Recognition of the Amazonian flora by Inception Networks with Test-time Class Prior Estimation

CMP submission to PlantCLEF 2019

Lukáš Picek[1], Milan Šulc[2], and Jiří Matas[2]

[1] Dept. of Cybernetics, Faculty of Applied Sciences, University of West Bohemia
picekl@kky.zcu.cz
[2] Visual Recognition Group, Faculty of Electrical Engineering, Czech Technical University in Prague
{sulcmila,matas}@cmp.felk.cvut.cz

Abstract. The paper describes an automatic system for recognition of 10,000 plant species, with focus on species from the Guiana shield and the Amazon rain forest. The proposed system achieves the best results on the PlantCLEF 2019 test set with 31.9% accuracy. Compared against human experts in plant recognition, the system performed better than 3 of the 5 participating human experts and achieved 41.0% accuracy on the subset for expert evaluation. The proposed system is based on the Inception-v4 and Inception-ResNet-v2 Convolutional Neural Network (CNN) architectures. Performance improvements were achieved by: adjusting the CNN predictions according to the estimated change of the class prior probabilities, replacing network parameters with their running averages, test-time data augmentation, filtering the provided training set and adding additional training images from GBIF.

Keywords: Plant Recognition, Computer Vision, Convolutional Neural Networks, Machine Learning, Class Prior Estimation, Fine-grained, Classification

1 Introduction

The paper describes an automatic system for visual recognition of plants among 10,000 species, developed for the the PlantCLEF 2019 plant identification challenge [4] organized in connection with the LifeCLEF 2019 workshop [5] at the Conference and Labs of the Evaluation Forum. Compared to previous PlantCLEF challenges [1,2,3], which contained mainly species living in Europe and North America, the 2019 task is focused on the recognition of species from "data deficient regions" - mainly the Guiana shield and the Amazon rain forest.

The proposed approach is based on CMP's winning submission to PlantCLEF 2018 [11]. Checkpoints of our models from PlantCLEF 2018 have been shared with other participants of PlantCLEF 2019 in order to provide a good starting point to all participants.

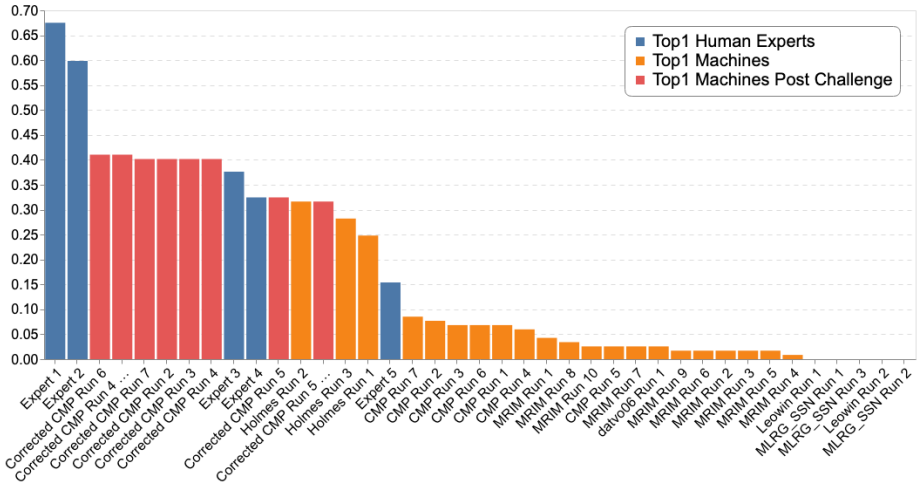

Fig. 1. Comparison of automatic plant recognition methods against human experts. The results of our method are shown in red as "Post Challenge" (our results submitted at the challenge deadline, shown in orange, were wrongly exported).

