Plant Identification with Large Number of Classes: SabanciU-GebzeTU System in PlantCLEF 2017

Sara Atito[1], Berrin Yanikoglu[1], and Erchan Aptoula[2]

[1] Faculty of Engineering and Natural Sciences, Sabanci University, Istanbul, Turkey
[2] Institute of Information Technologies, Gebze Technical University, Kocaeli, Turkey
{saraatito,berrin}@sabanciuniv.edu
eaptoula@gtu.edu.tr

Abstract. We describe the plant identification system that was submitted to the LifeCLEF plant identification campaign in 2017 [1], as a collaboration of Sabanci University and Gebze Technical University. Similar to our system that got a very close second place in 2016, we fine-tuned two well-known deep learning architectures (VGGNet and GoogLeNet) that were pre-trained on the object recognition dataset of ILSVRC 2012 and used an ensemble of 4-9 networks using score-level combination. Our best system was obtained with a classifier fusion of 9 networks trained with some differences in train settings, achieving an average inverse rank of 0.634 on the official test data, while the first place system achieved an impressive score of 0.92.

Keywords: plant identification, deep learning, convolutional neural networks

1 Introduction

Automatic plant identification addresses the identification of the plant species in a given photograph. Plant identification challenge within the Conference and Labs of the Evaluation Forum (CLEF) [2,3,4,5,6,7,1] is the most well-known annual event that benchmarks content-based image retrieval of plants. The campaign has been run since 2011, with plant species and number of training images almost doubling every year, reaching to 10,000 classes in the 2017 evaluation. Considering very high similarities between species and a large variety of imaging and plant conditions, the problem is rather challenging.

Our team participated in the PlantCLEF 2017 campaign under the name of *SabanciU-GebzeTU*. In all of our runs, we used an ensemble of 4-9 convolutional networks, with different classifier combination criteria. The base networks were pre-trained deep convolutional neural networks of GoogLeNet [8] and VGGNet [9] that were fine-tuned with plant images. The campaign organizers provided two separate data sets: the main training set consisted of 256,203 images with clean labels (collected from the *Encyclopedia of Life (EOL)*) and the web crawled

data consisted of around 1.6 million of images with noisy labels. The test set was sequestered until a few weeks before results submission. Details of the campaign can be found in [1].

The rest of this paper is organized as follows. Section 2 describes our approach based on: fine-tuning GoogLeNet and VGGNet models for plant identification and applying score-level classifier fusion. Section 3 describes the data sets and experimental results. The paper concludes in Section 4 with the summary and discussion of the utilized methods and obtained results.

2   Approach

Our approach was fine-tuning and fusing of two successful deep learning models, i.e. GoogLeNet [8] and VGGNet[9], using the implementations provided in the Caffe deep learning framework [10]. These models are, respectively, the first-ranked and second-ranked architectures of the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) 2014–both trained on the ILSVRC 2012 dataset with 1.2 million labeled images of 1,000 object classes.

In this work, we fine-tuned the GoogLeNet and VGGNet models starting from the learned weights of our PlantCLEF2016 system [11]. In the first network, we used only the training portion of EOL with internal augmentation (during training at each iteration a random crop of the image is used and randomly mirrored horizontally), to get some quick results. This network was the VGGNet architecture with all but the last layer of weights being fixed. In fact, in all of the experiments, we could only fine-tune the last 1-2 layers, as learning was very slow otherwise. This network achieved 41% accuracy.

After getting the base system running, we started using 8-fold external augmentation for training and later we started to incorporate images from the noisy dataset into the training data: as the web crawled data is not reliable, we tested 200,000 images from the noisy data set using the best networks we had thus far and took only those images for which prediction matched the groundtruth.

