

SemantI-Nav: Human-Guided Imitation Learning for Safe Quadruped Navigation on Construction Sites

Liqun Xu¹, Dharmaraj Veeramani², and

Zhenhua Zhu³

^{1,3}Department of Civil and Environmental Engineering, University of Wisconsin-Madison, USA

²Department of Industrial and Systems Engineering, University of Wisconsin-Madison, USA
liqun.xu@wisc.edu, raj.veeramani@wisc.edu, zzhu286@wisc.edu

Abstract –

The U.S. construction industry continues to face persistent labor shortages that contribute to project delays and productivity loss. Mobile robots are a promising way to supplement human labor by automating repetitive and physically demanding tasks. Among these platforms, quadruped robots are well suited for construction sites because legged locomotion enables traversal of uneven terrain, stairs, and confined spaces. However, safe navigation at scale remains difficult in dynamic jobsites with rich semantic and social cues. Many existing systems rely on predefined rules or cost-based traversability maps derived from LiDAR/camera occupancy grids. These methods often reduce the world to “passable” and “impassable” space and can miss critical context, such as cones and caution tape, evolving work zones, and operating envelopes of heavy equipment. This paper presents SemantI-Nav, an imitation learning approach that learns navigation behaviors directly from expert demonstrations. We build a high-fidelity virtual construction site and collect multimodal demonstrations using a joystick-controlled quadruped robot, recording multi-view RGB-D observations, operator commands, and robot state. We train a policy that produces context-sensitive local velocity commands from these inputs. We evaluate the learned policy in validation scenarios with static layouts, dynamic workers, and dynamic machinery using success rate, navigation time, path length, and minimum distance to the goal. Results show high success across scenarios and performance comparable to a human operator in dynamic machinery settings, while remaining less efficient in some static and human-traffic cases.

Keywords –

Imitation learning (IL); Quadruped robot; Construction robotics; Robot navigation

1 Introduction

The U.S. construction industry continues to face persistent labor shortages, which contribute to productivity loss and project delays. Reports by the Associated General Contractors of America indicate that more than 75% of construction firms struggle to fill essential job positions [1,2]. To mitigate these shortages, construction companies are increasingly exploring robotic systems to supplement human labor effectively [3] and to improve productivity by automating repetitive and physically demanding tasks [4]. Among mobile platforms, quadruped robots are particularly promising for construction sites because legged locomotion enables traversal of uneven terrain, stairs, and confined spaces that can challenge wheeled or tracked systems [5].

Quadruped robots have already shown value for construction management tasks that require site mobility and frequent data capture. Prior work has demonstrated their utility in progress monitoring [6] and in inspection applications such as safety equipment inspection [7]. Realizing these benefits at scale, however, requires autonomous navigation that remains safe and reliable under complex and dynamic jobsite conditions [8].

Many current navigation solutions rely on predefined rules or cost-based traversability maps built from LiDAR and camera occupancy grids [9,10]. These pipelines often reduce the environment to “passable” and “impassable” space. This abstraction can miss safety-relevant context on construction sites, such as symbolic boundaries (e.g., cones and caution tape), evolving work zones, and equipment operating envelopes (e.g., an excavator swing radius). Because these constraints are diverse and situation dependent, it is difficult to hand-code

comprehensive rules for safe operation.

This limitation motivates navigation intelligence that is robust to the semantic and social complexity of construction. A policy must respond not only to geometry but also to cues that indicate intent, hazard, and right-of-way. Rather than encoding such cues as fixed rules or manually tuned costs, IL provides a direct path to ground robot behavior in expert decision-making observed under realistic conditions.

In this paper, we propose SemantIL-Nav, an IL approach that learns safe navigation behaviors directly from expert demonstrations. The goal of this study is to examine whether IL-based navigation is feasible in construction-site settings and whether a quadruped can learn a usable local navigation policy from human demonstrations under semantic and dynamic site conditions. We collect multimodal demonstrations in a high-fidelity virtual construction site, capturing how operators regulate speed in narrow passages, yield to moving machinery and nearby workers, and comply with symbolic safety markers. We then train a goal-conditioned local policy that maps multi-view RGB-D observations and robot state to continuous velocity commands for robust autonomy in semantically rich site environments. The method focuses on short-horizon local planning and control rather than explicit long-horizon path planning.

