Human-Inspired Inverse Reinforcement Learning for Autonomous Excavator Operations in Lunar Environments

Minguk Kim[1] and Youngjib Ham[2]

[1]Department of Construction Science, Texas A&M University, USA
[2]Department of Civil and Environmental Engineering, Seoul National University, South Korea
mingukkim@tamu.edu, yj.ham@snu.ac.kr

Abstract –

Construction robots must perform complex construction tasks robustly under extreme and unpredictable conditions. Although current reinforcement learning methods, including imitation learning, show promise for this goal, they face inherent challenges, such as time-consuming reward function design or substantial data requirements that limit their generalizability to unstructured environments. These limitations are exacerbated by energy constraints of deep neural network-based algorithms, making them less viable for lunar surface construction, where data collection and power supply are significantly constrained. This study integrates inverse reinforcement learning with neuroscientific theories of human decision-making to equip construction robots with robust autonomy. The proposed approach modularizes a given task into simpler subtasks governed by learned parameters and explicitly models state features such as spatial and temporal distances, enabling the agent to anticipate and adapt to diverse scenarios while reducing computational overhead. Experimental results indicate that the agent modeled by the proposed approach achieved a 73% success rate in a lunar-like simulation environment, exceeding the 30 % success rate of an agent modeled by a conventional method. The findings demonstrate that the proposed approach can enhance adaptability for autonomous infrastructure construction, especially in lunar surface construction missions constrained by minimal human oversight and limited resources. This study contributes to the advancement of autonomous construction robotics, offering a novel and scalable solution for extreme construction settings.

Keywords –

Construction Robotics; Lunar Surface Construction; Inverse Reinforcement Learning; Autonomous Excavation

1 Introduction

Driven by prospects such as abundant resource acquisition and advances in cutting-edge technologies, the global space exploration ecosystem is shifting its paradigm toward human habitation and resource extraction on the Moon [1]. This ambitious shift hinges on developing infrastructure and capabilities to support sustainable operations beyond Earth. However, the lunar environment is harsh, dynamic, and fraught with unpredictable events [2,3], thereby necessitating the deployment of construction robots capable of minimizing human intervention while reliably performing complex construction tasks on the lunar surface.

Nevertheless, it remains challenging to imagine current commercial construction robots operating stably under such dynamic and adverse conditions. These robots have been primarily designed to perform repetitive tasks in factories or other relatively controlled and static environments, thereby depending substantially on human oversight. Indeed, human operators are often required to intervene in unexpected situations directly or engage in indirect collaborative forms, such as sequential task execution. In light of these considerations, the development of fully autonomous construction robots that can function robustly in demanding and unpredictable lunar environmental settings becomes imperative.

Reinforcement learning (RL) approaches, including imitation learning (IL), present a potential avenue to surmount these challenges, as they allow robotic agents to be trained through interactions with their environment to accomplish assigned tasks, either by exploring and exploiting independently or by utilizing cues provided by experts. By leveraging such approaches, robots can potentially deliver robust and adaptive performance, facilitating stable execution of complex construction tasks in dynamic and harsh environments [4]. However, conventional RL methods have inherent limitations, such as the time-consuming process of reward function design and an excessive dependence on training data [5,6].

Prevailing IL methods still require extensive datasets of high-quality demonstrations, even though they circumvent these limitations [7]. These constraints may prevent construction robots from overcoming scenarios that deviate from their training set or from performing reliably in environments characterized by limited resources and accessibility [8]. Moreover, many state-of-the-art algorithms, which inherently rely on deep neural networks (DNN), demand substantial energy consumption. These challenges pose significant obstacles to lunar surface construction where accessibility is highly constrained.

To mitigate these challenges, this study draws inspiration from the remarkable adaptability and responsiveness that humans exhibit across diverse conditions [9]. To embed these capabilities into construction robots, this study proposes an approach grounded in robust biological evidence, integrating a human-like decision-making architecture with inverse reinforcement learning (IRL) framework. Through this integration, the resulting construction robots will be endowed with the capacity to operate robustly under lunar environments, ultimately contributing to the realization of autonomous infrastructure construction in outer space.

