Understanding the Amazon from Space with Multi-task CNNs

Allen Zhao, Chenyao Yu, Shawn Hu
{arzhao, chenyaoy, shawnghu}@stanford.edu

Background & Introduction

The Amazon Basin accounts for the largest share of human deforestation, contributing to reduced biodiversity, habitat loss, climate change, and other devastating effects. Machine understanding of satellite imagery can provide a better idea than traditional methods of how and where deforestation happens and can help governments and local stakeholders respond more quickly and effectively.

Problem Statement and Evaluation

Our objective is to build a neural network that classifies satellite images of the Amazon rainforest according to the atmospheric conditions (e.g, cloud cover), geography (e.g, rivers, lakes) and visible human impact (e.g, slash-and-burn, roads, mining, cultivation) in that location.
Each image may be assigned one or multiple of 17 labels, hence, our problem is a *multi-task-learning* problem.

We will evaluate our model's performance with the competition's *F2 metric:*

$$(1+\beta^2)\frac{pr}{\beta^2 p + r}$$

where $\beta = 2$, p represents the precision, and r represents the recall. This score is the weighted harmonic mean of p and r, with a preference for recall and hence aversion to false negatives.

Dataset

Our data comes from Kaggle and is sourced from Planet, a company that designs and builds the world's largest constellation of Earth-imaging satellites. The dataset consists of 40,479 labeled training images given in both JPG and TIF formats and a preliminary test set of 40,668 images. Labels occur with frequencies ranging from over 90% (primary) to less than 1% (mining, slash and burn, blowdown).

A few examples of the dataset are the following:

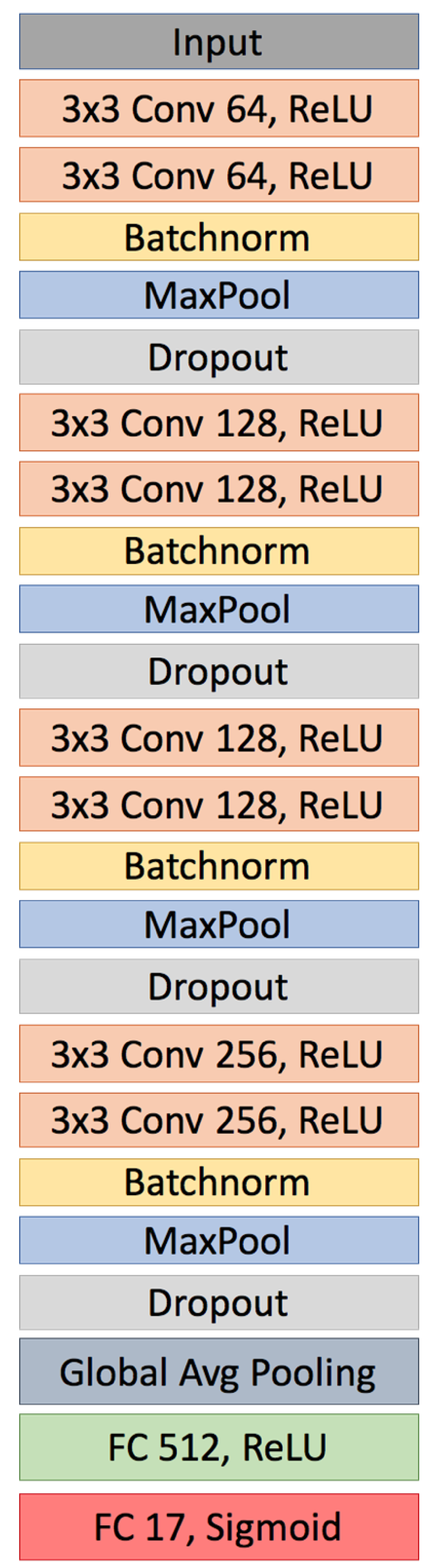
Label: blooming clear cultivation habitation primary slash_burn

Label: conventional_mine partly_cloudy primary

Label: cloudy

Label: agriculture artisinal_mine clear primary road water

Architecture



The first convolutional layer in each pair is padded such that the output size is equal to the input size. The second conv layer

All max pool layers are 2x2 pool stride 2 layers.

