Weak Constraint Leaf Image Recognition based on Convolutional Neural Network

Euncheol Kang, Il-Seok Oh
Division of Computer Science and Engineering
Chonbuk National University
ec.kang@jbnu.ac.kr, isoh@jbnu.ac.kr

Abstract

Recently the computer vision and machine learning research communities pay a great attention to the leaf image recognition problem. Our literature survey focusing on the user interaction aspect reveals that two schemes of image acquisition have been used, one with strong constraint and the other with no constraint. The strong constraint interaction asks users to capture images by placing a leaf on a uniform background such as white paper while the unconstrained interaction allows any form of image capturing. The former one gets a high performance sacrificing the user convenience while the latter one provides a great convenience sacrificing the recognition performance. Our scheme is weakly constrained in the middle of two extremes. The proposed interaction scheme only asks users to center the leaf on smartphone camera screen. The leaf may be on the tree or off the tree. When the leaf is picked off the tree, it is recommended to place it against rather uniform background such as sky, soil, or tree bark. By fine-tuning the pre-trained CNNs(Convolutional Neural Network), we obtained a practical performance, 96.08% top-1 and 99.81% top-5 accuracies. The dataset is publicly open and the recognition system is released as an Android App.

Keywords: Automatic leaf recognition, convolutional neural network, deep learning, fine tuning.

1. Introduction

Many people often wonder the name of a plant when they walk on the road. However, it is often not an easy job to know the plant name right away. As everyone has own smartphone, there is a lot of demand for automatic plant recognition. Recently, research has been carried out in various groups to facilitate the problem of plant recognition. Some studies try to recognize plants with flowers and fruits, etc., but others focus on leaf recognition. Leaves are easier to collect than flowers and fruits and less affected by seasonal change.

The Leafsnap has started an interest in the leaf recognition [1]. The interaction of image acquisition used by the Leafsnap App is strongly constrained in the sense that a user has to carry a white paper when going in the wild. The Clef group initiated a big project whose interaction is unconstrained in the sense that a user freely acquires the leaf images [2, 3]. Some sample images can be seen in Fig.1. Currently between these two extremes, most of the researches belong to the strong constraint [4, 5].

In this paper, we propose a weak constraint user interaction that balances well the user convenience and system performance. Our constraint asks users only to center the leaves on the camera screen.

We collected 2,739 leaves in the wild with weak constraints on 63 species of trees with various natural backgrounds. We implemented CNN using the caffe framework[6]. We fine-tuned AlexNet[7], VGGNet[8], and GoogLeNet[9] using our dataset. Top1 accuracy of 96.07% and top5 accuracy of 99.81% was obtained. We developed an Android-based application.

2. Related Works

In building a leaf recognition system, the dataset plays a critical role and influences heavily on system performance. So we first summarize the well-known datasets and then describe a brief survey of recognition methods.

2.1 Datasets

The ImageCLEF is a big project initiated in Europe to collect images and to develop methods for automatic annotation and multimodal retrieval [3]. The public donation through Internet is a main way of the image collection. The plantCLEF is a part of ImageCLEF and it contains a large volume of images from a large variety of plant species. The images include various parts of the trees such as fruit, leaf, flower, bark, branch, and whole tree. The dataset has 113205 images from about 1000 species collected around France and neighbor countries.

The Leafsnap team who released the first plant recognition App provides the leaf image dataset[1]. The images were collected from 185 species on Northeastern part of USA. The dataset contains 30866 images which consist of 23147 lab images and 7719 field images. The images were captured as the leaf being placed on a white paper.

An attempt with the aim of scientific archiving rather than collecting the training set is also available. Agarwal et al. constructed a collection of type specimens at the Smithsonian Institution Department of Botany[10].

Table 1 summarizes 5 datasets including the fore mentioned 2 datasets and Flavia, MalayaKew[5] and our own dataset. Our dataset is named as Moip. Figure 1 shows sample from five datasets.

2.2 Recognition methods

The methods can be categorized into 4 groups depending on two criteria of the background uniformity and the way of feature extraction. With respect to the first criterion some methods require uniform background which corresponds to the strong constraint interaction. The other allows non-uniform background which corresponds to the weak and no constraints. With respect to second criterion, we can divide the methods into hand-crafted feature design and automatic feature learning. The automatic feature learning is fulfilled with the deep learning.

