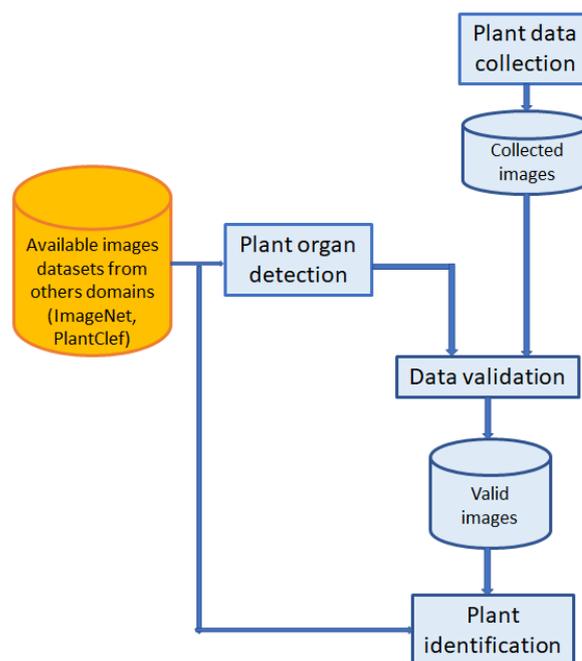


Fig. 3. The proposed framework.



(a) *Camellia sinensis*



(b) *Terminalia catappa*

Fig. 4. Some images of data collection for two species: (a) *Camellia sinensis*, (b) *Terminalia catappa*. First row shows images are collected by manual image acquisitions, second row shows images are collected by crowdsourcing.

B. Organ Detection and Plant Identification Method Based on Deep Learning and Transfer Learning

Before describing the method of organ detection and plant identification, we would like to recall the aim of these tasks. Automatic plant organ detection aims at determining the organ present in a given image. An image may contain several plant organs. However, in our work, we are interested in the most dominant organs. Therefore, one image will be classified in one of six classes that are leaf, flower, fruit, stem, branch, non-plant. The purpose of the plant identification is to determine the name of the species in a given image.

Thanks to the impressive results obtained in different tasks of computer vision, deep learning is widely used in computer

vision. However, deep learning methods often require a lot of training data as they have to train millions of parameters. In reality, many real-world problems typically do not have a lot of data. Therefore, to solve this problem, a commonly used technique is transfer learning [28]. The general idea of transfer learning is to use knowledge learned from a task where a number of labeled training data is available, in a new related task where data is limited. The advantages of transfer learning are that it allows to reduce the training time and often obtain better performance.

In this paper, we propose to apply deep learning and transfer learning for building organ detection and plant identification methods. In order to distinguish two networks (one for organ detection and one for plant identification), we name the two networks OrganNet and VnPlantNet respectively.

OrganNet and VnPlantNet are built by applying transfer learning methods on GoogLeNet architecture. Among different architectures, GoogLeNet that won in ImageNet Large Scale Visual Recognition Competition (ILSVRC) 2014 [25] is employed. GoogLeNet is an inception architecture that combines the multi-scale idea and dimension reduction layers so it has deeper and wider architecture than many other CNNs such as AlexNet, VGGNet. This net allows reducing a large number of training parameters in the network. It includes roughly 6.8 million parameters, 22 layers with nine inception modules, two convolutional layers, one convolutional layer for dimension reduction, two normalization layers, four max-pooling layers, one average pooling, one fully-connected layer, and a linear layer with Softmax activation as the classifier on top. The inception module uses parallel 1×1 , 3×3 , 5×5 convolutional layers with other layers to reduce dimensionally. This architecture is also called network in the network. During training, GoogLeNet connects two auxiliary classifiers to the intermediate layers of the network to effectively perform backpropagation through all the layers. CNNs are generally trained based on the prediction loss minimization. Let input x (images) and corresponding output (class labels). The purpose of the training is to iteratively minimize the average loss value $L(w)$ defined as follows:

$$s = f(x_i; w) \quad (1)$$

$$L_i = -\log P(Y = y_i | X = x_i) = -\log \left(\frac{e^{s_{y_i}}}{\sum_j e^{s_j}} \right) \quad (2)$$

$$L(w) = \frac{1}{N} \sum_{i=1}^N L_i + \lambda R(w) \quad (3)$$

where x_i is an input image, y_i is class label of x_i , N is the number of data instances (mini-batch) in every iteration, L_i is the loss function, f is the predicted output of the network depending on the current weights w , R is the

regularization term or weight decay with the Lagrange multiplier λ . The stochastic gradient descent (SGD) is used to update the weights.

