Herbarium 2021
Half-Earth Challenge – FGVC8

Ricardo B. Sousa
Faculdade de Engenharia da Universidade do Porto
R. Dr. Roberto Frias, 4200-465 Porto, Porto, Portugal
up201503004@edu.fe.up.pt

Abstract

Work still in progress (abstract will be elaborated only for the final report)...

1. Introduction

In the world, there are approximately 3000 herbaria that represent massive repositories of plant diversity data. All specimens dating back hundreds of years include their reproductive state, collection dates and locations, and the name of the person who collected the specimen. This data represents not only the plant diversity but also snapshots of plant diversity through time [1].

The Herbarium 2021: Half-Earth Challenge [1] intends to identify vascular plant specimens. It is hosted as part of the Eight Workshop on Fine-Grained Visual Categorization (FGVC8) and sponsored by the New York Botanical Garden (NYBG). The data is provided by several herbariums around the world including more than 2.5M images representing approximately 65,000 species, making the Herbarium 2021 [1] competition an image classification problem.

Deep Convolutional Neural Networks (DCNN) have become state-of-the-art for image classification, given that DCNN exploits multiple layers of nonlinear information processing for feature extraction and later classification [4]. DCNN-based networks such as AlexNet [3], VGG16 [5], GoogLeNet [6], ResNet [2], and EfficientNet [7] were tested on the ImageNet dataset. The accuracy results demonstrated that indeed using deep networks allow the improvement in image classification problems.

This work focus on studying the different variants of ResNet [2] (for now, the ResNet-18 and ResNet-34; later, it will be used the ResNet-50) for the large scale classification problem Herbarium 2021 [1]. Future steps are the use of data augmentation to prevent the overfitting of the model to the training data and comparison with the EfficientNet [7]. The later will only be performed if there is time to implement it and train the network.

2. Related Work

Image classification focus on classifying a specific label to an image. Deep Convolutional Neural Networks (DCNN) are usually used to this task. Due to the use of multiple convolutional layers and nonlinear processing, these networks are capable of extracting high-level features to correspond a specific label to an image [4]. AlexNet [3] (Krizhevsky) was the pioneer in optimizing the GPU use for training a CNN winning the ImageNet 2012 competition with a 84.6% top-5 and a 63.3% top-1 accuracy. Since then, deeper CNN were used to achieve higher accuracy results for the image classification problem. VGG16 [5] (Simonyan and Zisserman) studied the increase of depth to 16–19 weight layers for large-scale image recognition (91.9% top-5 and 74.4% top-1 accuracy in ImageNet). GoogLeNet [6] (Szegedy *et al.*) improved the computing resources inside the network while using a 22 layers deep network (2014 ImageNet winner with 74.8% top-1 accuracy). ResNet [2] (He *et al.*) introduced identity shortcut connections to skip one or more layers to prevent degrading the network's performance when added more layers (hundreds or even thousands). ResNet-50 and ResNet-152 achieved a 78.25%/93.95% and a 78.57%/93.29% top-1/5 accuracy in ImageNet, respectively. More recently, Tan and Le proposed in 2019 the EfficientNet [7] that focused on compound scaling to achieve better accuracy and efficiency (EfficientNet-B7 achieved a 84.4%/97.1% top-1/5 accuracy), while using less parameters than other networks.

3. Preliminary Experiments

Given that the Herbarium [1] competion is a plant classification problem, it is require to train from scratch the models (most of them were trained using the ImageNet dataset). This requirement influences the risks of development (using models already available in standard PyTorch or TensorFlow packages versus importing models from GitHub) because Kaggle only offers a weekly GPU quota of 38 hours and a maximum of 9 hours of continuous use (e.g., training 1 epoch with a ResNet-18 model requires at least 6 hours and 30 minutes using the Herbarium 2021 [1] dataset). While my initial objective was to use Colab for this competition (because Colab offers higher quotas for using the GPU), I was unable to import the data (dataset occupies approximately 150GB versus the approximately 120GB of disk space in Colab)[1]. So, I focused on using the ResNet [2] model and comparing the accuracy of different variants (available in `torchvision.models`); specifically, the ResNet-18, ResNet-34 and ResNet-50. The code is available at `https://www.kaggle.com/ricardobarbosasousa/herbarium-2021-rbs-resnet`.

3.1. Dataset Herbarium 2021

The dataset provided for the competition has 2,500,779 images with 64,500 species of vascular plants. The data has been approximately split in 80%/20% for training/test (2,257,759/243,020 images). Overall, the dataset contains a minimum of 3 images per species, with some species represented by more than 100 images. Although the training set contains species with hundreds of examples, the test set has a maximum of 10 examples per species. Given that the submission only provides the accuracy results, the accuracy measure is the evaluation metric considered in this work.

