DeepRootz: Classifying satellite images of the Amazon rainforest

Scott Longwell, Tyler Shimko, and Alex Williams

Dataset analysis

Raw data

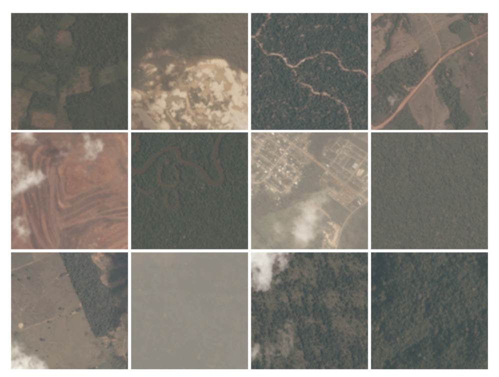

Fig. 1 - Example "chips" from the Kaggle Planet challenge dataset.

- 40,479 ~1km x 1km "chips"
- Each chip 256px x 256px
- 4 channels
  - Red
  - Green
  - Blue
  - Near-IR

Atmospheric/ground label co-occurrence

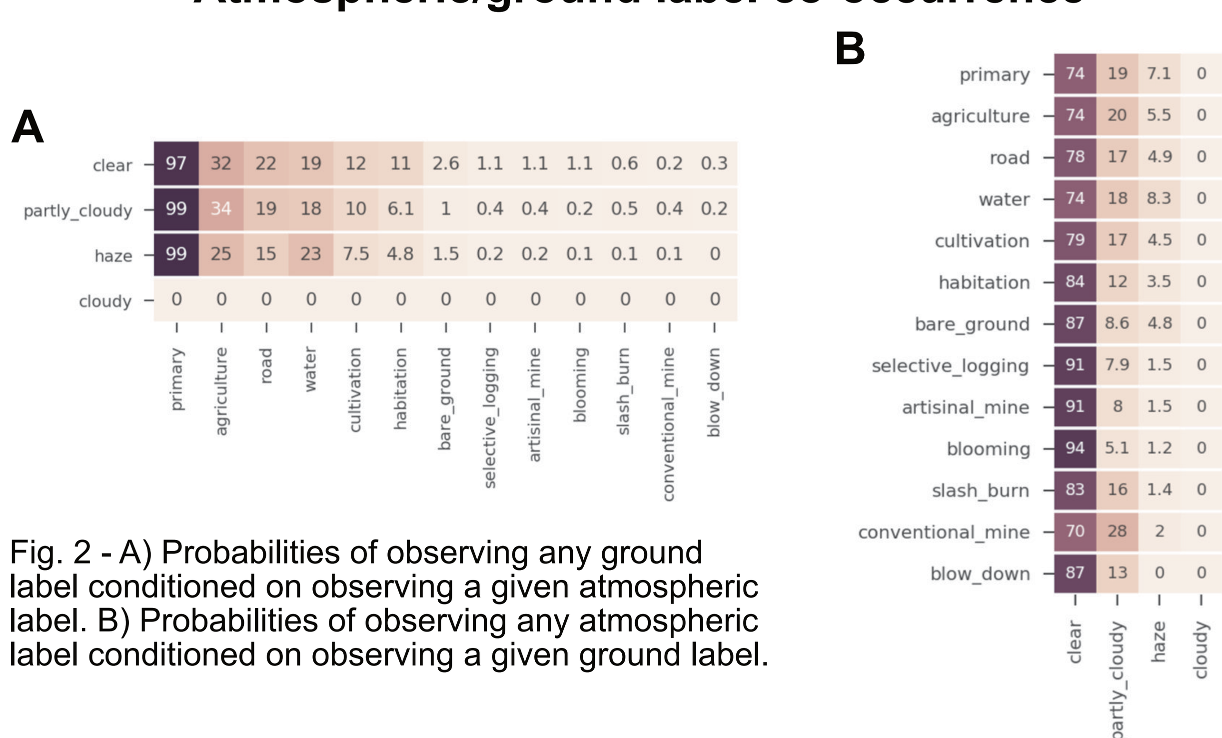

Fig. 2 - A) Probabilities of observing any ground label conditioned on observing a given atmospheric label. B) Probabilities of observing any atmospheric label conditioned on observing a given ground label.

Each chip is labelled with exactly one atmospheric condition label and one or more ground condition labels. An atmospheric label of "cloudy" negates any possible ground labels, as the ground cannot be observed from the satellite imagery.

Ground label co-occurrence

Certain ground labels are highly correlated. For instance, if an area has agricultural activity, it is also likely to have roadways and other signs of human settlement.

