

A MACHINE LEARNING PIPELINE ARTICULATING SATELLITE IMAGERY AND OPENSTREETMAP FOR ROAD DETECTION

M. A. Zurbaran^{1*}, P. Wightman², M. A. Brovelli¹

¹, Dept. of Civil and Environmental Engineering, Politecnico di Milano, Piazza Leonardo da Vinci, 32, 20133 Milano MI, Italy -
(mayraalejandra.zurbaran, maria.brovelli)@polimi.it

² Dept. of Systems Engineering, Universidad del Norte, km 5 via Pto. Colombia, Atlántico, Colombia -
pwightman@uninorte.edu.co

Commission IV, WG IV/4

KEY WORDS: Machine Learning, Artificial Intelligence, OpenStreetMap, Remote Sensing, Satellite Imagery

ABSTRACT:

Satellite imagery from earth observation missions enable processing big data to gather information about the world. Automatizing the creation of maps that reflect ground truth is a desirable outcome that would aid decision makers to take adequate actions in alignment with the United Nations Sustainable Development Goals. In order to harness the power that the availability of the new generation of satellites enable, it is necessary to implement techniques capable of handling annotations for the massive volume and variability of high spatial resolution imagery for further processing. However, the availability of public datasets for training machine learning models for image segmentation plays an important role for scalability.

This work focuses on bridging remote sensing and computer vision by providing an open source based pipeline for generating machine learning training datasets for road detection in an area of interest. The proposed pipeline addresses road detection as a binary classification problem using road annotations existing in OpenStreetMap for creating masks. For this case study, Planet images of 3m resolution are used for creating a training dataset for road detection in Kenya.

1. INTRODUCTION

The overall growth of generated data has propelled the renaissance of Artificial Intelligence (AI) for new approaches to solve problems in different areas. The increased availability of high resolution satellite imagery in recent years, enabled earth observation analysis in ways that were not possible before. Among the many possibilities, identifying features automatically from remote sensing imagery (RS) is an important task to take advantage of this data for updating existing maps.

Using satellite imagery for automatic road detection is of utmost importance for map-making; however, access to high resolution images, pre-processing and the lack of annotated masks ready to use in Machine Learning models (ML) play an important role for scalability. (Demir et al., 2018) state that “satellite images are only recently gaining attention from the [computer vision] community for map composition”. In (Demir et al., 2018) is presented the DeepGlobe 2018 challenge, urging competitors for ML models to “parse the earth through satellite images”, identifying roads among other features. The challenge provides the dataset ready to use; however, this behavior often present in challenges, inevitably biases the creation of the models to work with the provided images; the models will learn patterns of the cities depicted in the training dataset, neglecting other areas of the world that were not included, specially those with a significantly different landscape.

The task of providing the necessary training datasets requires bridging knowledge from computer vision as well as RS processing; there needs to be a methodology that enables the computer vision community to obtain datasets ready to use for ML models from anywhere in the world.

In order to address the lack of annotated datasets for solving hard problems with ML, a pipeline is presented to generate annotated masks for road detection. The proposed pipeline is built with pyQGIS, using the Overpass API to consume OpenStreetMap (OSM) data for an area of interest. This process may produce masks with some inaccuracy as opposed to human labelling, but with the possibility of generating a high volume of training data. Even if road extraction is not perfect due to challenging surroundings in the images, like dry rivers that can be confused with unpaved roads, the identification of roads through ML serves also to direct a human mapper focus on specific areas in the imagery to map.

1.1 Case Study

For this paper, a methodology that enables the creation of a dataset for training a ML model for road detection in Kenya is intended. Kenya was selected due to its challenging landscape for identifying roads, many of which are unpaved. These roads when seen from an aerial image are hard to identify for the human eye, even more when contrasted with bare soil or dry rivers due to seasonality, making labelling a complex task.

For this case study the creation of the training dataset is decisive in what the model will learn, other road detection models might not work due to the different landscape characteristics of Kenya.

The lack of extensive, accessible, and authoritative data sources is also a drawback in this area, being OSM and satellite imagery the most reliable ones.

