

Mapping Urban Villages Using Fully Convolutional Neural Networks

J. Mast^a, C. Wei^b, M. Wurm^c

^a Department of Remote Sensing, Department of Geography and Geology, University of Würzburg, Würzburg, Germany

^b School of Geography and Planning, Sun Yat-Sen University, Guangzhou, Guangdong, China

^c German Remote Sensing Data Center (DFD), Deutsches Zentrum für Luft- und Raumfahrt (DLR), Oberpfaffenhofen, Wessling, Germany

Johannes Mast; Johannes.mast@stud-mail.uni-wuerzburg.de; Department of Remote Sensing, Julius-Maximilians-Universität, 97074 Würzburg

“This is an Accepted Manuscript of an article published by Taylor & Francis Group in Remote Sensing Letters 27/05/2020, available online:

<http://www.tandfonline.com/10.1080/2150704X.2020.1746857>

Please cite the article as follows:

Mast, J., Wei, C., Wurm, M. (2020): *Mapping Urban Villages using fully convolutional neural networks*. *Remote Sensing Letters* 11(7), pp. 630-639. Taylor & Francis
<https://doi.org/10.1080/2150704X.2020.1746857>

Mapping Urban Villages Using Fully Convolutional Neural Networks

Urban villages are a characteristic settlement type characterized by preserving their morphological characteristics embedded in sharp contrast in modern, high-rise developments found especially in fast growing urban agglomerations of China. They serve very important socioeconomic functions in terms of the provision of cheap housing for rural-urban migrants, but they are also considered controversial for local governments. Due to the unprecedented pace of urban growth, especially in the Pearl River Delta region (PRD), up-to-date information on the size and location of urban villages are mostly missing. Large-area but highly detailed data from earth observation platforms can provide crucial information for mapping urban villages based on their characteristic morphologies. This study deploys fully convolutional neural networks for mapping urban villages in the city of Shenzhen. Results of the underlying experiments show that very high mapping accuracies of 84% can be achieved.

1. Introduction

Spatial urbanization processes transform our world and our societies significantly. This is especially the case in China, where the interference of the spatially expanding cities with the registration system has given rise to a peculiar phenomenon. Scattered throughout the sprawling cityscapes of the Pearl River Delta (PRD), we find pockets of seemingly chaotic and highly dense clusters of buildings that stand morphologically in stark contrast to the structured patterns, broad streets, and modern skyscrapers of the formally developed urban areas surrounding them (Chung 2010). Deep inside this concrete forest, the streets between face-to-face buildings are full of life, packed with shops, grocery stores, and service outlets. Due to the extremely high densities which are estimated to reach 90% (Chung 2010, Taubenböck, Kraff, and Wurm 2018), public spaces are scarce, and infrastructure is insufficient (Liu et al. 2010). These spaces are urban in all but name, and so are the people who inhabit them.

The origin of this phenomenon lies in the Chinese dual registration system, to which the concept of hukou is central. Under this system, most land and its people are registered

as either 'rural' or 'urban'. Changing one's hukou is very difficult due to strong restrictions on registration (Zhang 2003). In a static world, this system may function well, but Chinese cities are anything but static; they sprawl at extremely high pace (Liu et al. 2010) and through this fast urban growth, they envelop formerly existing villages which lie in their surrounding area. These villages, surrounded by planned urban areas, often retain their rural designation and irregular morphologic structure (Lin, de Meulder, and Wang 2011). The villages' inhabitants turn from agriculture towards a new source of income: large quantities of unauthorized extra dwellings (Zhang et al. 2003) are constructed and rented to new rural migrants who are excluded from formal urban housing by economic and institutional barriers (Hao 2012).

