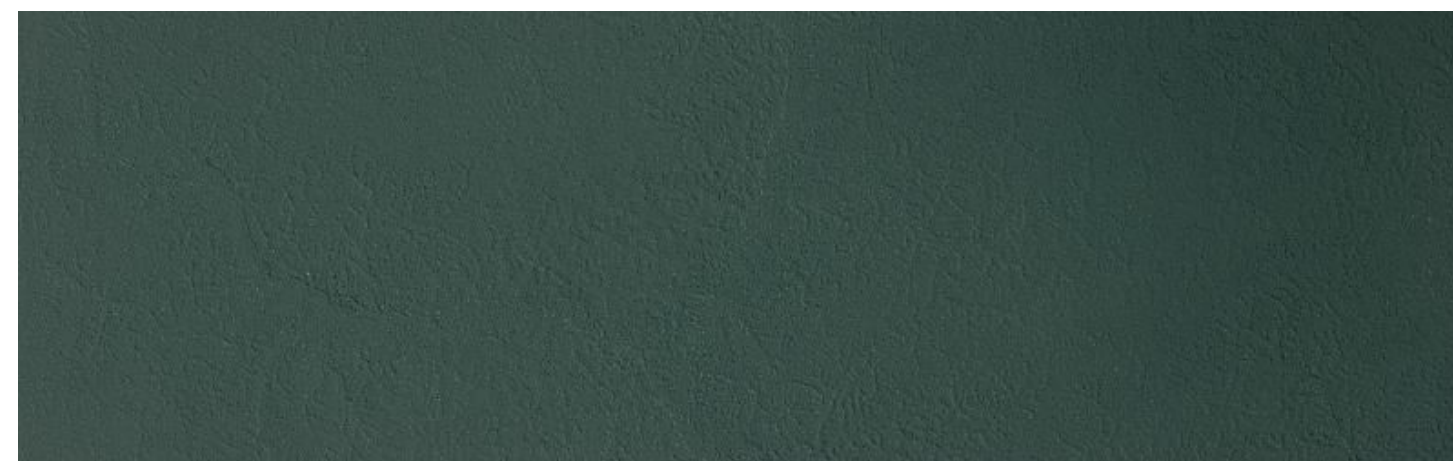


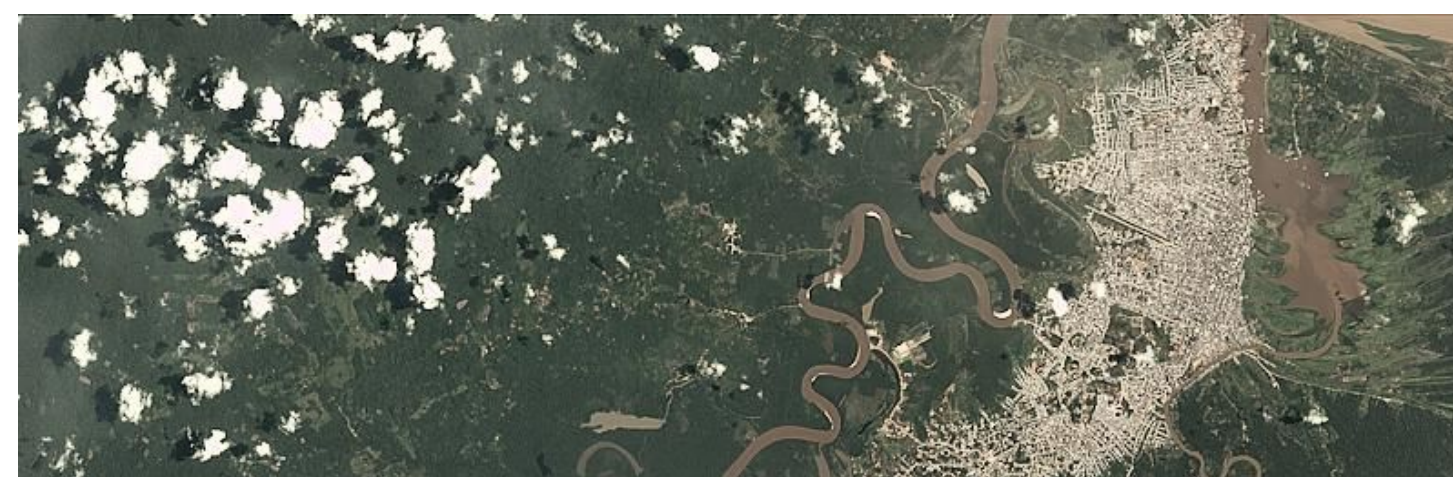
## Intro

The Amazon Rainforest contains roughly 390 billion trees belonging to 16,000 species. The forest contains over half of the earth's rainforests by land area. The Amazon basin contains 5,500,000 square kilometers of rainforest. However, even with its massive size, deforestation is a large problem. Because the Amazon basin contains such a large amount of biodiversity, the losses are severe.

