INTRODUCTION AND BACKGROUND

Currently, the SOC Team tasks high-resolution satellite imagery providers to supply county-scale imagery for their projects, including tracking the progress of construction. Since the team works with imagery from counties and regions, this amounts to an enormous amount of data to purchase, process, and analyze. To illustrate this further, consider a single target area, like Harris County in Texas. To get data for this county, the SOC Team will need to procure 1,780 square miles of imagery. This is expensive for many reasons, including the cost to obtain the data, the time and processing cost for storage and retrieval, and cost in processing the imagery. Even with all the data coming from a large metropolitan area, there will be vast swaths of the landscape that will not get developed. This leads to unnecessary captures of high-resolution imagery. In the case of Harris County, construction activity may only be around 200 square miles, or just over 11% of the total area. Clearly, the SOC Team's current data procurement and processing strategy involves a large outlay of resources, some of which is going to waste.

Problem Statement

There is a need to limit the extent of the proposed capture area so that the SOC Team can focus their monetary and compute resources on regions that are actually undergoing construction.

METHODOLOGY

To help the SOC Team, this project will require the use of satellite-derived orthoimagery, image modeling tools like PyTorch, and libraries that are optimized for handling geospatial information.

I intend to use the Python API for Google Earth Engine (GEE) to collect Sentinel-2 imagery. GEE is free to use with approval and has open access to Sentinel-2 data, along with others like Landsat, MODIS, and NAIP. Sentinel-2 imagery's visible and near-infrared (NIR) bands have a resolution of 10m (all other bands have resolutions larger than 10m). NIR is useful for calculating NDVI (Normalized Difference Vegetation Index). Areas with high NDVI values are those that are rich in vegetation. My assumption, implicit with using NDVI, is that areas with low- to no NDVI values are those that are undergoing excavation and thus all measurable plant life is removed.

To start with a simple baseline approach, I will train a PyTorch model that performs gross image differencing and thresholding. I will then bring in spectral data with NDVI which can help find areas where vegetation is cleared in preparation for excavation and construction; i.e., finding areas where the NDVI is low indicates places that are devoid of healthy vegetation.

I will need to account for seasonal changes, as these may appear as excavation, when in fact it is a dry season. This can be a later stage in developing the model's capabilities. Further refinements will involve deep learning segmentation.

Supporting Tools

Besides GEE to retrieve orthoimages, I will take advantage of PyTorch's image models to segment satellite imagery for candidate construction zones. Sklearn may be useful for clustering or feature extraction tasks.

Libraries like H3 and Rtree offer efficient geospatial indexing. Using DuckDB with Parquet files is fast and efficient at querying data without the need for a separate database to store everything. I should mention that the geospatial toolkit is rapidly evolving and the choice for one tool over another may only become clear once the project is underway.

TIMELINE AND MILESTONES

This project will be completed in two months with three distinct phases. There will be a clear deliverable produced at the end of each phase.

Initial Phase - Weeks 1 - 3

- Literature review
- Acquisition of image data and pre-processing
- Replication of baseline methods from prior work(s)
- →Deliverable: progress report on initial results

Development Phase - Weeks 4 - 6

- Implementation of change detection algorithms
- Experimentation with different segmentation approaches (from thresholding to more complex deep learning segmentation)
- →Deliverable: report detailing performance of models

Refinement and Testing Phase - Weeks 7 - 8

- Fine-tune model to improve results
- Benchmark performance
- →Deliverable: final report on task and code delivery

EXPECTED OUTCOMES

At the end of the two-month period, I will provide the SOC Team with a proof-of-concept model that can identify candidate areas likely undergoing construction, given an area of focus.