Ivan Cano, Byungheon Jeong

Professor Schwartzman

DSC 180a A16: Satellite Imaging

December 12, 2021

Brief Overview of Quarter 2's Project

Introduction

One of the great limitations of macro analysis of human activity is that we have historically been limited to a subset of the population, using statistical methods to design surveys within an acceptable error margin. This is certainly reasonable as attempting to manually count every single business or family or acre of land would be a singularly uneconomical usage of resources. However, this abstraction causes cracks, especially in relation to public policy, into which the most vulnerable and powerless members of our planet fall. This can be seen when surveyors cannot reach those without adequate access to telecommunications infrastructure. Happily, the advances in image processing in satellite imagery and machine learning has presented a tantalizing possibility for truly analyzing the entire population instead of a subsample of it. After all, nearly human activity occurs on the surface of the planet and satellites are directly recording that activity.

Indeed, the field of urban activity detection has been very active in the last two decades, and current literature uses mixtures of cutting edge algorithms and pattern recognition systems to identify the human footprint on the earth.

Problem

However, the subfield that directly deals with illegal, or unsanctioned urban development is not as developed in using cutting edge methods. In quarter two, our team will be using these tools to address one of the areas of most significant potential: measuring urban growth. Though in many wealthy countries urban growth occurs in areas set forth by urban planners, the poor in many countries are unable to afford legal developments. Therefore, many settle in makeshift urban territories by building their own buildings and homes.

In order to reduce discriminatory data-collection and help bring about a fuller and more equitable data collection, we will be developing a classification service that will predict the urban growth of areas in unsanctioned areas by analyzing satellite images alongside relevant statistical data, in order to test the reliability of statistical surveys of these urban areas. We may also be looking at using time series analysis to predict future potential growth but at the moment we are not certain whether such methods could work in these contexts.

Using our data pipeline and implementing cutting-edge algorithms currently used in the field, we aim to automate the generation of time series data concerning urbanization and unofficial settlements.

Additionally, in our initial exploration of this topic, we have struggled to find open source labeling services that deal with satellite imagery. Rather, most use curated datasets such as spacenet. We want to address this gap by creating a service, whether in the form of an open-source github project or a cloud based Saas solution, that allows researchers to easily generate datasets from google earth engine.

Solution Process

Just like our quarter 1 project we will be re-implementing techniques and datasets that have previously been employed, such as our use in the landsat 7 satellite image repository and also adapting new ones. The reason we have chosen to continue using landsat 7 is because it provides a reliable set of images that span from 1999 to present day.

We will also be re-employing some data processing strategies we have employed in our quarter 1 project, such as the conditional use of different EM bands on satellite images. Just as the different bands of the electromagnetic spectrum helped us distinguish ice and non-ice objects on mountain ranges, we hope that using a similar technique will aid us in tracking the growth of unsanctioned urban centers.

Some adjustments to our implementation, such as the use of the NDSI (Normalized-Difference Snow Index), will be adjusted to urbanization specific metrics as we will no longer be tracking snow or ice. Instead one possible band that we could implement is NDVI (Normalized Difference Vegetation Index), and something equivalent for concrete and asphalt.

Outputs

One of the prime outputs of this project is a versatile labeling tool that works specifically with Google Earth engine's landsat data. Using an augmented image band, the labeling service will segment the portion of the satellite image and use that label to generate a multi-dimensional data set, with each dimension consisting of a band of the electromagnetic spectrum. This tool should be able to take in parameters directly from Google Earth engine and provide satellite imagery data sets. Our current prototype works only with glacier images and can support two labels: ice and no ice. In our final implementation, we want to support as many labels as the researcher wishes. Next quarter we will be working to universalize this prototype into a production ready labeling solution for Google Earth engine.

Another product would be trained models that can take in satellite imagery of unsanctioned urban areas and label urban footprint. At the current moment our ice identification models work with pixel-wise data which does not retain spatial information. That is why we were limited to linear models and tree models to classify our images. In quarter 2 we aim to support spatial information within our data set so that we may be able to use models that use convolutions or clustering algorithms.

Finally, we aim to build time series models that can take in the information from the aforementioned labeling models and translate a series of them into historical analysis and

predictions on its future. For the moment, this part remains the most theoretical of our project proposals though we are excited to delve deeper into the literature and explore ways of achieving these goals.

Progression

Ivan's progression for Week 2:

This week I read two academic papers that explored using different filters to track the growth of urbanization. These two academic papers were, "A Genetic Algorithm-Based Urban Cluster Automatic Threshold Method by Combining VIIRS DNB, NDVI, and NDBI to Monitor Urbanization," and "The Environmental Effects of Urban Development in Hanoi, Vietnam from Satellite and Meteorological Observations from 1999–2016." These papers were useful in bringing to light bands that my team had not previously considered, such as NDBI and NTL. I also dedicated some time to research the different gdp growths of developing regions of the world, such as Mexico, El Salvador, Vietnam, and China. I concluded that China is an optimal place to test our project on since it has had some of the most rapid and impressive growth in modern history and would therefore have a clear progression of development throughout the landsat images we will be using when classifying this urban development.

Josephs' progression for Week 2:

This week I worked on refactoring last quarter's codebase into something that can form the foundations of our final project. Specifically, I burned out the scratch code and things like jupyter notebooks and arranged the code into modules. Finally, I created a new repo to store the code.

Also, I was working on adding new features into the application so that we can generalize our labeler. As of now, only 'snow' and 'not-snow' labels can be used. The labeler must be able to accept any number and types of labels. As of now, I have not been able to finish it, but have found the scope of the problem and the modules that I need to make.

In regards to challenges, getting the generalized labeler to work has been frustrating, as the data structure that should be used is something not yet figured out. I was thinking of using a queue but a hash map may work better.

Finally, as the underlying labeling application is another open source project, I am in a bit of a pickle in regards to how to arrange my wrappers.

Projects Tasks and Challenges

Operation of front to end google earth engine labeling service

Tasks

1. Adjust service to take multiple labels
    a. Make labeling service accept multiple labels · Issue #2 · byungheon-jeong/DSC180A-Q1 (github.com)
2. Move alpha to new public repo
3. Connect service with GEE
4. Figure out how to move service on remote service (as a download or web-based?)
5. Build CI/CD pipeline
6. Deploy on Nautilus or AWS

Challenges

1. How to structure script modules?
2. Where to put tests?
3. Integrate db?