## Data



Primary forest



Cloudy, habitation, roads



Water, primary



Agriculture, roads

# Understanding the Amazon from Space

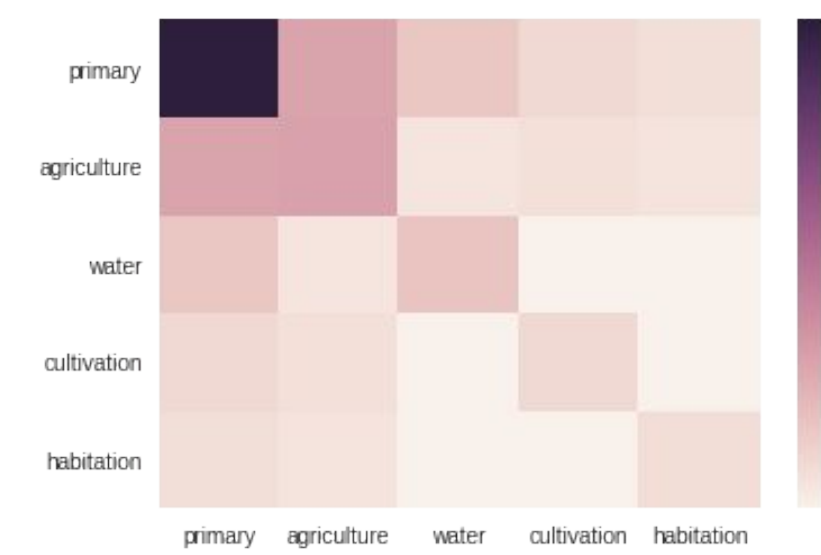
Jeff Pyke

## Methods

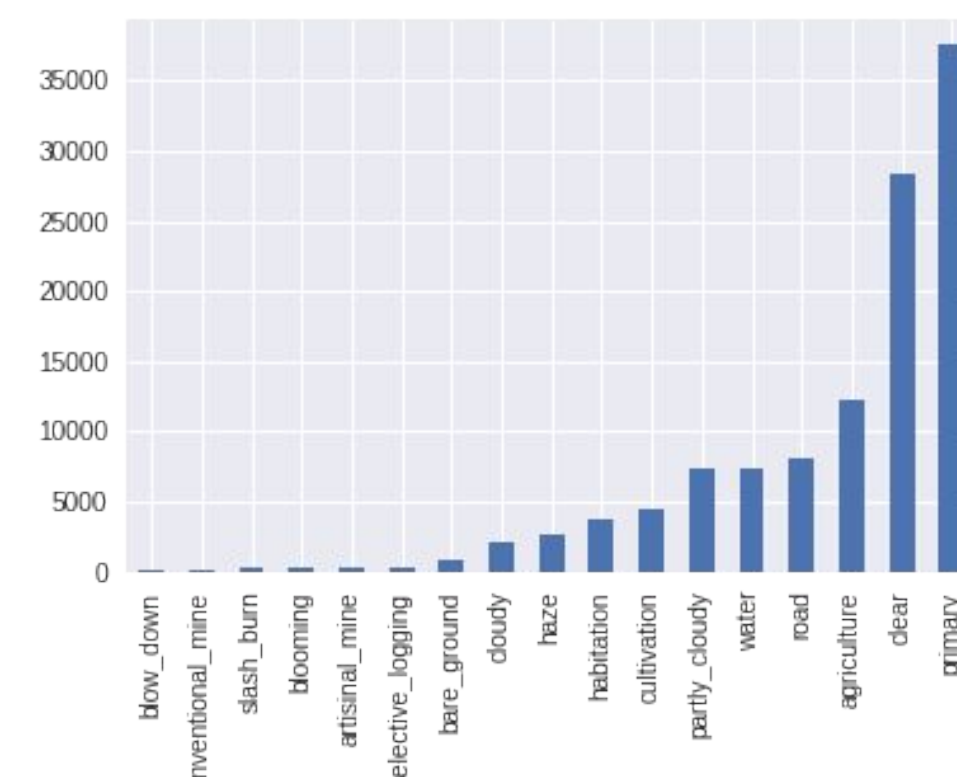
I used the keras framework (wrapper around tensorflow) to run a multi-layered net with various strategies such as conv layers, pooling, dropout, normalization, and fully connected layers. Google Compute Engine was used, along with a GPU, to train the model.

