Revealing landslide exposure of informal settlements in Medellín using Deep Learning

Michael Wurm[1*], Raphael Tubbesing[1], Thomas Stark[1], Marlene Kühnl[1,2], Marta Sapena[1], Wolfgang Sulzer[3], Hannes Taubenböck[1,4]

[1]*German Remote Sensing Data Center (DFD), German Aerospace Center (DLR),* Oberpfaffenhofen, Germany
[2]*Company for Remote Sensing and Environmental Research (SLU), Munich, Germany.*
[3]*Institute of Geography and Regional Planning, University of Graz, Graz, Austria.*
[3]*Institute for Geography and Geology, Julius-Maximilians-Universitat Würzburg*, Würzburg, Germany
* E-Mail: michael.wurm@dlr.de

2023 Joint Urban Remote Sensing Event (JURSE) | 978-1-6654-9373-4/23/$31.00 ©2023 IEEE | DOI: 10.1109/JURSE57346.2023.10144128

Abstract— Large areas of informal settlements on the slopes of Medellín are exposed to landslide risk, but there exists no accurate and up-to date data set on the location and size of informal areas. It is thus difficult to develop mitigation strategies to reduce the risk for the local population. Here, we tackle the issue of inaccurate geodata and apply a CNN for the extraction of individual building footprints from orthophotos. With it we achieve a more reliable data base for a more precise estimation of the amount of exposed population in informal areas towards landslides.

Keywords—CNN, building extraction, landslide, slums

I. Introduction

The city of Medellín, Colombia, is situated in a valley surrounded by steep slopes. The fast industrial and economic growth in the middle of the 20th century resulted in a significant influx of migrants from the countryside who settled informally on the margins of the city. In later years, informal housing was further intensified due to conflicts between paramilitary forces and guerilla groups [1].

The uncontrolled, informal growth expanded the city into highly inaccessible areas on the steep slopes. Frequently occurring heavy rainfall events and the bedrock consisting of weak and highly erosive rocks make these informal settlements of low-quality building fabric highly prone to landslide hazards [2]. To counter the prevailing risk by targeted mitigation strategies, it is important to know *where* the areas at risk are and *how many people* would be affected in case of a landslide event. Official population data for Medellín incorporate a precise geolocalization of formal residents, but they lack large parts of more recent areas of informal settlements on the steep slopes [3] which leads to severe underestimation of the exposed population.

Remote sensing imagery of high spatial resolution allow for the detection of areas of the urban poor. Their morphologic appearance of small-scaled, complex, dense urban structures is characteristic for informal settlements [4]. In this paper, we deploy a convolutional neural network (CNN) for the detection of informal buildings to create an updated, more comprehensive building mask compared to official data sources. This improved building mask is then applied to estimate the number of the exposed population to landslides, and compare to official data to fill current spatial knowledge gaps.

II. Study area and Data

A. Medellìn

The study area extends to the very edges of the *urban zone* of the city of Medellín which is formed by 16 *comunas*. In recent years the informal settlements have evolved especially in the high and steep areas of the surrounding mountains.

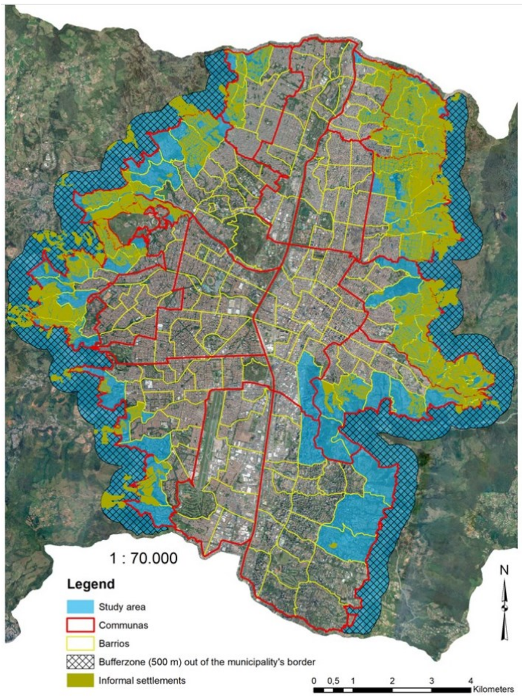


Fig. 1. The study area and the neighborhoods of interest. The map also includes the buffer zone with dense built-up structures beyond the official border with many informal settlements.

Here, we focus specifically on the barrios in the boundary area of the official administrative city area and 500 m beyond to include settlements which are today located outside the boundary as those areas are exposed to the highest landslide risk (Fig. 1).