The first group uses the uniform background and hand-crafted features. The Leafsnap segments a leaf image of white background using EM algorithm[1]. After segmentation, a feature vector is extracted by computing the multi-scale curvatures along the contour. The classification is accomplished by using kNN algorithm. The recognition accuracy of Leafsnap was reported to be 58.9% in[5]. Lots of the papers including recent results belong to this group[11, 12].

The second group recognizes the uniform background images using deep learning. The Deep-plant uses their own dataset acquired with uniform background. They train CNN using data augmentation that crops venation patches without using leaf shapes[5]. The Treelogy combines hand-crafted features and features extracted by CNN. The latter features are obtained from fine-tuning the pre-trained CNN[4].

The third group uses the non-uniform background and hand-crafted features. Since the non-uniform background presents vast amount of variability, the hand-crafted features cannot properly cope with this difficulty. So, a little paper could be found. The most intensive works are a comparative study of preprocessing and segmentation of leaves from images of non-uniform background[13]. They concluded that the Guided Active Contour segmentation algorithm achieved the best performance. The effect of user's input stroke was also analyzed as a mean to enhance the performance.

The final group is the most challenging and reflects recent research trend. As the deep learning becomes popularized due to its high performance and automatic feature learning, the recognition approaches were naturally moved from the hand-crafted feature methods to the deep learning methods [14]. In addition, the publicly available high quality and easy-to-use deep learning software platforms such as Caffe, Theano, and Tensorflow resulted in a quick transition from hand-crafted method to the deep learning method. A lot of methods can be found from the LifeCLEF plant challenge [2, 3]. The participants used to adopt the pre-trained models such as GoogLeNet, VGGNet, and AlexNet. The champion of 2016 was the Bluefield system[3]. The system used a VGGNet with spatial pyramid pooling, parametric ReLU. It provides unknown class rejection option based on the minimal prediction score[15]. Our approach belongs to the last group.

3. Our Approach

The traditional approach to the leaf recognition uses the hand-crafted segmentation and features. However due to the uneven surfaces of leaves, the approach cannot completely solve even the uniform background case. The deep learning is expected to be only the viable way of constructing a system for automatically recognizing leaves in the wild. Using the deep learning technology, we build an end-to-end system without suffering from separately designing the hand-crafted segmentation and feature extraction.

In the following, we will describe our deep learning approaches to recognizing the leaf images taken in the wild with the weakly constrained user interaction. Before presenting our CNN implementation, our dataset named Moip is described.

3.1 Collection of Moip dataset

Our image acquisition scheme asks a collector only to center the leaf on camera screen. Our collection is divided into two ways, one acquiring images as the leaf being on tree and the other picking the leaf off the tree and using nearby objects as background. The collector used the sky, soil, tree bark, and white paper as background.

We collected around 45 images from 63 woody plant species and constructed the Moip dataset having 2739 images. We used iphone and galaxy smartphones. The area of collection is the campus of Chonbuk National University located Korea. Figure 2 illustrates one sample for each of 63 species which was acquired on tree.

3.2 CNN implementation

The CNN is currently the most successful deep learning model, especially for the image recognition tasks[14]. The prominent advantage of the CNN is the transfer learning. The network pre-trained using the large-sized natural images can be transferred to a new task by fine-tuning to the small-sized dataset collected from the new task. Very conveniently, three high-end pre-trained CNNs are available, AlexNet[7], VGGNet[8], and GoogLeNet[9]. In our implementation, all of three CNNs available on Caffe platform were used. The performance of three networks will be also compared in Section 4.

Since our dataset size is 2739, using the original set is prone to overfitting. So in order to regularize the CNNs, we augmented the dataset by rotation transformation. Since the leaf is centered on image, we used 0, 15, 30, 150, 165, 180, 195, 210, 330, 345 degrees to avoid severely extending out of the boundary. The empty space is filled with 0. By this transformation, the dataset has been augmented 10 times.

3.3 Visualizing operation details of CNN

To overcome the limitations of viewing CNN operations only as a blackbox, many visualization tools have been proposed and made available publicly. To scrutinize what our CNNs are doing in deep layers with leaf images, we used the tool proposed by Yosinski[16]. Figure 3 shows the feature maps (first row) and deconvolution results (second row) produced at each of convolution layers. The figure shows only for AlexNet. From each layer, only four maps are chosen and displayed in a large resolution so that the map details can be easily identified. A feature map illustrates the convolution result by learned mask.

