The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLVIII-4/W5-2022
7th International Conference on Smart Data and Smart Cities (SDSC), 19–21 October 2022, Sydney, Australia

ESTIMATING SPATIO-TEMPORAL URBAN DEVELOPMENT USING AI

M. A. Waseem, M. A. Basheer*, M. Uppal, M. Tahir

Department of Electrical Engineering, Lahore University of Management Sciences, Lahore 54792, Pakistan
(m_waseem, muhammad.basheer, momin.uppal, tahir)@lums.edu.pk

Commission IV, WG IV/9

KEY WORDS: Building Counts, Spatio-temporal profile, Digital Maps, Semantic Segmentation, Satellite Imagery, Remote Sensing, Urban data

ABSTRACT:

Estimating the spatio-temporal profile of a building's construction using high-resolution satellite images is a critical problem since it can be utilized for a variety of data-driven urban initiatives. One strategy to achieve this is to extract building footprints and track them in multi-temporal data as observed in SpaceNet's Challenges. Although several unique solutions have been presented for this problem, this task can become extremely difficult for partially obscured buildings with densely overlapping boundaries, such as those found in underdeveloped countries like Pakistan. Consequently, in this paper we propose a framework to address this problem by merging built-up area segmentation with digital maps. In the first step, satellite image is passed to a deep learning model that predicts segmentation masks over the built-up area following which building construction profiles are generated by overlaying digital maps over these predicted masks. We compare the results with ground truth profiles and our results show that the proposed method extracts building counts and construction profiles with an accuracy of 95%.

1. INTRODUCTION

Urban planning has become more crucial than ever because of the rapidly changing urban environment and human development patterns. To meet the demands of the present and future communities, urban planners will need to be more data-driven in their planning to enable optimal land and infrastructure solutions. In this regard, constructing a spatio-temporal profile of the development of buildings is vital since it is used in a variety of applications, such as urban sprawl analyses, population estimation, mobile targeting, managing infrastructure deployment, and enhancing citizens' access to services. Traditional methods for these types of tasks are usually based on onsite measurements and surveys that require a lot of human effort, time, and resources. With the advancements in remote sensing technologies, it is now possible to extract these spatio-temporal profiles from high-resolution satellite images.

The task of extracting the spatio-temporal construction profile of buildings has been tackled primarily by three types of methods. These include classical methods, building footprint extraction methods, and regression methods. A brief review of each type of method will be provided in Section 2. Although the results of these methods are excellent and they have been used in a variety of applications effectively, their performance is still uncertain for developing countries like Pakistan. The main reason for this is that these developing countries have diverse development patterns, including a lot of partially occluded buildings that are densely packed together. Therefore, the state-of-the-art building extraction and object detection methods fail to perform well. This problem is represented in Figure 1, where similar segmentation models have been used for the very high resolution (V.H.R) images of USA, as well as the relatively low resolution and densely packed areas of Pakistan. The buildings footprints have been extracted easily and accurately for USA, while the model is unable to accurately detect building boundaries for densely packed areas of Pakistan. In addition, the unavailability of very high to ultra high-resolution images (0.1m - 0.01m per pixel) publicly makes it difficult to extract building footprints with high accuracy.

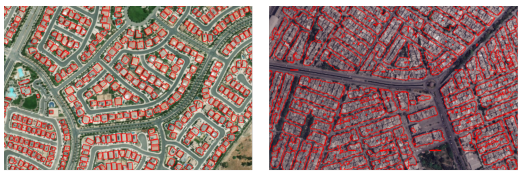

Figure 1. A comparison of building footprint extraction in different areas. On the left side, successful building extraction has been performed using deep learning models with high-resolution imagery in the USA (ESRI, 2020). On the right, similar methods have been applied but with much lower resolution and densely packed areas of Pakistan.

To overcome the above-mentioned problems, we propose a two-step novel approach that utilizes digital maps, along with state-of-the-art deep learning methods to establish spatio-temporal building profiles at any given location. At first, we train a deep learning model based on DeepLabV3plus architecture (Chen et al., 2018a) for semantic segmentation of built-up areas in satel-lite imagery. The built-up segmentation masks are then overlaid with digital maps to extract the construction profile for each building in that area. The details of our proposed methodology will be explained in Section 3. To obtain better results, we pre-pare a small dataset of 730 satellite images which include over 25,000 buildings spanning around 57 sq. km of the land area of Lahore, the second largest city of Pakistan.