B. Remote Sensing Data

The image data for the extraction of the informal buildings is a true orthophoto mosaic (TOP) from the year 2019 downloaded from GeoMedellín service (https://www.medellin.gov.co/mapas/rest/services/ServiciosImagen/Ortofoto_Medellin_2019/ImageServer). The images are terrain corrected and have three spectral bands (blue/green/red) and a geometric resolution of 8 cm.

C. Building Cadastral Data and Census data

For training, test and validation of the CNN as well as for generating estimates of exposed population, we use the official building cadaster from the city of Medellín. It contains the spatial footprints for most buildings in Medellin as well as building height. However, the official building cadaster lacks many informal and/or recently constructed buildings. For the population modeling, we use the official census data.

978-1-6654-9373-4/23/$31.00 ©2023 IEEE

D. *Landslide hazard map*

We use the mass movement threat map from the year 2014 (https://geomedellin-mmedellin.opendata.arcgis.com/datasets/gdb-pot-acuerdo48-de-2014). It contains five categories of criticality: "very low threat" (Probability 0 – 0.2%); "low threat" (Probability 0.2 – 3.2%); "medium threat" (Probability 3.2 – 5.5%); "high threat" (Probability 5.5 – 16.5%).

III. Deep Learning for Building Detection

For the detection and classification of individual informal buildings in the orthophotos, we train and apply a CNN. CNNs allow for semantic segmentation for the prediction of pixel-dense outputs in images [5]. These networks are able to learn representational features to an increasing abstract degree from the original input image. A key property, which is a major advantage of deep learning compared to conventional classification algorithms, is the ability of the algorithm to automatically learn meaningful features. Although hand-crafted textural, geometric, and spectral attributes could also lead to accurate classification results, it is hardly possible to achieve an optimal balance between discriminability and robustness of image object details [6].

CNNs consist of multiple layers or levels of representation, based on the raw image data. They are obtained by simple, non-linear functions applied to the pixel values by applying a moving kernel on an image tile. Every single created feature map or layer of representation of these tiles transforms the previous layer into a higher, slightly more abstract level using a general-purpose learning procedure [7]. Semantic segmentation using CNNs has proven to yield high classification accuracies for mapping of informal settlements on the patch level in very high resolution (VHR) imagery [8]. Semantic segmentation of individual buildings in VHR imagery, however, poses further challenges, e.g. varying roof material/types, high built-up densities, and roof pitches can severely complicate this task, especially in structurally complex built-up areas on steep slopes. Various FCNs in related work have successfully achieved the task of building extraction in various urban environments, e.g. Mask-RCNN [9], or U-NetInceptionResnetv2 [10]. Here, we build upon previous work and use an adaption of the popular U-Net architecture as it yielded high accuracies [10]. Originally developed for medical image analysis [11], it was soon applied to other image data. Its architecture is based on the FCN for semantic segmentation [5] and has been adapted to deal with a small amount of training data, which is a typical constraint in biomedical tasks, and also relevant for the task of building extraction (Fig. 2).

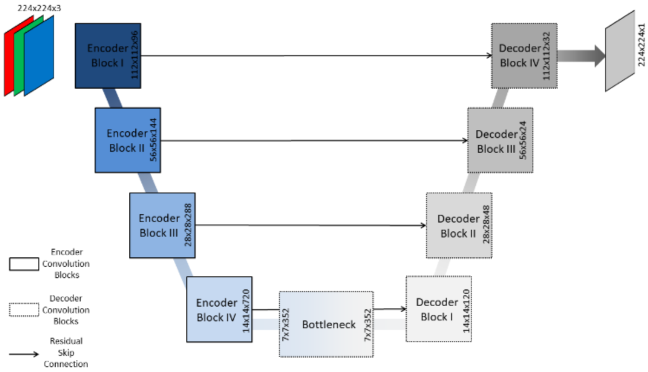


Fig. 2. The U-NetEfficientNetB2 architecture.

The network consists of a contracting (encoder) and an expanding part (decoder), whereas the latter makes the main difference compared to the classical FCN. After reducing the resolution and extracting deep features of the original input image, the FCN has one up-sampling module to transfer the features to the original resolution, performing a pixelwise prediction. In contrast, the U-Net consists of a decoder, which is nearly equally structured as the encoder. Combined with skip connections, features of all four stages in the encoder and therefore, features with the whole spectrum of abstraction and resolution will be added to the decoder in the respective stage. This enables a more effective and detailed localization of the objects in the image. In addition, [11] employed considerable data augmentation to deal with less training samples and improve generalization ability. We modify the traditional U-Net architecture in combination with an EfficientNet variant as a more modern backbone in order to both benefit from the reliability of the established U-Net architecture and to take into account recent advancements in the field of deep learning using segmentation models [12]. EfficientNet [13] has been designed to be more efficient in depth, width, and resolution while being lighter and more precise in its predictive ability.