First, some time was spent exploring the data using kaggle provided examples. Heatmap graphs are shown below.

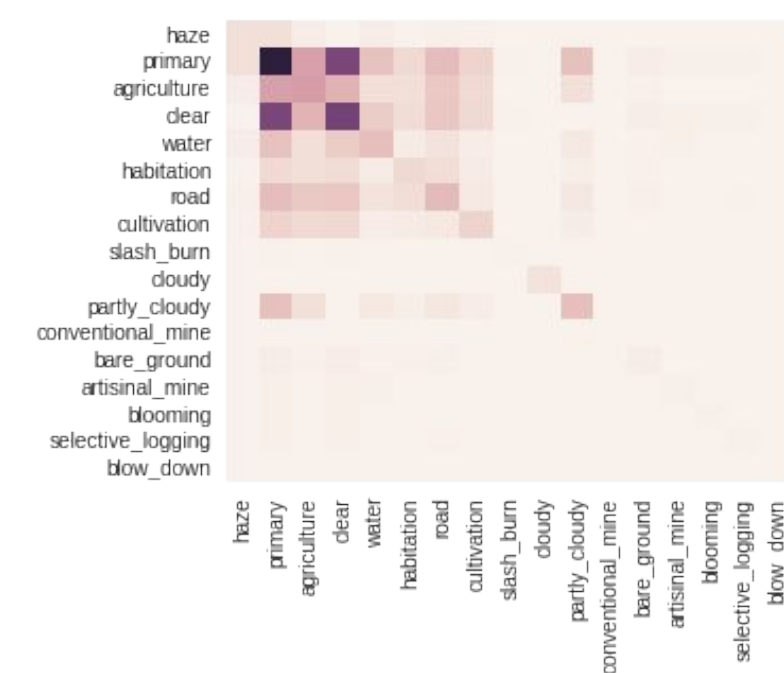
The basic framework for the neural net is based on a starter kernel provided by the kaggle user anokas. The keras framework is used to provide a high level library on top of tensorflow. The initial network is visualized to the right. The NN uses binary cross-entropy loss with the adam optimizer.



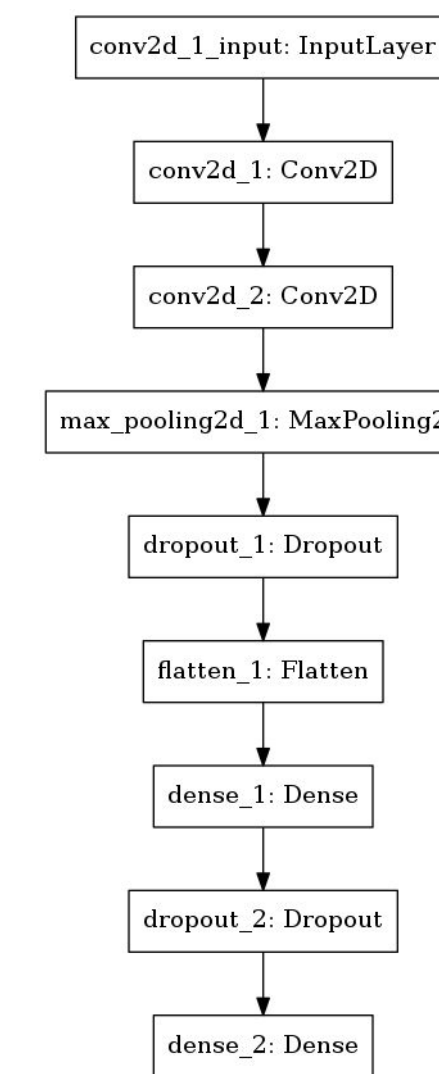
Land co-occurrence matrix



Label frequency



Full co-occurrence matrix



## Results

Accuracy of 93% was achieved on the validation set. A threshold of 0.2 was used for determining whether a label should be applied. Results have not yet been uploaded to Kaggle, because all of the so far shown results have been run on the jpg train/test set which is incomplete. For the final submission, the tif data must be used.

