

Open-set Plant Identification Using an Ensemble of Deep Convolutional Neural Networks

Mostafa Mehdipour Ghazi¹, Berrin Yanikoglu¹, and Erchan Aptoula²

¹ Faculty of Engineering and Natural Sciences, Sabanci University, Istanbul, Turkey

² Institute of Information Technologies, Gebze Technical University, Kocaeli, Turkey
{mehdipour,berrin}@sabanciuniv.edu
eaptoula@gtu.edu.tr

Abstract. Open-set recognition, a challenging problem in computer vision, is concerned with identification or verification tasks where queries may belong to unknown classes. This work describes a fine-grained plant identification system consisting of an ensemble of deep convolutional neural networks, within an open-set identification framework. Two well-known deep learning architectures of VGGNet and GoogLeNet, pre-trained on the object recognition dataset of ILSVRC 2012, are fine-tuned using the plant dataset of LifeCLEF 2015. Moreover, GoogLeNet is fine-tuned using plant and non-plant images, to serve for rejecting samples from non-plant classes. Our systems have been evaluated on the test dataset of PlantCLEF 2016 by the campaign organizers and the best proposed model has achieved an official score of 0.738 in terms of the mean average precision, while the best official score was announced to be 0.742.

Keywords: open-set recognition, plant identification, deep learning, convolutional neural networks

1 Introduction

Automated plant identification is a fine-grained image classification problem concerned with small inter-class and large intra-class variations. As with many other problems, recent research in this area has concentrated on deep learning methods and significant improvements have been obtained compared to traditional methods [1,2,3,4]. Raw data is fed into these networks in multiple levels, allowing the system to automatically discover high-level features for plant species recognition; however, due to their high computational complexity for training from scratch, the existing pre-trained deep networks are fine-tuned for plant identification purposes [2,3,5].

Plant identification challenge of the Conference and Labs of the Evaluation Forum (CLEF) [6,7,8,9,10] is among the most well-known annual event that benchmark content-based image retrieval from structured plant databases including photographs of leaves, branches, stems, flowers, and fruits. The latest annotated plant dataset has been provided by the LifeCLEF 2015 campaign

with over 100,000 pictures of herbs, trees, and ferns belonging to 1,000 species collected from Western Europe.

Open-set recognition is a challenging task in computer vision which deals with identification or verification problems where samples from unknown classes may be presented to the system [11,12]. To create a similar scenario, the PlantCLEF 2016 campaign has provided a test image query which is very different from the previous years' CLEF datasets in nature [16]. This dataset contains unknown plant species and non-plant objects; hence it is an open-world plant dataset. Therefore, the task challenges are not limited to automatically recognizing the known plant species, but also rejecting those unknown objects/plants.

Our team participated in the PlantCLEF 2016 challenge under the name of *SabancıUGEbzeTU* and came a very close second overall. In this work, we fine-tuned the pre-trained deep convolutional neural networks of GoogLeNet [13] and VGGNet [14] for plant identification using the LifeCLEF 2015 plant task datasets. We augment this data using different image transforms such as rotation, translation, reflection, and scaling to overcome overfitting while training and to improve the performance during testing the system on highly noisy test data. The overall system is then composed from these two networks, using score-level averaging. To enable rejections, we trained another deep learning system to separate plants from non-plants, by fine-tuning GoogLeNet using the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) 2012 dataset as negative examples (after removing the potted plant category).

The rest of this paper is organized as follows. Section 2 describes the proposed methods based on the fine-tuning of GoogLeNet and VGGNet models for plant identification and open-set rejection, data augmentation, and classifiers' fusion. Section 3 is dedicated to the description of the utilized dataset and presentation of designed experiments and their results. The paper concludes in Section 4 with the summary and discussion of the utilized methods and obtained results.

2 Proposed Method

Our proposed method for automated plant identification is based on fine-tuning and fusing of two successful deep learning models, i.e. GoogLeNet [13] and VGGNet[14]. 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.

GoogLeNet [13] contains 57 convolutional layers, 14 pooling layers, and one fully-connected layer while VGGNet [14] involves 16 convolutional layers, five pooling layers, and three fully-connected layers. Both networks take color image patches of size 224×224 pixels as the input and connect a linear layer with Softmax activation in the output.

We used the plant task datasets of LifeCLEF 2015 for fine-tuning the pre-trained models. Data augmentation is applied to decrease the chance of overfitting during training and to improve performance while testing. For this purpose,

we randomly extract K square patches from each original image around its center. The original image is also rotated by $\pm R$ degrees and the largest square image is cropped from the center of each rotated image. All the extracted patches as well as the original image are then scaled to 256×256 pixels and the mean image is subtracted from them. Finally, five patches of size 224×224 pixels are extracted from four corners and the center of each image. These patches are then reflected horizontally, resulting in $10 \times (K + 2 + 1)$ patches per image in total. Training parameters consisted of weight decay of 0.0002, a base learning rate of 0.001, and a batch size of 20.

Score-level averaging is applied to combine the prediction scores assigned to each class for all the augmented patches within a single network. Finally, the obtained scores from every deep network classifier are combined using score-level averaging.

For the unknown-class rejection task within the unseen data, we separately fine-tuned the GoogLeNet model for a binary classification problem, i.e. plants vs. non-plants. The training samples for this purpose consist of plant images from LifeCLEF 2015 and non-plant images from ILSVRC 2012. Using the obtained results from the binary classifier, we reject those images that are classified as non-plant and receive a low confidence score in the main plant identification system.