We also tried VGGNET using Batch Normalization and GoogleNet architecture, with roughly similar performance. In both of these networks, all of the layers were fixed except for the last one due to scarce computing resources. Another network concentrated on the most common 1000 species and while we found that this network only achieved a 27% accuracy, it helped improve the performance of the ensemble like all other networks. In this fashion, each successive network (for a total of 9 different ones) was trained for either more iterations, or with new data added, or with different network architecture. At last, we trained one of the previous networks with *all* available training data, merging the validation set to the training set. This was done for only one network given the limited time.

Score-level averaging is applied to combine the prediction scores assigned to each of the augmented patches within a single network. As for the final systems, the obtained scores from all networks are combined using Borda count [12] or based on the maximum score of different classifiers.
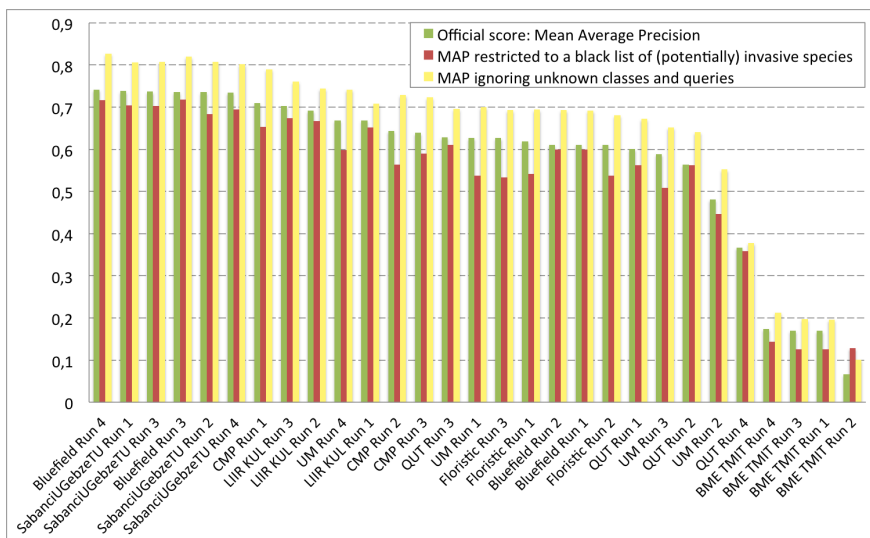
Fig. 1. The official released results of PlantCLEF 2017

Our main problem was computational resources, faced with a very large number of classes and large amount of data. All trains and tests were run on a Linux system with a Tesla K40c and 12GB of video memory and in most cases training a network took 2-3 days.

3   Experimental Results

For training and validating our system, we used the EOL data consisting of 256,203 images of different plant organs, belonging to 10,000 species. Specifically, we randomly divided the *training* portion of the dataset into two subsets for training and validation, with 174,280 and 81,923 images, respectively. The test portion of the dataset consists of a separate set of 25,170 images that was sequestered by the organizers, until the last weeks of the campaign. We will call these three subsets train, validation and test subsets respectively in the remainder of this paper.

The base accuracy of the networks trained with all of the 10,000 classes ranged from 41% to 48.4% and the combined accuracy was 61.03%, on the validation subset. The combination was helpful even with highly correlated networks and taking less successful networks from the ensemble always reduced the performance The most successful network, based on the accuracy of the validation set, was the VGGNet using the largest training set (the train subset and around 60,000 samples from noisy data) and with a large batch size (60).

The submitted runs are described below and the results (mean inverse rank) released by the campaign organizers are shown in Figure 1 and given in [1].

More details of the used training set in our experiments with rank comparison
are shown in table 1.

