Mapping Human Settlements with Multi-seasonal Sentinel-2 Imagery and Attention-based ResNeXt

Chunping Qiu[1], Michael Schmitt[1], Hannes Taubenböck[2], Xiao Xiang Zhu[1,2]
[1]Signal Processing in Earth Observation, Technical University of Munich (TUM)
Arcisstr. 21, 80333 Munich, Germany
[2]Earth Observation Center (EOC), German Aerospace Center (DLR)
Münchener Str. 20, 82234 Wessling, Germany

***Abstract*—This paper explores the potential of multi-spectral Sentinel-2 imagery for human settlement mapping, using deep learning based methods. We show first results of a study area in central Europe, with an attention-based ResNeXt to better exploit the spectral information. Reasonable mapping accuracy has been achieved, compared to the state-of-the-art products. Based on the results and comparison with the existing products, we discuss two interesting questions: how can human settlement mapping be made consistent with or complementary to the existing human settlement maps and how can further improvement in human settlement mapping be achieved by exploring deep learning-based approaches?**

***Index Terms*—Sentinel-2, classification, attention, convolutional neural network (CNN), human settlement (HS) mapping**

I. Introduction

Mapping human settlements across the world is of great importance, since the availability of accurate, reliable and up-to-date human settlement maps is essential to a large number of issues including housing and sustainable urban development, poverty reduction, climate change, biodiversity conservation, ecosystem services provision, as well as disaster management [1]. This is especially true in the era of rapid urbanization. Currently, there are only few global products in this regard available, for example the Global Urban Footprint (GUF) [2] produced from TerraSAR-X and TanDEM-X Synthetic Aperture Radar (SAR) data and the Global Human Settlement Layer (GHSL) produced with global, multi-temporal archives of fine-scale satellite imagery and other auxiliary data [3]. While all the products show operational application value in assessing and monitoring human presence [4], improvement is still needed regarding aspects such as mapping accuracy, update frequency, production costs etc. For example, high resolution satellite images such as those from SPOT 5 and WorldView used in [5] are not freely available on a global scale. In addition, in areas such as Africa, where not much reference data is available, the accuracy assessment of each product tends to be difficult, which makes cross comparison among different products necessary. More importantly, these products are not mapping exactly the same targets, even though they are all related to human settlements or built-up areas. This is due to the differing definitions of each product regarding the classes “urban”, “human settlement” or “built-up”. For example, GUF represents the built-up areas marked by the presence of vertical structures, while GHSL represents the built-up areas marked by the presence of buildings. Another reason might relate to the employed data source (optical or SAR satellite images), which carries different kinds of information with specific potentials and limitations [5]–[7].

The current situation motivates us to explore more possibilities on human settlement mapping from space, taking into account the powerful feature learning ability offered by deep learning based methods. Furthermore, we want to eventually explore the potential to monitor settlement areas based on data with frequent updating capabilities. For this, we propose to first focus on the globally available multi-spectral images provided by the Sentinel-2 mission [8]. Back in 2016, [9] already assessed its value for detecting built-up areas, showing its potential regarding the mapping of thematic contents compared to Landsat and Sentinel-1 images. In addition, we recently investigated the use of Sentinel-2 imagery with respect to Local Climate Zone (LCZ) mapping at large scale [10], using deep learning based methods. For this non-trivial task, mapping results with state-of-the-art accuracy have been achieved, exploiting both the multi-spectral Sentinel-2 imagery and the powerful feature learning ability of a Residual Convolutional Neural Network based architecture [11]. Adding to these first results, this work is meant to give insights into the potential of Sentinel-2 imagery for the mapping of human settlements using deep learning based methods, especially in the case where no ground truth data is available for the target area, thus domain adaptation is needed. To this end, we investigated the attention-based ResNeXt architecture, followed by discussions on the potential, limitations and possible solutions.

II. A case study in Central Europe

A. Convolutional block attention module-based ResNeXt for human settlement mapping

For deep learning-based HS mapping, there are two options. The first is to train a semantic segmentation neural network,

This work is jointly supported by the China Scholarship Council (CSC), the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation program (Grant Agreement No. ERC-2016-StG-714087, Acronym: So2Sat), the Helmholtz Association under the framework of the Young Investigators Group SiPEO (VH-NG-1018, www.sipeo.bgu.tum.de) and the Bavarian Academy of Sciences and Humanities in the framework of Junges Kolleg.