2 State of the Art

RL approaches have recently gained traction in construction robotics because of their ability to learn robust and adaptive behaviors through continuous interaction with the environment ([5,10–12]). By updating policies or value functions based on trial-and-error, agents can theoretically handle dynamic and uncertain scenarios, making them appealing candidates for tasks conducted in extreme environmental settings, such as those on the lunar surface. Generally, they employ DNN to approximate policies, and in many instances, they leverage expert trajectories to enhance learning efficiency [13].

However, these predominantly policy-based RL approaches suffer from several drawbacks that can be especially pronounced in extraterrestrial construction contexts. First, they often rely on extensive, high-quality datasets—an issue exacerbated by the inaccessibility of the lunar surface, where data collection is highly constrained [5]. Second, their dependence on DNN results in high energy consumption, necessitating substantial computational capabilities in hardware [4]. Lastly, in conventional RL methods, the need for sophisticated reward function design and the challenge of sparse rewards can make it difficult for agents to obtain the necessary signals to guide their learning in a timely manner [6], particularly when tackling complex, long-duration tasks under extreme conditions. As a result, these challenges can impede the scalability and reliability of construction robots, preventing them from swiftly attaining the necessary performance for diverse and uncertain tasks, such as those encountered in the lunar environment.

In contrast, IRL has not yet been widely explored within construction robotics despite its promising potential to provide computational models of human and animal behavior [14]. Unlike policy-based RL approaches, IRL aims to uncover the underlying reward structure from expert demonstrations, thereby distilling implicit cues that can guide learning toward higher-level objectives [15,16]. This characteristic of IRL makes it especially attractive for complex tasks where reward design is difficult and for scenarios where sample efficiency is paramount due to data collection constraints. In lunar construction, these strengths of IRL could enable robust adaptation to diverse or novel circumstances, alleviating the burdens of large-scale data gathering and intricate reward engineering.

A particularly relevant extension is modular IRL, which draws from neuroscientific findings indicating that humans decompose tasks into simpler subtasks that can be evaluated separately [17]. By framing a complex construction activity as multiple subtasks, the approach can assign simplified Q-values to each subgoal. These modular Q-values can then be estimated using Bayesian IRL [18]. For instance, Zhang et al. [19] demonstrated the potential of such an approach for reproducing human sensorimotor decisions by incorporating a discounting factor that regulates impulsive behavior in virtual reality settings. Compared to policy-based RL approaches, this value-based technique shows greater sample efficiency and a reduced computational footprint [4], making them particularly well suited for physically and energetically constrained environments like the Moon. Nevertheless, while modular IRL has proven its capabilities primarily in spatial navigation tasks, its broader application to industrial contexts, including automation in construction, has yet to be fully explored.

In summary, contemporary research into RL-based robotic control indicates great potential for autonomous operation in construction tasks. Still, multiple hurdles, ranging from data scarcity to energy constraints, must be addressed to realize robust performance in lunar environments. In this regard, an IRL-based approach integrating human-like decision-making processes can be a promising way to overcome the hurdles. Building on these insights, the next section proposes a novel approach that integrates modular IRL to endow construction robots with robust, adaptive capabilities suited to the rigors of lunar surface construction.

3 Methods

Lunar construction imposes unique constraints due to limited on-site supervision, power restrictions, and complex terrain. To demonstrate the proposed IRL-based approach under these demanding conditions, this study focuses on an excavator swing operation, a frequent maneuver for material transport and debris clearance. Continuous horizontal swinging has been identified as a primary source of struck-by hazards on Earth-based sites, and this risk is magnified in off-Earth contexts where immediate human intervention is challenging [20]. As such, the excavator swing task represents an ideal testbed for validating an autonomous decision-making scheme that effectively balances safety and task efficiency.

3.1 Problem Formulation

Initially, the excavator's swing is modeled using a Markov Decision Process (MDP):

- States (S): Each state describes the excavator's relevant kinematic features (joint angles and velocities of boom, stick, cab) and key environmental cues (e.g., nearby obstacles, target locations).
- Actions (A): Discrete maneuvers include moving the boom up/down, the stick in/out, rotating the cab left/right, or performing no movement. Actions that exceed joint limits or risk an immediate collision are excluded from consideration.
- Reward (R): A single reward function penalizes collisions and rewards movement toward a defined dump area, implicitly treating all safety and positioning objectives as one.
- Discount Factor (γ): A scalar in (0,1) that weighs future returns; γ near 1 values cautious, long-term strategies, whereas smaller γ emphasizes immediate outcomes.

