**Sam Ehrenstein**

### 26 August 2019

Engineers for Exploration

**Using train\_fc.py**

# INTRODUCTION

## OVERVIEW

Once features have been extracted using extract.py, we can train a classfier. We found that the best approch, by far, is to use a multilayer perceptron (MLP) network. Briefly, this is a small (4-layer) fully-connected neural network that takes a feature vector as input and outputs a classification: whether the image that produced the vector is mangrove or not.

## PREREQUISITES

A computer with a reasonable CPU, or a GPU. Almost all versions of my MLP network trained in under 30 seconds on my laptop’s CPU, an Intel i7-8750H.

If using unlabeled data, a local copy of your input directory to extract.py that you have write access to. If the file names are not the same, the sorting process will not work!

Python 3 with the following modules:

* numpy
* scipy
* tensorflow
* scikit-learn
* tqdm
* PyYaml

# NOTES

I have tried my best to be unambigious, but just to be explicit: All input directories to train\_fc.py must be output from extract.py, unless noted. By design, train\_fc.py does not process raw images.

train\_fc.py can train from multiple directories of labeled output from extract.py. This is in order to facilitate training on multiple sites. In order to do this, we use a YAML config file. The config file should have a single field, called train, that has a list of the paths to each directory of output you wish to use for training. The trained model will be saved in the first directory listed.

## STEPS

1. Get the output of extract.py onto the computer you’re using to run train\_fc.py.
2. Run train\_fc.py. If you have not run the model on this training set before, you must pass the -r flag to (re)train the model. Otherwise, a saved model will be loaded from the first directory in the list in the config file.

It has two modes: validate and sort. Validate is used to validate the model against labeled feature vectors. It is used like this:

$ python train\_vc.py -v –cfg=<config file> --test=<path to directory to validate on>

Validate mode will print out a classification report and confusion matrix for the test. 0 is mangrove, 1 is non-mangrove.

The second mode is sort. Sort requires that the -f flag was used with extract.py for this dataset, and is used to sort a directory of unlabeled images. This is the ONLY time a directory of raw images is going to be passed into train\_fc.py. Given the directory of images and the directory of output from extract.py when those images were processed, train\_fc.py will sort the images into m and nm directories. It is used like this:

$ python train\_fc.py –sort –cfg=<config file> --test=<path to directory of feature vectors> --indir=<path to directory of images> --outdir=<path to directory for sorted output; defaults to indir>

## SPECIFICATIONS

The default model is fully connected, and has a layer of 256 neurons followed by a 50% dropout layer, then a single sigmoid output neuron.

# REFERENCES

Inspired by this paper: <https://www.hindawi.com/journals/js/2018/6257810/>

[https://keras.io/applications/#vgg16](https://keras.io/applications/" \l "vgg16)

<https://towardsdatascience.com/transfer-learning-from-pre-trained-models-f2393f124751>