AMAZON SATELLITE IMAGE CLASSIFICATION

FOR NGOS, NONPROFITS AND GOVERNMENT AGENCIES

I. THISTED FOR SPRINGBOARD CAPSTONE PROJECT

PROJECT OVERVIEW

- The Amazon is the largest surviving rainforest ecosystem in the world and currently home to 10% the world's species[1]. It plays an important role in regulating rainfall cycles in South America[2] and in thermo-regulating the atmosphere above the region
- By some projections, a loss of as little as 20 to 25 percent of original forestland could tip the entire system into an unstoppable transition to a drier, savanna-like ecosystem.[3]
- The WWF estimates that 27 percent of the Amazon biome will be without trees by 2030 if the current rate of deforestation continues.
- Climate change, wildfires, cattle ranching, mining, logging, and farming all contribute to deforestation compounded by ineffective or counterproductive government policies and lack of resources that have left many of these activities unchecked.
- Better data about the location of deforestation and human encroachment on forests can help governments and local stakeholders to respond more quickly and effectively.

DATA

- Satellite imagery with 3 to 5-meter resolution of the Amazon basin was made available on Kaggle
- The images were labeled with crowd sourcing and collected by Planet’s Flock 2 satellites between January 1, 2016 and February 1, 2017. They cover an area of 30,000,000 hectares
- Each image chip can contain one or more labels for associated atmospheric conditions and various classes of land cover and land use.
- There are approximately 40,000 16-bit images in the training set, each one has dimensions 256x256 pixels.
  - Images are available in 2 formats .jpg and tif, with the latter containing near infra-red data in addition to RGB. For the purposes of this project and because of computational power limitations, only the .jpg images will be utilized.
- A .csv file was provided containing the names of the images and its associated labels
  - No missing data

DATA

- 17 possible labels, each image can contain multiple labels
  - **Atmospheric conditions:** haze, clear, cloudy, partly_cloudy
  - **Vegetation and land use**: primary, agriculture, water, road, cultivation, habitation, bare_ground, selective_logging, slash_burn, blooming, blow_down, conventional_mine, artisinal_mine
- Only one atmospheric label is assigned to each image

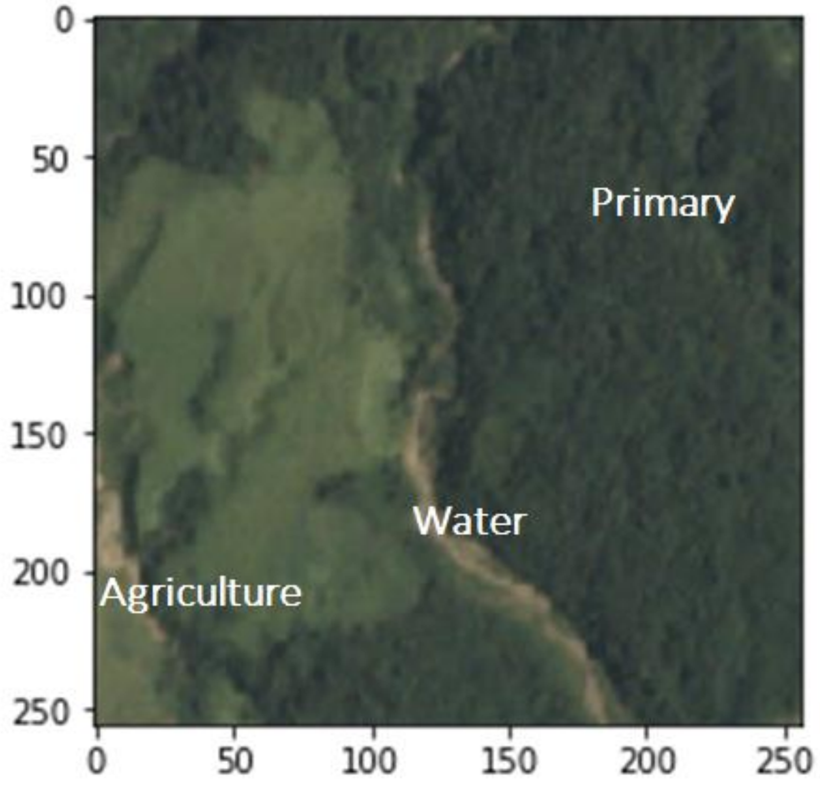
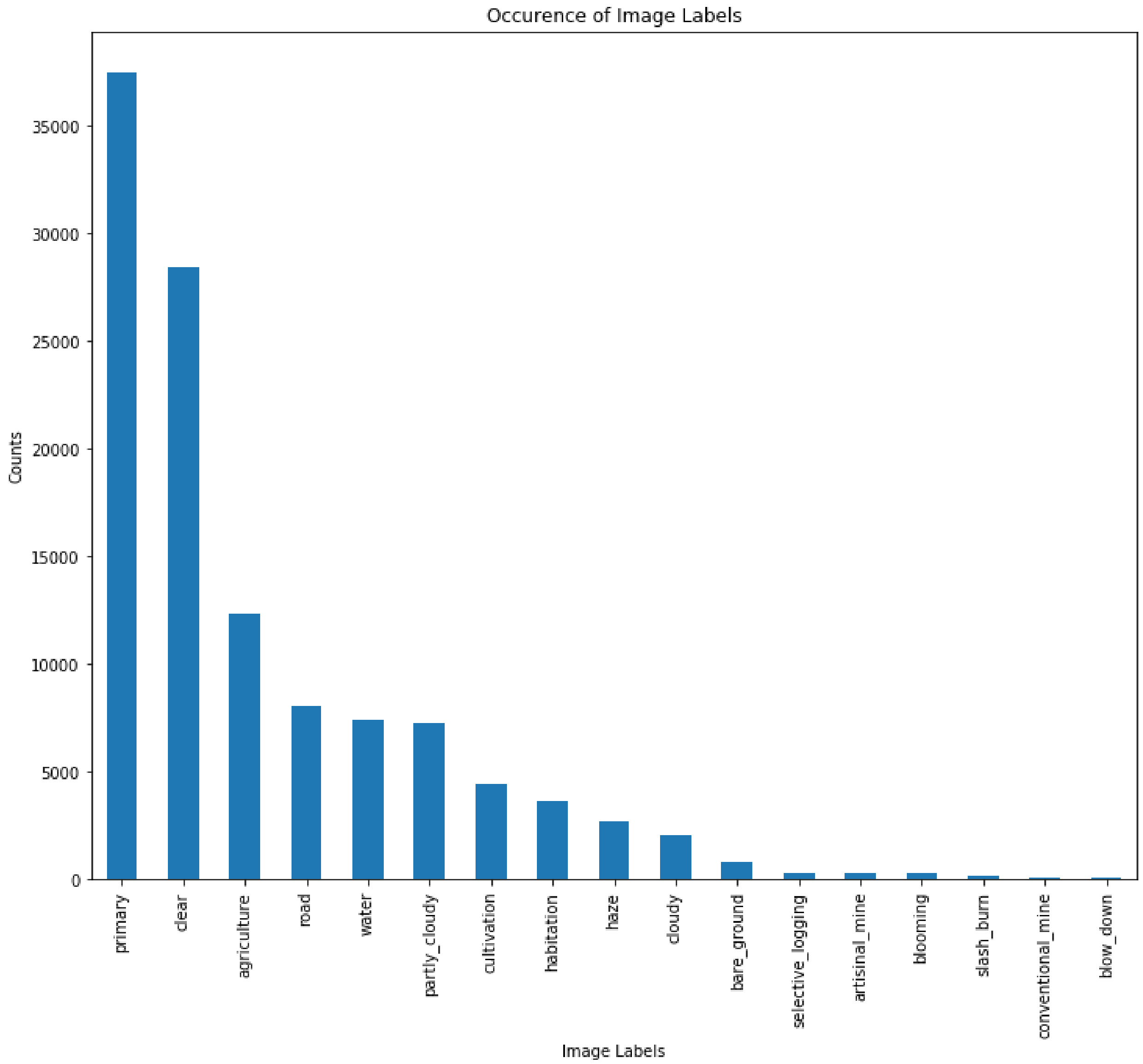


Figure 1 - Sample Image and Labels

EXPLORATORY DATA ANALYSIS

- The predominant cloud coverage in the images is “clear” (28,431), followed by partly cloudy, haze and cloudy.
- The majority of the images (37,513 images) in the dataset are also labeled as “primary”, which represents the primary rainforest and is characterized by dense tree cover.
- The second most commonly occurring land use label is agriculture with 12,315 occurrences.
- Road and water (rivers and lakes) are also commonly occurring and appear in approximately 8,000 images each.
- Artisanal mine, blooming (blooming of flowering trees), selective logging and slash and burn each have fewer than 500 instances
- Blowdown (toppled trees resulting from naturally occurring microbursts) and conventional mines each have fewer than 100 instances.



