Fine-Grained Plant Classification Using Convolutional Neural Networks for Feature Extraction

Niko Sünderhauf, Chris McCool, Ben Upcroft, and Tristan Perez

Agricultural Robotics Program, Queensland University of Technology
2 George Street, Brisbane QLD 4001, Australia
http://www.tiny.cc/agrc-qut
Corresponding author: niko.suenderhauf@qut.edu.au

Abstract. We present an overview of the QUT plant classification system submitted to LifeCLEF 2014. This system uses generic features extracted from a convolutional neural network previously used to perform general object classification. We examine the effectiveness of these features to perform plant classification when used in combination with an extremely randomised forest. Using this system, with minimal tuning, we obtained relatively good results with a score of 0.249 on the test set of LifeCLEF 2014.

Keywords: convolutional neural network, extremely random forest, plant classification

1 Introduction

Future food security presents a serious challenge. To sustain the projected world population of over 9 billion in 2050, the current worldwide production of food will have to almost double. This is a significant challenge given that the land allocation dedicated to agriculture has already peaked in most countries. In addition, agriculture requires a significant amount of energy and water, and it is a large contributor to green-house-gas emissions. To overcome these problems requires alternative approaches such as the use of new crop varieties with increased yield and robustness to climate change, the adoption of policies leading to sustainable practices, and the development of new technologies to increase the efficiency of farms.

Robot technology in farms will play an important role in the near future in several applications. For both infield and crop management, robots could contribute to the efficient use of energy, herbicides, pesticides, water, and fertiliser by measuring plant growth and detecting weeds/pests. In harvesting of horticultural produce, robots could contribute to increased yield and quality by automating the picking and grading processes. To do this requires sophisticated algorithms that could identify plants in their various states of growth.

The Plant Identification Task [5] of the LifeCLEF challenge [7] is an established benchmark for fine-grained plant classification. The task asks the participating teams to correctly identify the images taken of 500 different herbs, trees, and fern species from France. The provided training dataset consists of over 47,000 individual images that are organized into *observations*. Each of these observations can contain multiple images from the same plant, but from different parts of it. These different parts are referred to as content categories and include: Branch, Stem, Leaf, LeafScan, Fruit, Flower and Entire.

2 Our Approach

We investigate the potential for generic features obtained from a well trained convolutional neural network (CNN) to perform the task of plant classification. This is a fine-grained image classification problem which has previously been addressed by deriving hand-crafted features. By contrast, our proposed system uses features from a CNN that was initially trained for general object classification [11] using millions of images from the ImageNet dataset. To classify these generic pre-trained features we make use of an extremely randomised forest. We briefly describe our features and classifier below.

2.1 Convolutional Neural Networks as Generic Feature Detectors

Convolutional neural networks (CNNs) were proposed in 1989 by LeCun et al. [9] to recognize hand-written digits. CNNs learn a sparser connection between regions of an image and the NNs by imposing spatial dependencies; this can reduce complexity. However, the broad applicability of CNNs has only recently shown promise most likely due to the availability of large datasets, growth in computational power, availability of GPUs and efficient algorithms such as rectified linear units [1, 6] which have been used to train these large networks.

Several research groups have explored the potential of these large CNNs to outperform more classical approaches to object recognition or detection that are based on hand-crafted features [2, 4, 8, 10, 11]. The CNN systems applied in these large-scale object recognition or detection tasks consist of a feature extractor (the actual CNN) followed by a classifier or regressor. Razavian et al. [10] showed that combining CNN features with a simple classifier such as a linear SVM is highly competitive or even superior to classical approaches for a variety of recognition and detection tasks. We therefore follow the approach of Razavian et al. [10] and apply a pre-trained CNN as a generic feature extractor to the images of the LifeCLEF Plant Task.

In our approach, we use the pre-trained Overfeat [11] features which is a CNN system. For feature extraction, we examine the effectiveness of using either the first or second fully connected layers – referred to as Layer 17 and Layer 19 respectively. To use the Overfeat features, we downsample the images so that the smallest dimension is 231 pixels wide or high respectively, this is

because the Overfeat network operates on a region of size 231×231. The extracted features are vectors of length $3,072$ or $4,096$ for Layer 17 and Layer 19 respectively. Overfeat will extract several feature vectors per image in a sliding window fashion unless the images are square (231×231). These feature vectors are then fed into an ensemble of extremely randomized trees that performs the actual classification.

2.2 Extremely Randomized Trees Classifier

An extremely randomized tree classifier (*extratree*) is a tree-based ensemble method for supervised classification. It is conceptually similar to random forest classifiers, but takes the idea of randomness one step further. In contrast to decision trees that are designed offline and derived from expert knowledge, an extremely randomized tree classifier learns the layout of its ensemble of trees from training data. The classifier output is a probability distribution over all classes, in our case this would over all 500 possible species in LifeCLEF. We refer the reader to the literature [3] for a further discussion of extremely randomized trees.

