A CONCEPTUAL MODEL FOR CONVERTING OPENSTREETMAP CONTRIBUTION TO GEOSPATIAL MACHINE LEARNING TRAINING DATA

H. Li[1,]*, A. Zipf[1,2]

[1] GIScience Chair, Institute of Geography, Heidelberg University, 69120 Heidelberg, Germany - hao.li@uni-heidelberg.de
[2] HeiGIT at Heidelberg University, Schloss-Wolfsbrunnenweg 33, 69118 Heidelberg, Germany - zipf@uni-heidelberg.de

Commission IV, WG IV/4

KEY WORDS: Volunteered geographical information, GeoAI, data quality, OpenStreetMap, machine learning, crowdsourcing.

ABSTRACT:

In the recent decade, Volunteered Geographical Information (VGI), in particular the OpenStreetMap (OSM), has helped to fill substantial data gaps in base maps, especially in Global South, thus has become a promising source of massive, free training data together with rich and detailed semantic information for geospatial artificial intelligence (GeoAI) applications. Although intensive works have explored the potential of generating training data from OSM, a systematic approach of harvesting OSM contribution as quality-aware training data for different GeoAI tasks is still missing. To fill this research gap, we proposed a conceptual model consisting of three major components: historical OSM and external datasets, quality indicators, and GeoAI models. As a proof of concept, we validated our conceptual model with an example task of detecting OSM missing buildings in Mozambique, where the impact of different error sources (e.g., completeness, alignment, rotation) in training data were compared and investigated in a quantitative manner. The lessons learned in this paper shed important lights on cooperating OSM data quality aspects with the development of more explainable GeoAI models.

1. INTRODUCTION

Due to the development of big data and crowdsourcing technology, Volunteered Geographic Information (VGI) as a special case of user-generated content continued to harvest big geographic data that was contributed voluntarily by mappers (Goodchild, 2007). As one of the most popular VGI projects, OpenStreetMap (OSM) has become an important source for geospatial data in many aspects, ranging from routing planning, land use/land cover (LULC) mapping to disaster management and humanitarian mapping, especially in the Global South. Examples of humanitarian mapping in OSM includes the 2010 Haiti earthquake (Zook et al., 2010), the 2014 West Africa Ebola outbreak (Dittus et al., 2016), and the 2019 Cyclone Idai and Kenneth in Mozambique (Li et al., 2020), where over millions of buildings and roads were mapped and added to OSM within a short period of time. In this context, OSM has contributed to fill substantial data gaps in base maps and to alleviate mapping inequalities across countries in worldwide (Albuquerque et al., 2016, Herfort et al., 2019).

More recently, major progress and achievements have been made in the field of big data analysis and geospatial artificial intelligence (GeoAI) (Janowicz et al., 2020), while the lack of high-quality training data has been identified one of the major bottlenecks for GeoAI since long. Fortunately, OSM was recently explored, making use of its rich semantic information (e.g., OSM tag and value) to extract customized geospatial objects, as well as generating geo-referenced training samples, in order to develop robust LULC mapping and geospatial object detection models (Schultz et al., 2017, Fonte et al., 2020, Chen and Zipf, 2017, Vargas-Munoz et al., 2020). Regarding LULC mapping, a very first attempt in (Schultz et al., 2017) successfully created the first OSMLULC map by training a random forest (RF) classification model with OSM training samples and Landsat MSI data. Similarly, (Fonte et al., 2020) investigated the potential of generating training data from OSM to classify Sentinel-2 time series data into distinct LULC classes. Moreover, for a more sophisticated object detection task, early work of (Chen and Zipf, 2017) showed stimulating results of deep learning from VGI, especially from OSM data, for a building detection task. Later in (Herfort et al., 2019), OSM data was adopted to fine-tune an object detection deep learning model to detect human settlements in rural areas and achieved competitive accuracy comparing to a crowdsourcing method (MapSwipe). OSM data has shown great potentials in offering a massive and freely available source of human-labeled geographical features as training data for GeoAI applications. However, a systematic approach of harvesting OSM contribution as training data for different GeoAI tasks is still missing. In this paper, we aim to fill this research gap by developing a conceptual model to convert OSM contribution into training data by incorporating its intrinsic data quality, which can be applied to different geospatial machine learning (ML) applications.