$$w_{t+1} = \mu w_t - \alpha \nabla L(w_t) \quad (4)$$

where μ is the momentum weight for the current weights w_t and α is the learning rate. The network weights are randomly initialized during training the network from scratch, while in fine-tuning deep models, the network weights are initially set to the pre-trained network weights.

Concerning the transfer learning, there are two main approaches. Firstly, we can fine-tune the weights of the pre-trained network by continuing the backpropagation. It is possible to fine-tune all the layers of the CNN or to keep some of the earlier layers fixed and only fine-tune some higher-level portion of the network. Secondly, we can use CNN as a fixed feature extractor. This means that we can remove the last fully-connected layer, then treat the rest of the CNN as a fixed feature extractor for the new dataset. Then, any classification method such as SVM can be applied on these feature vectors. In our study, we will evaluate these two approaches for OrganNet. For VnPlantNet, we apply only the first one. Both OrganNet and VnPlantNet use a GoogLeNet pre-trained model which is trained on a very larger ImageNet dataset (which contains 1.2 million images with 1,000 categories). Then we adjust some parameters such as number output, batch size, learning rate, momentum, weight decay, etc to fit our dataset and fine-tune the weights through the training process. Additionally, we use OrganNet to extract feature vector of average pooling layer (pool5/7x7_s1), the layer is nearly the fully connected layer, the output is 1024 dimensions vector for every image. This layer extracts the most abstract features and provides the best representation of objects in the image.

C. Data Validation

As collected data may contain noise, the main purpose of this task is to remove the invalid plant images while keeping the valid ones. In this work, the invalid images are images that do not contain at least one interested plant organs. Based on the observation that is even the distribution of plant species on the planet is diverse, plant species share common features: they all have organ types such as leaf, flower, etc. Therefore, we build an OrganNet that was trained on the existing image datasets (ImageNet and PlantClef) as an automatic filter for data validation.

IV. EXPERIMENTAL RESULTS

In this section, we evaluate both OrganNet and VnPlantNet for organ detection and plant identification as well as the proposed framework for building an automatic plant identification without an available dataset. To evaluate the performance of these networks, we use the accuracy measure Acc_{rank_k} at rank k that is defined as follows:

$$Acc_{rank_k} = \frac{T_{rank_k}}{N}$$

T_{rank_k} is a number of true detection/identification in the first position k , N is the total number of query images. In this study, we compute the accuracy at two first ranks ($k = 1, 2$) for organ detection and five first ranks ($k = 1, 5$) for plant identification as an effective method should bring relevant results at the first ranks.

We have used Caffe [29] - a framework for deep learning where the state-of-the-art CNN architectures and models are available. We used the data augmentation technique implemented within Caffe library that consists of cropping randomly a 224×224 pixels image and mirroring images horizontally. Data augmentation is applied to reduce the overfitting and to improve classification results during testing. All experiments are run on a computer with a GeForce GTX 1080 Ti GPU.

A. Organ Detection Results

To train and evaluate OrganNet, we prepare the dataset by taking images of PlantClef 2015 dataset [17] and the collected dataset from the Internet. PlantClef 2015 consists of seven organs of a plant: leaf on a complex background, leaf on a simple background (leafscan), flowers, fruits, stems, branches, and entires. Due to the large similarities between leaf and leafscan, branch and entire, we combine these pairs together. With non-plant images, we collected from the Internet. Table III presents the detail information of the dataset. We perform three experiments. The two first experiments are evaluated the effectiveness of OrganNet with two approaches of transfer learning while the last one is to compare our network with the state-of-the-art work.