To illustrate the proposed methodology for the creation of annotated masks, some steps in the pipeline as well as the selection of the data sources were considered in particular for the

*Corresponding author

case study scenario to obtain better results, even though other general solutions like RoboSat¹ are available.

1.2 Related Work

The term ML refers to the study of algorithms and statistical models that enable computers to perform a specific task without explicit instructions, but to be more specific, the task of road detection, as it is image recognition, is better addressed by Deep Learning (DL) algorithms. These are part of ML, but in particular, DL algorithms are the ones that “allow computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction” according to (LeCun et al., 2015). DL models use the backpropagation algorithm through the layers of a neural network to better adjust the model parameters. DL algorithms are usually convolutional neural networks composed by many layers, (LeCun et al., 2015) states that the key aspect of DL is that these layers of features are not designed by human engineers, they are learned from data using a general-purpose learning procedure.

Road detection with ML is not trivial due to the diverse shape of the roads as stated in (Schweitzer, Agrawal, 2019), the subject has nevertheless been studied before and has recently gained more attention due to the raise in computational power for computer vision algorithms.

Road detection studies have focused on the architecture for the model, in most cases using neural networks (Mokhtarzade, Zoj, 2007) or more recently, convolutional neural networks as in (Schweitzer, Agrawal, 2019) where authors use a synthetic aerial dataset for emulating the dessert. Successful trials have mostly been done with cities that have paved roads and with very high resolution imagery e.g. National Agriculture Imagery Program (NAIP) or images with masks provided in challenges (Demir et al., 2018).

Paved roads produce a high contrast against their surroundings, facilitating their detection. For this case study, on the contrary, unpaved roads in Kenya are easily confused with bare soil or dry rivers due to the similar reflectance and it is not trivial to differentiate from one another.

2. DATA COLLECTION AND PRE-PROCESSING

In order to generate binary masks for road detection, two main sources are needed: RS imagery and road topology. The following subsections will depict why the selected sources were chosen but also noting that another kind of imagery is possible to use with the proposed pipeline while analysing other areas.

2.1 Satellite Imagery

The new generation of satellites produce periodical high resolution imagery from the world. This periodical data is useful for training ML models for inferences about specific areas, considering not only one image per area but many in order to generate training data for different seasons. In the context of our case study, this comes useful, since producing a time series training dataset for each area of interest, aids in establishing a baseline to handle the seasonality of roads; including images for monsoon as well as draught periods, and also to produce a larger training dataset.

High resolution images are necessary for identifying features for mapping; specifically features in artificial land, like roads. Among the available possibilities for satellite imagery, Planet offers limited downloads of 3m resolution satellite imagery for research purposes; this is the highest resolution available for free from a satellite. There are also other alternatives, like drone aerial images available from OpenAerialMap, but this source is not as complete and not enough for the case study, even though the resolution is higher. For these reasons, Planet imagery (Daily Satellite Imagery and Insights — Planet, n.d.) will be used in the the case study.

Images are filtered from Planet using a bounding box for an area of interest to use the ones with less cloud coverage (about 15%), while allowing to query as well different time periods for the same area. In this way 2 composites for each area of interest are included for creating the training dataset. The size of the area of interest will determine the number of tiles that will be generated for the model. For this work, 3 towns were considered, therefore, 3 bounding boxes were used to filter the satellite images and 6 composites were produced from them in order to consider the seasonality of the roads.

2.2 OpenStreetMap Road Data

The OSM platform serves as a rich source of data that can be used for creating annotated masks for road detection. The crowd-sourced geospatial road data is sometimes the only national-level source available for some countries according to (Barrington-Leigh, Millard-Ball, 2017).

There have been plenty of research directed toward the assessment of the OSM road networks, focusing on positional accuracy and completeness. The first studies were done in 2009, in their works (Ather, 2009), (Haklay et al., 2010) and (Koukletsos et al., 2011) assessed the positional accuracy of OSM roads in England with a visual comparison of a limited number of roads with an authoritative dataset and a statistical approach based in the work of (Goodchild, Hunter, 1997). In order to obtain the completeness of the dataset covering England, the authors compared the lengths of the roads in OSM with those of the Ordnance Survey vector datasets. (Kounadi, 2009) did a similar experiment in Athens, where he considered around 300 roads and found that an average difference between OSM and official roads of about 6m and an average overlap of nearly 80%, which is in line with what Haklay concluded.

