

SignLoc: Robust Localization using Navigation Signs and Public Maps

Nicky Zimmerman¹, Joel Loo¹, Ayush Agrawal² and David Hsu¹



Fig. 1: Global localization using **SignLoc** with a Boston Dynamics Spot robot. The robot detects navigational signs and extracts directional and locational cues. To localize, it matches these cues to a large-scale, indoor-outdoor navigation graph, constructed from publicly available maps.

Abstract—Navigation signs and maps, such as floor plans and street maps, are widely available and serve as ubiquitous aids for way-finding in human environments. Yet, they are rarely used by robot systems. This paper presents *SignLoc*, a global localization method that leverages navigation signs to localize the robot on publicly available maps—specifically floor plans and OpenStreetMap (OSM) graphs—without prior sensor-based mapping. *SignLoc* first extracts a navigation graph from the input map. It then employs a probabilistic observation model to match directional and locational cues from the detected signs to the graph, enabling robust topo-semantic localization within a Monte Carlo framework. We evaluated *SignLoc* in diverse large-scale environments: part of a university campus, a shopping mall, and a hospital complex. Experimental results show that *SignLoc* reliably localizes the robot after observing only one to two signs.

Index Terms—Localization, Semantic Scene Understanding.

I. INTRODUCTION

LOCALIZING and navigating in the open world remains a challenge for robots due to the diversity and complexity of human environments. Prior work enables robust localization, particularly for small-scale indoor environments, by exploiting geometric features [46] along with some semantic features like objects [15] or in-the-wild text [49]. However, these face challenges scaling up to large, heterogeneous

environments, particularly those spanning both indoors and outdoors. Existing methods also typically require in-situ sensor data to build representations for localization, necessitating an additional pre-deployment mapping step.

Humans have addressed these challenges by designing wayfinding aids such as *navigational signs* and *publicly available human-centric maps*—e.g., floor plans and street maps. These two aids are complementary, with both encoding symbolic locational and coarse directional information about manmade environments. Signs are informative features capturing global information about places and their relative locations. They can be directly matched to features in existing human-centric maps, enabling localization without pre-deployment mapping. Since they are ubiquitous both indoors and outdoors, and are purposefully placed in large-scale spaces to support human navigation, they are ideal semantic features for localizing in large-scale, heterogeneous scenes. However, both signs and human-centric maps remain under-utilized sources of information in robot systems.

We present **SignLoc** (Fig. 1) a global localization approach that leverages navigational signs and publicly available human-centric maps—namely floor plans for indoor spaces and OpenStreetMap (OSM) graphs for outdoor areas. Our approach rests on two key observations. First, human-centric maps encode key locations and the paths between them in a coarse but directionally accurate way, which can be extracted in the form of a *navigational graph*. Second, navigational signs often indicate the *direction of shortest travel* to a given location. This property is a distinctive feature for localization, as it lets us narrow down the set of possible locations for the sign by comparing the sign’s directions with the directions of shortest paths computed from all candidate locations in the

Manuscript received: Aug, 24, 2025; Revised Nov, 10, 2025; Accepted Nov, 26, 2025.

This paper was recommended for publication by Editor M. Vincze upon evaluation of the Associate Editor and Reviewers’ comments. This work was supported by ... and the DSO Graduate Fellowship.

¹Smart Systems Institute, National University of Singapore.

²Khoury College of Computer Sciences, Northeastern University.

Digital Object Identifier (DOI): see top of this page.

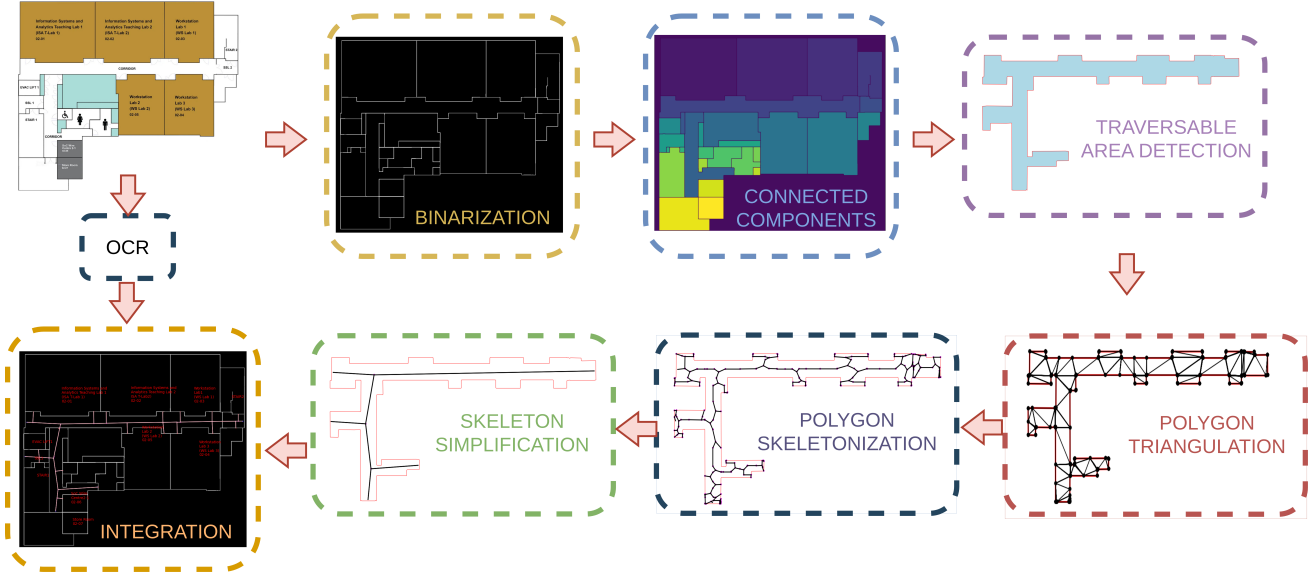


Fig. 2: A diagram of the map extraction pipeline, starting with a venue map or a floor plan, and extracting the navigational graph. After the venue map is binarized and the text and symbols are extracted, computational geometry methods are applied to extract the traversable nodes and the place nodes of the navigational graph. Then symbol and text information is integrated into the graph, connecting portal nodes and assigning labels to place nodes.

navigational graph.