Different from traditional authorized geospatial data sources, one major advantage of OSM data is the availability of its full history. In (Barron et al., 2014), it was reported that approximate statements (without external reference) on OSM data quality are generally possible. Therefore, by exploring the OSM historical data, intensive existing works have been dedicated to assess the OSM data quality in an intrinsic manner (Mooney et al., 2010, Minghini and Frassinelli, 2019, Grinberger et al., 2021, Schott et al., 2021). Quality measurements of OSM data usually follow the principles of International Organization for Standardization (ISO) under ISO 19113 and ISO 19157, which consist of multiple quality aspects, such as completeness, position accuracy, and logical consistency. For example Completeness describes how complete the OSM data is, and a lack of data is referred to as "Error of Omission" (Barron et al., 2014). More recently, early attempt started to investigate the effect of

* Corresponding author

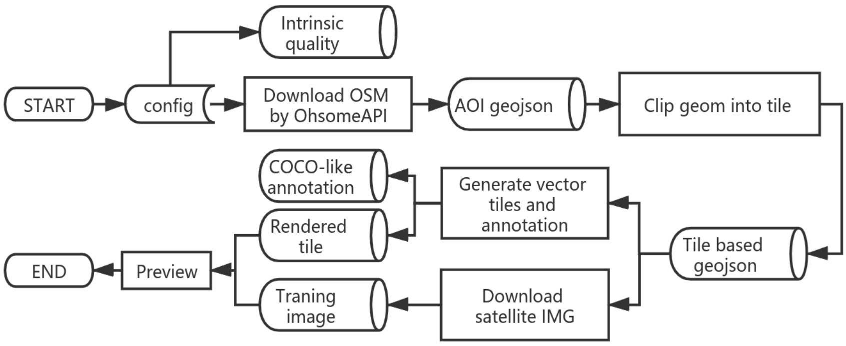

Figure 1. The workflow of ohsome2label as a prototype of the proposed conceptual model.

such errors (e.g., "Error of Omission") in training data when applying ML methods to earth observation (EO) data (Elmes et al., 2020). To this end, we argue that the consideration of OSM data quality is definitely necessary and potentially beneficial for the effective and efficient development of geospatial ML models.

This study aims to develop a conceptual model of converting OSM contribution to quality-aware geospatial ML training data, and further investigate the impact of training data quality in spatial ML models. Specifically, we first provided a systematic overview of how to properly extract OSM features depending on the type of ML tasks (e.g., classification, object detection, semantic segmentation); we then identified quality aspects of training data that might affect the model performance of GeoAI, and define potential indicators based on historical OSM data or external data, which can measure the intrinsic/extrinsic data quality of the training data; finally, we evaluated the conceptual model by fine-tuning a building detection on simulated OSM training data and examines the impact of training data in model performances.

As a prototype of our conceptual model, we developed an open-source python tool named ohsome2label[1] (Wu et al., 2020). As its name says, ohsome2label was built on the ohsome API[2], which provides a flexible and fast way of analyzing the rich data source of OSM history, therefore it allows us to calculate many intrinsic quality indicators (see a list of endpoints in ohsome API[2]) besides extracting OSM geometry. In this paper, we selected the OSM missing building detection as an example geospatial ML task, and evaluated our conceptual model across a residential area in Mozambique by fine-tuning a Faster RCNN building detector on different sets of simulated OSM training data, such as complete or incomplete, well-aligned or misaligned, etc. Our study therefore makes two major contributions: (1) a systematic and generic approach of harvesting OSM contribution as quality-aware training data for different geospatial ML applications; (2) by testing on simulated OSM training data, the impact of errors in training data on model performance is examined in a quantitative manner.

The remainder of this paper is organized as follows. Section 2 describes the conceptual model and Section 3 presents the experimental design and datasets used in the OSM missing building detection example. Section 4 elaborates on the results and discusses the future directions, then Section 5 summarizes the lessons learned and concludes the paper.

2. CONCEPTUAL MODEL

Besides the prototype workflow of ohsome2label (Figure 1), we elaborate on the overall design of our conceptual model in Figure 2, where the model consists of mainly three major components:

Historical OSM and external datasets - The first component refers to the full-history of OSM data as well as external data sources (e.g., authorial GIS data (Fan et al., 2014) or open-source building layers (Li et al., 2020)), based on which intrinsic or extrinsic data quality analysis of OSM feature can be conducted. Moreover, the historical OSM data serves as a basis of extracting target features as training data for GeoAI.

