Machines vs. Human Experts: Contribution to the ExpertLifeCLEF 2018 Plant Identification Task

Mario Lasseck

Museum fuer Naturkunde Berlin, Germany
`Mario.Lasseck@mfn.berlin`

Abstract. This paper presents deep learning techniques for image-based plant identification at very large scale. Deep Convolutional Neural Networks (DCNNs) for image classification are fine-tuned to identify 10,000 species. The proposed approach is evaluated in the ExpertLifeCLEF 2018 campaign. Results are compared with identifications by human experts and the best models of the LifeCLEF 2017 plant identification task. Performance comes close to the most advanced human expertise and surpasses previous state-of-the-art by 6 % on the official test set.

Keywords: Plant Species Identification, Biodiversity, Deep Convolutional Neural Networks, Deep Learning, Image Classification, Data Augmentation.

1　Introduction

Automated plant identification is a promising tool supporting agriculture automation, biodiversity monitoring and education. It can help professionals and amateurs identifying plants in the field or searching large image collections of flora.

For training automated systems, the data sets of previous LifeCLEF campaigns are provided. They mainly consist of a 'trusted' set of ca. 260,000 images coming from several public databases or institutions dedicated to botany (e.g. Wikimedia, iNaturalist, Flickr, etc.) and a much larger 'noisy' data set of over 1.4 million images collected by searching the web.

In 2017, models and model ensembles based on DCNNs achieved impressive identification results with accuracies near 90 % on the PlantCLEF 2017 test set. This raises the question of how far such automated systems are from human expertise. For this to find out, a subset of the test set is given this year also to plant experts for identification. Results are compared with those of automated systems. Similar man vs. machine experiments were conducted previously in this domain and show significant performance increase over the last few years thanks to recent advances in deep learning [4],[5]. They now come close to the most advanced human expertise. More information on data collection and task can be found in the ExpertLifeCLEF 2018 lab overview [1] and the task summary [2].

This paper is building on previous solutions to the LifeCLEF plant identification task. More details regarding data preparation and augmentation methods can be found in the corresponding working note of 2017 [3].

2   Implementation Details and Model Training

New models are trained via PyTorch [6] instead of MXNet [7] or Digits/Caffe [8],[9] used in 2017. The first model is based on a 152 layer Residual Network architecture (ResNet-152) [10]. The second model is based on a 92 layers deep Dual Path Network (DPN-92) [11]. Both models were pre-trained on ImageNet [12]. ResNet-152 on the ILSVRC [13] version of the data set containing 1.2 million images of 1000 object categories and DPN-92 on a larger version covering 5k object classes, a subset of ImageNet10K [14]. The models are publically available and can be imported in PyTorch via torchvision or pretrainedmodels [15] python packages.

A major difference of PyTorch compared to MXNet is the data loading and augmentation pipeline. Besides random horizontal flip, random rotation and applying color jitter to saturation and brightness for models in 2017, a random resized crop is applied for augmentation in PyTorch. This technique was successfully used for ImageNet training by [16]. It crops various patches at random position of the image with an evenly distributed size between 8 and 100 percent of the image area and additionally changes aspect ratio between 3/4 and 4/3. Furthermore color augmentation is extended by also applying jitter to contrast (factor: 0.3) and hue (factor 0.01) of the image.

For training, images from both, the trusted and noisy data sets are used. A fixed learning rate policy is applied starting with a base learning rate of 0.01. It is decreased by a factor of 0.1 whenever performance on the validation set stopped improving. This is done twice within the training process. Stochastic gradient descent with Nesterov momentum of 0.9 and weight decay of 1e-4 is used for optimization. Although not really necessary for training in PyTorch, all images were resized such that the shorter side becomes 256 pixel while preserving the original aspect ratio.

Another difference to 2017 is the number of crops extracted from the test images. The 5 crops from middle and each corner of the image are doubled by also adding their mirrored version (horizontal flip). To identify the test set, for each plant observation predictions are hierarchically averaged over:

- all 10 patches cropped and mirrored from the image
- all images belonging to the same observation
- all models within the ensemble (for run 1 & 4)

3   Results

Identification performance is evaluated on the whole ExpertLifeCLEF 2018 test set and the expert subset. Results are summarized in table 1 for both sets.

Run 1. To better compare results and to find out if and how much progress was made on identification performance since last year, the best performing ensemble of PlantCLEF 2017 was left unchanged and used in the first run to predict the 2018 test sets. The 12 models in the ensemble consist of different network types (GoogLeNet, ResNet, ResNeXT), are trained on different data sets and use various data augmentation techniques (see [3] for details).