CO-OCCURRENCE OF LABELS

- There are close to 28,000 images with co-occurrence of "clear" cloud cover and "primary" rainforest.
- "Primary" and "agriculture" represent the second most commonly occurring co-occurring labels with close to 12,000 examples, followed by "agriculture" and "clear"
- We can also confirm here that atmospheric labels have a zero rate of co-occurrence with each another

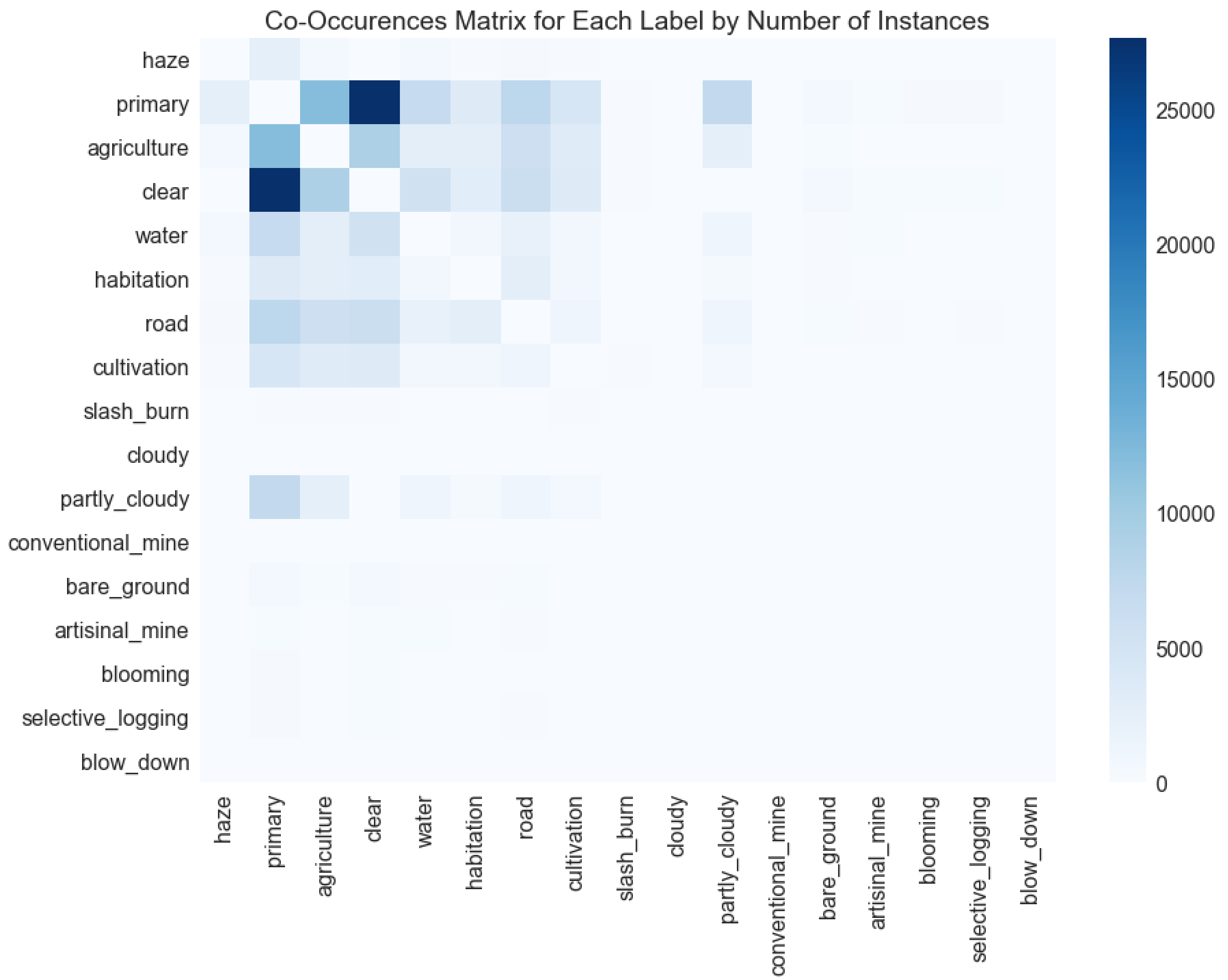
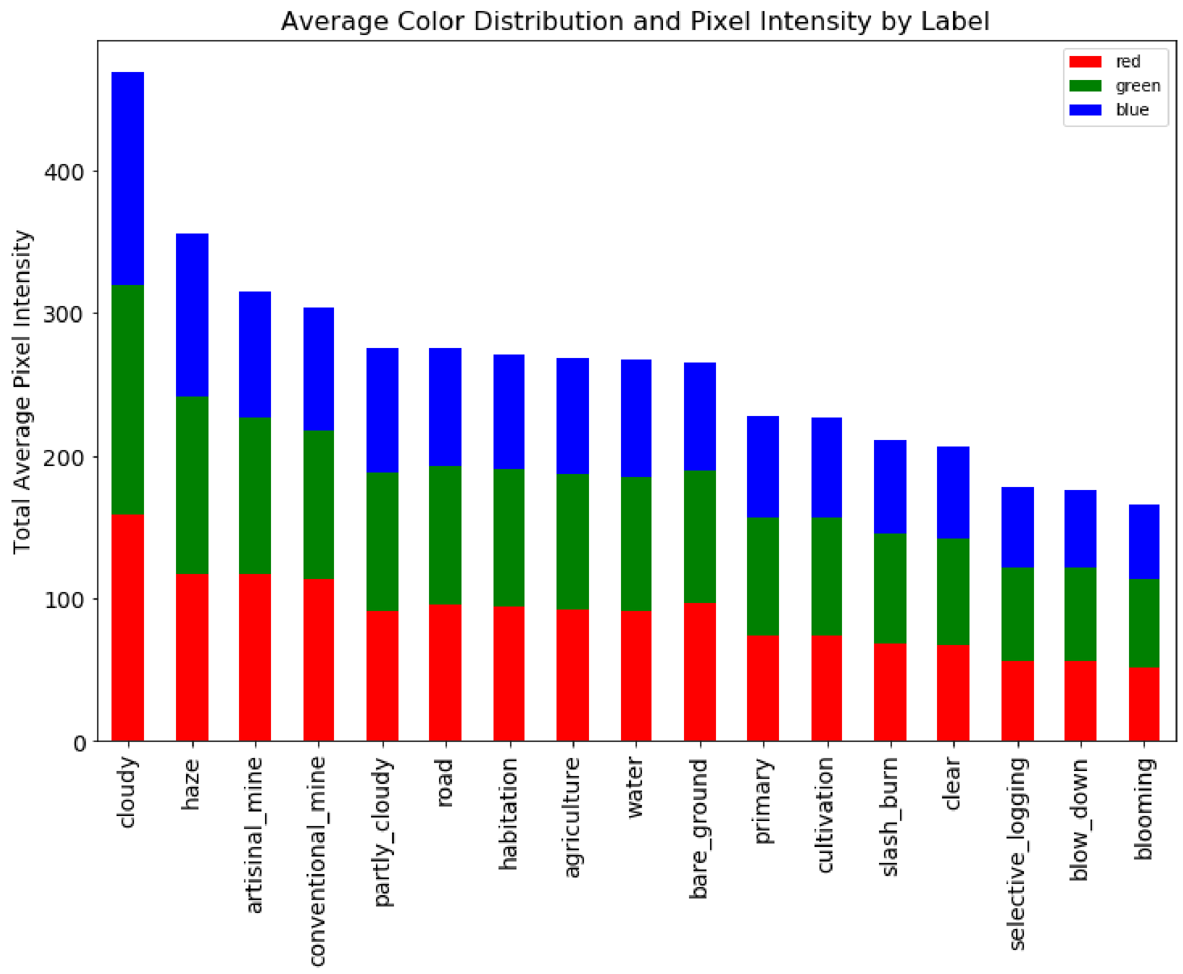


