



COMPASS: Learning Global Spatial Context for Long-Range Goal Navigation

Jose Lavariega¹, Erica Tevere², Jiaxu Xing³, Rudolf Rieter³, Patrick Spieler²

¹ETH Zurich; ²NASA Jet Propulsion Laboratory, California Institute of Technology;

³Robotics and Perception Group, University of Zurich



Jet Propulsion Laboratory
California Institute of Technology

ETH zürich

1. Introduction & Motivation

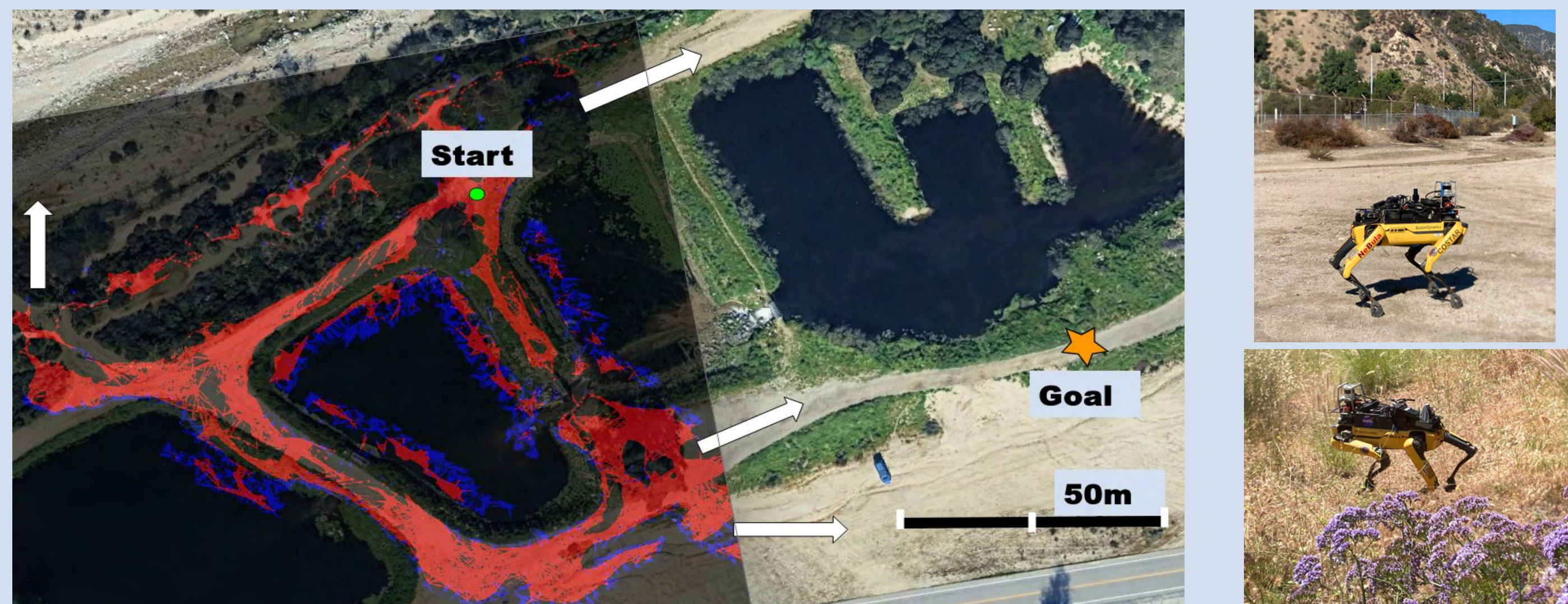


Fig.1: (Left): Long Range Navigation problem: Which frontiers to open in unexplored space, to reach the goal? Visualized with our Navigation Graph (Right) Data Collection and Deployment in a semantically-sparse outdoor environment; Dirt Path with Open Field (top) Urban with low-brush (bottom)

