Plant Identification with Deep Learning Ensembles in ExpertLifeCLEF 2018

Sara Atito[1], Berrin Yanikoglu[1], Erchan Aptoula[2], İpek Ganiyusufoğlu[1], Aras Yıldız[1], Kerem Yıldırır[1], Barış Sevilmiş[1], and M. Umut Şen[1]

[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. This work describes the plant identification system that we submitted to the ExpertLifeCLEF plant identification campaign in 2018. We fine-tuned two pre-trained deep learning architectures (SeNet and DensNetwork) using images shared by the CLEF organizers in 2017. Our main runs are 4 ensembles obtained with different weighted combinations of the 4 deep learning architectures. The fifth ensemble is based on deep learning features but uses Error Correcting Output Codes (ECOC) as the ensemble. Our best system has achieved a classification accuracy of 74.4%, while the best system obtained 86.7% accuracy, on the whole of the official test data. This system ranked 4th place among all the teams, but matched the accuracy of one of the human experts.

Keywords: plant identification, deep learning, convolutional neural networks

1 Introduction

Automatic plant identification is the problem of identifying the given plant species in a given photograph. Plant identification challenge of the Conference and Labs of the Evaluation Forum (CLEF) [1,2,3,4,5,6,7,8] is the most well-known annual event that benchmarks the progress in identification of plant species. The campaign has been running since 2011, with plant species reaching 10,000 classes in the 2017 evaluation.

The emphasis of the campaign changes slightly from year to year, while the core of the campaign is to benchmark plant identification progress. This year's emphasis was on measuring automatic systems' performances with that of human experts. For that reason, a subset of the test data was labelled by human experts and the systems were evaluated on their accuracy on the whole test set, as well as their performance on the subset. The details of the plant identification and the overall LifeCLEF campaigns are described in [8] and [9] respectively.

We have been participating into this campaign since 2011, first with traditional approaches and carefully selected features [10,11,12] and then with deep learning approaches [13]. While the traditional approaches worked well on the simpler problem of leaf based identification (leaf images on simple backgrounds), deep learning approaches brought a significant increase in accuracy despite much increased problem complexity (unrestricted photographs and 10,000 classes).

This year our team participated in the ExpertLifeCLEF2018 challenge under the name of *SabanciU-GTU*. In our main 4 runs (Runs 1, 3, 4, 5), we have used an ensemble of four convolutional networks according to different combination weights. The networks were pre-trained deep convolutional neural networks of SeNet [14] and DensNetwork [15] that were fine-tuned with plant images. In the fifth system, we took the deep learning features (last convolutional layer activations) of our SeNet system and trained 200 different binary classifiers to form an Error Correcting Codes (ECOC) ensemble.

The training data was obtained from CLEF, as a combination of data collected from the Encyclopedia of Life (EOL) and images collected from the web and shared by CLEF in 2017. This latter set is noisy as it is not verified by experts for correctness. The submitted systems were different combination schemes applied to the four models.

The rest of this paper is organized as follows. Section 2 describes the proposed methods based on the fine-tuning of SeNet and DensNetwork models for plant identification, data augmentation, and classifiers' fusion. Section 4 is dedicated to the description of the utilized dataset and presentation of designed experiments and their results. The paper concludes in Section 5 with the summary and discussion of the utilized methods and obtained results.

2 Core System

Our approach was based on fine-tuning and fusing of two successful deep learning models, namely SeNet [14] and DensNetwork[15]. 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.

SeNet [14], Winner of ImageNet 2017 Classification Task [16], introduces a building block for convolution neural networks that improves channel interdependencies. The main idea is to weight each channel adaptively based on its importance. SE-block is flexible which means that it can be integrated into any modern deep learning architecture. In this work, we utilized SE-blocks with ResNet-50 [17] module.

DensNetwork [15] are built from dense blocks and pooling operations where there is a connection between each block to every receding blocks. Thus, with n blocks, there are $n(n+1)/2$ direct connections. Input of each dense block is an iterative concatenation of previous feature maps. One of the advantages of DensNetwork is that it lessens the vanishing-gradient problem which makes it easy to train.

Score-level averaging is applied to combine the prediction scores assigned to each class for all the augmented patches within a single network and then for combining the scores obtained for different images of the same unique plant (called an "observation" in the campaign terminology).

All training and tests were run on a linux system with a Titan X Pascal GPU and 12GB of video memory.

