# Using external knowledge for geographically-intelligent semantic segmentation on satellite images

**Hayley Song (USC ID #: 52760478267)**

## 1. Problem Description

As a final project for CSCI 699, I propose to develop a semantic segmentation algorithm that incorporates external information and geospatial reasoning to improve a specific visual task of extracting road networks from satellite images. My motivation to integrate external knowledge into the segmentation pipeline comes from how humans perceive a visual world. In interpreting what we see, we use not only the visual patterns acquired from seeing millions of images in the past but also our knowledge about the world that we have acquired from other sources and activities such as reading a book and interacting with physical objects. As humans, this external knowledge is crucial in understanding what we see and achieving, even with few data, our level mastery on various visual tasks. My project will focus on the task of detecting roads in four areas (Vegas, Paris, Shanghai and Khartoum) using the Wikipedia and the OpenStreetMap as sources of external knowledge in a structured form. My main goal is to build a model that learns to extract relevant knowledge from the external databases and use it to learn features that are geospatially meaningful in addition to visual patterns as other CNN models would do. This model would generalize better to different geographic regions and would be able to extract knowledge most relevant to the road detection task as a natural byproduct of its learning process.

## 2. Example Scenario

Figure 2 illustrates how we might use our knowledge about the world in detecting roads from a satellite image. If we know that the picture we are looking at the picture is taken from Paris, then we have some expectations on what road networks will look like (urban, tiled roads as opposed to unpaved grounds). We also have intuitions about how roads behave through our physical experiences: roads are continuous; highways may appear layered, but city streets should not. We also have understandings on how roads and other objects interact: if a road is next to houses, it is likely to be a residential type. If it’s very close to a hospital or a mall, it is likely to be a service road. In between two blocks of small buildings, there is likely to be a street, but a highway is less likely. Near a plaza, there is likely a pedestrian way.

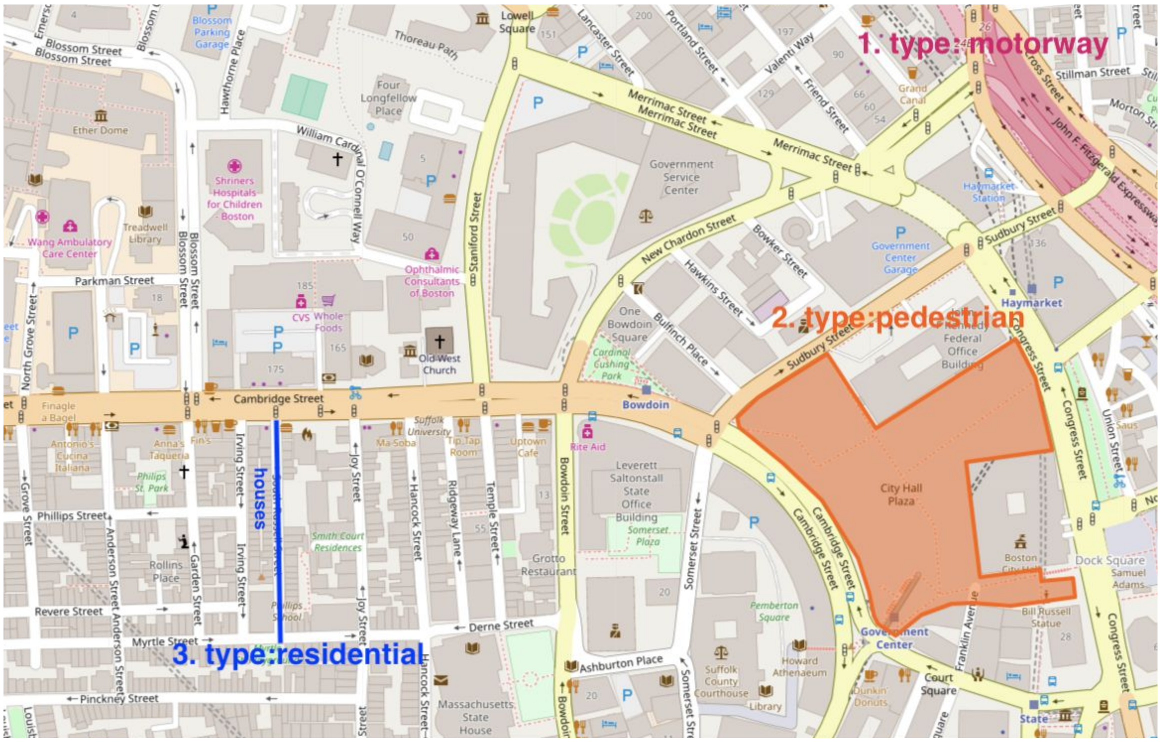


Figure 2a: Example of road types and geospatial features from the OpenStreetMap database