Quality Indicators - Herein, the concept of quality indicators ranges from extrinsic measurements (e.g., completeness, position accuracy, etc.) to intrinsic factors (e.g., currentness, saturation, etc.), thus besides existing indicators this can be easily extended to novel indicators of OSM data quality. For more details, one can refer to existing works of OSM quality assessment in (Barron et al., 2014) and (Fan et al., 2014, Zhou et al., 2019) from either an intrinsic or extrinsic perspective. More importantly, we aim to cooperate the quality prior knowledge into the training of GeoAI models, and investigate the impact of quality aspects on model performances.

GeoAI models - The proposed conceptual model is designed to be independent of the specific GeoAI task or model, which should support the training of diverse GeoAI models based on OSM features, for instance, classification, semantic segmentation, and object detection in Figure 3. In general, GeoAI models consume OSM-based training data as well as their corresponding quality prior knowledge and seek to learn predictive capability w.r.t target features (e.g., building, road, and LULC). In this context, our conceptual model could facilitate the development of explainable GeoAI (Hu et al., 2019, Xing and Sieber, 2021) with explicit prior knowledge of training data. Moreover, our conceptual model highlights the potential of generating ML predictions with high accuracy levels and relying on these mapping results during extrinsic quality analysis. An early attempt in this direction was conducted in (Li et al., 2019).

As a proof of concept, we validate the proposed conceptual model on an example GeoAI task of detecting OSM missing

[1] https://github.com/GIScience/ohsome2label
[2] https://docs.ohsome.org/ohsome-api/v1/

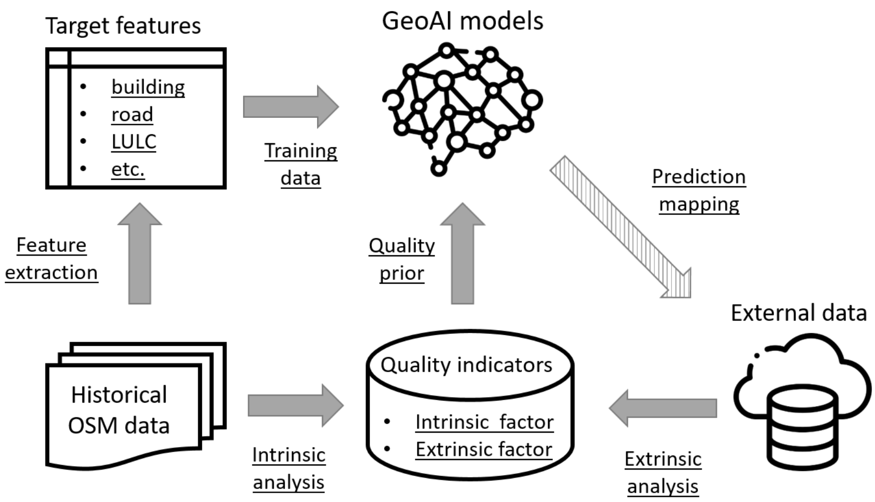

Figure 2. The conceptual model of converting OSM contribution to quality-aware training data of GeoAI models.

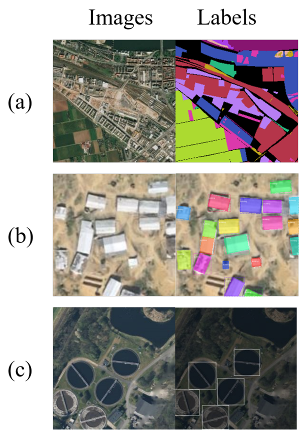

Figure 3. Examples of OSM training data generated by ohsome2label. (a) LULC classification; (b) semantic segmentation of human settlements; (c) object detection of wastewater treatment plants.

building. Specifically, we select a test area from a rural region of Mozambique, then explicitly simulate different error sources (i.e., completeness, alignment, rotation) in the training data, and further examine their impact on the OSM missing building detection results.

3. DATA AND EXPERIMENT

3.1 Study area and model

The experimental setup of the OSM missing building detection task (Figure 4) consists of mainly two steps. First, a well-mapped area (i.e., buildings) in Tanzania was selected as the training area for the base model to detect buildings. Next, a target area in Mozambique, where buildings are completely missing in OSM, was identified as a test area to validate the proposed conceptual model. By considering the fact that buildings from two areas might appear differently (e.g., size, shape, appearance) and lead to poor model performance of the base model, 80% of the target area was split for the model fine-tuning purpose and 20% was kept for testing. In this context, we carefully digitized all building geometries based on the Bing satellite imagery in the target area, which leads to 740 buildings for fine-tuning and 183 buildings for testing. Based on these building geometries, we further simulated three different error sources during the fine-tuning, namely completeness, alignment, and rotation.