Our main contribution is a global localization approach that uses navigational signs as spatial semantic features to localize with respect to a topo-semantic navigational graph extracted from floor plans and OSM graphs. We build on existing work in sign understanding [1], and propose a probabilistic observation model to match directional cues extracted from signs, with the navigational graph. This model is used for global localization with a particle filter. We also propose an approach that extracts and stitches together navigational graphs from floor plans and OSM graphs.

In sum, we make three key claims: Our approach is able to (i) Extract navigational information from signs; (ii) Utilize human-centric map priors in the form of floor plans and OSM graphs to establish a unified world representation; (iii) Localize robustly in large human-centric environments, with seamless transition between outdoor and indoor spaces. Experiments across diverse large-scale environments—ranging from a mall to a campus to a hospital—support these claims. In particular, we find that **SignLoc** reliably localizes the robot after observing only one to two signs.

II. RELATED WORK

Localization for mobile robots has been studied extensively [34], [35], and geometric approaches gained substantial popularity [11], [12], [19]. In the recent decade, there is a noticeable shift towards utilizing semantic cues [5], [28], [48] and considering new map representations [4], [29], [39].

Semantic cues. The need for reliable and informative, yet compact features drives the use of semantic cues in localization. Recent works explored various cues, from specific high-level features like lane markings [32], to broad semantic features like affordances [23], objects [48] and semantic part understanding of scenes [18]. Signs are navigational aids

commonplace in human-oriented spaces. While a large body of work exists on detecting and classifying signs, both in indoor [2], [38] and on-road settings [10], [14], [37], [47], the application of signs to localization remains limited. Cui et al. [8] and Zimmerman et al. [49] detect and associate textual cues on signs to locations in floor plans, while Escalera et al. [9] classify traffic signs for on-road localization. Liang et al. [20] goes further by parsing text and direction information from signs, using them to guide local mapless navigation.

We use navigational signs as semantic features for global localization in large-scale, heterogeneous scenes. Specifically, we propose an observation model based on the property that signs often reflect the direction of shortest travel to given locations, to localize on human-centric maps. Unlike Liang et al. [20]’s rule-based approach, we follow Agrawal et al. [1] and use VLMs for more open-ended sign understanding.

Publicly available, human-centric maps. Most robot localization approaches leverage dense maps, e.g. occupancy or voxel grids [25], built from sensor data taken during a pre-deployment mapping run. In contrast, humans localize and navigate effectively in unseen spaces by using abstract, semantic maps like floor plans, road maps etc., which are made publicly available to aid wayfinding. Recent works explore harnessing human-centric maps for localization and navigation, with many focusing on floor plans for indoor scenes, and OpenStreetMap (OSM) information for outdoor scenes. Several approaches seek to segment and parse floor plans [21], [33], [41], [45], but often fail to handle floor plans that deviate from the architectural standard, such as the visitor and venue maps or operate under the Manhattan world assumption. Xie et al. [40] proposed to generate a navigational graph from a floor plan to localize and guide the visually impaired, but their map extraction fails to generalize to different floor plans and is not suitable for global localization. Cheng et al. [6] propose a

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

localization approach centered around detecting the topology of a junction, but their hand-crafted classifier can only detect a closed set of topologies and they restrict their approach to “regular buildings”, a criteria that many public facilities architectures fail to meet. To localize in unseen outdoor scenes, Yan et al. [42] use building outlines as features, while Ruchti et al. [30] rely on road network topologies. Other visual place recognition approaches, [26], [36], [44] exploited Google Street View for image matching.

We use publicly available human-centric maps—floor plans and OSM graphs—as sources of prior map information to globally localize in unseen environments. We propose a method to extract and stitch a unified navigational graph from floor plans and OSM graphs, which can scale to large areas and span both indoor and outdoor scenes.

III. SIGNLOC

We use navigational signs to globally localize with respect to publicly available human-centric maps—floor plans and OpenStreetMap (OSM) graphs. We decompose this problem into three main steps: (i) *map extraction*, (ii) *sign parsing*, and (iii) *sign-centric localization*. Map extraction (Sec. III-A) extracts and stitches together a unified navigational graph \mathcal{G} from floor plans and OSM graphs. Sign parsing (Sec. III-B) extracts navigational cues—associated textual and directional features—from navigational signs. Sign-centric localization (Sec. III-C) describes our novel observation model for matching navigational cues to \mathcal{G} for global localization.

A. Navigational Graph Construction based on Priors

We propose an approach to extract and stitch together a single navigational graph \mathcal{G} from floor plans and OSM graphs. Navigational graph construction takes as input floor plans as 2D top-down RGB images, and OSM graphs. The output is the navigational graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, inspired by [43]. The graph nodes \mathcal{V} comprise three types: **(i) intersection nodes** where two or more distinct navigable paths meet (e.g. turn in corridor, T-junction etc.); **(ii) place nodes**, which correspond to named regions in the input maps and store the region’s textual label as an attribute; **(iii) portal nodes** that connect disjoint areas (e.g. doors, lifts, stairs). The graph edges \mathcal{E} represent connectivity between the nodes. The extracted graph \mathcal{G} need not be metrically accurate, but it should preserve coarse directional information. Specifically, **SignLoc** discretizes directions that edges $e \in \mathcal{E}$ can take into the set of 8 cardinal directions \mathcal{D} . If metrically accurate OSM maps are input, the \mathcal{G} should be metrically embedded in the coordinate frame of the OSM map.

We apply computational geometric techniques to obtain graphs from an indoor floor plan. If an OSM graph is present, it can serve as the “backbone” of the navigational graph, being used to coarsely embed subgraphs from individual floor plans into a consistent global coordinate frame. This yields a single navigational graph \mathcal{G} spanning from indoors to outdoors.