A SoftMax policy, as shown in Equation (1), governs action selection:

$$\pi(s,a) = \frac{\exp\left(Q(s,a)/\eta\right)}{\sum_{a'\in A}\exp\left(Q(s,a')/\eta\right)} \tag{1}$$

where $Q(s,a)$ is the action-value function, and $\eta \in (0,1]$ manages exploration versus exploitation. Although this single-task MDP is straightforward, it does not fully capture human-like strategies that partition tasks into sub-goals and consider spatial and temporal dimensions when avoiding collisions.

3.2 Task Modularization with Explicit State Features

Drawing on empirical findings in human motor control and decision-making [17,19], a main task is replaced with multiple subtasks. For instance, the swing operation might split into:

- Obstacle Avoidance (Subtask 1): Minimizing collisions.
- Targeted Material Placement (Subtask 2): Reaching or depositing material in the designated area.

Each subtask n is assigned:

- A subtask weight $r^{(n)} \in, with\ \sum_n r^{(n)} = 1$.
- A unit reward $u^{(n)}$, reflecting baseline significance for that objective.
- A sub-discount factor $\gamma^{(n)} \in (0,1)$, enabling each subtask to handle near- vs. long-term gains differently.

Hence, the total action-value $Q(s,a)$ is shown in Equation (2).

$$Q(s,a) = \sum_{n=1}^{N} Q^{(n)}(s^{(n)},a) \tag{2}$$

where $Q^{(n)}$ focuses on a particular sub-goal. This modularization can make the overall decision process more transparent and potentially more adaptable, as each subtask can be updated independently.

Beyond subdividing tasks, psychological distance [21] is introduced to emulate how skilled operators anticipate future collisions or time-critical steps:

- Spatial Distance: The Euclidean distance between the excavator's bucket and key objects (e.g., a rock, the dump area).
- Temporal Distance: The estimated time until a collision (or successful arrival) if a chosen action is executed.

These distances are computed at each timestep, using a brief forward simulation. By incorporating distance-based features explicitly, one can reduce reliance on DNN, thus lowering computational demand—a crucial advantage in energy-constrained lunar construction missions. This approach also strengthens safety, as the agent "looks ahead" before finalizing an action.

3.3 Parameter Estimation

The following steps outline the process of estimating subtask weights $r^{(n)}$ and discount factors $\gamma^{(n)}$ from expert demonstrations:

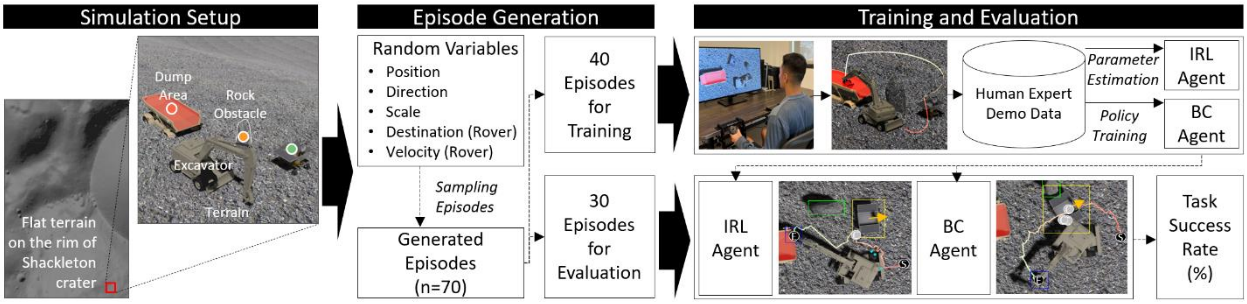

Figure 1. Overall experimental steps for training and evaluating IRL and BC agents in a simulation environment