2 Methodology

2.1 Cleaning and extending the training dataset

The PlantCLEF 2019 training set covers 10,000 species and consists of:

- PlantCLEF 2019 EOL: 72,260 images covering 4,197 classes from the Encyclopedia of Life[3]
- PlantCLEF 2019 Google: 68,254 images covering 6,262 classes automatically retrieved by web search engines.
- PlantCLEF 2019 Bing: 307,557 images covering 8,666 classes automatically retrieved by web search engines.

The average number of images per specie decreased dramatically from PlantCLEF 2018. One fifth of species contains less then 10 images and some of them contains only 1 image.

[3] `http://www.eol.org`

Fig. 2. Randomly selected images from the LifeCLEF 2019 training set (top) and test set(bottom).

A brief manual inspection showed that the provided training set is afflicted with noisy samples - wrongly labeled images, including images of non-flora objects. Examples of noisy samples are in Figure 3. We therefore decided to detect non-flora images by a pre-trained Darknet53 448x448 [8] classifier. Out of 428,702 images from the official training set, we removed 6,181 images detected as non-flora, After that our training data missed approximately 2000 classes, so we had to gather additional training images to fill that gap. We created a new training set[4] including external training data downloaded from GBIF[5], described in Table 1. Changes in the dataset statistics are visualized in Figure 4.

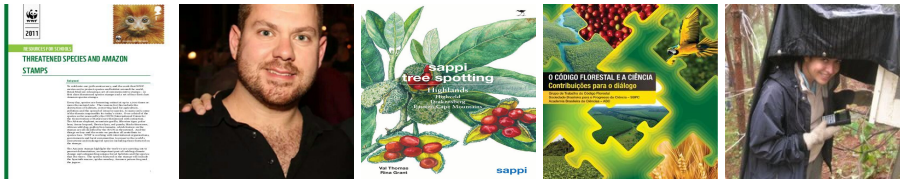

Fig. 3. Randomly selected noisy images from the LifeCLEF 2019 training set.

To make sure that none of the additional training images (or its resized or cropped versions) downloaded from GBIF appear in the test set, we used the image retrieval pipeline of Radenovic et al. [7] with VGG-16 and whitening. The

[4] For full reproducibility, a list of removed samples as well as an archive with additional training images are shared at `http://cmp.felk.cvut.cz/~sulcmila/LifeCLEF2019/`

[5] `http://www.gbif.org/`

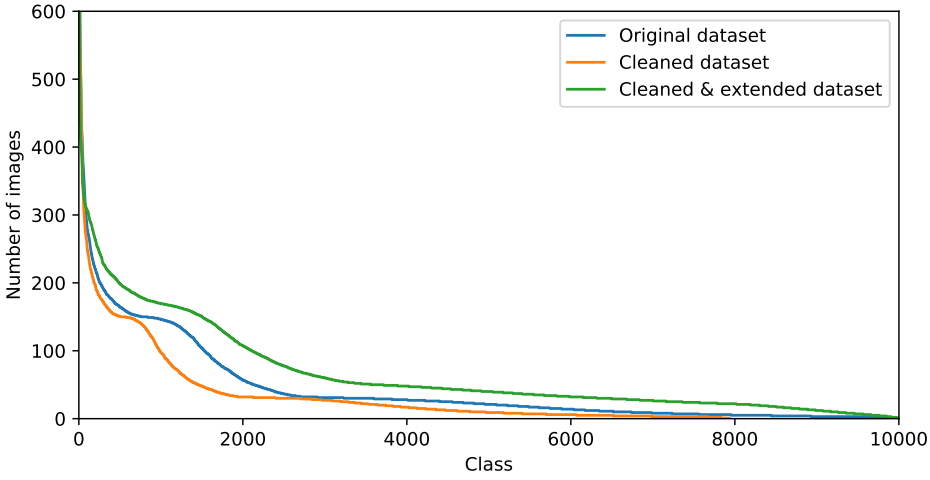

Fig. 4. Numbers of training images per class in the original dataset (blue), cleaned dataset (orange) and cleaned and extended (green), sorted for each dataset separately.

