The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLIX-B1-2026
XXV ISPRS Congress 2026 “From Imagery to Understanding”, Commission I, 4–11 July 2026, Toronto, Canada

Urban-Graph: Bridging Local SLAM and Global Earth Observation for Fine-Grained Urban LCLU Mapping

Minghao Yu[1,2], Chenyang Wang[3], Youchen Tang[1], Xiao Wu[4], Jian Zhou[1,2*]

[1] Wuhan University, State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, 430070 Wuhan, China - (yuminghao, youchentang, jianzhou)@whu.edu.cn
[2] Hubei Luojia Laboratory, Wuhan University, Wuhan, China
[3] State Key Laboratory of Marine Thermal Energy and Power, Wuhan Second Ship Design and Research Institute, 430070 Wuhan, China - wang_chenyang@outlook.com
[4] Wuhan University of Science and Technology, Key Laboratory of Metallurgical Equipment and Control Technology, Ministry of Education, 430081 Wuhan, China - wuxiao@wust.edu.cn

Keywords: Land-cover/Land-use, Earth Observation, SLAM, Scene Graph, Semantic Mapping

Abstract

Urban scene understanding requires both global geographic context and local structural detail. Earth Observation (EO) imagery supports large-scale land-cover and land-use (LCLU) mapping, but in urban areas it often merges heterogeneous surfaces into broad built-up classes. Vehicle-based sensors such as LiDAR and cameras recover these local structures, but their maps can drift and often remain in a local coordinate frame. We present urban graph, which combines overhead EO priors, vehicle observations, and fixed roadside anchors in a hierarchical semantic scene graph. Coarse georeferenced regions from EO data are updated with local observations, while a factor graph jointly optimises SLAM constraints and global geodetic constraints. The resulting graph is projected back to the overhead layer to separate coarse urban classes into finer semantic components. Experiments in CARLA show improved global alignment, reduced drift, and more detailed projection of local semantics into EO space.

1. Introduction

Earth Observation satellites provide consistent observations for land-cover and land-use mapping at regional and global scales. Their synoptic coverage makes them indispensable for urban monitoring, environmental assessment, hydrological modelling, and long-term change analysis. This advantage weakens in complex cities, where coarse EO classes still aggregate multiple materials and functions within the same pixel. Roads, roofs, paved open spaces, bare ground, and tree canopy are frequently merged into one built-up class. That aggregation reduces semantic specificity and weakens downstream interpretation in studies of urban ecology, runoff generation, thermal environment, and green infrastructure. The validation problem is equally severe. Many medium-resolution and even part of the high-resolution LCLU products still rely on sparse reference samples, manual interpretation, or small local surveys. These protocols provide only partial coverage and limited temporal continuity. Urban environments intensify the problem because built-up pixels are internally mixed and structurally complex. A single correctness label often conceals substantial variation in material composition, vegetation cover, and human-made structure. This mismatch between coarse EO labels and fine urban reality constrains both product assessment and algorithm development. Local sensing offers complementary information. Vehicle-mounted LiDAR, cameras, and inertial sensors recover urban scenes at much finer spatial detail. Modern SLAM pipelines reconstruct trajectories, local geometry, semantic surfaces, and object instances in dense and structured form. Scene-graph-based methods further encode topology, containment, and spatial relations among roads, facades, trees, poles, and other urban entities. These outputs contain exactly the information that coarse EO classes fail to preserve. Figure 1 illustrates this type of semantic scene-graph representation in the simulator.

Figure 1. Example semantic scene graph produced by urban graph in the simulator.

The difficulty lies in scale and georeferencing. Local maps accumulate drift over long trajectories and often remain tied to an internal coordinate frame. Even when local accuracy is high, the resulting maps may not align well with global geographic reference systems. This prevents direct use of local maps as reliable validation layers for EO products. A stable mechanism is therefore needed to link local detail to a global frame and to coarse overhead semantics.