TABLE III: DATASET USED FOR EVALUATING ORGAN DETECTION METHOD

	Training	Testing
Leaf	25,972	2,911
Flower	28,225	8,327
Fruit	7,720	1,423
Stem	5,476	584
Branch	24,365	8,201
Non-plant	807	189

Experiment 1: In this experiment, to understand the role of fine-tuning, we evaluate the performance of OrganNet with two initialization strategies: randomly initialized weights and pre-trained weights from ImageNet. We fine-tune our network and adjust some parameter such as batch size: 32, initial learning rate: 0.001, momentum: 0.9, weight_decay: 0.0002. Table IV presents the results corresponding to two weighted initialization strategies. The results show that using the weighted training set on a large such as ImageNet allows to obtain an improvement +5.08% for the rank 1 and +2.54% for the rank 2 over the case of randomly weight initialization. The accuracies of the proposed method are 87.18% and 97.46% at rank 1 and rank 2 respectively. This result is very promising as the working images are mainly captured in a complex background. This proves that deep learning is capable of learning well with natural images. To understand the behavior of the proposed network for each organ, the confusion matrix is shown in Table V. The accuracies at rank 1 are non-plant (**95.02%**), stem (**92.98%**), flower (**91.26%**), branch (**86.97%**), leaf

(**82.72%**) and fruit (**70.20%**). The detection of non-plant images for the highest result, because the non-plant images in this dataset have very different shape, color, background among plant organs images. Stem is the most distinguishable plant organ thanks to its specific texture and color. Moreover, stem images usually do not contain other organs. Fruit is the most difficult organ to detect in the images as many fruit images are similar to flower and branch images. Fig. 5 illustrates some examples of wrong classification. Images are not correctly identified because of the following reasons:

- The shape and color of images of different organs are similar. For example, some images of flower and fruit have similar appearance. Determining which organ is present in these images is challenging even for human.
- Images usually contain more than one organ especially entire/branch images. They are often classified as leaf, flower, fruit, stem images and vice versa.
- Some leaf images are taken at a very close distance. In this case, due to the vein of the leaf, leaf images can be identified as stem images.
- The wrong identification results are also derived from the ambiguity of manual annotation provided in PlantClef 2015.

TABLE IV: THE ORGAN DETECTION PERFORMANCE OF THE ROPOSED ORGANNET WITH DIFFERENT WEIGHT INITIALIZATION

Weighted initialization strategy	Acc _{rank_1} (%)	Acc _{rank_2} (%)
Randomly generated weight	82.10	94.92
Pre-trained on ImageNet	87.18	97.46



Fig. 5. Some examples for wrong identification.

TABLE V: CONFUSION MATRIX FOR PLANT ORGAN DETECTION OBTAINED WITH THE PROPOSED ORGANNET (%)

		Prediction					
		Leaf	Flower	Fruit	Stem	Branch	Non-plant
Ground -truth	Leaf	82.72	2.1	1.48	0.65	12.95	0.10
	Flower	0.29	91.26	1.73	0.14	6.57	0.01
	Fruit	2.32	14.13	70.2	0.49	12.79	0.07
	Stem	1.37	0.86	0.86	92.98	3.77	0.17
	Branch	3.13	7.95	1.37	0.56	86.97	0.02
	Non-plant	0.50	0.50	2.49	1.00	0.50	95.02

To confirm the robustness of the chosen network architecture for OrganNet, we compare results of OrganNet with two different architecture AlexNet, VGG-16. The accuracies at rank 1 are 85.64% (AlexNet), 81.79%

(VGG-16) and 87.18% (OrganNet). OrganNet obtains the best results among the three evaluated architectures. Additional, to visualize the decisions of these networks, we apply the visualization method in [30]. Fig. 6 shows results of the AlexNet, OrganNet and VGG-16 networks on two input images. Red pixels are evidence for a class and blue ones against it. The obtained results clearly show that the predicted areas are in the center of the object, while VGG-16 does not focus on the center of the object but scatter around the object or interest in the background of the object.