The main contributions of our work are: (1) We create a dataset that allows us to capture the building patterns in complex

* Corresponding author

urban scenarios, and (2) We develop a tool for extracting and tracking building construction profiles in time-series. This technique could be employed by policy-makers and urban planners for evidence based and data driven policy making.

The remainder of this paper is organized as follows. Section 2 highlights the diverse building extraction and object detection methods being used by different researchers followed by methodological framework. Section 3 explains the end-to-end pipeline used in this research for developing building construction profiles. After this, Section 4 quantifies the efficacy of our proposed method by comparing the generated results with ground truth data. Section 5 then delineates the main findings of this research and explains how our proposed pipeline can be used to estimate the growth in number of buildings in different regions of Pakistan over the time. Finally, we conclude our paper in Section 6 by outlining our contribution and providing future research directions

2. RELATED WORK

In this section, we will discuss various approaches that have been employed in existing literature for obtaining building profiles or estimating urban sprawl. As discussed earlier, three major methods have been used to extract and count buildings from satellite data, i.e., classical methods, building footprint extraction methods, and regression methods. Let us look into each one briefly.

2.1 Classical Methods

Several classical methods that do not include deep models have been proposed for estimating urban sprawl and population density estimation. Most of these methods use Landsat's images to perform land use land cover (LULC) classification, and the number of pixels per class gives a rough estimate of urban growth (El Garouani et al., 2017; Shah et al., 2021; Sahana et al., 2018). However, such methods cannot give a quality estimate at a building level because of the low resolution (maximum of 15m per pixel) of the Landsat images. In contrast to Landsat images, some methods also try to estimate urban sprawl through population density estimates (Zhang, 2003; Terzi and Kaya, 2008), which are also rarely available for many underdeveloped and developing countries. Moreover, linear modeling was applied on fine-resolution (1m per pixel) LiDAR data along with satellite images to estimate building counts (Silvan-Cardenas et al., 2010), but again the approach is quite expensive and non-scalable.

2.2 Building Footprint Extraction Methods

With the advent of remote sensing technology, researchers are effectively employing computer vision techniques to extract building footprints. Several mathematical models have been proposed so far (Ok et al., 2013; Huang et al., 2014; Chen et al., 2018b) for this purpose, but they require very high resolution images (0.01m per pixel) to give good results. To overcome this issue, researchers began to use deep learning approaches for the task of building detection and footprint extraction. The availability of freely available datasets and open challenges (Wang et al., 2016; Maggiori et al., 2017; Ji et al., 2018; ISPRS 2D Semantic Labeling Contest, n.d.; Van Etten et al., 2018; Gupta et al., 2019 have also boosted the interests in this area. The Space-Net challenges (Van Etten et al., 2018), in particular, have demonstrated the feasibility of extracting buildings from medium-resolution satellite images (1–4m per pixel). Hence a lot of solutions, including segmentation along with post-processing (Yuan, 2018; Liu et al., 2018), instance segmentation (Wen et al., 2019; Zhao et al., 2018), generative adversarial networks (GANs) (Li et al., 2018; Shi et al., 2018), customized networks (Hui et al., 2019; Liu et al., 2019), and graph-based networks (Qin et al., 2018) have been proposed in recent years. Moreover, trained models are also available on ArcGIS (ESRI, 2020) which can be used directly to extract building footprints. However, as explained in Section 1, these methods still struggle to deliver effective results for areas with varying architectural designs, and tightly packed buildings with no visible gaps between consecutive buildings.

2.3 Regression Methods

Apart from the above two methods that count buildings in an area implicitly, several approaches in the literature directly count the number of buildings from satellite images and other GIS data. For example, micro-scale data for spatio-temporal modeling of building population estimation (Greger, 2015) was done for highly urbanized areas. A mathematical model was formulated that tries to count the number of buildings by counting the number of objects of a certain class in a desired region (Meng et al., 2021). A deep learning-based regression model was proposed (Shakeel et al., 2019) for counting built-up areas in satellite imagery. To adapt the pre-trained building counting models on the developed countries for the under-developing countries with unlabelled data, counting consistencies have been used (Zakria et al., 2021). Although these methods have proved to be quite effective, most of these provide counts for a region (either an image or a whole area) and do not provide within region densities, i.e., estimating the construction profile at the level of individual building.