2 Related Work

2.1 Quadruped Robots in Construction Operations

Quadruped robots are increasingly used as mobile sensing platforms for construction operations because they can traverse rough terrain, stairs, and cluttered spaces. These capabilities support routine site tasks that require mobility and close-range observation. As a result, quadrupeds are being explored for inspection, documentation, and logistics support across diverse jobsite conditions [4].

Inspection has been a primary driver of quadruped deployment on construction sites. Recent studies used quadrupeds to inspect safety equipment and site conditions using onboard vision and sensing [7,11]. In addition to vision-based approaches, some studies explore multimodal inspection (e.g., acoustic or contact-based cues) for subsurface or material-condition assessment [12,13]. These efforts show that legged platforms can reach locations that are difficult for wheeled robots and can collect rich visual evidence for managers.

Progress monitoring and as-built data capture are also common use cases. Quadrupeds have been combined with SLAM and 3D reconstruction to support progress

tracking and building verification [14,15]. However, most prior systems still depend on manual operation or conservative navigation constraints, which limits autonomy in dynamic work zones.

2.2 Navigation in Construction Environments

Autonomous navigation is the key capability that enables quadrupeds to execute site tasks with minimal supervision. A typical navigation stack includes localization, mapping, global planning, and local obstacle avoidance. These components must operate under moving obstacles, temporary structures, and irregular terrain that are common on construction sites.

Most deployed systems rely on geometry-based navigation that represents the environment with occupancy grids or cost maps. Such representations are effective for avoiding hard obstacles, but they often treat the scene as only “free” or “occupied” space. This simplification can miss construction-specific constraints, such as restricted areas, temporary markers, and equipment operating envelopes [16,17].

Semantic and human-aware navigation methods extend geometric planning by adding context about terrain, work zones, and human activity. For example, researchers have used terrain semantics, BIM, and task knowledge to improve traversability estimation and path selection [18,19]. Human-centered approaches also model worker behavior and interaction patterns to improve safety during close collaboration [11–13].

2.3 Learning-based Navigation

Learning-based navigation aims to reduce manual rule design by training policies from data. Reinforcement learning has shown strong performance in simulated navigation, but it can be sample-intensive and safety-critical when applied on real robots [14,15,20]. IL provides a safer alternative because it learns directly from expert behavior and can be trained offline [21].

Recent benchmarks also highlight the importance of data and evaluation for navigation under human motion and social constraints. SCAND, for example, provides human demonstration data for social navigation and reports 8.7 hours of on-robot trajectories collected with joystick teleoperation [22]. Ratatouille further shows that careful IL design can reduce collisions and improve success without changing the data [23]. These findings support the use of demonstration-driven policies when direct exploration is unsafe or impractical.

For legged platforms, hierarchical designs that separate locomotion from high-level navigation can improve robustness in dynamic environments. Learning to Walk by Steering learns high-level navigation commands from IL while learning low-level gait skills with reinforcement learning, which enables perceptive

quadrupedal locomotion in dynamic scenes [21]. However, most learning-based benchmarks focus on indoor or pedestrian environments and do not capture the semantic and operational constraints of construction sites. SemantI-Nav addresses this gap by training a goal-conditioned policy from multi-view RGB-D demonstrations collected in a high-fidelity virtual construction site.

3 Methodology

3.1 Overview of SemantI-Nav

In this study, the learning target is local planning and control rather than explicit long-horizon path planning. A separate global planner is assumed to provide the goal or waypoint signal, while the policy learns how to generate safe short-horizon motion commands under semantic and dynamic site conditions.

SemantI-Nav follows a learning-from-demonstration pipeline to acquire navigation behaviors that are robust to the semantic and social complexity of construction sites. The key idea is to replace hand-crafted navigation rules with an IL policy that maps multimodal observations and a goal specification to continuous velocity commands for a quadruped robot. The policy is trained to reproduce expert control behaviors, so that it can implicitly learn safety-relevant responses such as slowing down near obstacles, selecting feasible passageways in clutter, and yielding under dynamic interactions.

This section introduces SemantI-Nav, a goal-conditioned IL method for autonomous navigation of a quadruped robot on construction sites. The method combines a high-fidelity virtual construction environment, teleoperated expert demonstrations, and a multimodal policy network that maps perception and robot state to local velocity commands. Figure 1 summarizes the policy architecture used in this study.

