Mapping Slums with Deep Learning Feature Extraction

Agatha Mattos[1], Michela Bertolotto[1] and Gavin McArdle[1]

[1]*School of Computer Science, University College Dublin, Ireland*

Abstract

Many real-world problems present challenges that still have not been solved by the machine learning community, despite the high availability of satellite imagery and recent advances in computer vision. In particular, techniques which are cheaper and less reliant on large data sets are needed to map slums in cities. This study presents preliminary results using deep learning feature extraction followed by clustering using k-means, an unsupervised method, to detect slums in Sentinel-2 satellite imagery. The clusters that represented deprived areas in cities are identified using a data set which contains information about the topology of the urban areas derived from crowd-sourced digital maps. Overall, the unsupervised method performed worse than the baseline, a fine-tuned ResNet18 model (a supervised approach). The mean Intersection over Union for the two investigated locations (Mumbai and Capetown) was 0.46 and 0.51 for the supervised model, and 0.27 and 0.31 for the unsupervised model. Results suggest that other strategies for dealing with such imbalanced data sets need to be investigated to improve the results obtained for the slum class, and also strategies to automatically identify the clusters that represent deprived areas/slums. The code used in this paper is available at: https://github.com/ml-labs-crt/slums-unsupervised.

Keywords

Deep learning Feature Extraction, Slums, Deprived Areas, Machine Learning, Earth Observation

1. Introduction

The last decade saw a surge in the availability of satellite imagery and the development of image processing techniques. With this increase, it was expected that more societal challenges would be solved using remote sensing data and machine learning. However, many important societal problems have not yet completely benefited from the higher availability of imagery or current developments in computer vision. Many factors contribute to this situation, especially the high cost of acquiring and processing very-high-resolution satellite imagery [1, 2], and the lack of labelled data related to many societal problems, required to train supervised machine-learning models [3].

This work investigates the potential of employing freely available medium-resolution satellite imagery and feature extraction using deep learning, an unsupervised approach that does not require labelled data, to detect deprived/slum areas in two cities (Mumbai and Capetown). Slums, according to the United Nations Habitat, are locations where residents lack at least one of the following: water, sanitation, housing durability, security of tenure or sufficient living area [4]. The UN-Habitat estimates that over one billion people live in such conditions, but because most of the information about these settlements comes from outdated census surveys [5], there is an interest to explore other forms of data collection and processing that could provide current estimates [5, 1, 3]. The next section outlines the literature pertinent to slum mapping and further details the motivations for this paper.

1.1. Related Work

Since 2012, there has been a popularisation of deep learning architectures, and they have been shown to perform well in many classification tasks. In line with this trend, the research to map slums moved from traditional image processing approaches to supervised learning methods using deep learning and high or very-high-resolution imagery [6]. In 2017, Mboga et al. [7] and Persello and Stein [8] demonstrated that convolutional neural networks outperformed feature extraction methods and since then, many works employing neural networks to map slums have been published.

However, the great majority of studies to date rely on supervised learning and costly high or very-high satellite imagery [6], and hence consider only small areas [5, 2]. Additionally, many researchers have found that models developed for one city do not generalise well to other areas [9, 10, 1]. For a global slum inventory to be possible, these issues need to be tackled, and unsupervised learning may be a suitable alternative.

Nonetheless, the literature on mapping slums with unsupervised learning techniques is limited. To the best of our knowledge, [11] and [12] are currently the most representative works, though both have limitations. Block et al. [11] employs high-resolution imagery and St. Amand [12] relies heavily on visual inspection for decision making. This paper presents our initial results of developing a pipeline to map slums using freely available medium-resolution satellite imagery, unsupervised learning and

CDCEO 2022: 2nd Workshop on Complex Data Challenges in Earth Observation, July 25, 2022, Vienna, Austria
agatha.hennigendemattos@ucdconnect.ie (A. Mattos); michela.bertolotto@ucd.ie (M. Bertolotto); gavin.mcardle@ucd.ie (G. McArdle)