To tackle this problem, we use urban graph, a graph-based fusion method. It combines three data sources: overhead EO priors, local vehicle observations, and fixed roadside anchors with known global positions. The central representation is a hierarchical semantic scene graph. EO data provides coarse georeferenced regions. Vehicle observations add local geometry, objects, and semantic evidence. Roadside infrastructure anchors the local map to the global reference system. A backend factor

* Corresponding author.

https://doi.org/10.5194/isprs-archives-XLIX-B1-2026-615-2026 | © Author(s) 2026. CC BY 4.0 License.

graph then optimises these components jointly and outputs a globally aligned semantic graph.

The method serves two linked objectives. For local autonomous systems, it produces a semantic map in a stable global frame with less drift. For remote sensing analysis, it provides a reference layer that reveals the internal composition of coarse urban classes. A built-up region can therefore be described with finer semantic components, including asphalt, concrete, facades, roofs, and tree canopy. The main contributions of this paper are summarised as follows:

1. A graph-based method that combines EO priors, SLAM-derived observations, and roadside anchors in a hierarchical semantic scene graph.
2. A geospatial alignment strategy based on a backend factor graph that reduces local drift and preserves spatial relations between local observations and overhead regions.
3. A projection procedure that maps the optimised graph back into EO space for urban class decomposition, evaluated in CARLA.

2. Related Work

2.1 EO-Based Urban LCLU Mapping and Validation

Large-scale LCLU products remain central to urban environmental analysis (Brown et al., 2022). Urban categories, however, are still difficult to classify and validate. Recent learning-based methods, including approaches built on foundation models, have improved medium-resolution EO mapping (Cong et al., 2022). Auxiliary data integration has also improved land-cover classification in heterogeneous landscapes (Shandu et al., 2025). Even so, many high-resolution products still preserve only a limited set of urban surface classes (Li et al., 2025). The built-up label often merges roads, roofs, industrial surfaces, residential paving, and fragmented vegetation into one class. This coarse representation makes complex urban morphology harder to interpret (Milovanović et al., 2024). It also limits downstream geospatial analysis, including green-infrastructure thermal studies (Shu et al., 2025) and urban microclimate analysis (Jamal et al., 2025).

Validation strategies are grounded in rigorous sampling theory (Olofsson et al., 2014), yet city-scale reference generation still depends heavily on manual work. This creates a major operational bottleneck, and recent global validation protocols report the same limitation (Tyukavina et al., 2025). Sparse point samples and manually labelled patches remain common, but they are poorly matched to rapidly changing urban environments. They also do not resolve the mixed-pixel problem (Shandu et al., 2025). A built-up pixel may be judged correct during validation while still hiding substantial variation in its internal structure. Recent studies therefore call for large-scale semantic reference data with finer urban detail (Lesiv et al., 2025). Our urban graph method addresses this need by generating geographically anchored local evidence that helps interpret coarse overhead classes.

2.2 SLAM, Semantic Mapping, and Scene Graphs

LiDAR-based SLAM systems have become standard tools for high-resolution urban reconstruction (Shan et al., 2020). More recently, the field has been moving from closed-set categorical mapping toward open-world semantic scene understanding (Miao et al., 2025). This shift has been supported by foundation models and by richer scene representations such as 3D Gaussian Splatting (Keetha et al., 2024). Dense representations preserve geometric detail, but higher-level spatial reasoning still requires structured abstractions.

Scene graphs provide this abstraction by storing semantic entities and their spatial relations explicitly. Early work introduced hierarchical scene-graph representations (Armeni et al., 2019). Later work extended this idea to dynamic scene graphs (Rosinol et al., 2020). Recent methods combine language-guided models with 3D scene graphs to support open-vocabulary mapping, where categories are not limited to a fixed label set (Zemskova and Yudin, 2025). Related work has also explored hierarchical open-vocabulary scene graphs for language-grounded robot navigation (Werby et al., 2024). These representations capture multi-scale relations, such as how roads contain lanes or how vegetation regions decompose into tree instances (Zhang et al., 2025). Most of these pipelines remain centered on the robot and are mainly evaluated on indoor benchmarks (Günther et al., 2026). Global georeferencing is often treated as a secondary issue. Urban graph uses the scene graph as the main fusion representation and treats geospatial consistency as a core requirement for EO mapping.