The pipeline consists of three stages. First, expert navigation demonstrations are collected while a human operator teleoperates the robot to reach specified goals in diverse site scenarios. During each episode, the system records synchronized multi-view RGB-D observations, compact robot state signals, and a relative goal representation, together with the expert velocity commands. Second, these demonstrations are converted into supervised learning pairs and used to train a neural policy that performs multi-view perception and goal-conditioned action prediction. The training objective minimizes the discrepancy between predicted and demonstrated velocity commands, with data augmentation and goal resampling used to improve robustness under limited data.

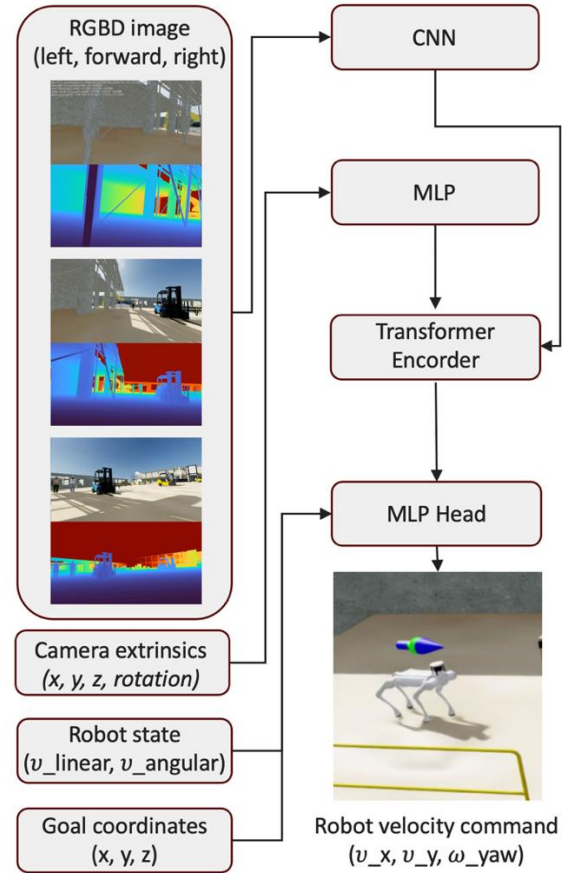


Figure 1. SemantI-Nav policy network for goal-conditioned navigation.

Finally, the trained policy is evaluated in different site scenarios that include both static clutter and dynamic agents (e.g., workers or moving machinery). We use modified site layouts, different worker and machinery trajectories, and different robot start-goal configurations from those used during demonstration collection. At deployment time, the policy runs in a closed loop: it consumes current observations and the goal representation, predicts velocity commands at a fixed control frequency, and continuously updates actions until the goal is reached or a termination condition is triggered. Performance is quantified using task-level and safety-related metrics, including success rate, time to goal, travel distance, and minimum distance to the goal.

3.2 Model Architecture

The policy network follows a compact “multi-view perception + fusion + control head” design. Each RGB-D image is encoded by a shared lightweight CNN composed of five stride convolution blocks with GroupNorm and SiLU activations; global average pooling produces a 256D feature per view, followed by a linear projection to the final image feature dimension.

This encoder is applied identically to the front/left/right cameras, producing three feature tokens per time step.

To make the multi-view fusion aware of camera geometry, each camera token is augmented with a learned embedding of fixed camera extrinsic. In the implementation, the embedding is built from a 3D camera position plus a yaw representation encoded as $(x, y, z, \sin(\text{yaw}), \cos(\text{yaw}))$, with predefined yaws for the front camera (0°), left camera ($+60^\circ$), and right camera (-60°). A small MLP maps these extrinsic descriptors into the same token dimension as the image features, and the embedding is added to each per-camera token before fusion. Although the camera rig is fixed, we use extrinsic-based embeddings instead of a pure view-index embedding because they provide geometrically meaningful cues rather than only arbitrary camera identities. This allows the fusion module to distinguish not only which camera produced a token, but also where that camera is located and how it is oriented relative to the robot.