Graph extraction from floor plans. When floor plans or venue maps are available, the indoor navigational graph can be extracted automatically. While previous works [40] attempted to construct navigational graph from binarized maps

in the pixel-space, we work in polygon-space. We skeletonize the binary image and detect enclosed areas using connected components analysis, and we capture these geometries as polygons. The nodes of the extracted, simplified skeleton are intersection nodes, while the enclosed areas are place nodes. We detect traversable spaces, such as corridors and lobbies, by their degree of connectivity to other rooms. For these traversable areas, we use a polygonal skeletonization approach [17] to extract the centerline on which we should navigate. For the non-traversable spaces, such as enclosed rooms, they are assigned a place node and are associated with a name that is extracted from the map using Paddle-OCR.

Portal nodes such as doors, lifts and stairs are detected using their name or a symbol in the floor plan. The detections of symbols, as well as removing clutter, text and colors to achieve the binary image that captures the geometry, are performed by querying a VLM. As the results may be imperfect, annotating a navigational graph or editing an existing graph is also available through our GUI application NavGraphApp. The GUI allows users to construct navigational graphs when prior maps are unavailable, too cluttered, outdated or lacking important details.

Registration to global coordinate frame. The procedure detailed above is aimed at a single source map, which usually describes one floor of a building. To address the construction of multi-building navigational graphs given separate floor plans, we propose an alignment process based on OpenStreetMap [13]. As we store a simplified exterior polygon of each floor in our floor nodes, we can register it against the polygon provided by OpenStreetMap. We find the similarity transform that maximizes the overlapping area between the two polygons. Given polygon P_{fp} extracted from the floor plan and polygon P_{osm} , we optimize the similarity transform M to maximize the IoU between the two polygons,

$$M_{opt} = \underset{M}{\operatorname{argmax}} \operatorname{IoU}(A(P_{osm}), A(MP_{fp})) \quad (1)$$

where $A(\cdot)$ donates a function that computes the area of a polygon and the similarity transform is defined as $M = RS + T$. R, S and T refer to rotation, scaling and translation respectively.

For multiple building with multiple floor plans or venue maps, we register each against its OSM counterpart, and transform each navigational graph to a global coordinate system where they are aligned. When the navigational graphs are aligned with the OSM coordinate frame, utilizing the road network to extend the navigational graph becomes trivial. The process of aligning floor plans and venue maps against the OSM polygons essentially embeds the topological graphs in 3D space, enabling also topometric localization.

B. Sign Parsing

The sign parsing module first identifies navigational sign candidates in RGB images, then processes them with a VLM-based sign understanding system [1] to extract the embedded navigational cues.

Shortlisting candidate signs. We seek to identify candidate signs relevant to localization from RGB input. Specifically,

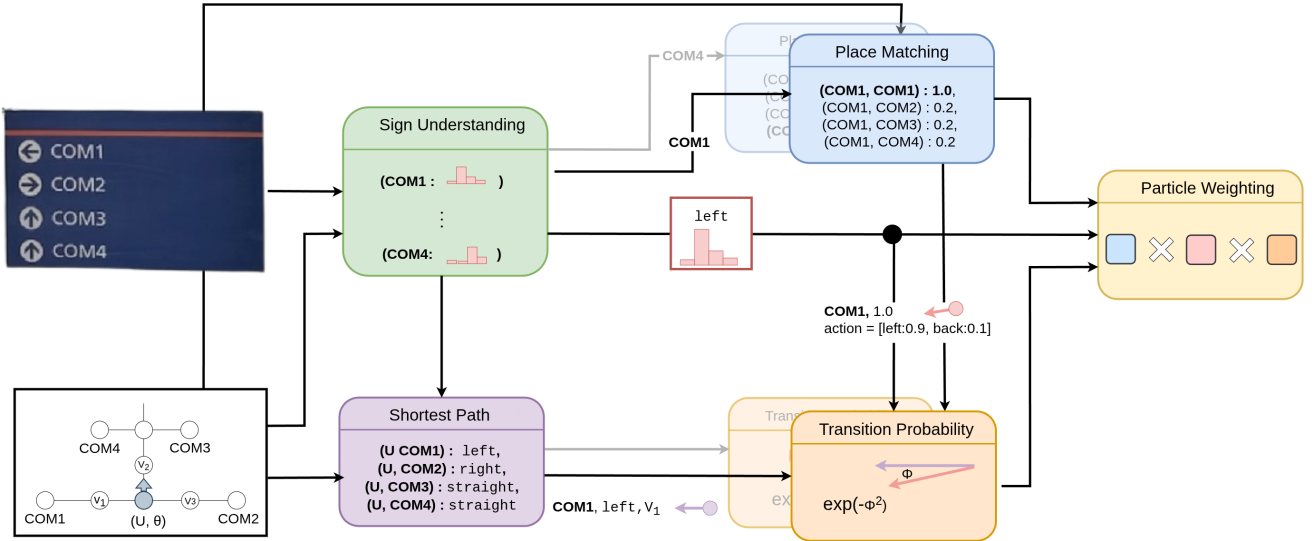


Fig. 3: An overview of the observation model. The input is the navigational graph and image of a directional sign. The observation model computes how likely it is to observe the sign for each particle with state (U, θ) , and reweights the particle accordingly.

we identify image inputs containing text strings of places that appear in prior maps. This is done using fast text spotting (e.g. PaddleOCR [7]) to extract observed text, and filter based on the list of place strings extracted from the prior map (Sec. III-A). Selected images are passed to an open-set object detector (GroundingDINO [22]) that is prompted to predict bounding boxes for “navigational signs” in the image. This produces candidate signs containing navigational cues useful for localization.

