# *WTPlant* (What's That Plant?): a Deep Learning System for Identifying Plants in Natural Images

Jonas Krause Gavin Sugita Kyungim Baek Lipyeow Lim
Department of Information and Computer Sciences, University of Hawai'i at Mānoa
1680 East-West Road, Honolulu, HI 96822, U.S.A.
(krausej|gsugita6|kyungim|lipyeow)@hawaii.edu

## ABSTRACT

Despite the availability of dozens of plant identification mobile applications, identifying plants from a natural image remains a challenging problem – most of the existing applications do not address the complexity of natural images, the large number of plant species, and the multi-scale nature of natural images. In this technical demonstration, we present the *WTPlant* system for identifying plants in natural images. *WTPlant* is based on deep learning approaches. Specifically, it uses stacked Convolutional Neural Networks for image segmentation, a novel preprocessing stage for multi-scale analyses, and deep convolutional networks to extract the most discriminative features. *WTPlant* employs different classification architectures for plants and flowers, thus enabling plant identification throughout all the seasons. The user interface also shows, in an interactive way, the most representative areas in the image that are used to predict each plant species. The first version of *WTPlant* is trained to classify 100 different plant species present in the campus of the University of Hawai'i at Mānoa. First experiments support the hypothesis that an initial segmentation process helps guide the extraction of representative samples and, consequently, enables Convolutional Neural Networks to better recognize objects of different scales in natural images. Future versions aim to extend the recognizable species to cover the land-based flora of the Hawaiian Islands.

## CCS CONCEPTS

• **Computing methodologies** → **Machine learning**; *Artificial intelligence*; Computer vision problems • **Object identification**

## KEYWORDS

Image Processing, Deep Learning, Convolutional Neural Network, Plant Taxonomy, Multi-Scale Analysis.

**ACM Reference format:**

Jonas Krause, Gavin Sugita, Kyungim Baek, Lipyeow Lim. 2018. *WTPlant* (What's That Plant?): a Deep Learning System for Identifying Plants in Natural Images. In *Proceedings of ACM ICMR conference, Yokohama, Japan, June 2018 (ICMR'18)*, 4 pages. DOI: 10.1145/3206025.3206089


ICMR'18, June 11-14, 2018, Yokohama, Japan.

ACM ISBN 978-1-4503-5046-4/18/06...$15.00.
https://doi.org/10.1145/3206025.3206089

## 1 INTRODUCTION

Plant identification has many important applications ranging from conservation to agriculture. Traditionally, botanists analyze different characteristics (visual and non-visual) of plants to be used as identification factors. However, identifying the species of plants accurately based on visual characteristics requires considerable expertise [1], which is almost impossible for the general public and challenging even for specialists. Previous efforts in using computer vision techniques for automated plant identification from controlled images have shown promising results [2,3,4,5,6]. However, a real-world plant identification application needs to be able 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 background, illumination, occlusions, shadows, and a rich local covariance structure that is usually present in the image. While human visual processing system navigates those factors with ease, an equivalent computational model for plant identification from natural images is still an open problem, the solution of which may have a substantial impact on botanical and agricultural fields.

Machine Learning (ML) approaches have shown promising results in various computer vision problems including the plant identification problem. Most previous efforts have used hand-designed features of leaves, flowers, and fruits [1,2,7,8,9] and are restricted to fairly controlled images with clean backgrounds. Identifying plant species relying on morphological characteristics extracted from well-controlled images is quite different from the noisy natural images that are found in real-world image classification applications.

More recently, Deep Learning (DL) methods have been introduced to this task [3,4,5,6,10,11,12,13,14] driven by the success of the Convolutional Neural Networks (CNNs). The deep convolutional approaches have been a growing trend in the computer vision field demonstrating impressive results in various tasks involving natural images. *WTPlant*, therefore, adopts the deep CNN approach and further extends it with the use of multiple stages or components of deep CNN in its architecture. In contrast to current plant identification methods that use hand-designed features, simple CNN architectures, and pre-trained models, *WTPlant* uses a carefully designed system of CNNs to classify plants and flowers in natural images.