We train a separate classifier for each of the content categories in the LifeCLEF dataset. Thus for each category, Branch, Leaf or LeafScan, we have a separate extremely randomized tree. To handle multiple samples from each image the classification results of the *extratrees* are combined to provide just one prediction per observation. This is achieved by averaging the output (probability distribution) for all features from an image and results in one probability distribution for each image.

PlantCLEF introduces a further difficulty in that an observation can have multiple images or even multiple content categories available. To combine the information from multiple images, or even multiple content categories, we treat the result of each image as a likelihood score and sum them into a single distribution of prediction scores over the 500 different classes for each observation.

3 Experimental Setup

To derive the hyper parameters for our system, we performed cross validation. The training dataset was split into two tests: the system is trained using 95% of the training observations and then evaluated on the remaining 5%. We note that our choice for forming the cross-validation dataset is overly simplified as we uniformly sample across observations and do not uniformly sample over classes (species) and observations; this means that we do not enforce an equal ratio between classes and observations of the training and the evaluation sets. Finally, we only used observations with a user-provided quality score of 3 or above in the training and validation stage.

The parameters of the ensemble classifier are tuned by sampling the parameter space. The resultant values are listed in Table 1. Our system is implemented in Python and uses the *extratrees* implementation of scikit-learn[1].

[1] `http://scikit-learn.org/`

Table 1: A table outlining the final system parameters used for the *extratrees* that were used.

Parameter	Value
size of ensemble	65 trees
used feature elements	64
tree depth	15
information gain measure	entropy

4 Results

A total of ten teams participated in the challenge and we ranked 4th in the team ranking. In Figure 1, we present the results using the LifeCLEF score metric where it can be seen that our approach scored 0.249; to facilitate interpretation of this plot we associated the runs of the top 4 teams with a different color. This is slightly worse than BME TMIT systems and considerably better than the FINKI systems.

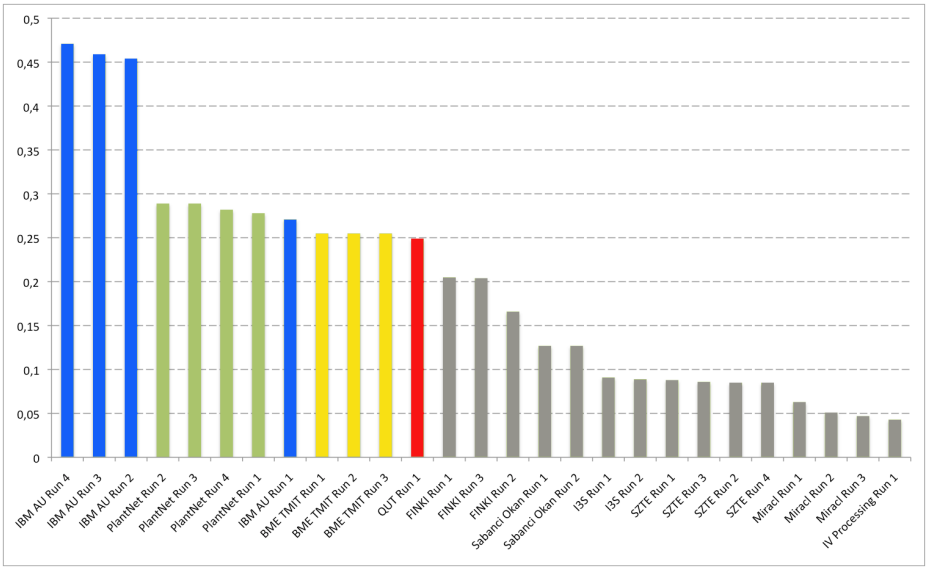

Fig. 1. The results of the LifeCLEF Plant Task 2014. Image adapted from the organizers' website. Our run (QUT) is marked in red, the runs committed by the top 3 teams are color coded in blue (IBM), green (PlantNet) and yellow (BME TMIT) respectively.

To understand better the performance of our system, we examined the recognition accuracy for each content category. In Figure 2, we present the recognition

accuracy for our system for each content category and also for using different layers of the neural network (Layers 17 and 19). We can see that our system is most accurate for the LeafScan category and achieves a rank-1 identification rate of more than 50%. The second best category is Flower with a rank-1 identification rate of almost 25%. After this, our rank-1 identification rate for the categories falls below 20% and performs worst on the Branch category. We believe that the considerable performance difference on the LeafScan category is due to the fact that these images contain a single leaf image which is usually well centered and has a homogeneous background; this greatly simplifies the feature extraction stage and consequently classification. By contrast, other categories such as Fruit, Branch and Leaf, have highly varying backgrounds and the object of interest is rarely well centered or at a consistent scale. Finally, we note that despite the considerable performance difference between LeafScan and the other categories, our overall system performance is acceptable. We believe this is because most of the images in the dataset appear to be LeafScan images.