The villages that are transformed in this manner are known as "chengzhongcun" (literally "villages encircled by the city") or "urban villages" (UV) (Zhang et al. 2003). While they share some morphological similarities with informal settlements / slums, their physical structure, characterized by moderately high residential buildings of square floor plan, with similar size and style, tends to be different enough to warrant a distinction between these two types of non-formal urban structures. Unlike most forms of informal settlements, the existence of UVs is the result of the strict adherence to China's land use system. Considering their socioeconomic function, the best describing concept is that of the "arrival city" (Saunders 2012; Taubenböck, Kraff, and Wurm 2018), a space that provides cheap housing for the poor and migrants as well as opportunities to become part of the urban society.

The Chinese urban government usually sees UVs in a negative light. UVs are criticized for their high crime rate, poor living conditions, and fire hazard due to the informally developed buildings. On the other hand, the city benefits from the supply of cheap housing and flexible and adaptive services the UVs can provide due to their

freedom from control and restrictions (Zhang et al. 2003), thereby enabling the stream of migrant workers which power the city's economy and – in providing social stability – playing a positive role in the rapid urban development (see Liu et al. (2010) for a detailed discussion). The UVs in turn benefit from the urban infrastructure surrounding them.



Figure 1. An urban village surrounded by high-rise buildings (a); the main street of an urban village (b); low-quality infrastructure (c); high-density built environment (d) (© 2012 Pu Hao, permission acquired).

The relationship between the cities and UVs is, thus, complex, and efforts by the urban governments to redevelop or upgrade UVs are consequently often controversial. This is, in part, because the spatial patterns and drivers of UVs development and metamorphosis are not sufficiently well understood. Hao et al. (2009) find that little is

known about the success of formal development and the patterns of non-conforming land use. Scientific interest focuses on the largest cities such as Shenzhen (Wang, Wang, and Wu 2009; Hao 2012), Shanghai (Wu et al. 2013), Wuhan (Huang et al 2015), and Guangzhou (Lin, de Meulder, and Wang 2011) but neglects the secondary cities across the country whose quick growth likely leads to the development and change of UVs, the monitoring of which would prove to be of great interest. That crucial insights can be gained from analysing UVs in their geographical context is demonstrated by Song and Zenou (2012), who explore the public perception of UVs via their impact on housing prices. Hao et al. (2009) further show that it is possible to gain understanding of the development and distributions of UVs using satellite imagery and GIS-based analyses. For such analyses to be relevant, they must be performed on data which is

- available on a large scale – because UVs occur throughout China
- consistent – to allow for cross-region comparisons
- up to date – to provide decision makers with up-to-date information on a

highly dynamic process.

Such up-to-date maps are often incomplete or unavailable and as the mega-cities of China cover very large areas, timely and complete detection of UVs just using the traditional method of fieldwork is impossible (Li, Huang, and Liu 2017). Recent developments in the application of machine learning to VHR optical remote sensing imagery offers a modern, automated approach and, thus, an alternative (Shekhar 2012, Zhu et al. 2019).

The application of machine learning approaches in the context of mapping specific urban morphologies has been demonstrated in former studies on a similar morphologic urban category, namely slums / informal settlements (e.g. Wurm et al.

2018), and, more recently, also deep learning methods such as fully convolutional neural networks (FCN), popularized by Shelhamer, Long, and Darrell (2016) have been deployed (Wurm et al. 2019). The specific case of mapping of UVs has been presented by Huang, Liu, and Zhang (2015), who used VHR satellite imagery to identify UV patches in WorldView-2 imagery via large numbers of both low- and high-level handcrafted features calculated on image scenes. This scene-based approach was further explored via the application of unsupervised deep neural network-based learned feature representations (Li, Huang, and Liu 2017) or landscape metrics (Liu et al. 2017). These scene-based methods of mapping UVs, while accurate, result in a relatively coarse resolution of around 120m, the size of one scene, making them very useful for the identification of UVs. However, no precise mapping in terms of semantic segmentation is performed. Higher resolutions can be reached through the fusion of optical and radar imagery, as shown by Wei, Blaschke, and Taubenböck (2016), albeit at lower accuracy. In this project, we explore the applicability of FCNs to the mapping of UVs in the city of Shenzhen.