Table 1. Training data (after cleaning and extending the provided training set) used in the experiments.

Data Source	Classes	Non EOL classes	Number of Images
EOL	4197	0	58548
Noisy Google	6262	3800	64863
Noisy Bing	8666	5069	305291
GBIF (additional)	9402	5734	238009
All	9998	5801	666711

nearest neighbours of test images among the downloaded images are vizualized in Figure 5.

2.2 Convolutional Neural Networks

The proposed system is based on two CNN architectures – Inception ResNet v2 and Inception v4 [12]. The TensorFlow-Slim API was used to adjust and fine-tune the networks from the publicly available[6] PlantCLEF 2018 winning checkpoints.

All networks in our experiments shared the optimizer settings enumerated in Table 2. The networks and their input resolutions are listed in Table 3.

The following image pre-processing techniques were used for training:

- Random image crop with aspect ratio range $(0.75, 1.33)$ and content at least 80% of origin image.

[6] `http://cmp.felk.cvut.cz/~sulcmila/LifeCLEF2018/`

Fig. 5. Six nearest couples of test set images (top) and GBIF images (bottom).

Table 2. Optimizer hyper-parameters, common to all networks in the experiments.

Parameter	Value
Batch size	32
Optimizer	rmsprop
RMSProp momentum	0.9
RMSProp decay	0.9
Initial learning rate	0.0075
Learning rate decay type	Exponential (stairs)
Learning rate decay factor	0.975

- Random left-right flip.
- Brightness and saturation distortion.

Table 3. Networks and hyper-parameters used in the experiments.

#	Net architecture	Input Resolution
1	Inception v4	299×299
2	Inception v4 (second)	299×299
3	Inception v4	598×598
4	Inception ResNet v2	299×299
5	Inception ResNet v2 (second)	299×299

2.3 Test time data augmentation

At test-time, 3 predictions per image are generated by using 3 crops:

- 1x Full image,
- 1x Central crop covering 80% of the original image,

- 1x Central crop covering 60% of the original image.

In submissions 4,5,6,7 the mirrored versions of all three crops were also evaluated.

2.4 Adjusting Class Priors at Test Time

The training set data distribution is highly unbalanced and we can not guarantee that the test images were drawn from the same distribution: as described in Section 2.1, the training set comes from different sources, where the class frequencies may not correspond with the test-time priors.

Following the notation of [10], the predictions $p(c_k|\mathbf{x}_i)$ of a network trained on a dataset with class prior probabilities $p(c_k)$ should be corrected in case of evaluation on a test set with different class priors $p_e(c_k)$:

$$p_e(c_k|\mathbf{x}_i) = \frac{p(c_k|\mathbf{x}_i)\frac{p_e(c_k)}{p(c_k)}}{\sum_{j=1}^{K} p(c_j|\mathbf{x}_i)\frac{p_e(c_j)}{p(c_j)}} \propto p(c_k|\mathbf{x}_i)\frac{p_e(c_k)}{p(c_k)} \tag{1}$$

Since the test-time priors $p_e(c_j)$ are unknown, we propose three different estimates of adjusting the predictions:

UNIFORM: As the simplest option, we adjust the test predictions by assuming a uniform prior for all classes.

MLE: As the second option, we compute a Maximum Likelihood Estimate of the test time prior $p_e(c_k)$ using the EM algorithm of Saerens et al. [9], comprising of the following two steps:

$$\textbf{E:}\quad p_e^{(s)}(c_k|\mathbf{x}_i) = \frac{p(c_k|\mathbf{x}_i)\frac{p_e^{(s)}(c_k)}{p(c_k)}}{\sum_{j=1}^{K} p(c_j|\mathbf{x}_i)\frac{p_e^{(s)}(c_j)}{p(c_j)}} \tag{2}$$

$$\textbf{M:}\quad p_e^{(s+1)}(c_k) = \frac{1}{N}\sum_{i=1}^{N} p_e^{(s)}(c_k|\mathbf{x}_i) \tag{3}$$