978-1-7281-0009-8/19/$31.00 ©2019 IEEE

using the ground truth data either in the region of interest or in a different area, and segment the whole image using the trained network. While straight-forward, this kind of HS mapping approach suffers from the unavailability of the ground truth data. The alternative is to train a patch-wise classification network, followed by the pixel labeling via a sliding window on the whole image, with the aiming ground spacing distance as the step of the sliding window. This way, the patch-wise ground truth is easier to obtain, for instance, through a propagation of urban land cover classes. However, the ground spacing distance of the resulting HS maps will be affected by the size of the input image patch and the step of the sliding window. This work will employ the second approach for HS mapping, and investigate two different sizes of the input image patch.

For the presented study, we have adapted the Convolutional block attention module (CBAM)-based ResNeXt for HS mapping, as it has been shown powerful through extensive evaluations [11], [12]. The input to the network is an image patch, and the output is a label indicating the class of the input image patch. The adapted network, hereafter referred to as *ResNeXt(8)* and *ResNeXt(32)* for the input size of 8 and 32, respectively. No big changes are made to the ResNeXt-50, except a CBAM is added for each residual block, in order to explore the potential of the attention-based neural network architecture for HS mapping.

B. Experimental data and setup

Our study areas are spread over seven cities located in the heart of Europe: Amsterdam, Berlin, Cologne, London, Milan, Munich and Paris. For each city, four cloud free Sentinel-2 images were downloaded from Google Earth Engine (GEE): one for each season from winter 2016/2017 to autumn 2017. Only 10 bands of Sentinel-2 imagery are used in this study: B2 (Blue), B3 (Green), B4 (Red) and B8 (Near-infrared) with 10 m Ground Sampling Distance (GSD) and B5 (Red Edge 1), B6 (Red Edge 2), B7 (Red Edge 3), B8a (Red Edge 4), B11 (Short-wavelength infrared 1) and B12 (Short-wavelength infrared 2) with 20 m GSD. The 20 m bands are up-sampled to 10 m GSD.

The ground truth labels available for selected neighborhoods in the seven cities are taken from the So2Sat LCZ42 dataset [13], which was hand-labeled for LCZ mapping. The 17 LCZs are: Compact high-rise, Compact mid-rise, Compact low-rise, Open high-rise, Open mid-rise, Open low-rise, Lightweight low-rise, Large low-rise, Sparsely built, Heavy industry, Dense trees, Scattered trees, Bush or scrub, Low plants, Bare rock or paved, Bare soil or sand and water, respectively. Since the focus of this study was on the detection of human settlements with a general notion rather than assigning different classes to different neighborhoods, we combined the LCZ classes 1 to 8 and 10 to a *human settlement* (*HS*) target class, because they all describe built-up areas. The LCZ A, B, C, D, F, G are considered as the background class. LCZ class 9 (*Sparsely built*) and LCZ E (*Bare rock or paved*) were not considered for the HS class or the background class, because they contain mostly natural surroundings, while the buildings and the streets (roads) parts are already contained in the *HS* class. In this way, we generate a binary classification scheme that distinguishes between human settlements and natural surroundings. For each ground truth pixel, a image patch was cutted around the corresponding position and the patch size is 8×8 and 32×32, as input for *ResNeXt(8)* to *ResNeXt(32)*, respectively. Because of the 100 meter GSD of the ground truth dataset, the 32×32 patches have some overlap.

In order to fully validate the proposed approach for a first proof-of-concept study, we designed a cross validation experimental setup, in which the city of Munich was left for testing while the other six cities were used for training. In this way, we can gain insights into the potential of this framework for large-scale human settlement mapping, where ground truth data is not sufficiently available.

After training the classifier, the human settlement map is produced by applying a sliding window on the Sentinel-2 image with a stride of one pixel, i.e., 10 meters. To have a completely independent evaluation, we used the building layer from OpenStreetMap (OSM) [14] [1] transferred to geotiff format with 10 meter GSD as ground truth.

