Plant Identification System based on a Convolutional Neural Network for the LifeClef 2016 Plant Classification Task

Sue Han Lee*, Yang Loong Chang*, Chee Seng Chan*, and Paolo Remagnino‡

*Centre of Image & Signal Processing, Fac. Comp. Sci. & Info. Tech.,
University of Malaya, Malaysia
‡Computer Science Dept., Kingston University, United Kingdom
{leesuehan,yangloong}@siswa.um.edu.my
cs.chan@um.edu.my
p.remagnino@kingston.ac.uk

Abstract. In this paper, we describe the architecture of our plant classification system for the LifeClef 2016 challenge [14]. The objective of the task is to identify 1000 species of images of plants corresponding to 7 different plant organs, as well as automatically detecting invasive species from unknown classes. To address the challenge [10], we proposed a plant classification system that uses a convolutional neural network (CNN).

Keywords: Plant classification, deep learning, convolutional neural network

1 Introduction

Plant classification has received particular attention in the computer vision field due to its important implications in agriculture automation and also for environmental conservation. For instance, botanical knowledge of plants is essential to improve agricultural development. Researchers in computer vision [19, 17, 16, 2, 6, 12, 20] have used variations of leaf characteristics as a comparative tool to classify plant. The reason is because leaf characters have been used extensively in traditional text-based taxonomic; they have been keys for plant identification since the early days of botanical science [5, 4]. Although the structural features of a leaf play an important role in the plant identification task, for certain plants, such as deciduous plants or semi-evergreen plants, leaves are not available over different periods of the years. Moreover, some species are hard to be differentiated using only their leaf organ as leaves in nature might have very similar shape and colour [20, 9]. Therefore, botanists usually extend their observation to more than one organ such as stems, flowers, branches or fruits. Since 2013, the LifeClef challenge [11] has provided the first multi-organ plant dataset that not only covers leaf-based images but different organs of given individual plants. Such images of plants were collected in an unconstrained environment, at different periods of time during the year, by different users. The objective of the plant identification task [10] in the LifeClef 2016 [14] challenge is to identify 1000 species of images of plants corresponding to 7 different organs, as well as automatically detecting invasive species from unknown classes.

Inspired by the deep learning breakthrough in image classification, more researchers have started to use deep learning models such as the CNN, to learn a robust plant image representation [1, 21, 3, 8, 24]. In this work, we employ a CNN model to build a plant classification system. We re-purpose the current state-of-the-art VGG net [22] to incorporate species and organ features and solve the multi-organ plant classification problem.

The rest of the paper is organized as follows. In Section 2, we present the methodology of our proposed architecture. Section 3 illustrates its training scheme. Section 4 shows the experiments and results for both the validation and testing set. Lastly, Section 5 presents conclusions and future work.

2 Method Description

The CNN model was initially designed to process multiple arrays of data such as colour (RGB) images, signal or sequences as well as video. Due to the availability of large scale image datasets, such as ImageNet [7], and, followed by the advancement of technology, such as Graphic Processing Units (GPUs), CNN is currently a commodity in the computer vision field. The VGG net [22] offers currently the best state-of-the-art result for image classification.

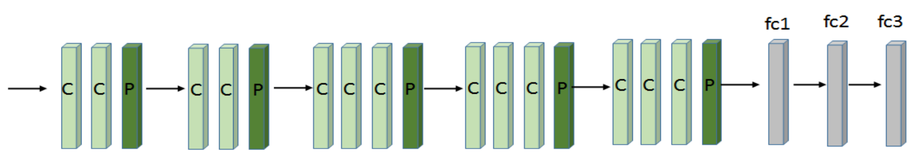


Fig. 1: VGG-net 16 layers architecture

Fig. 1 depicts the configuration of the proposed 16 layers VGG net. In this work, we initialize our model architecture based on the VGG net and modify its higher level convolutional layer to learn the combined species and organ features. Fig. 2 shows our proposed architecture. We do not handcraft any feature descriptor for the fusion features, but introduce convolution layers to learn the filters themselves. We could view these filters as the learned feature descriptors encoding the distinctive fusion structures.

