

PlantClef 2019 - Plant Identification on Amazonian flora

Lukáš Pícek¹

1 Introduction

Automated identification of plants has improved significantly in the last few years thanks to recent research contributions in Deep Learning. In the scope of LifeCLEF², we already achieved an impressive performance with automatic system for identification of 10K plant species. However, these 10K species, mostly growing in Europe and North America, only represent the tip of the iceberg. The vast majority of the species in the world (~369K species) actually lives in data deficient countries and the performance of state-of-the-art machine learning algorithms on these species is unknown and presumably much lower. Thus, the main focus of the 2019 edition of PlantCLEF was to evaluate automated identification on the flora of such data deficient regions.

2 Challenge Description

The goal of the task was to predict the most likely species for each observation of the test set (an observation is a set of images of the same individual plant). A small part of the observations in the test set was re-annotated by several experts in order to allow comparing the performance of the evaluated systems with the one of highly skilled experts.

3 Data Collection

Provided dataset was consist out of 10K species (~300k images) mainly focused on the Guiana shield and the Amazon rainforest (known to be the largest collection of living plants and animal species in the world). The average number of images per specie in that new dataset was much lower than the dataset used in the previous editions of PlantCLEF (about 10 vs. 100). Many species contained only a few images and some of them even contained only one. At the same time, approximately 80% of the data were “labeled” as noisy (10-20% wrongly annotated) and that makes image recognition even harder.

4 Proposed System Description

Our best submission was achieved by an ensemble of 4 CNNs trained on preselected PlantCLEF data extended with additional data from GBIF³. All models (2x Inception-V4, 2x

¹ PhD. student, University of West Bohemia, Faculty of Applied Sciences, Department of Cybernetics, Computer Vision specialization.

² LifeCLEF lab is part of the Conference and Labs of the Evaluation Forum: CLEF 2019.

³ GBIF is an international network and research infrastructure funded by the world’s governments and aimed at providing anyone, anywhere, open access to data about all types of life on Earth.

Inception-ResNet-v2; trained with different hyper-parameters; first two epochs were trained only FC layers) were initialized from ours PlantCLEF2018 checkpoints that we shared to all competitors with help of organizers to make the “starting point” same for all the participants.

In post-processing we included test-time image augmentation (crops and mirroring) and automatic test set class-prior estimation from the CNN outputs

5 Results

Our system for plant identification based on image recognition scored 1st in this year LifeCLEF classification challenge. We achieved 41% accuracy in recognition of 10,000 extremely rare Amazonian flora species. While testing against humans, our system scored better than 3 out of 5 experts. At the other hand two best experts outshone our system with accuracy over 60%.

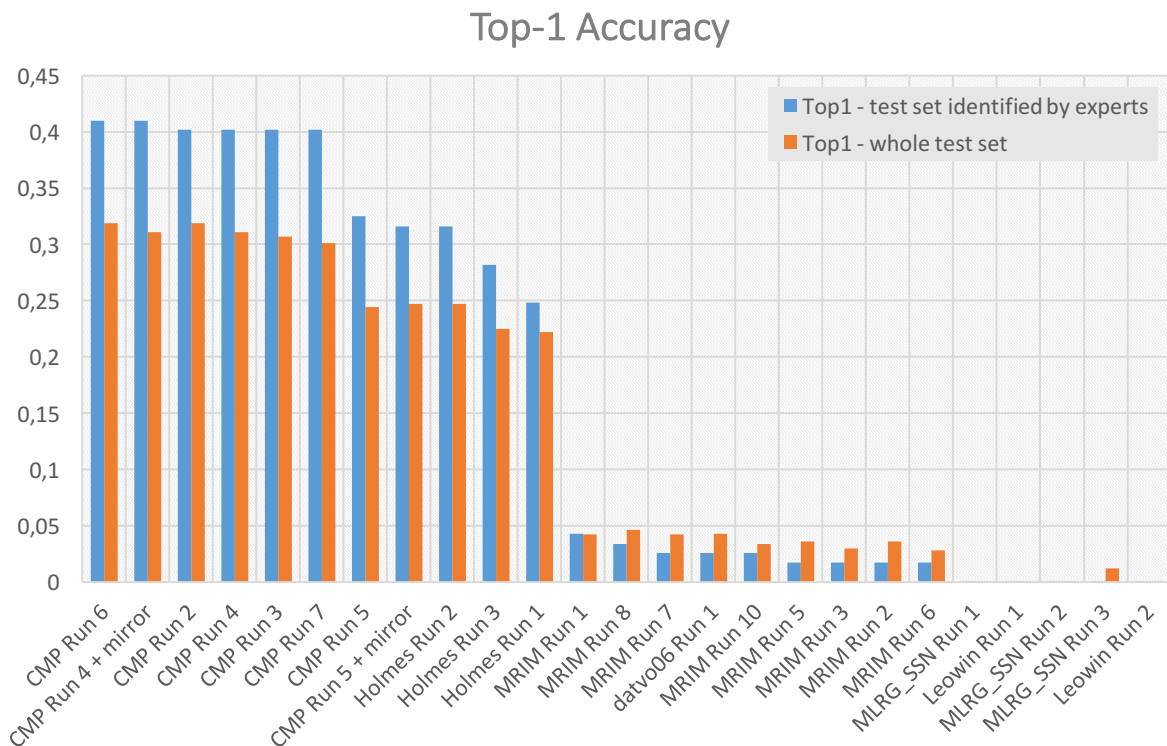


Figure 1: The following figure reports the comparison of the submissions made by 6 participants for the PlatClef-2019 challenge.

Acknowledgment

Access to computing and storage facilities owned by parties and projects contributing to the National Grid Infrastructure MetaCentrum provided under the programme "Projects of Large Research, Development, and Innovations Infrastructures" (CESNET LM2015042), is greatly appreciated.

References

- Šulc M., Pícek L., Matas J. (2018) Plant Recognition by Inception Networks with Test-time Class Prior Estimation, CLEF 2018
- Szegedy, Ch., Ioffe, S., Vanhoucke, V., Alemi, A. (2016) Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning, eprint arXiv:1602.07261.