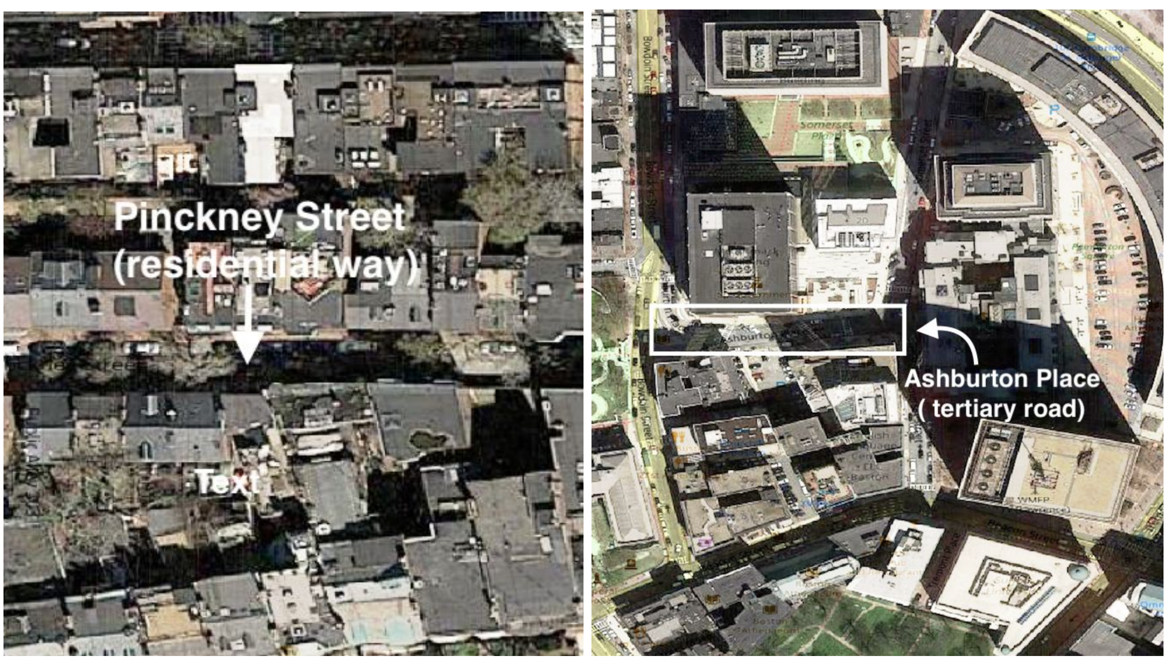


Figure 2b: OSM data overlayed on satellite images. Geospatial features such as the type of nearby buildings and the spatial layout of the surrounding objects can help to detect roads

Figure 2: Example of how geospatial knowledge and statistics that could help detect roads and distinguish the types

## 3. Dataset

I will use the SpaceNet challenge dataset [1] collected from DigitalGlobe which covers over 8000 km of roads from four areas of interest (AOIs): Vegas, Paris, Shanghai and Khartoum. Each area dataset comes with two types of data, the raster images for multispectral satellite images and vector files for the annotated road networks. First, each satellite image contains 8 bands acquired from DigitalGlobe's WorldView-3 satellite with the resolution of 30cm. The vector data for road labels are manually annotated and quality-controlled as instructed in the challenge's guideline [2]. It contains seven road types: motorway, primary, secondary, tertiary, residential, unclassified and cart track. Please refer to [1] for more details on the dataset. Below, Figure 3 shows sample images from the SpaceNet challenge [1].

Figure 3a: Paris

Figure 3b: Shanghai

Figure 3c: Vegas

Figure 3d: Khartoum, Sudan

Figure 3: Sample images from the dataset. Coloured lines indicate the manual annotations

In addition to the SpaceNet challenge dataset, I have satellite images from two other cities in South Sudan: Runkek and Malakal. As these cities do not resemble any other cities geographically, it will be interesting to see the model's performances in these regions. Unfortunately, there is no reliable label, so I will use them for a visual, qualitative evaluation of the model's generalizability. Figure 4 shows samples from Runbek and Malakal datasets.

Figure 4a: Runbek, South Sudan Figure 4b: Malakal, South Sudan

Figure 4: Sample images from Runbekk and Malakal, South Sudan

## 4. Evaluation Metrics

I propose to use the following two metrics to evaluate the segmentation performance.

