What's That Plant? *WTPlant* is a Deep Learning System to Identify Plants in Natural Images

Jonas Krause[1]
krausej@hawaii.edu
Kyungim Baek[1]
http://www2.hawaii.edu/~kyungim
Lipyeow Lim[1]
http://www2.hawaii.edu/~lipyeow
Gavin Sugita[1]
gsugita6@hawaii.edu

[1]Dept. of Information and Computer Sciences
University of Hawai'i at Mānoa
1680 East-West Road
Honolulu, HI 96822, USA

Abstract

Automatic identification of plant species from natural images is a challenging problem with many practical applications in multiple disciplines. An accurate and automated system for plant species identification has important implications in addressing botanical taxonomy gaps, identifying new species, controlling the balance of ecosystems, and estimating yield and resource requirements in agriculture. However, identifying plants from uncontrolled natural images is a challenging problem due to the complexity of natural images, a large number of plant species, inter-species similarity, and the large-scale variance in appearance. In this work, we present a system called *WTPlant*, specifically designed for identifying plants in natural images. By assembling a collection of Convolutional Neural Networks with stacked/residual blocks and a preprocessing stage for multiscale analysis, *WTPlant* presents itself as a highly discriminative deep learning approach for this image classification problem.

1 Introduction

Knowledge of plant species is essential to protect the biodiversity of any flora. Traditionally, botanists analyze different characteristics of plants as identification factors. But identifying the plant species accurately based only on visual characteristics requires considerable expertise [1], which is almost impossible for the general public and challenging even for specialists. Therefore, an automated system to identify plants has important implications for the society at large not only in the preservation of ecosystem biodiversity including public education, but also in agricultural activities such as automatic crop analysis, species variability analysis, analysis phylogenetic relationships, identification of pests and diseases, and identification of invasive species. The improvement in these agricultural efforts can, in turn, lead to better crop control and management, higher yielding food production, and possibly a reduction in pesticide use.

Approaches using computer vision techniques for automated plant identification from controlled images have shown promising results [2, 3, 4, 5, 6]. Nevertheless, a realworld plant identification application needs to handle natural images, which is a big challenge for automated computer vision systems. Analysis of unconstrained natural images can be extremely difficult due to factors related to complex background, illumination, occlusions, shadows, and a rich local covariance structure that is usually present in these images. While human visual sysem deals with those factors with ease, an equivalent computational model for plant identification from natural images is still an open problem.

In the past decades, Machine Learning (ML) methods have shown promising results in various computer vision tasks including plant identification. Most of the previous efforts for identifying plants used hand-designed features of leaves and flowers [1, 2, 7, 8, 9] and are restricted to fairly controlled images with clean backgrounds. However, identifying plant species based on morphological characteristics extracted from well-controlled images is quite different from handling the noisy natural images that are found in real-world problems. More recently, Deep Learning (DL) approaches [3, 4, 5, 6, 10, 11, 12, 13] have been introduced to analyze plant images driven by the success of the Convolutional Neural Networks (CNNs). The use of deep convolutional approaches has been a growing trend in computer vision, demonstrating impressive results in various tasks using natural images.

In this work, we present a DL-based plant identification system called *WTPlant* that uses a novel framework consisting of multiple preprocessing stages and multiple CNN pipelines. In contrast to existing plant identification methods that use handdesigned features, simple CNN architectures, and pre-trained models, *WTPlant* uses a collection of CNNs to classify leaves and flowers separately, and then combine their predictions to achieve more accurate identification results.

In summary, *WTPlant* consists of multiple pipelines and stages of different CNN components designed to extract deep multi-scale discriminatory features. These pipelines segment the query image, preprocess the regions of interest into samples of different sizes and classify them using deep CNNs. The separation between the processing of leaves and flowers allows the networks to learn filters specifically for each task in order to analyze the different types of input better. The results from each pipeline are then combined to obtain more accurate predictions in a process reminiscent of ensemble techniques. This version of *WTPlant* is trained to classify 100 different plant species found on the campus of the University of Hawai'i (UH) at Manoa. Preliminary experiments show that the initial segmentation process helps guide the extraction of representative samples and, consequently, enables CNNs to better recognize plants at different scales in natural images.

Section II of this paper presents the related work by describing two of the most famous methods for each of previously studied feature extracting approaches. In Section III, *WTPlant* system is described in greater detail by explaining each pipeline, its structure, the novel preprocessing stage for multi-scale analysis, and the implemented CNN architectures. Section IV presents the experimental results showing the prediction accuracy of our system and comparing it with other commonly used methods. In the end, Section V concludes the presentation of this system and describe future work.