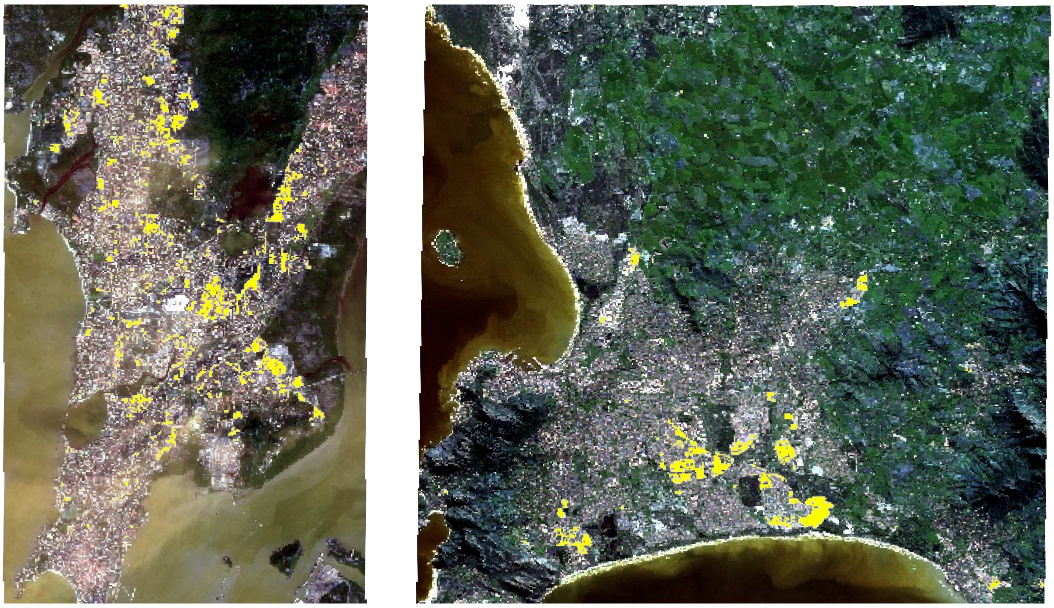

Figure 1: Left: Sentinel-2 image with ten meters resolution of Mumbai, India. The yellow polygons were annotated as slum areas and were used as ground truth in the supervised model. Right: Similar to the image on the left, but of Capetown, South Africa.

automated classification of slum clusters using topological information derived from crowd-sourced digital maps. In the next section, the methodology used in this study is described.

2. Methodology

Two locations were used to investigate the potential of feature extraction using deep learning and posterior clustering: Mumbai, in India, and Capetown, in South Africa. The satellite imagery was collected by Gram-Hansen et al. [1] and consists of Sentinel-2 images with ten metres resolution. These cities have been investigated by other researchers, and hence are ideal for testing the proposed unsupervised method. As in Block et al. [11]'s experiments, three bands were used (blue, green and red) and the imagery was scaled from 16-bit to 8-bit. Figure 1 shows the satellite imagery of the locations.

The imagery was split into tiles of 20 by 20 pixels (approximately 200 x 200 metres), slightly bigger than those used by Taubenböck et al. [13], who also adopted medium-resolution imagery in their research. The baseline model to which the unsupervised approach was compared was a supervised model trained with ground-truth data collected by Gram-Hansen et al. [1]. For a tile to be considered as belonging to a certain class, at least 50% of the pixels in that tile would have to be from that class. As expected, this is a hugely imbalanced data set, as only 3% of the tiles are slums in Mumbai and less than 1% in Capetown. Table 1 presents an analysis of the areas covered in this paper.

As suggested by other researchers [14] and to mimic a real-world scenario, only 20% of the slum tiles were used to train the model. Also, the non-slum class was undersampled with a proportion of 4 to 1, in an attempt to account for the imbalance in the data set. The remaining 80% of the slum tiles and non-slum tiles were used to test the model. As a result, the baseline was trained with 399 tiles (80 slums and 319 non-slums), in the case of Mumbai, and with 532 tiles (106 slums and 426 non-slums) for Capetown. The code used in this paper is available at https://github.com/ml-labs-crt/slums-unsupervised.

2.1. Baseline

The model adopted as the baseline was a fine-tuned ResNet18, trained initially on ImageNet images. The supervised model choice followed from the results obtained by Bell and Veeraraghavan [15], who tested ResNet models of different sizes. Both the supervised model (baseline) and the unsupervised model were implemented in PyTorch 1.10.2. The supervised model was trained using a batch size of 8 and for 50 epochs. Early stopping was triggered when the average loss of the validation set was 20% higher than the average of the last 10 epochs. The re-

Table 1
Analysis of the Areas

Location	Height # of pixels	Width # of pixels	Non-slum # of tiles	Slum # of tiles	Non-slum % of total tiles	Slum % of total tiles
Mumbai	3920	1980	18831	573	97.0%	3.0%
Capetown	6080	5300	79799	761	99.1%	0.9%