3. METHODOLOGY

From the arguments in the preceding section, it is clear that the deliberate or subliminal aim in existing literature is to generate building profiles. Each suggested solution however has certain limitations. For instance, some methods need high-quality data, which prevents them from being extended to other areas, while others provide urban growth detection at higher urban scales and do not offer estimates at the individual building level. In this section, we will go over our suggested methods for addressing these challenges. First, we will discuss the data requirements, where we demonstrate how easily accessible data can be used to solve the issue. Next, we will discuss the process of model training, where we use the collected data to train a deep learning model for built-up area semantic segmentation. After that, we will describe the end-to-end pipeline we utilize to predict the construction profile of each (individual) building. The pipeline makes use of the trained model as well as additional data extracted from digitized maps.

3.1 Data Requirements

Two key data sources are required to use our suggested pipeline: (i) RGB satellite images of the region/area of interest, and (ii) geo-tagged digitized map data with the locations of all available plots in that area. We used satellite imagery from Google's

repository, which is freely available for academic usage. To create a semantic segmentation dataset for our region, we label a fraction of images using GIS tools. As a result, our deep model works incredibly well for our areas. The digital map information was taken from the openly downloadable housing plans for about 350+ societies in Lahore, Pakistan. Let us now discuss each individual step in detail.

3.1.1 Labelling The trained models for the developed world do not perform well in segmenting out the built-up regions in the under-developing nations as has been highlighted in the preceding sections. It is, therefore, necessary to gather segmentation data for these regions. To deal with this problem, we develop a dataset by manually marking building footprints for a portion of Lahore, a metropolitan city in Pakistan. We selected the region of the Defence Housing Authority (DHA) from Lahore, which is roughly 57 sq. km in area. Using satellite imagery, we manually map the footprints of 24,928 building structures that are located in this region. Figure 2 displays the overview of the markings in the indicated location. Since we are working on multi-temporal data, therefore we also mark a portion of historical images as well.

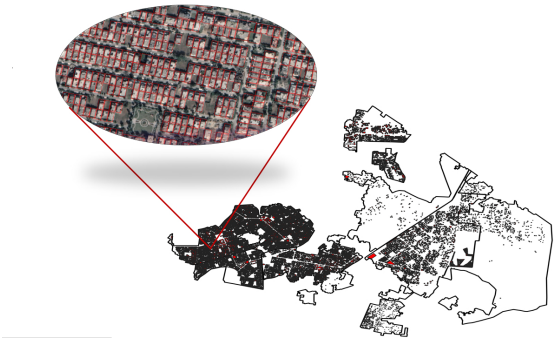

Figure 2. Details of the marked Area. The complete marked area of Defence Housing Authority (DHA), Lahore, along with a zoomed in look at the marked footprints

3.1.2 Extracting Segmentation Data The marked/digitized polygons along with satellite image are then used to extract segmentation data. The next step is to transform this data into tiny images and masks so that they can be fed directly to the deep model. For this, the entire region is divided into 300m x 300m sized tiles from which images and their respective masks are extracted. Additionally, we mask off the area's component that was not a part of the marked zone. Figure 3 depicts the masking procedure and tile extraction. We prepared a dataset of 740 images for the DHA region using this procedure.

3.1.3 Collecting Digitized Map Data The data gathered can be leveraged to train the deep model and, in some way, be used to make predictions about built-up regions. However, another form of data—the geo-locations of the accessible plots in the specified region of interest—is required for our pipeline to make inferences at the scale of buildings. To do this, we first gathered the society maps (in the form of images) for more than 350 societies in Lahore. After that, we vectorize these maps of the societies using raster to vector conversion tools from the GDAL package. Then, as shown in Figure 4, we extract the center points from each polygon on the vectorized map, which provides us with the Geo-locations of the available plots. Note that the map includes information about plots of land only; it does not indicate whether a particular plot of land has construction on it.

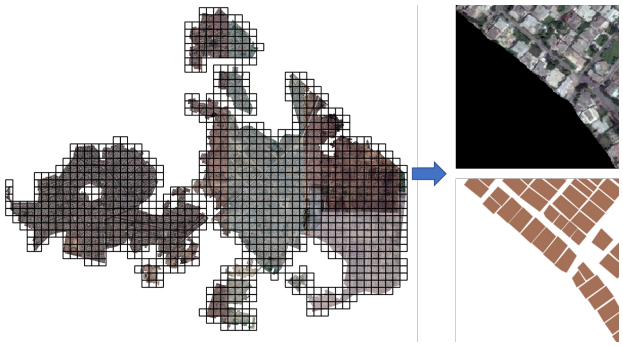

Figure 3. Tiling of Images. Here we show our process of extracting images from Geo-tiff file. On left, grid of 300m x 300m is overlaid over the marked region. On the right, the extracted image (upper right) and it's corresponding segmentation mask (lower right) from one of the tiles is shown. The black area shows the unmarked region.