Sign understanding. We adapt the sign understanding pipeline of Agrawal et al. [1] to extract symbolic navigational cues from candidate signs. In their work, sign understanding involves (1) extracting locations and directional symbols from the sign, and (2) associating each location with a direction. To address (2), prior works often make assumptions not valid across signs in the wild (e.g. one-to-one location-arrow correspondences, fixed sign layouts) [1]. VLMs provide the semantic reasoning and common-sense to handle such variations. We extend their definition of navigational cues to tuples (l, D) , where l is the location label and D is a probability distribution over the set of direction categories \mathcal{D} , discretized into 8 cardinal directions. Each element of $D \in \mathbb{R}^8$ represents the likelihood that l corresponds to that direction. The pipeline sends the same parsing prompt to the VLM multiple times, using its responses to empirically construct the uncertainty distribution D . The output is the list of navigational cues present in the sign, $\mathcal{C} = \{(l^1, D^1), \dots, (l^J, D^J)\}$.

C. Sign-Centric Localization

The localization module globally localizes the robot with respect to the map obtained in Sec. III-A, using navigational cues from signs, \mathcal{C} , obtained in Sec. III-B. Our approach builds on the Monte Carlo localization framework [11], which estimates the robot’s state x_t in a given map m based on motion priors u_t and observations z_t . In this particle filter formulation, when motion priors are available, the particle’s

state is updated via the **motion model**. The **observation model** reweights the particle based on the likelihood of the observation given the particle’s state. During the **resampling** step, particles are sampled based on their weights, and particles with higher weight are more likely to be selected, while weak hypotheses perish. An overview of the localization can be seen in Fig. 4.

In our sign-centric localization, the map m takes the form of the navigational graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. We define the robot’s state $x_t = (v, \theta)$, where $v \in \mathcal{V}$ and θ is the robot’s heading. The observation z_t are the navigational cues \mathcal{C}_t , obtained from signs. Specifically, we design a particle filter for global localization, and describe the associated observation and motion models.

Observation model. Given observed navigational cues $z_t = \{z_t^1, \dots, z_t^J\} = \{(l^1, D^1)_t, \dots, (l^J, D^J)_t\}$, the observation model (Fig. 3) is:

$$p(z_t | x_t, m) = p(\mathcal{C}_t | x_t, \mathcal{G}) = \prod_{j=1}^J p(z_t^j | x_t, \mathcal{G})^{\frac{1}{J}} \quad (2)$$

Specifically, the observation model is the geometric mean of observation likelihoods for each $z_t^j = (l_t^j, D_t^j)$. Each likelihood models how likely it is for a location corresponding to l_t^j to be found in a given direction d , with respect to the robot’s current location at node $v_t \in \mathcal{V}$ with heading $\theta_t \in [-\pi, \pi)$:

$$\begin{aligned} p(z_t^j | x_t, \mathcal{G}) &= p((l_t^j, D_t^j) | (v_t, \theta_t), \mathcal{G}) \\ &= \sum_{u \in \mathcal{V}_{\text{top-}k}} p(l_t^j = \text{label}[u]) \sum_{d \in \mathcal{D}} p(\text{toward}(u) = d | (v_t, \theta_t), \mathcal{G}) \end{aligned} \quad (3)$$

Due to ambiguous place names or parsing errors, the extracted label l_t^j may not match exactly with any node labels, or may match with multiple node labels in \mathcal{G} . To handle this, we consider the top- k most similar node labels, i.e.

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

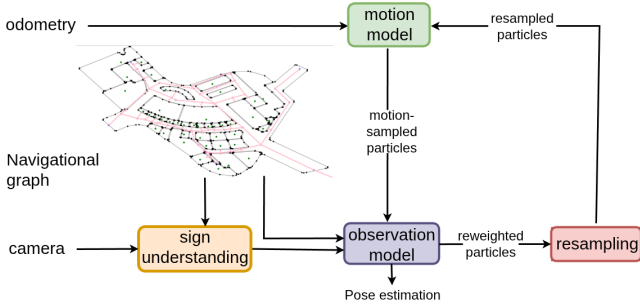


Fig. 4: An overview of our localization approach, based on our navigational graph and utilizing navigational cues from signs.

$\mathcal{V}_{\text{top-}k} = \underset{v \in \mathcal{V}}{\operatorname{argmax}_k} p(l_t^j = \text{label}[v])$, where $p(l_t^j = \text{label}[v])$ is a normalized fuzzy matching score based on Levenshtein distance [24].

We assume that signs indicate the *direction of shortest travel* to a goal. Thus, for each matched node $u \in \mathcal{V}_{\text{top-}k}$, we want to model the likelihood that moving in direction d takes us along the shortest path from v_t toward u . In **SignLoc**, the set of directional categories for edges, \mathcal{D} , is the 8 cardinal directions corresponding to the relative headings $\{0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}, \pi, -\frac{3\pi}{4}, -\frac{\pi}{2}, -\frac{\pi}{4}\}$ when the robot is in a pose facing the sign head-on. This likelihood is modelled as:

$$p(\text{toward}(u) = d \mid x_t, \mathcal{G}) = p(d_{\text{edge}} = d) \exp(-(d_{\text{edge}} \cdot d_{\text{act}})^2) \quad (4)$$

$d_{\text{edge}} \in \mathcal{D}$ denotes the normalized direction vector of the edge (v_t, w) in the world frame, where w is the next node along the shortest path from v_t toward u . The term d_{act} represents the relative heading d composed with the robot's current heading θ_t , capturing the robot's expected orientation after turning to face w . Intuitively, $(d_{\text{edge}} \cdot d_{\text{act}})^2$ measures directional consistency via cosine similarity, while the negative exponential term converts this into a valid probability. The weight $p(d_{\text{edge}} = d)$ represents the prior probability that the direction of shortest travel toward u corresponds to $d \in \mathcal{D}$, obtained from the sign-parsing distribution D_t^j .

Resampling. After the particles have been re-weighted, we resample the particles using a mixture of reciprocal sampling [27] and low-variance sampling [34], with an efficient sample size criteria [3] of $\frac{N}{2}$, where N is the number of particles. We assume sign detections happen in proximity to a node rather than an edge, therefore sampling new particles close to nodes. The nodes and orientations are sampled proportionately to weight of their particles, meaning a highly likely particle (u, θ) would be sampled multiple times with some Gaussian noise, while particles with low weight would perish.