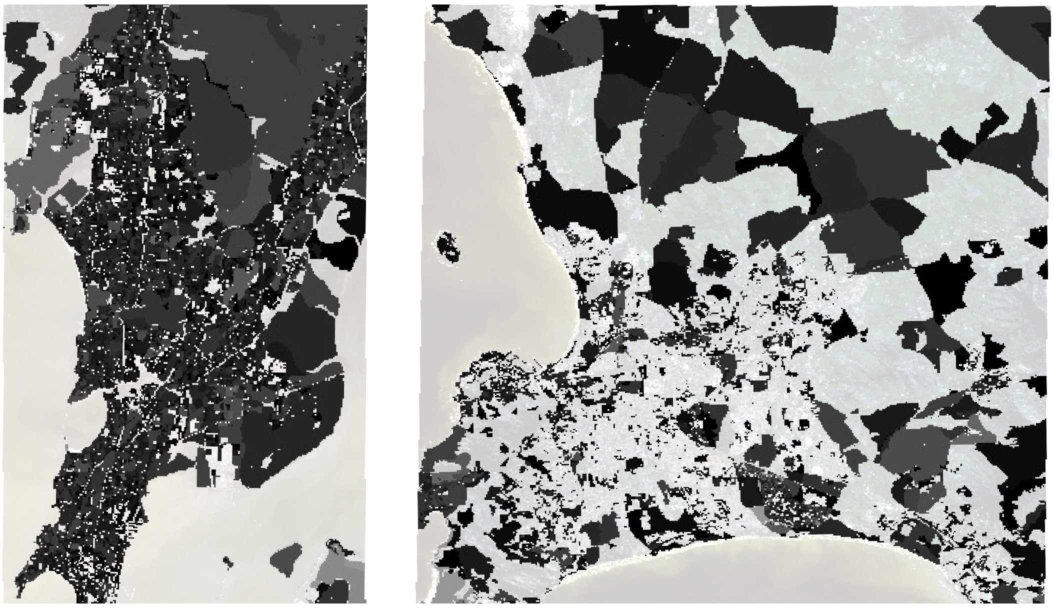

Figure 2: Left: Complexity scores for Mumbai, India. Darker polygons denote smaller complexity scores. In the background, Sentinel-2 satellite imagery of the same city. Right: Similar to the image on the left, but of Capetown, South Africa.

sults were evaluated using Intersection over Union (IoU), as is commonly done in the related literature.

2.2. Unsupervised Approach

The unsupervised model's features were extracted using a ResNet18 model pre-trained with the ImageNet data set. Care was taken so that the exact same tiles were used to train both models. The extracted features for each tile (a vector with 1000 rows) were subsequently fitted to a k-means model initialised using sklearn's default initialisation and 100 repetitions. The number of repetitions was set following from Fränti and Sieranoja [16]. The number of clusters chosen was seventeen, and it was selected based on Taubenböck et al. [17]'s work, who analysed satellite imagery of 110 cities worldwide using the Local Climate Zones Classification Scheme (that has seventeen different climate zones).

Lastly, to decide which clusters should be considered slums and which should be labelled as non-slums, the complexity score designed by Soman et al. [18] was leveraged. Figure 2 shows the complexity score for the two areas investigated in this paper. The complexity score for Mumbai ranged between 0 and 20, and for Capetown, between 0 and 18. Lower scores denote less developed areas. This complexity score was set based on information available on OpenStreetMap. For this reason, some locations within the city do not have a complexity score. In the case of Mumbai, 41% of all pixels did not have a complexity score (mostly areas where water bodies are) and in Capetown that was the case for 63% of the pixels. The median complexity score of each cluster was calculated using the average complexity score of the pixels in each tile. Subsequently, clusters with the lowest values of median complexity score were assigned as "slum clusters" (see details of each ones on Table 2). In the next section, the results are discussed.