2.3 Global-Local Fusion and Geospatial Anchoring

Global-local fusion has long been used to stabilise local vehicle trajectories. Cross-view geo-localisation has evolved from image retrieval to bird’s-eye-view matching methods (Ye et al., 2024). Geometric disentanglement has further improved cross-view geolocalization (Zhang et al., 2024). Recent work has also explored resource-efficient drone geo-localization (Sun et al., 2025). In road environments, persistent roadside infrastructure and refined HD map elements provide practical anchor sources for global referencing (Guo et al., 2025). Recent work therefore uses landmarks such as poles and traffic lights for accurate autonomous-vehicle localisation (Dong et al., 2023).

The use of geospatial constraints to refine urban LCLU products remains limited. Existing cross-view and map-fusion studies usually focus on localisation accuracy or multimodal fusion for vehicle navigation. Urban graph extends this line of work to remote-sensing analysis. In our setting, global alignment registers the vehicle trajectory and supports projection of local semantic detail back into overhead space. This makes it possible to decompose coarse built-up regions and evaluate them more systematically.

3. Method

3.1 System Overview

As summarised in Figure 2, urban graph decomposes coarse EO pixels using a graph representation that spans multiple spatial scales. The system is defined as a five-tier hierarchical semantic scene graph

$$\mathcal{G} = (\mathcal{V}, \mathcal{E}), \tag{1}$$

where the vertices represent five physical and semantic layers of the urban environment, and the edges represent both within-layer topology and between-layer associations.

The system processes an EO prior map, LiDAR point clouds, RGB images, and IMU measurements. These observations

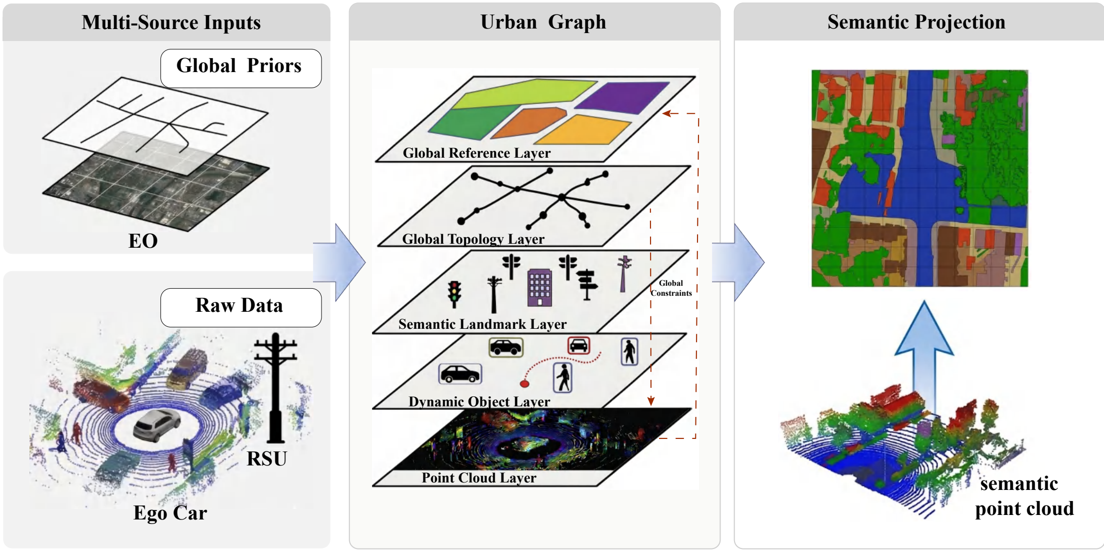


Figure 2. Overview of the urban graph pipeline.

update the graph incrementally. Static local observations are grouped into landmark nodes, and RSU landmarks with known global positions provide the absolute reference used to align the local graph with the EO layer. Dynamic objects are filtered before backend optimisation. The remaining static graph is jointly optimised and then projected back into the global frame for LCLU decomposition.