MAP: As the third option, we use the Maximum a Posteriori estimate proposed in [10]:

$$
\begin{aligned}
\mathbf{P}^{\text{MAP}} &= \arg\max_{\mathbf{P}} p(\mathbf{P}|(\mathbf{x}_1, .., \mathbf{x}_N)) \\
&= \arg\max_{\mathbf{P}} p(\mathbf{P}) \prod_{i=1}^{N} p(\mathbf{x}_i|\mathbf{P}) \\
&= \arg\max_{\mathbf{P}} \log p(\mathbf{P}) + \sum_{i=1}^{N} \log p(\mathbf{x}_i|\mathbf{P}) \\
&\text{s.t.} \sum_{k=1}^{K} P_k = 1;\ \forall k : P_k \geq 0
\end{aligned}
\tag{4}
$$

We model the prior knowledge about the categorical distribution $p_e(c_k)$ by the symmetric Dirichlet distribution:

$$
p(\mathbf{P}) = \frac{1}{B(\alpha)} \prod_{k=1}^{K} P_k^{\alpha-1} \tag{5}
$$

where the normalization factor for the symmetric case is $B(\alpha) = \frac{\Gamma(\alpha)^K}{\Gamma(\alpha K)}$. As in [10], we use $\alpha = 3$.

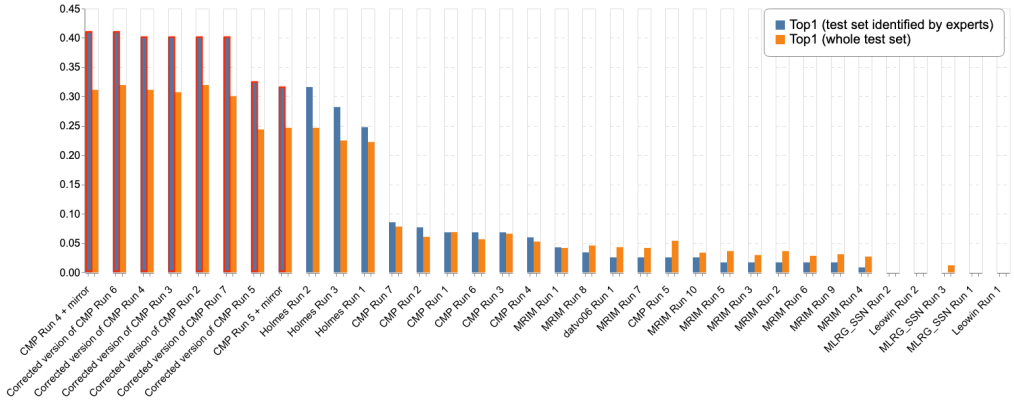

Fig. 6. Comparison of automatic plant recognition methods on the PlantCLEF 2019 test set. "Post Challenge" submissions are marked with red border.

3 Results

Table 4 describes eight final runs used for the evaluation. An ensemble of all five networks from Section 2.2 was used in all runs and predictions were averaged over all networks and all test image augmentations from Section 2.3.

Table 4. Description of our (corrected/post-challenge) submissions.

Run description			Test accuracy			
Name	Test-time augm.	Prior est.	Top1	Top1 Exp.	Top5 All	Top5 Exp.
CMP Run 2	3×scale	(none)	0,244	0,325	0,356	0,410
CMP Run 3	3×scale	uniform	0,247	0,316	0,360	0,419
CMP Run 4	3×scale	MAP	0,301	0,402	0,453	0,573
CMP Run 5	3×scale	MLE	0,307	0,402	0,451	0,573
CMP Run 6	3×scale + mirrors	(none)	0,311	0,402	0,454	0,538
CMP Run 7	3×scale + mirrors	uniform	0,311	0,410	0,461	0,564
CMP Run 4*	3×scale + mirrors	MAP	0,319	0,402	0,468	0,581
CMP Run 5*	3×scale + mirrors	MLE	**0,319**	**0,410**	**0,470**	**0,581**