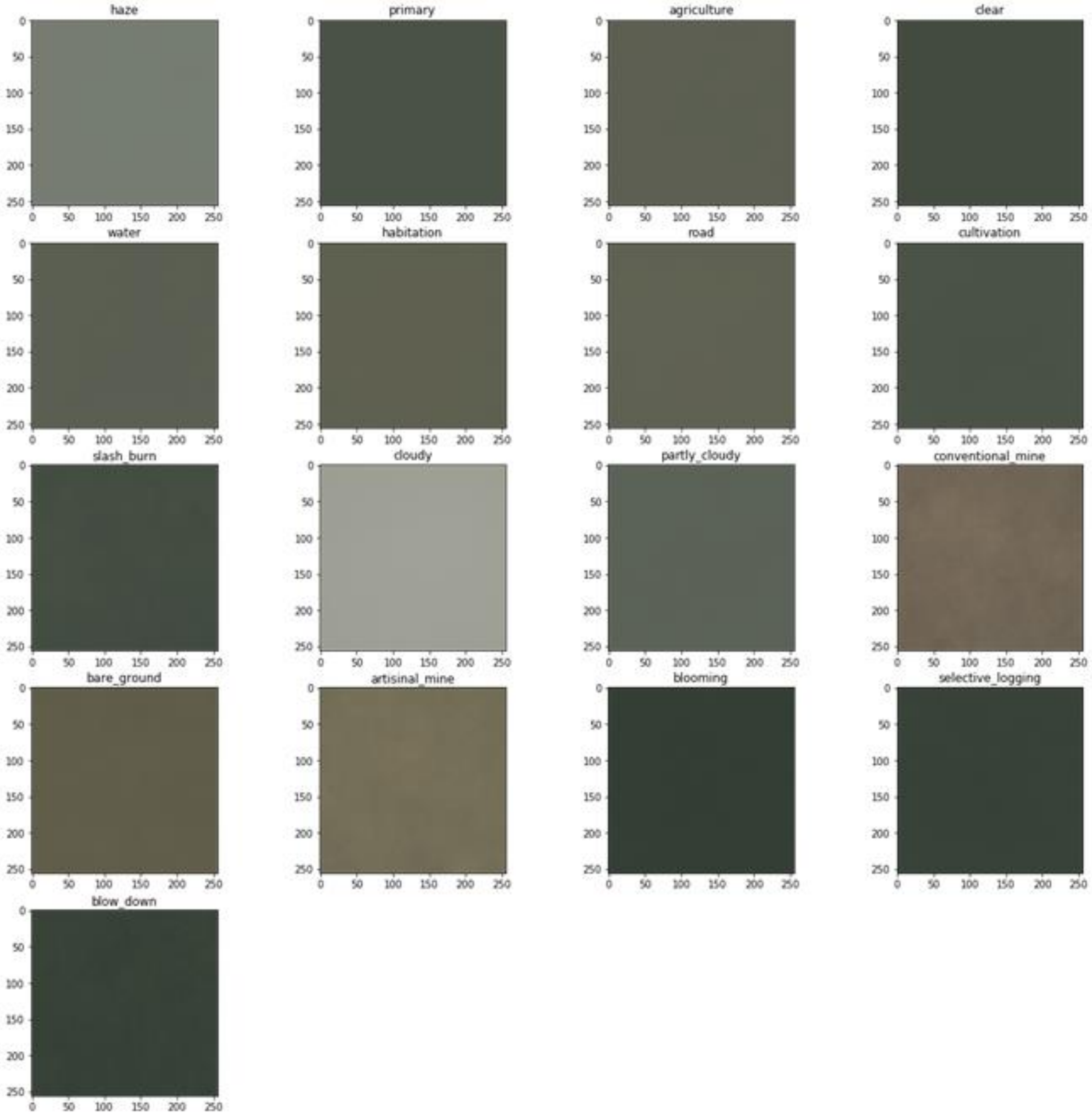
IMAGE DECOMPOSITION

- The average RGB values for all of the images tagged with a given label were calculated
- Green is the dominant color for the majority of the labels, with the exception of conventional mines, artisanal mines, and bare ground for which red is the dominant color.
- Cloudy and hazy are the most saturated labels and this is expected given they are white or gray dominant.
-



AVERAGE IMAGES

- Artisanal mine, conventional mine, and bare ground appear as shades of brown while all other images appear as shades of green or gray.
- Cloudy and haze images have distinct gray colors.
- Because the images can contain a variety of labels, are not oriented in any specific way, and information regarding the location of the labeled features within the image is not available these images are not particularly helpful.



MACHINE LEARNING – RANDOM FOREST

- For this project, we ultimately would like to explore convolutional neural networks for the classification of the various labels in the images, but as a starting point, a baseline Random Forest model was created.
- The labels for all the images in the training set were one-hot-encoded and the average R, G, and B values per image were calculated.
- This information was used to train a Random Forest model and the following parameters were optimized using Grid Search CV and 5-fold cross validation:
  - Max depth = 10
  - Number of estimators = 200
- The training set consisted of 20,000 images and the testing set consisted of 5,000 images.
- The overall F1 score obtained with this model is **0.704**, however the **average F1 score is 0.21** - this is because the model did not perform well on the rare occurring labels.
- This model served as a baseline so that further optimization using thresholding or more image data was not performed.

	Recall	Precision	F1
haze	0.039813	0.548387	0.074236
primary	0.988687	0.956234	0.972190
agriculture	0.494366	0.657329	0.564318
clear	0.938177	0.831550	0.881651
water	0.026211	0.660000	0.050420
habitation	0.014196	0.818182	0.027907
road	0.278067	0.617162	0.383393
cultivation	0.000000	0.000000	0.000000
slash_burn	0.000000	0.000000	0.000000
cloudy	0.520958	0.746781	0.613757
partly_cloudy	0.000844	0.250000	0.001682
conventional_mine	0.000000	0.000000	0.000000
bare_ground	0.014925	0.400000	0.028777
artisinal_mine	0.078125	0.625000	0.138889
blooming	0.000000	0.000000	0.000000
selective_logging	0.000000	0.000000	0.000000
blow_down	0.000000	0.000000	0.000000