3.2 Five-Tier Semantic Scene Graph

The graph has five layers. At the coarsest scale, the Global EO Layer ($\mathcal{V}_{eo}$) provides the geographic reference frame. It initialises coarse semantic priors from degraded satellite imagery. A macro-node is defined as

$$v_{eo} = \{g_{eo}, \boldsymbol{p}_{eo}\} \in \mathcal{V}_{eo}, \tag{2}$$

where g_{eo} represents the 2D global polygon footprint and $\boldsymbol{p}_{eo}$ is the categorical prior distribution associated with a coarse built-up region.

The next layer is the Global Topology Layer ($\mathcal{V}_{topo}$), derived from high-resolution EO or OpenStreetMap data. It extracts a road topology in which nodes represent intersections and edges represent road segments. It also partitions continuous point-cloud observations into geographic chunks, which keeps city-scale updates manageable.

The Semantic Landmark Layer ($\mathcal{V}_{lmk}$) links the global and local frames. It stores stable 3D entities extracted from local perception, such as facades, traffic lights, and utility poles. A landmark node is parameterised as

$$v_{lmk} = \{\boldsymbol{b}_{lmk}, l_{lmk}, \boldsymbol{\Sigma}_{lmk}\} \in \mathcal{V}_{lmk}, \tag{3}$$

where $\boldsymbol{b}_{lmk}$ denotes the 3D bounding box, l_{lmk} is the semantic class label, and $\boldsymbol{\Sigma}_{lmk}$ captures pose uncertainty. Roadside Units (RSUs) appear in this layer as anchors that tie the local frame to the global coordinate system. These RSU-constrained landmarks provide the connection used to align local observations with EO polygons.

Moving entities are stored separately in the Dynamic Object Layer ($\mathcal{V}_{dyn}$). A dynamic node is written as

$$v_{dyn} = \{\boldsymbol{x}_t, \boldsymbol{v}_t\} \in \mathcal{V}_{dyn}, \tag{4}$$

where $\boldsymbol{x}_t$ and $\boldsymbol{v}_t$ describe the position and motion state of a moving object. This separation removes transient objects before backend optimisation.

The finest layer is the Point Cloud Layer ($\mathcal{V}_{pts}$), which stores dense geometry captured by the vehicle. A micro-node is defined as

$$v_{pts} = \{\boldsymbol{p}_{pts}, \boldsymbol{f}_{pts}, \boldsymbol{y}_{pts}\} \in \mathcal{V}_{pts}, \tag{5}$$

Here $\boldsymbol{p}_{pts}$ is the 3D coordinate, and $\boldsymbol{f}_{pts}$ denotes local geometric features such as surface normals. The term $\boldsymbol{y}_{pts}$ is the local semantic probability vector. Guided by topology chunks from $\mathcal{V}_{topo}$, this layer preserves local detail while feeding semantic refinements back to the EO layer.

3.3 Cross-Scale Graph Construction

The construction of $\mathcal{G}$ combines top-down management with bottom-up extraction. First, the global topology layer defines the spatial chunking structure. As the vehicle traverses the environment, the perception frontend identifies moving targets through bounding-box detection and assigns them to $\mathcal{V}_{dyn}$. These targets are removed from the backend graph, while the remaining static geometry populates $\mathcal{V}_{pts}$. Structural clustering then groups dense points into semantic landmarks in $\mathcal{V}_{lmk}$.

Using the landmark layer as the bridge between local and global frames, the system creates weighted associations between local static nodes and global EO polygons. For a local node i and a global polygon j, the association score is computed as

$$s_{ij} = \alpha\, \mathrm{IoU}(g_i^{proj}, g_j) + \beta\, \boldsymbol{y}_i^\top \boldsymbol{p}_j, \tag{6}$$

Here g_i^{proj} is the projected 2D footprint of the local node, and g_j is the candidate EO polygon. The vector $\boldsymbol{y}_i$ stores the semantic probabilities of the local observation.

https://doi.org/10.5194/isprs-archives-XLIX-B1-2026-615-2026 | © Author(s) 2026. CC BY 4.0 License.

The scores are then converted into normalised association weights with a softmax:

$$a_{ij} = \frac{\exp(s_{ij})}{\sum_k \exp(s_{ik})}. \tag{7}$$

This weighting models boundary uncertainty before geometric convergence and keeps multiple overhead candidates active during optimisation.