The evaluation results are shown in Figures 1,6. From the class prior estimation methods, MAP estimation with the Dirichlet hyperprior achieves the best results. This corresponds to the results of [10], where adding the hyperprior brought noticeable improvement over the MLE estimation, which may have a tendency to overfit. Note that the results from Table 4 are the official post-challenge evaluation not included in the challenge leaderboard, as our predictions were wrongly exported into the challenge run-files.

4 Conclusions

The proposed system achieves the best accuracy on the PlantCLEF 2019 test set - 31.9% on the full set and 41.0% on the test subset for plant identification experts. The results show that even for "data-deficient" plant species, automatic image recognition systems achieve human expert accuracy in visual recognition of plants: The proposed method performed better than 3 of the 5 participating experts in plant recognition. Although the results are promising, there are many opportunities for further improvement of automatic plant recognition systems for data-deficient species, such as one-shot learning and open long-tailed recognition [6] methods.

The increasing precision of the automated plant recognition methods should allow for a better assistance to both nature lovers and biological experts in the fields. For example, showing a shortlist of potential species candidates can decrease the time needed for decision and potentially increase the recognition rate.

Acknowledgements

LP was supported by the UWB project No. SGS-2019-027. MŠ and JM were supported by OP VVV project CZ.02.1.01/0.0/0.0/16019/000076 Research Center for Informatics. We'd like to thank Tomáš Jeníček for his assistance with the image retrieval pipeline in Section 2.1.

References

1. Goëau, H., Bonnet, P., Joly, A.: Plant identification in an open-world (lifeclef 2016). In: CLEF working notes 2016 (2016)
2. Goëau, H., Bonnet, P., Joly, A.: Plant identification based on noisy web data: the amazing performance of deep learning (lifeclef 2017). CEUR Workshop Proceedings (2017)
3. Goëau, H., Bonnet, P., Joly, A.: Overview of expertlifeclef 2018: how far automated identification systems are from the best experts? In: CLEF working notes 2018 (2018)
4. Goëau, H., Bonnet, P., Joly, A.: Overview of lifeclef plant identification task 2019: diving into data deficient tropical countries. In: CLEF working notes 2019 (2019)
5. Joly, A., Goëau, H., Botella, C., Kahl, S., Servajean, M., Glotin, H., Bonnet, P., Vellinga, W.P., Planqué, R., Stöter, F.R., Müller, H.: Overview of lifeclef 2019: Identification of amazonian plants, south & north american birds, and niche prediction. In: Proceedings of CLEF 2019 (2019)
6. Liu, Z., Miao, Z., Zhan, X., Wang, J., Gong, B., Yu, S.X.: Large-scale long-tailed recognition in an open world. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2537–2546 (2019)
7. Radenović, F., Tolias, G., Chum, O.: Fine-tuning cnn image retrieval with no human annotation. IEEE transactions on pattern analysis and machine intelligence (2018)
8. Redmon, J.: Darknet: Open source neural networks in c. `http://pjreddie.com/darknet/` (2013–2016)
9. Saerens, M., Latinne, P., Decaestecker, C.: Adjusting the outputs of a classifier to new a priori probabilities: a simple procedure. Neural computation **14**(1), 21–41 (2002)
10. Sulc, M., Matas, J.: Improving cnn classifiers by estimating test-time priors. arXiv preprint arXiv:1805.08235v2 (2019)
11. Sulc, M., Picek, L., Matas, J.: Plant recognition by inception networks with test-time class prior estimation. Working Notes of CLEF 2018 (2018)
12. Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A.A.: Inception-v4, inception-resnet and the impact of residual connections on learning. In: Thirty-First AAAI Conference on Artificial Intelligence (2017)