C. Experimental results

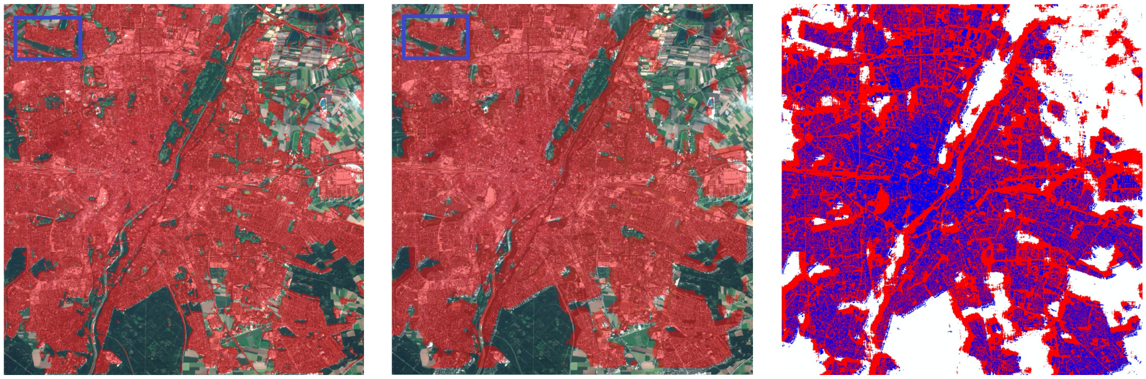
Table I shows the classification accuracy of the two approaches, evaluated against the OSM building layer, as well as the manually labeled ground truth. In order to avoid human-induced bias, an equally distributed grid is generated for the test city, in the city center area, with 2000 meter distance between each two points. These manually labeled grid-based checking points (MLGCPs), with a size of $20m \times 20m$, are manually classified into HS or non-HS, allowing for a meaningful spatial assessment of the HS mapping results. Furthermore, the state-of-the-art products, GUF, Global HBASE, and the GHS built-up grid, were chosen for comparison, for better evaluation of the investigated HS mapping approach.

The comparative classification results from *ResNeXt(8)* and *ResNeXt(32)* can be seen in Fig. 1, where the corresponding OSM building layer is also shown, as a reference. Figure 2 shows a closer view of the subset in the upper-left corner of the study area in Fig. 1

TABLE I: Accuracy assessment of the HS mapping results. Overall accuracy (OA), Kappa, commission error (cme), omission error (ome) are with respect to the MLGCPs, and the recall is with respect to the OSM building layer.

	MLGCPs-based				OSM-based
	OA	Kappa	cme	ome	recall
ResNeXt(8)	0.91	**0.79**	0.25	**0.02**	**0.99**
ResNeXt(32)	0.88	0.71	0.29	0.09	0.96
GHSL	0.88	0.69	0.21	0.24	0.84
GUF	**0.92**	0.78	**0.08**	0.24	0.87
HBASE	0.84	0.64	0.33	0.15	0.89

[1] The employed OSM data copyrighted OpenStreetMap contributors and available from https://www.openstreetmap.org.

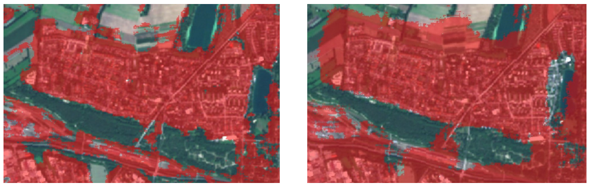


(a) The result of *ResNeXt(8)*.

(b) The result of *ResNeXt(32)*.

(c) OSM overlaid on (a).

Fig. 1: The mapping results overlaid on the Sentinel-2 images, or overlaid with the OSM building layer, in the city center area of Munich, Germany.



(a) The result of *ResNeXt(8)*.

(b) The result of *ResNeXt(32)*.

Fig. 2: A closer view of the subset indicated by the blue polygon, in Fig. 1.

III. Discussion

A. Evaluation of the Classification Results

Both the classification accuracy and the mapping results in Section II show promising performance of the proposed framework for human settlement mapping. Comparable accuracy to the state-of-the-art products is achieved. The major part of the human settlement in the city of Munich is successfully mapped based on the visual comparison to the OSM building layer in Fig. 1. The higher commission error in Fig. 1 (c) can be possibly due to the bigger pixel of the HS mapping result, the different definition of the HS than the building layer in OSM, as well as the potentially inaccurate and unreliable of the OSM [15].

From Tab. I, it can be seen that a small input patch size (*ResNeXt(8)*) is better based on both of the evaluation methods, with higher recall for class *HS*. Besides, *ResNeXt(8)* provides lower omission error and commission error than *ResNeXt(32)*. This can also be proved by Fig. 2, where the result of *ResNeXt(32)* shows a clear false positive on the boundary of the building. This is partly due to the larger patch size of the input for *ResNeXt(32)*, thus the classification result of a pixel is possibly negatively affected by its neighborhood within 320 meter.

