Multi-label Classification of Satellite Images with Deep Learning

Daniel Gardner & David Nichols
Department of Computer Science, Stanford University

Project Summary

Up-to-date data on deforestation locations is vitally important to scientists and governments working to preserve the Amazon rainforest. We implement a Convolutional Neural Network model to perform multi-label classification of Amazon satellite images. Our model identifies the weather conditions and natural terrain features in the images as well as man-made developments such as roads, farming, and logging. We show that a relatively simple CNN model can achieve good results (0.85 F score), allowing stakeholders to pinpoint where current deforestation is taking place.

Background / Data

Deforestation of the Amazon basin has occurred at a rapid pace over the past four decades. Governments and scientists, concerned about consequences ranging from habitat loss to climate change, need a way to monitor where and when the deforestation is occurring.

Planet, a satellite imaging company, recently released a dataset of 40,000 images from the Amazon basin and sponsored a Kaggle competition involving labelling the atmosphere and ground features in the images. Each image is 256x256 pixels and has rgb and infrared channels. Notably, these images have at least ten times greater resolution than any earth image data used previously in tracking deforestation, with each pixel representing only three to five meters.

We build a model using Convolutional Neural Networks to analyze each image and classify it with one or more of 17 feature labels. Note that each image has exactly one atmosphere label and zero or more ground labels. Some ground features are human-related (habitation, slash burn) while others are natural (blooming, blow down).

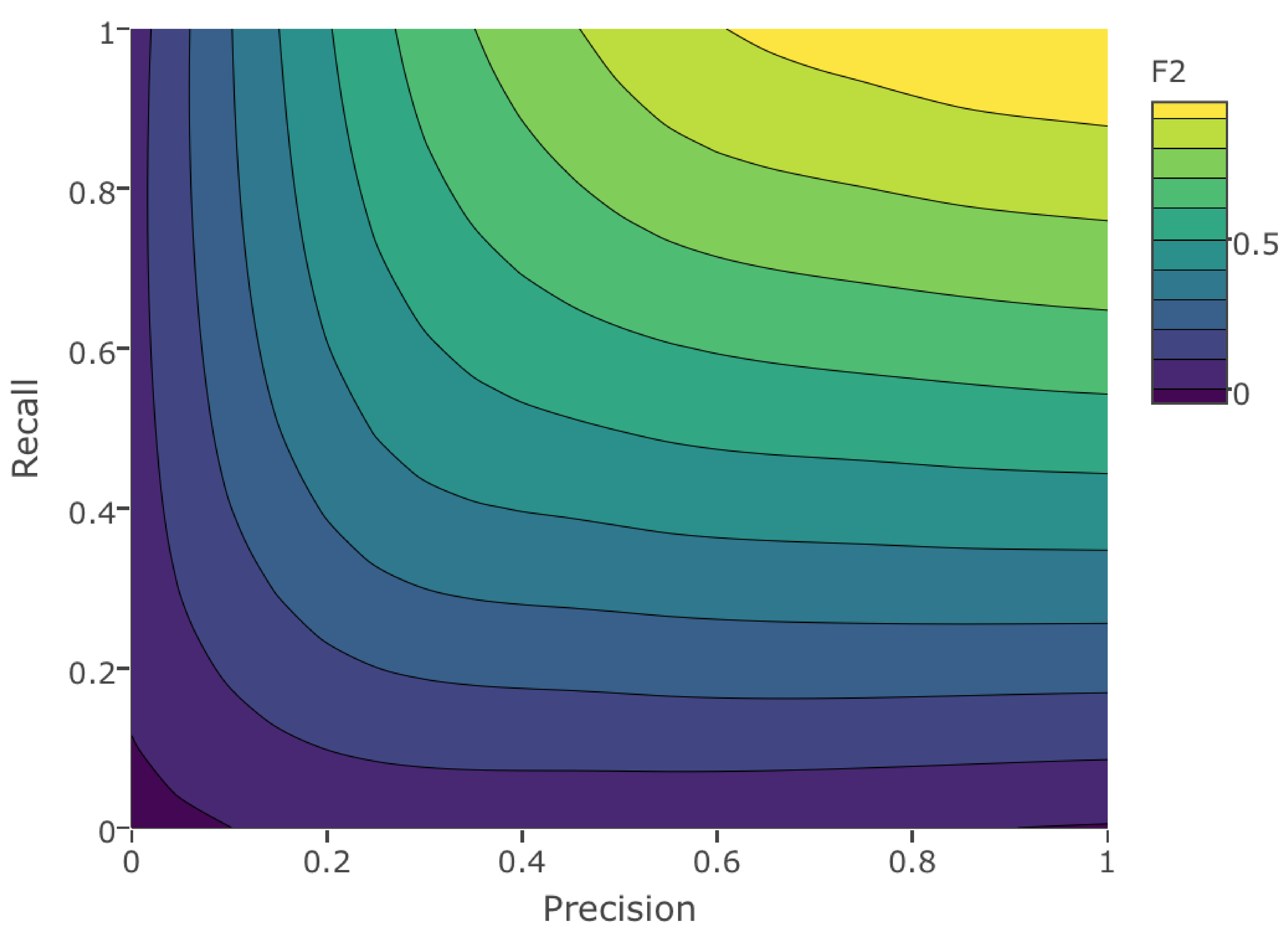
F2 Metric

The evaluation metric for our model is F2 score, which is defined as:

$$F_2 = (1 + 2^2) \cdot \frac{\text{precision} \cdot \text{recall}}{(2^2 \cdot \text{precision}) + \text{recall}}$$

where Precision $= \frac{TP}{TP+FP}$ and Recall $= \frac{TP}{TP+FN}$.