Run 2. For the second run a single model was trained via PyTorch by fine-tuning a pre-trained ResNet-152 network. Training was done with a batch size of 144 and took about one week for 84 epochs using 3 GPUs (Nvidia Geforce 1080 Ti). The model provides the most competitive run to identify images of the expert test set.

Run 3. For the third run a Dual Path Network pre-trained on ImageNet-5k was used for training. With a batch size of 102 training took about two weeks for 141 epochs (also using 3 GPUs). Although performing better on ImageNet-1k compared to ResNet-152 (top-1 accuracy of 79.4 % for DPN-92 vs. 78.4 % for ResNet-152 [15]) performance regarding plant identification is slightly lower on the whole test set and much lower on the expert subset (see table 1).

Run 4. An ensemble of all 14 models from the previous runs (PlantClef 2017 models, ResNet-152 & DPN-92) was created for the fourth run by averaging all predictions. This is the best performing run regarding the complete test set.

Table 1. Official scores on the ExpertLifeCLEF 2018 test sets.

Run	# Models	Top-1 Accuracy [%] Whole test set	Top-1 Accuracy [%] Expert test set
1	12	82.6	76.0
2	1	84.8	**78.7**
3	1	84.7	77.3
4	14	**87.5**	77.3

Figure 1 compares scores of all submissions to the ExpertLifeCLEF 2018 task. The above described models ranked second place among all participating teams. In Figure 2 performance is compared with those of human experts. Only 5 out of 9 experts achieve better identification results than the submitted automated systems even if only a single model is involved. More results and evaluation details can be accessed via the ExpertLifeCLEF 2018 webpage [17] and the crowdAI leaderboard [18].
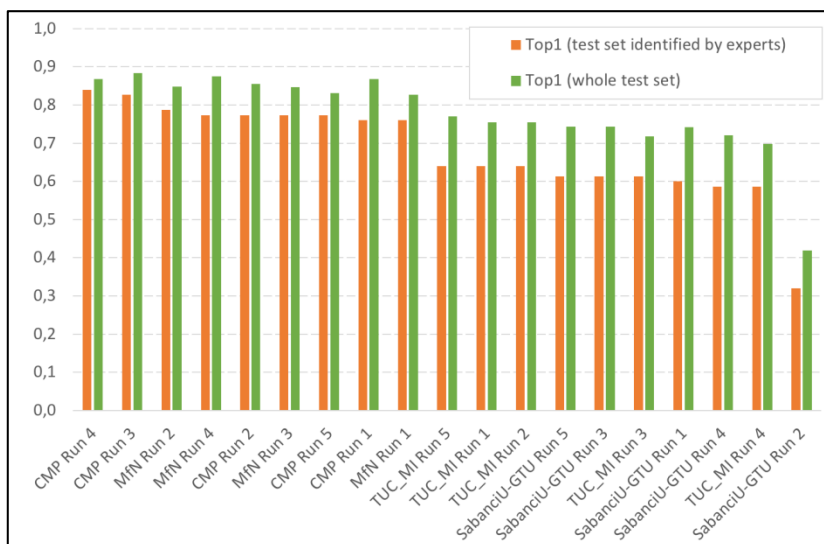
Fig. 1. Official scores (top-1 accuracy) of the ExpertLifeCLEF 2018 plant identification task. The above described methods and submitted runs belong to MfN.
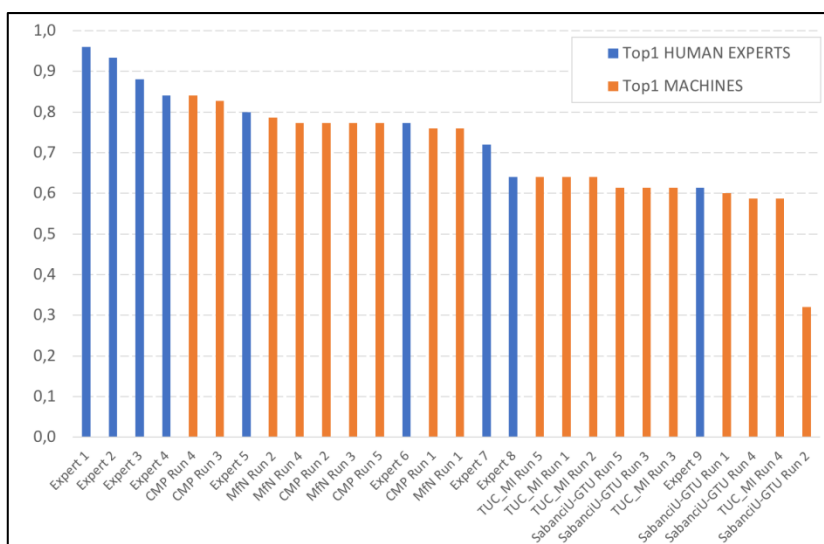

Fig. 2. Comparison of top-1 accuracy between "machines" and the human experts. The above described methods and submitted runs belong to MfN.

