The Herbarium 2022: Flora of North America

Alejandro Barreiro, Patrick Indri

Technische Universität Wien
Vienna University of Technology
TU Wien Informatics

Climate change on plants

Nowadays, climate change is already a reality that is being expressed across the globe. The Intergovernmental Panel on Climate Change, (the United Nations body for assessing the science related to climate change) states in its most recent report, the unequivocal increase in average air and ocean temperatures, changes in precipitation patterns, widespread melting of ice and rising sea levels on a global scale. In recent years, the need for increased efforts to combat climate change has become even more apparent, and organisations and administrations around the world have adopted various commitments and measures, which is where the Kaggle competition addressed in this poster comes into play.

There are approximately 400,000 known vascular plant species with an estimated 80,000 still to be discovered. With the threats of climate change, we need new tools to quicken the pace of species discovery. This is more pressing today as a United Nations report indicates that more than one million species are at risk of extinction, and amid this dire prediction is a recent estimate that suggests plants are disappearing more quickly than animals.

A flora in a Neural Network

In botany, Flora is a systematic account of plants of a defined geographical area and provides keys and descriptions of plants for identification. A flora is a tool used to identify plants and can be useful in fields like Biodiversity management.

A neural network, is a computational learning system that uses a network of functions (neurons), to understand and translate a data input into a desired output. They have evolved over the years with different architectures and paradigms, which have advanced the state of the art in multiple domains, helping in different ways to solve many real-life problems.

In this task, we will focus on the image recognition problem, specifically fine-grained image recognition. The Fine-Grained Image Classification task focuses on differentiating between hard-to-distinguish object classes.

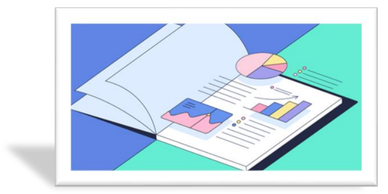
I have been working with a dataset that strives to represent all known vascular plant taxa in North America (only vascular land plants (lycophytes, ferns, gymnosperms, and flowering plants)), using images gathered from 60 different botanical institutions around the world.

The purpose of our model is to take advantage of the potential of AI in botanical applications to quicken the pace of species discovery simplifying the work of taxonomists.

Furthermore, the success of the model will lead to understanding plant diversity on a massive scale and learning how it has changed over time, which is an integral first step to speed the pace of species discovery and save the plants of the world.

Research on the field → Feature Engineering → Model management → Model training

Papers



Notebooks

kaggle

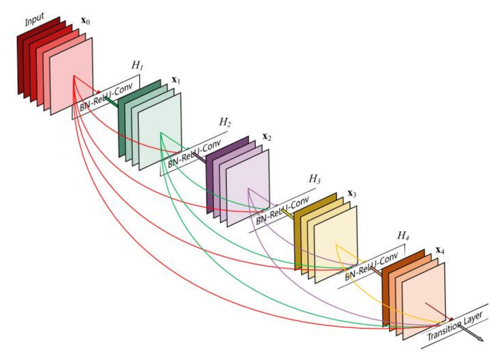
Dataset

Augmentation: Resizing, Cropping, Rotation, Flipping, Normalizing

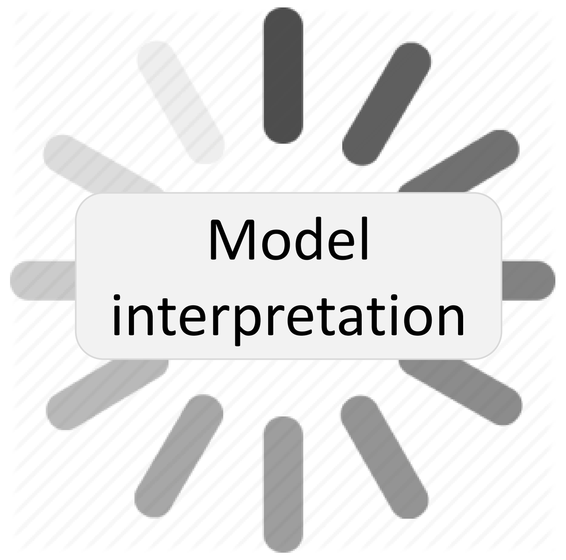
Distilled

BiT: S, M

ResNet: 18, 50, 101



PyTorch



Methodology

The dataset, "The Herbarium 2022: Flora of North America", comprises 1.05 M images of 15,500 vascular plants, which constitute more than 90% of the taxa documented in North America. The number of images per taxon is as few as seven and as many as 100 images. In order to obtain insight on the field, I have been studying some papers on fine-grained image recognition, specifically on plants. This research along with the notebooks and discussions published on Kaggle, have been fundamental to the development of my approach, helping me to educate myself on the deep learning field as well as discovering the state of art architectures on the task.

In the feature engineering section, I have constructed a pipeline with the different configurations that have been found to be successful in "research on the field" for the augmentation of the images of the dataset. Different experiments have been designed in order to adapt the configurations to the task, considering the different types of images available on the dataset and the computational possibilities of the resources provided.

I have selected different models that have shown a successful performance on fine grained recognition. All of them are different versions of the famous architecture based on Residual Networks, ResNet. Following the base implementation available on Pytorch, I kept on working with different layer sizes (18, 50, 101). I must also stress that I am working with the pretrained version of ImageNet, having adapted the classification head for the number of herbarium classes. Delving deeper into the literature, I have found Big Transfer (BiT), which attains state-of-the-art performance at multiple tasks on image recognition. The architecture has three different models (S, M, L), differing on the size of the dataset they have been pretrained. Nevertheless, there is a growing discrepancy in image recognition between big sized models that achieve state-of-the-art performance and models that are affordable in practical applications, therefore, the BiT team have published a compressed model called distilled BiT, thanks to the Teacher-Student architecture.

The model training has been one of the bottlenecks of the competition, the constrains on performance of the computer resources have caused the inability to test some architectures like BiT-L BiT-M, BiT-S and the distilled versions. I have fixed a low batch size with a mediocre number of epochs. The optimizer consists of a Stochastic Gradient Descent with Nesterov momentum on top of a ReduceLROnPlateau scheduler. The loss function used is the torch Cross Entropy Loss.

I have tried different explainability techniques for interpreting the model and its results but unfortunately, the computer resources aren't enough for computing them. Is important to highlight one downside of Kaggle's competitions, it does not provide you the classes of the test set so you can´t focus on the misclassifications of the model.

Results

With a fixed number of 10 epochs, the model that gave better results was a ResNet18 with a 0.56441 of accuracy on Kaggle, managing to predict properly more than half of the species. The other ResNets need more epochs in order to get a better understanding of the images due to their complexity. The only BiT model I was able to test was the BiT-S-101x1 with the predefined data augmentation pipeline of the model, which performed with a poor accuracy (0.31922).

Future work

This problem is characterized, like most fine-grained recognition problems, by the immense depth with which it can be tackled, coupled with the time constraints that characterize CNNs. Therefore, this section could be extremely extensive, but I will stick to a few key points.

- Training deeper models with a larger number of epochs
- Correctly tuning BiT to solve this task.
- Analyze the results of the models by creating a test partition from the train images provided.
- Search for more optimal explanatory techniques that can explain the models obtained.
- Merge different models such as Resnet, Efficient, Swin... to achieve an improvement in accuracy