Sidewalk Measurements from Satellite Images: Preliminary Findings

Maryam Hosseini[1,4,*], Iago B. Araujo[2],
Hamed Yazdanpanah[2], Eric K. Tokuda[2],
Fabio Miranda[3], Claudio T. Silva[1], and Roberto M. Cesar Jr.[2]

[1] New York University (NYU), New York, NY, United States
[2] University of São Paulo (USP), São Paulo, SP, Brazil
[3] University of Illinois at Chicago (UIC), Chicago, IL, United States
[4] Rutgers University, Newark, NJ, United States

Abstract. Large-scale analysis of pedestrian infrastructures, particularly sidewalks, is critical to human-centric urban planning and design. Benefiting from the rich data set of planimetric features and high-resolution orthoimages provided through the New York City Open Data portal, we train a computer vision model to detect sidewalks, roads, and buildings from remote-sensing imagery and achieve 83% mIoU over held-out test set. We apply shape analysis techniques to study different attributes of the extracted sidewalks. More specifically, we do a tile-wise analysis of the width, angle, and curvature of sidewalks, which aside from their general impacts on walkability and accessibility of urban areas, are known to have significant roles in the mobility of wheelchair users. The preliminary results are promising, glimpsing the potential of the proposed approach to be adopted in different cities, enabling researchers and practitioners to have a more vivid picture of the pedestrian realm.

Keywords: sidewalks · semantic segmentation · urban analytics · accessibility.

1 Introduction

Pedestrian infrastructure has a significant impact on the everyday life of people, specifically those with special needs, for whom such infrastructures are the primary means of accessing public spaces [14]. The presence, condition, width, and shape of sidewalks are shown to impact pedestrians' safety and accessibility [12]. The curvature is one of the critical factors in ensuring safe navigation for wheelchair users [1]. Despite their importance, there is a significant lack of city-wide fine-level sidewalk data, which poses challenges to the assessment and planning of pedestrian infrastructures [3]. The time and cost-intensive nature of in-field data collection have for long been a limiting factor in expanding the

---

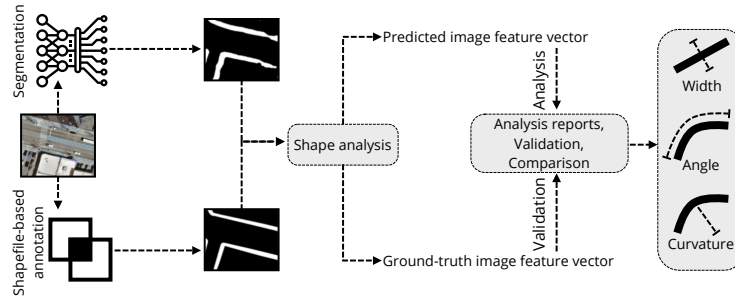* Corresponding author: `maryam.hosseini@nyu.edu`

Fig. 1: Proposed method with both analysis and validation components.

research on the built environment [13]. The availability of new sources of data such as aerial and street-level images, together with the advent of new computer vision techniques, opened new frontiers in measuring the physical form of the cities. [7,14]. Sidewalks, however, have been noticeably overlooked. The majority of the sidewalk detection models using satellite images suffer from low prediction accuracy since sidewalks occupy a very small portion of the visual information in satellite images compared to roads and buildings, are often obstructed by trees, bridges, and other urban structures, and their reduced area are particularly affected by sunlight condition [5,15]. A promising approach was using street-level images to tackle the obstruction issue; however, the resulting sidewalks were represented as polyline features, making them unsuitable for various analyses such as width measurement [8].

To address these challenges, we introduce a method for detecting and morphological analysis of sidewalks from orthorectified aerial images. To overcome the high cost of pixel-wise annotation, we use publicly available planimetric data of New York City (NYC) (described in Section 2) to create accurate ground truth annotations for the obtained orthoimages. We train a multi-scale attention-based semantic segmentation model [16] to detect roads, sidewalks, and buildings from aerial images. The specific architecture of the model enables detecting the sidewalks with very high precision, while the properties of the planimetric sidewalk data allow the model to make a correct prediction for various instances of the occluded sidewalks. The trained model is then employed to extract the sidewalk data from unlabeled images of Manhattan to be used for further analysis.