2 Related Work

Wäldchen and Mäder [1] systemically analyzed previous studies on producing an automated plant identification system. They list 120 papers that used only hand-designed features, showing the relationship between each extracted feature and the identification

factor analyzed in the plants. Most of the reported approaches rely on shape identification factors to correctly classify leaf images. This is mainly due to the fact that sample images were created by placing each leaf on a flat background and taking individual pictures. This approach has been used for the last few decades and yielded good applications such as the LeafSnap [2]. Other reviewed papers focused on flower images classification, redirecting the feature extraction from morphological features to textural ones. Approaches that rely only on flower images to identify species are not very common due to the short flowering period of the plants. This is a strong indicator that a system combining individual leaf and flower analyses may result in a more robust method. Despite this observation, most of 120 reviewed approaches require very specific leaf pictures to work properly, which makes them unsuitable for identifying plants in natural images.

2.1 Hand-Designed Feature Approaches

Two of the most famous plant identification applications that use hand-designed features are LeafSnap [2] and Folia [8], where the former uses images of leaves on a plain background while the latter works by segmenting leaves from natural images and then extracting hand-designed features.

LeafSnap is probably the most well-known mobile app for leaf image classification. It all started in 2003 with computer science professors from Columbia University and the University of Maryland[1]. Yet, faced with the difficulty of analyzing natural images, they introduced the idea of taking pictures of a single leaf in a solid light-colored background to facilitate shape discrimination. Presented by Kumar et al. [2] in a more recent paper, LeafSnap is described as the first framework created to classify plants using automated computer vision methods. The description of this end-to-end application details the process of classifying leaf images among 185 tree species. This system relies mainly on hand-designed features for this classification, but other computer vision techniques were also applied. An interesting characteristic of this application is that it saves GPS coordinates and timestamp of each photo taken, hoping to be able to map the biodiversity of a region over space and time. While LeafSnap can be used in the field, a limitation is that it requires a single leaf specimen to be photographed against a plain background such as a sheet of plain paper and hence LeafSnap is still not a general solution to identifying plant species from natural images.

Folia is another plant identification application that uses hand-designed features to segment and classify leaves in natural images. Cerutti et al. [8] presented this application that collects the same leaf features that botanists use to classify tree species. They include leaf size, global shape, venation, basal and apical shapes, type of margins, number of lobes and others. Although natural images with natural backgrounds were used, authors selected only non-compound simple leaf images with several lobes, centered and vertically-oriented. With 50 plant species studied, they reported good classification performance when compared with other non-DL methods. Even with these restrictions, Folia may be the application that best targets a leaf in natural images and some of its approaches have inspired the implementation of the *WTPlant* system. Nonetheless, segmenting plants from natural images is by itself a big challenge, and for this, state-of-the-art methods such as DL networks are yielding satisfactory results.

[1] ssec.si.edu/stemvisions-blog/leafsnap-turns-students-hands-botanists

2.2 Deep Learning Approaches

In recent years, Convolutional Neural Network (CNN), a type of DL model, has been successfully used for image analysis and segmentation tasks. Several different CNN architectures have been used to address the plant identification problem and two of them are discussed below. Most of them are adaptations of standard CNNs to work with plant images; few have presented new approaches designed to address specific aspects of the plant identification problem.

Pl@ntNet[2] is a world-scale participatory platform and information system dedicated to the monitoring of biodiversity through image-based plant identification approaches [14]. In 2015, its classification method was migrated from handdesigned features classifiers to DL ones. This project started in 2010 and has evolved during the last few years with iterative developments based on multimedia information retrieval, data aggregation, and integration by a growing community of volunteers [15, 16]. Nevertheless, a huge improvement on the plant recognition performance was only observed when CNNs were introduced in their classification process [11]. Using an existing CNN architecture called GoogLeNet [17], Pl@ntNet fine-tunes its pre-trained network periodically. The main advantage of this application is that it collects thousands of images from different datasets of European, Indian Ocean, South America, and North Africa floras, and classify plants in natural images among 10K species. Pl@ntNet continues expanding to cover North American species. However, it uses only one predefined CNN as its classification engine and hence is somewhat limited in its capability. In contrast, our system employs a novel combination of multiple CNNs to increase classification accuracy and robustness.