Table 1. Comparison of 5 leaf image datasets

	plantCLEF 2015 [3]	Leafsnap [1]	Flavia [10]	MalayaKew [5]	Moip
Number of images	113205	30866	1907	2816	2739
Image size	various	about 700x525	1600x1200	256x256	300x400 or 300x533
Number of species	1000	185	33	44	63
Background	nature	white	white	black	nature, bark, sky, white
Segmentation ground truth	X	O	X	O	X
Image acquisition constraint	None	Strong	Strong	Strong	Weak

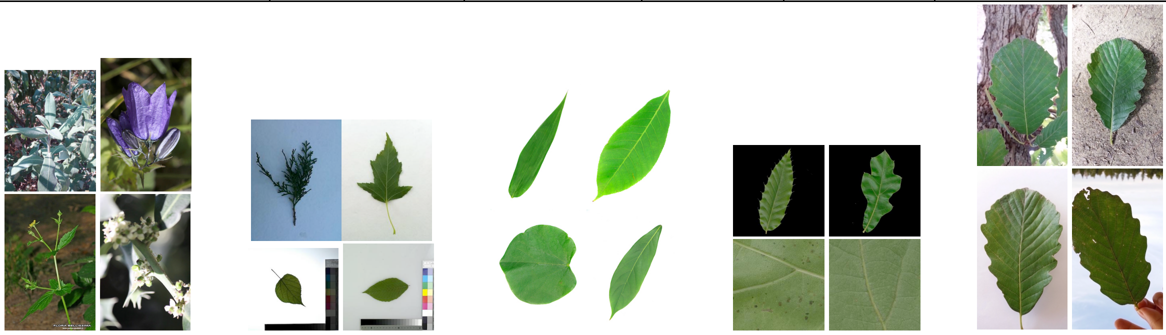

Figure 1. Sample images from five datasets

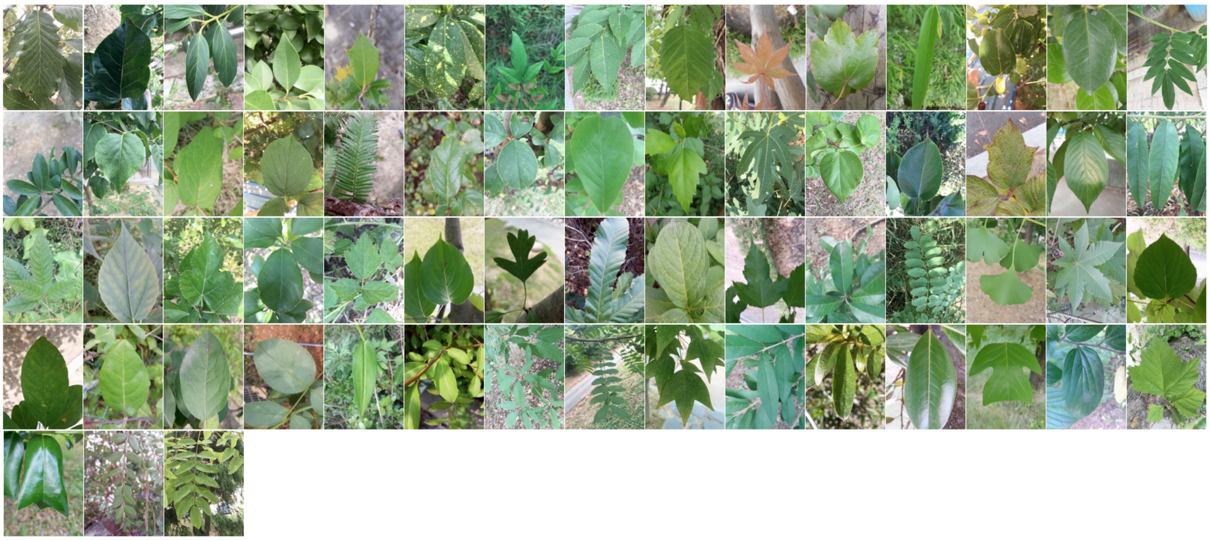

Figure 2. Moip dataset (one sample image per each of 63 species captured as being on the tree