IV. Work flow

A. *U-NetEfficentNetB2 model training and parameters*

The urban morphology in the Medellín area varies significantly between formally developed structures with mostly high but structured built-up densities in the urban core of the city, and informal, unstructured settlement patterns on the slopes with high built-up densities. A decreasing density gradient with increasing distance from the city center and altitude can be observed [14]. To account for these varying building morphologies, the U-NetEfficientNetB2 needs to be trained with samples from each characteristic urban morphology. Reference data has been collected manually by visual delineation of individual building footprints in the orthophoto. The reference data set was split into training, testing and validation data in the proportion 60/20/20. The orthophoto data was split into regular image tiles of 224 x 224 pixels and an overlap of 1/3 between the tiles to reduce border effects induced by the convolutions of the U-Net architecture.

In order to find the best fitting model parameters, we empirically tested various set-ups regarding the data input. The original geometric resolution of the orthophoto is with 8 cm very high but in 224 x 224 pixels per image tile, this results in a very small field of view of ~18 m and thus lacks spatial context for the building extraction. Therefore, we test for model performance at 16 cm and 32 cm input data resolutions. Initial models have been trained using the same data augmentation, learning rate (0.0001), batch size of 4, class weights (buildings: 2, background 0.5), and a training time of 40 epochs. These parameters are further empirically optimized, after the best performing image resolution was found. As a final step, the trained models are predicted on the test tiles using a majority vote for the parts of the image tiles areas with multiple predictions due to the 1/3 overlap. For the resulting building footprints, a detailed accuracy assessment was performed using the reference data set. The most accurate mask was selected for the subsequent population disaggregation and exposure analysis.

B. Landslide exposure analysis

The main goal of the current study is to quantify the amount of population in informal settlements which is exposed to landslide risk. Therefore, the extracted building footprints are used as spatial proxy information to estimate the number of dwellers per housing unit. Census data are disaggregated onto the building footprints to assess their geographic location and relate it to landslide exposure. Estimation of the exposed population at hazardous sites is performed by the following equation:

$$P_r = \frac{Ba_r}{Ba_t} \times P_t \tag{1}$$

where P_r and Ba_r are the population and the built-up area at risk, respectively, and P_t and Ba_t are the population and the built-up area in total for each barrio, respectively.

The landslide prone areas are represented by the medium and high-hazard areas of the landslide hazard map (see section II-D).

V. RESULTS

A. Building extraction with U-NetEfficientNetB2

In a first suite of experiments, the optimal balance between the geometric resolution and the field of view in terms of model performance for various urban morphologies was tested. The results are summarized in Table 1: these indicate that high image resolutions alone are not the most important factor for building extraction. The field of view for the highest resolution of 8 cm is too small when 224 x 224 image tiles are fed into the model (Fig. 3). In this case almost the entire tile is covered by roofs and thus lacking spatial context. Precision is reported highest for the 8 cm data but with lowest recall. Thus, we use the harmonic mean 8 (F1-score) of both accuracy values. It is reported highest for the 16 cm data resolution and thus we apply this data for further analysis. In a next step, hyperparameters were empirically optimized to find the best fitting combination. It has been found that the change of the class weights to 1.5 (buildings) and 0.5 (background), enhanced data augmentation and a combination of the cross-entropy and the Jaccard loss function could increase the performance of 0.2 as measured by the F1 score. Here, the network is able to better balance between foreground and background during the training process.

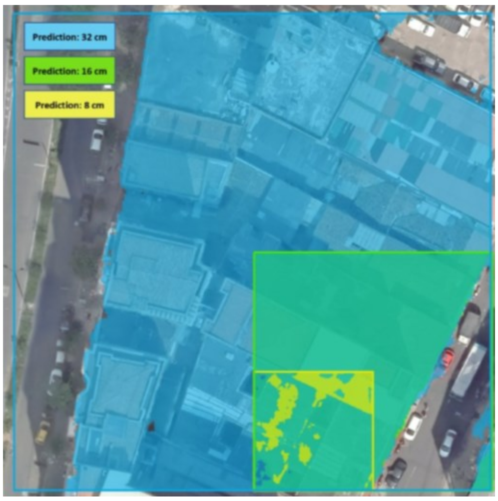


Fig. 3. Prediction effects of the field of view in dependency of the geometric resolution (yellow: 8 cm, green: 16 cm, blue: 32 cm).