Cross-camera fusion is performed by a Transformer encoder operating over the three camera tokens. The

zdefault configuration uses a Transformer Encoder with $\text{tf_layers} = 2$ and $\text{tf_heads} = 4$, producing updated tokens that exchange information across views. Finally, the fused visual vector is concatenated with the 6D robot state (linear and angular velocity) and the goal vector and passed through an MLP head with dropout to regress the 3D action command.

4 Implementation and Evaluation Setup

4.1 Virtual Construction Environment

We created a high-fidelity virtual construction site in



Figure 3. Teleoperation demonstration collection

NVIDIA Isaac Sim to represent an active jobsite with complex geometry and clutter as shown in Figure 2. The site spans approximately $110 \text{ m} \times 110 \text{ m}$ and includes representative areas such as partially built structures, material staging zones, and access corridors. The environment also contains common safety controls, such as barriers and railings, to reflect the symbolic boundaries that influence human navigation decisions. We organized all scene assets into semantic groups to support systematic simulation control and reproducible experiments. The groups include temporary works, materials and inventory, permanent works, equipment and machinery, site infrastructure and utilities, safety systems and controls, and workers. This grouping allows us to enable or disable subsets of assets and to define consistent collision filters during validation. Then we collected data in this virtual environment as shown in Figure 3.

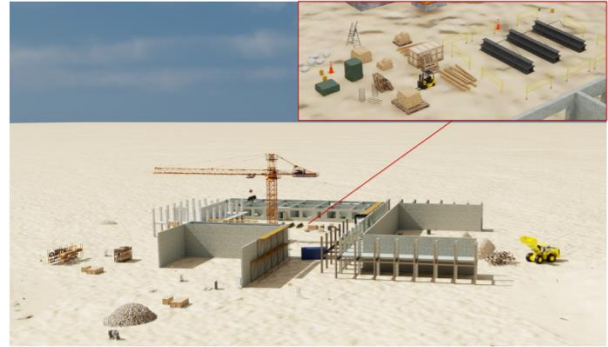


Figure 2. High-fidelity virtual construction site in Isaac Sim ($110 \text{ m} \times 110 \text{ m}$)

4.2 Policy Training Procedure

We trained SemantI-Nav with supervised behavioral cloning to predict base velocity commands from observations and a goal signal. At each timestep, the policy receives multi-view RGB-D images and low-dimensional robot state, together with the relative goal representation. The model outputs a three-dimensional action vector $(v_x, v_y, \omega_{\text{yaw}})$ that matches the operator’s command at that timestep.

To make effective use of limited demonstrations, we used goal relabeling and data augmentation during training. Instead of always training on a fixed-distance goal, we relabeled training targets by sampling goal distances from 1 m to 20 m along the recorded trajectories. This expands the range of goal conditions without collecting new data and helps the policy generalize to longer navigation tasks. We also applied

lightweight appearance and sensor-noise augmentations to improve robustness to visual and depth variation.

4.3 Validation Setup and Scenarios

We evaluated the learned policy in three scenario families that isolate key interaction types on construction sites as shown in Figure 4. For each scenario, we ran 5 episodes. The static scenario tests navigation through

cluttered layouts with stationary materials, tools, and safety markers. The dynamic-people scenario introduces walking workers with crossing, head-on, and same-direction motion relative to the robot. The dynamic-machinery scenario introduces moving equipment, such as a truck, with analogous motion patterns. Across scenarios, the start and goal were placed to create a nominal goal distance of about 10 m.