Table 2
Number of Tiles in Each Cluster and Median Complexity of Clusters. In Bold, Clusters That Were Identified as “Slum Clusters”

	Mumbai				Capetown			
Cluster_ID	Non-slum # of tiles	Slum # of tiles	Tiles per Cluster %	Median Complexity	Non-slum # of tiles	Slum # of tiles	Tiles per Cluster %	Median Complexity
0	1812	37	9.8%	**3.01**	1452	35	1.9%	**2.87**
1	698	18	3.8%	**3.00**	6553	72	8.3%	3.00
2	1252	25	6.8%	3.89	3760	2	4.7%	3.00
3	272	5	1.5%	**2.95**	1413	9	1.8%	**2.21**
4	1708	32	9.2%	3.49	5853	4	7.3%	3.00
5	1887	62	10.3%	**3.00**	5147	56	6.5%	**2.77**
6	6	0	0.0%	3.75	4588	6	5.8%	3.00
7	1461	69	8.1%	**3.00**	4276	42	5.4%	**2.58**
8	1922	28	10.4%	3.61	8238	54	10.4%	3.00
9	1038	21	5.6%	**3.00**	7036	26	8.8%	3.00
10	1951	22	10.5%	3.73	3121	37	4.0%	3.00
11	18	3	0.1%	3.77	6944	37	8.7%	**2.91**
12	1133	35	6.2%	4.00	5088	73	6.5%	**2.49**
13	1102	35	6.0%	3.60	3006	5	3.8%	3.00
14	556	9	3.0%	3.79	5808	32	7.3%	3.00
15	829	22	4.5%	**3.07**	5167	106	6.6%	**2.16**
16	730	36	4.1%	**3.02**	1741	13	2.2%	**2.74**
Total	18375	459	100%		79191	609	100%	

3. Results and Discussion

The extraction of features using deep learning was carried out for two locations (Mumbai and Cape Town). For Mumbai, the percentage of tiles assigned to each cluster was in the range of 0.03% to 10.5%, and for Capetown it was in the range of 1.8% to 10.4%. Using the ground-truth data, it was possible to observe that some clusters did contain most of the slum tiles; for instance, clusters 5 and 7 for Mumbai contained 13.5% and 15% of the total slums tiles. Similarly, clusters 1, 12 and 15 for Capetown contained 11.8%, 12.0% and 17.4% of all slum tiles. Table 2 describes the number of tiles assigned to each cluster.

As mentioned in Section 2, the decision of which clusters would be considered “slum clusters” took into consideration the average complexity of the pixels of each tile in that cluster. Though Soman et al. [18] suggests in their paper that areas with a complexity score smaller than 5 or 6 could be considered informal settlements, in the cities covered in this study, this would result in all clusters being labelled as slums. For example, for Capetown the median complexity for all clusters was in the range of 2.16 to 3.0. In the case of Mumbai it was in the range of 2.95 to 4.0. For this reason, only clusters that had a complexity below the median cluster complexity for each location were considered ”slum clusters”. In the case of Mumbai, it meant clusters with a median complexity below 3.49 and for Capetown clusters with a median complexity below 3.0 (see Table 2). All tiles in the so-called “slum clusters” were then assigned a slum label and compared with the ground truth to obtain Intersection over Union (IoU) scores that could be compared to the baseline results obtained with the supervised model. Due to all clusters having a non-negligible amount of non-slum tiles in them (see Table 2), overall, the unsupervised learning model performed worse than the supervised method. Figure 3 shows a visualisation of the clusters and Table 3 has the intersection over union (IoU) for each class and for each model.

Both models had an intersection over union (IoU) below 0.10 for the slum class, caused by tiles being classified as slums even when they were not labelled like that in the ground-truth data. The obtained results suggest that oversampling the non-slum areas with a 4 to 1 ratio may not be an appropriate strategy for dealing with the huge imbalance in this problem. Moreover, the use of complexity scores needs further investigation to determine the best strategy to set the complexity threshold for each location. In the way that it was employed in this experiment, it did not help identify the less developed/slums clusters. Other parameters set in the experiment may need to be reviewed to increase performance, such as the tile dimension and number of clusters.

Nonetheless, the mean IoU of the unsupervised method outperformed the results obtained by Gram-Hansen et al. [1] in the case of Capetown (0.17 versus 0.31) and was only slightly worse than the case of Mumbai (0.40 versus 0.27). The intersection over union (IoU) for the slum class, however, was smaller than obtained by Gram-Hansen et al. [1] for both locations. Still, Gram-Hansen et al.