1. Data Collection: Human experts run the excavator in a simulation, completing the swing operation multiple times under simulation environments. This yields state–action pairs $\{(s_1, a_1), \dots, (s_T, a_T)\}$, where T is the end of an episode.
2. Likelihood Computation: For a trial parameter set $\Theta = \{r_1, \dots, r_N, \gamma_1, \dots, \gamma_N\}$, the likelihood $\mathcal{L}$ of human expert demonstration data is calculated by Equation (3), where Q sums up each subtask's value function $Q^{(n)}$, weighted by $r^{(n)}$ and discounted by $\gamma^{(n)}$.

$$\mathcal{L} = \prod_{t=1}^{T} \frac{exp^{Q(s_t,a_t)\eta}}{\sum_{a\in A} exp^{Q(s_t,a)/\eta}} \tag{3}$$

3. Parameter Set Sampling: Each subtask's reward and discount factor are estimated by a sampling method. This study uses a Hamiltonian Monte Carlo (HMC) technique to explore the parameter set spaces. Key constraints, $\sum_n r^{(n)} = 1$ and $0 < r^{(n)}, \gamma^{(n)} < 1$, are enforced to ensure valid solutions.
4. Convergence to Final Parameter Set: Parameter sets aligning closely with expert actions are accepted, leading to a convergent estimate $\widehat{\Theta}$. This final parameter set grants the excavator robot agent.

4 Experimentation

4.1 Experimental Goal and Setup

The primary goal of this experiment is to determine whether the IRL agent modeled by the proposed IRL-based approach can robustly execute a construction task compared to an agent modeled by behavior cloning (BC), a conventional IL method. To do so, as shown in Figure 1, A total of 70 episodes were devised in a simulated lunar environment constructed based on real lunar terrain data [22] in the Unity game engine [23], with 40 episodes for training and 30 episodes for evaluation. Each episode included five key objects—an excavator, a dump truck, a rock obstacle, a moving rover, and the surrounding terrain, which are randomly arranged to generate varied spatial layouts and dynamics using the Latin Hypercube Sampling method [24]. To ensure dynamic interaction throughout each episode, constraints were applied during the generation process. These constraints included maintaining a minimum distance between the moving rover and other objects, as well as selecting episodes where the rover's linear path did not result in collisions with other objects, thus preserving the intended episode dynamics. Episodes are terminated upon the excavator bucket reaching the designated dump area or upon reaching the time limit of 150 timesteps, which is approximately three times longer than an expert requires. The workstation on which the experiments were conducted has the following specifications: Intel Xeon W-2123 CPU, 64 GB RAM, NVIDIA GeForce RTX 2080 GPU.

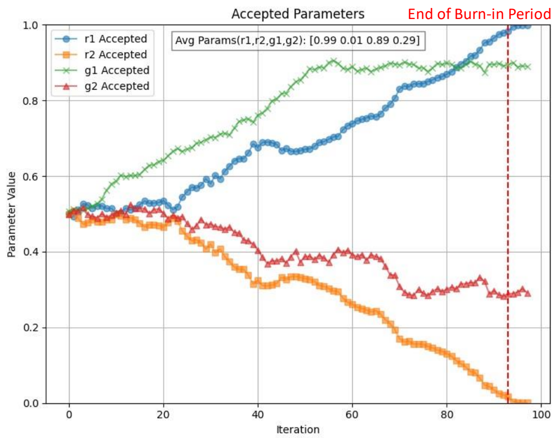

Figure 2. Accepted parameter trajectories during HMC sampling

4.2 Experimental Procedure

Human expert demonstrations were recorded in a simulation equipped with joystick controls and a first-person view. Each demonstration began with five seconds of situational awareness from an overhead perspective, allowing participants to plan their route