Motion models. We provide two motion models, for purely topological localization and for topometric localization. In the topological setting, the motion model is based on a discretized action space, comprising motions in the 8 cardinal directions. We use these actions as the motion prior u_t , to update our state x_t . The robot state is updated only when reaching the end of the current traversed edge.

We also extend the framework to handle the topometric

TABLE I: Evaluating sign understanding across 3 different environments.

	Cue Precision	Cue Recall	Sign Accuracy
Mall	0.824	0.847	0.692
Hospital	0.830	0.849	0.428
Campus	0.820	0.837	0.583

setting, where the navigational graph \mathcal{G} can be embedded in 3D Euclidean space. In this case, we track the robot pose $p_t = [x, y, z, \theta]$ and use the robot's odometry as u_t , sampling from the proposal distribution $p(p_t \mid p_{t-1}, u_t)$. p_t can be discretized to obtain the state $x_t = (v_t, \theta_t)$.

IV. EXPERIMENTAL EVALUATION

In this work, we propose a novel localization approach that exploits navigational cues from signs, enabling deployment of robots in new environments without prior sensor-based mapping. We conduct experiments to verify the performance of our method. The results of our experiments also support our key claims, which are: (i) Extract navigational information from signs; (ii) Utilize various sources of priors to establish a unified representation; (iii) Localize robustly in large human-centric environments, with seamless transition between outdoor and indoor spaces.

A. Experimental Setup

We recorded an evaluation dataset for global localization across multi-storey, multi-building indoor and outdoor public spaces, including a hospital, a mall, and a university campus. Indoor areas had floor plans and venue maps, while outdoor regions had OpenStreetMap data. Each trajectory contains odometry and RGB image streams. The hospital and mall data were captured with an iPhone, and the campus data with an arm-mounted camera on a Boston Dynamics Spot robot. Due to the Spot camera's low mounting, we employed the active perception module from Agrawal *et al.* [1] to reposition the arm for optimal sign visibility.

B. Sign Understanding

We evaluate sign understanding on 10 trajectories across the 3 test environments. The trajectories have human-annotated ground truth navigational cues for each sign in each trajectory. The cue precision/recall metrics are adapted from the sign recognition metrics in [1], and reflect accuracy of inferring navigational cues, aggregated over all signs in the 10 trajectories. Cue precision is the fraction of accurately inferred cues out of the total number of cues inferred by the sign understanding pipeline, while cue recall is the fraction of accurately inferred cues out of the total number of ground truth annotated cues. We also compute sign accuracy, which denotes the fraction of signs across all 10 trajectories which are perfectly parsed. We define a sign to be perfectly parsed if the set of inferred navigational cues for that sign exactly matches the annotated ground truth.

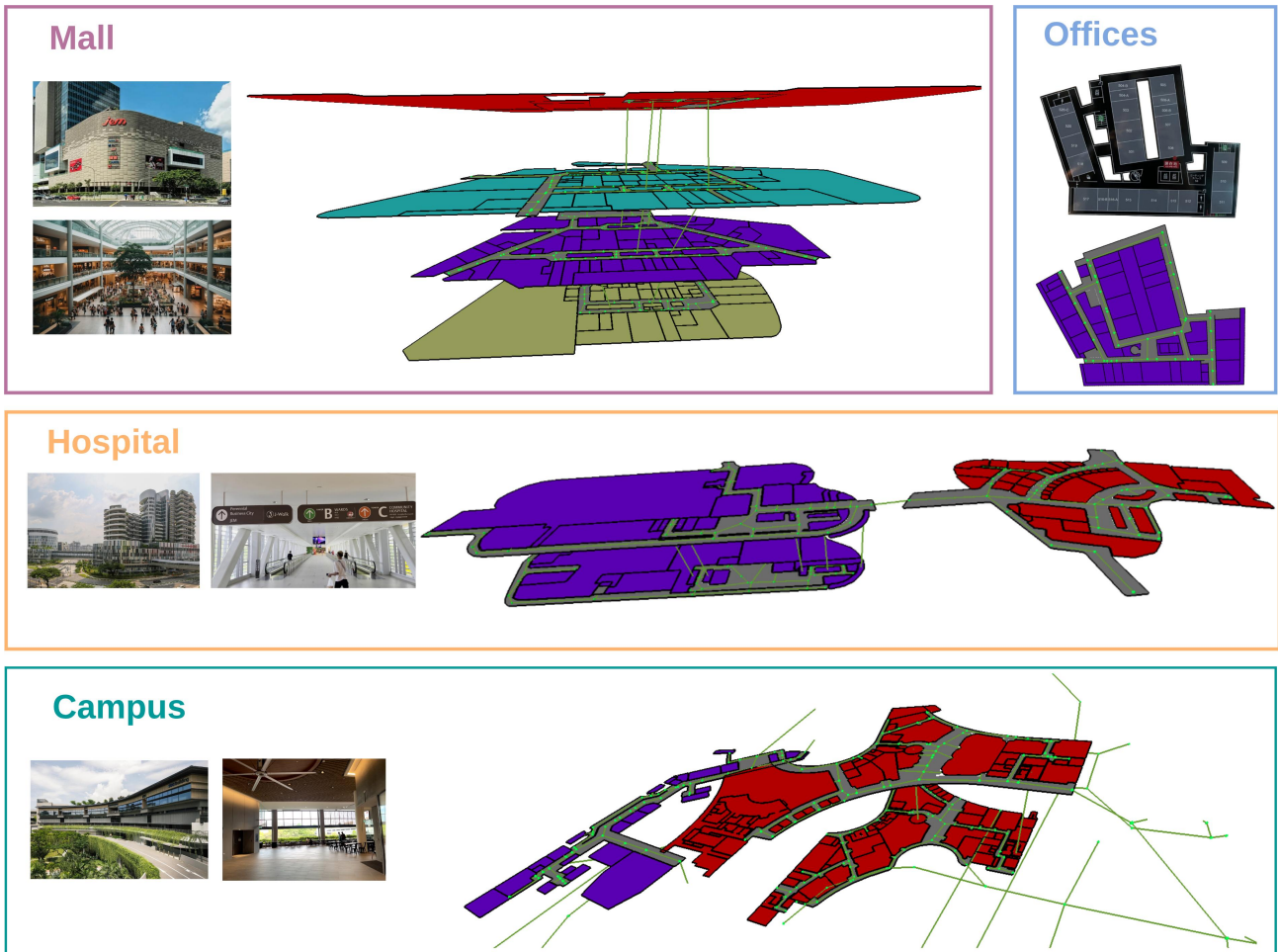


Fig. 5: Qualitative results of our map extraction approach with multi-floor, multi-building navigational graphs extracted from publicly available maps. In the top right (offices) is our constructed navigational graph for the floor plan used by Xie et al. [40].

