Open GeoSpatial Data as a Source of Ground Truth for Automated Labelling of Satellite Images

Marjan Alirezaie, Martin Längkvist, Andrey Kiselev, and Amy Loutfi

Applied Autonomous Sensor Systems, Örebro University, Sweden,
{marjan.alirezaie,martin.langkvist,andrey.kiselev,amy.loutfi}@oru.se

Abstract. In this paper, we summarize the practical issues met during our interaction with OpenStreetMap for the purpose of automatically generating labelled data used by data classification methods.

1 Introduction

Despite the recent advances in qualitative processing of geodata, scene parsing (labelling) in the field of geo remote sensing is still relying on data driven methods. More specifically, augmenting raw data (e.g., satellite imagery data) with a layer of qualitative features is not possible unless the data is initially processed by classification (labelling) and segmentation methods. On the other hand, a challenge in supervised learning methods which are known as a robust classification technique, is to provide a labelled training set (i.e., a set of data associated with ground truth[1]). Moreover, even if the imagery data is already classified, we still need to provide enough (labelled) data for further purposes. This is particularly true for dynamic changing datasets for example, a recently flooded city. In general, a ground truth used in a classification process can vary in size and structure depending on the classification method. For instance, given a satellite imagery scene containing several objects, a pixel-based classification method may need a ground truth in the scale of the scene's number of pixels, whereas a ground truth in the scale of the number of objects in the scene suffices for an object-based classifier.

Among the methods used for computer vision tasks, Convolutional Neural Networks (CNNs) [1] as a recently emerging technique has shown a good performance in scene parsing [2] due to its adaptability to the input data. CNNs are trained with supervised learning and therefore require a large amount of labelled data to learn the model parameters. Providing a reliable and precise ground truth for large city-wide size satellite imagery data on a pixel-level is non-trivial and time consuming.

In our previous work [3] we manually annotated the satellite imagery data of Boden[2] with rudimentary labels such as *building*, *road*, *railroad*, etc. In order to provide the ground truth, we developed an interactive software tool for the labeling process. In order to speed-up the process, the data was segmented into regions using a standard image segmentation method. However, the process was

[1] In remote sensing, *ground truth* refers to information collected on location.

[2] Boden is a small city located in north of Sweden.

still time consuming and there was no guarantee the labels were 100% accurate due to the human errors associated with manual work.

In this paper, we summarize our experience on using online geo information as a help for data-driven analysis methods. Our focus is on the content of OpenStreetMap®[3] to share our experience with the community whose concern is about the quality of the representation of open spatial data.

2 OpenStreetMap® in Ground Truth Generation Process

OpenStreetMap® (OSM) contains a large amount of actionable information in the form of roads, buildings, landscapes, etc. This information is provided from different sources including manual imports by human users. There is a representational (vector-based) layer underlying the data that facilitates the process of map annotation with spatial (meta) data in OSM. More specifically, each piece of information, as an OSM data primitive, is represented as a vector indicating a point, line or polygon. In the following we briefly explain how we exploited the OSM contents to generate the ground truth used in our scene labelling process.

2.1 Parsing OSM Data

The second imagery data set processed for our scene labelling purposes belongs to an area as large as about 67 km^2 (16384 × 16384 pixels with 0.5 meter resolution) located in central part of Stockholm. The ground truth generation process starts by extracting the OSM-XML file containing information related to this specific area. This file represents points (of interest) as well as non-point shapes (i.e., *way*) as a stream of coordinates which indicate lines and polygons in a similar way[4]. Each data point provides information for merely a single point (equivalent to a pixel) regardless of the shape and the size of the structure located there in reality. We therefore only investigate polygons and lines.

The set of rudimentary labels chosen to label the scene includes *Building*, *Ground*, *Vegetation*, *Parking*, *Water*, *Road* and *Railroad*. The list of shapes including lines and polygons that are extracted from the osm file is further investigated to categorize the items into one of the aforesaid initial labels. This categorization process is accomplished by checking the type of amenity[5] assigned to each shape. Given the list of labelled shapes, the ground truth generation process would be ideally completed if all the internal pixels of each shape could be effortlessly labelled with the same label assigned to the shape. However, the process of filling polygons with labels pixels (i.e., transforming vector representation to raster representation) needs to deal with a number of issues stemmed in the representational structure of geodata online sources. These issues are summarized in the following:

Polygons versus Lines: A polygon is a suitable geometrical shape to represent each labelled region on the map. In OSM, however, roads, rail-roads and in general any structure categorized as a *path* are represented in the form of a

[3] http://www.openstreetmap.org/

[4] The shape is a polygon only if the first and last coordinates are the same.

[5] It is a description assigned to each tag of data that can indicate its affordance. For further information check: `<http://wiki.openstreetmap.org/wiki/Key:amenity>`

line (and not a polygon), lied in the middle of the structure. In order to reduce the false negative labelled pixels (i.e., pixels of a road which are positioned out of its representative vector-line), we therefore need to make a raster representation of the road by widening each road line from its both sides to transform it into a polygon. However, the roads are available in different width size. On the other hand, extra meta data indicating the exact width of roads lacks. Although statically widening roads with a fix number of pixels can reduce the rate of false negative labelled pixels, it can give rise to the false positive rate at the same time. The second solution to this problem is adjusting the width of the polygon according to the type and location of the road.