2. Materials and Methods

2.1. Study Area

Since the central government chose it as the site of China's first special economic zone (SEZ) in 1980, Shenzhen has become one of China's biggest and most important cities (UN 2018). We selected the western part of Shenzhen as the area of interest (AOI) for this study because of the availability of a reference map on UV for 2009 (referenced in Hao 2012), the presumed presence of many UVs in different stages of development (see Figure 2 for examples), and a large body of research and scientific interest in the socioeconomic and political background of the UVs. We sub-divided the AOI into 4

quadrants (Figure 2), of which the densely developed coastal quadrants 1, 2 and 4 were used as training region and the inland quadrant 3 as testing region.



Figure 2. Overview over the study area. 1-4: sub-regions. (a)-(d): Examples of urban villages: (a) Dense village with some residential buildings, at the outskirts of the city. (b) Area which is currently being developed, bordered by two UVs (c) Dense village in the heart of the SEZ. (d) Village containing a large fraction of low-rise buildings

2.2. Inputs

For the task of mapping informal settlements, both the studies of Huang et al. (2015) and Wurm et al. (2019) found that very high resolution (VHR) imagery produced the best results. As a variety of VHR sensors exist, we synthesized imagery by sampling Google Earth imagery at 2 m resolution – a resolution which is representative of what is reached and surpassed by most commonly available VHR satellites (see Wang et al. (2019) for a comparable procedure). On this basis, we created a reference map of the

previously described AOI following a standard image analysis procedure developed for informal settlements (e.g.: Wurm & Taubenböck 2018; Friesen et al. 2018). Using a random forest classifier trained on visually derived objects, we first created a map of the land cover classes Water, Vegetation, and Artificial land (including all formal and bare surfaces which do not correspond to UVs). Further, we visually mapped UVs by delineating their boundaries in high resolution imagery. To facilitate the mapping procedure, we incorporated maps created by the Urban Planning and Design Institute of Shenzhen and the School of Urban Design of Wuhan University in 2005 (Hao et al 2009) as well as a map created by the Shenzhen Municipal Building Survey in 2009 (Hao 2012) for reference data generation. The final reference map covers an area of 905.4 km² and includes 333 patches of UVs. The thematic accuracy of the reference map including four land-use/land-cover classes has been independently evaluated by visual inspection of a non-stratified random sample of 200 test points over the entire area reporting an overall accuracy of the reference map at 95%.

2.3. Preprocessing

While the FCN architecture is capable of whole-image predictions they are not feasible for remote-sensing images due to their comparatively large size. For this reason, we split both the satellite image and the land cover map into 21,426 overlapping tiles of 128 by 128 pixels each, corresponding to 256 by 256 m each, which are considered sufficient to capture the spatial extent of UVs based on the findings of Friesen et al. (2018), who report a mean patch size of informal settlements of 1.6 ha. Due to class imbalance in the data set, we additionally performed data augmentation on tiles with a significant fraction of UV to artificially increase the amount of UV image tiles (Stiller et al. 2019). Our experimental set-up of the mapping procedure consists the tiles of the three training regions which are randomly split into a training (70%) and a validation

(30%) dataset to perform independent performance evaluation (see Figure 3 for an illustration of the workflow).

2.4. Modeling

The basis of the model is the VGG19 network created by the Visual Geometry Group of Oxford University (Simonyan and Zisserman 2014). It was adapted into a fully convolutional network (FCN-VGG19) by discarding the fully connected layers and adding a series of deconvolutional layers which upsample the prediction (Shelhamer, Long, and Darrell 2016).

