Estimating OpenStreetMap Missing Built-up Areas using Pre-trained Deep Neural Networks

Hao Li
GIScience Chair
Heidelberg University
Im Neuenheimer Feld
Heidelberg, Germany
hao.li@uni-heidelberg.de

Benjamin Herfort
GIScience Chair
Heidelberg University
Im Neuenheimer Feld
Heidelberg, Germany
herfort@uni-heidelberg.de

Alexander Zipf
GIScience Chair
Heidelberg University
Im Neuenheimer Feld
Heidelberg, Germany
zipf@uni-heidelberg.de

Abstract

Although built-up areas cover only a small proportion of the earth's surface, these areas are closely tied to most of the world's population and the economic output, which makes the mapping of built-up areas a vital challenge. Thanks to the generous contribution of volunteers, OpenStreetMap shows great capability in addressing this challenge, while the missing of maps is still a major concern. In this study, we propose a built-up areas mapping method by fine-tuning pre-trained deep neural networks, which aims to estimate OpenStreetMap missing built-up areas in a large-scale humanitarian mapping scenario. Specifically, we train an object detection network using very high resolution satellite images and corresponding OpenStreetMap building features. Then, we employ task-level labeling algorithms to produce the built-up estimation results and compare their accuracy performances with state-of-art baseline data sets. Considering the model transferability during scaling up to larger areas, we select two geographical independent areas in north Tanzania, Africa, for training and testing, respectively, where finished MapSwipe projects are available. Experiment results confirm that the pre-trained networks could yield high quality built-up maps and competitive estimation performances, which lead to over 75% of missing areas detection and over 92% of estimation overall accuracy.

Keywords: deep learning, crowdsourcing, built-up areas, OpenStreetMap, pre-trained deep neural networks, humanitarian mapping.

1 Introduction

Since the ever growth in the need of humanitarian support over the worldwide, missing maps project was launched in 2014 by American Red Cross, British Red Cross, Humanitarian OpenStreetMap Team (HOT) and Médecins Sans Frontières (MSF), which aimed at “mapping the most vulnerable places in the world” (Scholz et al., 2018). Taking advantages of satellite imagery, the mapping task of MapSwipe project is to collect volunteered geographical information (VGI), which indicates the demanding base map information (human settlements and roads) (Albuquerque et al., 2016).

VGI platforms, specifically OpenStreetMap (OSM), show great potential to support such humanitarian mapping tasks, while the availability of VGI data might remain a major concern (Barron et al., 2014). Fan et al. (2014) confirmed the high completeness and semantic accuracy of OSM building features in urban areas of Munich. While the OSM data availability in suburban areas mostly remains unrevealed.

Recently, deep learning methods, in particular, deep neural networks (DNNs), have been successfully employed for remote sensing tasks (Zhu et al., 2017), such as land cover classification, semantic segmentation, and object detection. However, the insufficiency of training samples is still one performance bottleneck of DNNs. OpenStreetMap has been believed as a powerful data source in providing massive and freely accessible geographic information, which can serve as a myriad of ground truth labeled samples for training. Mnih and Hinton (2012) successfully labeled satellite image with OSM vector data of streets and trained DNNs models for pixel-level street segmentation. In Kaiser et.al (2017), OSM was employed as training samples and compared with manual labels in semantic segmentation networks of buildings and roads. In Chen et.al (2018), an active learning framework named MC-CNN was proposed to incorporate multiple VGI data for humanitarian mapping in Malawi. Based on the aforementioned work, one interesting finding is that the sheer amount of VGI based training data could significantly compensate for their uneven quality and accuracy.