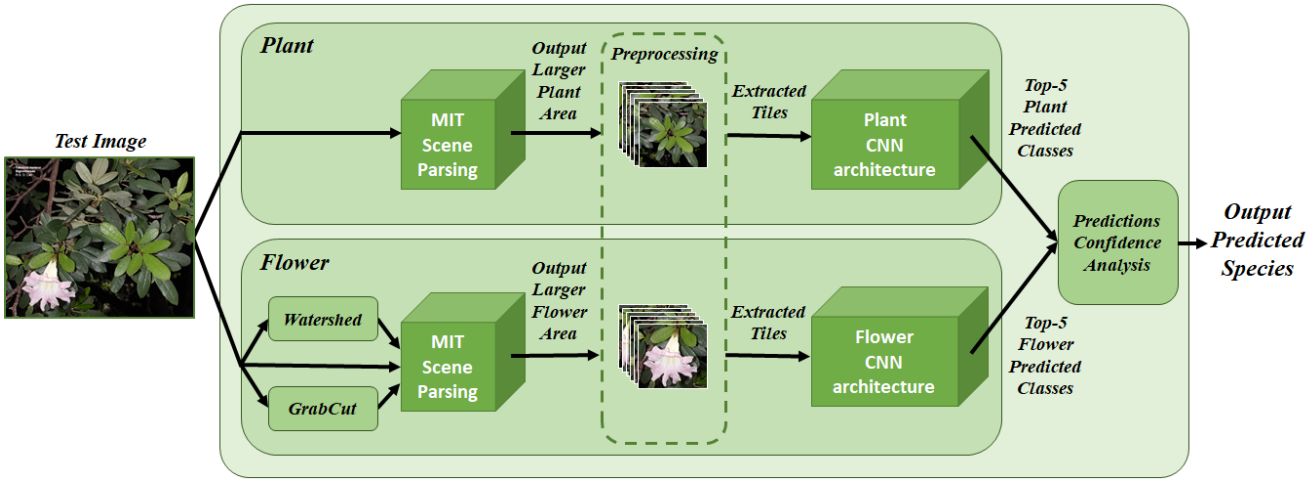


**Figure 1: Block diagram of the *WTPlant* system.**

The *WTPlant* system of CNNs consists of multiple pipelines and multiple stages of CNN components designed to extract deep discriminatory features. Each pipeline will segment the query image, preprocess the regions of interest into tiles of different sizes and classify the tiles using deep CNNs. The results from each pipeline are then combined to obtain more accurate predictions in a process reminiscent of ensemble techniques. The initial version of *WTPlant* presented here is trained to classify about 100 different plant species found on the campus of the University of Hawai'i (UH) at Mānoa. Preliminary experiments show that the initial segmentation process helps guide the extraction of representative samples and, consequently, enables Convolutional Neural Networks to better recognize objects of different scales in natural images.

In Section 2, we describe *WTPlant* in greater detail by explaining the system capabilities, the innovative preprocessing stage for the multi-scale problem, and the implemented CNN architectures. Section 3 describes how the system can be demonstrated. Section 4 reports initial results and Section 5 concludes the presentation of the system.

## 2 THE *WTPlant* SYSTEM

By training individual CNN architectures, one for plants and one for flowers, the *WTPlant* system is able to handle natural images with plants only, flowers only, or both together. In the latter case, the system will consolidate the predictions for plant and flower to output the final classification species. This collection of CNNs brings state-of-the-art DL architectures working together to provide the most accurate plant identification system. The accuracy will be measured by counting the correctly classified top-5 plants and flowers versus the incorrect ones during initial experiments. In the first version of *WTPlant* presented at this demo, the system is able to identify the 100 plant species present on the UH Mānoa campus. In future versions, this system will be extended to all plant species in the state of Hawai'i.

Figure 1 presents a block diagram of the *WTPlant* and details the workflow during the classification process of a query (unseen) image. The process begins by replicating the query image to the pipeline for plants and the pipeline for flowers. One of the key problems in computer vision is called scene parsing, i.e., recognizing and segmenting objects in a natural image. We used the MIT Scene Parsing method from Zhou et al. [15] which is a cascade segmentation module based on a CNN architecture with stacked convolutional blocks. MIT Scene Parsing is trained to segment 150 different objects from a scene, including plants and flowers. The *WTPlant* system uses the MIT Scene Parsing to segment and localizes regions potentially containing plants and flowers. For flowers, two pre-segmentation algorithms are used to assist the main segmentation process performed by the MIT Scene Parsing. After segmentation, the largest segmented areas of plant and flower are identified and passed to the preprocessing stage to create multi-position and multi-scale representative tiles. These tiles are individually classified by the CNN architectures that output the prediction results. In the final stage, the predicted classes for plant and flower are analyzed and combined according to the highest prediction values of each of the multi-position and multi-scale tiles. Based on these predicted high-confidence results, the top-5 final plant species are determined.