The main contributions of this paper are: (1) A method that leverages a rich, publicly-available data set as the ground truth to train and test a recently developed semantic segmentation model capable of detecting sidewalks with high accuracy. (2) Introduction of a robust approach to estimate meaningful sidewalk features (width, curvature, and angle) from the segmentation results.

2   Materials and method

To detect sidewalks from satellite images, we trained an attention-based semantic segmentation model using orthorectified images and annotation masks created from planimetric shapefiles of sidewalks, roads, and buildings. The trained model
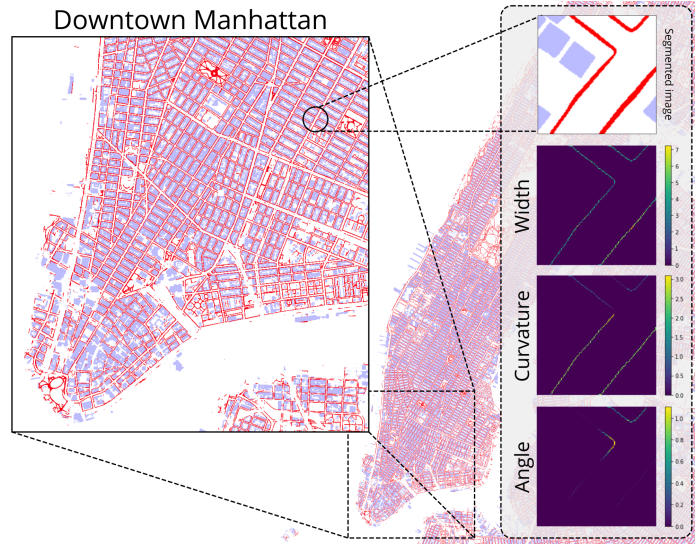
Fig. 2: New York City segmented sidewalks (red) and buildings (blue). All other pixels are labeled white to highlight the sidewalks segmentation. The figure also highlights examples of the features extracted from a segmented tile.

is employed to extract the sidewalks from unlabeled images. Shape analysis algorithms are then applied to the extracted features to produce the indicators. To evaluate the performance of our approach, we also compute each indicator for the ground-truth annotation labels and validate our results against these extracted metrics. The proposed shape analysis method is summarized in Figure 1. We detail the main components of our approach next.

Data description. Raw satellite images have inherent distortions that cause feature displacement and scaling errors, resulting in inaccurate direct measurement of distance, angles, areas, and positions. Such distortions are corrected in the process of orthorectification to create accurately georeferenced images while preserving the distances between geographical features [17]. Planimetric data are created from orthorectified aerial images due to their high accuracy in representing Earth's surface. Hence they are suitable choices for creating annotation masks. The majority of the NYC planimetric data is manually digitized [10]. To create the training set, we obtained 15,400 orthorectified tiles captured in 2018 from Manhattan and Brooklyn [11]. For each tile, an annotation mask was created from planimetric sidewalks, roads, and buildings shapefiles clipped to the geographical extent of the tile. The image and annotation data were split into training (60%), validation (20%), and test sets (20%).

Semantic segmentation. For the specific task at hand, we adopt the Hierarchical Multi-Scale Attention [16] with HRNet-OCR [18] backbone. HRNet connects high-to-low resolution convolutions via parallel and repeated multi-scale fusions to better preserve low-resolution representations alongside the high-resolution ones compared to previous works [19] and has shown superior performance across

segmentation benchmarks. The network is then trained for 200 epochs (batch size of 16); Rectified Adam as optimizer with polynomial learning rate policy.

Sidewalk metrics. Using our trained model, we made predictions on the entire Manhattan's unlabeled data, which amounts to roughly 20000 tiles. The predictions are then used to estimate the sidewalks' width, angle, and curvature by employing geometry and image processing techniques [2]. We use the *skeletonization* technique to measure the attributes of interest. In colloquial terms, the skeleton of a binary shape is a thin line equidistant to the shape's borders. Here, the skeleton is obtained by a systematic sequence of morphological *thinning* of the shape. For each point in the skeleton, we define the *width* of the shape at a given point in the skeleton as twice the distance to the borders of the shape. The sidewalk *angle* is estimated by its slope at a given point of the skeleton. A finite-difference approach is used to estimate this measurement which relies on the determination of the orientation of the line connecting the query point to a neighboring one in the skeleton. The proposed approach is heavily dependent on the parameter h: the distance between the point of interest and its neighbor in the finite-difference operation, which defines the scale of the measurement. The sidewalk *curvature* estimation considers the osculating circle passing through the query point and its two neighboring points in the skeleton [4]. The curvature of a shape at the query point is defined as the inverse of the radius of the obtained circle.