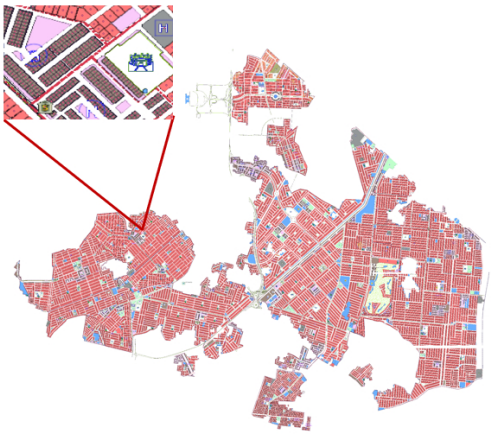

Figure 4. Extracting Geo-Locations. An example of one of the geo-referenced digitized map, along with the extraction of center points using raster to vector and other GIS tools.

3.2 Model Training

Once the data is obtained, we train a deep learning model based on DeepLabV3plus (Chen et al., 2018a) architecture with Resnet50 encoder for semantic segmentation of built-up regions. Since our dataset is really small, therefore training a deep network like DeepLabV3plus, with millions of parameters, from scratch may lead to over-fitting. To prevent this potential problem, we initialize the model with pre-trained weights on the image-net dataset. After the initialization, we set the number of classes of the last convolutional layer to 2 (building or non-building). Once this is complete, we unfreeze all the layers and then fine-tune the model using our small dataset. We train the model with a batch size of 4 and a learning rate of 0.00008, and we train it for 80 Epochs.

3.3 Proposed Pipeline

Once our deep model is trained, we have completed every requirement for our proposed pipeline. Hence, we can now com-

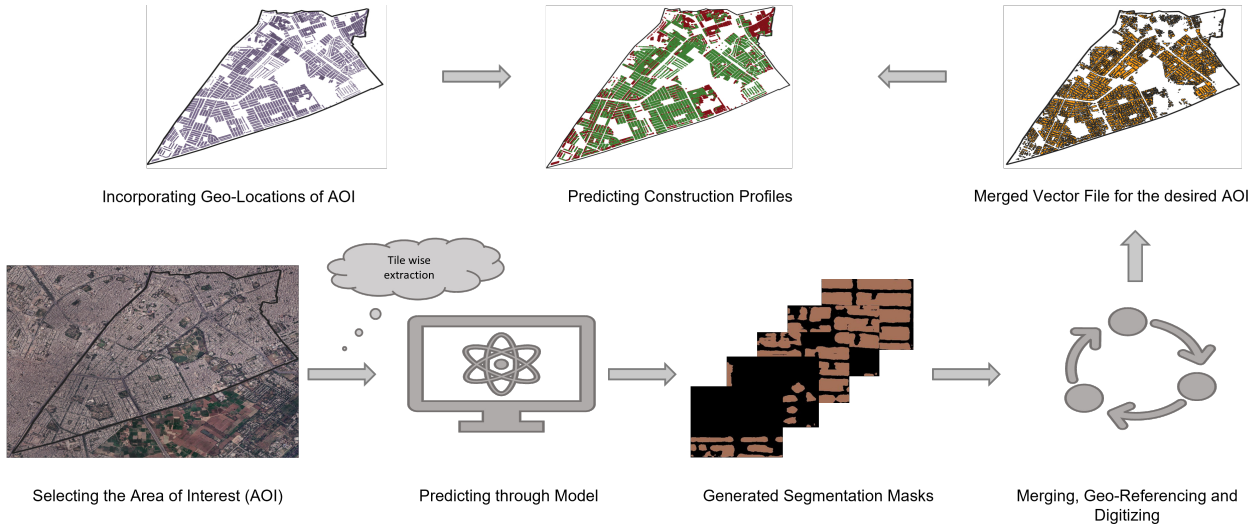

Figure 5. The Proposed Pipeline. The first step is the selection of AOI and providing its satellite image to the trained model through tiling. After that, the generated masks are merged, geo-referenced, and digitized into a vector file. In the end, the vectorized file is overlaid with the extracted geo-locations for that AOI, and profiles are developed. The green dots in the final image represent the construction of that specific building, while the red dots show that they weren't constructed.