One of the few papers that addressed the classification of entire plants and trees in natural images is presented by Sun et al. [10]. Their work uses high-resolution images of 100 plant species with individual bushes and trees collected from the Beijing Forestry University campus. The collection, called BJFU100 dataset, is available online[3]. In their images, it is more evident how challenging the classification of plants in natural images can be: The images come in a variety of backgrounds with different illumination, different focal points, different shadows – it is not always possible to clearly identify a leaf of the analyzed plant. In the light of these challenges, a modified version of the Residual Network (ResNet) [18] architecture was proposed to classify these images. ResNets were developed to extract even deeper discriminative features by adding the previous input layer along with the extracted features. Sun et al. used a DL architecture where a pre-trained ResNet is used as a bottleneck between an initial convolutional block and the last layers of the DL network. A key weakness of their approach is the somewhat simplistic preprocessing stage: the original images of 3120x4208 pixel resolution are drastically downscaled to fit their first convolutional block that expects a 224x224 input area. Such aggressive downscaling results in a significant loss of relevant information that negatively impacts the classification accuracy. Our approach avoids such drastic downscaling by segmenting and tiling the most relevant regions of the input image.

3 *WTPlant* System

WTPlant (What's That Plant?) is based on deep learning approaches (more specifically, a collection of CNNs) carefully designed to address the problem of identifying plants in

[2] identify.plantnet-project.org
[3] pan.baidu.com/s/1jILsypS

natural images. Imposed challenges consist of segmenting the plant to be analyzed from a complex background, dealing with the scale problem, and developing a suitable architecture deep enough to extract discriminative features among similar species. Respectively, this system addresses these issues by using stacked convolutional blocks for the segmentation process, a novel preprocessing stage for multi-scale analyses, and residual blocks to extract deeper and more discriminative features. By designing two classification pipelines, one for leaves and one for flowers, *WTPlant* is able to more accurately classify plants during their flowering season by exploiting both the visual characteristics of the flowers and of the foliage.

This collection of CNNs brings state-of-the-art DL architectures working together with a common goal: correctly classify plants by analyzing their leaves and flowers. The accuracy of the method is measured by counting the correctly classified top-1 and top-5 species versus the incorrect ones during the testing stages. The initial scope is limited to plants present in Hawai'i. Two main datasets were collected for this task, the UH Manoa Campus plant dataset provided by the Botany Department of the University of Hawai'i and a larger one scraped from the bing.com search engine website.

Fig. 1 presents a diagram of the *WTPlant* system and details the workflow while testing a new image. The workflow begins by sending a copy of the test image to the pipeline for leaves and the pipeline for flowers. Both pipelines start by segmenting the image, but in the flower pipeline, two additional pre-segmentation algorithms are used to augment the main segmentation process. The largest segmented areas for leaves and flowers are then preprocessed to create multiscale representative samples. These samples are then classified by CNNs individually trained for each classification problem. In the final stage, the prediction confidence values from each pipeline are aggregated (somewhat similar to ensemble methods) to output the final plant species prediction.

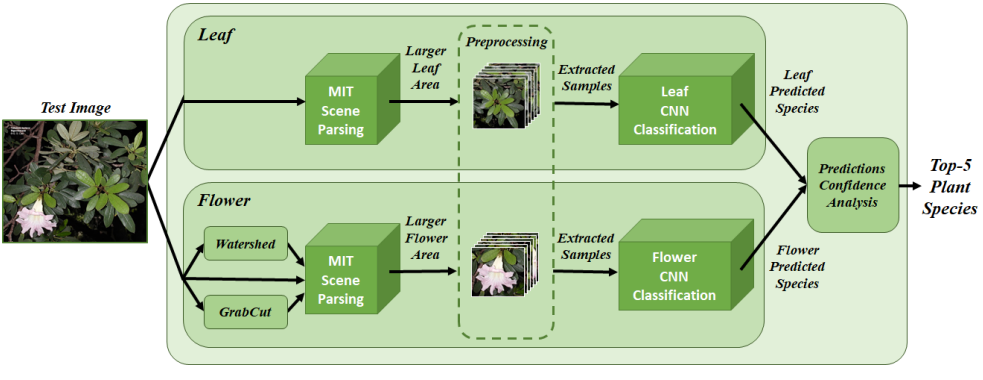

Figure 1: Block diagram of the *WTPlant* system [19].