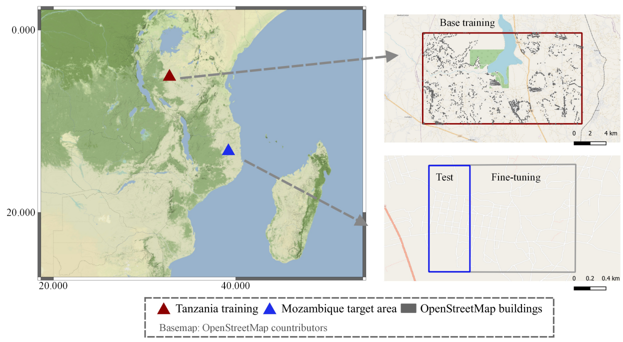

Figure 4. The overview map of study areas as a proof of concept. (1) the Tanzania training area for the base model (well-mapped in OSM); (2) the target area Mozambique split by 80% for fine-tuning and 20% for test (completely missing in OSM).

As for the building detection model, we followed a good practice in (Li et al., 2019) and implemented a Faster R-CNN (Ren et al., 2016) using ResNet-50 (He et al., 2016) as a backbone network, which was pre-trained on Microsoft COCO dataset (Lin et al., 2014). The pre-trained parameters were downloaded from the Tensorflow Detection Model Zoo (Tensorflow, 2020). For more details of the base model training on OSM, one could refer to a technical walk-through of automatic building detection with ohsome2label and Tensorflow (Wu et al., 2021).

3.2 Error sources of training data

Based on the 740 fine-tuning buildings in Mozambique, we then simulated three typical types of training data errors as follows:

Completeness error - As one can see column-wise in Figure 5, we explicitly considered different level of completeness in the fine-tuning building geometries. Specifically, such errors refer

https://doi.org/10.5194/isprs-archives-XLIII-B4-2022-253-2022 |

Table 1. Validation accuracies of the OSM missing building detection in Mozambique regarding different error sources in training data.

Methods	Completeness(%)	Predictions	Recall(%)	Precision(%)	F1
Correct	30	71	32.42	83.46	0.4669
	50	131	58.72	82.05	0.6844
	70	169	74.61	80.79	0.7758
	90	192	84.40	80.23	0.8226
	100	198	86.54	79.72	0.8299
Misalignment	30	35	17.43	90.47	0.2923
	50	81	38.53	87.50	0.5350
	70	127	58.40	84.14	0.6895
	90	155	70.03	82.37	0.7570
	100	150	67.89	82.52	0.7450
Rotation	30	51	24.46	86.96	0.3818
	50	114	52.29	83.82	0.6440
	70	153	68.50	81.45	0.7442
	90	174	77.37	81.09	0.7919
	100	200	85.93	78.49	0.8204

to the absence of a certain amount of buildings in the training data.

Alignment error - Besides the completeness, we also considered different levels of position accuracy, for which building geometries were moved into four random directions by 5 meters distance, thus simulated the misalignment in the training data.

Rotation error - Last, different levels of Shape accuracy were included, where we rotated building geometries by random angles based on their geometry center. This type of error simulated the rotation of training features.

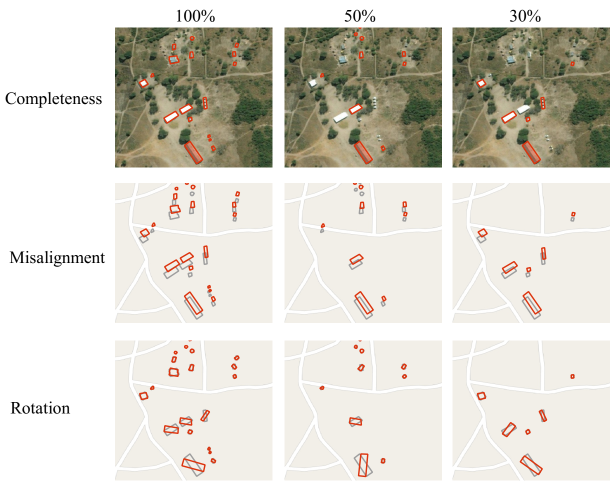

Figure 5. Illustrations of simulated error sources in training data. Basemaps contrast to Bing aerial imagery.