Autonomous long-range navigation in unstructured, unknown environments is an essential capability for search-and-rescue and goal-driven exploration. A robot must choose a frontier in explored space to continue navigation towards the goal (Fig1). To increase adaptability to unseen, semantically poor, environments, our method uses visual and geometric features without visual heuristics. Work using graph memory for navigation has so far been tested on feature-rich environments. [1,2]

2. Method

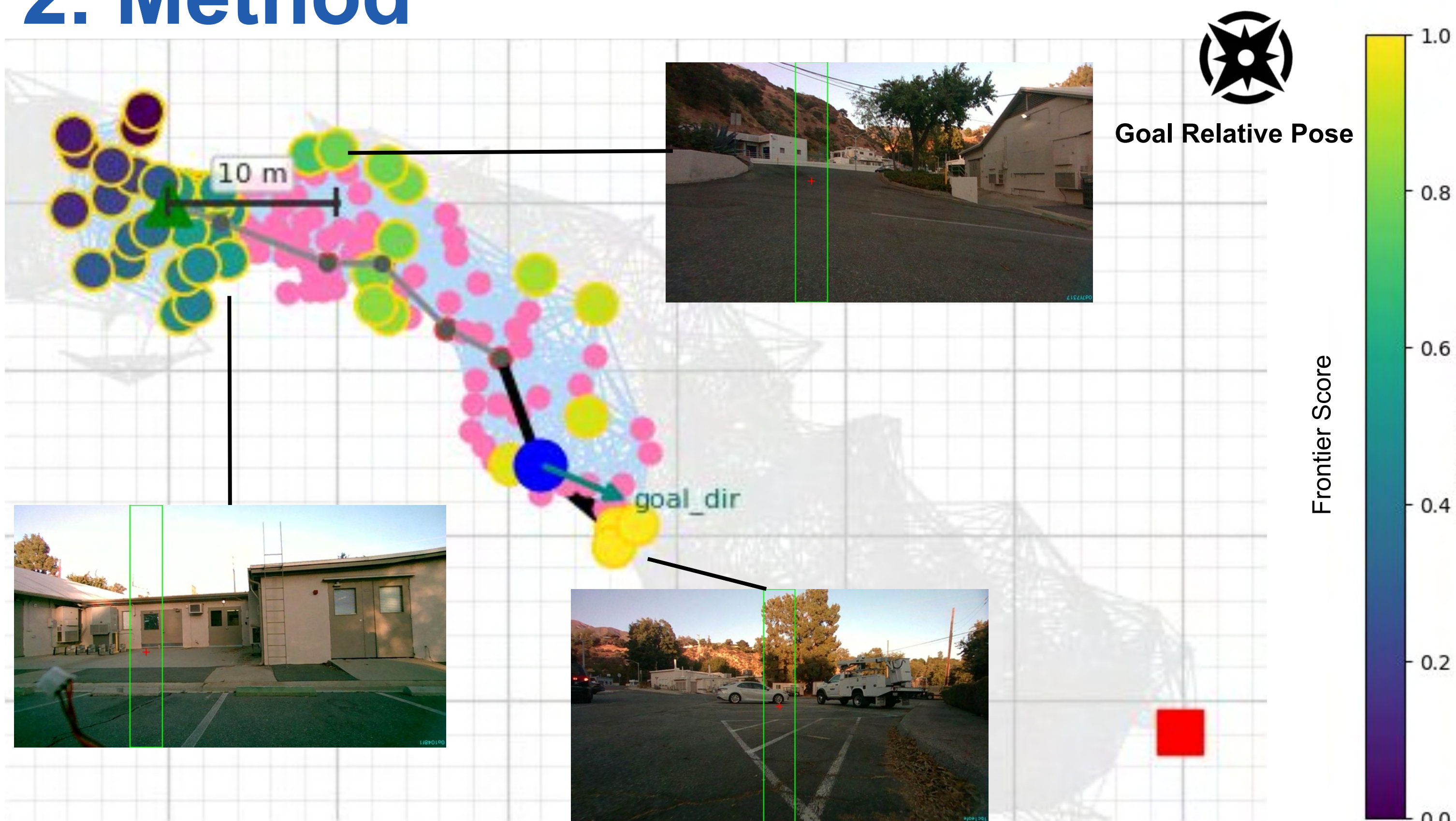


Fig.2: COMPASS: The currently-observed navigation graph (light blue) is passed through a Graph Convolutional Network that scores the current frontiers. Each of the frontiers are projected from graph node position and encoded into a DINO foundation model. Geometric and image features are scored to select the next frontier to open (viridis-colored nodes).

COMPASS (Fig.2) fuses a LiDAR-constructed navigation graph (navgraph) and visual features to encode traversability, explored space, and safe paths in a compact manner for long-range navigation decision making.

- Firstly**, we use a Graph Convolutional Network (GCN) to score the current frontiers in the navigation graph according to cost-to-go to get to goal. Through graph structure, node degree, free-space radius and inter-node traversability costs, the GCN learns macro-structures like dead-ends.(Fig.3)
- Secondly**, we use a DINOv3[3] foundation model to encode raw segments of camera image. Each image segment is centered around a frontier projected from global position onto pixel values.
- Thirdly**, we fuse visual, geometric and a goal-alignment feature to generate a learning environment for RL, deployed on a simple simulator from field data.
- Finally**, we train a Deep Q-Network policy to pick the frontier that reduces cost-to-go towards the goal.