Tab. I shows that the sign understanding pipeline performs consistently well on navigational signs across diverse environments. The main failure mode observed is incorrect direction category prediction due to complex sign structure in the wild, which may make association of a textual label to a directional arrow ambiguous. Overall, the high cue precision/recall values indicate strong performance of the pipeline, and suggest that each observed sign has a high proportion of accurately parsed navigational cues. While some inaccuracies in parsing individual cues lead to lower sign accuracy values, **SignLoc** is able to utilize the remaining accurately parsed cues from each sign for localization, enabling strong performance in localization.

C. Map extraction

We qualitatively evaluate our map extraction on several publicly available maps, including a hospital, a shopping mall and a university campus, as well as office buildings from different countries. Overall, the map extraction was tested on 13 separate floor plans and venue maps, from 7 different buildings. We also test the augmentation of our navigational graph with information from OpenStreetMap’s road network, and show that the navigational graph can also be extended seamlessly to describe outdoor urban environments.

We benchmark against the work of Xie et al. [40], who also attempted to extract a navigational graph from a floor plan. While our approach successfully extracted a navigational graph from the map used by Xie et al. [40], as seen in Fig. 5, their approach failed to produce navigational graphs from the publicly available maps we used for evaluation and localization. Our tested locations include multi-floor buildings, the alignment of connected buildings with separate floor plans and the extension of the navigational graph to outdoor spaces, all which were not addressed by Xie et al. [40]. The qualitative results of our map extraction approach can be seen in Fig. 5 and Fig. 1. We show that the extracted navigational graphs can support robust sign-centric localization in Sec. IV-D. We observe that the most challenging part for our approach is extracting symbols and text from the maps, despite the recent advances of VLMs and OCR approaches.

D. Localization

We evaluate on 10 sequences from the dataset described in Sec. IV-A. For global localization, particles are uniformly initialized across all traversable nodes of the navigational graph. Thus, the number of particles differs for each graph, with the largest graph initialized with 4448 particles.

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

We evaluate topological localization on nine sequences, each containing 5 sign sightings. Sequence S10 has 7 sign sightings, and is used to evaluate topometric localization. S1-S3 are collected in a mall, with a graph containing 4 retail floors. Trajectories traverse escalators to change floors. S4-S6 are collected in a hospital, and capture transitions between two buildings (Towers A and B). S7-S9 are collected on a university campus, across 2 buildings with transitions between floors. S10 starts outdoors at the basement level and traverses up to level 1, across two buildings, measuring at around 300 m long. More details about the metadata for the navigational graphs used for localization is included in Tab. II. For the topological localization, we use actions $\{\uparrow, \leftarrow, \downarrow, \rightarrow, \dots\}$ taken at each node for the motion model, while for the topometric localization we rely only on odometry from the robot.

To evaluate the performance of the approach, we compare the localization estimation at sign sighting locations, where ground truth localization is manually annotated based on the known location of the signs. To determine correctness of pose estimation, the predicted node must match the GT node, and the orientation difference must be lower than $\frac{\pi}{4}$. Localization is considered successful if the prediction matched the ground truth before the end of the sequence, and if the localization remained converged until the end of the sequence.

As can be seen in Tab. III, the global localization converges to the correct topological node and the correct (discretized) orientation in all sequences. The convergence occurs after the first or second sign sighting in 80% of the cases, but more challenging sequences can require more signs. Specifically, slower convergence on S3 and S6 are attributed to noise from locations referenced in signs that were not reflected in the public map. Also, S6 featured hard-to-parse signs with ambiguous location-direction associations, causing noisier localization. For all sequences, once the pose estimation converged, it remained correct until the end of the sequence, resulting in 100% success rate.

We evaluate localization performance with ground-truth sign annotations, and find it to be effectively identical to that with signs parsed by our pipeline (Tab. III). This indicates that perception noise has limited impact. The primary source of noise is missing navigational cues—while some cues are omitted, the remaining ones are generally accurate and often sufficient to localize the robot reliably.

As our approach can utilize both venue maps and OSM information, it was difficult to compare it with an appropriate baseline. While VPR approaches such as Lalaloc++ [16] can localize in floor plans, they cannot localize outdoors and also are challenged by environments with high geometric symmetry. Approaches such as OrienterNet [31] can globally localize outdoors with a single RGB image and OSM, but cannot be extended to indoor spaces.

E. Runtime

We evaluated our approach on an Nvidia Jetson Orin. For each sign, we batch the multiple parsing prompts in a single VLM query. We measure runtime on the campus navigational graph, containing 605 nodes, using 4448 particles to

TABLE II: Map metadata for the navigational graphs that were used for localization.

	nodes	edges	floors	buildings	ind. area	out. area
Mall	641	2416	4	1	51 338 m ²	-
Hospital	309	1200	2	2	30 012 m ²	-
Campus	605	2145	2	2	16 905 m ²	160 000 m ²

TABLE III: Localization performance for 3 different environments. The metric i/n reports at which sign sighting i^{th} of the n visible signs the approach pose estimation converged to the ground truth.