TABLE I. PERFORMANCE COMPARISON OF VARYING RESOLUTIONS AND URBAN MORPHOLOGIES FOR THE THREE TESTED IMAGE RESOLUTIONS AND THE TUNED PARAMETERS FOR THE 16 CM DATA. LSSD=LOW SOCIOECONOMIC STATUS DENSE, LSSL: LOW SOCIOECONOMIC STATUS DENSE, C&I=COMMERCE&INDUSTRY, MSS=MEDIUM SOCIOECONOMIC STATUS.

		8 cm	*16 cm*	*32 cm*	*optimized (16 cm)*
LSSD	PRECISION	0.87	0.85	0.81	0.89
	RECALL	0.79	0.96	0.98	0.95
	F1	0.82	0.89	0.89	0.92
LSSI	PRECISION	0.97	0.66	0.49	0.78
	RECALL	0.39	0.95	0.99	0.98
	F1	0.56	0.79	0.66	0.87
C&I	PRECISION	0.92	0.92	0.68	0.95
	RECALL	0.94	0.97	0.96	0.94
	F1	0.88	0.94	0.91	0.95
MSS	PRECISION	0.98	0.95	0.94	0.94
	RECALL	0.67	0.94	0.98	0.90
	F1	0.79	0.95	0.96	0.92
AVERAGE ACC.	PRECISION	0.94	0.85	0.73	**0.89**
	RECALL	0.67	0.96	0.98	**0.94**
	F1	0.76	0.90	0.86	**0.92**

The best performing model was applied on all ~700k image tiles for the exposure analysis (Fig. 4).

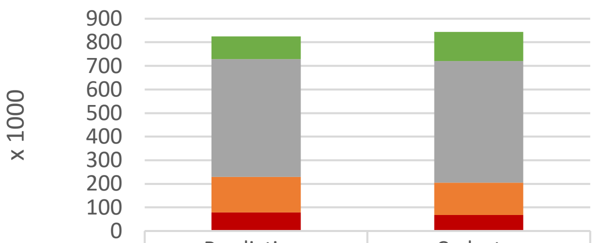
Fig. 4. Examples of prediction results for varying urban morphologies.

B. Exposure analysis

We aim at estimating the potentially affected population in informal settlements by landslide hazards. To do so, we use population counts from the official census data and we model individual population for each of the extracted building footprints. These up-dated population data are subsequently overlaid with the hazard map for a quantitative assessment of exposed population.

Results show a total area of 9.54 km² of informal buildings as an outcome of the CNN compared to 6.70 km² from the official cadaster. This reveals a difference of ~42% in the data. Consequently a severe underestimation of affected population in official data sets is observed. Further, the spatial analysis shows that a majority of the informal areas are exposed to a much higher landslide risk compared to formal settlements: about 35.2% of informal areas are located in areas of high and medium risk, while this accounts only for 8.3% of the formal areas.

Naturally, a higher number of identified informal buildings in the areas of high and medium landslide risk lead also to a higher number of potentially threatened population. While the exposure analysis using official cadaster data claims about 205,000 slum dwellers being exposed, the use of a more complete and up-to-date building data base reveals that a much higher share of the population may be affected by landslide risk. The exposure analysis for high and medium risk shows that about 230,000 slum dwellers are living in endangered areas, which accounts for a difference of about +12%. Fig. 5 depicts the number of affected population in informal settlements for each hazard class and for both data sets: the official building cadaster and the extracted building footprints using the U-NetEfficientNetB2.



	Prediction	Cadaster
very low risk	97	124
Low risk	499,321	515,203
Medium risk	151,183	137,874
High risk	78,066	67,563

Fig. 5. Results of the exposure analysis for the estimates of the affected population by landslides based on the building footprints extracted by the U-NetEfficientNetB2 and the official building cadaster.

VI. Conclusions

The results reveal that official population data is especially inaccurate for the informal settlements and that the percentage of the population living in areas with medium and high landslide risk is considerably higher than in the cadaster-based analysis. Therefore, we conclude by stressing the importance of up-to-date geodata for the estimation of the exposed population. Dwellers in informally settled areas are not only disadvantaged by their poor living conditions, but also by the higher landslide risk due to their location in hazardous zones [15], and finally also by the fact, that official authorities are lacking precise data on their quantities and locations which makes it difficult to provide action plans in case of an event. High resolution imagery and current Deep Learning methods can make important contributions to foster risk reduction.