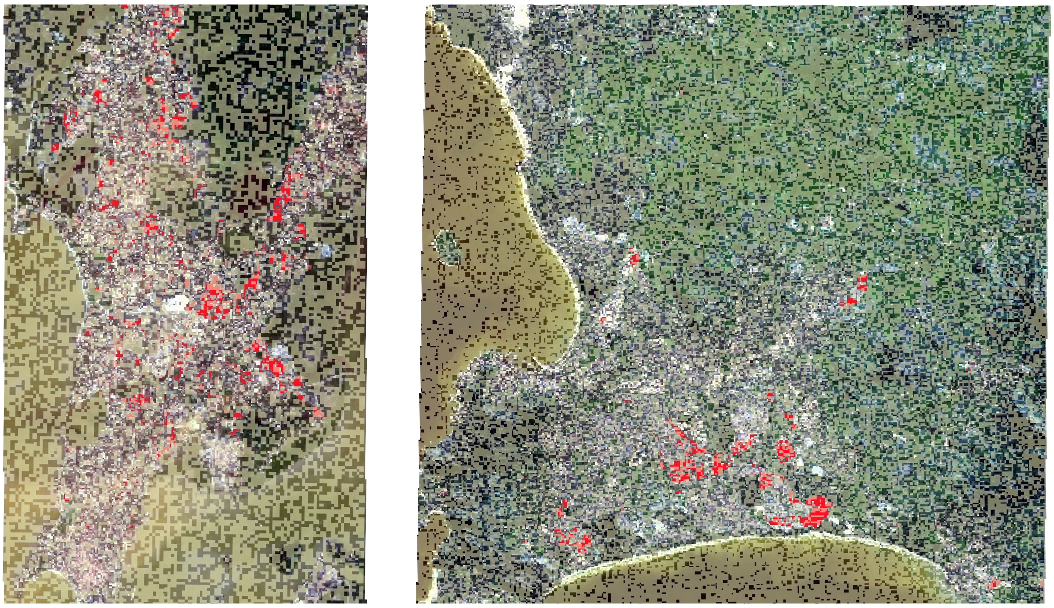

Figure 3: Left: Areas in yellow were predicted as non-slums (unsupervised approach). Areas in red were labelled as slums (ground truth). In the background, satellite imagery of Mumbai. Right: Similar to the image on the left, but of Capetown.

Table 3
Results of the Binary Classification of Urban Areas into Slum/Non-slum Classes Using Intersection over Union (Iou)

	Supervised learning			Unsupervised learning		
Location	IoU Non-slum	IoU Slum	mean IoU	IoU Non-slum	IoU Slum	mean IoU
Mumbai	0.84	0.08	0.46	0.52	0.03	0.27
Capetown	0.95	0.07	0.51	0.60	0.01	0.31
All locations	0.90	0.08	0.49	0.56	0.02	0.29

[1] used convolutional neural networks and very-high-resolution imagery (30cm per pixel) in their experiments, which indicates that unsupervised learning and freely available medium-resolution imagery can be promising for this real-world application.

4. Conclusions and Future Work

This experiment presents the initial results of an attempt to use deep learning feature extraction and unsupervised learning to map slums. Results demonstrate that the proposed method performed worse than the baseline, a supervised learning approach.

Looking to the future, it would be desirable to investigate strategies to improve the results for the slum class, such as oversampling the slum class to the point of eliminating the imbalance, as suggested in [19], or adopting more sophisticated sampling for the non-slum class. It is also possible that more traditional image processing techniques could be used to mask out regions that are clearly not urban, such as water and vegetation. These changes would reduce the total number of non-slum tiles and potentially make the problem less imbalanced. Additionally, the adoption of block complexity derived from crowd-sourced digital maps requires further investigation to determine its usability as a tool to identify clusters that represent deprived areas/slums. Performing feature extraction using a deep learning model pre-trained with a remote sensing data, as opposed to ImageNet, may also be beneficial. Also, it would be interesting to see a comparison of the deep features extracted from medium-resolution satellite imagery and very-high-resolution imagery for the same location with the intention of confirming that the former can satisfactorily be employed for mapping slums using unsupervised learning. Lastly, to develop a global slum inventory, the analysis developed here could be extended to estimate the population living in the areas identified as deprived/slums.

Acknowledgments

This publication has emanated from research supported in part by a grant from Science Foundation Ireland under Grant number 18/CRT/6183. For the purpose of Open Access, the author has applied a CC BY public copyright license to any Author Accepted Manuscript version arising from this submission.

References

[1] B. J. Gram-Hansen, P. Helber, I. Varatharajan, F. Azam, A. Coca-Castro, V. Kopackova, P. Bilinski, Mapping Informal Settlements in Developing Countries using Machine Learning and Low Resolution Multi-spectral Data, in: Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society, AIES '19, Association for Computing Machinery, New York, NY, USA, 2019, pp. 361–368. doi:10.1145/3306618.3314253.