Base on this design, we considered a range of completeness levels from 100% to 90%, 70%, 50%, and 30%, together with alignment and rotation error. As a result, a number of 15 building geometries was used to fine-tune the base model trained on OSM data in order to investigate the impact of different errors in training data. The fine-tuning process in Mozambique was run for 20,000 epochs with an initial learning rate of 0.0004. All experiments ran on a Linux server with 4 GeForce RTX 2080Ti graphical processing units (GPUs), each with 12 GB of memory.

Based on the building detection results, we used distinct validation metrics to assess the general mapping accuracy against the 183 reference buildings. Specifically, we considered the metrics of Precision, Recall, and F1 score (F1). All metrics are derived from the number of False Negatives (FN), False Positives (FP), True Negatives (TN), and True Positives (TP), where we used the Intersection over Union (IoU) between building geometries and all prediction boxes as a criterion and empirically set the threshold to 0.5.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{1}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{2}$$

$$\text{F1} = \frac{2\text{TP}}{(2\text{TP} + \text{FP} + \text{FN})} \tag{3}$$

4. RESULT AND DISCUSSION

Table 1 presents a quantitative comparison of building detection accuracies w.r.t different error sources in training data. Based on the numerical results, one can observe the following key findings.

First and most straightforward, we noticed that a higher completeness level mostly led to a better model with higher accuracies, where a 100% completeness in the training data outperformed other variations regardless of the misalignment or rotation errors. Actually, the importance of completeness in the training data of GeoAI was recognized since long, where huge efforts were dedicated to establishing high-quality and complete benchmark datasets, for instance, NWPU VHR-10 (Cheng et al., 2016), DOTA (Ding et al., 2021), LCZ42 (Zhu et al., 2020), and FAIR1M (Sun et al., 2022).

Figure 6. Prediction maps of OSM missing building in the test area w.r.t different simulated error sources in training data. Basemaps contrast to Bing aerial imagery.

Moreover, from an intrinsic quality perspective, the completeness of OSM data could be estimated via intrinsic analysis, for example, buildings of a specific area can be stated "nearly complete" when the temporal change of new contributions or user edits is close to saturation (Barron et al., 2014). Such an estimation of OSM data completeness is of substantial benefit when training a GeoAI model with OSM data, especially in global south countries where external datasets are less available. While from an extrinsic quality perspective, a recent work in (Zhang et al., 2022) proposed a promising method of assessing OSM building completeness based on population data.

Next, regarding the alignment error, though the impact of 5 meter misalignment varies in buildings of different sizes, one can observe a significant decrease in detection accuracies with much fewer building predictions than training on correct geometries. Specifically, the misalignment in training data resulted in the lowest recall value (17.43%) combined with a completeness level of 30%. Therefore, the position accuracy of training data definitely plays an important role in establishing a robust GeoAI model. A good example of assessing the position accuracy of OSM data can be found in (Fan et al., 2014).

Last but not the least, though rotation errors are sometimes even more obvious than misalignment (Figure 5), the impact of rotation errors was relatively trivial compared to misalignment. Especially in the case of 100% completeness, our building detection models achieved a similar level of recall, precision and F1 score. While it is not supervising to find that the pre-trained Faster R-CNN was robust to these rotation errors, since rotation was usually considered as a data augmentation technique when lacking of enough training data. Therefore, a lesson learned from our results was the GeoAI model (at least the Faster R-CNN used in this work) can compromise various shape accuracies to a certain extent when completeness and position accuracy are ensured.

A visual comparison of models with different error sources in Figure 4 provides important insights into the aforementioned findings. In this first row, by contrasting satellite images, we can observe a trend of missing buildings when models were trained with less complete data. Moreover, when comparing the second and third rows, the model was more robust to rotation errors than alignment errors and generated much more building predictions in the third row, which also confirmed our previous statement of different impact of these two types of errors w.r.t detecting buildings of various sizes and shapes.