### 2.1 Capabilities

*2.1.1 Modularity.* The collection of CNNs is modularly designed to incorporate auxiliary methods that improve each stage of the system (segmentation, preprocessing, and classification). Moreover, each CNN can be easily upgraded (fine-tuned, retrained, or replaced) and incorporated back into the system. Therefore, new versions of the *WTPlant* system will be more accurate with the addition of new and more powerful CNNs. These new architectures may also be scaled up to a broader range of plant species, covering a wider ecosystem.

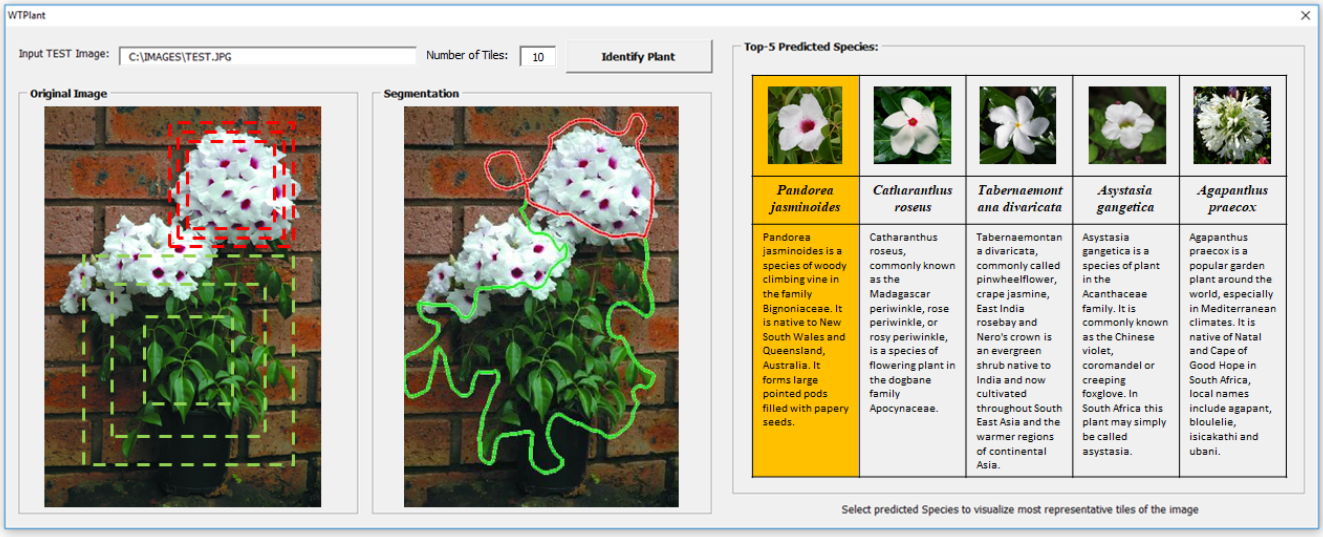


**Figure 2: Graphical User Interface (GUI) of the *WTPlant* system.**

*2.1.2 Pre-Segmentation.* Two pre-segmentation methods are incorporated into the *WTPlant* to assist the MIT Scene Parsing on the detection of flowers. They are needed due to the fact that some plant images have small flowers that are not initially captured by the MIT Scene Parsing. Hence, by separating background from the foreground using the Watershed Transform [16] and the GrabCut [17] algorithms the small flowers become more evident for the scene parsing.

*2.1.3 Preprocessing.* After the segmentation process performed by the MIT Scene Parsing, Regions of Interest (RoI) delimitating the larger plant and flower areas are collected from the inputted image. Searching for the most representative square areas within these RoI, a preprocessing step is implemented to extract multi-position and multi-scale representative tiles from the segmented plant and flower. The novelty of extracting tiles with multi-scale properties and using them to train the CNNs gives this system a better scale generalization capability when compared with common preprocessing methods such as resizing and random crop. Some of the reviewed approaches [6,10,13] suggest that simply downscaling the entire image is good practice. But CNN architectures generally receives small area images and a drastic downscale of a natural image will inevitably result in the loss of valuable information. Therefore, a preprocessing approach is needed to correctly handle the segmented RoI. The proposed method is aligned with previous approaches [3,4,5,14] that divided their analyzed regions into small tiles. However, instead of randomly crop these tiles, *WTPlant* performs a guided tile extraction which collects multi-location and multi-scale representative samples of the plant and flower to be analyzed.