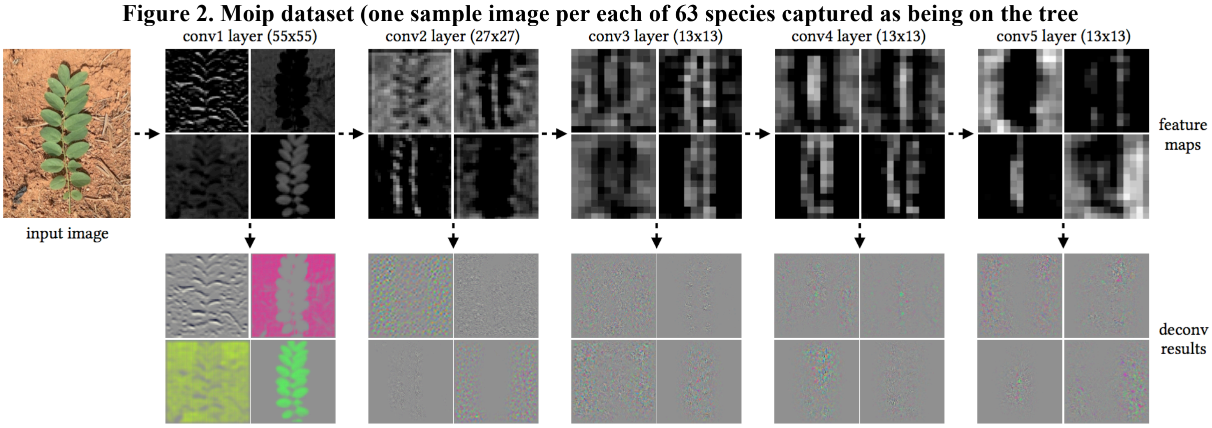

Figure 3. Feature maps (first row) and deconvolution results (second row)

3.4 Application

Using the trained CNN, we built a server-based Android App named Moip. The Moip is available on the Play store. The App provides top-5 predictions. It asks a user to focus by tapping the center of smartphone screen and to center the leaf on the screen. We believe that a user of our system willingly is faithful to this form of interaction in the hope of getting a right answer from the system.

4. Performance Evaluation

In our experiments, 80% of Moip dataset is used for training and the remaining is used for test. Our evaluation was done for each of on- and off-the-tree data. Another evaluation for the whole dataset is also done.

Table 2 provides the accuracy data obtained by AlexNet. The accuracy is notated as Top-1/Top-5. Since the on-the-tree images have much more clutter on the background, the on-the-tree data reveals lower accuracy. The difference from the off the tree is about 9% for top-1 case while the difference for top-5 is reduced to about 4%. For the whole dataset, top-1 and top-5 accuracies are 90.07 and 99.09%, respectively. The last row of 'whole without fine-tuning' shows the accuracies when AlexNet is retrained after initializing all layers.

Table 3 compares three CNNs after fine-tuning with whole dataset. The VGGNet was the best with top1 and top-5 accuracies of 96.08% and 99.81%.

Table 2. Accuracies by AlexNet (top-1/top-5 accuracies in %)

Data set	Accuracy
On tree	82.59 / 95.38
Off tree	91.45 / 99.45
Whole	90.07 / 99.09
Whole without fine-tuning	47.31 / 79.50

Table 3. Comparison of AlexNet, VGGNet, and GoogleNet (top-1/top-5 accuracies in %)

CNN	Accuracy
AlexNet	90.07 / 99.09
VGGNet	96.08 / 99.81
GoogLeNet	93.50 / 99.50

5. Concluding Remarks

This paper proposed a nice user interaction scheme called the weak constraint for the leaf recognition. The scheme balances well between user convenience and system performance. The experiments with the pre-trained CNNs showed high accuracies that are close to the field deployment. The App and dataset are released publicly with the aim of expanding the plant recognition research community.

The most important future work is focused on the study of field user's requirement and satisfaction. Increasing the number of species recognizable by the system is a vital element. One-shot learning is a good way since the burden of leaf image collection is greatly reduced [17]. We think that the leaf recognition problem would be a good application of the one-shot learning because one leaf image allows usually to identify the species and factor of variations is limited compared to other natural and man-made objects. Another aspect regarding the user satisfaction is the precision of recognition confidence and the rejection option. The field users would input much more diverse forms of images compared to our Moip dataset. The diversity may include the non-leaf, unconstrained images, blurred images, and species not supported. Ensemble of three or more CNNs is a simple method of increasing the belief on the confidence. We can use the belief in implementing the rejection option.