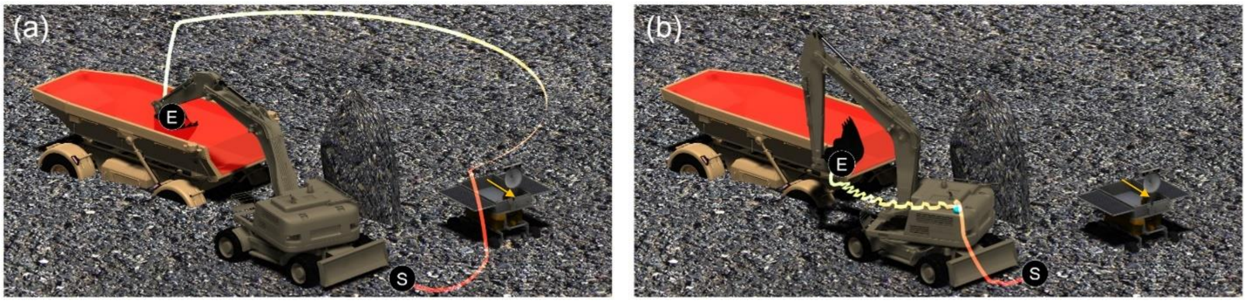

Figure 3. Trajectory comparison of a successful episode between (a) a human expert and (b) the corresponding IRL agent employing different detour strategies

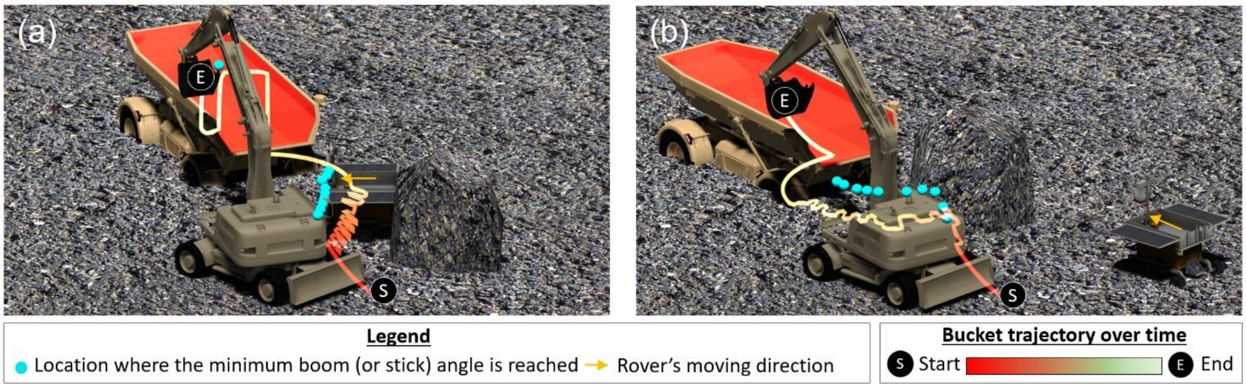

Figure 4. Failure cases of IRL agents: (a) and (b) illustrate inefficient movement leading to episode termination

[25,26]. Participants then navigated the excavator bucket to the dump area using what they perceived as the safest, most efficient path. The collected state–action pairs formed the training data for parameter estimation in the IRL agent and supervised training in the BC agent.

Afterward, the IRL agent's parameters, $\Theta = \{r^{(1)}, r^{(2)}, \gamma^{(1)}, \gamma^{(2)}\}$, were inferred from the recorded demonstration data via HMC. After discarding an initial burn-in, the final accepted samples were aggregated to define a robust parameter set. Figure 2 illustrates the evolution of these parameters over iterations, highlighting their convergence after the burn-in phase. On the other hand, the BC agent was trained using a neural network architecture with optimal hyperparameters determined through a grid search: dropout probability of 0.1, learning rate of 1e-4, neural network dimension of 256, and depth of 3.

Both agents were subsequently evaluated in the 30 remaining episodes, and the task success rate was measured as the proportion of runs where the bucket reached its target without a collision.

5 Results and Discussion

5.1 Robustness Compared to BC Agent

The IRL agent demonstrated a success rate of 73%, significantly outperforming the BC agent, which achieved only 30%. This notable difference underscores the limitations of relying solely on policy imitation, as seen in the BC agent's performance in dynamic environments. In contrast, the IRL agent exhibited superior adaptability to diverse layouts, including a moving rover. Additionally, the IRL agent achieved a traveled distance of 15.71 meters on average (S.D.=3.91), compared to the BC agent's average traveled distance of 17.98 meters (S.D.=4.61). These findings highlight that the proposed IRL-based approach enables construction robots to robustly perform construction tasks more efficiently than the existing RL methods. This finding aligns with prior literature emphasizing that conventional policy-based IL methods can struggle in unseen or high-variability contexts [8]. On the other hand, the IRL agent exhibited robust performance and more efficient path selection.