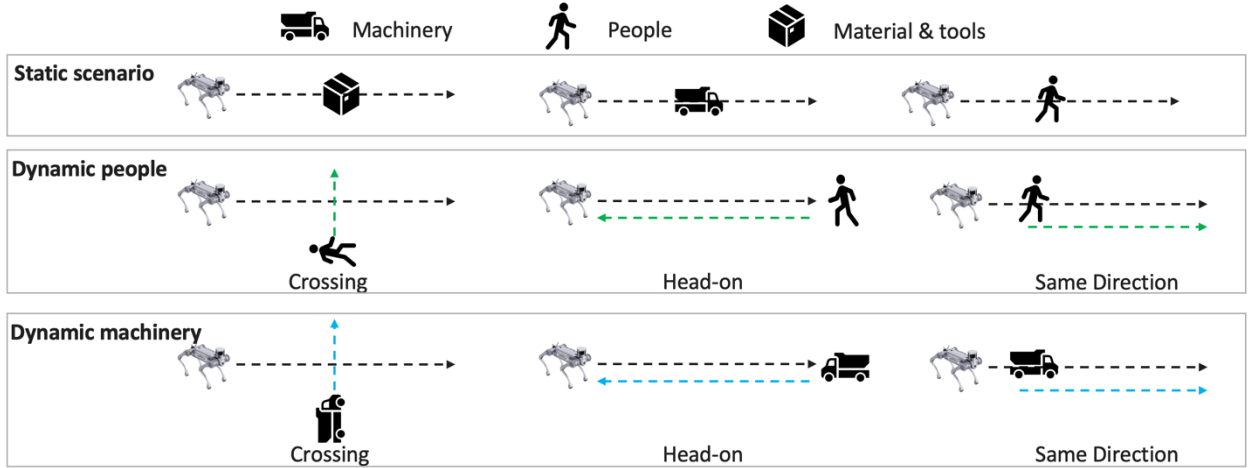


Figure 4. Validation scenario suite for static layouts, dynamic-worker interactions, and dynamic-machinery interactions (crossing, head-on, and same-direction).

To contextualize performance, we also recorded a human operator baseline on the same validation tasks. The operator teleoperated the quadruped under the same scenario configurations, start-goal settings, and dynamic agent behaviors. During evaluation, the operator was provided with the same egocentric visual input as the robot through the robot-mounted camera views and did not have an additional global site view. The human operator was the same person who provided the expert demonstrations and was familiarized with the simulated environment before evaluation. However, we used one trial per scenario configuration to reduce potential memory effects that could arise if the same operator repeated the same task multiple times. This baseline provides a practical reference for task efficiency and helps separate limitations of the learned policy from constraints imposed by the environment.

We report episode-level results using metrics that capture reliability and efficiency. A trial is marked successful when the robot reaches within the specified goal tolerance, and it is marked as a timeout when it exceeds the maximum episode duration. For each trial, we record navigation time, traveled distance, and the travel distance over the episode. We also log collisions

between the robot and non-ground objects to identify safety-critical failures. To assess safety, we log collisions between the robot and non-ground objects, including workers, machinery, and site obstacles. In the current study, safety is defined operationally as collision-free task execution; no additional minimum-clearance threshold or explicit rule-based safety criterion is imposed during evaluation.

5 Results and Discussion

5.1 Quantitative Results

This section summarizes navigation performance of SemantI-Nav and a human operator across the three scenario families described in Section 4.5. Table 1 reports success rate, navigation time, and traveled distance, where each entry is presented as SemantI-Nav / human operator. These metrics reflect both task completion and efficiency under static clutter, dynamic workers, and dynamic machinery.

In terms of task completion, SemantI-Nav achieves high success in all scenarios but remains below the human operator. The agent succeeds in 93.2% of static trials, 92.4% of dynamic-people trials, and 90.3% of

dynamic-machinery trials, whereas the operator succeeds in 100% of trials. Failures are primarily associated with timeouts or collisions, which reflect challenges in tight spaces and in resolving right-of-way interactions.

Efficiency metrics show that the human operator consistently reaches the goal faster and along shorter paths in static and dynamic-people settings. In static scenes, SemantI-Nav requires 17.7 s and travels 18.2 m on average, compared with 15.3 s and 14.4 m for the operator. In dynamic-people scenes, SemantI-Nav requires 17.4 s and travels 16.9 m, compared with 14.3 s and 13.2 m for the operator. In dynamic-machinery scenes, the two are comparable in time (19.8 s vs. 19.9 s) and closer in traveled distance (17.5 m vs. 16.0 m), suggesting that the learned policy can handle structured vehicle interactions but still makes suboptimal detours in clutter and pedestrian traffic.

Table 1 Navigation Performance Comparison: SemantI-Nav Agent (average of 5 trials) vs. Human Operator

Scenario	Static layout	Dynamic people	Dynamic machinery
Success (%)	93.2/ 100	92.4/ 100	90.3/ 100
Nav time (s)	17.7/ 15.3	17.4/ 14.3	19.8/ 19.9
Trav dist (m)	18.2/ 14.4	16.9/ 13.2	17.5/ 16.0

The results indicate that SemantI-Nav maintains a high task success rate across all scenarios, including interactions with dynamic workers and moving machinery. This is important for construction deployment because safe navigation is primarily defined by consistent goal-reaching behavior under changing traffic conditions and limited visibility. In dynamic machinery settings, the agent achieves performance comparable to the human operator in navigation time and travel distance, suggesting that imitation learning can capture conservative yielding and speed regulation strategies around large equipment. In contrast, the higher navigation time and longer travel distance in static and worker-traffic scenarios suggest residual inefficiency, which is preferable to unsafe aggressiveness but still indicates room for improvement in path efficiency and anticipation.