3 Experiments and Results

For training and validating our system, we used the plant dataset of LifeCLEF 2015 [15] consisting of 113,204 images of different plant organs belonging to 1,000 species of trees, herbs, and ferns. Specifically, we randomly divided the *training portion* of the dataset into two subsets for training and validation, with 70,904 and 20,854 images, respectively. The test portion of the dataset consists of a separate set of 21,446 images; however the ground-truth for the test dataset was only recently released and thus used in a limited fashion in our system. We will call these three subsets train, validation and test subsets respectively, in the remainder of this paper.

The PlantCLEF 2016 test dataset includes 8,000 samples submitted by the users of the mobile application Pl@ntNet [16]. This dataset is highly noisy and essentially different from the plant task datasets of LifeCLEF 2015 since it contains pictures of unknown plant/non-plant objects. Therefore, we use the data augmentation approach explained in Section 2 and set $K = 5$ and $R = 10$ to train and test 80 patches per image. In some experiments, we also set $K = 9$ and $R = \{10, 20\}$ to train and test 140 patches per image.

To fine-tune GoogLeNet for the open-set unknown-class rejection task, K is set to zero and no rotation is applied so as to augment data with only 10 patches per image. The training data for this problem is obtained by combining the earlier training subset of LifeCLEF 2015 for the plant class samples and an equal number of non-plant object samples from ILSVRC 2012, giving about 140,000 samples in total. In addition, the validation set was obtained from combining the

earlier validation subset of LifeCLEF 2015 and an equal number of non-plant object samples of ILSVRC 2012, giving about 40,000 samples in total. Images used in the training and validation subsets contained distinct samples.

We implemented GoogLeNet and VGGNet models using the Caffe deep learning framework [17] with pre-trained weights obtained from Caffe Model Zoo provided by the Berkeley Vision and Learning Center (BVLC). In the rest of this section, we explain our conducted experiments, their validation results, and the prepared runs.

So we have not used additional plant images for training any of the systems. Furthermore, all systems were fully automatic except for Run 3 where we manually removed 90 reject images in order to obtain scores that would not be affected by rejects.

Run 1. In this run, we first fine-tuned the pre-trained GoogLeNet and VGGNet models using the 70,904-image train subset of LifeCLEF 2015 that we augmented with 140 and 80 patches per image till 600,000 and 500,000 iterations, respectively. Next, we fine-tuned the pre-trained GoogLeNet with almost all of the data (training+test+validation/2) augmented with 80 patches per image until 200,000 iterations. We fused all the obtained scores for augmented image patches from different networks' classifiers. With this system, we achieved an accuracy of 79.80% on the given validation set (validation/2).

To reject test samples of unknown classes, we fine-tuned the pre-trained GoogLeNet with the obtained plant/non-plant dataset until 100,000 iterations and achieved an accuracy of almost 100% on the given validation set. Next, we tested the combined deep model on the non-plant validation set; from the obtained plant identification prediction scores it was inferred that these scores have a uniform-like distribution. We set the rejection threshold to $T = 0.4$ based on the top-1 score for maximally rejection of non-plant samples and minimum rejection of plant data. In the test stage, we utilized our combined system for score prediction as well as rejection of samples whose top-1 score was less than 0.4. We rejected 480 images from the test dataset in this manner and obtained our first run.

Run 2. All the implemented steps for achieving this run were the same as those performed for Run 1. The only difference was that no image rejection was performed in obtaining this run.

Run 3. All the steps utilized in preparing this run were common with Run 1 except for the rejection method. In this experiment, we manually reviewed the text images and put aside 90 non-plant images.

Test Results. We submitted the classification results of the aforementioned systems on the official test set of the PlantCLEF 2016. The utilized metric for evaluation was the mean average precision. In other words, all prediction scores

Acknowledgments. This work is supported by the Scientific and Technological Research Council of Turkey (TUBITAK) under the grant number 113E499. Dr. Aptoula was at Okan University when this work was conducted.

References

1. Chen, Q., Abedini, M., Garnavi, R., Liang, X.: IBM research Australia at LifeCLEF 2014: Plant identification task. In: CLEF (Working Notes). (2014)
2. Choi, S.: Plant identification with deep convolutional neural network: SNUMedinfo at LifeCLEF plant identification task 2015. In: CLEF (Working Notes). (2015)
3. Champ, J., Lorieul, T., Servajean, M., Joly, A.: A comparative study of fine-grained classification methods in the context of the LifeCLEF plant identification challenge 2015. In: CLEF (Working Notes). (2015)
4. Ghazi, M.M., Yanikoglu, B., Aptoula, E., Muslu, O., Ozdemir, M.C.: Sabanci-Okan system in LifeCLEF 2015 plant identification competition. In: CLEF (Working Notes). (2015)
5. Ge, Z., McCool, C., Sanderson, C., Corke, P.: Content specific feature learning for fine-grained plant classification. In: CLEF (Working Notes). (2015)
6. 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)
7. 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)
8. 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)
9. 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)
10. Goëau, H., Bonnet, P., Joly, A.: LifeCLEF plant identification task 2015. In: CLEF (Working Notes). (2015)
11. Scheirer, W.J., Jain, L.P., Boulton, T.E.: Probability models for open set recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **36**(11) (2014) 2317–2324
12. Bendale, A., Boulton, T.: Towards open world recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. (2015) 1893–1902
13. 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)
14. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. *Computing Research Repository (CoRR)* (2014) arXiv: 1409.1556.
15. Joly, A., Goëau, H., Spampinato, C., Bonnet, P., Vellinga, W.P., Planqué, R., Rauber, A., Palazzo, S., Fisher, B., Müller, H.: LifeCLEF 2015: multimedia life species identification challenges. In: *Experimental IR Meets Multilinguality, Multimodality, and Interaction*. (2015) 462–483
16. Goëau, H., Bonnet, P., Joly, A.: Plant identification in an open-world (lifeclef 2016). In: *CLEF working notes 2016*. (2016)
17. 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