After some years, other studies were done based on the buffer zone methodology (Koukletsos et al., 2012), (Zielstra, Zipf, 2010), (Wang et al., 2013), (Siebritz, Sithole, 2014), (Graser et al., 2014). The studies show different results for different areas included in the experiments. In Europe mostly, the spatial accuracy and the completeness were good enough, while for places like South Africa, the dataset did not meet the accuracy requirements for the integration with the authoritative database.

In (Helbich et al., 2012), authors did a case study of a German city, they found that areas with high accuracy of OSM were primarily located in higher populated parts of the city, and concluding that these areas were subject to more frequent validation, followed by correction of errors, than rural areas.

In (Brovelli et al., 2016) and (Brovelli et al., 2017), authors introduced an automatic method based on geometrical similarity and a grid-based approach for the evaluation of road completeness and positional accuracy. The procedure demonstrated that

¹RoboSat - <https://github.com/mapbox/robosat>

OSM is a reliable source for Paris and it is flexible enough to be reused in other cities.

In (Girres, Touya, 2010), features representing the same object in the OSM and authoritative datasets were selected and matched manually to avoid errors. Differences in position were then computed on each pair of homologous objects. While the mean distance was acceptable, the standard deviation was larger than the reference accuracy used for official datasets, showing heterogeneity in the quality of the data. Regarding completeness, the authors found that using as an indicator the number of objects, OSM was complete only by around 10%; however, the completeness improved when they considered the comparison between the total length/area of the objects, obtaining an average value around 40%. This result shows that shorter objects are more likely to be absent, suggesting that volunteer contributors tend to map the most important elements in the road network.

In (Jovanovic et al., 2019) is stated that in developing countries, OSM is usually the most complete source. They performed studies for the case of Kenya, where after performing a completeness and positional accuracy analysis against the Road dataset published by Digital Chart of the World (DWC) provided by the United Nations Economic Commission for Africa and Global Roads published by Columbia University and provided by the Centre for International Earth Science Information Network (CIESIN), OSM proved to be more reliable with a 4.5 ratio against DWC and a ratio of 2.5 against CIESIN. Authors in (Jovanovic et al., 2019) go further and state that OSM also proved to have more positional accuracy in Kenya when compared to the mentioned datasets.

2.3 Pre-Processing

Satellite imagery was obtained from Planet and processed with Google Earth Engine to generate rendered composites as the one in Figure 1. For this, each image obtained from Planet for the area of interest was uploaded as a Google Earth Engine (GEE) asset (since Planet is not yet a GEE source) and loaded as an ImageCollection for visualization and export as a single composite. Planet imagery has 4 bands, with the first 3 corresponding to blue, green, and red; the order and the bands to use while processing was: b3, b2, b1.

In order to produce the composites from Planet images, minimum and maximum statistics were calculated for each band to adjust visualization parameters. The gamma value used was 1.5, 1.3, 1.3 for each BGR band, respectively. The javascript code for this processing is available in the github page for this project².

The composite then serves to generate smaller tiles like the one seen in Figure 2, that can be used in a ML model, since the complete raster may not be processed at once. For this, the zoom level 16 was used for running the *gdal2tiles* command, but a different zoom level is possible to use in the script. This will ensure that the same tiles are generated for the composite as well as for the mask while obtaining a good detail from both. The composite needs to be saved as a rendered raster and projected to pseudo mercator before running this command. The structure of the file directory output can then serve to easily identify which areas contain roads in the map in order to direct mappers attention to them.

²osm2mask - <https://github.com/mazucci/osm2mask>



Figure 1. Sample of a rendered composite



Figure 2. Sample of a rendered composite tile

While generating the tiles there may be some cases in which the image does not fill the entire tile size near the corners of the composite. In these cases, smaller tiles like seen in Figure 3 are generated to avoid creating the corresponding masks with false information, i.e. having 0 values where no data is present.