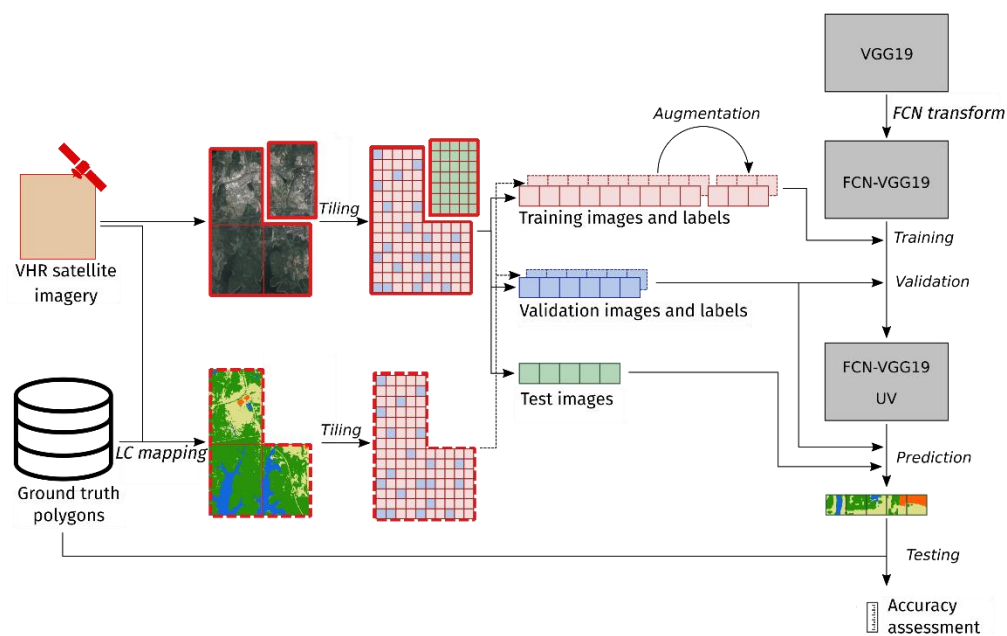


Figure 3. Overview over the data processing and modeling workflow.

We adapted the weights of the VGG19 network and trained the FCN-VGG19 network on the previously created dataset for 8 epochs equaling 146,296 iterations, when the training and validation error converged.

2.5. Accuracy Assessment

Accuracy assessment was performed both on the validation tiles – which were randomly selected and held out tiles in the training region – as well as all the tiles of the testing region. This twofold analysis thus measures the capability of the model both to detect UVs in familiar surroundings, as well as its ability to transfer to different environments. We calculated the overall accuracy for all land cover classes to assess the general performance of the algorithm. To evaluate performance on the UV class we merged all non-UV classes into a background class and calculated the Cohen's kappa coefficient (Kappa) as well as the Intersection over Union (IoU) metric considering just the UV and background classes.

3. Results

Figure 4 displays selected areas of the predicted maps. Fig 4a and 4b are selected from sub-regions 4 and 1 (compare Figure 2) and represent the training region. Figure 4c is selected from the test region 3 and shows along its northern boundary a dense village that was not recognized as well as a patch of bare soil that was falsely classified as UV, an example of a falsely positive prediction.