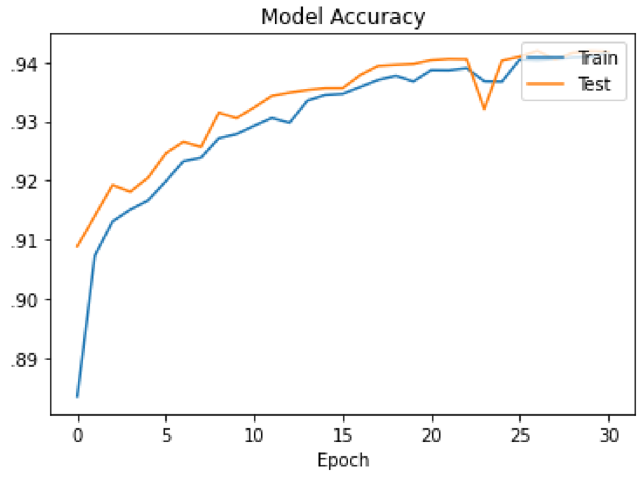
CONVOLUTIONAL NEURAL NETWORKS

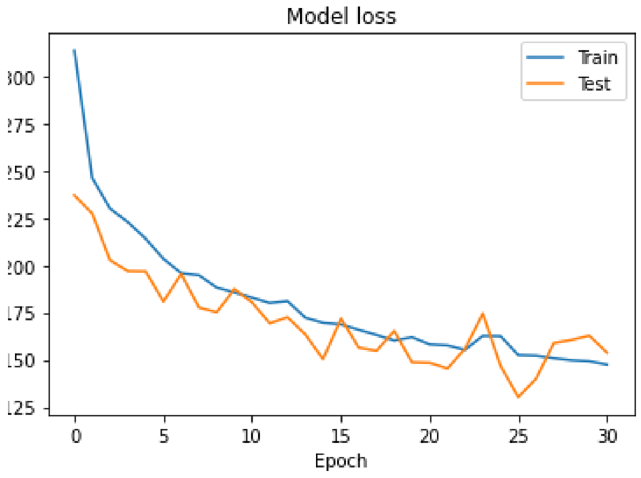
- Convolutional Neural Networks (CNNs) are a class of neural networks that have been successfully proven overtime for image classification problems.
- Implemented using Keras (Tensorflow backend)
- Our base CNN model:
  - Image resolution: 64x64 (3 channels)
  - Training set size: 20,000
  - Validation set size: 7,000
  - Test set size: 5,120 (multiples of data generator batch size of 256)
  - Batch size: 256
  - Early stopping: monitoring validation loss with patience of 5
  - Metric: F1 score
  - Adam optimizer
  - Loss: binary crossentropy
  - Epochs: 50 (early stopping kicked in at the 26th epoch)

```
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d_5 (Conv2D)            (None, 57, 57, 32)        6176
_________________________________________________________________
conv2d_6 (Conv2D)            (None, 50, 50, 64)        131136
_________________________________________________________________
max_pooling2d_2 (MaxPooling2 (None, 12, 12, 64)        0
_________________________________________________________________
conv2d_7 (Conv2D)            (None, 10, 10, 64)        36928
_________________________________________________________________
dropout_3 (Dropout)          (None, 10, 10, 64)        0
_________________________________________________________________
flatten_2 (Flatten)          (None, 6400)              0
_________________________________________________________________
dense_2 (Dense)              (None, 128)               819328
_________________________________________________________________
dropout_4 (Dropout)          (None, 128)               0
_________________________________________________________________
preds (Dense)                (None, 17)                2193
=================================================================
Total params: 995,761
Trainable params: 995,761
Non-trainable params: 0
```

BASE CNN MODEL

- Maximum Validation Accuracy: .9409
- Minimum Validation Loss: .1302
- Maximum Validation F1: 0.8115
- Epochs: 26 (patience of 5)





BASE CNN MODEL RESULTS

	Best Threshold	F1
haze	0.350	0.665
primary	0.550	0.977
agriculture	0.450	0.762
clear	0.540	0.946
water	0.270	0.465
habitation	0.250	0.479
road	0.400	0.627
cultivation	0.220	0.447
slash_burn	0.030	0.041
cloudy	0.270	0.717
partly_cloudy	0.220	0.851
conventional_mine	0.090	0.222
bare_ground	0.140	0.252
artisinal_mine	0.120	0.414
blooming	0.050	0.125
selective_logging	0.040	0.071
blow_down	0.010	0.012
Average_F1	0.235	0.475

- Blowdown, slash and burn, conventional mine, selective logging and blooming only get predicted at very low threshold values and their F1 score is also low.
- Average F1 score is still very low when all labels are weighted equally. Initially an overall F1 score had been computed but found to be misleading since it did not do well at performing the low occurrence labels

CNN FILTER VISUALIZATION - BASE CNN

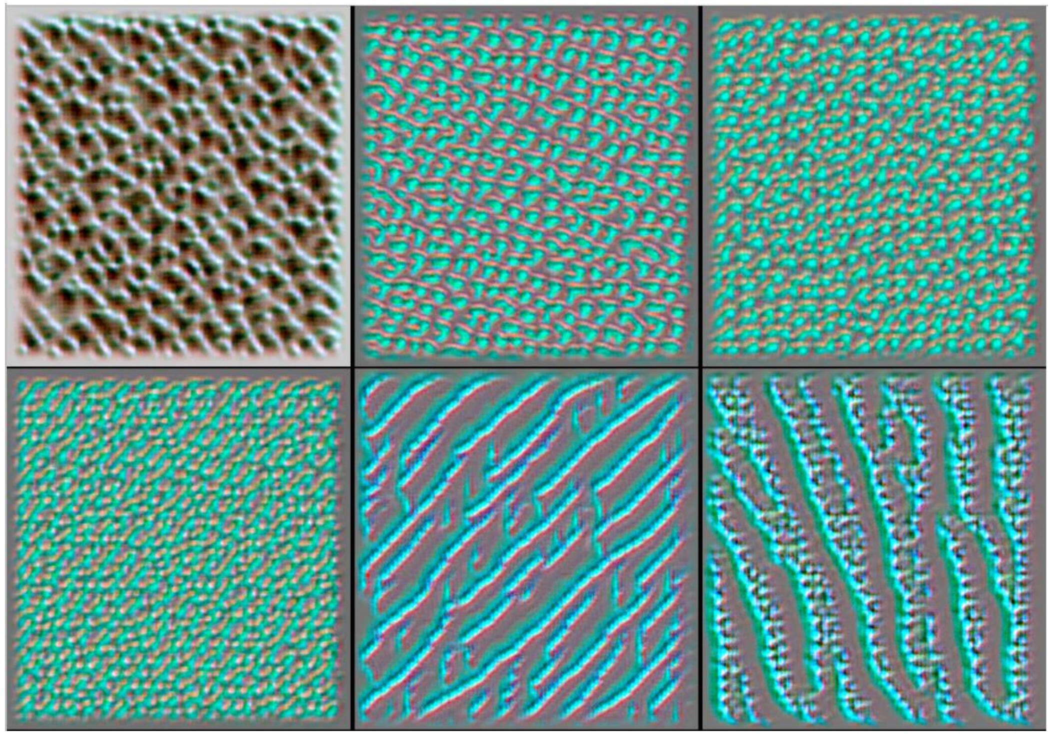


Figure 3 - Filter Visualization for the last convolutional layer (conv2d_7) of the baseline model

