Deep Multi-Label Classification for High Resolution Satellite Imagery of Rainforest

Christian Castellanos, Andrew Nguyen, Robert Sun

CS231N, Stanford University

Objective

- **Label** satellite images of rainforest with **multiple** land/air conditions
- Automated labeling can **track** *where/how/why* of **deforestation**

- Use **CNNs** (Pretrained and Random) of various architectures
- Evaluate and improve performance of dataset labels

Dataset *(Planet: Understanding the Amazon from Space)* [1]

- Dataset of 40,000+ high resolution satellite images
- Images given up to **17** potential labels (some rare, some common)
- Images contain *specific* **and** *general* features

Methods and Experiments

Used a variety of pre-trained and randomly initialized state-of-the-art NN models (*VGG*, *ResNet*, *Inception*) as well as our own architecture for multi-class labeling of satellite imagery with the Keras Framework (Tensorflow) [2].

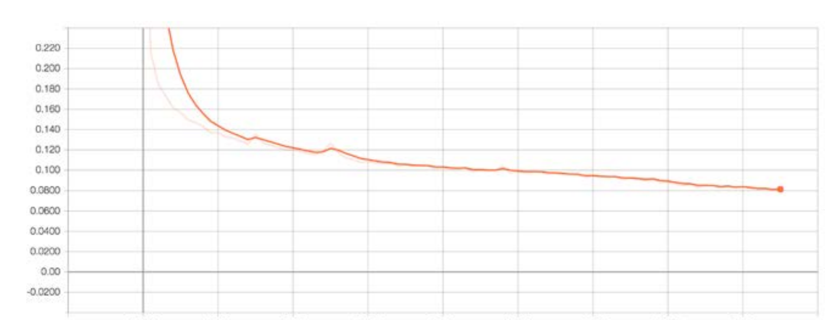
Models were set up to output **17** probabilities [0,1] that represent the presence of a given label belonging to an image, with averaged binary cross-entropy loss. A threshold is applied to create generated labels.

Model performance is based on validation accuracy and mean F_2 score

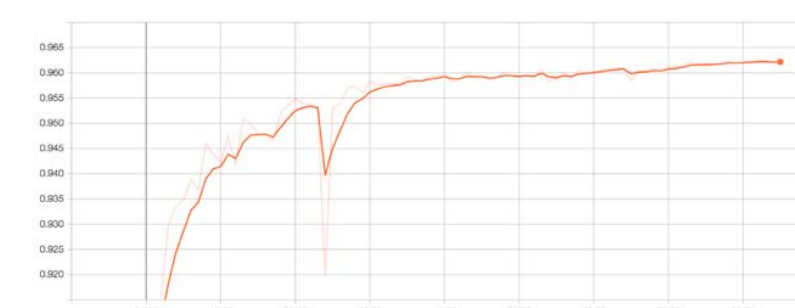
$$F_\beta = (1+\beta^2)\frac{Prec \, * Recall}{(\beta^2 * \, Prec) + Recall}$$

Precision: the ratio of true positives to predicted positives
Recall: the ratio of true positives to actual positives
F_2 score weights recall higher than precision

Training History (Best Performing)



