

WildOS: Open Vocabulary Object Search in the Wild

Hardik Shah^{1,2}, Manthan Patel², Erica Tevere¹, Deegan Atha³, Marcel Kaufmann¹, Shehryar Khattak³,
Marco Hutter², Jonas Frey^{2,4,5}, Patrick Spieler¹

¹Jet Propulsion Laboratory, Caltech ²RSL, ETH Zurich, ³Field AI ⁴Stanford University ⁵UC Berkeley

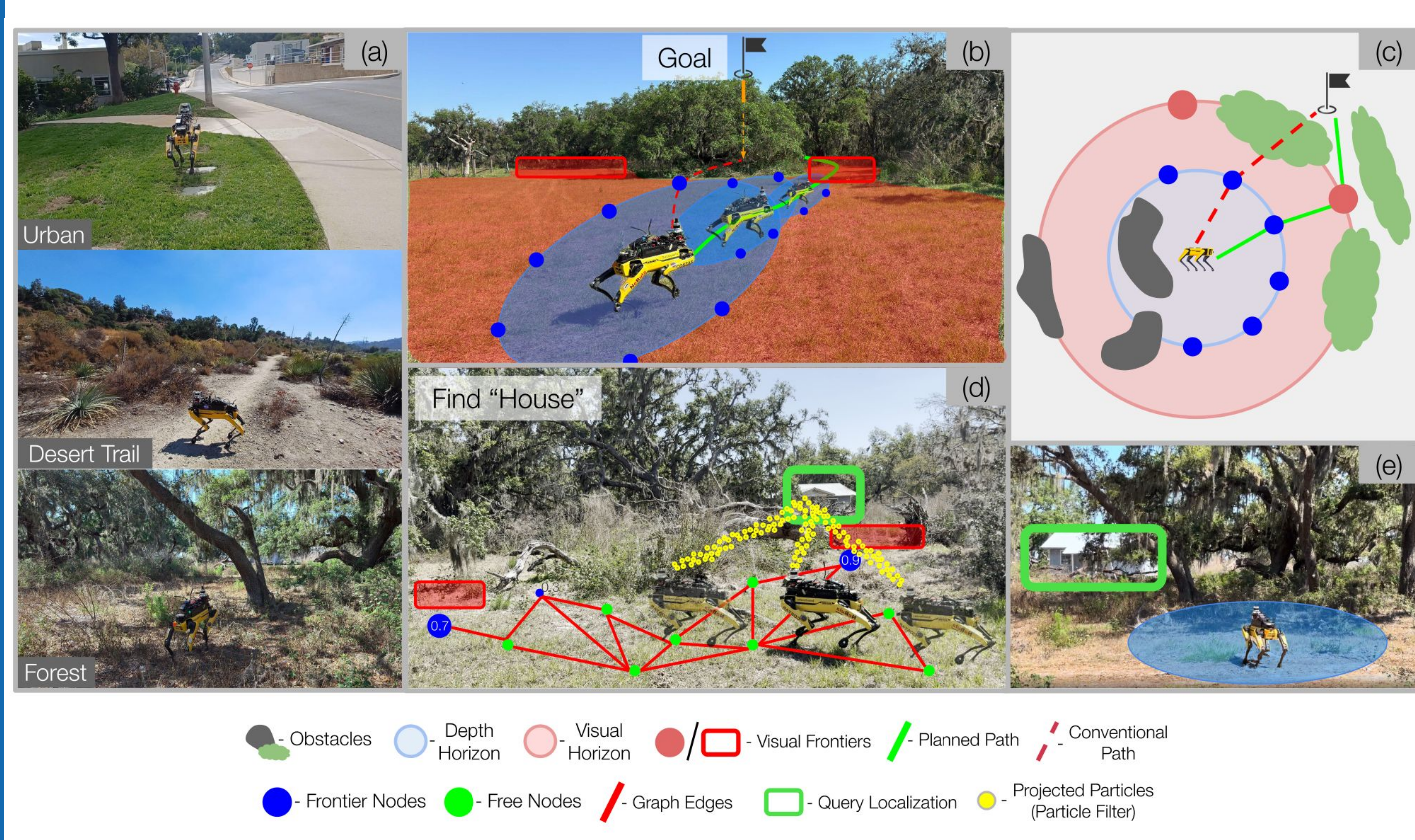


1 Introduction

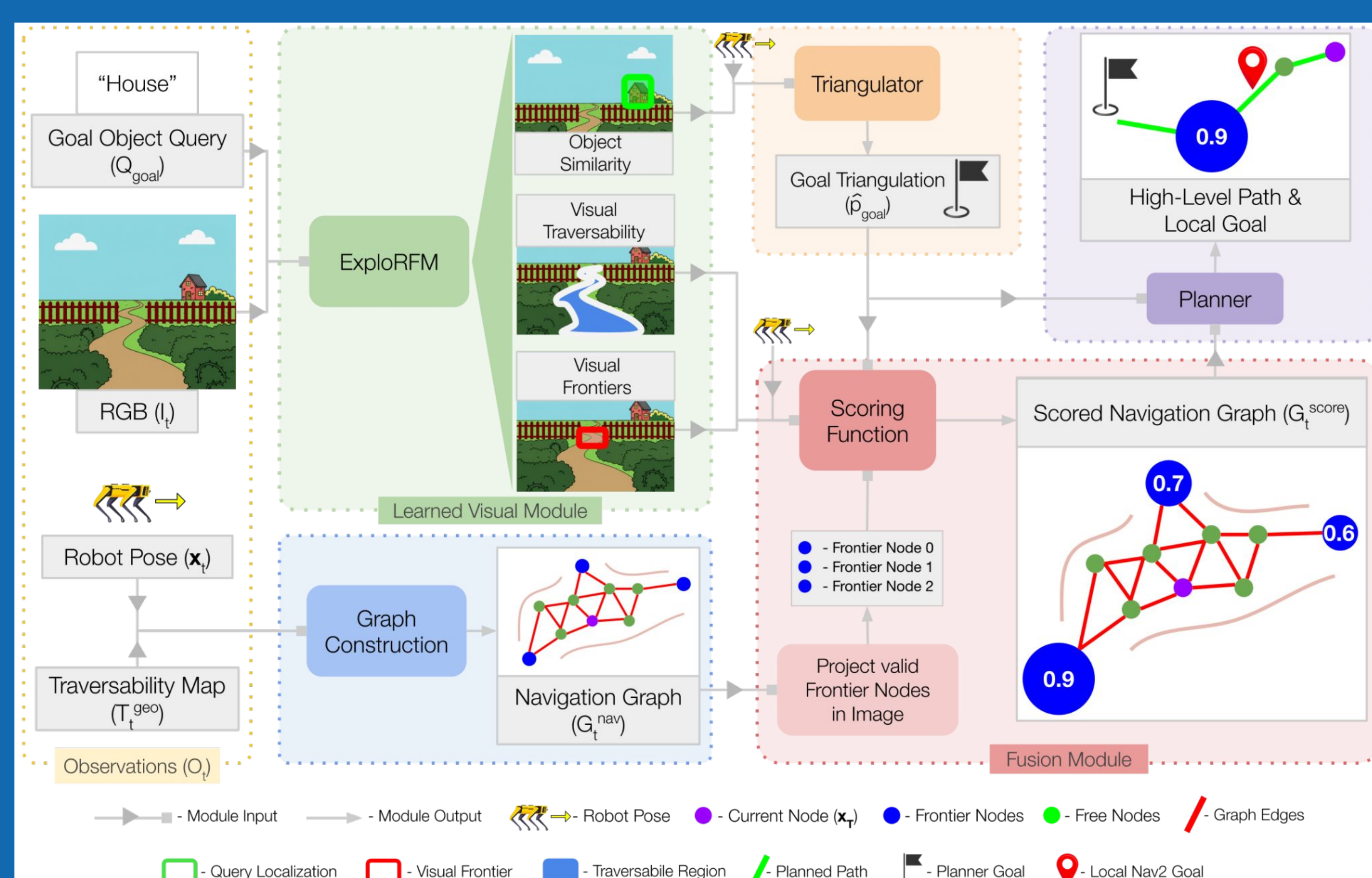
Motivation: Autonomous navigation in complex outdoor environments require robots to operate over long ranges without prior maps. Purely relying on geometric frontiers is insufficient and the robots need to reason semantically about *where to go* and *where is safe to traverse* for robust efficient exploration.

Objective: We present WildOS, a unified system for long-range, open-vocabulary object search that combines safe geometric exploration with semantic visual reasoning. WildOS builds a sparse navigation graph to maintain spatial memory, while utilizing a foundation-model-based vision module, ExploRFM, to score frontier nodes of the graph.