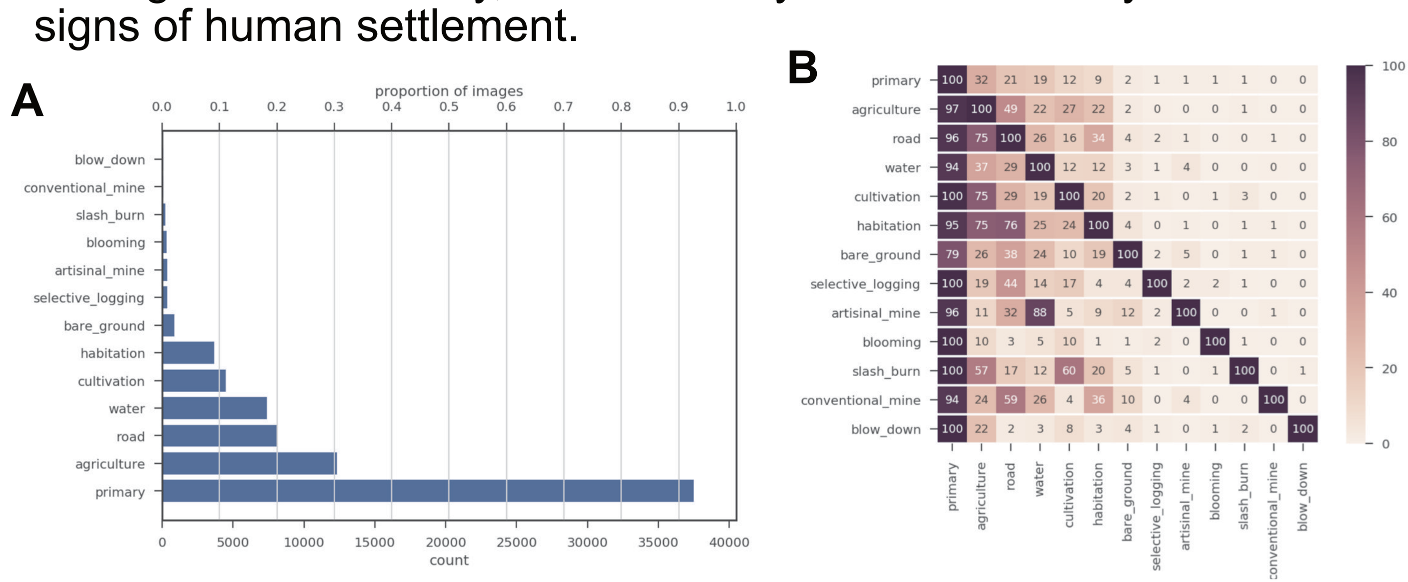

Fig. 3 - A) Total number of images labelled with a given label. B) Co-occurrence matrix for all ground labels.