3.2. ResNet without data augmentation

The models used for the classification problem were the ResNet-18 and the ResNet-34. These models were trained with a virtual machine hosted by Kaggle with a Intel Xeon CPU @ 2.00GHz and a Nvidia Tesla P100-PCIE-16GB. The optimizer chosen was the Adam optimization algorithm, and the criteria was the cross entropy loss. The learning rate was set to 4×10^{-4}. In terms of batch size, the training of ResNet-18 and ResNet-34 models used 512 and 384 images, respectively. It was not possible to train ResNet-34 with a batch size higher than 384 due to graphics memories constraints. All the training data provided in the Herbarium 2021 [1] dataset (2,257,759 images) were used for the training state, and the images were resized to the maximum size allowed by ResNet (224x224 resolution).

The preliminary results of training the two models, ResNte-18 and ResNet-34, are illustrated in figure 1 and table 1. The figure 1 illustrates the training of the models. Given the time require to train 1 epoch, the evolution of the cost function is shown after each batch of images is pro-

[1] I tried using my UPorto Google drive due to the free available space. However, mounting the drive within the virtual machine at Colab also occupies disk space. So, I ended up to using Kaggle notebooks because the data is alread available in the virtual machine

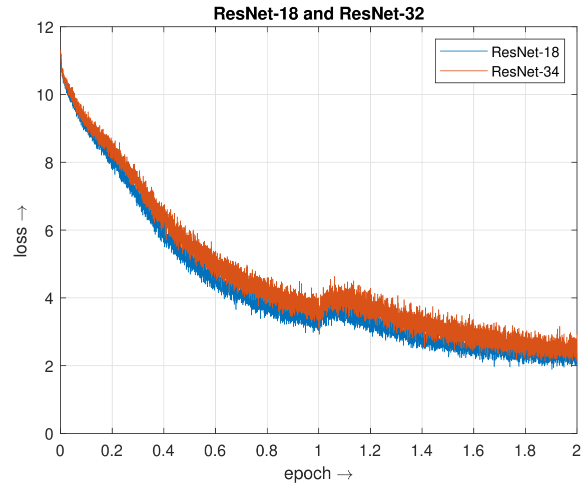

Figure 1. Train of ResNet models

cessed and with only 2 epochs. The models were saved after each epoch because Kaggle has a maximum continous utilization of 9 hours (and the training plus inference requires at least 7 hours and 30 minutes). In table 1, it is shown the numerical results of train and test sets. The results of the train set consider the average and best results after each of the two epochs. As for the test results, it is only possible to obtain the automatically processed accuracy by Kaggle.

Even though figure 1 is not relative to the overall epoch, it seems that both ResNet-18 and ResNet-34 can obtain better results if we train these models one or two more epochs. However, given the drop in accuracy noted from train to validation on approximately 40% (on the second epoch for both models), it seems that both models are overfitting to the training data. Therefore, the future steps is to implement data augmentation to reduce this effect. The augmentation will probably not be focused on color transformations, given that plant species have some dependency on the color.

4. Next steps

The results presented in this intermediate report are a first test of using DCNN for the Herbarium 2021 [1]; specifically, the ResNet [2] model. The existent code already handles the loading the pre-processing of the dataset (dataloaders, labels encoding, etc.), creation of the model, training, and inference on the test set. Given the preliminary results, the next steps will be train a ResNet-50 (evaluate if a deeper model results on better accuracy) and finishing the train on the two other models, the implementation of data augmentation, and, as a final optional step, compare ResNet [2] with the EfficientNet [7] within the scope of the Herbarium 2021 [1] competition. The later step will require further changes in the code because the EfficientNet [7] is not supported by the PyTorch package `torchvision.models`.

		Train				Test
		Avg. loss	Avg. acc.	Best loss	Best acc.	Acc.
Model	Epoch		(%)		(%)	(%)
ResNet-18	1	5.7913	19.11	2.9748	44.92	9.487
	2	2.7901	47.21	1.8959	62.70	22.47
ResNet-32	1	6.1600	15.93	3.2792	41.93	7.479
	2	3.1264	43.69	2.0428	59.64	18.45

Table 1. Preleminary experimental results using ResNet models

References

[1] FGVC8 and CVPR 2021. Herbarium 2021 - Half-Earth Challenge - FGVC8 - Kaggle, 2021. `https://www.kaggle.com/c/herbarium-2021-fgvc8/overview`.

[2] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, June 2016. `https://doi.org/10.1109/CVPR.2016.90`.

[3] Alex Krizhevsky. One weird trick for parallelizing convolutional neural networks. *CoRR*, abs/1404.5997, 2014. `http://arxiv.org/abs/1404.5997`.

[4] Waseem Rawat and Zenghui Wang. Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review. *Neural Computation*, 29(9):2352–2449, 09 2017. `https://doi.org/10.1162/neco_a_00990`.

[5] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In *3rd International Conference on Learning Representations (ICLR) 2015*, San Diego, CA, USA, May 2015. `http://arxiv.org/abs/1409.1556`.

[6] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–9, 2015. `https://doi.org/10.1109/CVPR.2015.7298594`.

[7] Mingxing Tan and Quoc Le. EfficientNet: Rethinking model scaling for convolutional neural networks. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97, pages 6105–6114. PMLR, Jun. 2019. `http://proceedings.mlr.press/v97/tan19a.html`.