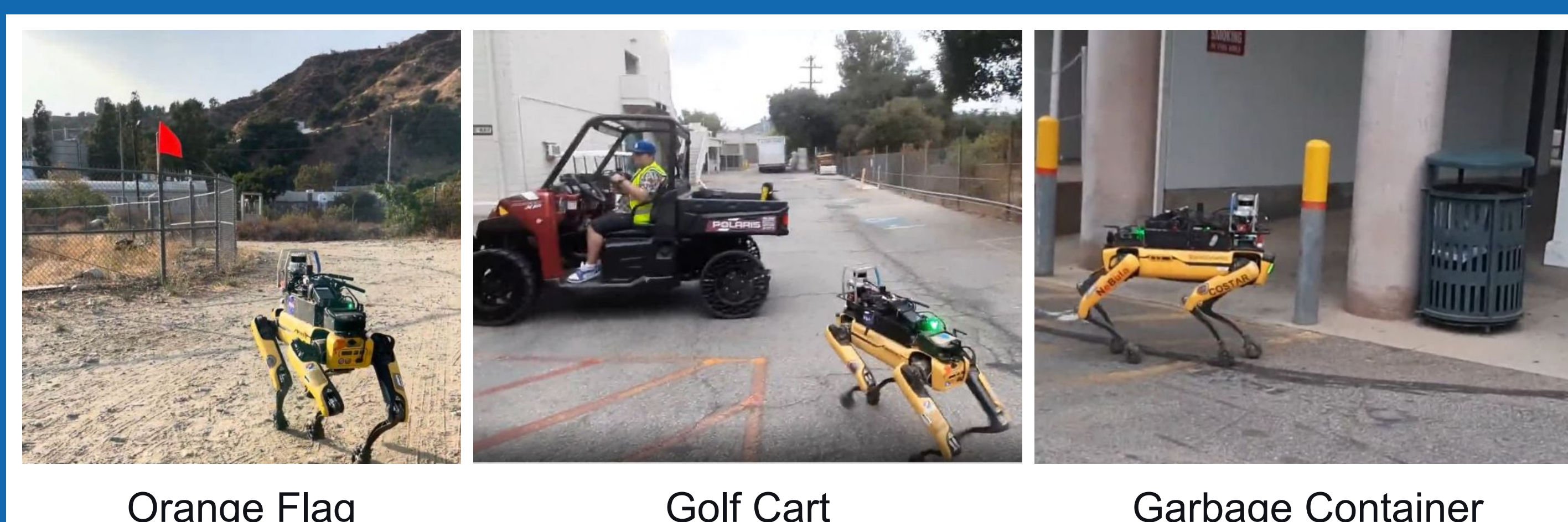
2 Summary



3 Method



4 Results - Open Vocab Object Search



- Our framework can handle any arbitrary Open Vocabulary Object query based on the similarity using Siglip-2 features

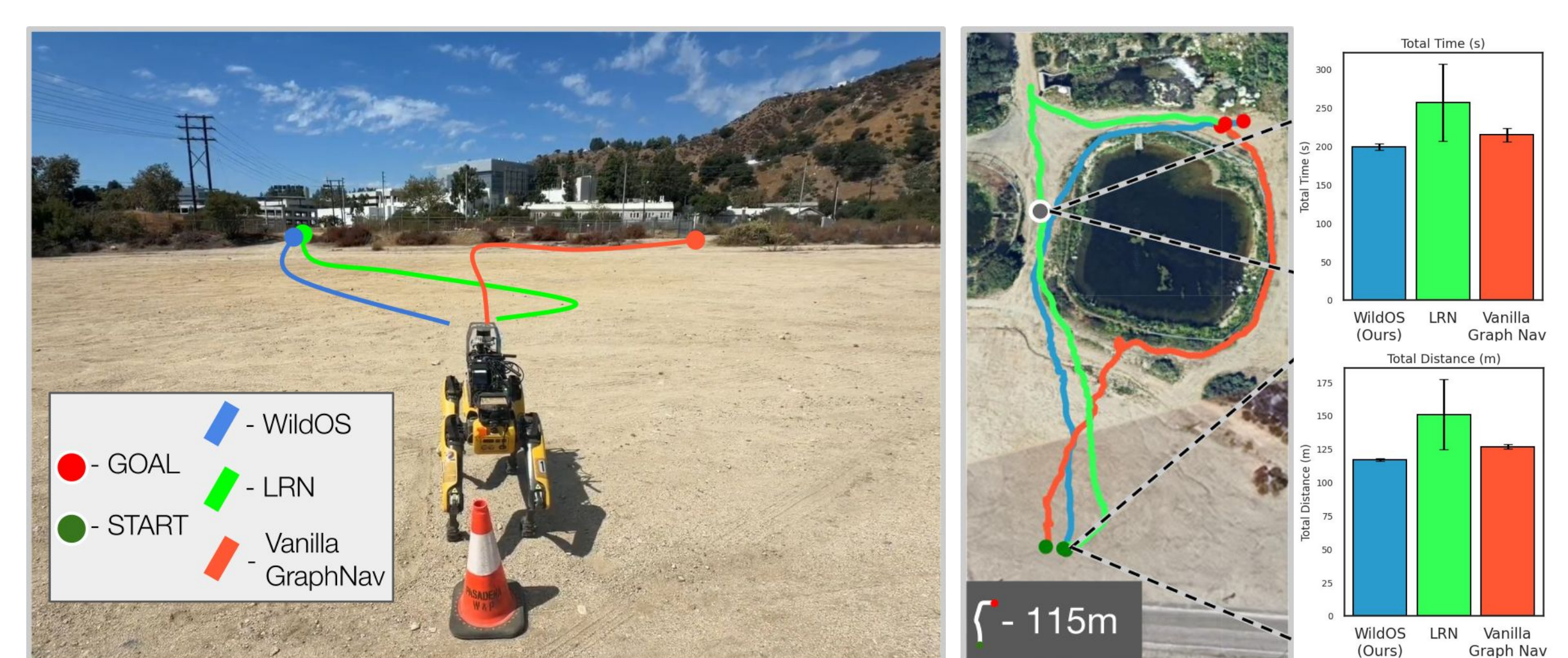
5 Results - Traversability and Visual Frontiers