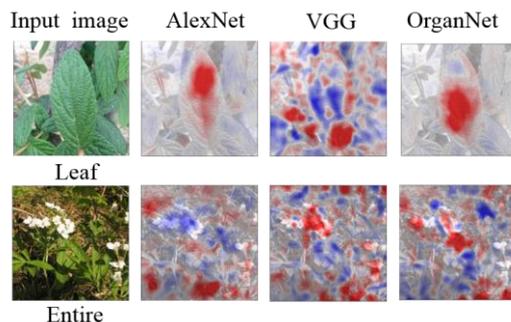


Fig. 6. Visualization of the prediction of different CNN architectures. Red pixels are evidence for a class, and blue ones against it.

Experiment 2: This experiment is to evaluate the second approach of transfer learning on our network. We use OrganNet as a feature extractor. We extract the 1024-dimensional feature vector of the last average pooling layer before the dropout layer. Ten classifiers are applied. They are Nearest neighbor (NNB), Linear Support Vector Machine (L_SVM), Non-Linear Support Vector Machine (we use Radial Basic Function kernel-SVM_RBF), Decision Tree (DT), Random Forest (RF), Neural network (NN), Ada Boost (AB), Naïve Bayes (NB), Quadratic discriminant analysis (QDA) and Softmax (SM) - the default classification method of the GoogLeNet. Fig. 7 shows the results of different classifiers. Among ten classification methods, Softmax, Linear SVM, and Neural network outperform the others. The Linear SVM classifier has achieved the best result (87.34%) because this method is effective for large datasets.

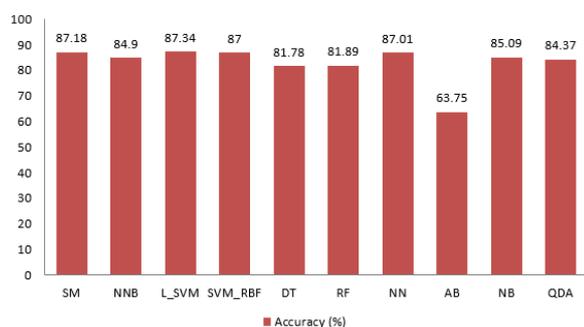


Fig. 7. Detection results of the OrganNet with different classification methods at the first rank ($k=1$).

Experiment 3: To compare the proposed network with the state-of-the-art work for organ detection, we re-implement and evaluate the method based on hand-crafted features presented in [3] on the same dataset. The method in [3] used GIST feature and SVM for leaf and non-leaf classification. In this method, from an input image, a 512-dimensional GIST feature vector is extracted. Then, the SVM classifier is

applied. The achievement of accuracy is **67.27%**, lower than 19.91% of our proposed network and lower than almost other classifiers in Experiment 2 except Ada Boost classifier. This proves the effectiveness of the proposed OrganNet for plant organ detection.

Fig. 8 shows the comparison of the results of the proposed method and methods [3] for each organ. The OrganNet outperforms the method in [3] for all organ types. The OrganNet can detect the fruit organ with 70.2% of accuracy for rank-1 while the method in [3] fails in detecting this organ.

B. Case Study: Vietnamese Medicinal Plant Retrieval (VnMed)

To illustrate the use of the proposed framework, we apply it for Vietnamese medicinal plant retrieval system named VnMed. This system is firstly introduced in [31]. The architecture of VnMed is shown in Fig. 9, Fig. 10 shows some snapshots of this application. This application provides different ways to search a plant of interest: text-based, biological icon-based and leaf image-based. This system has two parts: client and server. On the client side, it includes textual dataset and text-based plant retrieval. On the server, it includes images dataset and plant identification module. The system provides detail information of 600 Vietnamese medicinal species including plant's common name, scientific name, introduction, description, distribution, harvesting, collecting and processing, chemical component, usage, and dosage. This system also supports image-based retrieval based on leaf image for 55 species. In the image recognition, this system only focuses on the leaf. Leaf images captured in complicated background need to go through an interactive segmentation. This system uses a modified Kernel descriptor for extracting feature and SVM classification [5]. However, leaf is not enough information for plant identification and user interaction with the system to segment images is time-consuming. Thus, in this paper, we add to this system one function-plant identification from images of many plant organs using the deep-learning method. As the multi-organ image dataset of the Vietnamese medicinal plants is not available, we apply the data collection method as described in Section III. A to collect images of 100 plants.

