International Cartographic Association
Association Cartographique Internationale

Waterloo Urban Scene Dataset: An Annotation-Efficient Dataset for Urban Scene Classification with Minimal Supervision

Hongjie He [a], Yiqing Wu [a], Xuanchen (Willow) Liu [a], Wenyi Shen[a], Shene Abdalla[b], Jianing Xu[a], Mingming Guo[a], Jonathan Li[a,b*]

[a] *Department of Geography and Environmental Management, University of Waterloo, Waterloo, ON N2L 3G1, Canada*
[b] *Department of System Design Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada*
Hongjie He-hongjie.he@uwaterloo.ca, Yiqing Wu-y537wu@uwaterloo.ca, Xuanchen (Willow) Liu-x646liu@uwaterloo.ca, Wenyi Shen-w39shen@uwaterloo.ca, Shene Abdalla-shene.abba@geomount.ca, Jianing Xu-j44xu@uwaterloo.ca, Mingming Guo-m64guo@uwaterloo.ca, Jonathan Li-junli@uwaterloo.ca.
* Corresponding author: Jonathan Li-junli@uwateloo.ca

Keywords: Waterloo urban scene dataset, Weakly supervised learning, Semi-supervised learning, Annotation-efficient

Abstract: High-definition (HD) urban scene mapping is crucial for urban applications and autonomous driving. However, achieving high performance in HD mapping requires large amounts of high-quality annotated data. While the SkyScape dataset is valuable, it is limited by its focus on lane markings in Germany. In this paper, we present the Waterloo Urban Scene Dataset, built upon the Waterloo Building Dataset, designed for minimal supervision deep learning. The dataset includes 907 well-annotated, 775 roughly annotated, and 23,172 intact patches (512 × 512 pixels, 0.12m/pixel resolution in RGB). Combined with the SkyScape dataset, it supports HD urban scene mapping with minimal supervision. Future work will focus on benchmarking and method development.

1. Introduction

Urbanization is rapidly increasing, with the United Nations projecting that 66% of the world's population will live in cities by 2050 (Kaluarachchi, 2022). This growth presents significant challenges, as cities are expected to generate about 80% of greenhouse gases, contributing to climate change and impacting urban infrastructure sustainability. Urban infrastructure (UI) is a complex system that includes not only physical structures but also connects governance, economic development, and environmental management. It encompasses water systems, energy networks, transportation, sanitation, and the built environment (El-Diraby and Osman, 2011). The complexity of UI management spans multiple levels of government and affects various stakeholders. Given this complexity, accurate automated mapping of urban areas at the centimeter level is crucial for the effective construction and management of smart cities.

These high-definition (HD) maps are essential for: efficient city management and planning, responding to emergencies and natural disasters, supporting insurance companies and municipal departments and developing autonomous vehicles and advanced driver assistance systems (Azimi et al., 2019). The traditional approach to creating HD maps relies heavily on mobile mapping systems equipped with various sensors (e.g., Radar, Light Detection and Ranging (LiDAR), cameras). While effective, these systems face challenges such as limited field-of-view and significant resource demands for mapping large urban areas. Aerial datasets, although promising, suffer from limitations like narrow class focus, suboptimal sensor angles, and inaccuracies in automated annotations (Azimi et al., 2019). Therefore, several challenges remain in automated HD mapping in urban scene: (1) Data limitations: There's a need for more comprehensive, freely available training data. (2) Image resolution: Obtaining imagery detailed enough for centimeter-level mapping can be difficult. (3) Classification accuracy: Distinguishing between different urban features accurately is complex, especially where they intersect or overlap. (4) Object size variability: The wide range of sizes in urban objects, from small road signs to large buildings, complicates consistent mapping. This paper focuses on addressing data limitations. The new dataset we propose aims to enhance the accuracy and applicability of HD urban scene mapping. It will also serve as a foundational resource for tackling the other three challenges in future research.