*2.1.4 Classification Architectures.* The classification engines of this first version of the *WTPlant* system are two *AlexNets* [18], a DL architecture well-studied in the literature. They were both implemented using the *MatConvNet*[1] toolbox, one was trained exclusively for the analysis of plant tiles and the second one focuses on the analysis of flower tiles only. The separation between plants and flowers allows the networks to learn specific filters for each task in order to produce a good analysis of plant leaves and flower petals individually. Furthermore, by training both networks with multi-scale tile samples extracted based on the guided segmentation process, these classification architectures are able to learn and analyze plant images at diverse scales.

*2.1.5 Confidence Analysis.* The classification engines output the top-5 predictions for each one of the preprocessed tiles. The analysis of the plant tiles is done in parallel of the flower tiles, combining their confidence results to determine the plant species. This strategy covers the seasonality of the flowers and enables the system to work with plant species that are known as non-flowering plants, such as ferns, mosses, and liverworts.

## 2.2 Graphical User Interface (GUI)

*WTPlant* user interface was created using MATLAB R2015a and is shown in Figure 2. The MIT Scene Parsing is also implemented using MATLAB, as well as the preprocessing stage and the *MatConvNets*. In the GUI, the user inputs the test image and clicks the "Identify Plant" button. Then the test image will be loaded into the "Original Image" area and the segmentation stage produces the plant and flower RoI. These regions function as a guide to the preprocessing of representative multi-location and multi-scale tiles of the plant. These tiles are then inputted into each trained CNN. After the *WTPlant* analyzes the prediction confidence for each tile, the top-5 results are outputted into the "Top-5 Predicted Species" designated area with a brief description of the plant taxonomy.

*2.2.1 User Interaction.* The *WTPlant* GUI was designed to be a simple and user-friendly interface with highly interactive

[1] http://www.vlfeat.org/matconvnet/

features. For example, to deeply explore the multi-location and multi-scale properties in high-resolution images, the user can change the number of analyzed tiles (default 10) and expand the analysis to more multi-scale perspectives. In addition, after the top-5 predictions, the user can visualize which extracted tiles were responsible for the top-1 prediction, as demonstrated in Figure 2 where the *Pandora jasminoides* species is predicted.

## 3 DEMONSTRATION

For this demonstration conference, attendees will have the chance to take live pictures of a small plant and identify its species using the *WTPlant* system. Users will be able to test the system accuracy by taking photos with different angles, illumination, scales, and backgrounds. Numerous test images of other species will also be available during the demonstration. After the plant species analysis, the user will be able to compare the predictions with the ground-truth species and visualize which of the preprocessed tiles were used to determine the correct plant species.

## 4 RESULTS AND DISCUSSION

Initial experiments showed that the *WTPlant* system was able to detect the presence of **99.3%** of plants in ~17,000 natural images. Using representative tiles extracted from this initial segmentation process, two CNNs were trained and yielded **61.87%** top-1 and **85.25%** top-5 accuracies when plant and flower results are combined. However, species classification accuracy is constantly improving as new CNN architectures are trained and upgraded as classification engines. Furthermore, these experiments aimed to compare our proposed multi-location and multi-scale method with the common resize and random crop approaches during the training and testing of CNNs. As a result, *WTPlant* performed **14.6%** better on the top-1 accuracy when compared with CNNs trained using a naïve resize approach, and **9.1%** better when compared with the random crop approach. A total of 278 unseen natural images with complex backgrounds from 100 different plant species were used to produce these results. *WTPlant* took an average of 20 seconds to analyze each image, which may vary according to its size. These initial results support the hypothesis that a preprocessed stage guided by a segmentation process may help the training of CNNs.

## 5 CONCLUSION

In this paper, we present the first version of the *WTPlant*, describing its capabilities and functionalities. The modularity of this system allows the incorporation of different CNN architectures and will be upgraded as soon as new networks are trained. The preprocessing stage implemented to collect multi-scale tiles is also an important characteristic of this system, which creates proper scale data for more accurate analysis. The described GUI presents these features and enables the user to easily interact with the system visualizing the most representative analyzed areas. Future work aims to expand current CNN architectures and explore new classification pipelines for fruits and seedlings in natural images.

## ACKNOWLEDGMENTS

This research was supported in part by the Brazilian National Council for Scientific and Technological Development (CNPq) under grant number 219438-14.5.

## REFERENCES

[1] Jana Wäldchen and Patrick Mäder. Plant Species Identification Using Computer Vision Techniques: A Systematic Literature Review. *Archives of Computational Methods in Engineering*, pages 1-37, 2017.