3 Experimental results

Figure 2 illustrates the segmentation and shape analysis results, and as can be seen, the model detected sidewalks and footpaths inside parks which are displayed in red. The model exhibits powerful detection capabilities in occluded areas since the annotation masks were based on the NYC planimetric data where sidewalks are mapped as continuous features even when occluded, given they are visible on both sides of the occluding element [10]. This property helped train the model to predict sidewalks correctly even when parts of them were occluded.

Table 1: Evaluation metrics.

Class	IoU	Precision	Recall
Building	0.83	0.92	0.90
Road	0.83	0.94	0.88
Sidewalk	0.79	0.91	0.86
Background	0.86	0.92	0.93
mIoU		0.83	

Table 2: Measurements.

Feature	Bin		N	RMSE
Width (pixels)	0	7	1291	3.12
	7	14	11045	1.85
	14	$+\infty$	529	6.07
Angle (degrees)	0	45	539	0.84
	45	90	6599	0.21
	90	135	4167	0.34
	135	180	1560	0.66
Curvature (pixels)	0.0	0.1	4069	0.09
	0.1	0.2	2926	0.10
	0.2	0.3	2688	0.10
	0.3	$+\infty$	3182	0.21

Table 1 shows the performance metrics of our semantic segmentation model on the held-out test set. Average and class-wise Intersection over Union (IoU) and precision and recall are presented. Sidewalks were detected with 79% IoU, and the average IoU (mIoU) across all four classes is 83%. Good results were attained in the identification not only of sidewalks but of all considered classes.

To further assess the sidewalks' condition, we measure three different average metrics: width, angle, and curvature as described in Section 2.

The average of each measurement is calculated across all the values of a given image and is used as a proxy for the overall characterization of the sidewalks visible in each image. The average width, angle, and curvature are computed independently for the ground truth and prediction masks. The values are partitioned into bins, and the root mean squared error is calculated for each measurement's bin (Table 2). The RMSE values vary significantly across the bins: wider sidewalks, sidewalks with angles between 0-45°, and the highly curved ones exhibit higher errors than the rest.

We also analyzed the relationship between the sidewalk attributes and land use at the tile level, calculated by doing a spatial join between the MapPLUTO data [9] and the tile extents. Table 3 shows the three sidewalk measurements aggregated by land use. Commercial areas have the highest width, in line with the NYC design guides. As expected, parks exhibit the highest angle and curvature due to various bending pathways. Residential areas have the second lowest curvature suitable for the navigation of people with different mobility levels.

Table 3: Aggregation by LU.

Land use (LU)	# Img.	Width	Angle	Curv.
Residential	4530	8.91	85.07	0.175
Commercial	1806	9.25	88.45	0.198
Industrial	75	7.83	83.9	0.156
Public facilities	934	8.54	91.9	0.20
Parks	1904	8.56	99.2	0.27

4   Conclusion

We present a method for the semantic segmentation of sidewalks from orthorectified images of NYC with highly dense urban areas and prevalent cases of sidewalk occlusion. The method shows promising detection power, achieving 83% mIoU over four classes with sidewalks having 79% IoU and 91% precision. Since the ground truth mapped sidewalks regardless of the shadow and occlusions, the evaluation metrics on the held-out test sets of 3000 images indicate that the model performed well in predicting sidewalks under different conditions, including occlusion cases. The trained model was used to extract the sidewalks, roads, and buildings for the whole of Manhattan. Morphological and discrete geometry operations were then applied to calculate the sidewalks' width, angle, and curvature for the ground-truth and model's predictions. The error rate for each measurement was evaluated per range of values. Finally, the sidewalk metrics were aggregated per land use.