Satellite imagery available directly from GEE may be used for other ML inferences that require less spatial resolution, facilitating access to images and allowing a high volume of composites to be generated with less user interaction.

3. OSM2MASK

3.1 Annotated Masks

The OSM data model for representing map objects is made of geometries with attributes that describe them. These geometries can be ways, nodes or relations. Ways are represented as lines, nodes are points, and relations are a collection of points or ways that represent a larger whole. Attributes are described as tags that can be part of a node, a way or a relation.

In Western Africa, where Kenya is located, the classification of roads for adding attributes in OSM is as explained in (East Africa Tagging Guidelines - OpenStreetMap Wiki, n.d.). The



Figure 3. Sample of a rendered cropped tile near the bounds of the composite



Figure 4. Sample of an annotated mask tile

document outlines what each attribute given to any way represents; this is useful for considering the width for each kind of road for generating the masks. Within the possible tags, there are some that can help identify highways, which are wider, these are: trunk, primary, secondary, and tertiary. Tags belonging to unclassified, residential, service, track, path or private access are considered minor roads and given the smaller width.

For the pipeline, some considerations were made in order to adapt the annotated masks from OSM data to match the satellite images as close as possible in the area of interest. From the rendered rasters, like the one in Figure 1, the width of roads is measured and explicitly set to road masks in the pipeline according to the classification given in OSM.

The OSM data is consumed using the Overpass API filtering by a bounding box for the area of interest and selecting only way features, the query for this is available in the project repository³. Ways are originally LineStrings but are later adjusted using a buffer to cover the surface of the road. In this way, LineStrings are converted to Polygons using a width of 6m for converting ways corresponding to minor roads and a width of 10m for highways. The Polygons representing the roads are used to generate a binary raster where pixels corresponding to a road have a value of 1 and pixels belonging to another kind of surface is marked as 0.

Having the two overlapping rasters, the mask is split into tiles in the same way as the composite to produce smaller overlapping images that can be used to train the ML model, since having the original larger images would make training the model too demanding. Figure 4 shows an annotated binary mask corresponding to the composite tile in Figure 1.

For handling the corners of the mask where no data may result, the same is done as with the composite, generating smaller tiles when no data is found.

3.2 Data Augmentation

In order to obtain a larger dataset for the model to learn better and take advantage of the obtained images, there is a technique

called data augmentation that has proved to enhance performance in deep learning algorithms as seen in (Ronneberger et al., 2015). Data augmentation finds its base in that the more data is available for a ML algorithm, the more effective it will be as stated in (Wang, Perez, 2017). This procedure consists in enlarging data, in this case, taking images of the training dataset to be altered in order to increase the volume of the dataset. For this, images can be shifted, flipped, rotated among other variations.

Data augmentation is used in both cases, to improve performance and when there is not enough reliable data to train a model. (Wang, Perez, 2017) states that the problem with small datasets is that models trained with them do not generalize well data from the validation and test set. Hence, these models suffer from the problem of over-fitting. Over-fitting refers to when a model learns too well the training data but performs poorly with new data. So in order to overcome this issue, which happens when few data is used, the initial dataset is augmented and the risk of over-fitting is reduced.

In this work, images and masks are rotated to increase four times the initial dataset, an example of this is showed in Figure 5.

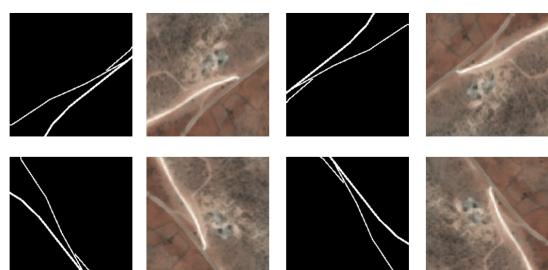


Figure 5. Sample of image augmentation by rotation

With the proposed pipeline it is possible to automatize the process for creating annotated masks for road extraction, the steps of the methodology can be summarized as follows:

1. Define a bounding box for the area of interest.
2. Use the bounding box to filter Planet images during drought and monsoon seasons - or the seasons available in the area of interest.
3. Upload the satellite images to GEE as assets and grouped into an image collection.
4. Use the makecomposite.js script to generate a composite with the satellite images.
5. Use the bounding box to query OpenStreetMap with the Overpass API for ways within the area of interest.
6. Create a buffer around each way with the width depending on the type of road as defined in the OpenStreetMap metadata.
7. Clip ways that stand outside the bounding box.
8. Add an *is_road* column to each polygon representing a road with the value of 1.
9. Rasterize the vector layer of the polygons using the *is_road* attribute to mark roads.
10. Use the gdal2tiles tool to generate tiles from both the composite and mask rasters.
11. Save rotated tiles from the composite and mask.
12. Use the output tiles to train a ML model³.

The scripts provided for GEE and pyQGIS automate this process.

Figure 6 explains the schema for the pipeline. In here, Planet imagery was used but this source is interchangeable with other satellite or aerial imagery.

For the case study, an image for an area of interest of 178.56 km², resulted in 527 images, each with a resolution of 256 × 256 pixels. When considering 2 composites for comparing between seasons, the number of output images was 1054 and 4216 after applying augmentation.

For the 3 towns considered, a total area of 983.9 km² covered by composites, resulted in 2857 images, after considering road seasonality and data augmentation the total output was 22856 with their corresponding masks.

3.3 Conclusions

This work demonstrates a methodology in which crowd-sourced data from OSM may be used as input for ML algorithms. A pipeline is detailed and explained in which satellite imagery and OSM data are merged to produce a training dataset for road segmentation. The quality of the generated masks is directly related to the accuracy of the mapped features present in the area of interest; nevertheless, for the case study area, OSM proved to be the one that is most reliable. For the

³Road segmentation model - <https://github.com/mahmoudmohsen213/airs>

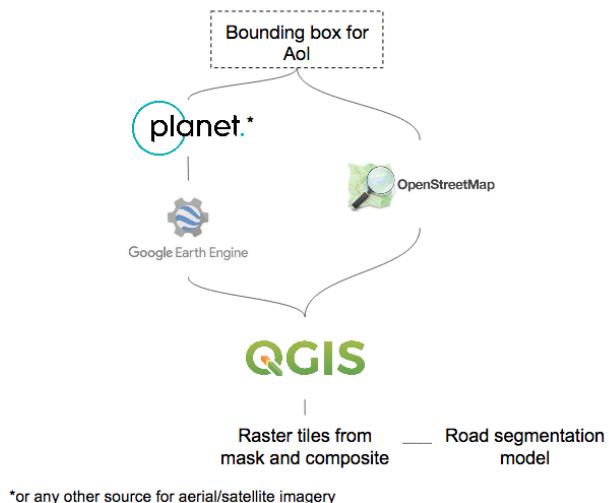


Figure 6. Schema for the pipeline

areas of interest considered with the 3 composites used and 3 more for checking both seasons, the resulting images are in the order of thousands; even if the annotations are not too precise, the volume produced enables training a ML model. For further research, it is proposed to work in cooperation with mapping teams like YouthMappers⁴ to verify the width of the roads and type of the roads for the considered area of interest to increase the accuracy of the output training dataset.

ACKNOWLEDGEMENTS

The authors would like to sincerely thank the Planet Team (Daily Satellite Imagery and Insights — Planet, n.d.) for providing access to their satellite imagery.