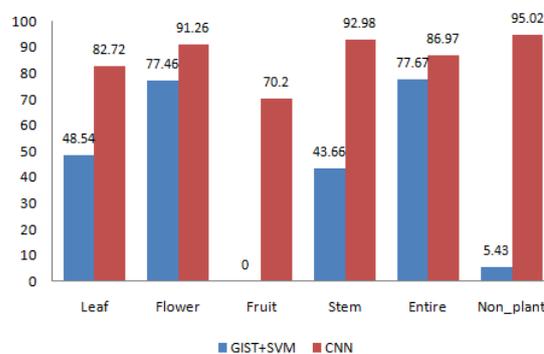


Fig. 8. Results obtained by the proposed OrganNet and the method in [3] for six organs.

As results, we have collected **4,873** images by manual image acquisition (named **VnDataset1**) and **15,772** images from Internet. These images are merged in the same dataset with the total number of images is **20,645** (named

VnDataset2). The number of images per species varies from 57 to 379. Each species has about 206 images on average. The **VnDataset2** contains noise (such as non-plant images) due to the automatic crawling process. Some images are shown in Fig. 4. We apply the OrganNet built in the previous section in **VnDataset2** in order to remove invalid images. The results for detecting Vietnam species database are shown in Table VI. The average accuracy of 6 organ detection is **79.54%**. If we are interest in two classes (valid and non-valid), the accuracy is **95,31%**, these results show that the proposed method works very well even with natural and challenging images. After this data validation, we remove invalid images from the **VnDataset2** to obtain **VnDataset3**. As the organ detection method is not perfect, **VnDataset3** still contains some invalid images. We then remove these images (636 images) manually to obtain **VnDataset4**. Among the four datasets, **VnDataset1** is the simplest while **VnDataset2** is the most challenging one. These datasets are shown in Table VII.

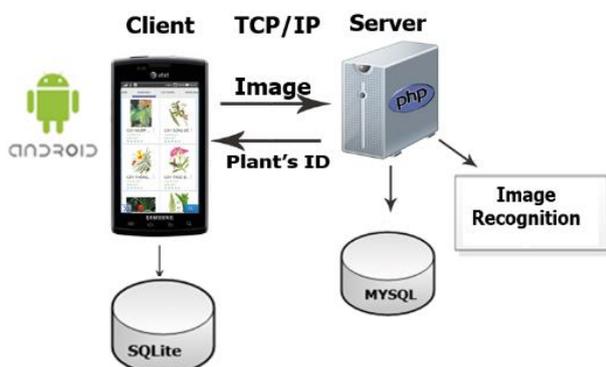


Fig. 9. Architecture of Vietnamese medicinal plant search system [31].

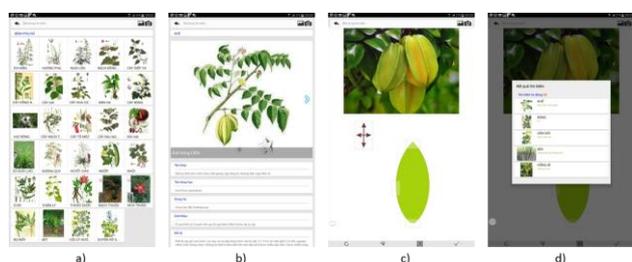


Fig. 10. Snapshots of VnMed; a) list of species for a group of diseases; b) a detail information for one species; c) a query image for plant identification; d) top five returned results.