3.1 Segmentation of Leaves and Flowers

One of the key problems in computer vision is called scene parsing, or recognizing and segmenting objects in an image. Using a CNN architecture with stacked convolutional blocks, Zhou et al. [20] developed a cascade segmentation module for the scene parsing problem (henceforth referred to as MIT Scene Parsing). They trained a three-level stacked

CNN using a dataset called ADE20K[4] to segment common background objects (sky, road, building, etc.), foreground objects (car, people, plant, flower, etc.) and object parts (car wheels, people's head and torso, etc.). The MIT Scene Parsing module is trained to segment 150 different objects from a scene, including plants. MIT Scene Parsing is distributed under the MIT License[5] with an open source initiative. Due to the highly accurate results reported on the segmentation of plants and the usage of stacked convolutional blocks in their process, this method was selected as the segmentation method of choice for the *WTPlant* system.

To better segment flowers in the images, *WTPlant* also uses two additional algorithms, the Watershed Transform [21] and the GrabCut [22], as a pre-segmentation step to the MIT Scene Parsing. The pre-segmentation is needed due to the fact that some plant images have very small flower areas that are not captured by the MIT Scene Parsing initially. Therefore, by roughly separating background from the foreground using the two algorithms, small flowers become more evident for the scene parsing [19]. Preliminary experiments support this approach and showed that these algorithms improve the flower segmentation process significantly.

3.2 Preprocessing

After the segmentation process is performed by the MIT Scene Parsing, Regions of Interest (RoI) delimitating the leaf and flower areas are collected from the input image. If more than one RoI is detected, only the RoI with the largest area is chosen to representing the plant in the image. Identifying multiple plants in an image is part of our future work. If any RoI is collected, meaning that a leaf, a flower, or both a leaf and a flower are detected in the image, the RoI is assumed to contain the most representative information of the plant and is used by the classifiers (individually trained to identify leaves and flowers) to determine the correct plant species. If no RoI is identified during the segmentation, the test image is considered as "No Plant Image". After the RoI (variable size) has been identified, a preprocessing step is implemented to extract fixed size sample images from this RoI to be input into the CNN architectures (which usually only receive images with a fixed size). Our preprocessing step searches for the most representative square areas within the RoI and extracts multiposition and multi-scale representative fixed size samples.

Some of the reviewed approaches [10, 12, 23] suggest that simply downscaling the entire image is a good practice. However, CNN architectures generally take small images as input and a drastic rescale of a natural image to lower resolutions will inevitably result in loss of valuable information. Therefore, a preprocessing approach is needed to properly handle the segmented RoI. Our proposed method is aligned with the reviewed approaches [3, 4, 5] that divided their analyzed regions into smaller samples. The size of these samples is generally the same size as the CNNs input images. This version of the *WTPlant* uses a 224x224 area for the sample extraction, which is the best configuration reported for the ResNet [18]. Since residual blocks are responsible to extract the discriminative features and a correct classification is the main objective, this system follows the specific configurations that this CNN architecture requires.

Extracting samples with multi-scale properties and using them to train the CNNs gives the *WTPlant* system a better scale generalization capability when compared with commonly used preprocessing methods such as resizing and random cropping [19]. To

[4] groups.csail.mit.edu/vision/datasets/ADE20K
[5] opensource.org/licenses/MIT

provide this scale analysis capability, different sized samples are systematically collected from various locations of a RoI. The collected samples represent different areas of the plant at multiple scales. In this way, from one natural image of a plant, numerous representative images are produced. This preprocessing step is responsible for collecting square-sized images for the training of CNNs and testing samples during the classification process.

As an example, Fig. 2 shows an image of the *Adansonia digitata (Bombacaceae)* and six of the samples to be extracted after the segmentation process. The green boundaries delimit the plants detected in the image. However, as previously described, only the largest area is considered for the extraction of representative samples. In this paper, *WTPlant* is set to extract ten different square samples from the largest RoI, one centered on the segmented area (1 in red), four samples dislocated one-third to the right, left, top and bottom (respectively 2, 3, 4, and 5 in orange), and five multi-scale samples covering an extended area (6 in yellow) to the borders of the image. Samples 1 through 5 are extracted with size 224x224 pixels and other larger areas are extracted and reduced to match this size, creating representative samples at various scales. The same approach is implemented if one or more flower areas are detected during the segmentation stage, resulting in ten different flower samples as well. This preprocessing technique is important to improve the robustness of the method addressing the problem of large-scale variance.