Table 1. Rank comparison of the CLEF2017 published results that used (EOL) and
(EOL+Noisy) data set

Run	Trusted (EOL)		
	Score	Top 1	Top 5
CMP Run 3	0.807	0.741	0.887
FHDO_BCSG Run 1	0.792	0.723	0.878
KDETUT Run 1	0.772	0.707	0.85
CMP Run 4	0.733	0.641	0.849
UM Run 1	0.7	0.621	0.795
PlantNet Run 1	0.613	0.513	0.734
SabanciUGebzeTU Run 2	0.581	0.508	0.68
Run	Trusted (EOL) + Noisy		
	Score	Top 1	Top 5
MarioTsaBerlin Run 4	0.92	0.885	0.962
KDETUT Run 4	0.853	0.793	0.927
KDETUT Run 3	0.837	0.769	0.922
UM Run 3	0.798	0.727	0.886
UM Run 4	0.789	0.715	0.882
SabanciUGebzeTU Run 4	0.638	0.557	0.738
SabanciUGebzeTU Run 1	0.636	0.556	0.737
SabanciUGebzeTU Run 3	0.622	0.537	0.728

- Run 1. In this run, the combination was done based on Borda count, with
  classifier confidence to break the ties.
- Run 2. This ensemble only used based systems trained with EOL data.
- Run 3. This system was the same as System 4 except for using a combination
  based on maximum confidence.
- Run 4. This system was the same as System 1 except for classifier combina-
  tion weights.

4   Conclusions

The main objective was to preserve the high scores we obtained in 2016, despite the 10-fold increase in the number of classes [11]. Unfortunately, the large number of classes and limited computational power made it impossible to successfully fine-tune the networks. While our results were significantly below the best performing system this year, our results are not too far from our results last year, despite 10-fold increase in classes. It was also a challenging exercise to deal with a large, real life problem.

References

1. Joly, A., Goëau, H., Glotin, H., Spampinato, C., Bonnet, P., Vellinga, W.P., Lombardo, J.C., Planqué, R., Palazzo, S., Müller, H.: Lifeclef 2017 lab overview: multimedia species identification challenges. In: CLEF 2017 Proceedings, Springer Lecture Notes in Computer Science (LNCS). (2017)
2. Goëau, H., Bonnet, P., Joly, A., Boujemaa, N., Barthelemy, D., Molino, J.F., Birnbaum, P., Mouysset, E., Picard, M.: The CLEF 2011 plant images classification task. In: CLEF (Notebook Papers/Labs/Workshop). (2011)
3. Goëau, H., Bonnet, P., Joly, A., Yahiaoui, I., Barthelemy, D., Boujemaa, N., Molino, J.F.: The ImageCLEF 2012 plant identification task. In: CLEF (Online Working Notes/Labs/Workshop). (2012)
4. Goëau, H., Bonnet, P., Joly, A., Bakic, V., Barthelemy, D., Boujemaa, N., Molino, J.F.: The ImageCLEF 2013 plant identification task. In: CLEF (Working Notes). (2013)
5. Goëau, H., Joly, A., Bonnet, P., Selmi, S., Molino, J.F., Barthelemy, D., Boujemaa, N.: LifeCLEF plant identification task 2014. In: CLEF (Working Notes). (2014)
6. Goëau, H., Bonnet, P., Joly, A.: LifeCLEF plant identification task 2015. In: CLEF (Working Notes). (2015)
7. Goëau, H., Bonnet, P., Joly, A.: Plant identification in an open-world (lifeclef 2016). In: CLEF working notes 2016. (2016)
8. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: IEEE Conference on Computer Vision and Pattern Recognition. (2015)
9. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. Computing Research Repository (CoRR) (2014) arXiv: 1409.1556.
10. Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T.: Caffe: Convolutional architecture for fast feature embedding. In: Proceedings of the 22nd ACM International Conference on Multimedia. (2014) 675–678
11. Mehdipour-Ghazi, M., Yanikoglu, B., Aptoula, E.: Open-set plant identification using an ensemble of deep convolutional neural networks. In: Working Notes of CLEF 2016 - Conference and Labs of the Evaluation forum, Évora, Portugal, 5-8 September, 2016. (2016) 518–524
12. Van Erp, M., Schomaker, L.: Variants of the Borda count method for combining ranked classifier hypotheses. In: In the seventh international workshop on frontiers in handwriting recognition. (2000)