- We train light-weight decoder heads on top of the Radio FM to predict traversability and visual frontiers
- The network generalizes to diverse scenarios despite being trained only on a few hundred images due to the frozen FM features



6 Results - Long Range Navigation

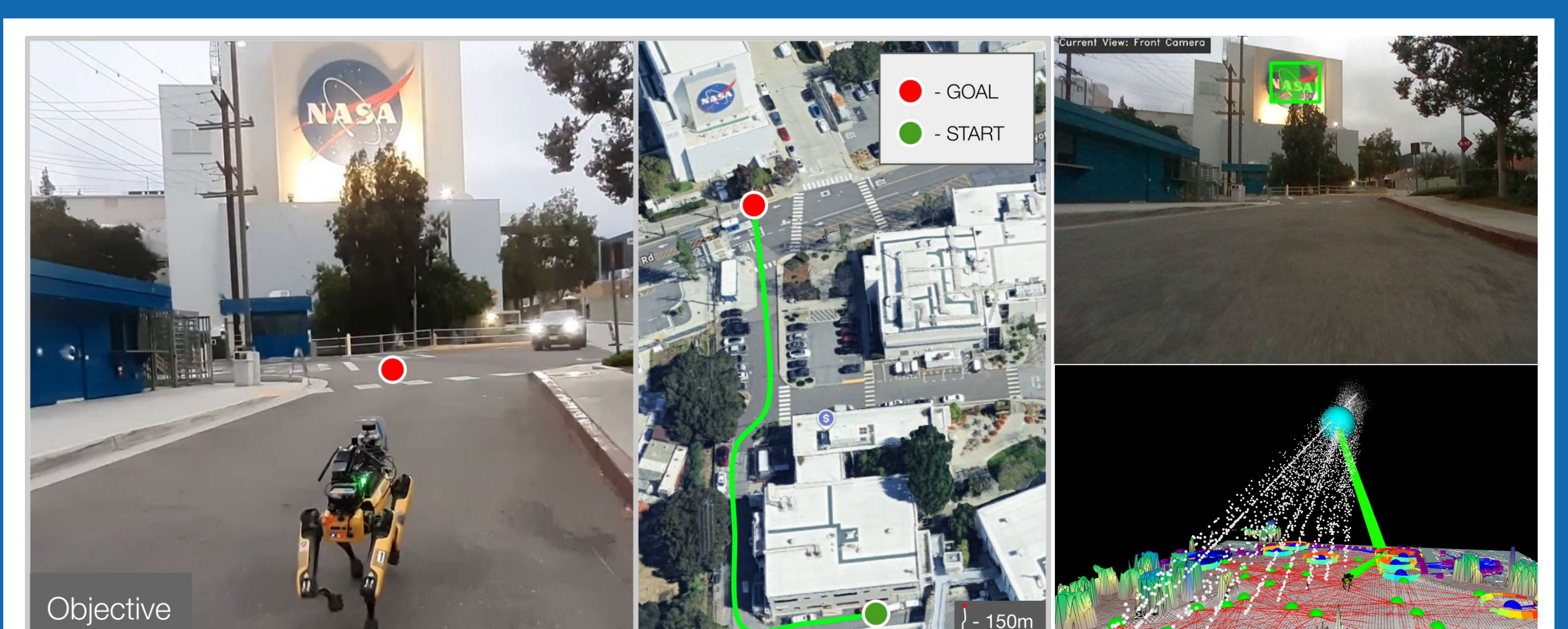
- Vision-based scoring with the nav. graph improves performance compared to pure-geometry or pure-vision approaches



- Graph structure provides us explicit memory enabling escape from dead-ends



7 Results - Full Framework Experiments



- Our framework enables long range object search navigation based on simple text query such as "Find the NASA logo" in the above scenario

8 Conclusion

- WildOS is a unified, real-time system for open-vocab object search and long-range navigation in unstructured outdoor environments
- Combines navigation graph with FM-based perception to achieve semantic, memory-aware exploration at scale
- Extensive closed-loop field experiments across diverse off-road & urban envs.