MAPPING INFORMAL SETTLEMENTS IN DEVELOPING COUNTRIES WITH MULTI-RESOLUTION, MULTI-SPECTRAL DATA

Patrick Helber[*,1,2], **Bradley Gram-Hansen**[*,3], **Indhu Varatharajan**[4], **Faiza Azam**[5], **Alejandro Coca-Castro**[6], **Veronika Kopackova**[7], **Piotr Bilinski**[3,8]
[1]German Research Center for Artificial Intelligence (DFKI), Germany
[2]TU Kaiserslautern, Germany
[3]University of Oxford, UK
[4]German Aerospace Center (DLR), Germany
[5]Independent researcher, Germany
[6]King's College London, UK
[7]Czech Geological Survey, Czechia
[8]University of Warsaw, Poland

ABSTRACT

Detecting and mapping informal settlements encompasses several of the United Nations sustainable development goals. This is because informal settlements are home to the most socially and economically vulnerable people on the planet. Thus, understanding where these settlements are is of paramount importance to both government and non-government organizations (NGOs), such as the United Nations Childrens Fund (UNICEF), who can use this information to deliver effective social and economic aid. We propose two effective methods for detecting and mapping the locations of informal settlements. One uses only low-resolution (LR), freely available, Sentinel-2 multispectral satellite imagery with noisy annotations, whilst the other is a deep learning approach that uses costly very-high-resolution (VHR) satellite imagery. This is in contrast to previous studies that only use costly VHR satellite and aerial imagery. To our knowledge, we are the first to map informal settlements successfully with LR satellite imagery. We extensively evaluate and compare the proposed methods. We make the code and the dataset publicly available at `https://frontierdevelopmentlab.github.io/informal-settlements/`.

1 INTRODUCTION

The United Nations (UN) and the Organisation for Economic Co-operation and Development (OECD) state that informal settlements are defined as follows OECD (2008); United Nations (2012):

1. Inhabitants have no security of tenure vis-à-vis the land or dwellings they inhabit, with modalities ranging from squatting to informal rental housing.
2. The neighborhoods usually lack, or are cut off from, basic services and city infrastructure.
3. The housing may not comply with current planning and building regulations, and is often situated in geographically and environmentally hazardous areas.

Slums are the most deprived and excluded form of informal settlements. They can be characterized by poverty and large agglomerations of dilapidated housing, located in the most hazardous urban land, near industries and dump sites, in swamps, degraded soils and flood-prone zones Kohli et al. (2016). In addition to tenure insecurity, slum dwellers lack basic infrastructure and services, public

[*]authors contributed equally

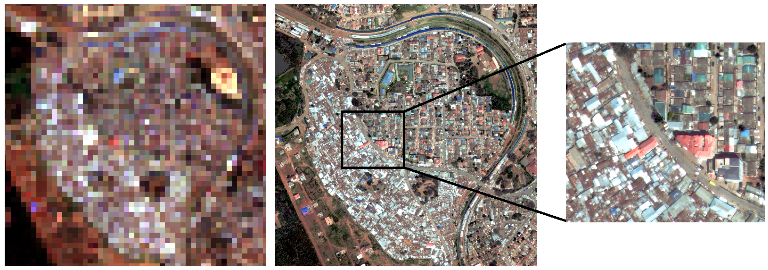

Figure 1: Two images of the same informal settlement in Kibera, Nairobi representing the difference between high and low resolution imagery. *Left*: The Sentinel-2 10 m resolution image. *Right*: A DigitalGlobe 30 cm VHR image. Also, a detailed view of a part of the VHR image is shown.

space and green areas, and are constantly exposed to eviction, disease and violence Sclar et al. (2005). The ability to map and locate these settlements would give organizations such as UNICEF and other related organizations the ability to provide effective social and economic aid Pais (2002). This in turn enables those communities to evolve in a sustainable way, allowing the people living in those environments to gain a much better quality of life.

1.1 Data

In this paper, we have annotated low-resolution (LR) and very-high-resolution (VHR) satellite imagery for the locations of informal settlements in parts of Kenya, South Africa, Nigeria, Sudan, Colombia and Mumbai.

Sentinel-2 Multi-spectral Satellite Data The Sentinel-2 mission is part of the Copernicus programme, a global earth observation service. The Sentinel-2 satellites map the entire global land mass on average every 5 days at various resolutions of 10 to 60 m per pixel. Sentinel-2 provides multiple-resolution, multi-spectral imagery using a multi-spectral instrument (MSI). MSI measures top of the atmosphere radiances in 13 spectral bands covering the visible, near infrared and the short-wave infrared part of the electromagnetic spectrum at different spatial resolution depending upon the particular band European Space Agency (2018); Zhang et al. (2017). We only use 10 bands due to atmospheric distortions. The resolution of up to 10 m denotes that each pixel represents a 10 $m \times 10$ m surface, which means that there is a certain amount of contextual information contained within one pixel. By observing the spectral signal, which provides us with the chemical composition of each pixel, we can extract this contextual information.