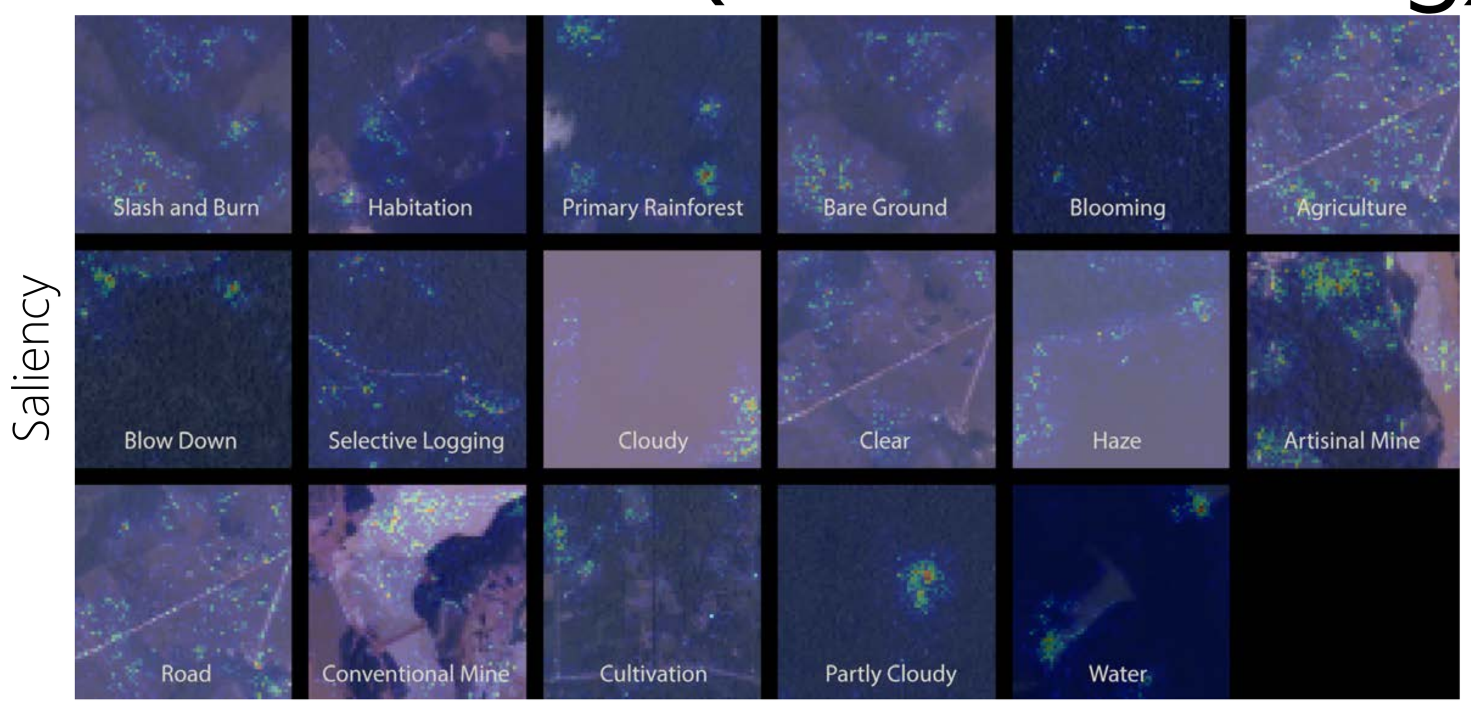
Loss



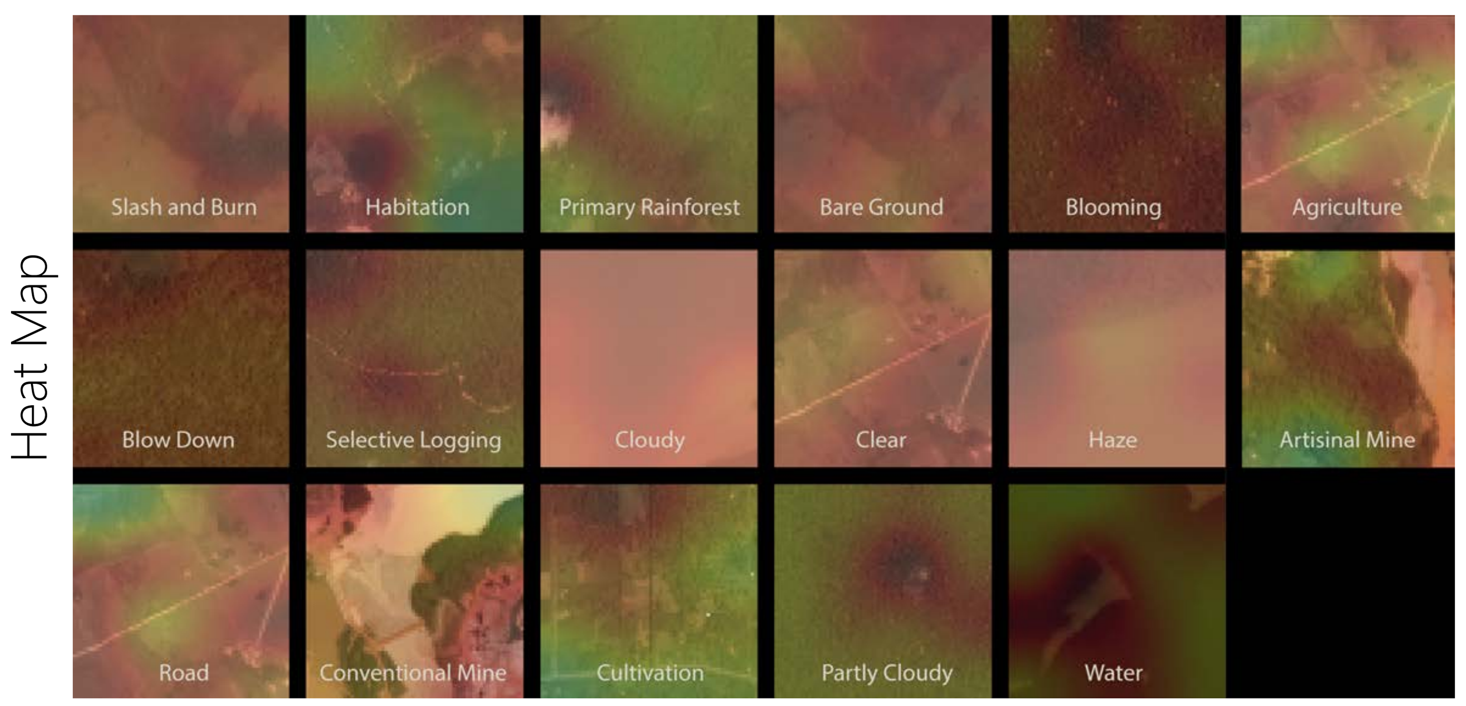
Accuracy

Visualizations (Best Performing)

Saliency



Heat Map



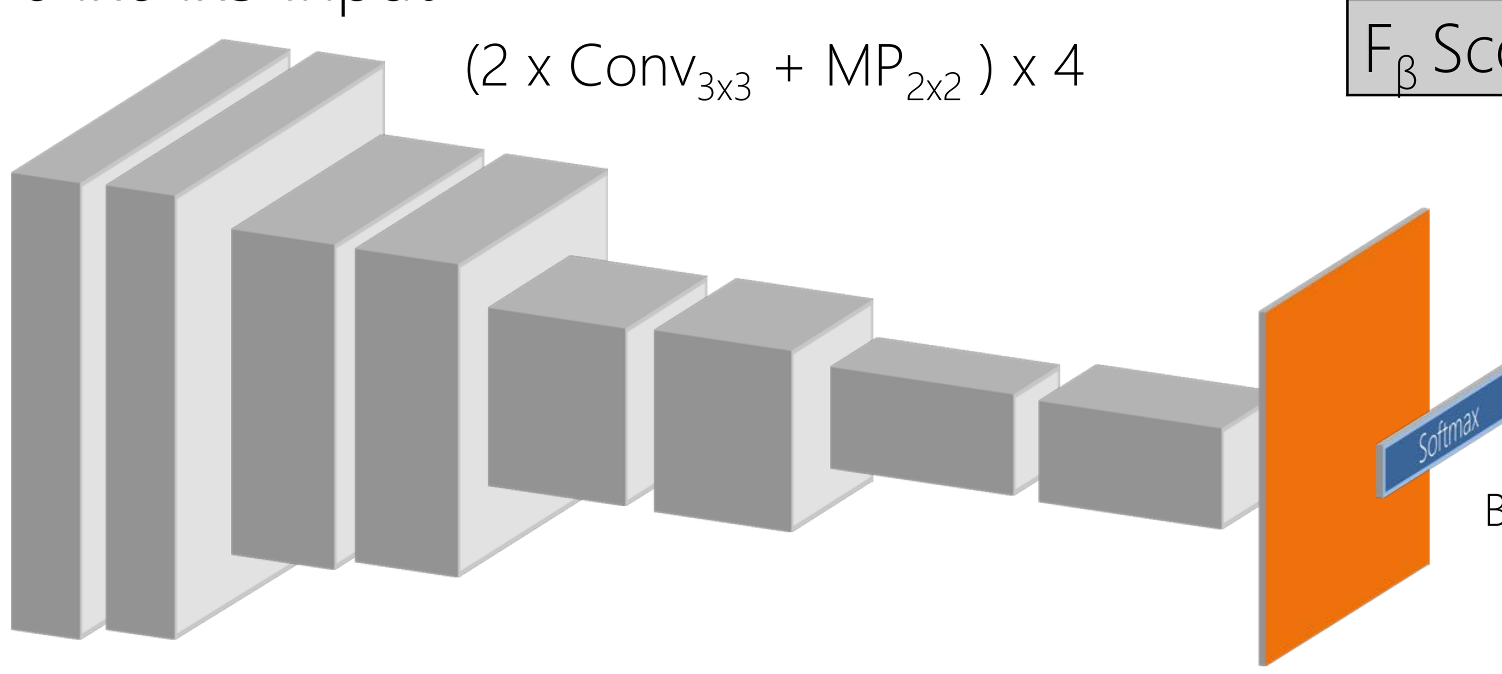
Saliency maps denote the pixels in an image that contribute most to a given label [5]
Heat maps denote the area in an image where the classifier looks for a given label[5]