[2] A. C. H. de Mattos, G. McArdle, M. Bertolotto, Mapping Slums with Medium Resolution Satellite Imagery: A Comparative Analysis of Multi-Spectral Data and Grey-level Co-occurrence Matrix Techniques, 2021. doi:10.48550/arXiv.2106.11395. arXiv:2106.11395.

[3] M. Burke, A. Driscoll, D. Lobell, S. Ermon, Using Satellite Imagery to Understand and Promote Sustainable Development, Working Paper 27879, National Bureau of Economic Research, 2020. doi:10.3386/w27879.

[4] U. N. Habitat, Tracking progress towards inclusive, safe, resilient and sustainable cities and human settlements (2018).

[5] R. Mahabir, A. Croitoru, A. T. Crooks, P. Agouris, A. Stefanidis, A Critical Review of High and Very High-Resolution Remote Sensing Approaches for Detecting and Mapping Slums: Trends, Challenges and Emerging Opportunities, Urban Science 2 (2018) 8. doi:10.3390/urbansci2010008.

[6] M. Kuffer, K. Pfeffer, R. Sliuzas, Slums from Space—15 Years of Slum Mapping Using Remote Sensing, Remote Sensing 8 (2016) 455. doi:10.3390/rs8060455.

[7] N. Mboga, C. Persello, J. R. Bergado, A. Stein, Detection of informal settlements from VHR satellite images using convolutional neural networks, in: 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2017, pp. 5169–5172. doi:10.1109/IGARSS.2017.8128166.

[8] C. Persello, A. Stein, Deep Fully Convolutional Networks for the Detection of Informal Settlements in VHR Images, IEEE Geoscience and Remote Sensing Letters 14 (2017) 2325–2329. doi:10.1109/LGRS.2017.2763738.

[9] Y. Gao, Assessing the spatial transferability of fully convolutional networks for slum mapping (2020).

[10] T. Stark, M. Wurm, X. X. Zhu, H. Taubenböck, Satellite-Based Mapping of Urban Poverty With Transfer-Learned Slum Morphologies, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 13 (2020) 5251–5263. doi:10.1109/JSTARS.2020.3018862.

[11] J. Block, M. Yazdani, M. Nguyen, D. Crawl, M. Jankowska, J. Graham, T. DeFanti, I. Altintas, An Unsupervised Deep Learning Approach for Satellite Image Analysis with Applications in Demographic Analysis, in: 2017 IEEE 13th International Conference on E-Science (e-Science), 2017, pp. 9–18. doi:10.1109/eScience.2017.13.

[12] F. St. Amand, Identification of Slums in Mumbai, India: Unsupervised Classification Techniques, Thinking Matters Symposium Archive (2014).

[13] H. Taubenböck, H. Debray, C. Qiu, M. Schmitt, Y. Wang, X. X. Zhu, Seven city types representing morphologic configurations of cities across the globe, Cities 105 (2020) 102814. doi:10.1016/j.cities.2020.102814.

[14] G. Leonita, M. Kuffer, R. Sliuzas, C. Persello, Machine Learning-Based Slum Mapping in Support of Slum Upgrading Programs: The Case of Bandung City, Indonesia, Remote Sensing 10 (2018) 1522. doi:10.3390/rs10101522.

[15] B. Bell, R. Veeeraraghavan, Locating informal urban settlements, in: AI for Social Good Workshop, 2020.

[16] P. Fränti, S. Sieranoja, How much can k-means be improved by using better initialization and repeats?, Pattern Recognition 93 (2019) 95–112. doi:10.1016/j.patcog.2019.04.014.

[17] H. Taubenböck, M. Weigand, T. Esch, J. Staab, M. Wurm, J. Mast, S. Dech, A new ranking of the world's largest cities—Do administrative units obscure morphological realities?, Remote Sensing of Environment 232 (2019) 111353. doi:10.1016/j.rse.2019.111353.

[18] S. Soman, A. Beukes, C. Nederhood, N. Marchio, L. M. A. Bettencourt, Worldwide Detection of Informal Settlements via Topological Analysis of Crowdsourced Digital Maps, ISPRS International Journal of Geo-Information 9 (2020) 685. doi:10.3390/ijgi9110685.

[19] M. Buda, A. Maki, M. A. Mazurowski, A systematic study of the class imbalance problem in convolutional neural networks, Neural Networks 106 (2018) 249–259. doi:10.1016/j.neunet.2018.07.011. arXiv:1710.05381.