CS231n Project Proposal

Brian Rohr

Vishal Subbiah

Brent Lunghino

April 23, 2017

Applying convolutional neural networks to label atmospheric conditions and landcover on satellite images of the Amazon rainforest. We will train our convnet to detect the cloudiness of an image, and what types of landuse are in the image, for example primary forest, water, or agriculture. This is an important challenge because it will improve automated classification of large satellite image datasets. Classified satellite data of the rainforest can be applied to compute deforestation rates, detect illegal mining/deforestation, and differentiate between natural and human-caused deforestation. Data will be provided by Planet for a Kaggle competition. The dataset is 150,000 images that are 256x256 pixels and 4 bands of color (RGB + near-infrared). Each image is 947 m x 947 m for a resolution of 3.7 m per pixel. The dataset covers 30 million hectares of Amazon rainforest.

We will train a convnet and use an F-score as our loss function

$$F_\beta = (1 + \beta^2)\frac{pr}{\beta^2 p + r} \quad \text{where} \quad p = \frac{tp}{tp + fp}, \quad r = \frac{tp}{tp + fn}, \quad \beta = 1 (F_1 score). \tag{1}$$

tp - true positive

fp - false positive

fn - false negative

p - precision

r - recall

Since the project, is based on a Kaggle competition we will be able to benchmark our work with others who have submitted solutions from around the world. Since every image could belong to multiple classes, we could study the correlation between such classes and if there is a causation to be identified.

tests: cross validation, and benckmark comparisons with other participants