3.4 Joint Backend Factor Graph Optimisation

Graph construction accumulates local drift over long trajectories. To correct this, we use a joint backend optimisation that excludes the dynamic layer and aligns only the static layers. The objective is formulated as

$$\begin{aligned} \min_{\boldsymbol{X},\boldsymbol{Y},\boldsymbol{A}} \quad & \sum \|\boldsymbol{r}_{slam}\|^2_{\Sigma_s} \\ & + \lambda_a \sum \|\boldsymbol{r}_{rsu}\|^2_{\Sigma_a} \\ & + \lambda_h \sum \|\boldsymbol{r}_{hier}\|^2_{\Sigma_h} \\ & + \lambda_{sem} \sum \|\boldsymbol{r}_{sem}\|^2_{\Sigma_{sem}}. \end{aligned} \tag{8}$$

The state variables include vehicle poses $\boldsymbol{X}$, local map geometry $\boldsymbol{Y}$, and association weights $\boldsymbol{A}$. The residual $\boldsymbol{r}_{slam}$ enforces LiDAR-inertial odometry and loop-closure consistency within $\mathcal{V}_{pts}$. The anchor residual $\boldsymbol{r}_{rsu}$ binds RSU landmarks to the absolute geographic frame. The hierarchical residual $\boldsymbol{r}_{hier}$ penalises geometric disagreement between local instances and their associated EO polygons, weighted by a_{ij}. The semantic residual $\boldsymbol{r}_{sem}$ encourages semantic smoothness among adjacent point-cloud nodes. Together, $\boldsymbol{r}_{rsu}$ and $\boldsymbol{r}_{hier}$ align EO and local observations: RSUs anchor landmark geometry in the global frame, and the corrected landmarks then support association with EO polygons. After optimisation, local drift is reduced and the cross-scale associations are updated using corrected geometry.

4. Experiments

4.1 Experimental Setup

We evaluate urban graph in the CARLA simulator. The vehicle is equipped with a LiDAR sensor and an IMU, producing high-frequency local odometry. For each urban scene, we extract a coarse EO prior to instantiate the global layer. To mimic the semantic gap found in EO products, we intentionally merge fine structures into broad built-up polygons.

To test the value of global anchoring, we compare our system with LIO-SAM, a widely used LiDAR-inertial odometry framework (Shan et al., 2020). Two configurations are considered:

1. **Unconstrained LIO-SAM:** the standard pipeline relying only on sequential sensor measurements.
2. **urban graph:** a graph-based formulation that augments the local frontend with sparse global constraints from roadside anchors such as RSUs.

4.2 Global Localisation and Drift Suppression

The primary quantitative evaluation focuses on the translational consistency of global localisation. We therefore report two metrics: Absolute Trajectory Error (ATE), reported as RMSE, and Maximum Position Error.

Scene	Method	ATE RMSE (m)	Maximum Position Error (m)
Town02	LIO-SAM	1.553	2.858
Town02	urban graph	1.272	2.266
Town10	LIO-SAM	1.717	4.911
Town10	urban graph	1.177	2.291

Table 1. Quantitative localisation comparison in Town02 and Town10.

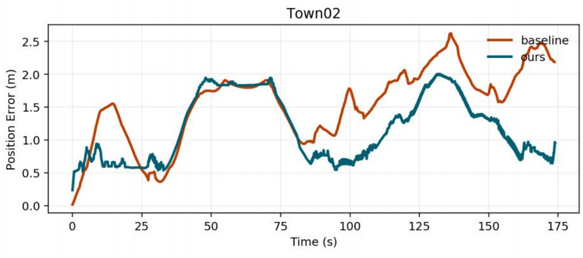
The unconstrained LIO-SAM baseline accumulates substantial error. Without global anchors, small angular errors in LiDAR registration and inertial integration grow into large translation errors over time. This effect is especially visible in long, open-ended trajectories, where the maximum position error becomes large.