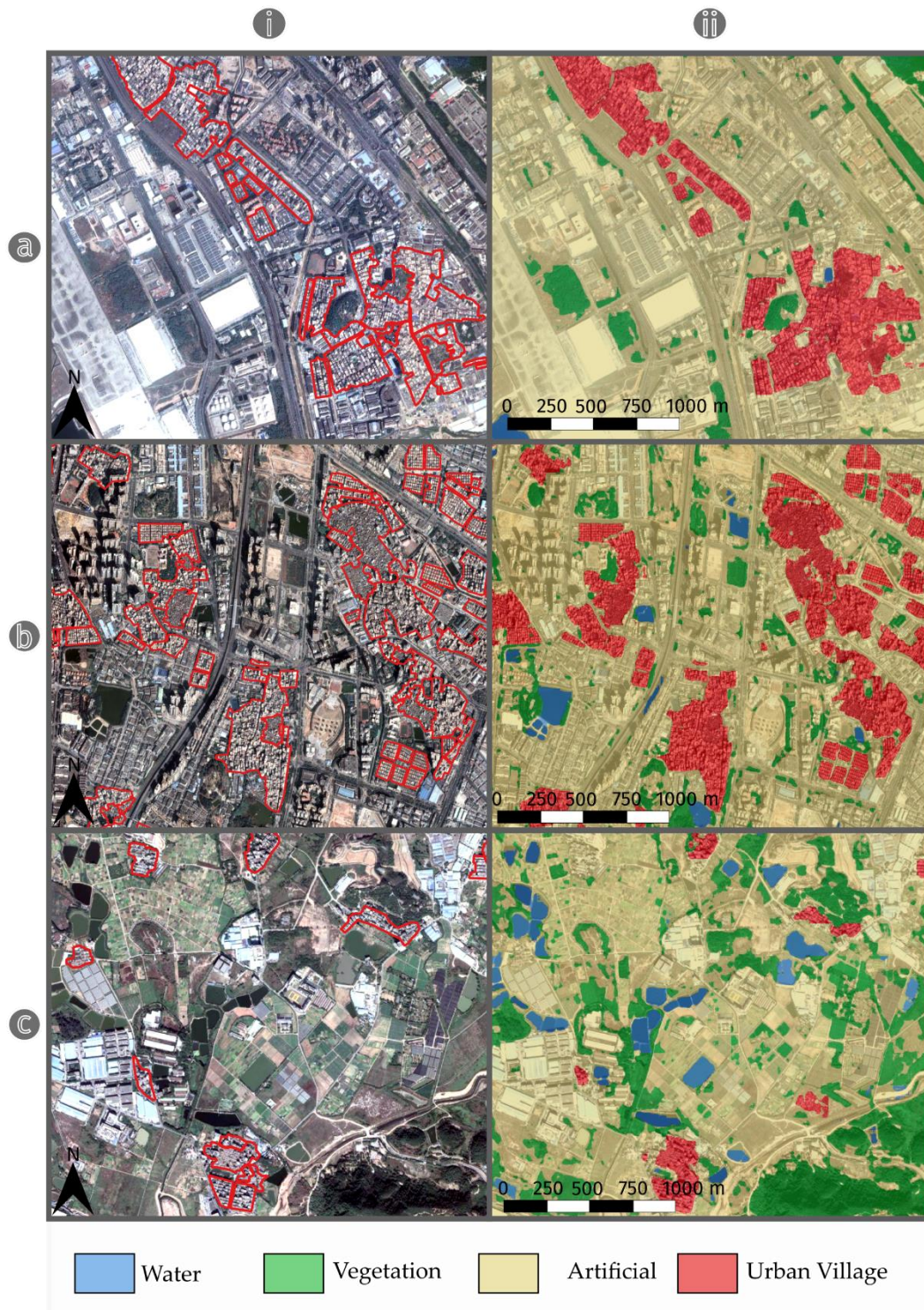


Figure 4. Results of predicted land cover of the model. (i): Imagery from the study area. (ii): corresponding predicted land cover of the FCN-VGG19-UV model. (a) and (b) represent the training region. (c) represents the testing region.

Table 1 displays the overall accuracy metrics as well as the Intersection over Union (IoU) specific to the UV class. As the mapping of the other three land cover classes is

not the primary focus of this study, further class specific accuracies are not displayed.

All accuracies are substantially higher for the validation dataset.

Table 1. Accuracy metrics for the validation and test datasets.

Data Set	Overall	Kappa coefficient	UV ¹ IoU ²
Validation	0.950	0.888	0.842
Test	0.917	0.803	0.680

¹ UV = Urban Village Class ²IoU = Intersection over Union

Table 2 displays the land cover fractions of the mapped area, split by the land area belonging to the SEZ and the land outside of it. Within the SEZ the fraction of UVs is much lower. Summing up over the entire AOI – which itself only makes up a portion of the city of Shenzhen – reveals that a total of 35.2 km² are covered by urban villages, constituting a proportion of 7.67% of the artificial land in the study area.