Model Design

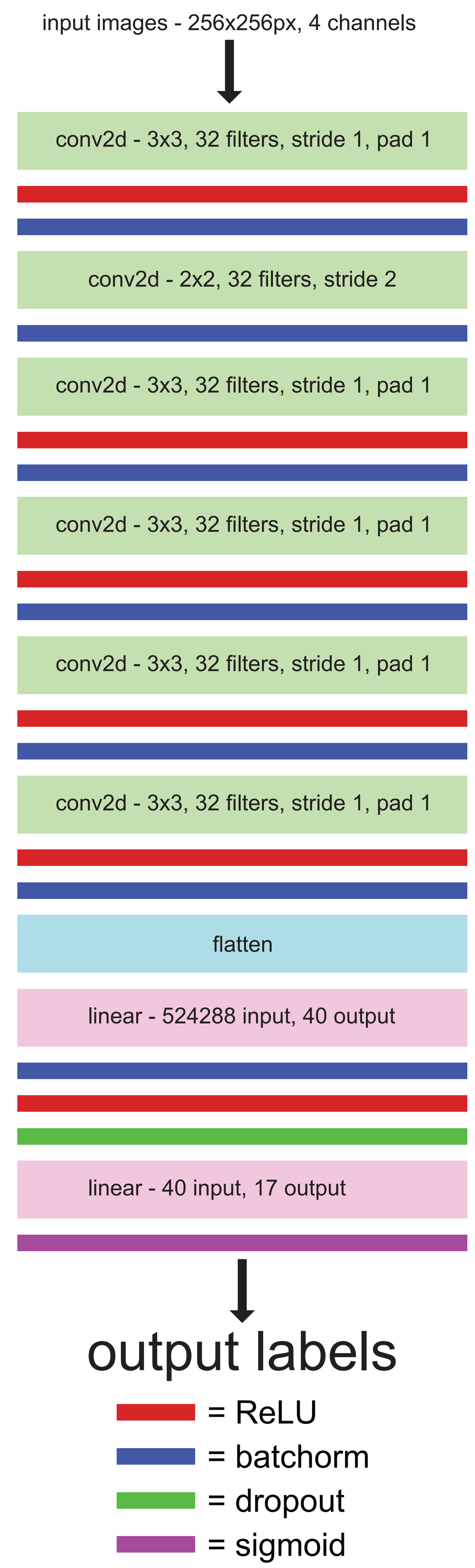

Fig. 4 - Current model architecture

Convolutional layers

Our model is composed of six convolutional layers, including one layer with filter size equal to stride to effectively downsample the image. Stacked layers with smaller filter sizes serve to increase effective receptive field size without the same parameter scaling as larger sized filters.

ReLU non-linearities

With the exception of the final layer of the model, ReLU units are used to add non-linearities to the model. This decision was based primarily on the performance superiority of ReLU units in similar image labelling tasks.

Batch normalization

Batch normalization was employed following every layer in the model, with the exception of the final layer. Batchnorm has been shown to both ease the training process as well as serve as a type of regularization.

Dropout

Dropout with a probability of 0.2 was employed on the second to last set of activations for regularization.

Loss function

Binary cross entropy loss was used for each possible label. Initial thresholds for label inclusion were set at 0.5, but were adjusted as training progressed to provide a boost in model performance.

F_2-measure

$$(1+\beta^2)\frac{pr}{\beta^2 p + r} \quad \text{where } p = \frac{tp}{tp+fp},\ r = \frac{tp}{tp+fn},\ \beta = 2.$$

Final entries are judged on mean example-wise F_2-measure upon submission.

Results

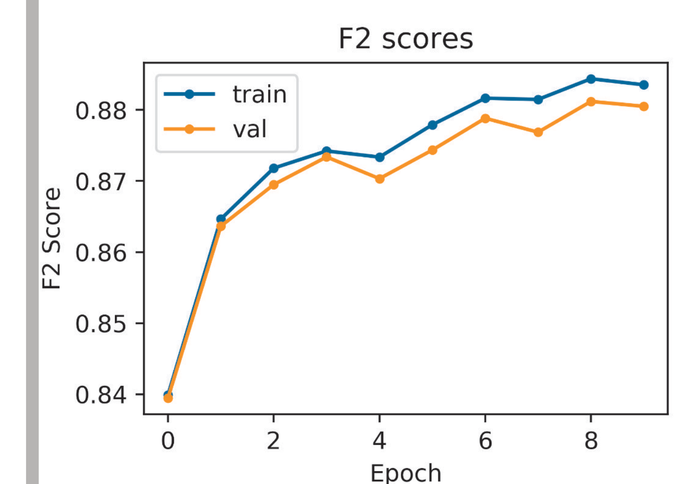

Fig. 5 - Training and validation F_2 scores per epoch.

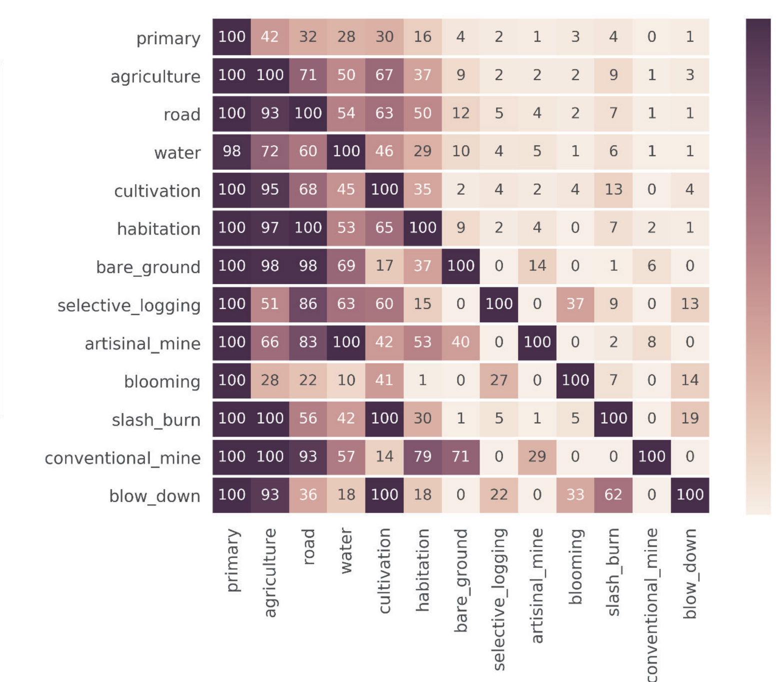

Fig. 6 - Conditional co-occurrence matrix for ground labels applied by our model.