Acknowledgements

This study was funded by the German Federal Ministry of Education and Research as part of the FONA Client II initiative, grant number 03G0883A-F, as part of the Inform@Risk project.

References

[1] J. Betancur, „Approaches to the regularization of informal settlements: the case of PRIMED in Medellin, Colombia.“, *Glob. Urban Dev. Mag.*, Bd. 3, Nr. 1, S. 1–15, 2007.

[2] J. Claghorn und C. Werthmann, „Shifting ground: Landslide risk mitigation through community-based landscape interventions“, *J. Landsc. Archit.*, Bd. 10, Nr. 1, S. 6–15, Jan. 2015, doi: 10.1080/18626033.2015.1011419.

[3] M. Sapena, M. Kühnl, M. Wurm, J. E. Patino, J. C. Duque, und H. Taubenböck, „Empiric recommendations for population disaggregation under different data scenarios“, *PLOS ONE*, Bd. 17, Nr. 9, S. e0274504, Sep. 2022, doi: 10.1371/journal.pone.0274504.

[4] M. Wurm und H. Taubenböck, „Detecting social groups from space – Assessment of remote sensing-based mapped morphological slums using income data“, *Remote Sens. Lett.*, Bd. 9, Nr. 1, S. 41–50, Jan. 2018, doi: 10.1080/2150704X.2017.1384586.

[5] J. Long, E. Shelhamer, und T. Darrell, „Fully Convolutional Networks for Semantic Segmentation“, *ArXiv14114038 Cs*, März 2015, Zugegriffen: 8. Oktober 2020. [Online]. Verfügbar unter: http://arxiv.org/abs/1411.4038

[6] L. Zhang, L. Zhang, und B. Du, „Deep Learning for Remote Sensing Data: A Technical Tutorial on the State of the Art“, *IEEE Geosci. Remote Sens. Mag.*, Bd. 4, Nr. 2, S. 22–40, Juni 2016, doi: 10.1109/MGRS.2016.2540798.

[7] Y. LeCun, Y. Bengio, und G. Hinton, „Deep learning“, *Nature*, Bd. 521, Nr. 7553, Art. Nr. 7553, Mai 2015, doi: 10.1038/nature14539.

[8] M. Wurm, T. Stark, X. X. Zhu, M. Weigand, und H. Taubenböck, „Semantic segmentation of slums in satellite images using transfer learning on fully convolutional neural networks“, *ISPRS J. Photogramm. Remote Sens.*, Bd. 150, S. 59–69, Apr. 2019, doi: 10.1016/j.isprsjprs.2019.02.006.

[9] D. Stiller, T. Stark, M. Wurm, S. Dech, und H. Taubenbock, „Large-scale building extraction in very high-resolution aerial imagery using Mask R-CNN“, in *2019 Joint Urban Remote Sensing Event (JURSE)*, Vannes, France, Mai 2019, S. 1–4. doi: 10.1109/JURSE.2019.8808977.

[10] M. Wurm, A. Droin, T. Stark, C. Geiß, W. Sulzer, und H. Taubenböck, „Deep Learning-Based Generation of Building Stock Data from Remote Sensing for Urban Heat Demand Modeling“, *ISPRS Int. J. Geo-Inf.*, Bd. 10, Nr. 1, S. 23, Jan. 2021, doi: 10.3390/ijgi10010023.

[11] O. Ronneberger, P. Fischer, und T. Brox, „U-Net: Convolutional Networks for Biomedical Image Segmentation“, *ArXiv150504597 Cs*, Mai 2015, Zugegriffen: 8. Oktober 2020. [Online]. Verfügbar unter: http://arxiv.org/abs/1505.04597

[12] P. Iakubovskii, „Segmentation Models“, *GitHub Repos.*, 2019, [Online]. Verfügbar unter: https://github.com/qubvel/segmentation_models

[13] M. Tan und Q. V. Le, „EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks“, 2019, doi: 10.48550/ARXIV.1905.11946.

[14] M. Kühnl, M. Sapena, und H. Taubenbock, „Categorizing Urban Structural Types using an Object-Based Local Climate Zone Classification Scheme in Medellín, Colombia“, gehalten auf der REAL CORP 2021, Vienna, 2021.

[15] I. Müller, H. Taubenböck, M. Kuffer, und M. Wurm, „Misperceptions of Predominant Slum Locations? Spatial Analysis of Slum Locations in Terms of Topography Based on Earth Observation Data“, *Remote Sens.*, Bd. 12, Nr. 15, S. 2474, Aug. 2020, doi: 10.3390/rs12152474.