There is a rising trend in taking advantages of OSM data to boost the object detection performance of DNNs. However, little work has been contributed towards estimating OSM missing areas utilizing pre-trained DNNs. In this study, we aim to address the research question as follows:

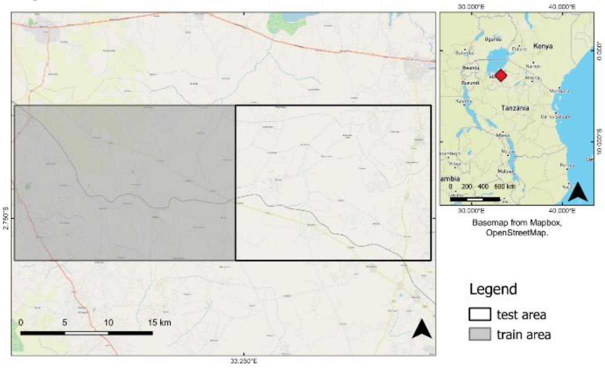
RQ: By fine-tuning pre-trained DNNs utilizing current OSM building data, to what degree can we estimate large-scale OSM missing built-up areas, which deserved further detailed mapping by volunteers?

We evaluate our proposed method in a case study of Tanzania and compare to state-of-art baseline data sets.

2 Data and Methods

We collect our experiment data from two areas in north Tanzania, Africa as the train and test areas in this study (Figure 1). Both areas have high levels of girls being subjected to female genital mutilation (FGM) and child marriage. In order to help NGOs to plan and implement outreach activities over this region, several MapSwipe mapping campaigns have been organized to collect VGI data of human settlements. In addition, it is necessary to notice the geographical independence of train and test areas, since the assumption is that few prior knowledge is available in the test area, which could be the most case in humanitarian mapping campaigns.

Figure 1: Overview of train and test area in Tanzania, Africa.



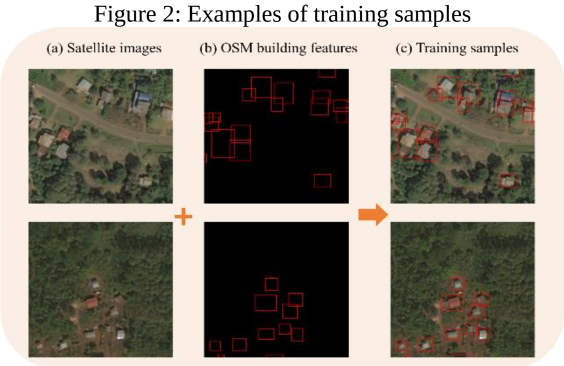
Source: OpenStreetMap and Mapbox.

2.1 Data sets

In this study, all satellite image tasks have been collected by requesting tile map service (TMS) from Bing satellite image at zoom level 18, which corresponds to the spatial resolution of roughly 0.6 meters. The size of all image tasks is 256*256 pixels.

Figure 2 shows examples of OSM based training samples. We extract the bounding boxes of OSM features with the Tag of "building", then label the corresponding images (For our object detection DNNs, object bounding boxes are required). The train area contains in total 19,256 tasks, while 3,574 tasks intersected with valid OSM building geometries (20,380 individual geometries) have been included as training samples.

Figure 2: Examples of training samples



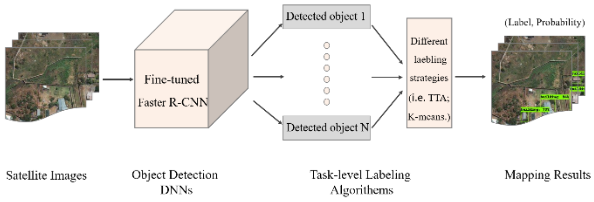
As for the test area, it contains in total 17,052 tasks, which covers approximately 417.77 km^2 rural areas. Since we aim to estimate the missing built-up areas in task-level, the reference data set considers not only the current OSM geometries but also an additional expert validation procedure. To be more specific, for tasks those do not intersect with any OSM geometry, a complementary crowdsourcing has been done by expert volunteers to validate whether the task contains at least one building. Therefore, this procedure provides us with a reliable reference.

In addition, we consider the MapSwipe mapping result, the Global Urban Footprint (GUF), and the High Resolution Settlement Layer (HRSL) as state-of-art baselines data sets in order to evaluate the estimation performance. The GUF has been derived automatically from TanDEM-X and TerraSAR-X radar images by the German Aerospace Center (DLR) (Esch et al., 2013). The HRSL consists of built-up areas generated from high resolution commercial satellite imagery by the Connectivity Lab at Facebook (Tiecke et al., 2017). The crowdsourced baseline of MapSwipe mapping results has been obtained using MapSwipe Analytics API (Heidelberg Institute for Geoinformation Technology. (2018)).