3 Error-Correcting Output Codes

As a second ensemble approach, we tried the Error Correcting Output Codes (ECOC) approach [18]. In ECOC, a number of binary classifiers are trained such that each one is assigned a separate dichotomy of the classes, which is defined by a given ECOC matrix. In the ECOC matrix M, the jth column indicates the dichotomy assigned for base classifier h_j. That is, a particular element $M_{ij} \epsilon \{+1, -1\}$ indicates the desired label for class c_i to be used in training the base classifier h_j. The ith row of M, denoted as M_i, is the codeword for class c_i indicating the desired output for that class.

A given test instance x is first classified by each base classifier, obtaining the output vector $y = [y_1, ..., y_L]$ where y_j is the output of the classifier h_j for the given input x. Then, the distance between y and the codeword M_i of class c_i is computed by using a distance metric such as the Hamming distance. The class c_k for which this distance is minimum, is chosen as the estimated class label:

$$k = argmin_{i=1...K}\ d(y, M_i)$$

We took the deep learning features (last convolutional layer activations) of our SeNet system (System2) and trained 200 different binary classifiers according to the predetermined ECOC matrix.

4 Experiments and Results

The first three systems are trained using SeNet-ResNet-50 architecture. For training the first system, we only used the EOL data consisting of $256,203$ images of different plant organs, belonging to 10,000 species. Internal augmentation was applied during training (at each iteration, a random crop of the image is used and randomly mirrored horizontally). For validation, we used the plant test dataset of LifeCLEF 2017 consisting of $25,170$ images.

For the second system, several data augmentation are applied to the training images like saliency detection [19], flip, and several rotation angles. In total, number of images in the training dataset after augmentation is around $4,500,000$ images and the system was trained over 10 epochs. For the third system, we trained using all of the available data with augmentation (EOL data, web-collected noisy data, and testing set of LifeCLEF2017) excluding $1,000$ images from test 2017 for validation. This system was trained over 25 epochs. The fourth system is trained using DensNetwork using the same training data as in System 3. Training DensNetwork was quite slow, therefore, we trained system 4 over only 5 epochs.

We implemented SeNet and DensNetwork models using the Caffe deep learning framework [20]. All the weights were fine-tuned, while the last layer was learned from scratch. We used the same learning rate for all of the system which is 0.01.

Run 1,3,4,5. Different weighted combinations of the same basic four deep learning systems described in Section 2. In System 5, that was the best performing system by a 0.001 margin, we used the image quality information that is given inside the metadata in the xml files. The score of each image is weighted using the quality information. In the absence of quality information, no weighting is applied.

Run 2. The ECOC ensemble where 200 base classifiers were trained on binary classification tasks set forth according to a predetermined, random ECOC matrix. The ECOC matrix was initialized randomly, and then simulated annealing was used to increase the Hamming distance between rows. As features, we used the deep learning features obtained from the last convolutional layer of first system described above, and trained 2-hidden layer shallow networks (500 hidden nodes at each layer) as base classifiers.

While the accuracy of this system fell short of the performance of the deep learning architectures, the system shows promise in that the accuracy increases as we increase the number of base classifiers: from 51% with 100 base classifiers, to 59% and 61% with 200 and 300 base classifiers, on the LifeCLEF 2017 test data. The training times are also less than one tenth of that of one deep architecture (around 2-3 hours per 100 base classifiers on an iMac).

As a promising and fast alternative, we are planning on working on improvements of the ECOC ensemble as proposed in [21] and [22].

Test Results. We submitted the classification results of the before mentioned systems on the official test set of the ExpertLifeCLEF 2018. The utilized official metric for evaluation was the average accuracy on a small subset of the test data that was also identified by human experts. Results on the whole test set were also provided. The released results by the challenge organizers are shown in Figure 1 and given in [9].

Our best system has achieved a top-1 classification accuracy of 74.4%, while the best system obtained 86.7% accuracy on the whole official test data. This system ranked 4th place among all the teams, but matched the accuracy of one of the human experts.

Our results for the small subset that is also labelled by human experts is 61.3%, while the 9 human experts scores range from 96% to 61.3%, on this subset. In other words, our best system has reached the top-1 identification accuracy of one of the human experts.

