Capstone Proposal

Schuman Zhang
January 8th, 2018

Domain Background

Deforestation is a global problem that humanity faces in the 21st century. Every minute, the world loses an area of forest the size of 48 football fields. Such widespread deforestation leads to reduced biodiversity and climate change. As such the ability to quickly and accurately pinpoint areas of deforestation will enable organizations and governments to respond in a timely and effective manner.

‘Planet’ is an earth-imaging satellite that collects daily imagery of Earth’s entire land surface at 3-5m resolution. Using Planet’s database of satellite images, we can track changes to natural and man-made features on the surface of the Earth. By combining Planet’s dataset with deep learning/computer vision capabilities, we can accurately detect areas of deforestation (both natural or otherwise) and thus enabling governments to response effectively.

An example of using convolutional neural networks or deep learning techniques to analyze satellite imagery can be found here - https://arxiv.org/pdf/1704.02965.pdf.

Problem Statement

Correctly classify and detect relevant features in Planet’s vast dataset of images of the Amazon basin, with the goal of tracking areas of man-made deforestation. This is a classification problem where the input will be individual images and the output will be features detected within the image, such features include clouds, haze, rainforest, water, habitation, agriculture, roads and cultivation etc. By detecting relevant features in satellite images to a high level of accuracy, we can automatically keep track of these features over time without the need for human analysts to manually go through thousands of satellite images.

Datasets and Inputs

The relevant dataset for this problem comes from Planet’s full-frame analytic scene products. The images in the dataset were taken from satellites in sun-synchronous and ISS orbit, and taken between January 1, 2016 and February 1, 2017. The ‘scene’ in our dataset captures the Amazon basin, an area spanning Brazil, Peru, Uruguay, Colombia, Venezuela, Guyana, Bolivia, and Ecuador. The scene is then broken down into individual 256 x 256 ‘chips’, which form the individual images in the dataset.

The dataset for this problem consists of 44,078 chips, each chip is labelled with the help of human experts and crowdsourced labor. There are a total of 16 classification labels, and each image may be tagged with multiple labels. The dataset provided via Kaggle is already split into a training set and a testing set. While training the CNNs, the training set will be further split into training and validation sets. The relevant classification labels are cloudy, partly cloudy, hazy, primary rainforest, water, habitation, agriculture, road, cultivation, bare ground, slash & burn, selective logging, blooming, conventional mining, artisinal mining, and blow down.

The source of the dataset can be found here -
https://www.kaggle.com/c/planet-understanding-the-amazon-from-space/data

Solution Statement

This is a supervised learning and multi-class classification problem in the field of computer vision. Convolutional neural networks (CNN) have shown tremendous promise in this field. The solution to is conduct transfer learning on several pre-trained CNN architectures, including Resnet50, Resnet101 and Resnet152. Weights from these pre-trained models will be transferred to a new CNN specifically used for classifying aforementioned satellite images.

Benchmark Model

A simple or naive benchmark model to compare against is a model which randomly guesses the labels of the images. There are a total of 16 labels and each image may contain one or more labels. The CNN proposed in the solution can then be compared to this model to ensure that a trained neural network can significantly outperform random guessing.

Evaluation Metrics

The evaluation metrics is to use F2 score. This balances recall and precision but a greater weight is placed on recall over precision. The formula for this metric is as follows:

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

Where *p* is precision, *r* is recall and *beta* = 2.

Project Design

The solution to this project involves building and training a convolutional neural network (CNN). Once the network is trained and the optimized weights are obtained, the model will be evaluated against the benchmark model as well as selected accuracy metrics. Once a satisfactory level accuracy is achieved, server-side code will be written to handle input images and produce a prediction based on the best model, and the result of the prediction will be displayed in a simple web-based application.

In terms of the deep learning aspect, a number of models/networks will be trained. Firstly a CNN built from scratch will be trained, and a number of different architectures will be tested. This will be a good chance to experiment with a number of hyper-parameters. Next transfer learning will be used on Resnet50, Resnet101 and Resnet152. Most of the layers used in these state-of-the-art architectures will be used and the last few layers will be altered for the purposes of this project. All trained networks will be evaluated against the test set and the optimal one will be chosen.

In terms of Infrastructure, GPU instances on AWS will be used to train a number of CNN architectures. Once the best one is selected, a separate EC2 instance on will host our server-side code with a number of endpoints. These endpoints will display the best CNN metrics, parameters and characteristics, as well as handle requests for producing new predictions based in new input images.

For the presentation of the project, a web-based application, written in either React or Angular will display the result in a user-friendly manner. This simple web application will have 2 parts. The first part will display the metrics of selected models, and the second part will enable the user to upload new images, and a prediction will be made for objects/features detected in that particular image.