Mall			Hospital			Campus			
S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
1/5	1/5	4/5	1/5	2/5	4/5	1/5	2/5	2/5	1/7

localize. The observation model (~ 0.025 s) and motion model (~ 0.012 s) execute in real-time. Each candidate navigational sign requires a VLM query, which takes ~ 2.5 s, during which the robot must temporarily halt. However, since signs are typically sparse, this cost is amortized over the trajectory and has limited impact on overall navigation efficiency, enabling real-time navigation. Our approach runs online on the Spot, and we enclose a video of online localization in our supplementary.

V. CONCLUSION

We present **SignLoc**, a global localization approach that matches directional cues from navigational signs to a navigational graph derived from publicly available human-centric maps—floor plans and OpenStreetMap (OSM) graphs. Because signs are ubiquitous in human environments and encode information closely aligned with these maps, they serve as ideal semantic features for large-scale localization spanning indoor and outdoor spaces. Experiments validate this, showing that **SignLoc** rapidly localizes in diverse routes after observing just one or two signs, and demonstrating its ability to globally localize and track the robot’s location on a long trajectory crossing both indoor and outdoor areas. Our code will be released on GitHub¹. The present system uses only signs to localize on public maps. In future work, our approach could be combined with methods like OrienterNet or LaLaLoc++, which leverage dense RGB features for localization.

REFERENCES

- [1] A. Agrawal, J. Loo, N. Zimmerman, and D. Hsu. Sign Language: Towards Sign Understanding for Robot Autonomy, 2025.
- [2] J.L. Almeida, F.C. Flores, M.N. Roecker, M.A. Braga, and Y.M. Costa. An indoor sign dataset (isd): An overview and baseline evaluation. *VISIGRAPP (4: VISAPP)*, pages 505–512, 2019.
- [3] M. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing (TSP)*, 50(2):174–188, 2002.
- [4] H. Bavle, J.L. Sanchez-Lopez, M. Shaheer, J. Civera, and H. Voos. Situational graphs for robot navigation in structured indoor environments. *IEEE Robotics and Automation Letters*, 7(4):9107–9114, 2022.
- [5] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. Reid, and J. Leonard. Past, Present, and Future of Simultaneous Localization And Mapping: Towards the Robust-Perception Age. *IEEE Trans. on Robotics (TRO)*, 32:1309–1332, 2016.

