Plant Species Identification by Bi-channel Deep Convolutional Networks

Guiqing He¹, Zhaoqiang Xia¹*, Qiqi Zhang¹, Haixi Zhang¹ and Jianping Fan²

¹School of Electronics and Information, Northwestern Polytechnical University, Shaanxi, China.
²Department of Computer Science, University of North Carolina at Charlotte, Charlotte, USA

*The corresponding author’s e-mail: zxia@nwpu.edu.cn

Abstract. Plant species identification achieves much attention recently as it has potential application in the environmental protection and human life. Although deep learning techniques can be directly applied for plant species identification, it still needs to be designed for this specific task to obtain the state-of-art performance. In this paper, a bi-channel deep learning framework is developed for identifying plant species. In the framework, two different sub-networks are fine-tuned over their pretrained models respectively. And then a stacking layer is used to fuse the output of two different sub-networks. We construct a plant dataset of Orchidaceae family for algorithm evaluation. Our experimental results have demonstrated that our bi-channel deep network can achieve very competitive performance on accuracy rates compared to the existing deep learning algorithm.

1. Introduction

Plants are tremendously important to human beings because they are source of food, clothing, medicines and other materials. In the past, plant species identification was the exclusive domain of professionals who identified the plants of interest by comparing them with previously collected specimens or by using books or identification manuals [1]. Obviously, learning about unknown plants is also an interesting adventure for amateur gardeners and outdoor enthusiasts. In recent years, image-based plant identification has been an area of growing interest. Plant classification and identification are generally based on the color, shape, texture and other characteristics of plants as the main basis, it always judges from the flowers, fruits, stems, leaves and other organs of the morphological characteristics. Leaf texture, color and morphological structure of plants are distinct, so they are the main basis for the distinction between plant species [2].

Due to the longer survival time of plant leaves and the convenience of image acquisition, the classification of plants by leaves has become the research direction of many scholars at present. In China, Wang et al. developed a method of identification of foliage from plants based on extraction of multiple features of leaf images. The system extracts the color, shape and texture features of the image of the foliage and implements identification based on the Support Vector Machine (SVM), achieving an identification rate of 91.41% [3]. Yang et al. analyzed the identification of seven Salix species using digital information analysis of leaf image characteristics and compared them with multiple features based on leaf-related features. The comprehensive identification accuracy reached 90.8% [4]. Based on the geometric and texture features of leaves, Zhang et al. proposed a classification method based on clonal selection algorithm and K nearest neighbors, which effectively improved the identification rate [5]. Besides, the CLEF (Cross Language Evaluation Forum) has organized multiple...
plant image identification competitions since 2011 and added identification conditions for different organs and complicated photographing situations in the 2014 competition. In 2015, the plant species database is up to 1000. In ImageCLEF 2011, Villena-Román et al. used SIFT keypoints for object detection in the field of plant image identification [6]. Backes et al. used a complex network to analyze leaf characteristics and identify them, which has good effects on leaf image identification with different resolutions and noisy [7]. A smart phone-based plant identification system called Leafsnap have been developed recently [8], which involves digitizing the plant image of a solitary flattened leaf and comparing it with a database of previously scanned leaves from some known plant species. By comparing the test plant image with the template images in the database, Leafsnap does not pay enough attention to both large intra-species visual variations and strong inter-species visual similarities which are typical in the botanical domain [9].

Although many advances have been made in plant identification research, there remain lots of problems. For example, in the process of leaf identification based on image analysis, the feature of extraction is usually determined by manual analysis. The differences among plant species are not solved, and the differences among plant data sets with the same features will be produced. In the case of complex background shooting, the identification accuracy of conventional methods decreased significantly. In addition, traditional leaf identification also has some disadvantages. The images contain too little information about the plants and do not have complex backgrounds. To this end, we have created a dataset of 27 different Orchidaceae plants that contain more complex information about the flowers, fruits, branches, stems and leaves.

Deep learning algorithm based on convolutional neural network can learn image features autonomously and reduce manual intervention. It can eliminate noise interference and improve image identification efficiency for leaf image with complex background. Therefore, we construct a bi-channel deep learning framework based on the convolutional neural network, train the model by using the Orchidaceae dataset that we created, and then conduct the test of identifying the plants. In order to verify the validity of the proposed framework, we will use some traditional methods to do the same experiment for comparative analysis. Experiments show that our proposed bi-channel deep learning framework has achieved good effect of identification.

2. Proposed method

2.1. Bi-channel Deep Learning Framework
The bi-channel deep learning framework consists of two sub-network layers: (1) VGG16Net based layer; (2) SqueezeNet based layer. The Figure 1 shows the architecture of our proposed network.

![Figure 1. The architecture of bi-channel deep learning framework](image-url)
2.1.1. **VGG16Net.** The VGG16Net was trained on the ImageNet1000 dataset and used for image classification. The format of input data is 244×224×3 pixels, after 13 rounds of processing of convolutional neural networks and pooling networks, we can get the feature data of 4096-dimension, and then through the 3 fully connected layers of neural network processing, finally we can obtain the standard softmax classification results. The size of overall convolution kernels is 3×3, concluding some special convolution kernels with 1×1, all the activation units in hidden layers are RELU. The pre-trained VGG16Net network is used while the output of the final layer is changed for the plant identification.