2. Data and Method

2.1 Data

We build our Waterloo Urban Scene Dataset upon our previous work, the Waterloo Building Dataset (He et al., 2021, He et al., 2022), by expanding it to include a broader range of urban scene features. The Waterloo Building Dataset was originally designed for building rooftop extraction and consists of 242 images covering the Kitchener-Waterloo area in Ontario, Canada. Each image in this dataset has a size of 8350 × 8350 pixels in the Red, Green, and Blue bands, with a spatial resolution of 0.12m/pixel. To create the Waterloo Urban Scene Dataset, we focus specifically on urban areas and select 86 images from the Waterloo Building Dataset.

https://doi.org/10.5194/ica-abs-10-104-2025

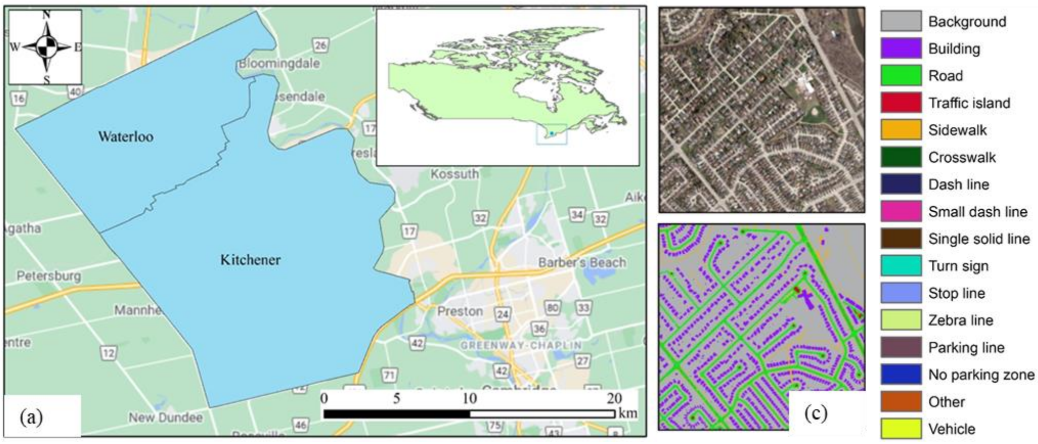

Figure 1. The dataset coverage area (a), An image sample (b), and annotations (c) (He et al., 2022)

2.2 Method

Considering that constructing a high-quality urban scene dataset is labor-intensive and time-consuming in practical situations, we aimed to minimize the supervision required by incorporating weakly supervised learning and semi-supervised learning. As a result, for the 86 images, we fully annotated 3 images, provided rough annotations for 10 images and left the remaining 73 images intact. This approach allows us to efficiently leverage minimal supervision while still ensuring the quality and diversity of the dataset.

For dataset construction, we organized an annotation team consisting of 7 master's students and 7 undergraduate research assistants. To ensure annotation accuracy, the process was conducted in multiple rounds. In the first round, all 14 students annotated the 86 collected images across 16 classes: Background, Building, Road, Traffic Island, Sidewalk, Crosswalk, Dash Line, Small Dash Line, Single Solid Line, Turn Sign, Stop Line, Zebra Line, Parking Line, No Parking Zone, Other, and Vehicle. In the second round, five master's students formed a quality control group to review the annotations, ensuring their accuracy and making necessary corrections. In the third round, a post-doctoral fellow joined the process to perform the final modifications and refinements.

Consequently, after splitting the large images into small patches of size 512 × 512, we obtained 907 well-annotated patches, 775 roughly annotated patches, and 23,172 intact patches. Combined with the SkyScape dataset, this small dataset can be effectively used for HD mapping with minimal supervision. We leave the benchmarking and method development for future research.

Acknowledgments

We acknowledge the contributions of Zarrin Tasneem, Jirui Hu, Tirtha Gajjar, Meixuan Li, Qianhui Xu, Leila Mirzaei, Yishan (Shirley) Li, and Chuhan (Han) Wang for their efforts in the first round of image annotation while they were working at the Geospatial Intelligence and Mapping Lab, affiliated with the University of Waterloo.

References

Azimi, S.M., Henry, C., Sommer, L. et al., 2019. Skyscapes fine-grained semantic understanding of aerial scenes. In Proc. IEEE/CVF ICCV, pp. 7393-7403.

El-Diraby, T.E. and Osman, H., 2011. A domain ontology for construction concepts in urban infrastructure products. Automat. Constr., 20(8), 1120-1132.

He, H., Jiang, Z., Tan, W., et al. (2021). Waterloo Building Dataset: A large-scale very-high-spatial-resolution image dataset for building rooftop extraction. Abstracts of the ICA, 3, 1-2.

He, H., Jiang, Z., Gao, K., et al. (2022). Waterloo building dataset: A city-scale vector building dataset for mapping building footprints using aerial orthoimagery. Geomatica, 75(3), 99-115.

Kaluarachchi, Y. (2022). Implementing data-driven smart city applications for future cities. Smart Cities, 5(2), 455-474.

https://doi.org/10.5194/ica-abs-10-104-2025