With global anchors, the graph reduces cumulative drift by incorporating sparse absolute position updates in the backend optimisation. These anchors limit low-frequency trajectory wander and stabilise the pose estimate. Table 1 shows lower ATE RMSE and Maximum Position Error in both CARLA towns. Figure 3(a) shows the position-error curves in Town02. Figure 3(b) shows the corresponding curves in Town10.

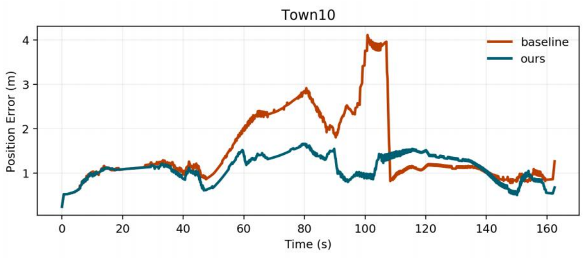
4.3 Fine-Grained Semantic Re-Projection

Figures 4(a) and 4(b) show the urban graph in Town02 and Town10, while Figures 4(c) and 4(d) show the corresponding fine-grained semantic re-projections. In both towns, urban graph projects local semantic observations back to the overhead frame and separates regions that are coarse in the EO prior into more specific components. Roads, vegetation, and roof-like structures appear as distinct regions within areas that were previously represented by a single built-up label.

These results show that the projected local map remains globally aligned while adding internal structure to coarse EO classes. The method improves localisation and also produces a more detailed overhead representation that is easier to interpret for urban LCLU analysis.

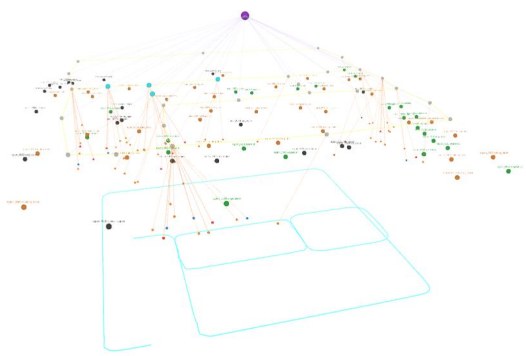


(a) Town02 position error comparison.

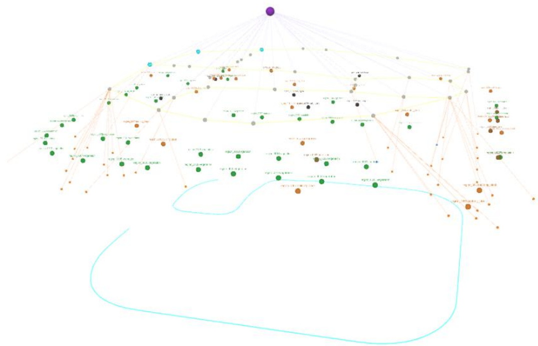


(b) Town10 position error comparison.

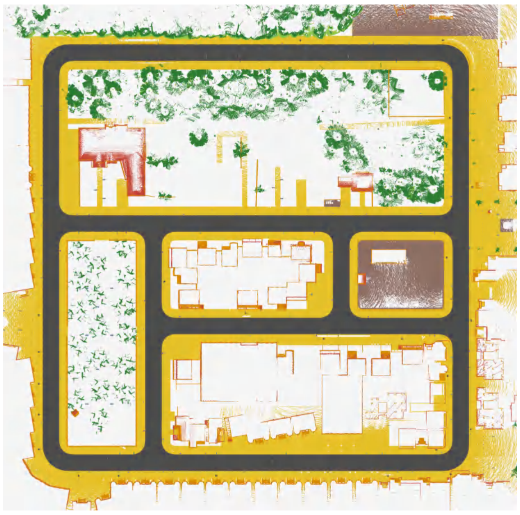
Figure 3. Position error over time for the unconstrained baseline and urban graph in two CARLA towns.



(a) Urban graph in Town02.



(b) Urban graph in Town10.



(c) Town02 fine-grained semantic re-projection.

(d) Town10 fine-grained semantic re-projection.