The contours for this evaluation metric are below:



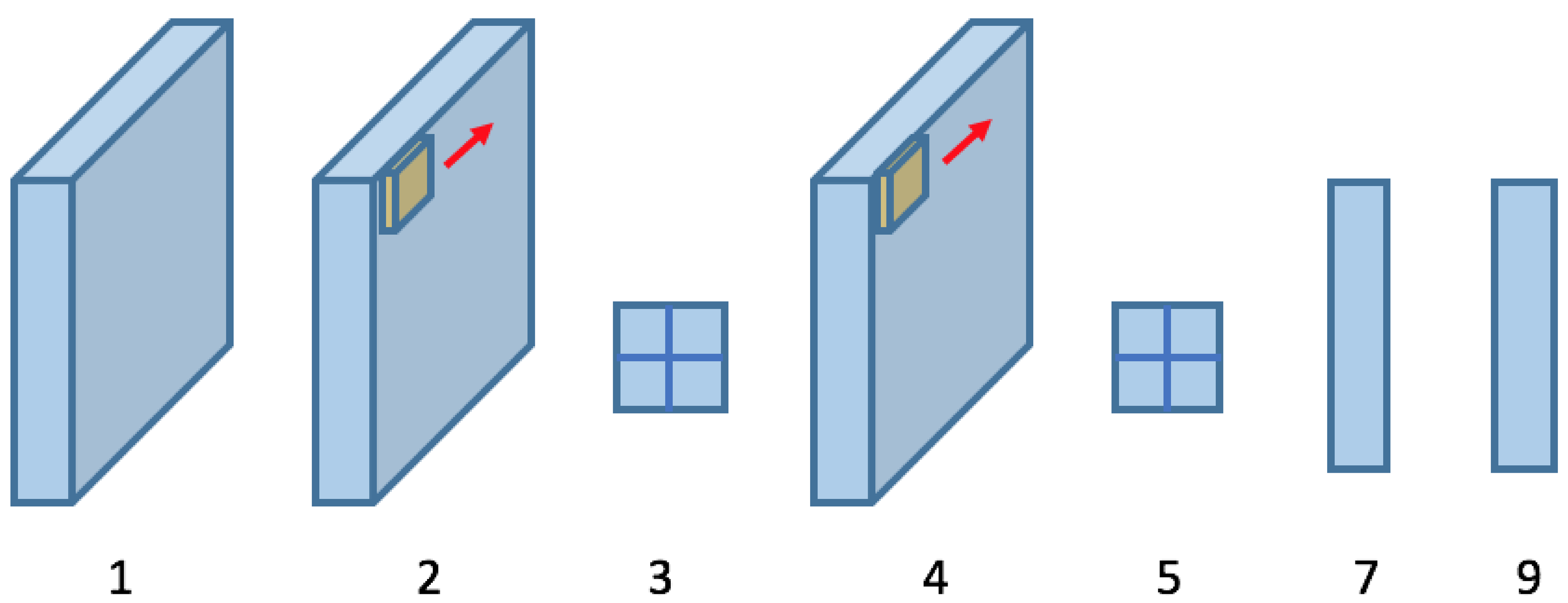
Classification Performance

Samples of satellite images and labels

	clear	partly_cloudy	haze	cloudy	primary	agriculture	road	water	cultivation	habitation	bare_ground	selective_logging	artisinal_mine	blooming	slash_burn	conventional_mine	blow_down
F2	0.961	0.777	0.709	0.755	0.987	0.812	0.738	0.597	0.498	0.512	0.024	0.000	0.186	0.000	0.000	0.000	0.000
Frequency	0.703	0.181	0.067	0.049	0.929	0.302	0.198	0.179	0.112	0.090	0.024	0.009	0.008	0.008	0.006	0.002	0.002
	Atmosphere				**Ground**												

Model performance by class

Model Structure



Layer	**Dimensions**
1) Input Image	256x256x4
2) Convolutional Layer – 32 3x3 filters, stride = 1, ReLU activation	254x254x32
3) Max Pooling Layer – 2x2	127x127x32
4) Convolutional Layer – 64 4x4 filters, stride = 1, ReLU activation	124x124x64
5) Max Pooling Layer – 2x2	62x62x64
6) Dropout – 0.25	62x62x64
7) FC Layer – 128, ReLU activation	1x128
8) Dropout – 0.5	1x128
9) FC Layer – 17	1x17

Preliminary Findings

Table: Prediction accuracy by model

Model	Epoch(s)	F2
Baseline Model	25	0.8529
ResNet50	1	0.8314
InceptionV3	1	0.8158

- Our Baseline model had the best performance, but ResNet and InceptionNet were trained for only one epoch and can likely do much better.

Discussion & Further Research

Remarks:

- We used binary cross-entropy loss, the Adam optimizer, and a learning rate of 0.001 for our training.
- After just one training epoch on the Baseline, we were able to achieve a relatively good F2 score (about 0.70). This is because the vast majority of the images have the clear and/or primary labels and our model classifies more than 95% of these correctly.
- Conversely, even after many training epochs, our model was unable to identify almost any of the rare ground features (those appearing in less than 5% of the training data).
- Despite being quite rare, our model performed well in identifying cloudy images, doing about as well as hazy and partly cloudy, which were much more common.

Further Improvements:

- Continue training with ResNet and InceptionNet on GPUs.
- Generate additional images with rare ground labels using data augmentation techniques, allowing our model to gain more experience recognizing these features.
- Experiment with a two-stage classifier that first identifies an image's atmosphere and then uses atmosphere-specific model weights to make ground feature classifications.
- Reformulate the model to focus more on identifying general human encroachment. Classifying the 17 labels is an interesting machine learning task, but the end goal of monitoring deforestation might be better achieved by zeroing in on destructive human activities and observing them over time.

dangard@stanford.edu / david.nichols@stanford.edu