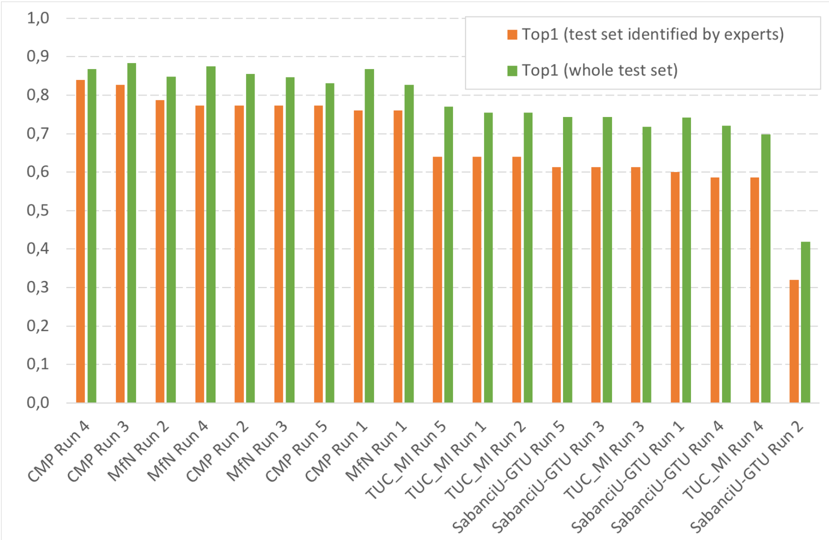

Fig. 1. The official released results of ExpertLifeCLEF 2018

5 Conclusions

The competition that has been running for several years now has seen a shift from hand-crafted features and to deep learning classifiers in the last years. Our goal this year was to use the best performing pre-trained architectures while diversifying the base classifiers within the ensemble. Considering the fact that we only had one machine with GPU, we consider the performance of our system (74.4% accuracy) satisfactory on such a complex problem (10,000 classes). For the future, we plan to work on better ensemble techniques with deep architectures, including improvements of the ECOC ensemble.

Acknowledgments. We gratefully acknowledge NVIDIA Corporation with the donation of the Titan X Pascal GPU used in this research.

References

1. 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)
2. 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)
3. 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)
4. 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)

5. Goëau, H., Bonnet, P., Joly, A.: LifeCLEF plant identification task 2015. In: CLEF (Working Notes). (2015)
6. Goëau, H., Bonnet, P., Joly, A.: Plant identification in an open-world (lifeclef 2016). In: CLEF working notes 2016. (2016)
7. 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)
8. 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
9. Joly, A., Goau, H., Glotin, H., Spampinato, C., Bonnet, P., Vellinga, W.P., Lombardo, J.C., Planqué, R., Palazzo, S., Müller, H.: Overview of lifeclef 2018: a large-scale evaluation of species identification and recommendation algorithms in the era of ai. In: Proceedings of CLEF 2018
10. Yanikoglu, B., Aptoula, E., Tirkaz, C.: Sabanci-Okan system at imageclef 2011: Plant identification task. In: CLEF (Working Notes). (2011)
11. Yanikoglu, B., Aptoula, E., Tirkaz, C.: Sabanci-Okan system at imageclef 2012: Combining features and classifiers for plant identification. In: CLEF (Working Notes). (2012)
12. Mehdipour-Ghazi, M., Yanikoglu, B., Aptoula, E.: Plant identification using deep neural networks via optimization of transfer learning parameters. Neurocomputing **235** (2017) 228–235
13. Yanikoglu, B., Aptoula, E., Tirkaz, C.: 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
14. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. arXiv preprint arXiv:1709.01507 (2017)
15. Huang, G., Liu, Z., van der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2017)
16. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C., Fei-Fei, L.: ImageNet Large Scale Visual Recognition Challenge. International Journal of Computer Vision (IJCV) **115**(3) (2015) 211–252
17. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. (2016) 770–778
18. Dietterich, T.G., Bakiri, G.: Solving multiclass learning problems via error-correcting output codes. Journal of Artificial Intelligence Research (1995) 263–286
19. Harel, J., Koch, C., Perona, P.: Graph-based visual saliency. In: Advances in neural information processing systems. (2007) 545–552
20. 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. (2014) 675–678
21. Zor, C., Yanikoglu, B., Windeatt, T., Alpaydin, E.: FLIP-ECOC: a greedy optimization of the ECOC matrix. In: Proceedings of the 25th International Symposium on Computer and Information Sciences, Springer (2010) 149 – 154
22. Zor, C., Yanikoglu, B., Merdivan, E., Windeatt, T., Kittler, J., Alpaydin, E.: BeamECOC: A local search for the optimization of the ECOC matrix. In: 23rd International Conference on Pattern Recognition, ICPR 2016, Cancún, Mexico, December 4-8, 2016. (2016) 198–203