- Best mean F_2 = 0.88
- Current leader achieves test set F_2 = 0.93
- Best decision threshold is rarely near 0.5
- Mid-frequency labels are hardest to recall

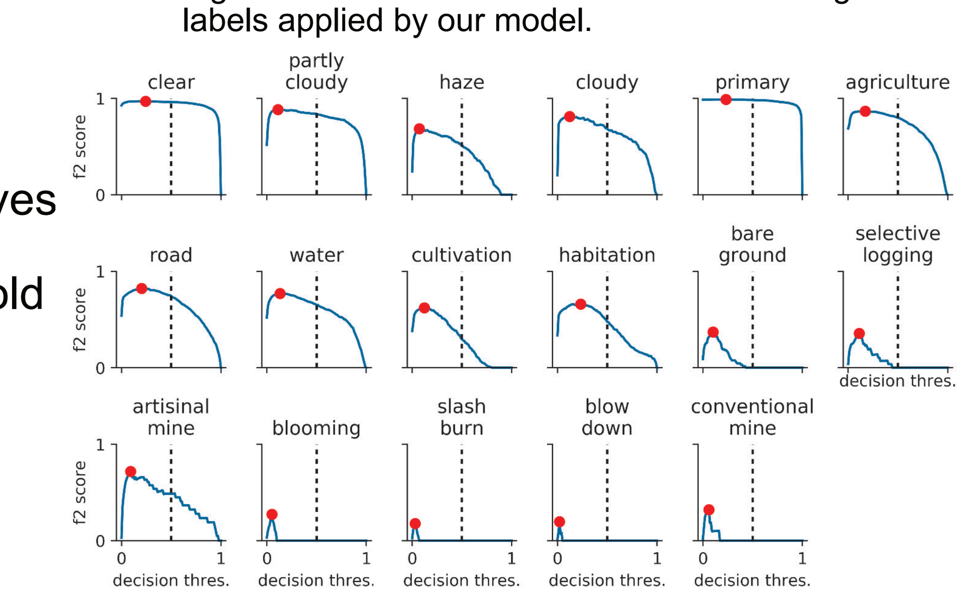

Fig. 7 - F2 score as a function of the decision boundary for every possible label. In most cases, F_2 score can be improved by adjusting the threshold from the default of 0.5.

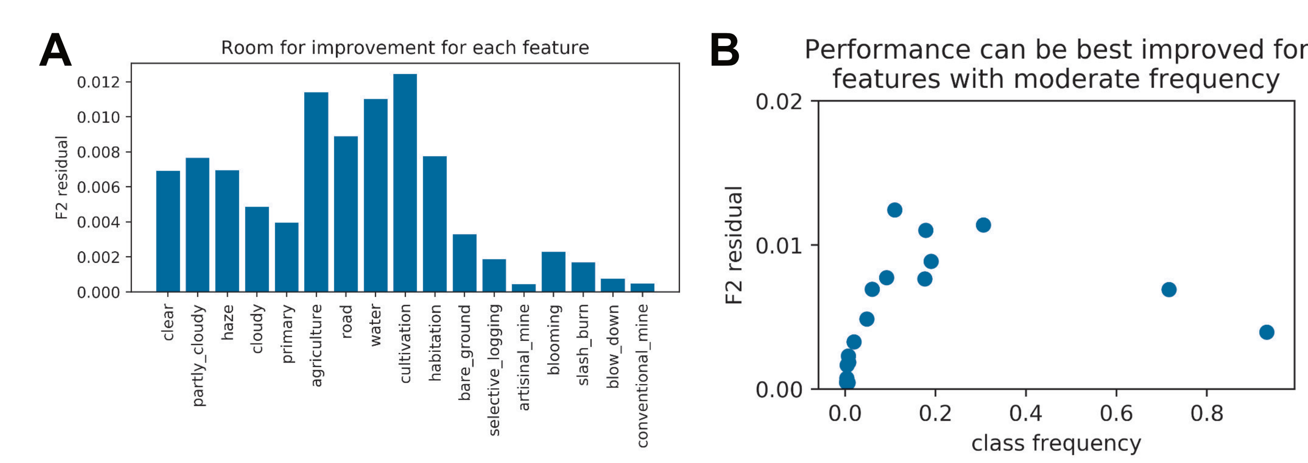

Fig. 8 - A) Total possible improvement if every instance of any particular label was correctly recalled. B) Possible improvement as a function of the frequency of any given label.

Future Directions

- Try larger models with further regularization
- Try residual network models to improve training speed/reliability
- Continue adjusting decision thresholds to improve training accuracy