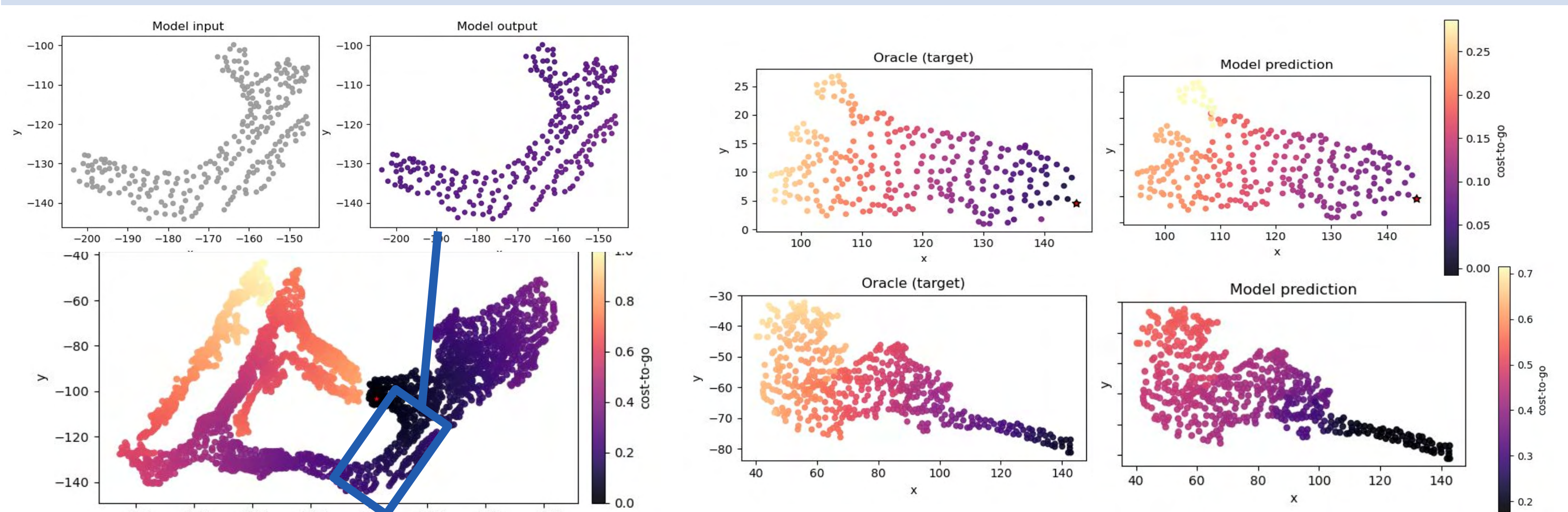


Fig.3: Our GCN is trained to match a goal-conditioned node score (right, Oracle columns) calculated from optimal path to goal. Subgraphs without connection to the goal (left, top) are generated as inputs to the GCN. (Left, bottom) shows a full scored graph, from which the input is subsampled. Inputs and oracles are normalized such that scores are always 0-1, thus capturing graph macrostructures (deadends, hubs) without dependency from absolute number of nodes in the subgraph.

3. Learning Environment

COMPASS is trained through 4 hours of field-collected data on semantically-sparse outdoor environments: open fields, hiking paths, wilderness with positive & negative obstacles, and urban environments.

We implement a simple simulator to train our policy, augmenting 4 hours of data into millions of distinct start-goal pair rollouts.

Our dataset is further diversified by including artefacts like lens flares, changes in lighting, and multiple points-of-view of the same geometric features(Fig.4).

In Fig.2 we allude to our policy observation and action space. Our robot observes the current navigation graph, and spatially-coupled images of current frontiers as observed from within explored space. Every frontier image is first processed by a DINO model. We store multiple points-of-view of the current frontiers to mitigate effects from occlusions or poor lighting in the saved image. The robot action is to choose the frontier which is scored highest for minimal time towards the goal.

Since the observation space spans all available frontiers, a key cost in COMPASS is an estimated cost of switching from the last-opened frontier to the next frontier to open.