Our architecture mainly comprises four components: (i) shared layers, (ii) organ layers, (iii) species layers, and (iv) fusion layers that handle those combinations of both species and organ features. We introduce shared layers for both species and organ components. The reasons are threefold. First, [25, 23] demonstrated that preceding layers in deep networks response to low-level features such as corners and edges. Since both higher level species or organ components

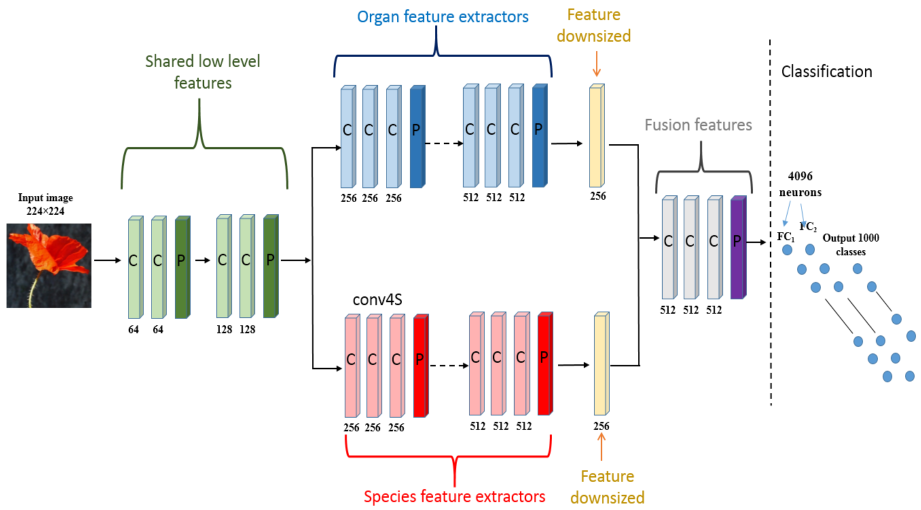
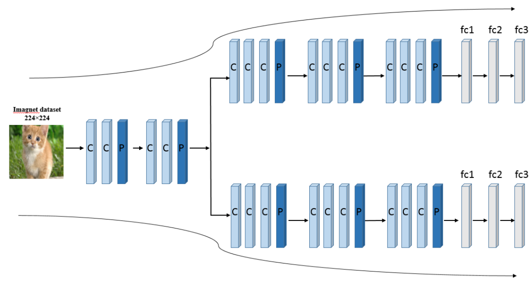


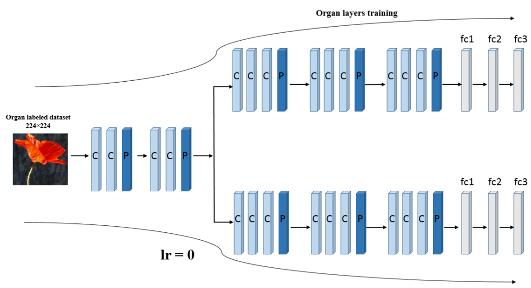
Fig. 2: Organ-species high level fusion architecture

require low-level features to build higher level features, we introduce shared preceding layer for both components. Second, according to [23], the shared layers may reduce both floating point operations and the memory footprint of the network execution, which are of importance for real world application. Lastly, using shared layers helps to reduce the number of training parameters, which is beneficial to the architecture performance.

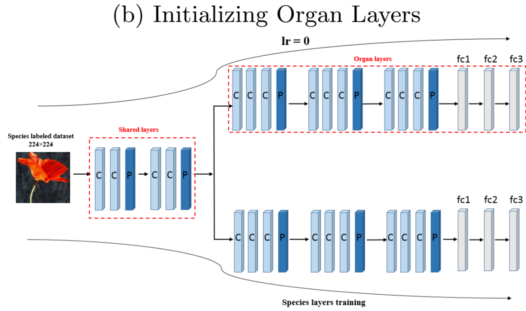
To incorporate both organ and species features, we firstly train the organ layers CNN based on organ classes. Then, we keep the shared and the organ layers unaltered, in order to reuse it to train the species layers. After we trained the species layers, we cascade both of them to learn the fusion features. Before cascading both features, a feature downsizing convolution layer is added in each layer to reduce the feature maps dimension and produce compact based features. This step is essential to reduce the number of training parameters, to compensate the overfitting issue. Last but not least, we train three fully connected layers as the classifier to classify input images to its corresponding species classes. To embed four components into one pipeline and jointly trained end-to-end, we employ the multiple steps training as outlined in Sec. 3.