In order to adapt to the VGG16Net model for our plant identification problem, we fine tune the model using the constructed Orchidaceae dataset. In the case of plant identification, we have 27 classes of plants. So the output of the last fully connected layer is changed from 1000 to 27, which is change the number of classified species. After making these changes, we resume the pretraining parameters on ImageNet1000 for all layers as initialized parameters and then continue training on our dataset to fine tune all the parameters.

2.1.2. **SqueezeNet.** SqueezeNet was trained on the ImageNet1000 dataset and used for image classification. There are totally 9 fire modules of SqueezeNet, interspersed with some of the max pooling layers, the last global average pooling layer (AVG POOL) instead the fully connected layer (FC) of the traditional network. These changes make the parameters greatly reduced. In the beginning and the end of the network, there are two simple conv layers to ensure that the input and output size can be mastered. Fire module is a special module which splits the original conv layer into two layers: squeeze layer and expand layer, each with Relu activation layer. There are 1x1 convolution kernels in the squeeze layer, 1x1 and 3×3 convolution kernels in the expand layer. Behind the expand layer, the convolution of 1×1 and 3×3 output feature maps are spliced together in the channel dimension.

In order to adapt to the SqueezeNet model for our plant identification problem, we fine tune the model using the constructed Orchidaceae dataset. So we also have to change the output of the last average pooling layer from 1000 to 27, which is change the number of classified species. After making these changes, we resume the pretraining parameters on ImageNet1000 for all layers as initialized parameters and then continue training on our dataset to fine tune all the parameters.

2.2. **Classification and Learning**

As shown in Figure 1, the common classification layer with fully-connections is used to fuse the predictions from two networks. The softmax function is employed to predict the probability of recognizing plants for two deep models. It is defined by:

\[
\gamma_c = \frac{\exp(W_i^T V)}{\sum_{c=1}^{C} \exp(W_i^T V)}
\]  

Where \(\gamma_c\) is the predicted probability of \(c\)th plants, and \(V\) denotes the output feature vector of last pooling layer.

The fusion operation is performed on the outputs of two deep models and the fusion results are used to predict the images. It is defined by:

\[
c_i = \gamma_i a_i + (1 - \gamma_i) b_i, \quad i = 1,2,\ldots,27
\]

Where \(c_i\) is the final output of the \(i\)th plants. \(a_i\) and \(b_i\) are the output of the two sub-networks respectively. \(\gamma_i\) is the fusion weights and can determined by grid search.

Two deep models based on VGG16Net and SqueezeNet are pre-trained in the ImageNet1000 database and then fused for the use in our Orchidaceae dataset. When training the model, the logloss function is used to learn the parameters by the Stochastic Gradient Descent (SGD) algorithm. For the test, we calculate the top1 and top5 accuracy to measure the final identification results. Top1 accuracy is the accuracy that the type of the highest output value is the correct one, and Top5 accuracy is the performance that the types of five highest output values contain the correct one.
3. Experiments

3.1. Experimental Setup

In our experiment, we have created a dataset of 27 different types of Orchidaceae. The training set contains 16,455 images, and the test set contains 1756 images. Images are of different backgrounds and different resolutions and contain a wealth of information, such as flowers, fruits, branches, stems, leaves and so on. All the images in this dataset are crawled from search engines such like Flickr, Google and Bing through keywords for plant species interpretation. These keywords are collected from a professional plant taxonomy website called “The Plant List” and shown in Table 1.

| Keywords                                                                 |
|--------------------------------------------------------------------------|
| Angraecum-calceolus, Bulbophyllum-ecornutum, Bulbophyllum-guttulatum, Calanthe-furcata, Cattleya-dormaniania, Cleisostoma-crochetii, Coelogyne-corymbosa, Cycnoches-cooperi, Dendrobium-acerosum, Dendrobium-furcatum, Dendrobium-tortile, Diuris-aurea, Epidendrum-stamfordianum, Eulophia-graminea, Lepanthes-bradei, Leptotes-bohnkiana, Masdevallia-racemosa, Mormodes-warszewiczii, Ophrys-vetula, Paphiopedilum-druryi, Phaia-wallichii, Phalaenopsis-pulcherrima, Platystele-compacta, Psychopsis-sanderae, Serapias-orientalis, Thelymitra-ixioides, Vanilla-planifolia |

We first crawl all the species names under Orchidaceae from the “The Plant List”, and then we use a python-based crawler to get images from different search engines through these keywords of Orchidaceae species. After that, a clustering-based rubbish image filter algorithm is used to filter out the noisy images. Finally all images are manually cropped and labelled following [10]. Some exemplar images are shown in Figure 2.

