Capstone 2 Project

SPRINGBOARD DATA SCIENCE CAREER TRACK
FLOWERCLASSIFIER

Kristopher Simino | March 2022

Introduction

Image classifiers can be trained to recognize images of most any object or group of objects. In this project, we will classify species of flowers with the focus on using pretrained convolutional neural networks. The goal is to determine how well a neural network, trained to recognize a thousand different objects (some very different), can be used to extract features, and classify types that are very similar.

Ultimately, some architectures are more effective than others and the distribution of the data has a major impact on the results. The results of this project were less than exceptional with the classifier tending towards the largest classes in the training set.

The details of the model implementation and summary metrics can be found at the following GitHub link.

- https://github.com/krsim/CapstoneTwo.git

Approach

DATA ACQUISITION AND WRANGLING

The data set used for this project was supplied by Kaggle and is called the “tpu-getting-started” dataset. It consists of 104 species of flowers broken into a shuffled training set of 12753 images, a stratified validation set of 3712 images, and an unlabeled test set of 7382 images. The same images were available in four different resolutions and 224 x 224 was used.

The data was well curated and needed no cleaning, however, because the dataset was originally intended for a Kaggle competition, the test set is unlabeled for submission. Therefore, the test set provided was not used, and instead, a new test set was drawn from the training data. This was done before any augmentation of the training set.

The remaining training data was extended by augmenting random images. This included, zooming, cropping, rotating, flipping, and adjusting saturation, brightness, and contrast of the images. The result was a training set of 114,912 images, a validation set of 3712 images, and a test set of 3177 images.

STORYTELLING AND INFERENTIAL STATISTICS

The flower images were mostly closeup and in a natural setting. A small subset can be seen below.

Figure 1. Subset of training data

The primary issue with the potential performance of any classification model was the data's class distribution. As *figure 2* shows, the data was highly skewed. The danger being that the classifier would focus on the features of the largest classes and simply place the rarer species in them. The top five most common species are iris, wild rose, wild geranium, common dandelion, and rose. They make up more than 25% of the total data. The bottom five however makeup less than 1% of the data.

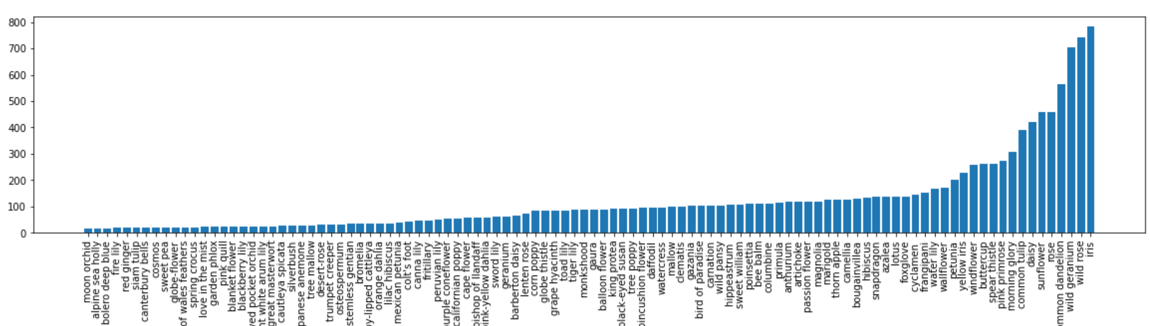
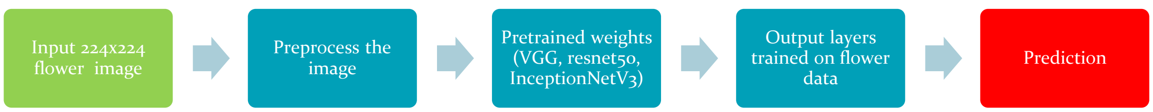


Figure 2. Class distribution

MODELING

Three separate architectures were implemented: VGG, ResNet50, and Inception Net V3. Each a deeper network than the last and all pretrained on the Imagenet dataset. A new head was created for each model so that they could be trained on the flower data and the output layer changed from the 1000 classes of the Imagenet dataset to the 104 classes of the flower dataset. A detailed summary of each architecture can be found in the flowerclassifier notebook in the provided GitHub link.

To fairly compare the three models, the same number epochs (30) were used to train, and the same learning rate schedule was used to fine tune. The results were very different for each model with ResNet50 performing the worst and Inception performing the best.