Interestingly, as shown in Figure 3, the IRL agent often exhibited detour strategies that differed from human experts. Both tended to detour around obstacles; however, the strategy employed diverged. While the human expert consistently chose the safest and farthest routes to avoid collisions, as depicted in Figure 3(a), the IRL agent opted to detour in the opposite direction, as shown in Figure 3(b). This strategic variation resulted in the IRL agent adopting a more efficient trajectory, reducing the required timesteps and minimizing redundant movements. Consequently, the IRL agent not only maintained a higher success rate but also enhanced overall task efficiency compared to the BC agent.

Figure 4 illustrates two distinct failure cases of the

IRL agent. In Figures 4(a) and 4(b), the IRL agent repeatedly reaches the minimum boom angle without progressing toward the target, leading to the termination of the episode at the maximum timestep limit. These failures occur not due to collisions but because the agent's movements become inefficient, highlighting a lack of comprehensive exteroceptive perception of the excavator's body. This inefficiency suggests that the IRL agent's decision-making process does not fully account for the excavator's physical constraints, leading to suboptimal movement strategies.

These observations underscore the critical role of parameter diversity within the IRL framework. Unlike traditional IL approaches that merely replicate observed behaviors, the proposed IRL approach models human-like decision-making strategies through modular subtask decomposition and explicit distance-based state features. This enables the agent to develop its strategic approaches to overcome harsh and complex environments and tasks. This capability to generate unique strategies, rather than simply imitating human actions, highlights the advantage of the proposed modular IRL approach in fostering more resilient and effective autonomous agents.

5.2 Summary and Implications

This study illustrates how the proposed IRL-based approach that models human-like decision-making can outperform a traditional IL approach in a demanding robotic construction task, echoing the importance of deeper reward representations and highlighting that direct behavior imitation can lead to failures when the environment is dynamic or partially observable.

In practical terms, the proposed approach could be adopted in extraterrestrial construction scenarios or similarly constrained environments where robotic autonomy must manage unpredictable conditions. By leveraging task modularization and employing explicitly modeled state features, the IRL agent demonstrated adaptability and robustness that could reduce mission risks and resource waste. In this regard, autonomous construction robots modeled by the proposed approach could perform construction tasks (i.e., excavation or debris clearance) with minimal human oversight, saving time and costs.

Several assumptions and abstractions were made to manage complexity and computational overhead. The excavation task was described with a limited number of subtasks, which may not capture all subtasks of the human decision-making process. Additional constraints, including simplified collision models and a uniform gravitational field, might not reflect every aspect of lunar physics. These simplifications allowed the method to converge more quickly and highlight possible limitations when extending to higher-fidelity simulations or more intricate tasks. Future work might consider expanding the subtask framework or incorporating fine-grained terrain and machine dynamics.

These experimental findings indicate that the proposed IRL-based approach can significantly improve construction robots' robustness. By recognizing both the strengths and constraints of the proposed approach, researchers and practitioners can refine this approach for even more robotic applications in extreme construction settings.

6 Conclusion

This study was motivated by the escalating complexity and risk inherent in construction robotics, particularly in extreme environments like the Moon. The current body of research emphasizes the need for construction robot modeling methods that can adapt to uncertain conditions with minimal human oversight. However, traditional RL approaches often replicate narrow expert behaviors and can fail when the environment diverges from the training set.

In response to these challenges, this study proposes an IRL-based approach integrated with neuroscientific theories of human decision-making. By identifying underlying reward structures rather than imitating actions outright, the IRL agent demonstrated a capacity for more adaptive and reliable decision-making, even in scenarios characterized by moving objects and variable layouts. The proposed approach’s novelty is that it leads to an agent that can synthesize multiple expert strategies and yield robust performance. The experimentations showed that the agent modeled by the proposed IRL-based approach outperformed an agent modeled by BC, a conventional IL approach regarding task success rates under a simulated lunar environment setting.