2.2 Methods

In this study, we propose a built-up areas mapping method based on pre-trained object detection DNNs (Figure 3). Our method intends to learn a robust building detection model from available OSM building features, which should be able to yield reliable prediction results even on independent test areas. Furthermore, the DNNs detection results are then employed for OSM missing built-up areas estimation purpose.

Figure 3: Workflow



For object detection DNNs, we select the Faster R-CNN (Ren et al., 2015) as a basic model. Since the heterogeneity of training samples is of crucial importance for robust generalization capability of DNNs, we consider initializing our Faster R-CNN with pre-trained hyperparameters on Microsoft COCO data set (Lin et al., 2014). Based on the OSM building features and satellite images, we then fine-tune the pre-trained Faster R-CNN model to detect only buildings. The raw prediction output would be a list of prediction bounding boxes together with confidence scores (Equation 1) for each individual task, which refers to the spatial coordinates and probability values of detected buildings. The implementation of pre-trained Faster R-CNN and the fine-tuning procedure are based on Python 3.6 and Tensorflow Object Detection API.

$$\begin{aligned} &\{P^i_{train} = (r^i_1, r^i_2, r^i_3 \dots) | i = 1,2,..,n\} \\ &\{P^j_{test} = (s^j_1, s^j_2, s^j_3 \dots) | j = 1,2,..,m\} \end{aligned} \qquad (1)$$

Where n, m refer to the number of train and test image tasks, and r^i, s^j denote to the confidence scores for train and test prediction objects, respectively.

Until here, the DNNs results in instance-level building detection, which already shows great potential in fine-grained building mapping. However, the first priority of humanitarian mapping is to identify task-level built-up areas, which would be further considered for detailed mapping in OSM. To achieve task-level classification, we modify the basic model with an additional task-level labeling algorithm (TLA).

Given the confidence scores (P^i_{train} and P^j_{test}) in Equation (1), TLA learns a classifier to generate the binary label of each task. In this study, we design a threshold transferring algorithm (TTA) as shown in Equation (2).

$$L^j_{test} = \begin{cases} 1 & if \max(P^j_{test}) \geq \theta_{train} \\ 0 & else. \end{cases} \quad (2)$$

Where L^j_{test} refers to the label of each test image task, and $\max(P^j_{test})$ is the corresponding probability value of positive built-up.

By learning an optimum threshold θ_{train} from the train area, we directly implement this threshold in the test area. Here, we optimize this threshold for the highest F1 score. To evaluate the transferring performance, we consider two variants of TLA for comparison: (1) TTA by using the optimal θ_{test} learned from test area TTA_θ_{test}; (2) K-means clustering binary labeling strategy (Lloyd (1982)).

To sum up, the proposed method enables us to map built-up areas in an unknown test area without any further training. The crowdsourcing workflow used in MapSwipe relies on the volunteer contribution to aggregate built-up labels, therefore, the data quality and mapping speed cannot be well controlled. In our case, the Faster R-CNN+TLA could be regarded as a machine volunteer, who gives consistent mapping results and keeps mapping without feeling tired. More importantly, we further exploit the potential of estimating OSM missing built-up area based on the machine-mapped results.

3 Experiments and Results

3.1 DNNs Building Detection

For Faster R-CNN, we employ the Inception V2 (Szegedy et al., 2016) as our base network and all parameters pre-trained in Microsoft COCO data set (Lin et al., 2014). Then, the fine-tuning procedure is run for 50,000 epochs with the initial learning rate of 0.00002 and a momentum value of 0.9. As for all TLA, a grid search is adopted to find the best threshold. To assess the performance of building detection, we consider the metrics precision, recall, f1 score, kappa coefficient (Cohen (1960)), overall accuracy (OA), as well as the numbers of false negatives (FN), false positives (FP), true negatives (TN), and true positives (TP). TTA_θ_{test} and k-means denote to two variants of TLA.