References

[1] N. Kumar, P. N. Belhumeur, A. Biswas, D. W. Jacobs, W. J. Kress, I.` C. Lopez, J. V. B. Soares, "Leafsnap: A computer vision system for automatic plant species identification" In *European Conference on Computer Vision*, vol 7573, pp. 502-516, 2012.

[2] A. Joly, H. Goëau, H. Glotin, C. Spampinato, P. Bonnet, W.P. Vellinga, R. Planqué, A Rauber, S. Palazzo, B. Fisher, H. Müller, "LifeCLEF 2015: Multimedia Life Species Identification Challenges" In *International Conference of the Cross-Language Evaluation Forum for European Languages*, vol 9283, pp. 462-483, 2015.

[3] A. Joly, H. Goëau, H. Glotin, C. Spampinato, P. Bonnet, W.P. Vellinga, J. Champ, R. Planqué, S. Palazzo, H. Müller, "LifeCLEF 2016: Multimedia Life Species Identification Challenges" In *International Conference of the Cross-Language Evaluation Forum for European Languages*, vol 9822, pp. 286-310, 2016.

[4] İ. Çuğu, E. Şener, Ç. Erciyes, B. Balcı, E. Akın, I. Önal, A. O. Akyüzz, "Treelogy: A Novel Tree Classifier Utilizing Deep and Hand-crafted Representations" *arXiv preprint arXiv:1701.08291*, 2017.

[5] S. H. Lee, C. S. Chan, P. Wilkin, P. Remagnino, "Deep-plant: Plant identification with convolutional neural networks" In *IEEE International Conference Image Processing*, pp. 452-456, 2015.

[6] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, T. Darrell, "Caffe: Convolutional architecture for fast feature embedding" In *Proceedings of the 22nd ACM international conference on Multimedia*, pp. 675-678, 2014.

[7] A. Krizhevsky, I. Sutskever, G. E. Hinton, "Imagenet classification with deep convolutional neural networks", In *Advances in neural information processing systems*, pp. 1097-1105, 2012.

[8] K. Simonyan, A. Zisserman, "Very deep convolutional networks for large-scale image recognition" *arXiv preprint arXiv:1409.1556*, 2014.

[9] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, "Going deeper with convolutions" In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1-9, 2015.

[10] G. Agarwal, P. Belhumeur, S. Feiner, D. Jacobs, W. J. Kress, et al, "First steps toward an electronic field guide for plants" In *Taxon*, vol 55, no 3, pp.597-610, 2016.

[11] S. G. Wu, F. S. Bao, E. Y. Xu, Y. X. Wang, Y. F. Chang, Q. L. Xiang, "A Leaf Recognition Algorithm for Plant Classification Using Probabilistic Neural Network", In *Signal Processing and Information Technology, 2007 IEEE International Symposium* on, pp. 11-16, 2007.

[12] J. Chaki, R. Parekh, S. Bhattacharya, "Plant leaf recognition using texture and shape features with neural classifiers" In *Pattern Recognition Letters*, vol 58, pp. 61-68, 2015.

[13] M. Grand-Brochier, A. Vacavant, G. Cerutti, C. Kurtz, J. Weber, and L. Tougne, "Tree leaves extraction in natural images: Comparative study of preprocessing tools and segmentation methods" In *IEEE Transactions on Image Processing*, vol 24, no 5, pp. 1549-1560, 2015.

[14] I. Goodfellow, Y. Bengio, and A. Courville, "Deep Learning", *MIT Press*, 2016.

[15] S. T. Hang, A. Tatsuma, M. Aono, "Bluefield (kde tut) at lifeclef 2016 plant identification task", In *Working notes of CLEF 2016 conference*, 2016.

[16] J. Yosinski, J. Clune, A. Nguyen, T. Fuchs, H. Lipson, "Understanding neural networks through deep visualization" *arXiv preprint arXiv:1506.06579*, 2015.

[17] B. M. Lake, R. Salakhutdinov, J. B. Tenenbaum, "Human-level concept learning through probabilistic program induction", In *Science*, vol 350, issue 6266, pp. 1332-1338, 2015.

[18] Z. Q. Zhao, B. J. Xie, Y. Cheung, and X. Wu, "Plant Leaf Identification via a Growing Convolution Neural Network with Progressive Sample Learning" In *Asian Conference on Computer Vision*, pp. 348-361, 2014.