

Reinforcement Learning for Wheeled Mobility on Vertically Challenging Terrain

Tong Xu, Chenhui Pan, and Xuesu Xiao

Abstract—Off-road navigation on vertically challenging terrain, involving steep slopes and rugged boulders, presents significant challenges for wheeled robots both at the planning level to achieve smooth collision-free trajectories and at the control level to avoid rolling over or getting stuck. Considering the complex model of wheel-terrain interactions, we develop an end-to-end Reinforcement Learning (RL) system for an autonomous vehicle to learn wheeled mobility through simulated trial-and-error experiences. Using a custom-designed simulator built on the Chrono multi-physics engine, our approach leverages Proximal Policy Optimization (PPO) and a terrain difficulty curriculum to refine a policy based on a reward function to encourage progress towards the goal and penalize excessive roll and pitch angles, which circumvents the need of complex and expensive kinodynamic modeling, planning, and control. Additionally, we present experimental results in the simulator and deploy our approach on a physical Verti-4-Wheeler (V4W) platform, demonstrating that RL can equip conventional wheeled robots with previously unrealized potential of navigating vertically challenging terrain.

I. INTRODUCTION

Autonomous off-road navigation has various safety, security, and rescue applications, such as search and rescue missions in hazardous or difficult-to-reach environments and scientific exploration in remote deserts or extraterrestrial planets [1]. One particular thrust in this area of research is the development of widely available wheeled robots capable of navigating vertically challenging terrain (e.g., steep slopes, rocky outcroppings, and uneven surfaces, Fig. 1 top) [2]. Achieving reliable and robust mobility in these environments is challenging due to the intricate nature of the terrain, the complex vehicle-terrain interactions, the adverse impact caused by gravity, and the potential deformation of the vehicle chassis.

Despite advancements in classical planning and control for off-road navigation, significant challenges remain. One major issue is the difficulty in precisely modeling vehicle-terrain interactions, which are highly variable and unpredictable in off-road, especially vertically challenge, environments. Implementing a high-precision kinodynamics or vehicle-terrain interaction model within a sampling-based motion planner can consume excessive computational resources onboard a mobile robot. Additionally, errors in these models can cascade into subsequent planning and control processes, leading to suboptimal performance. Furthermore, integrating multiple sensors and control algorithms increases system complexity and makes it challenging to generalize and scale across different terrain and applications.

All authors are with the Department of Computer Science, George Mason University {txu25, cpan7, xiao}@gmu.edu

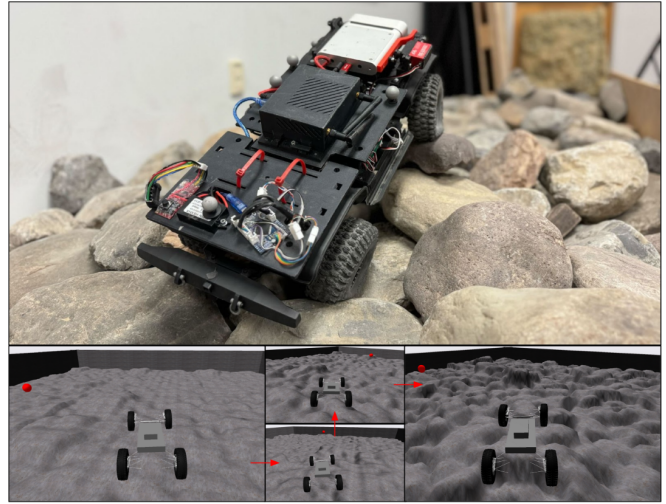


Fig. 1: VW-Chrono: Simulator for Wheeled Mobility on Vertically Challenging Terrain with Increasing Difficulty (Lower Left to Lower Right).

To address these challenges, an increasing number of research efforts have introduced RL methods into off-road navigation. RL algorithms, such as Proximal Policy Optimization (PPO) [3], enables autonomous vehicles to learn and adapt to complex terrain through trial and error in simulation, without the need for costly real-world or expert demonstration data. Learning from a high-precision physics model in a simulator with RL in advance can also alleviate onboard computation during deployment.