Table 1 shows the detailed mapping performance of three variants of the proposed method in comparison to current OSM data (OSM_Raw). The reference data is further validated based on current OSM data, so it consequently yields the highest Precision value. To this end, the FN of OSM Raw data reveals

Table 1: Detection Performance

Method	OSM Raw	Faster R-CNN		
		TTA	TTA θ_{test}	k-means
TP	2,218	4,048	3,994	4,131
FP	0	838	756	1,051
TN	12,362	11,524	11,606	11,311
FN	2,356	526	580	443
Precision	**100.00**	82.85	84.08	79.72
Recall	48.49	88.50	87.32	**90.32**
F1 score	65.31	85.58	**85.67**	84.69
Kappa	0.58	**0.80**	0.80	0.78
OA	86.09	91.95	**92.11**	91.18

the fact that 2,356 tasks are actually missing build-up areas, thus need further detail mapping. Aiming at automatic building detection, the proposed method could decrease the FN to below 600 tasks, which successfully detects over 75% of missing areas. Regarding recall, k-means variant leads to the highest value of 90.32%, while it also produces the most FP number of 1,051. It is easy to understand that by giving more positive predictions, one could result in higher Recall. Otherwise, higher precision, kappa coefficient as well as f1 score are achieved utilizing two TTA variants, which further confirms the effectiveness of the parametric TTA variants.

Due to the imbalanced distribution of building tasks in our study areas, it is obvious that most tasks are classified into no building class, therefore lead to general high Accuracy of over 85%. Therefore, the improvement of OA is less significant.

Figure 4 demonstrates the confusion maps of the proposed method as well as OSM Raw data. By filling most missing areas of current OSM data, it is believed that the Faster R-CNN+TLA could effectively map large-scale areas and produce high quality built-up areas maps in task-level, which ensures further accurate estimation of current OSM missing areas.