The results show promising potentials of the method, offering several exciting opportunities for future work. We plan to aggregate the measurements on the sidewalk segment level to provide a more informative overview of the sidewalks' condition. We also plan to create a more generalizable model by expanding the training data to include cities with varying topological characteristics and shadow patterns [6].

References

1. Beale, L., Field, K., Briggs, D., Picton, P., Matthews, H.: Mapping for wheelchair users: Route navigation in urban spaces. The Cartographic Journal **43**(1), 68–81 (2006)
2. Costa, L.F., Cesar-Jr., R.: Shape Classification and Analysis: Theory and Practice. Taylor & Francis (2009)
3. Deitz, S., Lobben, A., Alferez, A.: Squeaky wheels: Missing data, disability, and power in the smart city. Big Data & Society **8**(2) (2021)
4. Flynn, P.J., Jain, A.K.: On reliable curvature estimation. In: Proc. of the IEEE Conf. on Comp. Vision and Pattern Recognition (1989)
5. Luo, J., Wu, G., Wei, Z., Boriboonsomsin, K., Barth, M.: Developing an aerial-image-based approach for creating digital sidewalk inventories. Transportation research record **2673**(8), 499–507 (2019)
6. Miranda, F., Doraiswamy, H., Lage, M., Wilson, L., Hsieh, M., Silva, C.T.: Shadow accrual maps: Efficient accumulation of city-scale shadows over time. IEEE Transactions on Visualization and Computer Graphics **25**(3), 1559–1574 (2019)
7. Miranda, F., Hosseini, M., Lage, M., Doraiswamy, H., Dove, G., Silva, C.T.: Urban Mosaic: Visual exploration of streetscapes using large-scale image data. In: Proc. of the 2020 CHI Conference on Human Factors in Computing Systems (2020)
8. Ning, H., Ye, X., Chen, Z., Liu, T., Cao, T.: Sidewalk extraction using aerial and street view images. Environment and Planning B: Urban Analytics and City Science (2021)
9. NYC DCP: PLUTO and MapPLUTO. Available: https://www1.nyc.gov/site/planning/data-maps/open-data/dwn-pluto-mappluto.page
10. NYC DoITT: New York City Planimetrics Data. Available: https://github.com/CityOfNewYork/nyc-planimetrics
11. NYC GIS: NYS Statewide Digital Orthoimagery Program. Available: https://gis.ny.gov/gateway/orthoprogram/index.cfm
12. Osama, A., Sayed, T.: Evaluating the impact of connectivity, continuity, and topography of sidewalk network on pedestrian safety. Accident Analysis & Prevention **107**, 117–125 (2017)
13. Rundle, A.G., Bader, M.D., Richards, C.A., Neckerman, K.M., Teitler, J.O.: Using Google Street View to audit neighborhood environments. American Journal of Preventive Medicine **40**(1), 94–100 (2011)
14. Saha, M., Saugstad, M., Maddali, H.T., Zeng, A., Holland, R., Bower, S., Dash, A., Chen, S., Li, A., Hara, K., Froehlich, J.: Project sidewalk: A web-based crowdsourcing tool for collecting sidewalk accessibility data at scale. In: Proc. of the 2019 CHI Conference on Human Factors in Computing Systems (2019)
15. Senlet, T., Elgammal, A.: Segmentation of occluded sidewalks in satellite images. In: Proc. of the 21st Int. Conf. on Pattern Recognition. IEEE (2012)
16. Tao, A., Sapra, K., Catanzaro, B.: Hierarchical multi-scale attention for semantic segmentation. arXiv preprint arXiv:2005.10821 (2020)
17. Tucker, C.J., Grant, D.M., Dykstra, J.D.: Nasa's global orthorectified landsat data set. Photogrammetric Engineering & Remote Sensing **70**(3), 313–322 (2004)
18. Wang, J., Sun, K., Cheng, T., Jiang, B., Deng, C., Zhao, Y., Liu, D., Mu, Y., Tan, M., Wang, X., Liu, W., Xiao, B.: Deep high-resolution representation learning for visual recognition. IEEE Trans. on Pattern Analysis & Machine Intelligence (2020)
19. Yu, F., Wang, D., Shelhamer, E., Darrell, T.: Deep layer aggregation. In: Proc. of the IEEE Conf. on Comp. Vision and Pattern Recognition (2018)