5. CONCLUSION

In this paper, we proposed a conceptual model of converting OSM contribution to quality-award training data of geospatial ML models. As a proof of concept, we conducted a series of experiments to validate the conceptual model regarding a specific geospatial ML task of detecting OSM missing buildings in Mozambique. To this end, we first trained a Faster RCNN in a well-mapped area in Tanzania, then fine-tuned it in our target area in Mozambique using a set of simulated training data consisting of three types of errors (i.e., completeness, misalignment, rotation), and investigated their impact of the model performance. Moreover, the proposed conceptual model could be easily extended to include existing or novel quality indicators via either intrinsic or extrinsic analysis based on lessons learned in this paper. One future direction could be applying this conceptual model with different GeoAI models (e.g., classification, semantic segmentation) and application contexts.

In short, the insight shared in this paper highlighted the huge potential of harvesting OSM data as GeoAI training data, while in the meantime suggested taking full consideration of OSM data quality for a more effective and efficient development of geospatial ML models. Our future work will focus on extending the quality indicators, as well as developing more explainable GeoAI models by incorporating quality as prior knowledge.

ACKNOWLEDGEMENTS

We thank the reviewers for their valuable comments and suggestions. This work has been partly supported by the Klaus Tschira Stiftung (KTS) Heidelberg.

REFERENCES

Albuquerque, J. P. d., Herfort, B., Eckle, M., 2016. The tasks of the crowd: A typology of tasks in geographic information crowdsourcing and a case study in humanitarian mapping. *Remote Sensing*, 8(10), 859. https://doi.org/10.3390/rs8100859.

Barron, C., Neis, P., Zipf, A., 2014. A comprehensive framework for intrinsic OpenStreetMap quality analysis. *Transactions in GIS*, 18(6), 877–895. https://doi.org/10.1111/tgis.12073.

Chen, J., Zipf, A., 2017. Deepvgi: Deep learning with volunteered geographic information. *Proceedings of the 26th International Conference on World Wide Web Companion*, 771–772.

Cheng, G., Zhou, P., Han, J., 2016. Learning Rotation-Invariant Convolutional Neural Networks for Object Detection in VHR Optical Remote Sensing Images. *IEEE Transactions on Geoscience and Remote Sensing*, 54(12), 7405-7415.

Ding, J., Xue, N., Xia, G.-S., Bai, X., Yang, W., Yang, M. Y., Belongie, S., Luo, J., Datcu, M., Pelillo, M. et al., 2021. Object detection in aerial images: A large-scale benchmark and challenges. *arXiv preprint arXiv:2102.12219*.

Dittus, M., Quattrone, G., Capra, L., 2016. Analysing volunteer engagement in humanitarian mapping: building contributor communities at large scale. *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, 108–118.

Elmes, A., Alemohammad, H., Avery, R., Caylor, K., Eastman, J. R., Fishgold, L., Friedl, M. A., Jain, M., Kohli, D., Laso Bayas, J. C. et al., 2020. Accounting for training data error in machine learning applied to Earth observations. *Remote sensing*, 12(6), 1034.

Fan, H., Zipf, A., Fu, Q., Neis, P., 2014. Quality assessment for building footprints data on OpenStreetMap. *International Journal of Geographical Information Science*, 28(4), 700–719. https://doi.org/10.1080/13658816.2013.867495.

Fonte, C. C., Patriarca, J., Jesus, I., Duarte, D., 2020. Automatic Extraction and Filtering of OpenStreetMap Data to Generate Training Datasets for Land Use Land Cover Classification. *Remote Sensing*, 12(20). 10.3390/rs12203428.

Goodchild, M. F., 2007. Citizens as sensors: the world of volunteered geography. *GeoJournal*, 69(4), 211–221.

Grinberger, A. Y., Schott, M., Raifer, M., Zipf, A., 2021. An analysis of the spatial and temporal distribution of large-scale data production events in OpenStreetMap. *Transactions in GIS*, 25(2), 622-641. https://doi.org/10.1111/tgis.12746.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Identity mappings in deep residual networks. B. Leibe, J. Matas, N. Sebe, M. Welling (eds), *Computer Vision – ECCV 2016*, Springer International Publishing, Cham, 630–645.

Herfort, B., Li, H., Fendrich, S., Lautenbach, S., Zipf, A., 2019. Mapping human settlements with higher accuracy and less volunteer efforts by combining crowdsourcing and deep learning. *Remote Sensing*, 11(15), 1799. https://doi.org/10.3390/rs11151799.

Hu, Y., Gao, S., Lunga, D., Li, W., Newsam, S., Bhaduri, B., 2019. GeoAI at ACM SIGSPATIAL: progress, challenges, and future directions. *Sigspatial Special*, 11(2), 5–15.