TABLE VI: CONFUSION MATRIX FOR DETECTION 6 ORGANS OF 100 VIETNAMESE SPECIES (%)

		Prediction					
		Leaf	Flower	Fruit	Stem	Branch	Non-plant
Ground-truth	Leaf	92.30	2.23	1.95	0.76	1.25	1.52
	Flower	1.24	87.99	5.75	0.00	3.99	1.04
	Fruit	11.21	5.45	71.25	0.41	3.41	8.26
	Stem	10.77	1.93	6.91	69.34	3.31	7.73
	Branch	5.75	8.06	7.18	0.45	75.92	2.64
	Non-plant	9.18	4.97	16.85	1.19	11.02	56.80

TABLE VII: FOUR VIETNAMESE MEDICINAL SPECIES DATASETS

	VnDataset1	VnDataset2	VnDataset3	VnDataset4
train	3,901	16,513	15,652	15,150
test	972	4,132	3,297	3,163

To understand how a model trained by using one dataset

perform to evaluate the performance of VnPlantNet, for each dataset, we take 20% for testing and the remaining for training.

Our VnPlantNet employs the pretrained weights from ImageNet and is fine-tuned with the following parameters: base learning rate: 0.001, batch size: 32. Four models are generated for four corresponding datasets (denote M_i model). A_{ij} is accuracy of model M_i on the test j . The results are shown in Table VIII. Three interesting conclusions can be made from the experiment results. Firstly, the training data plays an important role in the performance of the network. The more heterogeneous the training data is, the more robust the model is. Among 4 models, M_1 outperform the other ones on VnDataset1 (accuracy at rank 1 is 81.58%). However, when testing with the other datasets, the performance of this model decreases dramatically. The other models obtain the results that are relatively lower than the model M_1 on **VnDataset1**. The accuracies at rank-1 of M_2 , M_3 , and M_4 are 76.03%, 78.70% and 79.63% respectively. However, these models still keep high accuracies when working with the others (**VnDataset1**, **VnDataset2**, **VnDataset3**). Secondly, the data validation is a crucial step in crowd-sourcing based dataset building. The experimental results show that the data validation obtains +1.38% (for automatic data validation with OrganNet) and +4.04% of improvement when combining both automatic and manual data validation. It is also worth noting that the automatic data validation with OrganNet allows to remove a significant part of the invalid images. This means that it can reduce a lot of human consuming time for data validation. Finally, the proposed framework can help to build an image-search function in the VnMed application that allows to search plants of interest by using images of different plant organs. This function is also satisfied the requirement of the real scenario where the query images of the users may be very diverse and complex.

TABLE VIII: RESULTS FOR VIETNAMESE MEDICINAL PLANT IDENTIFICATION

Testing dataset from	Accuracy (%)	M_1	M_2	M_3	M_4
VnDataset1	Rank 1	81.58	76.03	78.70	79.63
	Rank 5	90.64	88.48	83.54	84.77
VnDataset2	Rank 1	26.06	54.35	54.61	54.42
	Rank 5	38.41	75.76	64.11	63.96
VnDataset3	Rank 1	28.48	55.14	55.99	56.51
	Rank 5	33.27	65.00	75.77	66.45
VnDataset4	Rank 1	29.62	56.50	57.73	58.46
	Rank 5	34.62	66.42	67.31	79.48

V. CONCLUSIONS AND FUTURE WORKS

The paper proposes an automatically plant identification system without an available database including steps: collecting data from various sources, organ detection, automatically evaluating data and plant species identification. The proposed methods (OrganNet for organ detection and VnPlantNet for plant identification) are based on transfer learning methods in CNNs and learn from the large available

datasets. We apply two transfer learning methods for OrganNet. OrganNet uses the weighted train on large dataset ImageNet which obtains the accuracy 87.18% at rank 1 higher than 5.08% of the case of randomly weight initialization. It also higher than 19.91% of the traditional method where GIST feature vector is extracted then the SVM classifier is applied. The experimental results also show that the set of features extracted from OrganNet is more efficient than hand-design features set (GIST) when passing through different classifiers. OrganNet is used in Vietnamese medicinal plant retrieval system as an automatic filter for collected data from many sources. Then we applied VNPlantNet to identify 100 species of plant. The method in the paper opens the ways for solving the problem automatically from validating data to plant identification. The paper is also a proposal for other specialized identification systems that do not an available database. In the future, we will focus to improve the results of plant identification. We will continue to collect more data with more species and we will build a new function which helps to collect plant image from users of VnMed.