Figure 2: Example of the preprocessing stage for the extraction of multi-location and multi-scale representative samples.

3.3 Classification Architectures and Predictions Analysis

The classification engines in this version of the *WTPlant* are two ResNets [18], one for leaves and another for flowers. Training two independent CNNs allows these networks to learn specific filters for each task, producing a good individual analysis of leaves and flowers [19]. After experimenting with various depths of the networks, the most accurate

CNN architectures were integrated into the proposed system. In addition, by training both networks with multi-scale samples extracted from the RoI from the segmentation process, these CNNs were able to learn discriminative features of leaves and flowers at various scales. After the CNN classification engines output the prediction confidence values for each one of the preprocessed samples, the prediction confidence values of the leaf samples are combined with values for the flower samples by summing their confidence results to make the final prediction of the plant species. This strategy also enables our system to work with flowering and non-flowering plant species, such as ferns, mosses, and liverworts as well.

A demonstration of this novel system was recently presented by Krause et al. [19], in which the first version of the *WTPlant* was developed using AlexNet [24] CNN architectures as classification engines. In this paper, the presented system was upgraded with deeper and more complex networks. Experiments with different CNN models (AlexNet and ResNet) with a various number of hidden layers were performed to select the best CNN architecture for each (leaf and flower) analysis. As a result, the classification accuracy increased considerably, as presented in the next section.

4 Results and Discussion

Previous experiments [19] showed that *WTPlant* was able to detect the presence of 99.3% of plants in ~17,000 natural images in the first stage. The high accuracy on the segmentation of plant images is due to the performance of the MIT Scene Parsing module [20] using stacked convolutional blocks and a cascade segmentation approach (which is not the focus of our work). Flower detection accuracy had an average of 72.7%, which improved to 77.5% after augmenting with the two pre-segmentation methods [21, 22]. Using a combination of datasets described in Section III, 100 plant species were selected and 45 images per species were included in the training set.

While AlexNets [24] were used in the previous experiments [19], the new version of *WTPlant* employs ResNet architectures [18] which yielded more accurate predictions in both Top-1 and Top-5 results. In particular, the improvement is significant for Top-1 results reducing the error rate by 19%. Table I (Top-1 results) and Table II (Top-5 results) present a comparison with other approaches commonly used to train CNNs, such as image resizing and random cropping. *WTPlant* and random cropping used ten times (x10) more training samples than the resizing approach. But only *WTPlant* creates multi-location and multi-scale samples for leaves and flowers simultaneously, which gives a great advantage over other training approaches. Presented results were obtained by measuring the classification accuracy of 278 unseen images (testing set) from the 100 species.

All the CNNs were trained using the same learning rate over 100 epochs. The final column of these tables describes the accuracy of the *WTPlant* as a whole, combining the leaf and flower pipelines by summing their prediction confidence. In bold are the best results for leaf and flower, where CNN architectures with 18 layers (ResNet18) for leaf and 34 layers (ResNet34) for flower pipeline outperformed other architectures. Even having deeper layers, ResNet50 did not show better results when compared with smaller residual architectures. These bigger networks generally require more training epochs and data to perform well. However, all presented experiments support the idea proposed by the *WTPlant* system on training CNNs with guided multi-scale samples. In all cases, this guided training process outperformed the commonly used ones and resulted in more generalized networks.

Due to the modular capability of the *WTPlant* system, different CNN architectures (in this case the ResNet18 for leaves and the ResNet34 for flowers) can be used together to predict the final plant species. Consequently, the *WTPlant* performance was further improved correctly identifying **69.09%** for the top-1 and **86.69%** for the top-5 plant species when using ResNet18 to analyze leaves and ResNet34 to analyze flowers. This result suggests that (1) modularizing the DL model and training each module separately is a viable approach, and (2) preprocessing to extract multi-scale representative samples guided by a segmentation process can help in the training of CNNs.

CNN	Resize	Crop	Leaf	Flower	*WTPlant*
ResNet18	39.21%	43.89%	**62.59%**	51.16%	68.35%
ResNet34	40.29%	44.60%	58.27%	**58.14%**	64.75%
ResNet50	28.42%	43.53%	56.47%	44.54%	60.07%

Table 1: Top-1 results.