Several assumptions were made in this study. The limited number of subtasks used in this study simplifies the complexity of real-world construction tasks, which often involve numerous interdependent and dynamic processes. Moreover, the high-level representations of state features may not fully translate to highly unstructured or rapidly changing environments without further customization. Future studies should consider extending the subtask framework, refining state features, or integrating more detailed physics to bridge the gap between simulated experiments and real deployment.

This study demonstrates the effectiveness of an IRL-based approach in automating construction tasks, particularly in scenarios like extraterrestrial excavator operations, where supervision is limited. The integration of human-inspired decision-making with IRL establishes the foundation for the development of robust and adaptable robotic systems capable of thriving in extreme environments. This work contributes to the broader vision of achieving fully autonomous infrastructure development in space by addressing the unique demands of extraterrestrial construction environments.

References

[1] NASA, The Artemis plan, NASA's lunar exploration program overview, 2020. https://go.nasa.gov/3WrOLjW (accessed January 5, 2024).

[2] C. Wong, E. Yang, X.-T. Yan, D. Gu, An overview of robotics and autonomous systems for harsh environments, in: 2017 23rd Int. Conf. Autom. Comput. ICAC, IEEE, Huddersfield, United Kingdom, 2017: pp. 1–6. https://doi.org/10.23919/IConAC.2017.8082020.

[3] D. Jud, P. Leemann, S. Kerscher, M. Hutter, Autonomous free-form trenching using a walking excavator, IEEE Robot. Autom. Lett. 4 (2019) 3208–3215. https://doi.org/10.1109/LRA.2019.2925758.

[4] J. Manuel Davila Delgado, L. Oyedele, Robotics in construction: A critical review of the reinforcement learning and imitation learning paradigms, Adv. Eng. Inform. 54 (2022) 101787. https://doi.org/10.1016/j.aei.2022.101787.

[5] R. Li, Z. Zou, Enhancing construction robot learning for collaborative and long-horizon tasks using generative adversarial imitation learning, Adv. Eng. Inform. 58 (2023) 102140. https://doi.org/10.1016/j.aei.2023.102140.

[6] H.B. Ammar, M.E. Taylor, K. Tuyls, G. Weiss, Reinforcement Learning Transfer Using a Sparse Coded Inter-task Mapping, in: M. Cossentino, M. Kaisers, K. Tuyls, G. Weiss (Eds.), Multi-Agent Syst., Springer Berlin Heidelberg, Berlin, Heidelberg, 2012: pp. 1–16. https://doi.org/10.1007/978-3-642-34799-3_1.

[7] M. Zare, P.M. Kebria, A. Khosravi, S. Nahavandi, A Survey of Imitation Learning: Algorithms, Recent Developments, and Challenges, ArXiv Prepr. (2023) 1–15. https://doi.org/10.48550/arXiv.2309.02473.

[8] K. Lee, B. Vlahov, J. Gibson, J.M. Rehg, E.A. Theodorou, Approximate Inverse Reinforcement Learning from Vision-based Imitation Learning, in: 2021 IEEE Int. Conf. Robot. Autom. ICRA, IEEE, Xi'an, China, 2021: pp. 10793–10799. https://doi.org/10.1109/ICRA48506.2021.9560916.

[9] Y. Qin, T. Menara, S. Oymak, S.N. Ching, F. Pasqualetti, Representation Learning for Context-Dependent Decision-Making, in: Proc. Am. Control Conf., American Automatic Control Council, Atlanta, Georgia, USA, 2022: pp. 2130–2135. https://doi.org/10.23919/ACC53348.2022.9867204.

[10] B. Belousov, B. Wibranek, J. Schneider, T. Schneider, G. Chalvatzaki, J. Peters, O. Tessmann, Robotic architectural assembly with tactile skills: Simulation and optimization, Autom. Constr. 133 (2022) 104006. https://doi.org/10.1016/j.autcon.2021.104006.

[11] C.-J. Liang, V.R. Kamat, C.C. Menassa, W. McGee, Trajectory-Based Skill Learning for Overhead Construction Robots Using Generalized Cylinders with Orientation, J. Comput. Civ. Eng. 36 (2022) 04021036. https://doi.org/10.1061/(ASCE)CP.1943-5487.0001004.