Janowicz, K., Gao, S., McKenzie, G., Hu, Y., Bhaduri, B., 2020. GeoAI: spatially explicit artificial intelligence techniques for geographic knowledge discovery and beyond.

Li, H., Herfort, B., Huang, W., Zia, M., Zipf, A., 2020. Exploration of OpenStreetMap missing built-up areas using twitter hierarchical clustering and deep learning in Mozambique. *ISPRS Journal of Photogrammetry and Remote Sensing*, 166, 41–51. https://doi.org/10.1016/j.isprsjprs.2020.05.007.

Li, H., Herfort, B., Zipf, A., 2019. Estimating openstreetmap missing built-up areas using pre-trained deep neural networks. *Proceedings of the 22nd AGILE Conference on Geographic Information Science*, 17–20.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C. L., 2014. Microsoft coco: Common objects in context. *European conference on computer vision*, Springer, 740–755.

Minghini, M., Frassinelli, F., 2019. OpenStreetMap history for intrinsic quality assessment: Is OSM up-to-date? *Open Geospatial Data, Software and Standards*, 4(1), 1–17.

Mooney, P., Corcoran, P., Winstanley, A. C., 2010. Towards quality metrics for openstreetmap. *Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems*, 514–517.

Ren, S., He, K., Girshick, R., Sun, J., 2016. Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE transactions on pattern analysis and machine intelligence*, 39(6), 1137–1149.

Schott, M., Grinberger, A. Y., Lautenbach, S., Zipf, A., 2021. The Impact of Community Happenings in OpenStreetMap—Establishing a Framework for Online Community Member Activity Analyses. *ISPRS International Journal of Geo-Information*, 10(3). https://www.mdpi.com/2220-9964/10/3/164.

Schultz, M., Voss, J., Auer, M., Carter, S., Zipf, A., 2017. Open land cover from OpenStreetMap and remote sensing. *International journal of applied earth observation and geoinformation*, 63, 206–213. 10.1016/j.jag.2017.07.014.

Sun, X., Wang, P., Yan, Z., Xu, F., Wang, R., Diao, W., Chen, J., Li, J., Feng, Y., Xu, T. et al., 2022. FAIR1M: A benchmark dataset for fine-grained object recognition in high-resolution remote sensing imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 184, 116–130.

Tensorflow, 2020. Tensorflow Detection Model Zoo: COCO-trained models.

Vargas-Munoz, J. E., Srivastava, S., Tuia, D., Falcao, A. X., 2020. OpenStreetMap: Challenges and opportunities in machine learning and remote sensing. *IEEE Geoscience and Remote Sensing Magazine*, 9(1), 184–199. https://10.1109/MGRS.2020.2994107.

Wu, Z., Li, H., Raifer, M., 2021. ohsome2label: historical OpenStreetMap objects to machine learning training samples. https://github.com/GIScience/ohsome2label.

Wu, Z., Li, H., Zipf, A., 2020. From historical openstreetmap data to customized training samples for geospatial machine learning. *Proceedings of the Academic Track at the State of the Map 2020 Online Conference*.

Xing, J., Sieber, R., 2021. Integrating XAI and GeoAI.

Zhang, Y., Zhou, Q., Brovelli, M. A., Li, W., 2022. Assessing OSM building completeness using population data. *International Journal of Geographical Information Science*, 0(0), 1-24. https://doi.org/10.1080/13658816.2021.2023158.

Zhou, Q., Jia, X., Lin, H., 2019. An approach for establishing correspondence between OpenStreetMap and reference datasets for land use and land cover mapping. *Transactions in GIS*, 23(6), 1420-1443. https://onlinelibrary.wiley.com/doi/abs/10.1111/tgis.12581.

Zhu, X. X., Hu, J., Qiu, C., Shi, Y., Kang, J., Mou, L., Bagheri, H., Haberle, M., Hua, Y., Huang, R., Hughes, L., Li, H., Sun, Y., Zhang, G., Han, S., Schmitt, M., Wang, Y., 2020. So2Sat LCZ42: A Benchmark Data Set for the Classification of Global Local Climate Zones [Software and Data Sets]. *IEEE Geoscience and Remote Sensing Magazine*, 8(3), 76-89.

Zook, M., Graham, M., Shelton, T., Gorman, S., 2010. Volunteered geographic information and crowdsourcing disaster relief: a case study of the Haitian earthquake. *World Medical & Health Policy*, 2(2), 7–33.