ACKNOWLEDGMENT

This research is funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 106.06-2018.23

REFERENCES

- [1] B. Pierre *et al.*, "Plant identification: Man vs. machine," *Multimedia Tools and Applications*, vol. 75, no. 3, pp. 1647-1665, 2016.
- [2] J. Alexis *et al.*, "Interactive plant identification based on social image data," *Ecological Informatics*, vol. 23, pp. 22-34, 2014.
- [3] K. Neeraj *et al.*, "Leafsnap: A computer vision system for automatic plant species identification," *Computer Vision-ECCV 2012*, Springer Berlin Heidelberg, pp. 502-516, 2012.
- [4] Phyu, K. Hninn, A. Kutics, and A. Nakagawa, "Self-adaptive feature extraction scheme for mobile image retrieval of flowers," in *Proc. Eighth International Conference on Signal Image Technology and Internet Based Systems (SITIS)*, IEEE, 2012.
- [5] T.-L. Le, D.-T. Tran, and V.-N. Hoang, "Fully automatic leaf-based plant identification, application for Vietnamese medicinal plant search," in *Proc. the Fifth Symposium on Information and Communication Technology*, ACM, 2014.
- [6] G. Herve, P. Bonnet, and A. Joly, "Plant identification based on noisy web data: the amazing performance of deep learning (LifeCLEF 2017)," *CEUR Workshop Proceedings*, 2017.
- [7] G. Rafađ, "How many species of seed plants are there?" *Taxon*, vol. 50, no. 4, pp. 1085-1090, 2001.
- [8] Gaston, J. Kevin, and M. A. O'Neill, "Automated species identification: Why not?" *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 359, no. 1444, pp. 655-667, 2004.
- [9] W. Jana and P. M älner, "Plant species identification using computer vision techniques: A systematic literature review," *Archives of Computational Methods in Engineering*, vol. 25, no. 2, pp. 507-543, 2018.
- [10] S. Oskar, "Computer vision classification of leaves from Swedish trees," M.S. thesis, Link öping University, 2001.
- [11] S. G. Wu *et al.*, "A leaf recognition algorithm for plant classification using probabilistic neural network," in *Proc. 2007 IEEE International Symposium on Signal Processing and Information Technology*, IEEE, 2007.
- [12] Leafsnap. (2018). [Online]. Available: <http://leafsnap.com/dataset/>
- [13] R. X. Hu *et al.*, "Multiscale distance matrix for fast plant leaf recognition," *IEEE Transactions on Image Processing*, vol. 21, no. 11, pp. 4667-4672, 2012.
- [14] L. S. Han *et al.*, "Deep-Plant: Plant Identification with convolutional neural networks," in *Proc. 2015 IEEE International Conference on Image Processing (ICIP)*, IEEE, 2015.
- [15] M.-E. Nilsback and A. Zisserman, "Automated flower classification over a large number of classes," in *Proc. 2008. ICVGIP'08. Sixth Indian Conference on Computer Vision, Graphics & Image Processing*, IEEE, 2008.
- [16] J. Deng *et al.*, "Imagenet: A large-scale hierarchical image database," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, 2009.
- [17] H. Go àu, P. Bonnet, and A. Joly, "LifeCLEF plant identification task 2015," *CLEF: Conference and Labs of the Evaluation Forum, CLEF2015 Working Notes*, vol. 1391, 2015.
- [18] Q.-K. Nguyen, T.-L. Le, and N.-H. Pham, "Leaf based plant identification system for android using surf features in combination with bag of words model and supervised learning," in *Proc. International Conference on Advanced Technologies for Communications (ATC 2013)*, IEEE, 2013.
- [19] K. Koray *et al.*, "Learning convolutional feature hierarchies for visual recognition," *Advances in Neural Information Processing Systems*, 2010.
- [20] K. Alex, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, 2012.
- [21] H.-J. Yoo, "Deep convolution neural networks in computer vision," *IEIE Transactions on Smart Processing & Computing*, vol. 4, no. 1, pp. 35-43, 2015.
- [22] M. Heidarysafa, K. Kowsari, D. E. Brown, K. J. Meimandi, and L. E. Barnes, "An improvement of data classification using random multimodel deep learning (RMDL)," *International Journal of Machine Learning and Computing*, vol. 8, no. 4, pp. 298-310, 2018.
- [23] K. U. R. Harliman, "Data- and algorithm-hybrid approach for imbalanced data problems in deep neural network," *International Journal of Machine Learning and Computing*, vol. 8, no. 3, pp 208-213, 2018.
- [24] S. Karen and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014.
- [25] S. Christian *et al.*, "Going deeper with convolutions," in *Proc. the IEEE Conference on Computer Vision and Pattern Recognition*, 2015.
- [26] G. Herv é *et al.*, "Lifeclef plant identification task 2014," *CLEF2014 Working Notes. Working Notes for CLEF 2014 Conference, Sheffield, UK, September 15-18, 2014*, CEUR-WS, 2014.
- [27] G. Herv é P. Bonnet, and A. Joly, "Plant identification in an open-world (lifeclef 2016)," *CLEF Working Notes*, 2016.
- [28] Y. Sawada and K. Kozuka, "Whole layers transfer learning of deep neural networks for a small scale dataset," *International Journal of Machine Learning and Computing*, vol. 6, no. 1, pp. 27-31, 2016.
- [29] Y. Q. Jia *et al.*, "Caffe: Convolutional architecture for fast feature embedding" in *Proc. the 22nd ACM International Conference on Multimedia*, ACM, 2014.
- [30] M. Z. Luisa *et al.*, "Visualizing deep neural network decisions: Prediction difference analysis," 2017.
- [31] T.-L. Le, V.-N. Hoang, H. Vu, and T.-H. Tran, "Vietnamese medicinal plant retrieval system for android," in *Proc. the 17th National Conference: Selected Problems about IT and Telecommunication*, Taynguyen, Vietnam, 2014.