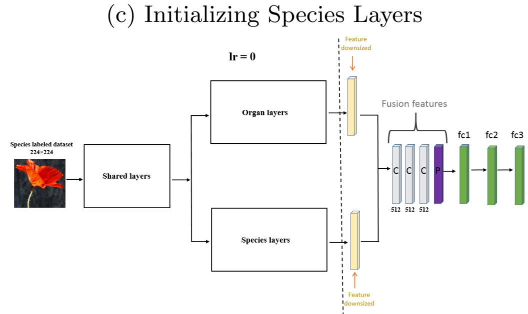
(a) Pre-Training Two-Path CNN



(b) Initializing Organ Layers



(c) Initializing Species Layers



(d) Initializing Fusion Layers

Fig. 3: Multiple steps training scheme for Organ-species high level fusion

3 Training

The algorithm below shows the training procedure of the proposed high level fusion architecture:

Algorithm of our proposed high-level fusion architecture:

---

procedure High-level fusion architecture training
- Step 1: Pre-Training Two-Path CNN
- Step 2: Initializing Organ Layers
- Step 3: Initializing Species Layers
- Step 4: Initializing Fusion Layers
- Step 5: Finetuning Organ-species high level fusion

Step 1: Pre-Training Two-Path CNN We initially design a two-path CNN for the purpose of training two different components (species and organ) as shown in Fig. 3a. Each path of the CNN configuration is similar to the VGG-net 16 layer architecture [22], except that each of them share the preceeding layers.

Step 2: Initializing Organ Layers After we have the two-path CNN pre-trained with Imagenet dataset, we re-purpose one of the path to train an organ layer as shown in Fig. 3b. We perform fine-tuning with seven organ labels: branch, entire, flower, fruit, leaf, leafscan and stem. These organ labels are annotated together with species in PlantClef Dataset. We finetune the VGG-16 network by replacing the final fully connected layer with a total of seven neurons corresponding to seven classes.

Step 3: Initializing Species Layers After we obtained the organ layers, we train the species layers based on the species labeled dataset as shown in Fig. 3c. As mentioned in Sec. 2, we allow both species and organ layers to share the common proceeding layers. Hence, to make the sharing of the filters possible, we keep the first two convolutional layers' weights to be consistent by setting the learning rate to zero during the species layers training. In addition to that, we set the organ layers learning rate to zero in order to avoid having their filters altered throughout the species layers training.

Step 4: Initializing Fusion Layers After having both organ and species components fine-tuned on the two-path CNN, we first migrate its convolutional layers to a new architecture as shown in Fig 3d. Then, we add a convolution layer to each component to reduce the feature map dimension and follow by another two stacks of convolutional layers to learn the fusion features. Lastly, we assign three fully connected layers for species classification. When training, we set the predefined layers' learning rate to zero and only train the newly assigned convolutional layers with species labeled dataset. Typically, we could see this third step as the fusion feature learning stage.

Step 5: Finetuning Organ-species high level fusion Finally, we finetune the whole architecture end-to-end using the same learning rate.

4 Experiments and Results

We evaluate the architectures on the PlantClef2015 dataset [15]. The models are trained using the *Caffe* [13] software. For the parameter setting in training, we employ fixed learning policy. We set the learning rate to 0.01, and then decrease it by a factor of 10, when the validation set accuracy stops improving. The momentum is set to 0.9 and the weight decay to 0.0001. The networks are trained with back-propagation, using stochastic gradient descent [18]. We run the experiments using multiple GPUs on two NVIDIA TitanX graphics cards.

4.1 Data augmentation

The images used for training capture living plants, where every object in the image can be captured at different size. Multi-scale training is therefore proposed. We isotropically rescale the training images into three different sizes: 256, 385 and 512, then randomly crop 224 *224 pixels from the rescaled images to feed into the network for training. Thus, the crop from the larger scaled images will correspond to the small part of the image or particularly subpart of the organ that may be an important feature for recognition. Besides that, we also increase the data size by mirroring the input image during training. Finally, we have a new set of training images that contains 272892 images and a validation set of 66711 images.

4.2 Experimental results on the validation set

For the evaluation of our validation set, we directly employ the softmax output from the model, i.e. we assign each test image the label with maximum softmax output from the classifiers and measure the numbers of correctly assigned labels over all the testing images.

Contribution of Imagenet pretraining. In this section, we evaluate the contribution of the Imagenet dataset [7] pretraining to the plant classification task. We perform transfer learning experiments on the PlantClef2015 dataset. We re-purpose a VGG-16 net by performing fine-tuning on the top fully connected layer. We compare it to the model trained directly from scratch using the PlantClef2015 dataset without any pretraining. The results of our analysis show that the VGG-16 net pretrained using Imagenet data improves by 9.5%, the original 61.7 % score. Hence, pretraining from a larger diversity dataset like Imagenet [7] is clearly beneficial, as it helps to improve the generalization accuracy of the model.