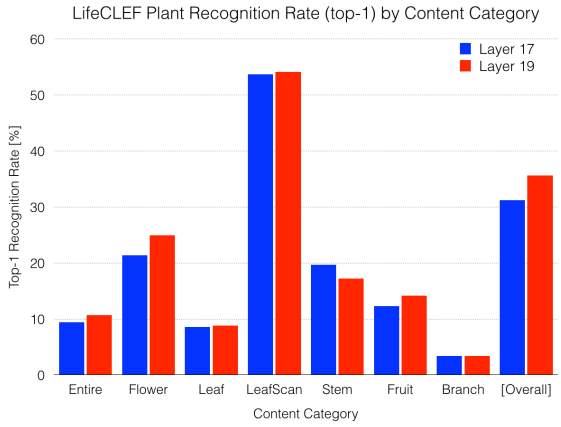

Fig. 2. Recognition accuracies (rank-1 identification rate) for the different content categories present in the dataset. The plot compares the performances of features extracted from the first and second fully connected layer presented in blue and red respectively.

Comparing the performance of the two different extracted feature sets, we can see in Figure 2 that Layer 19 provides the best overall score. Layer 17 is the first fully connected layer and Layer 19 is the second fully connected layer. Examining the performance on a per category basis, we can see that Layer 17 and 19 have similar performance except for Flower and Fruit where Layer 19 outperforms Layer 17. By contrast, Layer 17 provides slightly better performance for the Stem category. We want to point out that Razavian et al. [10] used the first fully connected layer (Layer 17) in their work.

5 Conclusions and Future Work

The initial results of our plant classification system using CNNs provides competitive performance. We have found that using Layer 19, rather than Layer 17, provides superior performance for the task of plant classification. Nevertheless, some issues are worth considering further.

One major issue is that we need to consider how to perform localisation of the object of interest within the image. We believe this is an important factor that degrades performance. One possible way to approach this problem is to search the image for the most salient parts in the image. Also, we should examine the properties of the CNN features, and how they could be re-trained to deal with the specific task of plant classification.

Acknowledgements

This work has been supported by the Department of Agriculture Fisheries and Forestry (DAFF) of the Queensland government through the Agricultural Robotics Program at QUT. We would also like to thank David Hall and Dr. Feras Dayoub for the fruitful conversations.

References

1. George E Dahl, Tara N Sainath, and Geoffrey E Hinton. Improving deep neural networks for lvcsr using rectified linear units and dropout. In *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, pages 8609–8613. IEEE, 2013.
2. Jeff Donahue, Yangqing Jia, Oriol Vinyals, Judy Hoffman, Ning Zhang, Eric Tzeng, and Trevor Darrell. Decaf: A deep convolutional activation feature for generic visual recognition. *arXiv preprint arXiv:1310.1531*, 2013.
3. Pierre Geurts, Damien Ernst, and Louis Wehenkel. Extremely randomized trees. *Machine learning*, 63(1):3–42, 2006.
4. Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. *arXiv preprint arXiv:1311.2524*, 2013.
5. Hervé Goëau, Alexis Joly, Pierre Bonnet, Jean-François Molino, Daniel Barthélémy, and Nozha Boujemaa. Lifeclef plant identification task 2014. In *CLEF working notes 2014*, 2014.
6. Geoffrey E Hinton, Nitish Srivastava, Alex Krizhevsky, Ilya Sutskever, and Ruslan R Salakhutdinov. Improving neural networks by preventing co-adaptation of feature detectors. *arXiv preprint arXiv:1207.0580*, 2012.
7. Alexis Joly, Henning Müller, Hervé Goëau, Hervé Glotin, Concetto Spampinato, Andreas Rauber, Pierre Bonnet, Willem-Pier Vellinga, and Bob Fisher. Lifeclef 2014: multimedia life species identification challenges. In *Proceedings of CLEF 2014*, 2014.
8. Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C.J.C. Burges, L. Bottou, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc., 2012.

9. Yann LeCun, Bernhard Boser, John S Denker, Donnie Henderson, Richard E Howard, Wayne Hubbard, and Lawrence D Jackel. Backpropagation applied to handwritten zip code recognition. *Neural computation*, 1(4):541–551, 1989.
10. Ali Sharif Razavian, Hossein Azizpour, Josephine Sullivan, and Stefan Carlsson. Cnn features off-the-shelf: an astounding baseline for recognition. *arXiv preprint arXiv:1403.6382*, 2014.
11. Pierre Sermanet, David Eigen, Xiang Zhang, Michael Mathieu, Rob Fergus, and Yann LeCun. Overfeat: Integrated recognition, localization and detection using convolutional networks. *arXiv preprint arXiv:1312.6229*, 2013.