5.2 Discussion and Failure Analysis

This section interprets the validation results and explains the observed gap between SemantI-Nav and an expert human operator. We link the quantitative metrics to qualitative behaviors and discuss what they imply about learning navigation policies for semantically rich, dynamic construction scenes.

Overall performance is strong across all scenarios. SemantI-Nav achieves success rates above 90% in static layouts, dynamic people interactions, and dynamic machinery interactions (Table 1). This indicates that the

policy learned reusable navigation behaviors from teleoperated demonstrations and can execute them under different traffic patterns and obstacle layouts. The result is encouraging because the environment includes heterogeneous materials, safety fixtures, and symbolic boundaries that are difficult to capture with hand-tuned costs or rules.

The main performance difference from the human operator is efficiency in relatively benign settings. In both the static scenario and the dynamic people scenario, the agent takes longer and travels farther than the human operator (Table 1). This suggests that the policy makes extra heading corrections and small detours. The behavior is consistent with a short-horizon velocity policy that must infer longer-horizon intent from local observations. When the correct decision requires early commitment, a purely reactive controller can delay the decision and then compensate late, which increases path length.

Dynamic machinery interaction is the most safety-critical setting and is therefore the most informative stress test. In this scenario, SemantI-Nav has a navigation time close to the human operator, but its success rate is slightly lower than in the static case (Table 1). This pattern suggests that the agent can move at a competitive speed once it selects a yielding or passing strategy. However, it is still sensitive to rare interaction geometries, such as tight clearances near moving vehicles or brief occlusions that hide the vehicle until the last moment. Improving robustness in these corner cases is essential for deployment.

Failure analysis reveals three recurring modes that explain most unsuccessful trials. First, the agent can circle or oscillate near the goal when depth cues are ambiguous and the heading correction becomes over-reactive. Second, the agent can become overly conservative in narrow passages, which slows progress and increases the likelihood of timeouts. Third, the agent can collide with lightweight site objects such as boxes or pallets. This third mode is strongly influenced by the demonstration distribution (Figure 5). During teleoperation, minor contacts in tight spaces may be acceptable and cause no damage, so the dataset can include aggressive maneuvers that reduce clearance. The learned policy can reproduce this behavior even when collision-free navigation is desired.



Figure 5. Failure mode: robot collide with box

These observations point to several targeted improvements that are compatible with the proposed pipeline. First, demonstration curation can filter segments with contacts, and it can label near-miss events to increase sensitivity to clearance. Second, a safety-oriented auxiliary objective can penalize actions that reduce predicted clearance to nearby obstacles. This encourages earlier avoidance manoeuvres and reduces late corrections. Third, a hierarchical strategy can separate global intent from local control by providing a longer-horizon goal signal or by combining the learned policy with a lightweight global planner for wayfinding. Finally, adding more demonstrations in high-risk geometries, such as narrow corridors and passing vehicles, should reduce the long-tail failure cases without requiring real-world data.

This study has several limitations that also define a clear research agenda. First, all experiments are conducted in simulation and within a single high-fidelity construction-site model, so real-world transfer is not evaluated in this paper. Second, the demonstration dataset is limited in both scale and diversity, which may constrain the generalizability of the learned policy. Third, safety in the current study is evaluated mainly through collision-free task execution, which does not fully capture the broader notion of safety in construction environments. Construction-site safety also involves domain-specific rules and contextual constraints, such as avoiding machinery operating zones, avoiding prolonged occupancy of machinery blind spots, and following site-specific interaction practices, and evaluation of such behaviors may require review by personnel with construction safety expertise. These aspects are not explicitly assessed in the present study. Nevertheless, simulation enables controlled variation of layouts and traffic conditions and supports repeatable comparisons with a human operator under identical tasks.