[2] Neeraj Kumar, Peter N. Belhumeur, Arijit Biswas, David W. Jacobs, W. John Kress, Ida C. Lopez, and João V. B. Soares. LeafSnap: A Computer Vision System for Automatic Plant Species Identification. In *Proceedings of the 12th European Conference on Computer Vision* (ECCV), pages 502-516, Florence, Italy, 2012.

[3] Sue H. Lee, Chee S. Chan, Paul Wilkin, and Paolo Remagnino. Deep-Plant: Plant identification with convolutional neural networks. *ArXiv Preprint*, 2015. arxiv.org/abs/1506.08425.

[4] Michael P. Pound, Jonathan A. Atkinson, Darren M. Wells, Tony P. Pridmore, and Andrew P. French. Deep Learning for Multi-task Plant Phenotyping. *bioRxiv Preprint*, 2017. doi.org/10.1101/204552.

[5] Michael P. Pound, Jonathan A. Atkinson, Alexandra J. Townsend, Michael H. Wilson, Marcus Griffiths, Aaron S. Jackson, Adrian Bulat, Georgios Tzimiropoulos, Darren M. Wells, Erik H. Murchie, Tony P. Pridmore, and Andrew P. French. Deep machine learning provides state-of-the-art performance in image-based plant phenotyping. *GigaScience*, 6(10):1-10, 2017.

[6] Pierre Barré, Ben C. Stöver, Kai F. Müller, and Volker Steinhage. LeafNet: A computer vision system for automatic plant species identification. *Ecological Informatics*, 40:50-56, 2017.

[7] Peter N. Belhumeur, Daozheng Chen, Steven Feiner, David W. Jacobs, W. John Kress, Haibin Ling, Ida Lopez, Ravi Ramamoorthi, Sameer Sheorey, Sean White, and Ling Zhang. Searching the world's herbaria: A system for visual identification of plant species. In *Proceedings of the 10th European Conference on Computer Vision* (ECCV), pages 116-129, 2008.

[8] Guillaume Cerutti, Laure Tougne, Julien Millea, Antoine Vacavant, and Didier Coquin. Understanding leaves in natural images – A model-based approach for tree species identification. In *Computer Vision and Image Understanding*, 117(10):1482-1501, 2013.

[9] Alexis Joly, Hervé Goëau, Pierre Bonnet, Vera Backić, Julien Barbe, Souheil Selmi, Itheri Yahiaoui, Jennifer Carré, Elise Mouysset, Jean-François Molino, Nozha Boujemaa, and Daniel Barthélémy. Interactive plant identification based on social image data. *Ecological Informatics*, 23:22-34, 2014.

[10] Yu Sun, Yuan Liu, Guan Wang, and Haiyan Zhang. Deep Learning for Plant Identification in Natural Environment. *Computational Intelligence and Neuroscience*, 2017. doi.org/10.1155/2017/7361042.

[11] Sarah T. Namin, Mohammad Esmaeilzadeh, and Mohammad Najafi, Tim B. Brown, and Justin O. Borevitz. Deep Phenotyping: Deep Learning For Temporal Phenotype/Genotype Classification. *bioRxiv Preprint*, 2017. doi.org/10.1101/134205

[12] Antoine Affouard, Herve Goeau, Pierre Bonnet, Jean-Christophe Lombardo, and Alexis Joly. Pl@ntnet app in the era of deep learning. In *Proceedings of the 5th International Conference on Learning Representations* (ICLR), pages 1-6, Toulon, France, 2017.

[13] Jordan R. Ubbens and Ian Stavness. Deep Plant Phenomics: A Deep Learning Platform for Complex Plant Phenotyping Tasks. *Frontiers in Plant Science*, 2017. doi.org/10.3389/fpls.2017.01190.

[14] Mario Lasseck. Image-based plant species identification with deep convolutional neural networks. In *Working Notes of Cross Language Evaluation Forum* (CLEF), 2017.

[15] Bolei Zhou, Hang Zhao, Xavier Puig, Sanja Fidler, Adela Barriuso, and Antonio Torralba. Scene Parsing through ADE20K Dataset. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 633-641, Honolulu, Hawaii, 2017.

[16] Jos B.T.M. Roerdink and Arnold Meijster. The Watershed Transform: Definitions, Algorithms and Parallelization Strategies. *Fundamenta Informaticae*, 41(1-2):187-228, 2000.

[17] Carsten Rother, Vladimir Kolmogorov, and Andrew Blake. "GrabCut": interactive foreground extraction using iterated graph cuts. *ACM Transactions on Graphics*, 23(3):309-314, 2004.

[18] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems (NIPS) 25*, pages 1097-1105, Lake Tahoe, Nevada, 2012.