Best-Performing Custom Network Architecture

VGG-Style Feedforward Network
(No Pre-Learned Weights)
Used: Batch Normalization, Dropout [2][3]

64x64x3 Input

$(2 \times Conv_{3x3} + MP_{2x2}) \times 4$



17 1-Hot Multilabel

Binary Cross-entropy "Multilabel" Loss

32 64 128 256 512 Feature Layer

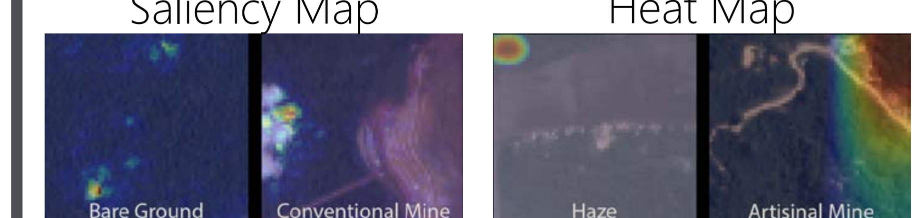
Metrics

Val Acc	96.26%
F_β Score (β=2)	0.9156

Other Architectures Attempted

VGG

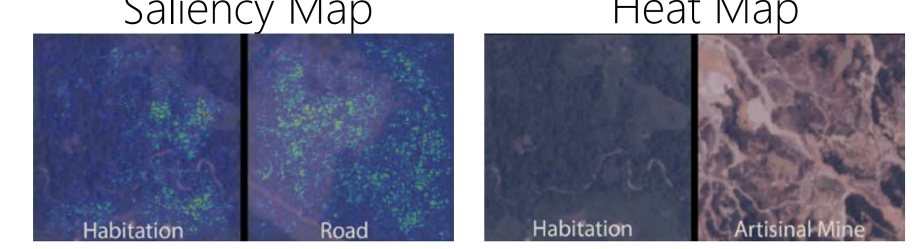
Saliency Map Heat Map



VGG is most similar to the best performing custom net, but deeper and more complicated.
Saliency and heat maps show accurate focus regions for some classes, but for others are completely wrong.

Inception

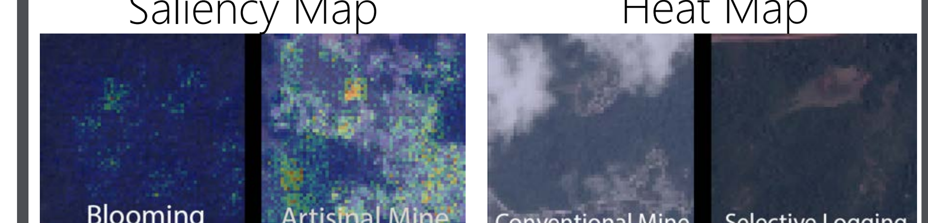
Saliency Map Heat Map



Inception fails to look for appropriate features in a given image (heat maps).
The saliency maps show that the network is looking at the wrong pixels for specific labels.
Inception is not suited for multi-label in this use case.

ResNet

Saliency Map Heat Map



The residual architecture acts too generally and performs poorly on specific features.
Heatmaps show inability to concentrate on many individual nuanced labels.

[1] Planet. **https://www.kaggle.com/c/planet-understanding-the-amazon-from-space**, Kaggle. 2017.
[2] F.Chollet. Keras. **http://github.com/fchollet/keras**, 2015.
[3] Ioffe, Sergey, and Christian Szegedy. "Batch normalization: Accelerating deep network training by reducing internal covariate shift." *arXiv preprint arXiv:1502.03167* (2015).
[4] Srivastava, Nitish, et al. "Dropout: A simple way to prevent neural networks from overfitting." *The Journal of Machine Learning Research* 15.1 (2014): 1929-1958.
[5] K. Rag. Keras-Vis. **https://raghakot.github.io/keras-vis/**, 2017.