Table 2. Land cover fractions of the mapped area

	Water	Vegetation	Artificial	UV ²
SEZ ¹ area	12.04%	37.38%	49.16%	1.42%
Non-SEZ ¹ area	4.69%	33.24%	57.04%	5.03%

¹ SEZ = Special Economic Zone ² UV = Urban Village Class

4. Discussion

The performed experiments showed that mapping of a complex urban morphological type such as the UVs could be performed on a large scale using state-of-the-art image analysis methods (Hao et al. 2009). Thus, we can assume that FCNs provide a fast and effective method for the task of mapping UVs as traditional models which use a large number of hand-crafted inputs. Further, the transfer of the model to mapping UVs in

other Chinese cities appears reasonable using this methodological framework. At the same time, however, we must acknowledge two important limitations that have become apparent in this study:

Firstly, any model will perform better on data that is similar to the data it was trained on. We can see this, as transferring the model from the training regions to the testing region comes with a drop of approximately 4% in overall accuracy and approximately 11% in UV IoU (Table 1). This is significant, considering that while the regions are structurally different in many ways, they are very close geographically. It is likely that the transfer of the model towards other cities – where differences in climate, culture, and urban planning have led to structural differences both in the structure of UVs and their surroundings – will come with further decreases in accuracy, unless a significant effort is invested into tuning the model to the new region. The reference data necessary for such tuning is the primary limitation to upscaling the methodology to a national level. While another limitation lies in the availability and cost of VHR imagery it can likely be overcome by using a more available source such as Sentinel 2, the lower resolution of which can be partially compensated by the transfer of features learned on VHR imagery (Wurm et al. 2019). Thanks to the comparative efficiency of the FCN approach (Shelhamer, Long, and Darrell 2016), the technical limitations to upscaling are low and a nation-wide application of the method – and thereby the realization of the goals stated in the introduction – are feasible.

Secondly, it must be accepted that not all UVs can be observed from space. UVs are, as mentioned above, highly dynamic, and can have – besides their morphological characteristics – many other aspects which are not physical and therefore not measurable by remote sensing, similar to the variations in morphologies of slums/informal settlements (Wurm & Taubenböck 2018). What we can map is a

common type of urban morphology that highly correlates with this complex phenomenon. To further explore its geographical dimension, additional data sources – such as household surveys, social media or census data – need to be integrated and will consequently allow for deeper insights. Remote sensing using VHR satellite data can only deliver the basis for such analysis; yet as this study shows, it can perform this task well.

The necessity for such analysis on informal settlements is highlighted by the United Nations Sustainable Development Goal 11 to make cities inclusive, safe, resilient and sustainable (UN 2015). As one of its indicators it features the proportion of urban population living in slums, informal settlements or inadequate housing for which China reports 25.2% (UN 2014). However, due to the aforementioned ambiguous legal status of UVs – which makes them unique among informal settlements – cross-country comparisons are problematic. Remote sensing can provide an alternative perspective which is based not on administrative designations, but on the physical structure of the settlements.

Due to the limitations of remote-sensing, additional efforts need to be taken to ensure that the information that is collected in this manner is being integrated into a larger framework of research on urban development; a framework which ought to include knowledge and methods from the fields of urban planning, urban climatology, and socioeconomics. Such efforts would not go to waste, as for the significant number of people involved in the urbanization process any insight that is gained can have great leverage.

5. Conclusion

Concluding the presented work, we present a deep learning approach for successfully mapping urban villages using a fully convolutional neural network. By learning to

recognize their characteristic spatial features, the network was able to distinguish the urban villages from the spectrally similar surrounding urban fabric. Yet, while state of the art accuracy was achieved, some important limitations also became apparent:

Firstly, the non-tuned transfer of the model to new regions comes with a noticeable decrease in accuracy due to differences in the morphology of urban villages and their surroundings.