Contribution of data augmentation Table 1 demonstrates the results of data augmentation. We can observe that the VGG-16 net gains 14.8% while the proposed high level fusion gains 14.5%. Hence, it can be deduced that data augmentation is important for the plant classification task, especially when training a large CNN with limited amount of data. Indeed, this enables models to expose to larger amount of data with higher diversity. However, the proposed high level fusion method has lower classification accuracy compared to the finetuned VGG-16 net. This might be because the VGG net uses species features only, performing better than the fusion of features, robust enough to represent plant images.

Table 1: Performance comparison using augmented and non-augmented dataset

Method	Non-Augmented (%)	Augmented (%)
Finetuned VGG-16 top layer	56.4	71.2
High level fusion (proposed)	**54.4**	**68.9**

4.3 Experimental Results on Test set

We have submitted four runs using the LifeClef 2016 - multi-organ plant dataset. The characteristics of each runs are stated as below:

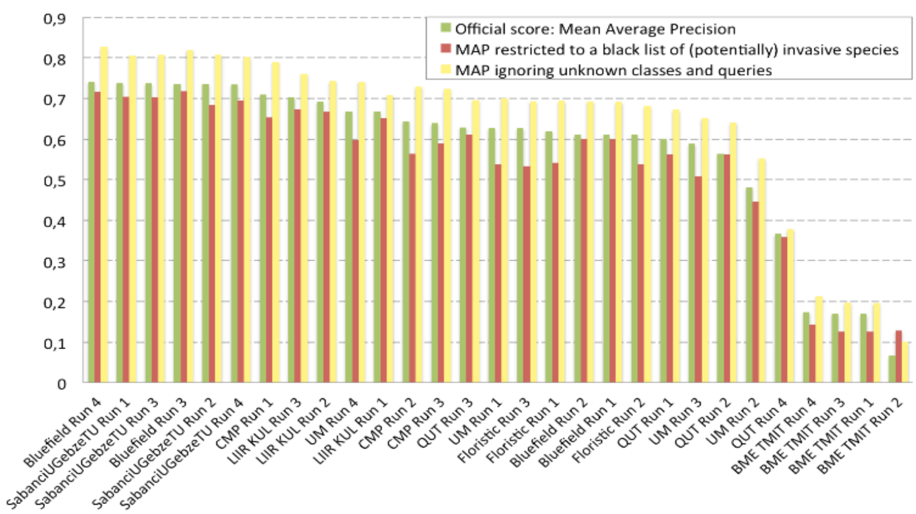


Fig. 4: Results of the LifeClef2016 multi-organ plant classification task

- UM RUN 1: pretrain VGG-16 net with Imagenet 2012, then finetune the top most layer with augmented PlantClef2015 training dataset.
- UM Run 2: train VGG-16 net from scratch with augmented PlantClef2015 training dataset.

– UM Run 3: train the proposed high level fusion architecture with augmented PlantClef2015 training dataset
– UM Run 4: finetune UM Run 1 architecture with validation set.

Fig. 4 shows the results of the mean average precision scores with the evaluation of the robustness of the system handling unseen categories. We observe that Run 4 is the best among the submitted Runs. The reason is because it was trained with a larger training dataset compared to other runs. In addition, Run 1 is better than Run 3 which is consistent with the results obtained in the validation set. Last but not least, Run 2 shows the lowest result among the submitted runs. This again shows the importance of the pre-training model with a large-scale dataset. Overall, all submitted runs obtain average results in the LifeClef2016 multi-organ plant classification task.

5 Conclusions and Future work

This paper proposed using the CNN model to incorporate species and organ features for the plant classification task. We described the methodology of our architecture and analyzed the results based on both validation and testing set. The results of the proposed high level fusion architecture is promising but still limited compared to the VGG net. In future, we will focus on exploring the CNN model for the plant classification task including its implication in the open-set recognition task.