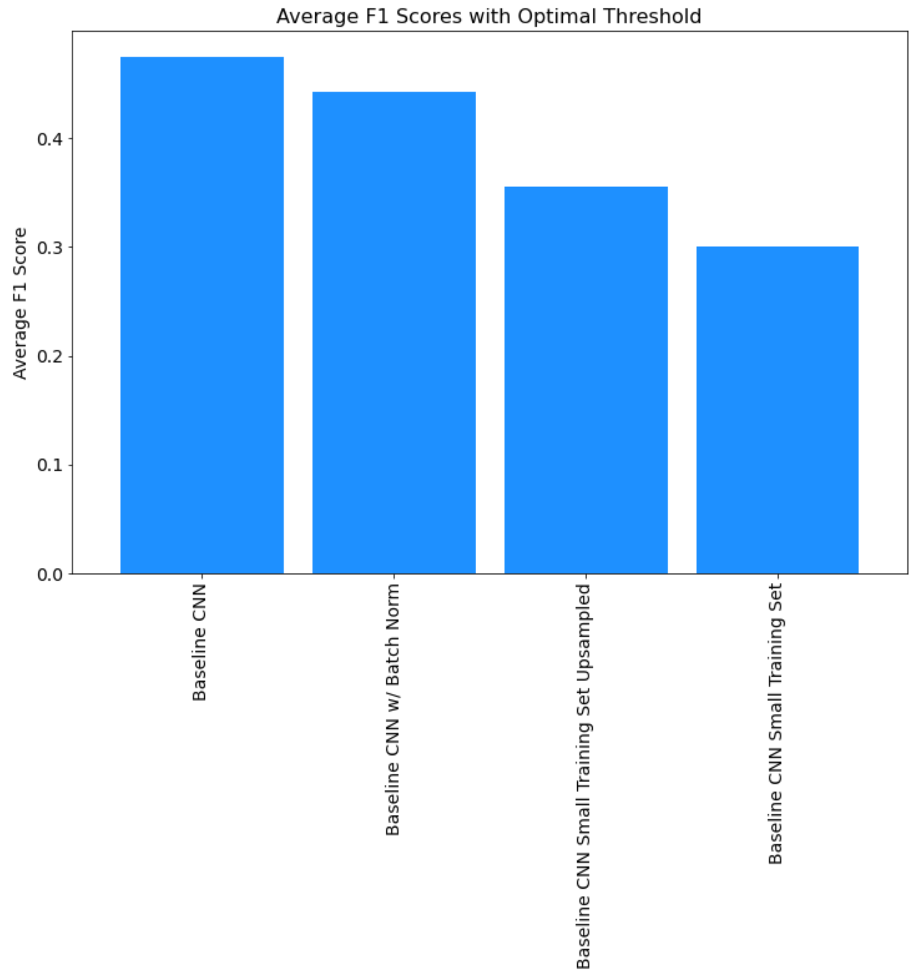
- One way to better understand how CNNs work is by visualizing the network's filters
- In order to create filter visualizations, a random image is created and fed to the network, the gradients of the loss for the image for a given filter are computed and the image is progressively updated to maximize that filter's activation function.
- The resulting images represent the inputs that maximize each filter's activation and can give us clues about what the network's filters are "looking for".
- The image on the left were created using the visualization process on our base CNN:
  - The filters remind us of some of the common patterns observed in our satellite images: tree cover, bare ground, agriculture, rivers and roads.
  - More broadly, these filters are likely extracting specific edge and gradient patterns from the input images.

CONVOLUTIONAL NEURAL NETWORK

- Several parameters and models were tested for comparison and to try and improve upon our baseline CNN classification performance:
  - **Batch Normalization**: an additional batch normalization layer was added to the model immediately following the first fully connected dense layer. This was done both to assess impact on training speed and to explore its impact on the network's classification performance.
  - **Softmax for cloud coverage prediction**: since the cloud coverage labels are mutually exclusive, the same baseline model architecture was used while modifying the target inputs to include only the 4 cloud coverage labels and the last (predictive) layer was changed to have a softmax activation function with 4 nodes (4 possible cloud classes). This was done to assess whether this model would perform better at exclusively predicting cloud cover and if so, utilize it for this purpose, separately from the land use and ground cover labels.
  - **Upsampling:** blowdown images from the entire training set were utilized and up-sampled to verify whether the model would do a better job predicting that label.
  - **Grayscale** images with higher resolution: the same baseline model was used with higher resolution grayscale images (1 channel) with 128x128 resolution to assess whether a single channel with better image resolution would do better at identifying and classifying the various labels.
  - **Baseline CNN with small training set**: the number of training samples was reduced to 2000 samples to assess the training set size's impact on overall model performance.

MODEL COMPARISON

- The best performing model, as measured by the average F1 score was the baseline convolutional neural net model.
- The addition of a batch normalization layer, the increase in image resolution provided by the grayscale images and the softmax activation layer for cloud coverage prediction (not shown) either decreased or did not have a significant impact in improving model performance.
- While upsampling increased the model performance over the small training set this might have been simple due to the fact that the up-sampling caused to it to have more training data, since it did not increase prediction performance for the up-sampled label.
- As expected, reducing the size of the training set had a significant detrimental impact to the average F1 score.



TRANSFER LEARNING

- Keras' applications interface offers us the ability to load pre-trained neural networks, thereby allowing us to use the knowledge gained from a problem to a different, related problem - this is known as transfer learning
- There are several available models that can be used in image recognition tasks, for our problem we utilized the VGG16 model to try and improve upon our predictions.
- The VGG16 model, developed by Oxford's Visual Geometry Group (VGG), consists of a 16 layer convolutional neural network as shown in the diagram below
- By specifying the parameter *weights = 'imagenet'* when the model is loaded, the network's initial weights will have been pre-trained using the ImageNet dataset
- The ImageNet dataset contains ~ 1.2 million images and the network weights were optimized to classify these input images into 1,000 different object categories

Figure 4 – Architecture of VGG16 transfer model

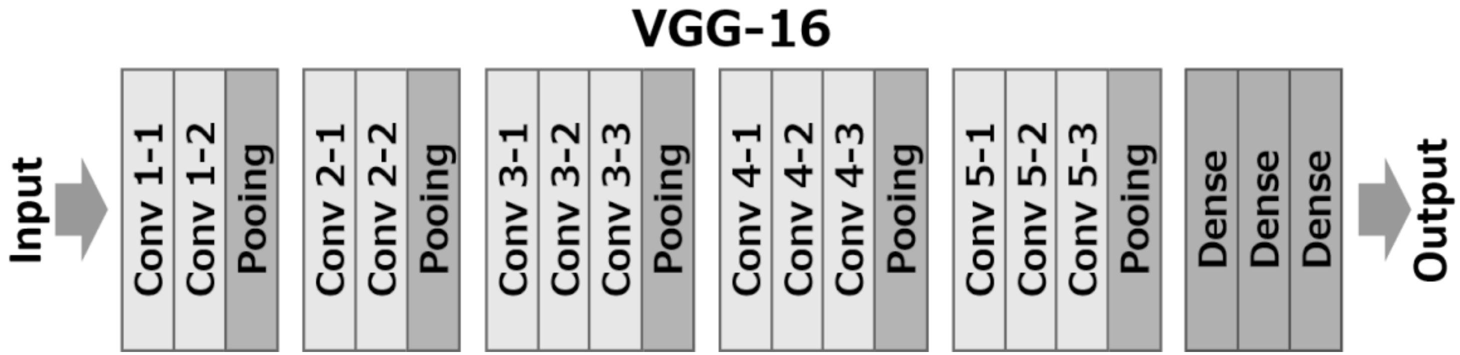


Image source: https://neurohive.io/en/popular-networks/vgg16/

TRANSFER LEARNING

- Some of the resulting images for the filters from the pre-trained VGG16 network on the ImageNet set are shown in the image on the right.
- Some of the images displayed resemble dogs, marbles, trees and chains.
- These filters are a direct result of the contents of the ImageNet dataset and the network's original purpose of identifying the 1,000 different labels and objects.