REFERENCES

- Ather, A., 2009. A quality analysis of openstreetmap data. *ME Thesis, University College London, London, UK*, 22.
- Barrington-Leigh, C., Millard-Ball, A., 2017. The world's user-generated road map is more than 80% complete. *PLoS ONE*, 12(8), e0180698. <https://dx.plos.org/10.1371/journal.pone.0180698>.
- Brovelli, M. A., Minghini, M., Molinari, M. E. et al., 2016. An automated GRASS-based procedure to assess the geometrical accuracy of the OpenStreetMap Paris road network. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 919–925.
- Brovelli, M. A., Minghini, M., Molinari, M., Mooney, P., 2017. Towards an automated comparison of OpenStreetMap with authoritative road datasets. *Transactions in GIS*, 21(2), 191–206.
- Daily Satellite Imagery and Insights — Planet, n.d. <https://www.planet.com/>.
- Demir, I., Koperski, K., Lindenbaum, D., Pang, G., Huang, J., Basu, S., Hughes, F., Tuia, D., Raska, R., 2018. DeepGlobe 2018: A Challenge to Parse the Earth through Satellite Images. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, IEEE, 172–17209.
- ⁴YouthMappers - <https://www.youthmappers.org/>

- East Africa Tagging Guidelines - OpenStreetMap Wiki, n.d. https://wiki.openstreetmap.org/wiki/East_Africa_Tagging_Guidelines.
- Girres, J.-F., Touya, G., 2010. Quality assessment of the French OpenStreetMap dataset. *Transactions in GIS*, 14(4), 435–459.
- Goodchild, M. F., Hunter, G. J., 1997. A simple positional accuracy measure for linear features. *International journal of geographical information science*, 11(3), 299–306.
- Graser, A., Straub, M., Dragaschnig, M., 2014. Towards an Open Source Analysis Toolbox for Street Network Comparison: Indicators, Tools and Results of a Comparison of OSM and the Official Austrian Reference Graph. *Transactions in GIS*, 18(4), 510–526.
- Haklay, M., Ather, A., Basiouka, S., 2010. How many volunteers does it take to map an area well? *Proceedings of the GIS Research UK 18th Annual Conference*, UCL, 193–196.
- Helbich, M., Amelunxen, C., Neis, P., Zipf, A., 2012. Comparative spatial analysis of positional accuracy of OpenStreetMap and proprietary geodata. *Proceedings of GI_Forum*, 24–33.
- Jovanovic, S., Jovanovic, D., Bratic, G., Brovelli, M. A., 2019. Analysis of Free Road Data in Tanzania, Uganda and Kenya Using Free and Open Source Software. *Proceedings of ISPRS Geospatial Week*.
- Koukoletsos, T., Haklay, M., Ellul, C., 2011. An automated method to assess data completeness and positional accuracy of openstreetmap. *GeoComputation*, 3, 236–241.
- Koukoletsos, T., Haklay, M., Ellul, C., 2012. Assessing data completeness of VGI through an automated matching procedure for linear data. *Transactions in GIS*, 16(4), 477–498.
- Kounadi, O., 2009. Assessing the quality of OpenStreetMap data. *Msc geographical information science, University College of London Department of Civil, Environmental And Geomatic Engineering*.
- LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. *Nature*, 521(7553), 436–444. <http://www.nature.com/articles/nature14539>.
- Mokhtarzade, M., Zoj, M. J., 2007. Road detection from high-resolution satellite images using artificial neural networks. *International Journal of Applied Earth Observation and Geoinformation*, 9(1), 32–40.
- Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9351, Springer, Cham, 234–241.
- Schweitzer, D., Agrawal, R., 2019. Semantic Segmentation of Complex Road Environments from Aerial Images Using Convolutional Neural Networks. *Proceedings - 2018 IEEE International Conference on Big Data, Big Data 2018*, IEEE, 3464–3469.
- Siebritz, L.-A., Sithole, G., 2014. Assessing the quality of openstreetmap data in south africa in reference to national mapping standards. *Proceedings of the Second AfricaGEO Conference, Cape Town, South Africa*, 1–3.
- Wang, J., Perez, L., 2017. The effectiveness of data augmentation in image classification using deep learning. *Convolutional Neural Networks Vis. Recognit.*
- Wang, M., Li, Q., Hu, Q., Zhou, M., 2013. Quality analysis of open street map data. *ISPRS Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci*, 2, W1.
- Zielstra, D., Zipf, A., 2010. A comparative study of proprietary geodata and volunteered geographic information for germany. *13th AGILE international conference on geographic information science*, 2010.