However, both approaches in this paper provide rather high commission error, which means many non-urban areas are classified into the *HS* class. While this might result from the limited generalization ability of the classifier trained with samples from six cities far from the test city Munich, it might also come from LCZ-derived *HS* class being non-optimally defined. A possible solution is to explore the semantic segmentation networks such as fully convolutional networks.

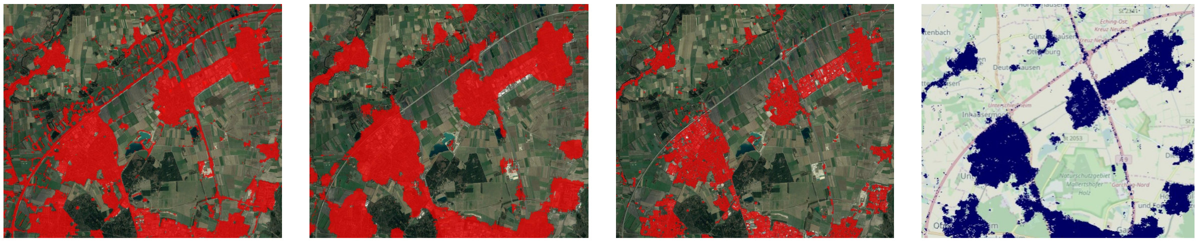
B. Difference between actual human settlement and the HS class

When overlying reference building footprints on the mapping result, as shown in Fig. 1(c), it becomes clear that the *HS* class is not only representing buildings. Instead, it contains also roads and some other falsely classified areas. This is partly due to the fact that the labeled *HS* class still contains streets in the building blocks, even though we excluded LCZ class E while preparing the reference samples. Besides, it is related to the spatial resolution of the employed Sentinel-2 images. The misclassification can be partly solved by suitable post-processing, while the challenge of distinguishing buildings from roads is non-trivial, when using the 10 meter GSD satellite images.

Similar misclassfication also exists in the existing products, as shown in Fig. 3. There are some ways to the exclude roads (streets) out of the human settlement mapping result, which is necessary at regional or even large-scale, even though we do not have standard definitions for “city”, “urban”, “built-up” or “human settlement” yet. The first is considering large and small neighborhood at the same time, as many parts of road is not classified as class *HS* when using *ResNeXt(32)* (Fig. 3(b)). Another solution can be fusing SAR and optical satellite images, since both GUF and the Built-up result from Sentinel-1 contain fewer road area (Fig. 3(c) and 3(d)). However in the next section, it shows that problems still exist even if the mapping results only contain buildings.

C. From multi-level classification to human settlement mapping

With human settlement mapping as an objective, it is not enough to only map buildings, since humans do not live in all kinds of buildings. In the LCZ scheme, there are 10 kinds of

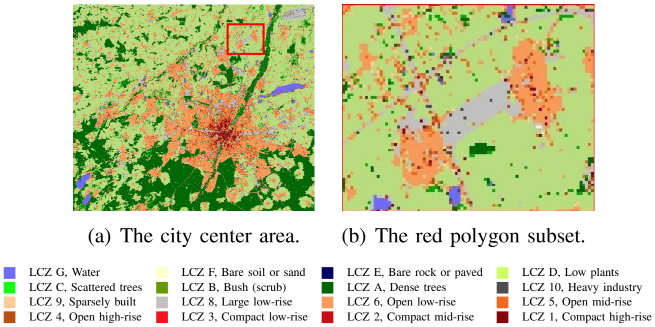


(a) The result of *ResNeXt(8)*. (b) The result of *ResNeXt(32)*. (c) GUF (d) Built-up Sentinel-1

Fig. 3: Comparison of the resulting maps (a, b), GUF (c) and GHSL (d), in a suburban area of Munich. GHSL Source: https://ghsl.jrc.ec.europa.eu/visualisation.php#. The satellite image data: Google, Image Landsat / Copernicus.

building areas, with difference in height, density and land use. Figure 4 shows the produced LCZ map of Munich, Germany. From the subset, we know that it contains at least two kinds of building areas: *Large-low rise* and *Open low-rise*. Therefore, classifying all buildings into one "human settlement" class is probably not enough in order to develop a global, people-based definition of cities and settlements.

From this example, it seems that a multi-level classification, i.e., from LCZ classification or other land cover classification to human settlement mapping, is a solution, which will be an important future research direction.



(a) The city center area. (b) The red polygon subset.

Fig. 4: The LCZ map of the city Munich, Germany.