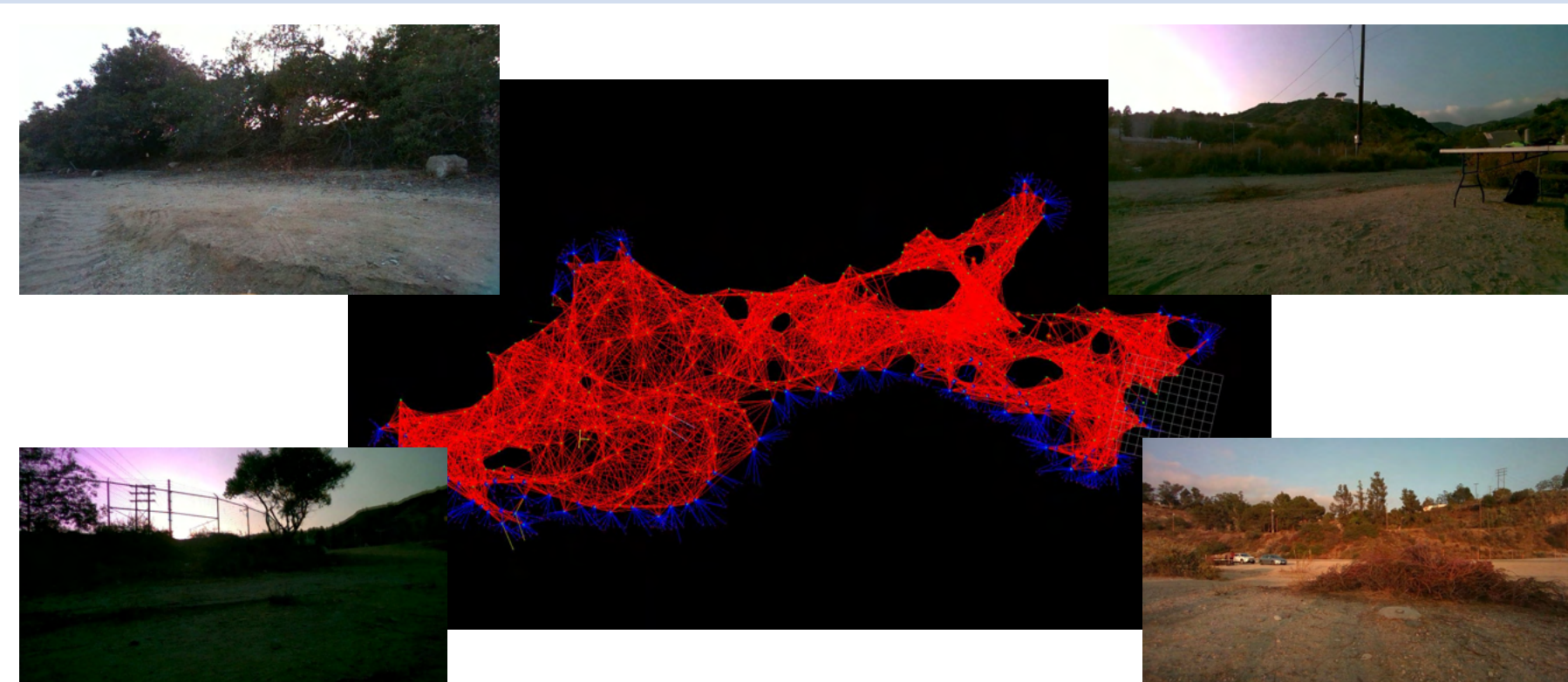