Thi Thanh Nhan Nguyen received the math information degree from Vietnam National University, in 2003, Vietnam and holds a master's degree in computer science from Thainguyen University, in 2007, Vietnam, and continues studying doctoral degree in computer science from Hanoi University of Science and Technology, Vietnam. She now is a lecturer at Information Technology and Communication University, Thainguyen University. Her current research interests include computer vision, image processing, pattern recognition.



Thi Lan Le received B.E. degree in information technology in 2003 and M.E. in computer science in 2005, both from Hanoi University of Technology, Vietnam. She received Ph.D. in computer science from the University of Nice - Sophia Antipolis at I.N.R.I.A in the ORION project team, 2009. She is interested in video surveillance indexing and retrieval, human-computer/robotic interaction, content-based image indexing and retrieval.



Hai Vu received B.E. degree in electronic and telecommunications in 1999 and M.E. in information processing and communication in 2002, both from Hanoi University of Technology. He received Ph.D. in computer science from Graduate School of Information Science and Technology, Osaka University, 2009. He is interested in medical imaging techniques, mainly video capsule endoscopy analysis

for diagnostic assistance; computer vision supporting visually impaired people, human-computer/robotic interaction; computer vision in agricultural engineering.



Van Sam Hoang received B.E. degree in forest resources management at Vietnam National University of Forestry in 1999 and M.E. in plant taxonomy and biodiversity conservation at Leiden University, the Netherlands. He received Ph.D. in plant taxonomy and biodiversity conservation at Leiden University, the Netherlands. He is a lecturer in Vietnam National University of Forestry, Vietnam. He is interested in

plant diversity and conservation, forest resources management, biodiversity conservation, National parks management, local people and forest resources.