bine these collected blocks in an end-to-end manner and predict the construction profiles of each individual plot in the provided region. Our proposed pipeline is shown in Figure 5, where we first pass images of our area of interest to the deep model. The deep learning model returns us segmentation masks for each of the passed images. We then apply a post-processing stage where we simply merge these images, geo-tag the merged file, and then digitize the merged raster to obtain vector/polygons over the built-up regions. In the end, we overlay the plot locations extracted from digitized society maps with these predicted polygons and develop the construction profiles of each building. We next describe each part of the proposed pipeline in detail.

3.3.1 Selection of AOI and tiling The first step in our pipeline is the selection of an area of interest (AOI) and providing the satellite image for that specific AOI. We extract 300m x 300m tiles from this large image using the step mentioned earlier. For a multi-temporal analysis like in our case, images at each of the specified times are collected, and the process is repeated for each of these images.

3.3.2 Inferences from the Model Once we have data in the form of images, we pass these images to our trained deep model. The model predicts a segmentation mask for each of these images, providing 1 for pixels labelled built-up and 0 otherwise. These binary masks are saved for each image.

3.3.3 Merging, Geo-Referencing, and Digitizing Once masks against each of the images have been predicted, the next task is to apply post-processing methods to convert them to vectors/ polygons so that they can be used as an overlay layer for the geo-locations of the plots. For this purpose, we employ tools from the Geospatial Data Abstraction Library (GDAL). We use the coordinates that were used to extract each tile for its geo-referencing and then combine each of the geo-referenced tiles. Once all tiles are geo-referenced, we vectorize this large image with the help of the value assigned to built-up pixels. Hence, we obtain a single vector file for our desired AOI that contains the information of the "built-up regions". To use this information to extract the construction profile of each building, we need to go through one more step, which is described next.

3.3.4 Extracting building profiles The built-up regions extracted in the previous step can alone be utilized to formulate many policies. However, they do not provide granular information about the construction or non-construction of each individual building in that area. Many footprint extraction methods try to add some classic post-processing stages, but the results are not good enough. We propose to solve this by formulating an easier problem than extracting footprints. We use the plots' geo-locations, extracted from the digitized maps, and assign them labels based on their position with respect to the generated built-up regions. Each plot of land is considered to have construction on it if it lies within any of the built-up regions, and un-built otherwise. Mathematically, this is akin to assigning a binary label y_i to each plot of land such that.

$$y_i = I(x_i, r) \quad \forall r \in R \tag{1}$$

Here, i represents the time index while $x_i \in L$ is the location of the i^{th} plot in that area. $r \in R$ is the built-up region extracted by our model, while $I(a, b)$ is an indicator function showing whether the point a lies within any of the regions in b or not. In particular, $I(a, b) = 1$ if $a \in b$, and is equal to 0 otherwise. Thus, $y_i \in [0, 1]$ is the estimated construction profile for that building at a specific time. Using this straightforward mathematical comparison, one can predict built-up labels for each of the plots in a given area at a specific time. As a result, our technique can be utilized to produce spatio-temporal development profiles of building construction for any specified metropolitan region with ease.

4. ACCURACY ASSESSMENT

In this section, we measure the performance of our model using different performance metrics. We compare the generated profiles from our pipeline, to the ground truth profiles. The ground truth data is simply obtained by using (1), but instead of predicted regions, we use ground truth building footprints as an overlay layer. First, we will check the performance of the model using a confusion matrix and afterward, we will see how does the model performs in estimating total number of building

counts for different spatial and temporal locations. In this way, we can better know about efficiency of the model as well as its consistency across spatial and temporal data.

4.1 Confusion Matrix

We use our ground truth footprints in (1) in place of predicted regions R to establish the ground truth profiles of the buildings. After that, we compare both predicted profiles and ground truth profiles to create the confusion matrix. For example, if a specific building's profile is labelled as built-up in ground truth, while it was labelled un-built by our pipeline, than it is termed as false negative, and so on. Using these settings, we obtain the confusion matrix as shown in Table 1.

		Predicted Label		
		$\hat{P}$	$\hat{N}$	**Total**
Ground Truth Label	**P**	25893	2144	28037
	N	2047	57292	59339
	Total	27940	59436	87376

Table 1. The confusion matrix for building construction profiles. $\hat{P}$ and $\hat{N}$ are the ground truth labels for built-up and un-built, respectively. P and N are the predicted la-bels for built-up and un-built, respectively.