¹<https://github.com/AdaCompNUS/SignLoc>

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

- [6] H. Cheng, H. Chen, and Y. Liu. Topological indoor localization and navigation for autonomous mobile robot. *IEEE Transactions on Automation Science and Engineering*, 12(2):729–738, 2014.
- [7] C. Cui, T. Sun, M. Lin, T. Gao, Y. Zhang, J. Liu, X. Wang, Z. Zhang, C. Zhou, H. Liu, Y. Zhang, W. Lv, K. Huang, Y. Zhang, J. Zhang, J. Zhang, Y. Liu, D. Yu, and Y. Ma. Paddleocr 3.0 technical report, 2025.
- [8] L. Cui, C. Rong, J. Huang, A. Rosendo, and L. Kneip. Monte-Carlo Localization in Underground Parking Lots Using Parking Slot Numbers. In *Proc. of the IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2021.
- [9] A. De la Escalera, J.M. Armingol, and M. Mata. Traffic sign recognition and analysis for intelligent vehicles. *Image and vision computing*, 21(3):247–258, 2003.
- [10] A. De La Escalera, L.E. Moreno, M.A. Salichs, and J.M. Armingol. Road traffic sign detection and classification. *IEEE transactions on industrial electronics*, 44(6):848–859, 1997.
- [11] F. Dellaert, D. Fox, W. Burgard, and S. Thrun. Monte carlo localization for mobile robots. In *Proc. of the IEEE Intl. Conf. on Robotics & Automation (ICRA)*, 1999.
- [12] D. Fox, W. Burgard, F. Dellaert, and S. Thrun. Monte Carlo Localization: Efficient Position Estimation for Mobile Robots. *Proc. of the Conference on Advancements of Artificial Intelligence (AAAI)*, 1999(343-349), 1999.
- [13] M. Haklay and P. Weber. Openstreetmap: User-generated street maps. *IEEE Pervasive computing*, 7(4):12–18, 2008.
- [14] B. Hoferlin and K. Zimmermann. Towards reliable traffic sign recognition. In *2009 IEEE Intelligent Vehicles Symposium*, pages 324–329. IEEE, 2009.
- [15] M. Hosseinzadeh, Y. Latif, T. Pham, N. Suenderhauf, and I. Reid. Towards Semantic SLAM: Points, Planes and Objects. *arXiv preprint*, 2018.
- [16] H. Howard-Jenkins and V.A. Prisacariu. Lalaloc++: Global floor plan comprehension for layout localisation in unvisited environments. In *European Conference on Computer Vision*, pages 693–709. Springer, 2022.
- [17] B. Kieler, W. Huang, J.H. Hauernt, and J. Jiang. Matching river datasets of different scales. In *Advances in GIScience: Proceedings of the 12th AGILE conference*, pages 135–154. Springer, 2009.
- [18] N. Kobyshev, H. Riemenschneider, and L.V. Gool. Matching features correctly through semantic understanding. In *2014 2nd International Conference on 3D Vision*, volume 1, pages 472–479, 2014.
- [19] J.J. Leonard and H.F. Durrant-Whyte. Mobile robot localization by tracking geometric beacons. *IEEE Transactions on robotics and Automation*, 7(3):376–382, 1991.
- [20] C. Liang, R.A. Knepper, and F.T. Pokorny. No map, no problem: A local sensing approach for navigation in human-made spaces using signs. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 6148–6155. IEEE, 2020.
- [21] C. Liu, J. Wu, P. Kohli, and Y. Furukawa. Raster-to-vector: Revisiting floorplan transformation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2195–2203, 2017.
- [22] S. Liu, Z. Zeng, T. Ren, F. Li, H. Zhang, et al. Grounding dino: Marrying dino with grounded pre-training for open-set object detection. *arXiv preprint arXiv:2303.05499*, 2023.
- [23] J. Loo and D. Hsu. Scene action maps: Behavioural maps for navigation without metric information. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 6354–6360, 2024.
- [24] F.P. Miller, A.F. Vandome, and J. McBrewhster. *Levenshtein Distance: Information theory, Computer science, String (computer science), String metric, Damerau-Levenshtein distance, Spell checker, Hamming distance*. Alpha Press, 2009.
- [25] H. Moravec and A. Elfes. High resolution maps from wide angle sonar. In *Proc. of the IEEE Intl. Conf. on Robotics & Automation (ICRA)*, 1985.
- [26] P. Panphattarasap and A. Calway. Automated Map Reading: Image Based Localisation in 2-D Maps Using Binary Semantic Descriptors. *arXiv preprint*, 2018.
- [27] A. Prorok, A. Bahr, and A. Martinoli. Low-cost collaborative localization for large-scale multi-robot systems. In *2012 IEEE International Conference on Robotics and Automation*, pages 4236–4241. Ieee, 2012.
- [28] A. Rosinol, M. Abate, Y. Chang, and L. Carlone. Kimera: an open-source library for real-time metric-semantic localization and mapping. In *Proc. of the IEEE Intl. Conf. on Robotics & Automation (ICRA)*, 2020.
- [29] A. Rosinol, A. Gupta, M. Abate, J. Shi, and L. Carlone. 3d dynamic scene graphs: Actionable spatial perception with places, objects, and humans. *arXiv preprint arXiv:2002.06289*, 2020.
- [30] P. Ruchti, B. Steder, M. Ruhnke, and W. Burgard. Localization on OpenStreetMap Data Using a 3D Laser Scanner. In *Proc. of the IEEE Intl. Conf. on Robotics & Automation (ICRA)*, 2015.
- [31] P.E. Sarlin, D. DeTone, T.Y. Yang, A. Avetisyan, J. Straub, T. Malisiewicz, S.R. Bulo, R. Newcombe, P. Kotschieder, and V. Balntas. Orienternet: Visual localization in 2d public maps with neural matching. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 21632–21642, 2023.
- [32] M. Schreiber, C. Knöppel, and U. Franke. Laneloc: Lane marking based localization using highly accurate maps. In *2013 IEEE Intelligent Vehicles Symposium (IV)*, pages 449–454, 2013.
- [33] K. Seongyong, Seula, K. Park, Hyeonjung, and Y. Kiyun. Deep floor plan analysis for complicated drawings based on styletransfer. *Journal of Computing in Civil Engineering*, 2020.
- [34] S. Thrun, W. Burgard, and D. Fox. *Probabilistic Robotics*. MIT Press, 2005.
- [35] I. Ullah, D. Adhikari, H. Khan, M.S. Anwar, S. Ahmad, and X. Bai. Mobile robot localization: Current challenges and future prospective. *Computer Science Review*, 53:100651, 2024.
- [36] O. Vysotska and C. Stachniss. Effective Visual Place Recognition Using Multi-Sequence Maps. *IEEE Robotics and Automation Letters (RA-L)*, 4:1730–1736, 2019.
- [37] J. Wang, Y. Chen, Z. Dong, and M. Gao. Improved yolov5 network for real-time multi-scale traffic sign detection. *Neural Computing and Applications*, 35(10):7853–7865, 2023.
- [38] S. Wang, X. Yang, and Y. Tian. Detecting signage and doors for blind navigation and wayfinding. *Network Modeling Analysis in Health Informatics and Bioinformatics*, 2:81–93, 2013.
- [39] L. Wiesmann, T. Guadagnino, I. Vizzo, N. Zimmerman, Y. Pan, H. Kuang, J. Behley, and C. Stachniss. Locndf: Neural distance field mapping for robot localization. *IEEE Robotics and Automation Letters*, 8(8):4999–5006, 2023.
- [40] L. Xie, A. Markham, and N. Trigoni. SnapNav: Learning Mapless Visual Navigation with Sparse Directional Guidance and Visual Reference. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, 2020.
- [41] Z. Xu, C. Yang, S. Alheejawi, N. Jha, S. Mehadi, and M. Mandal. Automatic floor plan analysis using a boundary attention-based deep network. *International Journal on Document Analysis and Recognition (IJ DAR)*, pages 1–12, 2024.
- [42] F. Yan, O. Vysotska, and C. Stachniss. Global Localization on OpenStreetMap Using 4-bit Semantic Descriptors. In *Proc. of the Europ. Conf. on Mobile Robotics (ECMR)*, 2019.
- [43] L. Yang and M. Worboys. Generation of navigation graphs for indoor space. *International Journal of Geographical Information Science*, 29(10):1737–1756, 2015.
- [44] A.R. Zamir and M. Shah. Accurate image localization based on google maps street view. In *Computer Vision-ECCV 2010: 11th European Conference on Computer Vision, Heraklion, Crete, Greece, September 5-11, 2010, Proceedings, Part IV 11*, pages 255–268. Springer, 2010.
- [45] Z. Zeng, X. Li, Y.K. Yu, and C.W. Fu. Deep floor plan recognition using a multi-task network with room-boundary-guided attention. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9096–9104, 2019.
- [46] J. Zhang and S. Singh. LOAM: Lidar Odometry and Mapping in Real-time. In *Proc. of Robotics: Science and Systems (RSS)*, 2014.
- [47] J. Zhang, X. Zou, L.D. Kuang, J. Wang, R.S. Sherratt, and X. Yu. CctsdB 2021: a more comprehensive traffic sign detection benchmark. *Human-centric Computing and Information Sciences*, 12, 2022.
- [48] N. Zimmerman, T. Guadagnino, X. Chen, J. Behley, and C. Stachniss. Long-Term Localization Using Semantic Cues in Floor Plan Maps. *IEEE Robotics and Automation Letters (RA-L)*, 8(1):176–183, 2023.
- [49] N. Zimmerman, L. Wiesmann, T. Guadagnino, T. Läbe, J. Behley, and C. Stachniss. Robust Onboard Localization in Changing Environments Exploiting Text Spotting. *Proc. of the IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2022.