Figure 4. Top row: urban graph in Town02 and Town10. Bottom row: corresponding fine-grained semantic structures re-projected from the drift-corrected local map into the overhead frame.

4.4 Experimental Analysis

Figure 3 and Table 1 provide a consistent picture of drift reduction. In both towns, the LIO-SAM baseline shows larger error values and higher peaks, whereas urban graph maintains lower error levels over time. The difference is especially clear in Town10, where the gap in maximum position error is the largest.

This more stable global alignment also supports the projection results in Figure 4. If the trajectory drifts, locally observed structures are projected into the wrong overhead regions and the semantic map becomes inconsistent. By reducing drift, urban graph maps roads, vegetation, and roofs back to the overhead frame more reliably.

The urban graph views in Figure 4(a) and Figure 4(b) also help explain the different error patterns across the two towns. Town10 contains longer connected corridors and larger open regions, so the unconstrained baseline accumulates drift over a broader spatial extent. After global correction, the projected semantics remain consistent with the surrounding road layout and nearby urban blocks in both scenes. This indicates that the method preserves network-level spatial organisation while refining local semantic structure.

5. Conclusion

This paper presents a method that combines global EO data with local SLAM observations through fixed roadside anchors. The method builds a city-scale semantic scene graph that serves two purposes. For autonomous systems, it provides a map in a global frame with less drift. For remote sensing, it provides a detailed reference layer that can support validation and help interpret coarse urban classes. It can also support future EO products by providing sub-pixel composition statistics for urban regions.

Future work will extend the method in several directions. One direction is to use crowd-sourced vehicle data so that the graph can be updated more frequently. This could support LCLU change detection when EO monitoring misses subtle changes or responds too slowly. We also plan to expand the semantic classes represented in the graph. This would move the method from land-cover labelling toward broader land-use analysis using signals such as traffic flow and time of day.

Acknowledgements

This research was supported by the National Natural Science Foundation of China (42471480) and the Key Research and Development Projects in Hubei Province (2024BAB078).

References

Armeni, I., He, Z.-Y., Gwak, J., Zamir, A. R., Fischer, M., Malik, J., Savarese, S., 2019. 3d scene graph: A structure for unified semantics, 3d space, and camera. *Proceedings of the IEEE/CVF international conference on computer vision*, 5664–5673.

Brown, C. F., Brumby, S. P., Guzder-Williams, B., Birch, T., Hyde, S. B., Mazzariello, J., Czerwinski, W., Pasquarella, V. J., Haertel, R., Ilyushchenko, S. et al., 2022. Dynamic World, Near real-time global 10 m land use land cover mapping. *Scientific data*, 9(1), 251.

Cong, Y., Khanna, S., Meng, C., Liu, P., Rozi, E., He, Y., Burke, M., Lobell, D., Ermon, S., 2022. Satmae: Pre-training transformers for temporal and multi-spectral satellite imagery. *Advances in Neural Information Processing Systems*, 35, 197–211.

Dong, H., Chen, X., Särkkä, S., Stachniss, C., 2023. Online pole segmentation on range images for long-term lidar localization in urban environments. *Robotics and Autonomous Systems*, 159, 104283.

Günther, M., Igelbrink, F., Lima, O., Niecksch, L., Renz, M., Atzmueller, M., 2026. A Scene Graph Backed Approach to Open Set Semantic Mapping. *arXiv preprint arXiv:2602.03781*.

Guo, Y., Zhou, J., Dong, Q., Li, B., Xiao, J., Li, Z., 2025. Refined high-definition map model for roadside rest area. *Transportation Research Part A: Policy and Practice*, 195, 104463.

Jamal, Y., Murdock, C. C., Baharia, R. K., Sharma, R., Vaishnav, K., Desai, V., Kohli, V., Mohanty, A. K., Pascual, M., Sharma, S. et al., 2025. Spatiotemporal patterns of urban heat in indoor and outdoor microclimates. *Environmental Research Letters*, 20(11), 114050.

Keetha, N., Karhade, J., Jatavallabhula, K. M., Yang, G., Scherer, S., Ramanan, D., Luiten, J., 2024. Splatam: Splat track & map 3d gaussians for dense rgb-d slam. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 21357–21366.