It is clear from the table that our model predicts the construction profiles with an accuracy of 95%. The precision & recall score is approximately 92%, which is excellent.

4.2 Spatial Consistency

In spatial consistency, we count the total number of constructed buildings in different areas and compare them with the ground truth count. In this way, the performance of our proposed pipeline in estimating building counts for a desired area can be evaluated. For this purpose, we sub-divide the marked area into 12 regions and we compare the total counts from model generated profiles with the ones generated from ground truth. To make calculations consistent over the size of regions, we divide the counts with the total number of plots of that region and we call it built-up ratio, as described in (2).

$$BR = \frac{\sum_i^N y_i}{N} \; \forall y_i \in Y \tag{2}$$

Here, BR is the output built-up ratio, Y is a set of profiles of buildings in that area, and N is the size of set Y. In this way, areas with 10,000 plots and 1000 plots can be evaluated using the same scale. Using this equation, built-up ratios for both ground truth and predicted profiles were calculated for each of the 12 regions. The results of the evaluation are shown in Figure 6.

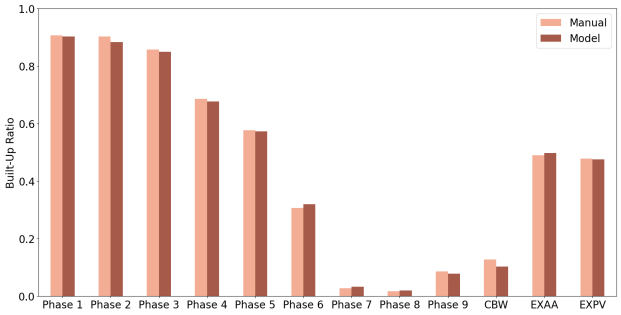

Figure 6. Spatial Consistency of the model: The results for spatial consistency performance of our model. On x-axis, we have different regions of DHA and on y-axis we show the estimated and ground truth BRs for that region.

The model is performing extremely well in almost all cases, with an average deviation of $\pm 0.01 (or\ 1\%)$ form ground truth built-up ratios.

4.3 Temporal Consistency

In temporal continuity, we perform the same analysis as in spatial consistency, but here we change the temporal dimension while keeping the area constant. In this way, the performance of the model over different dates on the same area can be evaluated. As we had marked a portion of past images to incorporate historical data in training as well, therefore we used that to develop ground truth profiles for that portion. We apply our pipeline to 17 different dates distributed between the year 2010 to the year 2020. The results for evaluation are shown in Figure 7.

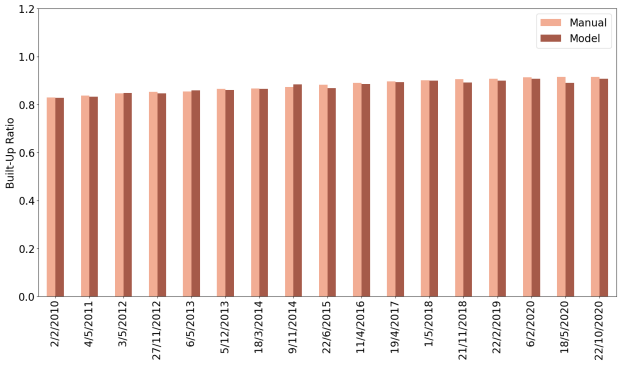

Figure 7. Temporal Consistency of the model: The results for temporal consistency performance of our model. On x-axis, we have images on different dates (dd/mm/yyyy format) for same regions (Phase 1) of DHA and on y-axis we show the estimated and ground truth BRs for that time.

The average deviation from ground truth built-up ratios is approximately $\pm 0.007 (or\ 0.7\%)$, which is extremely good.

5. ANALYSES AND RESULTS

In this section, we will use our proposed pipeline to analyze and quantify urban sprawl of DHA from 2010 to 2021. Here, we will be counting the growth in the number of buildings over the years. Figure 8 shows the a change in one of the regions of

Figure 8. Growth in Phase-5 DHA: The analysis results on one of the regions (Phase 5) of DHA. The left most image is for 2010, the center image is for 2014, and the last image is for 2020.