Figure 4: Confusion Maps

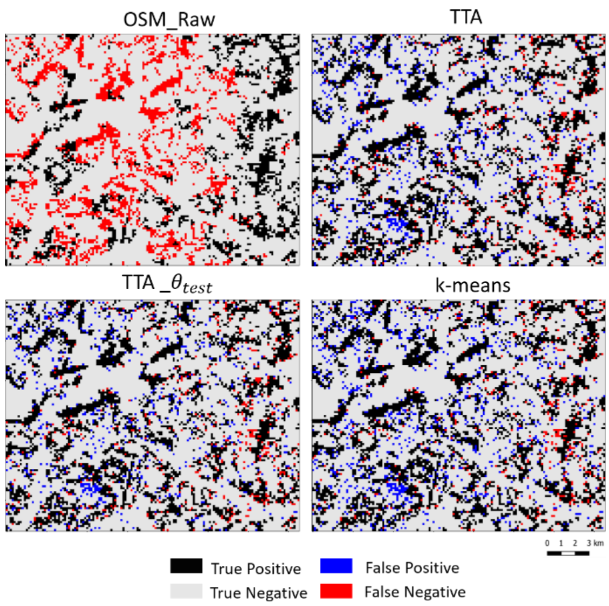


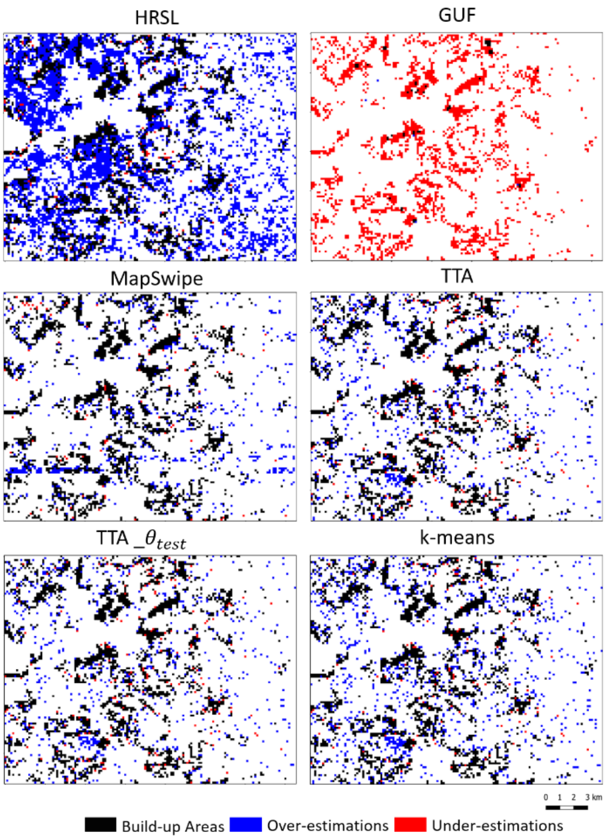
Table 2: Estimation Performance

Method	HRSL	GUF	MapSwipe	Faster R-CNN		
				TTA	TTA θ_{test}	k-means
OE	4278	5	748	838	756	1051
UE	118	2302	92	173	200	127
Precision	34.35	**91.53**	75.17	72.26	74.04	67.96
Recall	94.99	2.29	**96.10**	92.66	91.51	94.61
F1 score	50.45	4.47	**84.35**	81.20	81.85	79.10
OA	70.13	87.33	**94.29**	93.13	93.50	92.00

3.2 Missing Areas Estimation

As a common challenge during managing new humanitarian mapping campaigns, it is extremely difficult to estimate the demanding of volunteer effort to complete an entire campaign (Mark (2018)). In this context, our proposed method aims at addressing this challenge by providing task-level estimation results, which indicates the geographical location and area amount of OSM missing built-up areas. Then, such areas should be priorities for future detailed mapping by volunteers. Since our target is the built-up areas that are missing in OSM, we first exclude the OSM mapped areas, and then calculate the estimation accuracy on the rest. Specifically, the additional evaluation metrics consist of the numbers of over-estimations (OE), under-estimations (UE).

Figure 5: Estimation Maps



As shown in Table 2, our proposed method significantly outperforms to the GUF and HRSL in detecting built-up areas, and successfully achieves a similar estimation performance comparing to the crowdsourced MapSwipe results. Although MapSwipe still reports the highest values of F1 score and OA, the crowdsourced results heavily rely on the intensive contribution of volunteers, which makes it impossible to estimate the demanding volunteer effort before or during mapping campaigns. Otherwise, it is still optimistic to consider fully automatic workflow and supplanting volunteers by DNNs, so our future efforts would focus on how to integrate crowdsourcing and deep learning in order to develop more efficient mapping workflow.

From the visual interpretation of Figure 5, it is obvious that neither HRSL nor GUF yields satisfying estimation maps in the rural area with heterogeneous building structures. This fact further confirms the effectiveness of our proposed method in underrepresented rural areas. Besides, one may notice the different error distributions between crowdsourced MapSwipe and our proposed method. For instance, there is an apparent band cluster of OE in the lower left part of MapSwipe, while the OE of the proposed method tends to be randomly distributed. The possible reasons could be undesired volunteer behaviors and satellite image quality issues. In general, the Faster R-CNN+TLA method is relatively independent of artificial disturbing factors, therefore could scale much better than volunteer contributions.

4 Conclusion

In this study, we propose the Faster R-CNN+TLA method to estimate OSM missing built-up areas in a case study of Tanzania. Firstly, fine-tuned on very high resolution satellite images and corresponding OSM training samples, the proposed method could generate accurate built-up areas maps. The preliminary results in Tanzania show that our method could significantly reduce the missing building tasks by over 75% and achieve around 85% F1 score. Next, we evaluate the estimation performance of the proposed method regarding OSM missing built-up areas, which leads to competitive estimation results (over 92% OA) in comparison to the crowdsourced MapSwipe baseline. With respect to the RQ, it is worthwhile to develop a machine volunteer collaborating workflow by combining both methods, especially when considering large-scale estimation in heterogeneous regions.

More importantly, we intentionally select independent train and test areas in order to develop robust and transferable DNNs, which could be easily implemented to unmapped areas. Nevertheless, the transferring capability of DNNs deserves further study. In the future, we would include more study areas and adopt different pre-trained DNNs. We would also investigate the factors that may affect the model transferability, such as population density, land use land cover, and geographical terrain.