CNN	Resize	Crop	Leaf	Flower	*WTPlant*
ResNet18	69.42%	70.14%	**85.25%**	82.17%	85.97%
ResNet34	68.35%	71.94%	80.94%	**83.72%**	83.81%
ResNet50	58.99%	69.71%	81.29%	81.51%	82.01%

Table 2: Top-5 results.

5 Conclusion

WTPlant is a new deep learning system specifically designed to identify plants in natural images. In this paper, we present an upgraded version of this system and describe how to deal with the main issues posed by this challenging classification problem. Using state-of-the-art Deep Learning models such as Convolutional Neural Networks, a systematic workflow is presented to analyze leaves and flowers samples collected through the segmentation and preprocessing stages. As shown in Fig. 1, *WTPlant* has two main pipelines carefully designed for plant identification based on leaves and flowers respectively. For the problem of large-scale variance present in natural image analysis, a preprocessing stage is implemented to generate representative samples of different scales. These samples are used to train the classification engines and to test new images, allowing CNNs to analyze plants at different distances. Focusing on the correct identification of leaves, 18-layer residual CNN (ResNet18) presented the most accurate results, while 34-layer residual CNN (ResNet34) outperformed the other tested architectures. ResNet50 may yield better results if trained with more data and for longer periods. The modularity of the *WTPlant* creates a broader analysis of the plant as a whole, where independent CNN architectures work in different regions of the image and combine their predictions to produce more accurate results. Our experiments support the idea that a collection of CNNs specifically designed to work together may overcome the limitations of commonly used methods and monolithic, non-modular DL architectures.

For future research, new DL models and architectures will be trained and incorporated into the system to improve the accuracy and robustness of predictions. New pipelines to analyze different plant organs (e.g. fruit, bark, and seedlings) can also be incorporated easily thanks to the modularity of the system. Data augmentation techniques such as

variation of hue, brightness, contrast, and saturation will be implemented for the final version of the *WTPlant* system. The incorporation of new DL networks, as well as the addition of new plant species and training images, allow this system to be constantly upgraded and improved. Ultimately, our vision is for *WTPlant* to be the most accurate automated plant identification system that is used to benefit society in the areas of conservation, botany, education, and agriculture.

References

[1] J. Wäldchen and P. Mäder, "Plant species identification using computer vision techniques: A systematic literature review," *Archives of Computational Methods in Engineering*, Jan 2017. [Online]. Available: https://doi.org/10.1007/s11831-016-9206-z

[2] N. Kumar, P. N. Belhumeur, A. Biswas, D. W. Jacobs, W. J. Kress, I. C. Lopez, and J. a. V. B. Soares, "Leafsnap: A computer vision system for automatic plant species identification," *in Proceedings of the 12th European Conference on Computer Vision - Volume Part II*, ECCV'12. Berlin, Heidelberg: Springer-Verlag, 2012, pp. 502–516. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-33709-3-36

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

[4] M. P. Pound, J. A. Atkinson, D. M. Wells, T. P. Pridmore, and A. P. French, "Deep learning for multitask plant phenotyping," *bioRxiv*, 2017. [Online]. Available: https://www.biorxiv.org/content/early/2017/10/17/204552

[5] M. P. Pound, J. A. Atkinson, A. J. Townsend, M. H. Wilson, M. Griffiths, A. S. Jackson, A. Bulat, G. Tzimiropoulos, D. M. Wells, E. H. Murchie, T. P. Pridmore, and A. P. French, "Deep machine learning provides state-of-the-art performance in image-based plant phenotyping," *GigaScience*, vol. 6, no. 10, pp. 1–10, 2017. [Online]. Available: http://dx.doi.org/10.1093/gigascience/gix083

[6] P. Barr, B. C. Stver, K. F. Mller, and V. Steinhage, "Leafnet: A computer vision system for automatic plant species identification," *Ecological Informatics*, vol. 40, pp. 50 – 56, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1574954116302515

[7] P. N. Belhumeur, D. Chen, S. Feiner, D. W. Jacobs, W. J. Kress, H. Ling, I. Lopez, R. Ramamoorthi, S. Sheorey, S. White, and L. Zhang, "Searching the world's herbaria: A system for visual identification of plant species," *in Computer Vision – ECCV 2008*, D. Forsyth, P. Torr, and A. Zisserman, Eds. Springer Berlin Heidelberg, 2008, pp. 116–129.