DHA over these years. We note that these type of visualizations can be performed with only segmentation model or even with the help of Landsat Images. However, to quantify the growth, we need to further post-process our results using (1) and (2). Let us now see quantified growth in number of buildings, as well as the built-up ratios, for the area of DHA.

5.1 Estimating urban sprawl

We collect satellite images for 17 different dates for Defence Housing Authority (DHA) to perform the urban sprawl analysis. We pass these images to our proposed pipeline and estimate building construction profiles for each of the mentioned date. The complete analysis show that that the building count for DHA was 14233 (built-up ratio 16.2%) in 2010, which has now increased to 27967 (built-up ratio 32%) in 2020. This shows that in only ten years, the number of buildings in DHA have doubled, which is a significant rise. This results are shown in Figure 9.

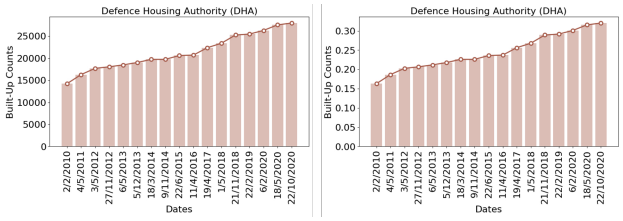

Figure 9. Urban Sprawl in DHA: The results of building counts (left) and built-up ratios (BRs) (right) with x-axis representing the date (dd/mm/yyyy format).

6. CONCLUSION AND FUTURE DIRECTIONS

Establishing construction profiles of buildings for a region is an essential problem in the process of collecting large-scale urban data; nonetheless, relatively few people have concentrated on directly tackling this problem. The suggested framework is based on datasets that are publicly available. The accuracy assessments for our pipeline indicate excellent performance across a variety of evaluation metrics. As a consequence of this, the findings are relevant to the work of both researchers and urban planners. In addition, the framework can further be put to use in the execution of spatio-temporal sprawl assessments, as was covered in the preceding sections. However, the availability of society maps can become a bottleneck in the use of our pipeline for many places, such as the slums (katchi abadis) of South Asia. Hence, for future study, we intend to focus on extracting such information without including society maps or plots geo-locations.

ACKNOWLEDGEMENTS

This work was supported financially by the Higher Education Commission (HEC) of Pakistan through a Grand Challenge Fund Grant No. GCF-521.

REFERENCES

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018a. Encoder-decoder with atrous separable convolution for semantic image segmentation. *Proceedings of the European conference on computer vision (ECCV)*, 801–818.

Chen, R., Li, X., Li, J., 2018b. Object-Based Features for House Detection from RGB High-Resolution Images. *Remote Sensing*, 10(3). https://www.mdpi.com/2072-4292/10/3/451.

El Garouani, A., Mulla, D. J., El Garouani, S., Knight, J., 2017. Analysis of urban growth and sprawl from remote sensing data: Case of Fez, Morocco. *International Journal of Sustainable Built Environment*, 6(1), 160-169. https://www.sciencedirect.com/science/article/pii/S2212609016300668.

ESRI, 2020. Building Footprint Extraction. https://www.arcgis.com/home/item.html?id=a6857359a1cd44839781a4f113cd5934.

Greger, K., 2015. Spatio-Temporal Building Population Estimation for Highly Urbanized Areas Using GIS. *Transactions in GIS*, 19(1), 129–150.

Gupta, R., Hosfelt, R., Sajeev, S., Patel, N., Goodman, B., Doshi, J., Heim, E., Choset, H., Gaston, M., 2019. xbd: A dataset for assessing building damage from satellite imagery.

Huang, X., Zhang, L., Zhu, T., 2014. Building Change Detection From Multitemporal High-Resolution Remotely Sensed Images Based on a Morphological Building Index. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(1), 105-115.

Hui, J., Du, M., Ye, X., Qin, Q., Sui, J., 2019. Effective Building Extraction From High-Resolution Remote Sensing Images With Multitask Driven Deep Neural Network. *IEEE Geoscience and Remote Sensing Letters*, 16(5), 786-790.

ISPRS 2D Semantic Labeling Contest, n.d. Accessed: May 27, 2022 [online]. Available: "http://www2.isprs.org/commissions/comm3/wg4/semantic-labeling.html".

Ji, S., Wei, S., Lu, M., 2018. Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery data set. *IEEE Transactions on Geoscience and Remote Sensing*, 57(1), 574–586.

Li, X., Yao, X., Fang, Y., 2018. Building-A-Nets: Robust Building Extraction From High-Resolution Remote Sensing Images With Adversarial Networks. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(10), 3680-3687.