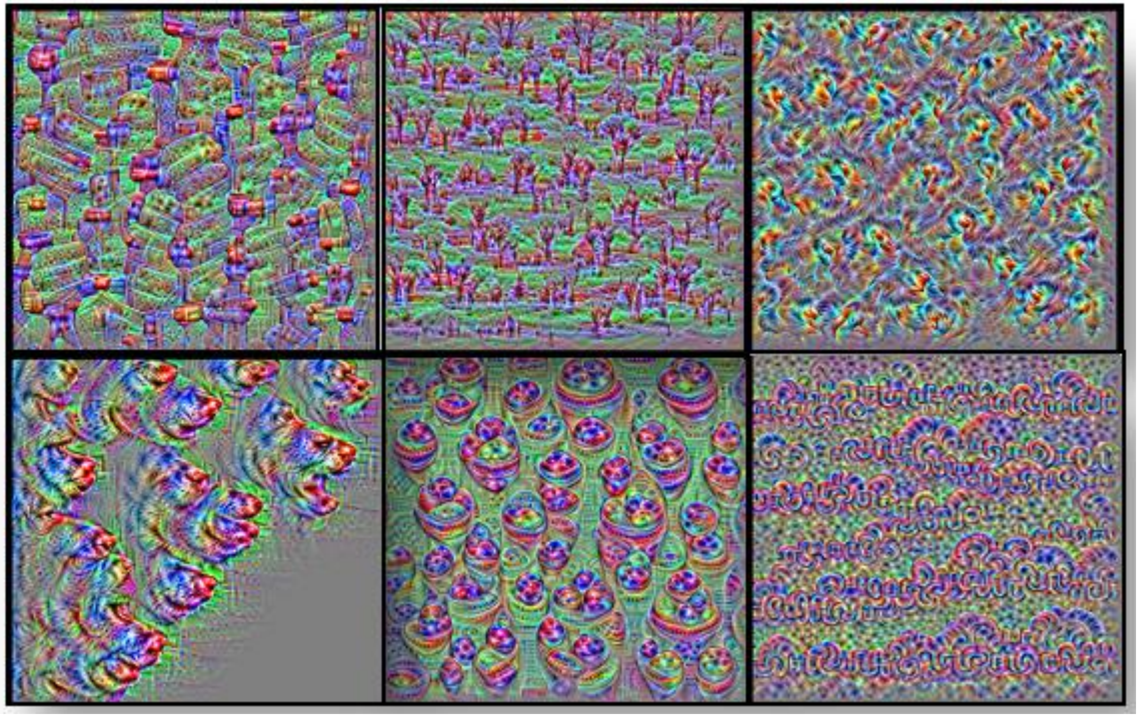


Figure 5 - Filter Visualization for the last convolutional layer for the pre-trained VGG16 model

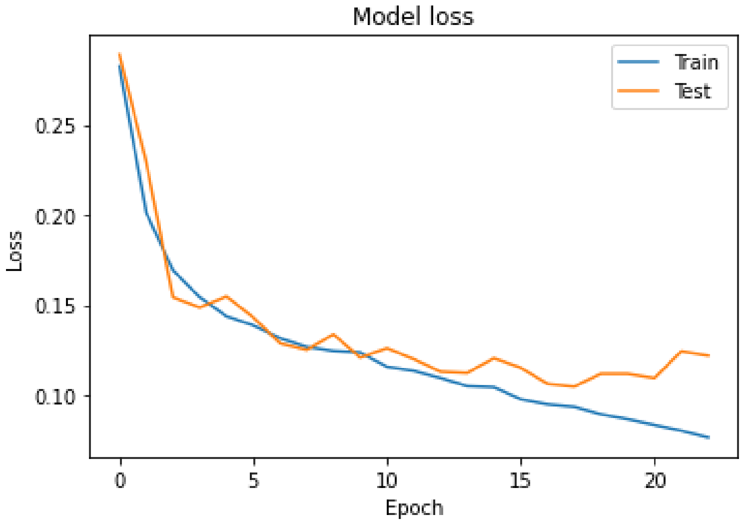
TRANSFER LEARNING

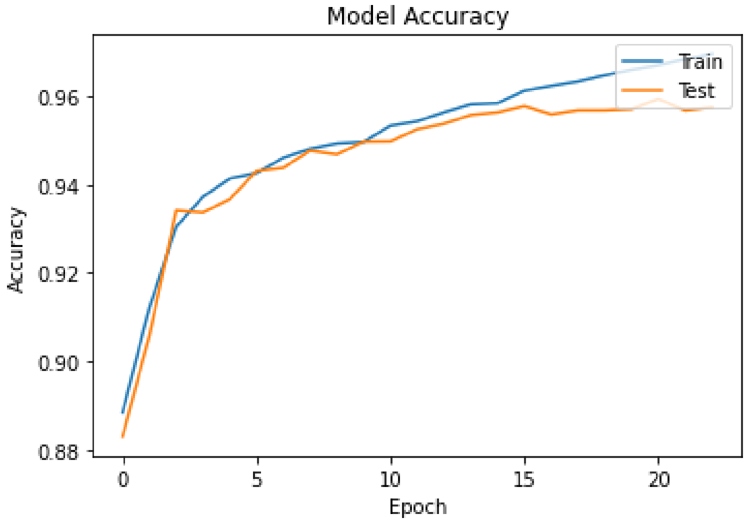
- To adapt the VGG16 model for our purposes, we integrated it into a new model, for which the VGG16 block essentially functions as a feature extractor.
- The output from the VGG16 network was flattened and fed into a final classifier type layer and retrained with the satellite images to predict our 17 labels of interest as shown in the summary architecture below.

```
Model: "sequential_1"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
batch_normalization_1 (Batch (None, 64, 64, 3)         12

vgg16 (Model)                (None, 2, 2, 512)         14714688

flatten_1 (Flatten)          (None, 2048)              0
_________________________________________________________________
preds (Dense)                (None, 17)                34833
=================================================================
Total params: 14,749,533
Trainable params: 14,749,527
Non-trainable params: 6
_________________________________________________________________
```

TRANSFER MODEL

- Maximum Validation Accuracy: .9567
- Minimum Validation Loss: .1047
- Maximum Validation F1: 0.8646
- Epochs: 19 (patience of 5)