Lesiv, M., Fritz, S., Dürauer, M., Georgieva, I., Buchhorn, M., Bertels, L., Tsendbazar, N., Van De Kerchove, R., Zanaga, D., Schepaschenko, D. et al., 2025. A global reference data set for land cover mapping at 10 m resolution. *Earth System Science Data Discussions*, 2025, 1–9.

Li, Z., Wang, H., Wang, Y., Chen, G., Chen, J. M., Chen, B., 2025. Large-Scale High-Resolution Essential Urban Land Cover Category Mapping Using a Semantic-Augmented and Noise-Tolerant Approach. *IEEE Transactions on Geoscience and Remote Sensing*.

Miao, L., Liu, W., Deng, Z., 2025. A Frontier Review of Semantic SLAM Technologies Applied to the Open World. *Sensors*, 25(16), 4994.

Milovanović, A., Cvetković, N., Šošević, U., Janković, S., Pešić, M., 2024. Synergies between land use/land cover mapping and urban morphology: a review of advances and methodologies. *Land*, 13(12), 2205.

Olofsson, P., Foody, G. M., Herold, M., Stehman, S. V., Woodcock, C. E., Wulder, M. A., 2014. Good practices for estimating area and assessing accuracy of land change. *Remote sensing of Environment*, 148, 42–57.

Rosinol, A., Gupta, A., Abate, M., Shi, J., Carlone, L., 2020. 3D dynamic scene graphs: Actionable spatial perception with places, objects, and humans. *arXiv preprint arXiv:2002.06289*.

Shan, T., Englot, B., Meyers, D., Wang, W., Ratti, C., Rus, D., 2020. Lio-sam: Tightly-coupled lidar inertial odometry via smoothing and mapping. *2020 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, IEEE, 5135–5142.

Shandu, I. D., Xulu, S., Gebreslasie, M., 2025. Enhancing land cover classification in the heterogeneous landscape by integrating auxiliary data with Sentinel-2 imagery using the random forest algorithm. *Frontiers in Remote Sensing*, 6, 1697897.

Shu, B., Chang, M.-E., Chang, H.-T., Ou, J.-H., Hsiung, T.-J., 2025. Research on thermal environment simulation improvement of green infrastructure at urban block scale. *All Earth*, 37(1), 1–21.

Sun, J., Liu, K., Zhang, C., Chen, C., Shen, J., Chen, C., Vong, C.-M., 2025. MobileGeo: Exploring Hierarchical Knowledge Distillation for Resource-Efficient Cross-view Drone Geo-Localization. *arXiv preprint arXiv:2510.22582*.

Tyukavina, A., Stehman, S. V., Foody, G. M., Bontemps, S., See, L., Olofsson, P., Tsendbazar, N., Radoux, J., Komarova, A., Serre, B. M. et al., 2025. Land Cover and Change Map Accuracy Assessment and Area Estimation Good Practices Protocol.

Werby, A., Huang, C., Büchner, M., Valada, A., Burgard, W., 2024. Hierarchical open-vocabulary 3d scene graphs for language-grounded robot navigation. *First Workshop on Vision-Language Models for Navigation and Manipulation at ICRA 2024*.

Ye, J., Lv, Z., Li, W., Yu, J., Yang, H., Zhong, H., He, C., 2024. Cross-view image geo-localization with panorama-bev co-retrieval network. *European Conference on Computer Vision*, Springer, 74–90.

Zemskova, T., Yudin, D., 2025. 3dgraphllm: Combining semantic graphs and large language models for 3d scene understanding. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 8885–8895.

Zhang, C., Delitzas, A., Wang, F., Zhang, R., Ji, X., Pollefeys, M., Engelmann, F., 2025. Open-vocabulary functional 3d scene graphs for real-world indoor spaces. *Proceedings of the Computer Vision and Pattern Recognition Conference*, 19401–19413.

Zhang, X., Li, X., Sultani, W., Chen, C., Wshah, S., 2024. GeoDTR+: Toward generic cross-view geolocalization via geometric disentanglement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(12), 10419–10433.