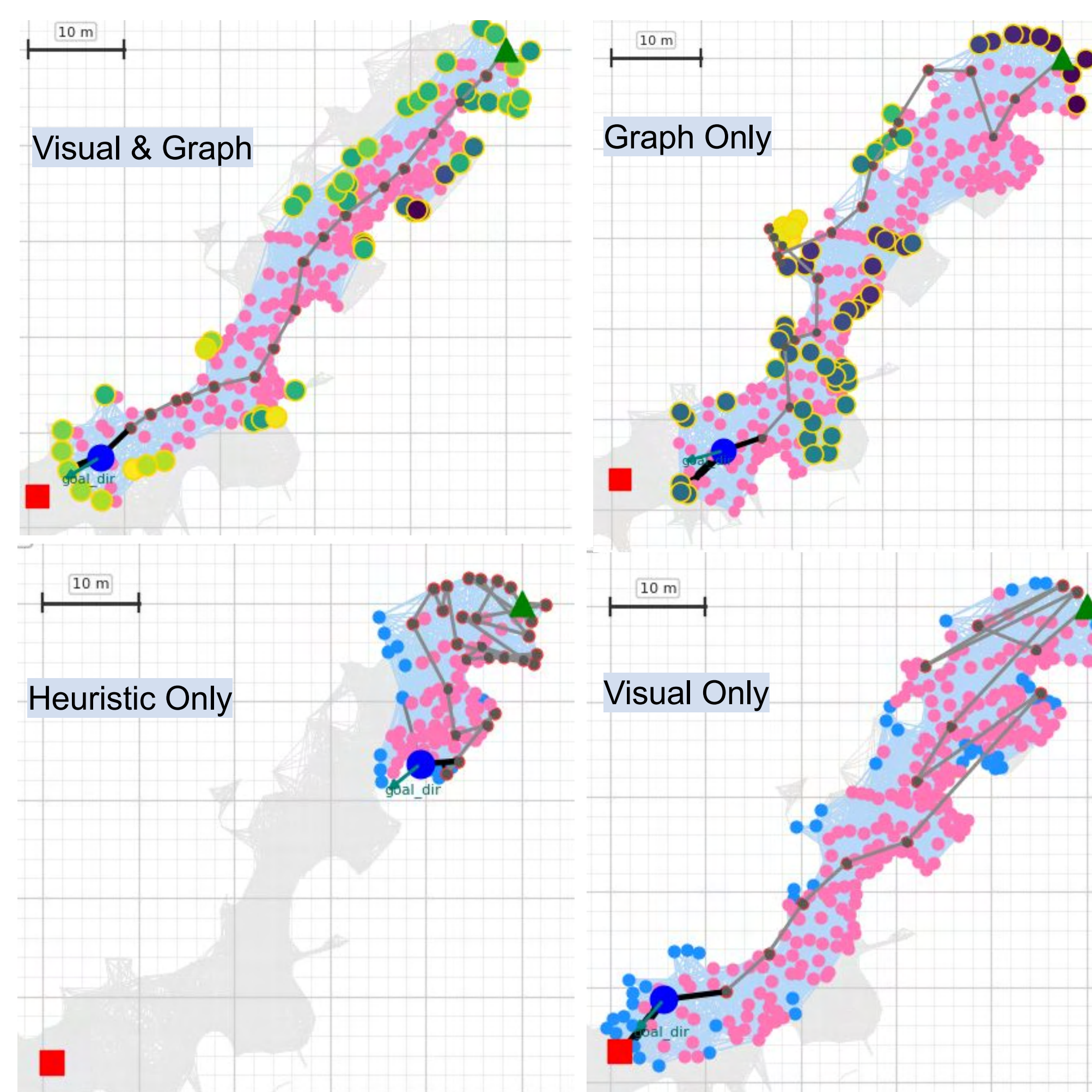