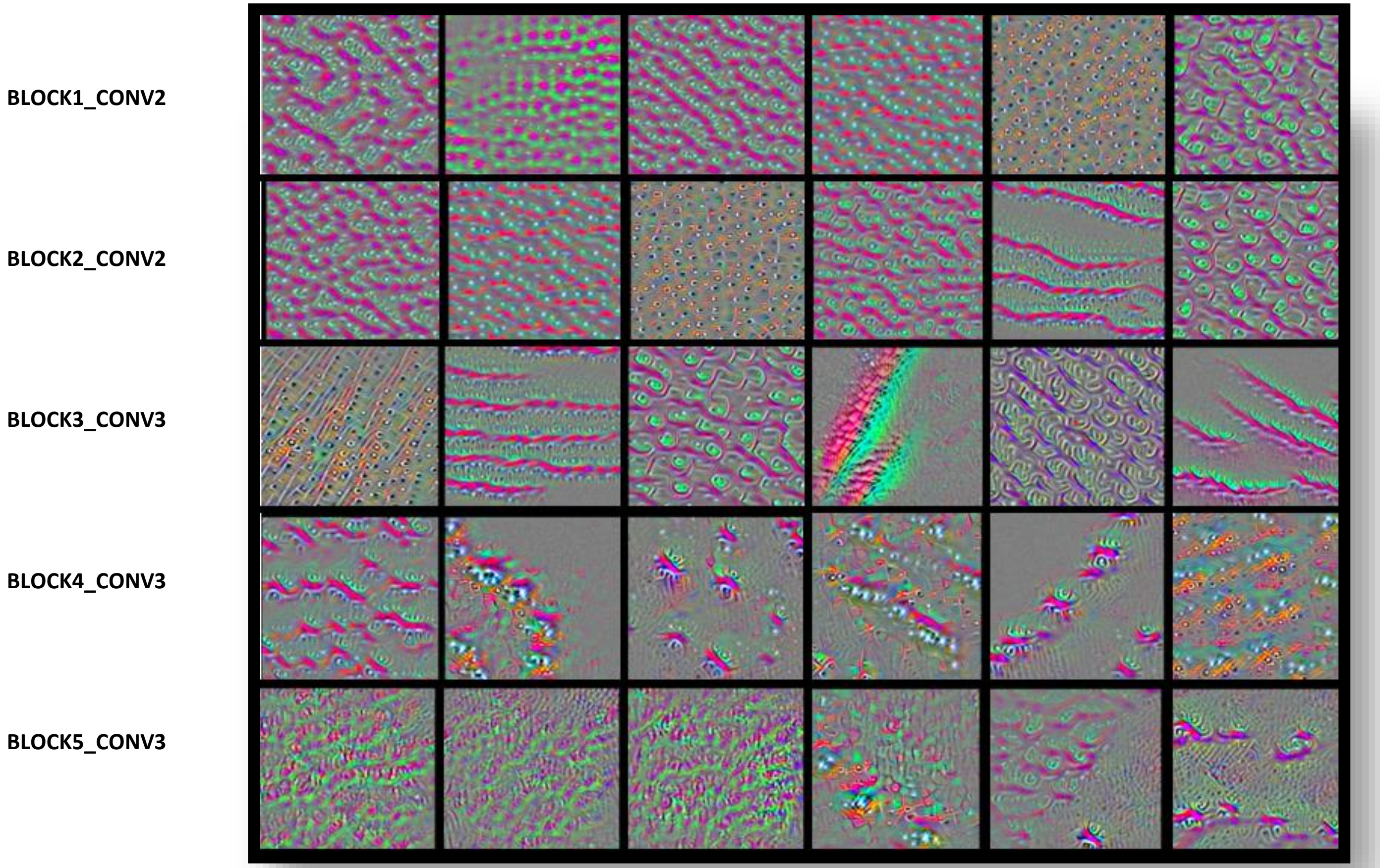
TRANSFER LEARNING MODEL RESULTS

- The transfer model yielded vast improvements in classification performance over the baseline convolutional neural net.
- For labels with low occurrence, the best threshold is quite low (slash and burn, conventional mines and blowdown).
- There is a noted difference in F1 scores between the commonly occurring labels and the rare labels, with slash and burn, conventional mines, bare ground, blooming, selective logging and blow down having best scores under 0.32.

	Best Threshold	Best_F1
haze	0.180	0.719
primary	0.380	0.983
agriculture	0.230	0.824
clear	0.690	0.964
water	0.280	0.727
habitation	0.240	0.549
road	0.280	0.727
cultivation	0.350	0.557
slash_burn	0.050	0.114
cloudy	0.500	0.812
partly_cloudy	0.460	0.909
conventional_mine	0.030	0.207
bare_ground	0.120	0.303
artisinal_mine	0.700	0.701
blooming	0.140	0.220
selective_logging	0.190	0.310
blow_down	0.090	0.190
average	0.289	0.577

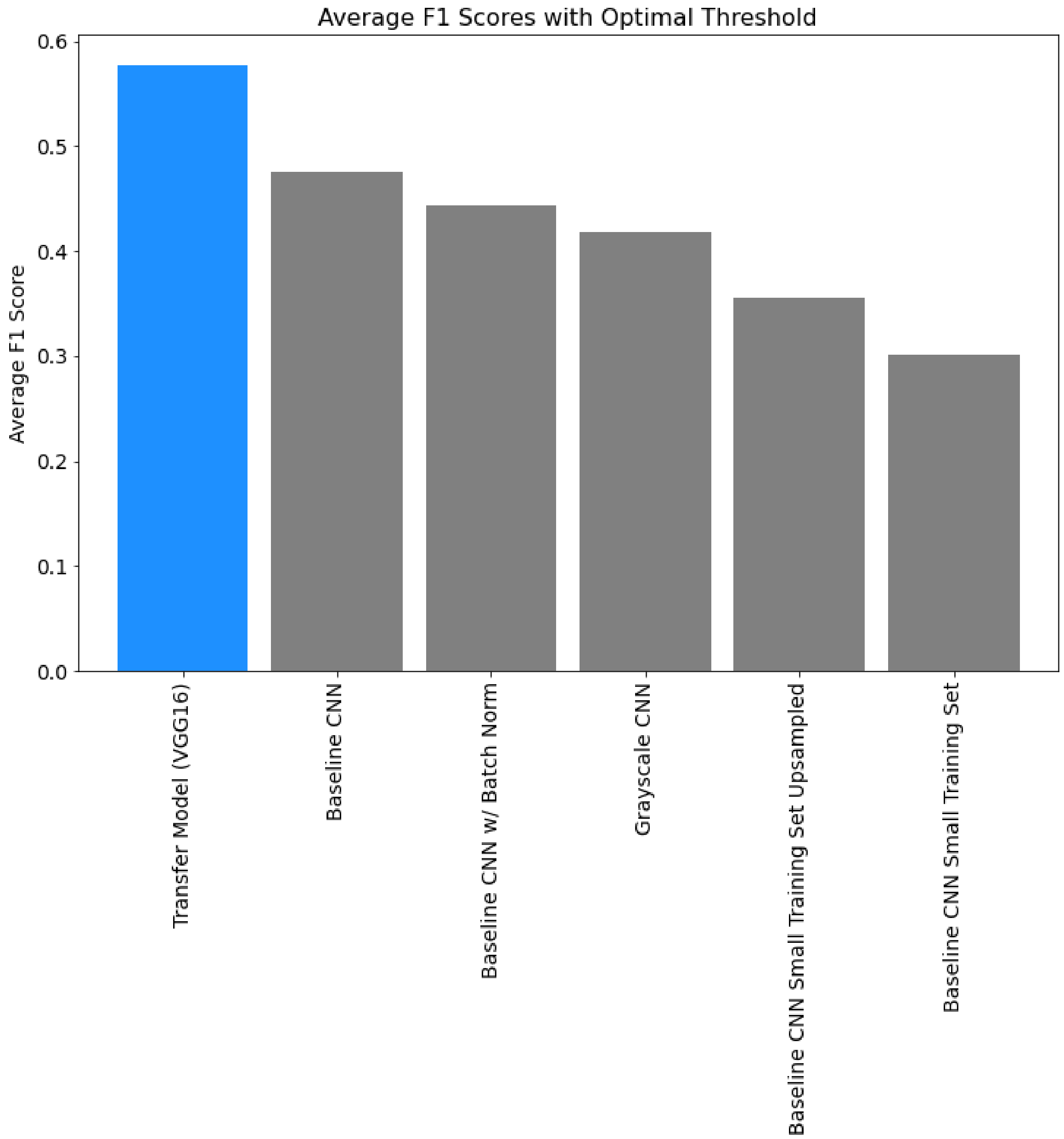
FILTER VISUALIZATION - TRANSFER MODEL

- The filter patterns causing maximum activation have significantly changed from the ones originally loaded in the VGG16 model trained on the ImageNet set.
- The deeper filters of the layers generally tend to look for more complex patterns while the earlier filters seem to be looking for general edges and patterns.
- The early filters also resemble the filters observed in our base CNN.
- The VGG16 model contains a larger number of layers with a total of 4224 convolutional filters, which is significantly more complex than our baseline CNN with only 160 filters.
- Its better performance is likely because it can both capture more patterns and more complex features.