IV. Summary and Outlook

This paper shows the potential of multi-spectral Sentinel-2 imagery for human settlement mapping, under the case where no ground truth data is used in the target area, using a attention-based ResNeXt. The preliminary results are promising and yet introduce some important challenges to be solved. First, the definition of human settlement should be well defined considering both the existing products and the characteristic of the employed data. As a first step towards automatic and efficient human settlement mapping, this work motivates us to study in directions such as the deep neural network design and multi-source data fusion in future work.

References

[1] UN General Assembly, "Transforming our world: the 2030 agenda for sustainable development," *New York: United Nations*, , no. 1, 2015.

[2] T. Esch, M. Marconcini, A. Felbier, A. Roth, W. Heldens, M. Huber, M. Schwinger, H. Taubenböck, A. Müller, and S. Dech, "Urban footprint processorfully automated processing chain generating settlement masks from global data of the tandem-x mission," *IEEE Geosci Remote Sens Lett*, vol. 10, no. 6, pp. 1617–1621, 2013.

[3] M. Pesaresi, D. Ehrlich, S. Ferri, A. Florczyk, S. Freire, M. Halkia, A. Julea, T. Kemper, P. Soille, and V. Syrris, "Operating procedure for the production of the global human settlement layer from landsat data of the epochs 1975, 1990, 2000, and 2014," *Publications Office of the European Union*, 2016.

[4] M. Melchiorri, A. J Florczyk, S. Freire, M. Schiavina, M. Pesaresi, and T. Kemper, "Unveiling 25 years of planetary urbanization with remote sensing: Perspectives from the global human settlement layer," *Remote Sens.*, vol. 10, no. 5, pp. 768, 2018.

[5] M. Pesaresi, H. Guo, X. Blaes, D. Ehrlich, S. Ferri, L. Gueguen, M. Halkia, M. Kauffmann, T. Kemper, L. Lu, et al., "A global human settlement layer from optical hr/vhr rs data: concept and first results," *IEEE J Sel Top Appl Earth Obs Remote Sens*, vol. 6, no. 5, pp. 2102–2131, 2013.

[6] B. Bechtel, M Pesaresi, L See, G Mills, J Ching, PJ Alexander, JJ Feddema, AJ Florczyk, and I Stewart, "Towards consistent mapping of urban structures-global human settlement layer and local climate zones," in *XXIII ISPRS Congress*, 2016, pp. 1371–1378.

[7] T. Esch, F. Bachofer, W. Heldens, A. Hirner, M. Marconcini, D. Palacios-Lopez, A. Roth, S. Üreyen, J. Zeidler, S. Dech, et al., "Where we livea summary of the achievements and planned evolution of the global urban footprint," *Remote Sens.*, vol. 10, no. 6, pp. 895, 2018.

[8] M Drusch, U Del Bello, S Carlier, O Colin, V Fernandez, F Gascon, B Hoersch, C Isola, P Laberinti, P Martimort, et al., "Sentinel-2: Esa's optical high-resolution mission for gmes operational services," *Remote Sens. Environ.*, vol. 120, pp. 25–36, 2012.

[9] M. Pesaresi, C. Corbane, A. Julea, A. J Florczyk, V. Syrris, and P. Soille, "Assessment of the added-value of sentinel-2 for detecting built-up areas," *Remote Sens.*, vol. 8, no. 4, pp. 299, 2016.

[10] C. Qiu, M. Schmitt, L. Mou, P. Ghamisi, and X. Zhu, "Feature importance analysis for local climate zone classification using a residual convolutional neural network with multi-source datasets," *Remote Sens.*, vol. 10, no. 10, pp. 1572, 2018.

[11] Saining Xie, Ross Girshick, Piotr Dollár, Zhuowen Tu, and Kaiming He, "Aggregated residual transformations for deep neural networks," in *CVPR*, 2017, pp. 1492–1500.

[12] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon, "Cbam: Convolutional block attention module," in *ECCV*, 2018, pp. 3–19.

[13] X. Zhu and et al., "So2sat lcz42: A benchmark dataset for local climate zones classification," 2018.

[14] OpenStreetMap contributors, "Planet dump retrieved from https://planet.osm.org ," https://www.openstreetmap.org, 2017.

[15] Cláudia M Viana, Luis Encalada, and Jorge Rocha, "The value of openstreetmap historical contributions as a source of sampling data for multi-temporal land use/cover maps," *IJGI*, vol. 8, no. 3, pp. 116, 2019.