Secondly, not all urban villages can be mapped, only a common physical manifestation that many villages tend to develop under influence of urbanization.

While these points may seem trivial, they indicate the direction for future research. In order for the previously stated goals to be realized; in order to match the force of urbanization, efforts must be undertaken to explore and optimize the transfer of models between geographical regions, to create a consistent database of China's arrival cities. If successful, such a database can then be the basis for linking local projects and for conducting large-scale studies on the dynamics and functions of urban villages. This would go a long way towards understanding and, thus, managing the complex urbanization process in the most populous country in the world.

Disclosure Statement

The authors declare no conflict of interest.

Funding

This research received no external funding.

References

- Chung, H. 2010. „Building an Image of Villages-in-the-City: A Clarification of China’s Distinct Urban Spaces: Debates and Developments“. *International Journal of Urban and Regional Research* 34 (2): 421–37. <https://doi.org/10.1111/j.1468-2427.2010.00979.x>.
- Friesen, J., H. Taubenböck, M. Wurm, and P. F. Pelz. 2018. „The Similar Size of Slums“. *Habitat International* 73: 79–88. <https://doi.org/10.1016/j.habitatint.2018.02.002>.
- Hao, P., R. Sliuzas, Q. Zhan, and S. Geertman. 2009. Land use deviations from the plan: a GIS based analysis of urban development in Shenzhen. Paper presented at the 11th international conference on Computers in Urban Planning and Urban Management, Hongkong, July
- Hao, P. 2012. „Spatial Evolution of Urban Villages in Shenzhen“. PhD thesis., University Utrecht.
- Huang, X., H. Liu, and L. Zhang. 2015. „Spatiotemporal Detection and Analysis of Urban Villages in Mega City Regions of China Using High-Resolution Remotely Sensed Imagery“. *IEEE Transactions on Geoscience and Remote Sensing* 53 (7): 3639–57. <https://doi.org/10.1109/TGRS.2014.2380779>.
- Li, Y., X. Huang, and H. Liu. 2017. “Unsupervised Deep Feature Learning for Urban Village Detection from High-Resolution Remote Sensing Images.” *Photogrammetric Engineering & Remote Sensing* 83 (8): 567–79. <https://doi.org/10.14358/PERS.83.8.567>.
- Lin, Y., B. de Meulder, and S. Wang. 2011. „Understanding the ‘Village in the City’ in Guangzhou: Economic Integration and Development Issue and Their Implications for the Urban Migrant“. *Urban Studies* 48 (16): 3583–98. <https://doi.org/10.1177/0042098010396239>.
- Liu, H., Huang, X., Wen, D., and Li, Jiayi. 2017. „The Use of Landscape Metrics and Transfer Learning to Explore Urban Villages in China”. *Remote Sensing* 9, no. 4 (2017): 365. <https://doi.org/10.3390/rs9040365>.
- Liu, Y., S. He, F. Wu, and C. Webster. 2010. „Urban villages under China’s rapid urbanization: Unregulated assets and transitional neighbourhoods“. *Habitat International* 34 (2): 135–44. <https://doi.org/10.1016/j.habitatint.2009.08.003>.
- Saunders, D. 2012. *Arrival City: How the Largest Migration in History Is Reshaping Our World*. New York: Vintage Books.
- Shekhar, S. 2012. „Detecting Slums from QuickBird Data in Pune using an Object Oriented Approach“. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XXXIX-B8 (July)*: 519–24. <https://doi.org/10.5194/isprsarchives-XXXIX-B8-519-2012>.
- Shelhamer, E., J. Long, and T. Darrell. 2016. „Fully Convolutional Networks for Semantic Segmentation“. *ArXiv:1605.06211 [Cs]*, Mai. <http://arxiv.org/abs/1605.06211>.
- Simonyan, K., and A. Zisserman. 2014. „Very Deep Convolutional Networks for Large-Scale Image Recognition“. *ArXiv:1409.1556 [Cs]*, September. <http://arxiv.org/abs/1409.1556>.
- Song, Y., and Y. Zenou. 2012. „Urban Villages and Housing Values in China“. *Regional Science and Urban Economics* 42 (3): 495–505. <https://doi.org/10.1016/j.regsciurbeco.2011.06.003>.
- Stiller, D., T. Stark, M. Wurm, Dech, S., and Taubenböck, H. 2019. Large-scale building extraction in very high resolution aerial imagery using Mask R-CNN. Joint Urban Remote Sensing Event (JURSE), 22th -24thMay, Vannes (FR).
- Taubenböck, H., N.J. Kraff, and M. Wurm. 2018. „The Morphology of the Arrival City - A Global Categorization Based on Literature Surveys and Remotely Sensed Data“. *Applied Geography* 92: 150–67. <https://doi.org/10.1016/j.apgeog.2018.02.002>.
- United Nations (UN). 2015. “Cities - United Nations Sustainable Development Action 2015. United Nations Sustainable Development (blog).” Accessed January 13, 2020. <https://www.un.org/sustainabledevelopment/cities/>.