Filled versus Non-filled polygons: Paths in OSM are not always represented in the form of lines. There are some paths shown as a closed line identified as a polygon. However, there is a conceptual difference between a polygon representing a non-path structure (e.g., a building) and a path polygon. In the case of the former, the user implicitly indicates a filled polygon as a building where all the pixels inside the polygon also belong to the building (or labelled as building); whereas, what indicates the latter as a path is only the boundary of the polygon (i.e., a closed line) and not the interior points. Due to this difference, we need to have a background information, that some structures should not be filled during the ground truth generation process even if they are represented as a polygon.

People's Perspectives: Differences in people's perspectives in labelling a structure or a region on the map can also be problematic for the ground truth generation process. For instance, there are a number of areas in OSM labelled as amusement park (i.e., building). However, our classifier labels these regions with *Water* as a large portion of them is (usually and not always though) covered by outdoor swimming pools. Unless a fine-grained structure (preferably an ontology) is provided for labels commonly used in cities, this problem exists.

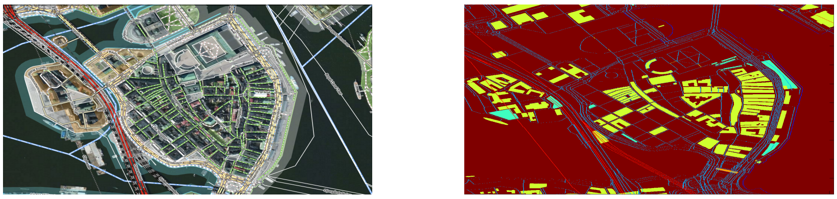
2.2 Precision

The area chosen for the scene processing contains 268,435,456 pixels. Dealing with the aforementioned issues, the ground truth generation process could label 54,919,330 pixels in total that covers about 25% of the size of the map. In this map, there are a number of significant vegetation areas that are not labelled by people. Figure 1a illustrates a snapshot of the map extracted from OSM[6] whose processed labels are also shown in Fig. 1b.

A CNN can be used to improve the shortcomings of the ground truth acquired from OSM. This is done by first training a CNN on the labelled areas and then classifying each pixel in the map using the trained CNN. Each pixel has a classification certainty that is related to the amount of labels from the OSM map. A sub-section of the labelled OSM map and the output from the CNN are shown in Fig. 2. As we can see in Fig. 2a, not all objects, roads, and categories are initially labelled. For example, there are some non-labelled buildings, areas of vegetation and non-infrastructure grounds. The results from the trained CNN

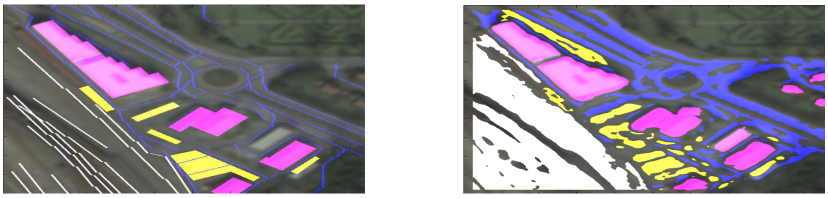
[6] https://www.openstreetmap.org/copyright

classifier is shown in Fig. 2b. As we can see, the roads and railroads that where initially only labelled by a road-centered polygons have become wider. There are some labelled buildings as well in the output. Due to the lack of annotation of vegetation and ground, the classifier has a low classification certainty of those areas and are therefore not added to the ground truth.



(a) Central of Stockholm (from OSM). (b) Labelled pixels.

Fig. 1: OSM data Processed to be used as Ground Truth.



(a) Labelled pixels from OSM. (b) Labelled pixels from a CNN.

Fig. 2: (a) A sub-section of the input data and the labels acquired from OSM. The categories are building (pink), parking (yellow), road (blue), railroad (white). (b) The labelled pixels from a trained CNN with above 99% classification certainty.

3 Discussion

In this paper, we summarized the practical issues met during our interaction with OSM for the purpose of automatically generating labelled data. Although the amount of retrievable information easily changes from one city to another depending on many factors including their fame, historical and touristy aspects, the representational issues mentioned in this paper are the same. However, despite these issues, we could generate a ground truth for our CNN classifier.

Acknowledgments. This work has been supported by the Swedish Knowledge Foundation under the research profile on Semantic Robots, contract number 20140033.

References

1. Lecun, Y. and Bottou, L. and Bengio, Y. and Haffner, P.: Gradient-based learning applied to document recognition. Pro. IEEE. 86(11):2278–2324 (1998)
2. Farabet, C. and Couprie, C. and Najman, L. and Lecun, Y.: Scene parsing with Multiscale Feature Learning, Purity Trees, and Optimal Covers. Pro. Int. Conf. on Machine Learning (ICML). 575–582 (2012)
3. Längkvist, M. and Kiselev, A. and Alirezaie, M. and Loutfi, A., Classification and Segmentation of Satellite Orthoimagery Using Convolutional Neural Networks. J. Remote Sensing 8(4):329 (2016)