[8] G. Cerutti, L. Tougne, J. Mille, A. Vacavant, and D. Coquin, "Understanding leaves in natural images - a model-based approach for tree species identification," *Comput. Vis. Image Underst.*, vol. 117, no. 10, pp. 1482–1501, Oct. 2013. [Online]. Available: http://dx.doi.org/10.1016/j.cviu.2013.07.003

[9] A. Joly, H. Goau, P. Bonnet, V. Baki, J. Barbe, S. Selmi, I. Yahiaoui, J. Carr, E. Mouysset, J.-F. Molino, N. Boujemaa, and D. Barthlmy, "Interactive plant identification based on social image data," *Ecological Informatics*, vol. 23, pp. 22 – 34, 2014, special Issue on Multimedia in Ecology and Environment.

[10] Y. Sun, Y. Liu, G. Wang, and H. Zhang, "Deep learning for plant identification in natural environment," Computational Intelligence and Neuroscience, vol. 2017, 2017. [Online]. Available: https://doi.org/10.1155/2017/7361042

[11] A. Affouard, H. Go¨eau, P. Bonnet, J.-C. Lombardo, and A. Joly, "Pl@ntNet app in the era of deep learning," *in ICLR 2017 - Workshop Track - 5th International Conference on Learning Representations*, Toulon, France, Apr. 2017, pp. 1–6. [Online]. Available: https://hal.archives-ouvertes.fr/hal-01629195

[12] J. R. Ubbens and I. Stavness, "Deep plant phenomics: A deep learning platform for complex plant phenotyping tasks," *Frontiers in Plant Science*, vol. 8, 2017. [Online]. Available: https://www.frontiersin.org/article/10.3389/fpls.2017.01190

[13] M. Lasseck, "Image-based plant species identification with deep convolutional neural networks," *in Working Notes of CLEF 2017 - Conference and Labs of the Evaluation Forum*, Dublin, Ireland, 2017. [Online]. Available: http://ceur-ws.org/Vol-1866/paper-174.pdf

[14] A. Joly, P. Bonnet, A. Affouard, J. Lombardo, and H. Go¨eau, "Pl@ntnet - my business," *in Proceedings of the 2017 ACM on Multimedia Conference*, MM 2017, Mountain View, CA, USA, October 23-27, 2017, 2017, pp. 551–555.

[15] A. Joly, H. Go¨eau, P. Bonnet, V. Bakic, J. Barbe, S. Selmi, I. Yahiaoui, J. Carr´e, E. Mouysset, J. Molino, N. Boujemaa, and D. Barth´el´emy, "Interactive plant identification based on social image data," *Ecological Informatics*, vol. 23, pp. 22–34, 2014.

[16] P. Bonnet, A. Joly, H. Go¨eau, J. Champ, C. Vignau, J. Molino, D. Barth´el´emy, and N. Boujemaa, "Plant identification: man vs. machine - lifeclef 2014 plant identification challenge," *Multimedia Tools Appl.*, vol. 75, no. 3, pp. 1647–1665, 2016.

[17] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. E. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," *CoRR*, vol. abs/1409.4842, 2014. [Online]. Available: http://arxiv.org/abs/1409.4842

[18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *in The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[19] J. Krause, G. Sugita, K. Baek, and L. Lim, "WTPlant (What's That Plant?): a deep learning system for identifying plants in natural images," *in Proceedings of the 2018 ACM on International Conference on Multimedia Retrieval (ICMR '18)*. ACM, New York, NY, USA, 517-520. DOI: https://doi.org/10.1145/3206025.3206089

[20] B. Zhou, H. Zhao, X. Puig, S. Fidler, A. Barriuso, and A. Torralba, "Semantic understanding of scenes through the ADE20K dataset," *CoRR*, vol. abs/1608.05442, 2016. [Online]. Available: http://arxiv.org/abs/1608.05442

[21] J. B. T. M. Roerdink and A. Meijster, "The watershed transform: Definitions, algorithms and parallelization strategies," 2001.

[22] C. Rother, V. Kolmogorov, and A. Blake, "Grabcut: Interactive foreground extraction using iterated graph cuts," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 309–314, Aug. 2004. [Online]. Available: http://doi.acm.org/10.1145/1015706.1015720

[23] P. Barr, B. C. Stver, K. F. Mller, and V. Steinhage, "Leafnet: A computer vision system for automatic plant species identification," *Ecological Informatics*, vol. 40, pp. 50 – 56, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1574954116302515

[24] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *in Advances in Neural Information Processing Systems 25*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2012, pp. 1097–1105.