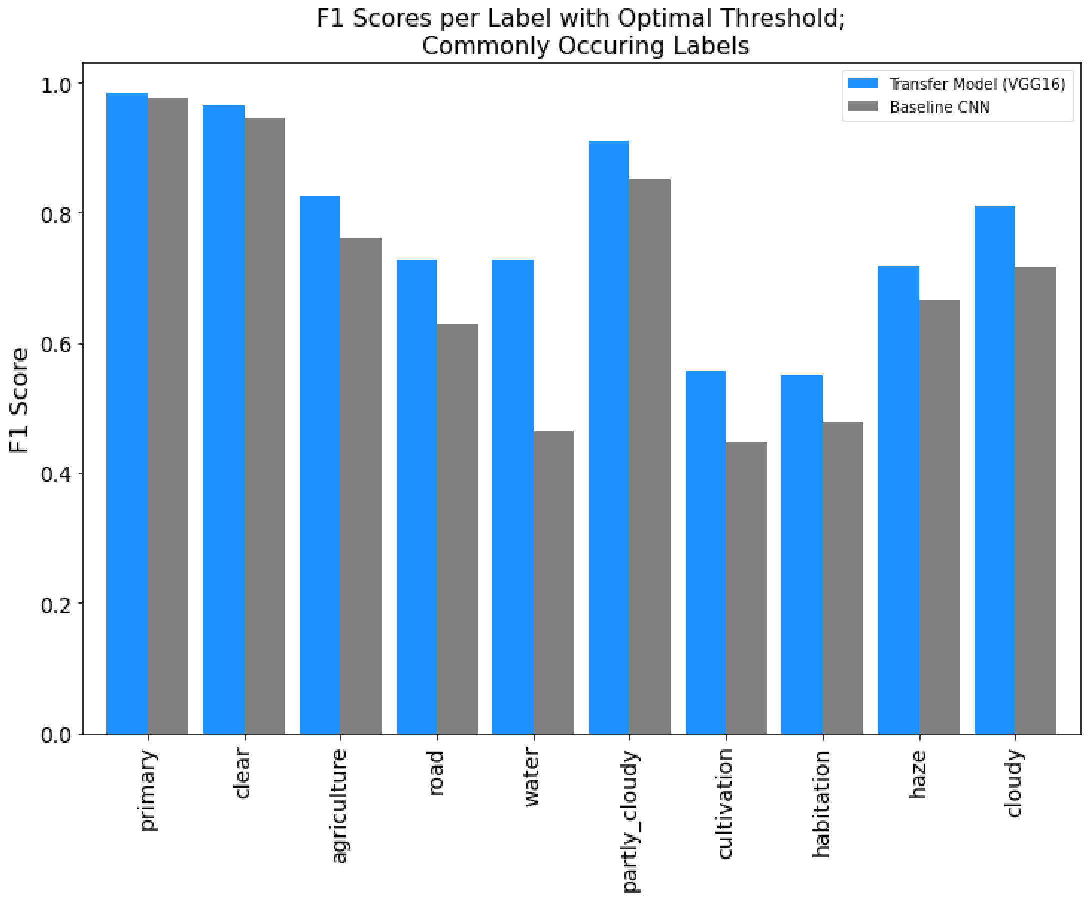
MODEL COMPARISON

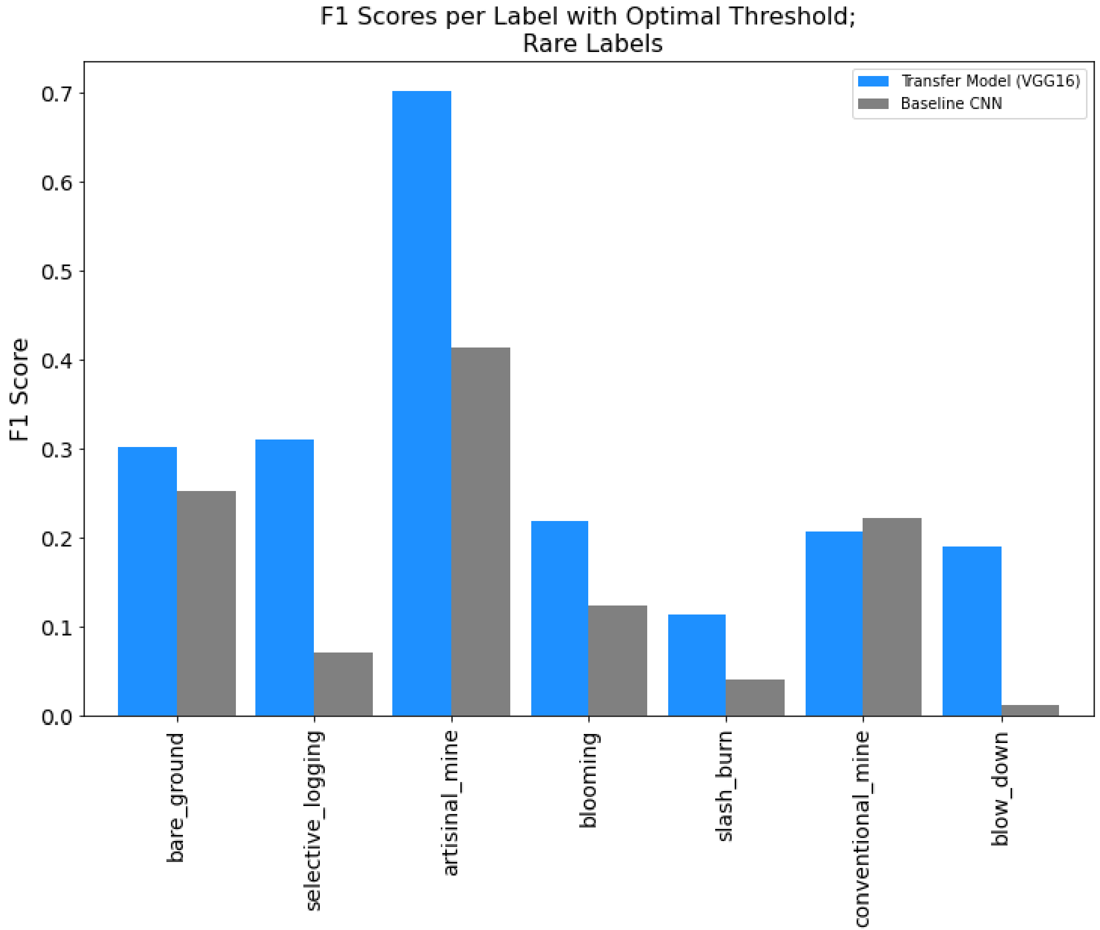
- The thresholds that yielded the best F1 scores was also computed for all of the deep learning models for each label, and the average F1 score for each model was calculated - the results are shown on the plot on the right.
- The VGG16 model has the highest average F1 score at 0.58 with the Baseline CNN model performing second best, with an average F1 score of 0.47.



MODEL COMPARISON

- The transfer model improved our results for both the high occurrence and low occurrence labels over our Baseline CNN. The most drastic improvement in performance were for predicting the low occurrence labels: artisanal mine, blow-down and selective logging
- The only label for which the transfer model had inferior results to the baseline CNN was “conventional mine”, the difference is small and not enough to justify using the baseline CNN model for the prediction of this class.





CONCLUSION

The model's predictions are useful for labeling the data, which in turn could help organizations make more informed decisions in the rainforest's preservation and management efforts.

The VGG16 model was the best performing model for all labels, with the most drastic improvements observed for the low occurring labels (blow down, selective logging, blooming, slash and burn and artisanal mine). This implies that the depth and complexity of the neural network, as well as the size of the training set play an important role in improving deep learning model performance for computer vision problems.

For future work, increasing the size of the training set by using all available images and using higher resolution .TIF images would be recommended, as well as artificially increasing the training set size through the use of image distortion, rotation and cropping, especially for the low occurrence labels.

REFERENCES

1. https://www.nationalgeographic.com/environment/2018/11/how-cutting-the-amazon-forest-could-affect-weather/
2. https://www.nationalgeographic.com/environment/2019/08/why-amazon-doesnt-produce-20-percent-worlds-oxygen/#close
3. https://advances.sciencemag.org/content/4/2/eaat2340
4. Simulated Changes in Northwest U.S. Climate in Response to Amazon Deforestation - https://journals.ametsoc.org/doi/10.1175/JCLI-D-12-00775.1
5. Team, K. (2020). Keras documentation: Visualizing what convnets learn. Retrieved 5 August 2020, from https://keras.io/examples/vision/visualizing_what_convnets_learn/