All dropout layers unlink 25% of all input neurons.

To handle the multi-class, multi-label objective, for each image we predict a score for each label and minimize the binary cross-entropy loss across all labels.

We determine whether a label is assigned based on whether the associated score exceeds a constant threshold. The value of these thresholds is found class-by-class by brute-force maximizing the F2 score on a validation set.

Conclusion

From our results, we conclude that standard multi-task sequential convolutional neural network architectures can perform quite well at the task of detecting atmospheric conditions and human impact on nature from satellite data. This has the potential to greatly reduce the amount of human effort needed to locate areas of deforestation anywhere around the world.

Future Considerations

- It may be helpful to augment our training data with simple image transformations to improve generalization.
- Our model does not currently exploit correlations between many of the labels - e.g, cloudy and hazy are mutually exclusive. It should be possible for a multitask network to learn these correlations, but given the limitations of our training set, it may help to explicitly learn them.

Experimental Evaluation & Findings

Our recent models achieved an F2 score of over 0.91 on Kaggle's preliminary test set, landing us in the top 100 of the leaderboard.

We tried a variety of changes to the architecture, but only some improved our results (shown in the table). Changes that did not help included: spatial pyramid pooling, additional FC layers, different layer depths, and increased dropout rates.

Architecture	Validation F2 Score
Starter Code	0.840
Improved Architecture + Hyperparameter Tuning	0.906
+ F2 Thresholding Heuristic	0.909
+ F2 Custom Loss	0.911
+ Batch Normalization	0.914
+ Global Average Pooling	0.917

References

[1] Rich Caruana. 1997. Multitask Learning. Mach. Learn. 28, 1 (July 1997), 41-75. DOI=http://dx.doi.org/10.1023/A:1007379606734
[2] Saikat Basu, Sangram Ganguly, Supratik Mukhopadhyay, Robert DiBiano, Manohar Karki, and Ramakrishna Nemani. 2015. DeepSat: a learning framework for satellite imagery. In Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems (SIGSPATIAL '15). ACM, New York, NY, USA, , Article 37 , 10 pages. DOI: https://doi.org/10.1145/2820783.2820816
[3] Jean, Neal, Marshall Burke, Michael Xie, W. Matthew Davis, David B. Lobell, and Stefano Ermon. "Combining Satellite Imagery and Machine Learning to Predict Poverty." Science. American Association for the Advancement of Science, 19 Aug. 2016. Web. 05 June 2017. <http://science.sciencemag.org/content/353/6301/790>.
[4] Jialei Wang, Peder A. Olsen, Andrew R. Conn, and Aurelie C. Lozano. 2016. Removing Clouds and Recovering Ground Observations in ... (April 2016). Retrieved June 5, 2017 from https://arxiv.org/pdf/1604.03915v1.pdf
[5] Emmanuel Maggiori, Guillaume Charpiat, Yuliya Tarabalka, and Pierre Alliez. 2016. Recurrent Neural Networks to Correct Satellite Image Classification Maps. IEEE Transactions on Geoscience and Remote Sensing (August 2016), 1–10. DOI:http://dx.doi.org/10.1109/tgrs.2017.2697453
[6] Ye, Nan, Kian Chai, Wee Sun Lee, and Hai Leong Chieu. "Optimizing F-Measures: A Tale of Two Approaches." *Proceedings of the 29 Th International Conference on Machine Learning* (2012)
[7] Lipton, Zachary C., Charles Elkan, and Balakrishnan Naryanaswamy. "Thresholding Classifiers to Maximize F1 Score." Thresholding Classifiers to Maximize F1 Score - Open Access Library.