4 Discussion

Both single models of the second and third run give better results than the ensemble of 12 models of the first run from the previous year. Although using the same data set for training, performance is significantly improved with PyTorch and its extended data augmentation methods. Table 2 compares performance of different network architectures from 2017 and 2018 models. All models use the exact same data sets for training and validation. The relatively low accuracy scores can be explained by the fairly high number of non-plant images in the validation set because both trusted and noisy data sets were used for training and validation. Also validation was carried out per image (one center patch per image) without augmentation.

Table 2. Performance of different network architectures.

Year	Framework	Network architecture	Top-1 Accuracy [%]
2017	Digits/Caffe	GoogLeNet	52.7
2017	MXNet	ResNet-152	52.5
2017	MXNet	ResNeXt-101-64x4d	53.0
2018	PyTorch	ResNet-152	59.4
2018	PyTorch	DualPathNet92_5k	59.0

Even with the same network type (ResNet-152) results are better with PyTorch. This shows once again the big influence of data augmentation methods. Another reason for better performance in PyTorch may be the pre-processing of images for training via MXNet. Image quality is slightly reduced when creating the MXNet Image RecordIO data set for faster reading, possibly resulting in a loss of some relevant information.

Nevertheless adding the less performing older models to the ensemble of new models, significantly improves results on the whole test set in the fourth run. On the contrary, results on the expert subset are not improved for this run.

The expert set seems to be much more challenging. Identification accuracies are consistently lower for all runs compared to the whole test set. Despite the difficulties, all submitted 'machines' could keep up with almost half of the human experts in this image-based plant identification task.

References

1. Joly A, Goëau H, Botella C, Glotin H, Bonnet P, Planqué R, Vellinga WP, Müller H (2018) Overview of LifeCLEF 2018: a large-scale evaluation of species identification and recommendation algorithms in the era of AI, In: Proceedings of CLEF 2018
2. Goëau H, Bonnet P, Joly A (2018) Overview of ExpertLifeCLEF 2018: how far automated identification systems are from the best experts? In: CLEF working notes 2018
3. Lasseck M (2017) Image-based Plant Species Identification with Deep Convolutional Neural Networks. In: Working Notes of CLEF 2017 Conference
4. Bonnet P, Joly A, Goëau H, et al. (2016) Plant Identification: Man vs. Machine, Multimedia Tools and Applications, vol. 75, number 3, pages 1647-1665, 2016
5. Bonnet P, Goëau H, Hang S, Lasseck M, Šulc M, Joly A (2018) Plant Identification: Experts vs. Machines in the Era of Deep Learning. In Joly A, Vrochidis S, Karatzas K, Karppinen A, Bonnet P (Eds.), *Multimedia Tools and Applications for Environmental & Biodiversity Informatics* (pp. 139-155). Springer International Publishing.
6. Paszke A, Gross S, Chintala S et al. (2017) Automatic differentiation in PyTorch. In: NIPS-W
7. Chen T et al. (2016) MXNet: A Flexible and Efficient Machine Learning Library for Heterogeneous Distributed Systems. In: Neural Information Processing Systems, Workshop on Machine Learning Systems, 2016
8. Digits Homepage, https://developer.nvidia.com/digits, last accessed 2018/05/29
9. Jia Y et al. (2014) Caffe: Convolutional Architecture for Fast Feature Embedding. In: arXiv:1408.5093, 2014
10. He K, Zhang X, Ren S, Sun J (2016) Deep Residual Learning for Image Recognition. In: CVPR, 2016
11. Chen Y, Li J, Xiao H et al. (2017) Dual Path Networks. arXiv preprint arXiv:1707.01629
12. Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L (2009) Imagenet: A largescale hierarchical image database. In: IEEE Conference on Computer Vision and Pattern Recognition, 2009. pp. 248–255 (2009)
13. Russakovsky O et al. (2014) ImageNet Large Scale Visual Recognition Challenge. In: arXiv:1409.0575, 2014
14. Deng J, Berg AC, Li K, Fei-Fei L (2010) What Does Classifying More Than 10,000 Image Categories Tell Us? In: Computer Vision–ECCV 2010, pp. 71-84
15. https://github.com/Cadene/pretrained-models.pytorch
16. Szegedy C et al. (2014) Going Deeper with Convolutions, In: arXiv:1409.4842, 2014
17. ExpertLifeCLEF 2018 Homepage, http://www.imageclef.org/node/231, last accessed 2018/05/29
18. crowdAI challenge page, https://www.crowdai.org/challenges/lifeclef-2018-expert, last accessed 2018/05/29