VHR Satellite Images In addition to freely available multi-spectral LR satellite images, we use VHR images with a resolution of up to 30 cm per pixel, kindly provided by DigitalGlobe through the Satellite Applications Catapult. See Figure 1 to view the difference in the resolution between Sentinel-2 images with a spatial resolution of up to 10 m per pixel and the VHR imagery with a spatial resolution of approximately 30 cm per pixel.

2 Methods

Our **first method** trains a classifier to learn what the spectral signal of an informal settlement is, using LR freely available spectral data. To do this, we employ a pixel-wise classification, where the classifier learns whether or not a 10-band spectra is associated to an informal settlement or the environment, which encompasses everything that is not an informal settlement.

When we require finer grained features, such as the roof size, or the density of the surrounding settlements to determine whether or not there exists an informal settlement, we demonstrate our **second method**, which trains a state-of-the-art semantic segmentation deep neural network that uses VHR satellite imagery. This is crucial when informal settlements do not have unique spectra when compared to the environment, like those in Al Geneina, Sudan, see Figure 3.

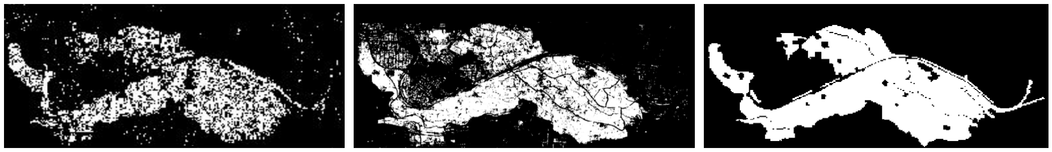

Figure 2: Predictions of informal settlements (white pixels) in Kibera, Nairobi. *Left:* The CCF prediction of informal settlements in Kibera on LR Sentinel-2 spectral imagery. *Middle:* CNN based prediction of informal settlements in Kibera, trained on VHR imagery. *Right:* The ground truth informal settlement mask for Kibera.

Pixel-wise Classification with Canonical Correlation Forest Canonical Correlation Forests (CCFs) Rainforth & Wood (2015) are a decision tree ensemble method for classification and regression. CCFs are the state-of-the-art random forest technique, which have shown to achieve remarkable results for numerous regression and classification tasks Rainforth & Wood (2015). Individual canonical correlation trees are binary decision trees with hyperplane splits based on local canonical correlation coefficients calculated during training. Like most random forest based approaches, CCFs have very few hyper-parameters to tune and typically provide very good performance out of the box. The only parameter that has to be set is the number of trees, n_{trees}. For CCFs, setting $n_{trees} = 15$ provides a performance that is empirically equivalent to a random forest that has $n_{trees} = 500$ Rainforth & Wood (2015), meaning CCFs have lower computational costs, whilst providing higher classification accuracy when compared to the classical random forest CCFs work by using canonical correlation analysis (CCA) and projection bootstrapping during the training of each tree, which projects the data into a space that maximally correlates the inputs with the outputs. This is particularly useful when we have small datasets, as for our case, because it reduces the amount of artificial randomness required to be added during the tree training procedure and improves the ensemble predictive performance Rainforth & Wood (2015).

Contextual Classification with CNNs Since informal settlements can also be classified by the rooftop size and the surrounding building density, we employ a state-of-the-art semantic segmentation neural network on optical (RGB) VHR satellite imagery to detect these contextual features. Contextual features are important when it is not possible to distinguish informal settlements from the environment by spectral signal in the same region. An example of such an informal settlement is shown in Figure 3. We see that the informal settlements in a rural region of Al Geneina, Sudan have a very low building density, and also the roof tops of both formal and informal settlements are built out of concrete, meaning they have the same spectral signal. This is in contrast to the dense slums in Nairobi and Mumbai.

Figure 3: VHR images comparing an informal, *left* and formal settlement, *right*, in Al Geneina, Sudan. The main distinguishing feature is the wider contextual information, as the material spectrums are the same.

Encoder-Decoder with Atrous Separable Convolution We use the DeepLabv3+ encoder-decoder architecture. DeepLabv3+ Chen et al. (2018b) is a deep CNN that extends the prior DeepLabv3 network Chen et al. (2018a) with a decoder module to refine the segmentation results of the previous encoder-decoder architecture particularly at the object boarders. The DeepLab architecture uses Atrous Spatial Pyramid Pooling (ASPP) with atrous convolutions to explicitly control the resolution at which feature responses are computed within the CNN. This ASPP module is augmented with image level features to capture longer range information. We use a Xception 65 network backbone in the encoder-decoder architecture. The beneficial use of this Xception model together with applying depth wise separable convolution to ASPP and the decoder modules have been shown in Chen et al. (2018b).

3 RESULTS

Experimental Setup For each region we have a 10 m LR Sentinel-2 image, the corresponding VHR 30-50 cm resolution image and the ground truth annotations. When training and validating a model on the same region we use a 80-20 split. We ensure that each class contains the same number of points, we then randomly sample 80% of each class to generate the training data and then use the remaining 20% of each class to construct our test set, which is comprised of a different set of points. We then center the training data (testing data accordingly) to have a mean of zero and standard deviation of one. We set the $n_{trees} = 10$ for training the CCF. For validating our methods we report both pixel accuracy, and mean intersection over union.