Findings

The chart below shows the resulting metrics for each model.

	Accuracy	Precision	Recall	F1 Score
InceptionNet V3	76.1%	77.0%	68.7%	71.1%
VGG	63.7%	62.1%	58.5%	58.7%
ResNet50	12.0%	4.0%	2.5%	1.8%

Figure 3. Resulting metrics

It is clear from the table that the InceptionNet V3 has the most promising results. Although the VGG model performed nearly as well. However, the results of the ResNet model do provide some insight into the outcome of all the models. Observe the confusion matrix below for the ResNet Model.

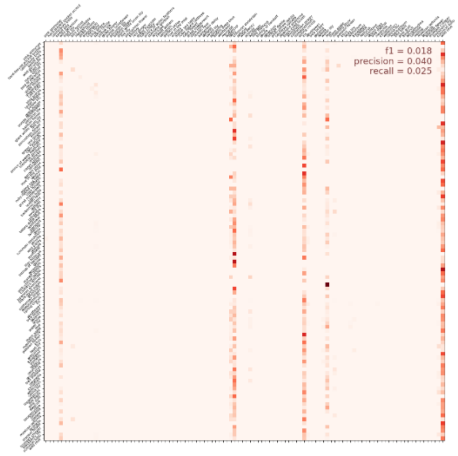


Figure 4 ResNet50 confusion matrix

Unlike the other two models that accurately classified the images more often than not, the ResNet model mostly placed the images into only five classes. Those classes were iris, wild rose, wild geranium, common dandelion, and rose; the classes that happen to be the top five largest in the training set. The highly skewed data had a predictable an impact on the outcome. It is clear on the ResNet model but after closer inspection of the other two, the same misclassifications can be seen, though at a far less degree.

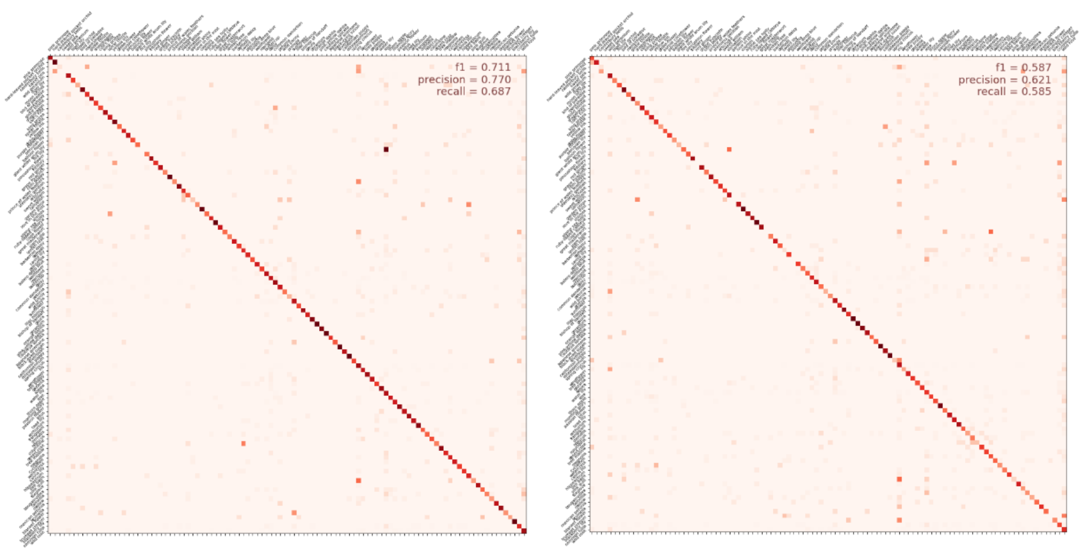


Figure 5 InceptionNet and VGG confusion matrix

The InceptionNet and VGG models have fair scores and can likely be improved greatly with some careful hyperparameter tuning. The ResNet model may not be suitable for this application.

Conclusions and Future Work

Pretrained networks appear to produce good results even on datasets with classes of very similar objects. Which architecture you choose and the distribution of classes in the data have a great impact on the results.

Given more time, the results of the flower classifier could have been improved by:

- a more balanced distribution of classes. Achieved by collecting more data of the rarer species or by weighting the augmentation process in the preprocessing phase to favor the rarer species.
- a careful selection of hyperparameters could further improve results.
- experimenting with different architectures.