References

1. 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 2015. vol. 1391 (2015)
2. Charters, J., Wang, Z., Chi, Z., Tsoi, A.C., Feng, D.D.: Eagle: A novel descriptor for identifying plant species using leaf lamina vascular features. In: Multimedia and Expo Workshops, 2014 IEEE International Conference on. pp. 1–6. IEEE (2014)
3. Choi, S.: Plant identification with deep convolutional neural network: Snumedinfo at lifeclef plant identification task 2015. In: Working notes of CLEF 2015 conference (2015)
4. Clarke, J., Barman, S., Remagnino, P., Bailey, K., Kirkup, D., Mayo, S., Wilkin, P.: Venation pattern analysis of leaf images. In: Advances in Visual Computing, pp. 427–436. Springer (2006)
5. Cope, J.S., Corney, D., Clark, J.Y., Remagnino, P., Wilkin, P.: Plant species identification using digital morphometrics: A review. Expert Systems with Applications 39(8), 7562–7573 (2012)
6. Cope, J.S., Remagnino, P., Barman, S., Wilkin, P.: The extraction of venation from leaf images by evolved vein classifiers and ant colony algorithms. In: Advanced Concepts for Intelligent Vision Systems. pp. 135–144. Springer (2010)
7. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: IEEE Conference on Computer Vision and Pattern Recognition, 2009. pp. 248–255 (2009)

8. Ge, Z., Mccool, C., Corke, P.: Content specific feature learning for fine-grained plant classification. In: Working notes of CLEF 2015 conference (2015)
9. Goëau, H., Bonnet, P., Barbe, J., Bakic, V., Joly, A., Molino, J.F., Barthelemy, D., Boujemaa, N.: Multi-organ plant identification. In: Proc. of the 1st ACM international workshop on Multimedia analysis for ecological data. pp. 41–44 (2012)
10. Goëau, H., Bonnet, P., Joly, A.: Plant identification in an open-world (lifeclef 2016). In: CLEF working notes 2016 (2016)
11. Goëau, H., Bonnet, P., Joly, A., Bakic, V., Barthélémy, D., Boujemaa, N., Molino, J.F.: The imageclef 2013 plant identification task. In: CLEF (2013)
12. Hall, D., McCool, C., Dayoub, F., Sunderhauf, N., Upcroft, B.: Evaluation of features for leaf classification in challenging conditions. In: Applications of Computer Vision (WACV), 2015 IEEE Winter Conference on. pp. 797–804 (2015)
13. Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T.: Caffe: Convolutional architecture for fast feature embedding. In: Proc. of the ACM International Conference on Multimedia. pp. 675–678 (2014)
14. Joly, A., Goëau, H., Glotin, H., Spampinato, C., Bonnet, P., Vellinga, W.P., Champ, J., Planqué, R., Palazzo, S., Müller, H.: Lifeclef 2016: multimedia life species identification challenges
15. Joly, A., Goëau, H., Glotin, H., Spampinato, C., Bonnet, P., Vellinga, W.P., Planqué, R., Rauber, A., Palazzo, S., Fisher, B., et al.: Lifeclef 2015: multimedia life species identification challenges. In: Experimental IR Meets Multilinguality, Multimodality, and Interaction, pp. 462–483. Springer (2015)
16. Kadir, A., Nugroho, L.E., Susanto, A., Santosa, P.I.: Leaf classification using shape, color, and texture features. International Journal of Computer Trends and Technology 1(3), 306–311 (July to August 2011)
17. Kalyoncu, C., Toygar, Ö.: Geometric leaf classification. Computer Vision and Image Understanding 133, 102–109 (2015)
18. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems. pp. 1097–1105 (2012)
19. Kumar, N., Belhumeur, P.N., Biswas, A., Jacobs, D.W., Kress, W.J., Lopez, I.C., Soares, J.V.: Leafsnap: A computer vision system for automatic plant species identification. In: Computer Vision-ECCV 2012, pp. 502–516. Springer (2012)
20. Lee, S.H., Chan, C.S., Wilkin, P., Remagnino, P.: Deep-plant: Plant identification with convolutional neural networks. In: IEEE International Conference on Image Processing 2015. pp. 452–456 (2015)
21. Reyes, A.K., Caicedo, J.C., Camargo, J.E.: Fine-tuning deep convolutional networks for plant recognition. In: Working notes of CLEF 2015 conference (2015)
22. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. CoRR, abs/1409.1556 (2014)
23. Yan, Z., Zhang, H., Piramuthu, R., Jagadeesh, V., DeCoste, D., Di, W., Yu, Y.: Hd-cnn: Hierarchical deep convolutional neural networks for large scale visual recognition. In: Proc. of the IEEE International Conference on Computer Vision. pp. 2740–2748 (2015)
24. Yanikoglu, B., Tolga, Y., Tirkaz, C., FuenCaglartes, E.: Sabanci-okan system at lifeclef 2014 plant identification competition. In: Working notes of CLEF 2014 conference (2014)
25. Zeiler, M.D., Fergus, R.: Visualizing and understanding convolutional networks. In: Computer vision–ECCV 2014, pp. 818–833. Springer (2014)