We provide a comparison of both the pixel-wise classification with CCFs and the contextual classification with CNNs for the detection and mapping of informal settlements, see on the left side of Table 1. The CCFs trained solely on freely available and easily accessible LR data perform well, although they are unable to match the performance of the CNN trained on VHR imagery, except for Kibera. Figure 2 shows the predictions of both methods and the ground truth annotations. Despite having access to VHR data, the CNN still manages to miss-classify structural elements of the informal settlements in Kibera. Whereas the CCF, although more granular, incorporates the full structure of the informal settlement in Kibera via only the spectral information.

Table 1: *Left-half*: Pixel accuracy and mean IOU (%) results for informal settlement detection using the CCF pixel-wise classification and the contextual classification with CNNs. CCFs are trained and tested on LR imagery, CNNs are trained and tested on VHR imagery. *Right-half*: Using the CCFs we train a model on Northern Nairobi (NN) and a model trained on Medellin (M), then predict on all regions. *ground truth annotations are less than 75% complete for the region.

	Pixel Acc.		Mean IOU		Pixel Acc.		Mean IOU	
Region	**CCF(LR)**	**CNN(VHR)**	**CCF(LR)**	**CNN(VHR)**	**NN**	**M**	**NN**	**M**
Kenya, Northern Nairobi	69.4	93.1	62.0	80.8	69.4	55.0	62.0	54.4
Kenya, Kibera	69.0	78.2	73.3	65.5	67.3	63.8	54.1	56.0
South Africa, Capetown*	92.0	-	33.2	-	41.3	71.5	43.1	32.0
Sudan, El Daien	78.0	86.0	61.3	73.4	14.2	1.1	37.9	34.0
Sudan, Al Geneina	83.2	89.2	35.7	76.3	27.1	6.0	34.9	41.0
Nigeria, Makoko*	76.2	87.4	59.9	74.0	59.0	77.0	37.8	34.6
Colombia, Medellin*	84.2	95.3	74.0	83.0	65.0	84.2	46.9	74.0
India, Mumbai*	97.0	-	40.3	-	37.9	63.0	32.4	34.4

4 CONCLUSION

In this work we have provided benchmarks for detecting informal settlements and have proposed two different methods for detecting informal settlements. The first method uses computationally efficient CCFs to learn the spectral signal of informal settlements from multispectral LR satellite imagery. The second trains a CNN with VHR satellite imagery to extract finer grained features. We extensively evaluated the proposed methods and demonstrated the generalization capabilities of the computationally efficient CCFs to detect informal settlements globally. In particular, to the best of our knowledge, we demonstrated for the first time that informal settlements can be detected effectively using only freely and openly accessible multi-spectral LR satellite imagery.

ACKNOWLEDGMENTS

This project was executed during the Europe program of the Frontier Development Lab (FDL). We gratefully acknowledge the support of Phi-Lab at ESA, Nvidia Corporation, Satellite Applications (SA) Catapult, Oxford University, Kellogg College, UNICEF and AIData. Patrick Helber was supported by the NVIDIA AI Lab program and the BMBF project DeFuseNN (Grant 01IW17002). Bradley Gram-Hansen was supported by the UK EPSRC CDT in Autonomous Intelligent Machines and Systems.

REFERENCES

Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4): 834–848, 2018a.

Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 801–818, 2018b.

European Space Agency. `https://sentinel.esa.int/documents/247904/685211/Sentinel-2_User_Handbook`, 2018. Accessed : 2018-08-27.

Divyani Kohli, Richard Sliuzas, and Alfred Stein. Urban slum detection using texture and spatial metrics derived from satellite imagery. *Journal of Spatial Science*, 61(2):405–426, 2016. doi: 10.1080/14498596.2016.1138247.

OECD. *OECD Glossary of Statistical Terms*. OECD glossaries OECD glossary of statistical terms. OECD Publishing, 2008. ISBN 9789264055087. URL `https://books.google.co.uk/books?id=iwTWAgAAQBAJ`.

Marta Santos Pais. Poverty and exclusion among urban children. 2002.

Tom Rainforth and Frank Wood. Canonical correlation forests. *arXiv preprint arXiv:1507.05444*, 2015.

Elliott D. Sclar, Pietro Garau, and Gabriella Carolini. The 21st century health challenge of slums and cities. *The Lancet*, 365(9462):901–903, Mar 2005. ISSN 0140-6736. doi: 10.1016/S0140-6736(05)71049-7. URL `https://doi.org/10.1016/S0140-6736(05)71049-7`.

United Nations. *State of the World's Cities 2012-2013: Prosperity of Cities*. United Nations Publications, 2012. ISBN 9789211324945. URL `https://books.google.co.uk/books?id=6cGzNAEACAAJ`.

Tianxiang Zhang, Jinya Su, Cunjia Liu, Wen-Hua Chen, Hui Liu, and Guohai Liu. Band selection in sentinel-2 satellite for agriculture applications. *2017 23rd International Conference on Automation and Computing (ICAC)*, pp. 1–6, 2017.