[12] L. Huang, Z. Zhu, Z. Zou, To imitate or not to imitate: Boosting reinforcement learning-based construction robotic control for long-horizon tasks using virtual demonstrations, Autom. Constr. 146 (2023) 104691. https://doi.org/10.1016/j.autcon.2022.104691.

[13] K. You, C. Zhou, L. Ding, Deep learning technology for construction machinery and robotics, Autom. Constr. 150 (2023) 104852. https://doi.org/10.1016/j.autcon.2023.104852.

[14] S. Russell, Learning agents for uncertain environments, in: Proc. Annu. ACM Conf. Comput. Learn. Theory, ACM, New York, NY, USA, 1998: pp. 101–103. https://doi.org/10.1145/279943.279964.

[15] S. Arora, P. Doshi, A survey of inverse reinforcement learning: Challenges, methods and progress, Artif. Intell. 297 (2021) 103500. https://doi.org/10.1016/j.artint.2021.103500.

[16] F. Torabi, G. Warnell, P. Stone, Behavioral cloning from observation, in: IJCAI Int. Jt. Conf. Artif. Intell., Stockholm, Sweden, 2018: pp. 4950–4957. https://doi.org/10.24963/ijcai.2018/687.

[17] C.A. Rothkopf, D.H. Ballard, Modular inverse reinforcement learning for visuomotor behavior, Biol. Cybern. 107 (2013) 477–490. https://doi.org/10.1007/s00422-013-0562-6.

[18] D. Ramachandran, E. Amir, Bayesian inverse reinforcement learning, in: IJCAI Int. Jt. Conf. Artif. Intell., IJCAI, Hyderabad, India, 2007: pp. 2586–2591.

[19] R. Zhang, S. Zhang, M.H. Tong, Y. Cui, C.A. Rothkopf, D.H. Ballard, M.M. Hayhoe, Modeling sensory-motor decisions in natural behavior, PLoS Comput. Biol. 14 (2018) 1–22. https://doi.org/10.1371/journal.pcbi.1006518.

[20] U.S. Department of Labor, Bureau of Labor Statistics, Census of Fatal Occupational Injuries, (2022). https://www.bls.gov/iif/fatal-injuries-

tables/fatal-occupational-injuries-table-a-9-2021.htm (accessed June 30, 2024).

[21] F. Pardo, A. Tavakoli, V. Levdik, P. Kormushev, Time Limits in Reinforcement Learning, in: Proc. Mach. Learn. Res., PMLR, Stockholm, Sweden, 2018: pp. 4045–4054. https://doi.org/10.48550/arXiv.1712.00378.

[22] M.K. Barker, E. Mazarico, G.A. Neumann, D.E. Smith, M.T. Zuber, J.W. Head, Improved LOLA elevation maps for south pole landing sites: Error estimates and their impact on illumination conditions, Planet. Space Sci. 203 (2021) 105119. https://doi.org/10.1016/j.pss.2020.105119.

[23] A. Juliani, V.-P. Berges, E. Teng, A. Cohen, J. Harper, C. Elion, C. Goy, Y. Gao, H. Henry, M. Mattar, D. Lange, Unity: A General Platform for Intelligent Agents, ArXiv Prepr. ArXiv180902627 (2018). https://doi.org/10.48550/arXiv.1809.02627.

[24] M.D. McKay, Latin hypercube sampling as a tool in uncertainty analysis of computer models, in: Proc. 24th Conf. Winter Simul. - WSC 92, ACM Press, Arlington, Virginia, United States, 1992: pp. 557–564. https://doi.org/10.1145/167293.167637.

[25] H.A. Taylor, S.J. Naylor, N.A. Chechile, Goal-specific influences on the representation of spatial perspective, Mem. Cognit. 27 (1999) 309–319. https://doi.org/10.3758/BF03211414.

[26] D. Sirkin, N. Martelaro, M. Johns, W. Ju, Toward Measurement of Situation Awareness in Autonomous Vehicles, in: Proc. 2017 CHI Conf. Hum. Factors Comput. Syst., ACM, Denver Colorado USA, 2017: pp. 405–415. https://doi.org/10.1145/3025453.3025822.