Future work will validate deployment on a physical quadruped robot under appropriate safety protocols. We will also expand the training dataset by collecting more demonstrations across a broader range of sites and operating conditions, incorporating demonstrations from multiple human experts, and leveraging data generated by planners with privileged information to augment supervision at scale. In addition, we will develop more comprehensive safety evaluation criteria tailored to construction environments.

6 Conclusion

The U.S. construction industry faces persistent labour shortages that drive delays and productivity loss. Quadruped robots can support jobsite work because they traverse uneven terrain, stairs, and confined spaces.

However, safe navigation remains difficult in dynamic, semantically rich sites. Many existing methods rely on rules or occupancy-grid cost maps that simplify scenes into “passable” and “impassable” space, which can ignore symbolic boundaries, changing work zones, and equipment operating envelopes. As a result, it is impractical to hand-code complete safety behaviours.

This paper proposes SemantI-Nav, an IL approach that learns safe navigation behaviors directly from expert demonstrations. We build a high-fidelity virtual construction site and collect multimodal demonstrations that capture how skilled operators regulate speed in narrow passages, yield to moving machinery and nearby workers, and comply with symbolic safety markers. We then train a policy that produces context-sensitive velocity commands from multi-view RGB-D observations and robot state. We evaluate the learned policy in multiple validation scenarios, including static layouts, dynamic workers, and dynamic machinery, using success rate, navigation time, path length, and minimum distance to the goal. Results show that SemantI-Nav achieves high success across scenarios and reaches performance comparable to a human operator in dynamic machinery settings, while remaining less efficient in some static and human-traffic cases.

References

- [1] 2024 Workforce Survey Analysis, Associated General Contractors of America, Arlington, VA, USA, n.d. https://www.agc.org/sites/default/files/Files/Communications/2024_Workforce_Survey_Analysis.pdf (accessed July 13, 2025).
- [2] 2024 Workforce Survey Results, Associated General Contractors of America, Arlington, VA, USA, n.d. https://www.agc.org/sites/default/files/users/user21902/2024_Workforce_Survey_National_FINALIZED.pdf (accessed July 13, 2025).
- [3] X. Ma, C. Mao, G. Liu, Can robots replace human beings? —Assessment on the developmental potential of construction robot, *Journal of Building Engineering* 56 (2022) 104727. <https://doi.org/10.1016/j.job.2022.104727>.
- [4] H. Yue, Y. Sun, N. Zeng, S. Chen, Y. Tan, Q. Wang, Legged Robots for Construction Management: Applications and Challenges, *Journal of Construction Engineering and Management* 151 (2025) 03125006. <https://doi.org/10.1061/JCEMD4.COENG-16185>.
- [5] P. Biswal, P.K. Mohanty, Development of quadruped walking robots: A review, *Ain Shams Engineering Journal* 12 (2021) 2017–2031. <https://doi.org/10.1016/j.asej.2020.11.005>.

