Semantic Segmentation of Aerial Imagery

Nikita Jayakar[1] **and Richanshu Jha**[2]

[1]**nmj303@nyu.edu**
[2]**rj1469@nyu.edu**

PROBLEM STATEMENT

These days, there are many applications which require the involvement of creation of geospatial representation of cities. It is possible to capture high resolution satellite images from satellites or drones for the goal of collecting this data. However, if one has to manually classify or segment the cities based on roads, buildings and green cover, it becomes a really time consuming task prone to human error. This applies not only to roads and buildings (The primary scope of this project), but also vegetation cover. This information is essential to generate the mapping of buildings and roads, and vegetation data is used to calculate the percentage of greenery in any city/geographical area and essential in forming various environmental metrics for the area. While vegetation in a city/geographical area may be found out reliably by using RGB-based thresholding algorithms, this is not true for roads/buildings. Due to this, deep learning image classification algorithms, such as Convolutional Neural Networks(CNN), become essential tools for satisfactory and feasible classification of roads and buildings.

LITERATURE SURVEY

The aim of this project is to target especially the urban geographical areas to analyze the mapping of roads, buildings and green cover. (Muruganandham S., 2016) It is a very useful technique to use deep learning to build image classifiers based on Fully Convolutional Networks(FCNs). Modified versions of CNNs can also be used for regularization, optimization and achieving higher model accuracies of the image classifier. Hence, an image classifier can be built on satellite images using deep learning techniques like CNNs. (Wurm and Taubenböck, 2019) Since analysis is to be done on urban areas, there is a journal paper which focused on a very specific use-case for analysis of urban cities. It focused on semantically segmenting the images of slum-areas which is also a big problem in some of the urban cities in the world. This segmentation on a very specific domain was performed using transfer learning on deep learning FCNs from one dataset to another. (Li and Yu, 2019) U-Net-based semantic segmentation can be used to get very less difference between the validation dataset and the actual required output of the building footprint extraction from the satellite images data. Building footprints can be classified and extracted quite accurately from the given images dataset. (Ishii T. and R., 2016) It is also vital to perform tuning on the threshold to be applied for getting optimal performance on the semantic segmentation of buildings after applying FCNs on the images.

DATA SET GENERATION

We did a fair amount of research on the internet with regard to the data sets available for our project. While segmented Geo spatial imagery data is readily available online, there are lots of issues with combining the (relative to the task) small data sets. Some are focused on rural areas, some on urban areas, and there is a lot of variation on file formats, sizes and resolutions.We have opted to stray from the convention of using pre-existing data and will be creating the data set as part of the project. For the task of creating the data set, our ground-truth will be extracted from google maps. Considerable progress has been made on the data extraction pipeline which has 3 modules (Google maps module, Image Extraction module, data preparation module) been presented in detail below.

The first module will be a browser-based google maps module that will have two instances of Google Maps open on the same screen. This will be done by calling Google APIs on a browser window. One instance of maps will display the satellite image of a predetermined area, and the other instance of the map will be programmed to show the roads and buildings of the same location. Segmenting all buildings into a single color channel was not possible using Google APIs. However, it is possible to configure the map such that they can be extracting via thresholding RGB values. These maps will be configured to follow a predetermined straight line path through a selected area. Both these maps will run simultaneously giving the satellite imagery and labelled roads/buildings view side by side. The image extraction module would be a simple Python command line program with the capability of taking screenshots or window captures of the Google maps module. These images would be taken at regular intervals based on the the panning speed of the google maps module and would perform basic image splitting and cropping to provide two raw images per screenshot (One for the satellite imagery and one for the

road/buildings view). The image extraction and google maps modules would run simultaneously to provide this raw data set.

The data set preparation module will be run after a significant amount of raw data has been collected. In this module, the raw pairs of images generated from the previous step will be processed here as follows. The 'road/buildings views' image will be threshold against various sets of values to extract the data of roads and buildings. The final format will be a greyscale image where pixels containing roads and buildings will be encoded as two discrete intensity values. The image and label would then undergo a process of data augmentation where the images are rotated, flipped etc. which will further increase the data size. We have built a working prototype of the Google maps and data extraction module. We have also tested the feasibility of extracting labelled data by applying RGB-threshold the outputs from these modules and were able to produce labelled data reliably.

MODEL

The primary deep learning task that this project is involved with is semantic segmentation of images. Thus, after considering multiple options such as Gated-SCNN, YOLACT, U-NET, Fast FCN etc. We have decided to opt for the U-NET Architecture. We will use PyTorch to implement this model. While this is not final, we choose to perform our initial experiments on this model. While the size and number of layers will be concluded after our initial tests, A simplified structure of the architecture pictured below:

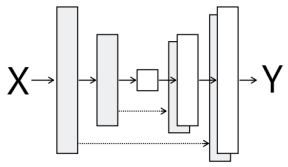

Figure 1. U-NET Model:(Zhao et al., 2019)

The input to this model will an image with the shape (MxNx3) Where MxN is the pixel size of the image. The image input will be in RGB, hence the 3 channels. The output of the model will be a label mapping of the image with the shape (MxNx1), each pixel having a value specifying its classification: Roads, buildings or neither. Given the feasibility of the task, this output will be complemented with a 4^{th} distinct value signifying vegetation and computed via Computer Vision-based non-ML techniques.

ANTICIPATED OUTCOMES

We expect that by the end of the project, we will have built a reliable satellite data extraction pipeline and trained a Convolutional Neural Network to semantically segment the satellite data from it into roads and buildings. If feasible, to give a more complete analysis of the location, we would complement that result with vegetation data we generate for the same input images via non-DL techniques. We anticipate some issues with the training of the model due to the high processing power/time needed to train such models. To alleviate this, we intend to train the model on Amazon's cloud services. As a backup, we have access to training the model on a machine with fair specs (Nvidia RTX 2080 GPU / 9th Generation intel processor). Another issue could be that of accuracy given the large variety of satellite data available. In order to keep a reasonable target for our model, we have constrained the scope of the project to classification of roads/buildings in urban/suburban environments only.

REFERENCES

Ishii T., Simo-Serra E., I. S. M. Y. S. A. I. H. and R., N. (2016). Detection by classification of buildings in multispectral satellite imagery. *23rd International Conference on Pattern Recognition (ICPR) Cancún Center*, (3344):3344–3349.

Li, W., H.-C. F. J. Z. J. F. H. and Yu, L. (2019). Semantic segmentation-based building footprint extraction using very high-resolution satellite images and multi-source gis data. *Remote Sensing*, (403):1–19.

Muruganandham S., R. M. (2016). Semantic segmentation of satellite images using deep learning. *Faculty of Electrical Engineering, Department of Cybernetics*, pages 51–62.

Wurm, M., S. T. Z. X. W. M. and Taubenböck, H. (2019). Semantic segmentation of slums in satellite images using transfer learning on fully convolutional neural networks. *ISPRS Journal of Photogrammetry and Remote Sensing*, (150):59–69.

Zhao, H., Qingyun, Y., Song, S., Ding, J., Lin, C.-L., Liang, and Zhang, M. (2019). Use of unmanned aerial vehicle imagery and deep learning unet to extract rice lodging. *Sensors*, 19:3859.