![Figure 2. Exemplar images from the dataset](image.png)

In order to evaluate the validity of the proposed framework, we use several traditional networks to conduct the same experiment and compare the experimental results. The first part of the comparison...
contains the traditional VGG16Net network [11] without any pretraining parameters (denoted as VGG16Net-1), the pretrained VGG16Net network without pretraining parameters updated (denoted as VGG16Net-2), and the fine-tune VGG16Net network with pretraining parameters updated (denoted as VGG16Net-3). The pretrained VGG16Net-2 network without pretraining parameters updated is set up as follows. First of all, each image in the training set and the test set is taken as the input data of the VGG16, and then we can get a 4096-dimensional feature vector as the output, directly replacing all the original data as the input of the behind layer. We add a 256-dimensional fully connected layer, a 27-dimensional fully connected layer and a softmax layer. The fine-tune VGG16Net-3 network updates the pretraining parameters of all the layers on our Orchidaceae dataset. The second part of the comparison contains AlexNet [12], VGG16 network [11], SqueezeNet (the three traditional networks are all fine-tuned) [13], and our bi-channel deep learning framework. The AlexNet network has 8 layers including 5 convolutional layers and 3 fully connected layers. The excitation function RELU and Local Response Normalization (LRN) are included in each convolutional layer.

We divide the training set into the training set and the validation set, using all the methods to train model on the training set, and validate the accuracy on the validation set after every training epoch. Finally, we conduct test on the test set, obtaining the accuracy as the assessment of these methods.

3.2. Experimental Results

As Table 2 shows, when using the traditional CNN network for classification, if the pre-training is not performed, the identification effect is poor at the same number of iterations. If feature extraction is performed on the basis of a pre-trained model and then the new layers are added for training, good results will be achieved. The advantage of this method is allowing us to use the existing model with a small amount of data and training time to achieve good results, which is more convenient to be used for simple classification problems with a small dataset. Moreover, the parameters of this network have been reduced, which cut down the training difficulty. But it's worth noting that if we update the pretraining parameters of all the layers, the identification result will be great. The benefit of fine-tuning is that it does not have to completely retrain the model, which increases efficiency, as the accuracy of new training models generally increases slowly from very low values, but fine-tuning allows us to get a better result with less iteration.

Table 2. The identification accuracy of VGG16 approaches for comparison on the Orchidaceae dataset.

|                | VGG16Net-1 | VGG16Net-2 | VGG16Net-3 |
|----------------|------------|------------|------------|
| Top1 accuracy  | 75.00%     | 88.55%     | 95.39%     |

As Table 3 shows, In most configures, the deep method outperforms the shallow models, while in some cases the shallow models are competitive with deep models, that might be due to the insufficient samples. Furthermore, our proposed method outperforms the other two deep models. Although VGG16Net achieved better performance than the SqueezeNet in ImageNet1000 classification, it can be seen from the experimental results that the performance of SqueezeNet and VGG16Net are quite for our small dataset, which shows that even a very small network can achieve the good CNN identification accuracy. By combining the two networks into a unified framework, the performance of plant identification can be improved.

4. Conclusion

We proposed a bi-channel deep learning framework for plant identification. In this model, the CNNs based on VGG16Net and SqueezeNet are employed to automatically extract the features. These two models are then learned jointly with shared classification layer. Models are pre-trained in the ImageNet1000. Experimental results assessed the effectiveness of our proposed approach which outperformed existing conventional networks.
Table 3. The identification accuracy of some fine-tuned CNNs and proposed Bi-channel Framework for comparison.

|                 | AlexNet | VGG16Net | SqueezeNet | Bi-channel Framework |
|-----------------|---------|----------|------------|----------------------|
| **Top1 accuracy** | 94.82%  | 95.39%   | 95.73%     | 96.81%               |
| **Top5 accuracy** | 99.51%  | 99.66%   | 99.26%     | 99.54%               |

Acknowledgments
This work is partly supported by the National Aerospace Science and Technology Foundation and the National Nature Science Foundation of China (No. 61402368 and No. 61702419).

References
[1] Cowan R, Chase M, Kress W, Savolainen V 2006 Taxon 55(3).
[2] Zhang S, Huai Y J 2016 J. Beij. Fore. Univ. 38(9).
[3] Wang L J, Huai Y J, Peng Y C 2015 J. Beij. Fore. Univ. 37(1).
[4] Yang T T, Pan X X, Mu L Q 2014 J. North. Fore. Univ. 42(12).
[5] Zhang N, Liu U W P 2013 J. Comp. Appl. 33(7).
[6] Villena Roman J, Lana Serrano S, Cristobal J C G 2011 Amsterdam: Clef Labs & Workshop.
[7] Backes A R, Casanova D, Bruno O M 2009 Patt. Reco. 42(1).
[8] Kumar N, Belhumeur P N, Biswas A, Jacobs D W, Kress W J, Lopez I C, Soares J V 2012 ECCV.
[9] Tera Botanica [www.tela-botanica.org](http://www.tela-botanica.org).
[10] Zhang H X, Fan J P 2017 IEEE MIPR.
[11] Parkhi O M, Vedaldi A, Zisserman A 2015 BMVC.
[12] Krizhevsky A, Sutskever I, Hinton G E 2012 NIPS 60(2).
[13] Iandola F N, Han S, Moskewicz M W, Ashraf K, Dally W J, Keutzer K 2016 preprint (arXiv:1602.07360).