- United Nations (UN), Department of Economic and Social Affairs, Population Division. 2018. "World Urbanization Prospects: The 2018 Revision, Online Edition."
- United Nations (UN), Department of Economic and Social Affairs, Statistics Division. 2014. "Millennium Indicators." Accessed January 13, 2020. <http://mdgs.un.org/unsd/mdg/Data.aspx>.
- Wang, J., Kuffer, M., Roy, D., & Pfeffer, K. (2019). Deprivation pockets through the lens of convolutional neural networks. *Remote sensing of environment*, 234, 111448.
- Wang, Y. P., Y. Wang, and J. Wu. 2009. „Urbanization and Informal Development in China: Urban Villages in Shenzhen“. *International Journal of Urban and Regional Research* 33 (4): 957–73. <https://doi.org/10.1111/j.1468-2427.2009.00891.x>.
- Wei, C., T. Blaschke, and H. Taubenböck. 2015. "Monitoring of 'Urban Villages' in Shenzhen, China from High-Resolution GF-1 and TerraSAR-X Data." In *SAR Image Analysis, Modeling, and Techniques XV*, 9642:964210. International Society for Optics and Photonics. <https://doi.org/10.1117/12.2194877>
- Wu, F., F. Zhang, and C. Webster. 2013. „Informality and the Development and Demolition of Urban Villages in the Chinese Peri-Urban Area“. *Urban Studies* 50 (10): 1919–34. <https://doi.org/10.1177/0042098012466600>.
- Wurm, M., T. Stark, X. X. Zhu, M. Weigand, and H. Taubenböck. 2019. „Semantic Segmentation of Slums in Satellite Images Using Transfer Learning on Fully Convolutional Neural Networks“. *ISPRS Journal of Photogrammetry and Remote Sensing* 150: 59–69. <https://doi.org/10.1016/j.isprsjprs.2019.02.006>.
- Wurm, M., and H. Taubenböck. 2018. "Detecting social groups from space - Assessment of remote sensing-based mapped morphological slums using income data." *Remote Sensing Letters* 9 (1), pp. 41-50. Taylor & Francis. <https://dx.doi.org/10.1080/2150704X.2017.1384586>
- Zhang, L., S. X. B. Zhao, and J. P. Tian. 2003. „Self-Help in Housing and Chengzhongcun in China's Urbanization“. *International Journal of Urban and Regional Research* 27 (4): 912–37. <https://doi.org/10.1111/j.0309-1317.2003.00491.x>.
- Zhu, X. X., D. Tuia, L. Mou, G.-S. Xia, L. Zhang, F. Xu, and F. Fraundorfer. 2017. „Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources“. *IEEE Geoscience and Remote Sensing Magazine* 5 (4): 8–36. <https://doi.org/10.1109/MGRS.2017.2762307>.