Fig. 4 Visual variety in a dataset obtained in open field with lens flares. Dataset navgraph visualized in center.

4. Results and Field Experiments

COMPASS is deployed on a commercially available quadruped equipped with additional LiDAR and visual cameras. We evaluate on the field in diverse outdoor environments: urban, open fields, hiking paths, and paths with elevation gain. Throughout all environments, we observe obstacles, dead-ends and occlusions. Comparison against other models [4] is pending at time of submission.

Ablation	Success Rate (75 Rollouts) $S_{rate} = \frac{1}{E} \sum_{i=1}^E S_i$	SR Weighted by Path Length (75 Rollouts) $SPL = \frac{1}{E} \sum_{i=1}^E S_i * \frac{l_i}{\max(l_i, p_i)}$
Visual & Graph	0.947	0.920
Graph-Only	0.973	0.642
Visual-Only	0.920	0.143
Heuristic-Only	0.800	0.476



Ablation studies are conducted on semantically-distinct pre-captured unseen outdoor environments (urban, hiking, open field) and COMPASS is evaluated online on hardware in representative samples of these environments.

For COMPASS on hardware, we also log number of interventions needed for successful arrival to goal.

Fig.5: Frontier choices taken during rollout for a pre-captured urban environment, target goal 70m away with obstacles (cars, buildings and barriers).

5. References & Acknowledgements

- [1] H. Li, Z. Wang, X. Yang, Y. Yang, S. Mei, and Z. Zhang, "Memonav: Working memory model for visual navigation," 2024. [Online]. Available: <https://arxiv.org/abs/2402.19161>
- [2] O. Kwon, N. Kim, Y. Choi, H. Yoo, J. Park, and S. Oh, "Visual graph memory with unsupervised representation for visual navigation," in Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), October 2021, pp. 15 890–15899.
- [3] O. Simeoni, H. V. Vo, M. Seitzer, F. Baldassarre, M. Oquab, C. Jose, V. Khalidov, M. Szafraniec, S. Yi, M. Ramamonjisoa, F. Massa, D. Haziza, L. Wehrstedt, J. Wang, T. Darcet, T. Moutakanni, L. Sentana, C. Roberts, A. Vedaldi, J. Tolan, J. Brandt, C. Couprie, J. Mairal, H. J'egou, P. Labatut, and P. Bojanowski, "Dinov3," 2025. [Online]. Available: <https://arxiv.org/abs/2508.10104>
- [4] H. Shah, E. Tevere, D. Atha, M. Kaufmann, S. Khattak, M. Patel, M. Hutter, J. Frey, and P. Spieler, "Wildos: Open-vocabulary object search in the wild," 2026. [Online]. Available: <https://arxiv.org/abs/2602.19308>

The research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration (80NM0018D0004). ©2026. All rights reserved