Liu, P., Liu, X., Liu, M., Shi, Q., Yang, J., Xu, X., Zhang, Y., 2019. Building Footprint Extraction from High-Resolution Images via Spatial Residual Inception Convolutional Neural Network. *Remote Sensing*, 11(7). https://www.mdpi.com/2072-4292/11/7/830.

Liu, Y., Zhang, Z., Zhong, R., Chen, D., Ke, Y., Peethambaran, J., Chen, C., Sun, L., 2018. Multilevel Building Detection Framework in Remote Sensing Images Based on Convolutional Neural Networks. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(10), 3688-3700.

Maggiori, E., Tarabalka, Y., Charpiat, G., Alliez, P., 2017. Can semantic labeling methods generalize to any city? the inria aerial image labeling benchmark. *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, IEEE, 3226–3229.

Meng, C., Liu, E., Neiswanger, W., Song, J., Burke, M., Lobell, D., Ermon, S., 2021. IS-COUNT: Large-scale Object Counting from Satellite Images with Covariate-based Importance Sampling. *arXiv preprint arXiv:2112.09126*.

Ok, A., Senaras, C., Yuksel, B., 2013. Automated Detection of Arbitrarily Shaped Buildings in Complex Environments From Monocular VHR Optical Satellite Imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 51, 1701-1717.

Qin, X., He, S., Yang, X., Dehghan, M., Qin, Q., Martin, J., 2018. Accurate Outline Extraction of Individual Building From Very High-Resolution Optical Images. *IEEE Geoscience and Remote Sensing Letters*, 15(11), 1775-1779.

Sahana, M., Hong, H., Sajjad, H., 2018. Analyzing urban spatial patterns and trend of urban growth using urban sprawl matrix: A study on Kolkata urban agglomeration, India. *Science of The Total Environment*, 628-629, 1557-1566. https://www.sciencedirect.com/science/article/pii/S0048969718305631.

Shah, A., Ali, K., Nizami, S. M., 2021. Spatio-temporal analysis of urban sprawl in Islamabad, Pakistan during 1979–2019, using remote sensing. *GeoJournal*. https://doi.org/10.1007/s10708-021-10413-6.

Shakeel, A., Sultani, W., Ali, M., 2019. Deep built-structure counting in satellite imagery using attention based re-weighting. *ISPRS journal of photogrammetry and remote sensing*, 151, 313–321.

Shi, Y., Li, Q., Zhu, X. X., 2018. Building footprint generation using improved generative adversarial networks. *IEEE Geoscience and Remote Sensing Letters*, 16(4), 603–607.

Silvan-Cardenas, J. L., Wang, L., Rogerson, P., Wu, C., Feng, T., Kamphaus, B. D., 2010. Assessing fine-spatial-resolution remote sensing for small-area population estimation. *International Journal of Remote Sensing*, 31(21), 5605–5634.

Terzi, F., Kaya, H. S., 2008. Analyzing urban sprawl patterns through fractal geometry: The case of Istanbul metropolitan area.

Van Etten, A., Lindenbaum, D., Bacastow, T. M., 2018. Spacenet: A remote sensing dataset and challenge series. *arXiv preprint arXiv:1807.01232*.

Wang, S., Bai, M., Mattyus, G., Chu, H., Luo, W., Yang, B., Liang, J., Cheverie, J., Fidler, S., Urtasun, R., 2016. Torontocity: Seeing the world with a million eyes. *arXiv preprint arXiv:1612.00423*.

Wen, Q., Jiang, K., Wang, W., Liu, Q., Guo, Q., Li, L., Wang, P., 2019. Automatic building extraction from Google Earth images under complex backgrounds based on deep instance segmentation network. *Sensors*, 19(2), 333.

Yuan, J., 2018. Learning Building Extraction in Aerial Scenes with Convolutional Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(11), 2793-2798.

Zakria, M., Rawal, H., Sultani, W., Ali, M., 2021. Cross-Region Building Counting in Satellite Imagery using Counting Consistency. *arXiv preprint arXiv:2110.13558*.

Zhang, B.-g., 2003. Application of remote sensing technology to population estimation. *Chinese geographical science*, 13(3), 267–271.

Zhao, K., Kang, J., Jung, J., Sohn, G., 2018. Building extraction from satellite images using mask r-cnn with building boundary regularization. *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 247–251.