- Pixel-wise difference evaluation (F1 score)
- Average Path Length Similarity (APLS)

The pixel-wise difference is not an ideal measure if our application is concerned about road connectivity. For instance, in a humanitarian planning, the pixel-level precision is less important than extracting correctly connected road networks. The second metric based on graph theory complements the limit of pixel-based metrics by incentivizing the creation of a continuous, valid road network. Figure 5 compares the pixel-based F1 scores on three detection results and demonstrates the limit of the pixel-based metrics. APLS gives a higher score for the second image. Please refer to *Van Etten, A et al* [3] for more details.

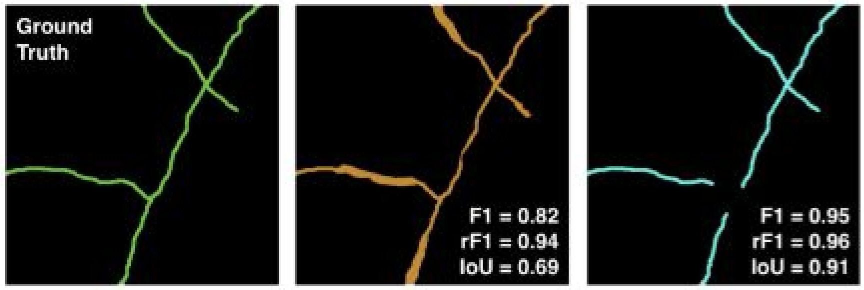


Figure 5: Limit of pixel-based metric. APLS favors the middle result over the rightmost result

## 5. Plan and Timeline

The main components of my model are (1) learn to extract relevant information from the Wikipedia and the OSM knowledge base, and (2) use that information to perform the segmentation task.

- First, as a preprocessing step, I need to organize the OSM knowledge base to a more organized graph structure so that I can implement graph searches efficiently.

- Enumerating through the entire knowledge graph (KG) and evaluate each entity to compare its relevance is not computationally feasible. I propose to use a Graph Search Neural Network as a way to walk through the KG and learn a per-node "important" score. I will investigate several networks such as *Gated Graph Sequence Neural Networks* [4] and *Graph Partition Neural Networks* [5]. *Marino et al* [6] tried a similar approach to image classification and showed improvements over previous works.

- I propose to use memory networks [7] with attention [8] over the OSM knowledge base to select relevant information.

Here is my timeline.

1. Establish baseline models -- [10/10 ~ 10/15]
    a. UNet, ResNet
    b. UNet, ResNet + transfer learning
2. Build a knowledge graph from the OSM database for the purpose of this project -- [by 10/20]
3. Graph Search Neural Network for efficient search over the knowledge graph and node selection; and/or attention model -- [by 10/24]
4. Incorporate into a single pipeline --[by 11/20]
5. Evaluation -- [by 11/30]
6. Finalize the write-up (Final Due: 12/09)

## 6. References

[1] SpaceNet on Amazon Web Services (AWS). “Datasets.” The SpaceNet Catalog. Last modified April 30, 2018. Accessed on October 10, 2018. https://spacenetchallenge.github.io/datasets/datasetHomePage.html.

[2] SpaceNet on Amazon Web Services (AWS). “The SpaceNet Roads Dataset labeling guidelines” Last modified April 30, 2018. Accessed on October 10, 2018. https://spacenetchallenge.github.io/assets/docs/SpaceNetRoads_labeling_rules_v2.html

[3] Van Etten, A., Lindenbaum, D., & Bacastow, T. M. (2018). SpaceNet: A Remote Sensing Dataset and Challenge Series. Retrieved from http://arxiv.org/abs/1807.01232

[4] Li, Y., Tarlow, D., Brockschmidt, M., & Zemel, R. (2015). Gated Graph Sequence Neural Networks. Retrieved from http://arxiv.org/abs/1511.05493

[5] Liao, R., Brockschmidt, M., Tarlow, D., Gaunt, A. L., Urtasun, R., & Zemel, R. S. (n.d.). *GRAPH PARTITION NEURAL NETWORKS FOR SEMI-SUPERVISED CLASSIFICATION*. Retrieved from https://www.microsoft.com/en-us/research/wp-content/uploads/2017/11/gpnn_arxiv.pdf

[6] Marino, K., Salakhutdinov, R., & Gupta, A. (n.d.). *The More You Know: Using Knowledge Graphs for Image Classification*. Retrieved from https://arxiv.org/pdf/1612.04844.pdf

[7] Weston, J., Chopra, S., & Bordes, A. (2014). Memory Networks. Retrieved from http://arxiv.org/abs/1410.3916

[8] Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural Machine Translation by Jointly Learning to Align and Translate. Retrieved from http://arxiv.org/abs/1409.0473