Towards integrating deep learning methods into the crowdsourcing applications for more intelligent workflow, our DNNs based method provides important insights into various applications. For instance, the accurate estimation of missing built-up areas could help project managers to make better plans before starting a new campaign. Moreover, the integration of machine-generated and crowdsourced data might significantly accelerate the mapping procedure while achieving similar or even higher accuracy as crowdsourcing workflow. The detailed workflow would be investigated in future work.

Acknowledgement: This work has been supported by the Klaus Tschira Stiftung (KTS) Heidelberg.

5 References

Scholz, S., Knight, P., Eckle, M., Marx, S. & Zipf. A. (2018) Volunteered Geographic Information for Disaster Risk Reduction—The Missing Maps Approach and Its Potential within the Red Cross and Red Crescent Movement. *Remote Sens.* 10, no. 8: 1239.

Albuquerque, J., 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), p. 859.

Barron, C., Neis, P., & Zipf, A. (2014) A Comprehensive Framework for Intrinsic OpenStreetMap Quality Analysis. *Transactions in GIS*, 18(6), 877–895

Fan, H., Zipf, A., Fu, Q. & Neis, P. (2014) Quality assessment for building footprints data on OpenStreetMap. *International Journal of Geographical Information Science*, 28, 700–719.

Zhu, X., Tuia, D., Mou, L., Xia, G., Zhang, L., Xu, F. & Fraundorfer, F. (2017) Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources *IEEE Geoscience and Remote Sensing Magazine*, vol. 5, no. 4, pp. 8-36.

Mnih, V. & Hinton, G. (2012) Learning to label aerial images from noisy data. *Proceedings of the 29th International Conference on International Conference on Machine Learning* (ICML'12), USA, 203-210.

Kaiser, P., Wegner, J. D., Lucchi, A., Jaggi, M., Hofmann, T. & Schindler, K. (2017) Learning Aerial Image Segmentation From Online Maps. *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 11, pp. 6054-6068.

Chen, J., Y. Zhou, A. Zipf & H. Fan. (2018) Deep Learning from Multiple Crowds: A Case Study of Humanitarian Mapping. *IEEE Transactions on Geoscience and Remote Sensing*. 1-10.

Esch, T., Marconcini, M., Felbier, A., Roth, A., Heldens, W., Huber, M., Schwinger, M., Taubenbock, H., Muller, A.& Dech, S. (2013) Urban Footprint Processor-Fully Automated Processing Chain Generating Settlement Masks From Global Data of the TanDEM-X Mission. *IEEE Geoscience and Remote Sensing Letters.* vol. 10, no. 6, pp. 1617-1621,

Tiecke, T.G., Liu, X., Zhang, A., Gros, A., Li, N., Yetman, G., Kilic, T., Murray, S., Blankespoor, B., Prydz, E.B. & Dang, H.A.H. (2017) Mapping the world population one building at a time. *arXiv:1712.05839*

Heidelberg Institute for Geoinformation Technology. (2018) MapSwipe Analytics API. *Available at: https://mapswipe.heigit.org/home/*

Ren, S., He, K., Girshick, R. & Sun., J. (2015) Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *arXiv:1506.01497*

Lin, T., Maire, M., Belongie, S.J., Bourdev, L.D., Girshick, R.B., Hays, J., Perona, P., Ramanan, D., Dollár, P. & Zitnick, C.L. (2014) Microsoft COCO: Common Objects in Context. *CoRR,abs/1405.0312.*

Lloyd, Stuart P. (1982) Least squares quantization in PCM. *IEEE Transactions on Information Theory 28.2: 129-137.*

Szegedy, C., Vanhoucke, V., Ioffe, S. & Shlens, J. (2016) Rethinking the Inception Architecture for Computer Vision. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA*

Cohen, J. (1960) A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 20(1):37-46.

Mark, W. (2018) Weaving automation into the mapping workflow: adding AI to the Tasking Manager. *Available at:https://medium.com/devseed/weaving-automation-into-the-mapping-workflow-adding-ai-to-the-tasking-manager-52172ed2ece8*