- [6] S. Halder, K. Afsari, J. Serdakowski, S. DeVito, R. King, Accuracy Estimation for Autonomous Navigation of a Quadruped Robot in Construction Progress Monitoring, (2022) 1092–1100. <https://doi.org/10.1061/9780784483893.134>.
- [7] A. Aziz, P. Herbers, H. Bayer, M. König, Fully Autonomous Fire Safety Equipment Inspection Missions on a Legged Robot, (2024) 804–812. <https://doi.org/10.1061/9780784485224.097>.
- [8] Safe and Robust Motion Planning for Autonomous Navigation of Quadruped Robots in Cluttered Environments | IEEE Journals & Magazine | IEEE Xplore, (n.d.). <https://ieeexplore-ieee-org.ezproxy.library.wisc.edu/document/10531249> (accessed September 3, 2025).
- [9] T. Guan, Z. He, R. Song, D. Manocha, L. Zhang, TNS: Terrain Traversability Mapping and Navigation System for Autonomous Excavators, (2022). <https://doi.org/10.48550/arXiv.2109.06250>.
- [10] I. Jeong, Y. Jang, J. Park, Y.K. Cho, Motion Planning of Mobile Robots for Autonomous Navigation on Uneven Ground Surfaces, *Journal of Computing in Civil Engineering* 35 (2021) 04021001. [https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0000963](https://doi.org/10.1061/(ASCE)CP.1943-5487.0000963).
- [11] S. Halder, K. Afsari, E.K. Chiou, R.N.C. Patrick, K. Hamed, Construction inspection & monitoring with quadruped robots in future human-robot teaming: A preliminary study, *Journal of Building Engineering* (2023). <https://doi.org/10.1016/j.jobe.2022.105814>.
- [12] Y. Jang, W. Seol, K. Lee, K.-S. Kim, S. Kim, Development of quadruped robot for inspection of underground pipelines in nuclear power plants, *Electronics Letters* 58 (2022) 234–236. <https://doi.org/10.1049/ell2.12414>.
- [13] H. Kolvenbach, D. Wisth, R. Buchanan, G. Valsecchi, R. Grandia, M. Fallon, M. Hutter, Towards autonomous inspection of concrete deterioration in sewers with legged robots, *Journal of Field Robotics* 37 (2020) 1314–1327. <https://doi.org/10.1002/rob.21964>.
- [14] X. Zhao, Y. Jin, N.M. Selvaraj, M. Ilyas, C.C. Cheah, Platform-independent visual installation progress monitoring for construction automation, *Automation in Construction* 154 (2023) 104996. <https://doi.org/10.1016/j.autcon.2023.104996>.
- [15] S. Halder, K. Afsari, J. Serdakowski, S. DeVito, M. Ensafi, W. Thabet, Real-Time and Remote Construction Progress Monitoring with a Quadruped Robot Using Augmented Reality, *Buildings* 12 (2022) 2027. <https://doi.org/10.3390/buildings12112027>.
- [16] P. Kim, J. Chen, J. Kim, Y.K. Cho, SLAM-Driven Intelligent Autonomous Mobile Robot Navigation for Construction Applications, in: I.F.C. Smith, B. Domer (Eds.), *Advanced Computing Strategies for Engineering*, Springer International Publishing, Cham, 2018: pp. 254–269. https://doi.org/10.1007/978-3-319-91635-4_14.
- [17] T. Ren, H. Jebelli, Efficient 3D robotic mapping and navigation method in complex construction environments, *Computer-Aided Civil and Infrastructure Engineering* 40 (2025) 1580–1605. <https://doi.org/10.1111/mice.13353>.
- [18] S. Karimi, R.G. Braga, I. Iordanova, D. St-Onge, Semantic Optimal Robot Navigation Using Building Information on Construction Sites, *ISARC Proceedings 2021 Proceedings of the 38th ISARC, Dubai, UAE (2021)* 57–64. <https://doi.org/10.22260/ISARC2021/0010>.
- [19] R.G. Braga, M.O. Tahir, S. Karimi, U. Dah-Achinanon, I. Iordanova, D. St-Onge, Intuitive BIM-aided robotic navigation and assets localization with semantic user interfaces, *Front Robot AI* 12 (2025) 1548684. <https://doi.org/10.3389/frobt.2025.1548684>.
- [20] Y. Tao, M. Popović, Y. Wang, S.T. Digumarti, N. Chebrolu, M. Fallon, 3D Lidar Reconstruction with Probabilistic Depth Completion for Robotic Navigation, in: *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2022: pp. 5339–5346. <https://doi.org/10.1109/IROS47612.2022.9981531>.
- [21] M. Seo, R. Gupta, Y. Zhu, A. Skoutnev, L. Sentis, Y. Zhu, Learning to Walk by Steering: Perceptive Quadrupedal Locomotion in Dynamic Environments, in: *2023 IEEE International Conference on Robotics and Automation (ICRA)*, 2023: pp. 5099–5105. <https://doi.org/10.1109/ICRA48891.2023.10161302>.
- [22] H. Karnan, A. Nair, X. Xiao, G. Warnell, S. Pirk, A. Toshev, J. Hart, J. Biswas, P. Stone, Socially CompliAnt Navigation Dataset (SCAND): A Large-Scale Dataset of Demonstrations for Social Navigation, *IEEE Robotics and Automation Letters* 7 (2022) 11807–11814. <https://doi.org/10.1109/LRA.2022.3184025>.
- [23] J.R. Han, M. Vanniasinghe, H. Sahak, N. Rhinehart, T.D. Barfoot, Ratatouille: Imitation Learning Ingredients for Real-world Social Robot Navigation, (2025). <https://doi.org/10.48550/arXiv.2509.17204>.