To advance off-road navigation solutions for wheeled robots on vertically challenging terrain using RL, we first develop a novel simulation environment developed within the Chrono multi-physics simulation engine [4]. This simulator allows RL for wheeled robots to navigate vertically challenging terrain, with subsequent deployment onto a physical Verti-4-Wheeler (V4W) [2]. We compare our navigation policy learned through PPO against an optimistic planner baseline and a classical planner with elevation approach, which shows the advantage of the RL-learned mobility. In summary, our contributions are outlined as follows:

- We create a simulator for wheeled mobility on vertically challenging terrain, VW-Chrono (Fig. 1 bottom), that procedurally generates four levels of increasing mobility difficulty to incorporate the principle of curriculum learning [5].
- We utilize PPO [3] combined with the Sliced-Wasserstein Autoencoder (SWAE) structure [6] to ef-

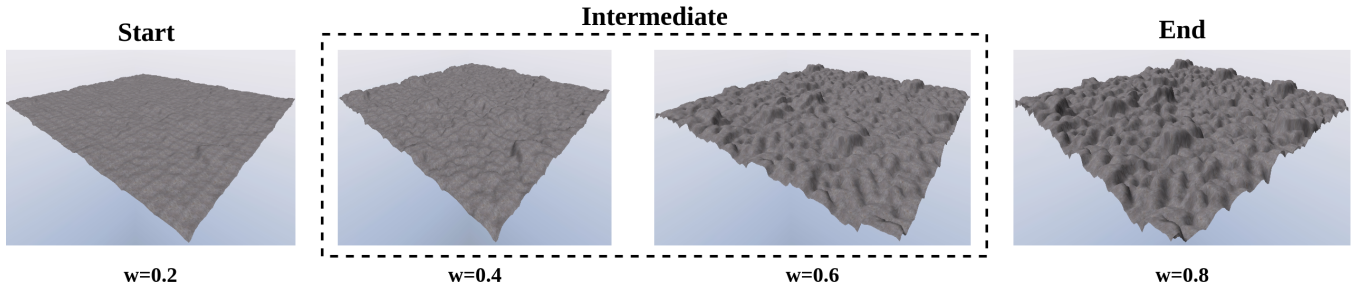


Fig. 2: Increasing Mobility Difficulty of Vertically Challenging Terrain by Interpolating Start and End with Weight w .

ficiently learn wheeled mobility in *VW-Chrono*.

- We present a comparative study between our RL-learned mobility and two baselines for autonomously driving wheeled robots over vertically challenging terrain.

II. RELATED WORK

In this section, we provide a comprehensive review of related work on off-road mobility, focusing on both classical approaches and recent advances in data-driven methods.

Off-road mobility presents significant challenges for autonomous robots due to the complexity and variability of unstructured terrains. Classical approaches have traditionally addressed these challenges by employing hand-crafted methodologies for perception [7], planning [8], modeling [9], and control [10]. These techniques often rely on heuristics and extensive domain expertise to handle environmental variations. While effective in controlled scenarios, these classical methods suffer from several notable limitations: they require significant engineering effort, are susceptible to cascading errors from upstream perception and planning modules, and struggle to adapt effectively to novel or unforeseen environments [11].

To overcome the shortcomings of traditional methods, data-driven approaches for off-road mobility have emerged as promising alternatives [12]. These methods leverage advances in machine learning to directly learn complex behaviors from data, offering adaptability in environments that are too intricate for manual engineering [12]. Among these methods, end-to-end learning of control policies has been explored extensively, where imitation learning [13] and reinforcement learning (RL) are used to learn robust navigation strategies from either expert demonstrations or trial-and-error interactions. Moreover, learning-based semantic perception methods have been employed to provide high-level scene understanding and terrain classification for improved mobility [14]–[19].

In addition to perception, recent efforts have also focused on learning kinodynamic models [20]–[27] that better capture the physical interactions between the robot and varying terrain types. Parameter adaptation approaches [28]–[32] have also been proposed to adjust system parameters on-the-fly based on perception feedback, providing greater robustness to environmental changes. Furthermore, learning-based cost function optimization [33]–[42] has contributed to improved

decision-making by enabling more nuanced and context-aware trajectory planning.

Despite their promise, data-driven approaches face notable challenges. Specifically, RL [43]–[45] and imitation learning [46]–[48] methods tend to be data-intensive, often requiring either millions of trial-and-error iterations or substantial expert-provided labeled datasets [49]–[52] for effective policy learning. Furthermore, ensuring generalization of learned models to diverse, unseen environments remains a critical open question. One potential solution lies in curriculum learning, where a sequence of progressively challenging tasks is presented to the agent [5], [53]. This strategy has shown potential for improving both sample efficiency and robustness of learned policies, thereby facilitating better generalization across different deployment settings.

III. METHOD

In this section, we present the design of *VW-Chrono* and its OpenAI Gym environment. We introduce our RL problem and training for wheeled mobility on vertically challenging terrain, as well as our SWAE-based elevation map encoder.

A. *VW-Chrono*

To ensure the simulated vertically challenging terrain resemble the real world, we first utilize our physical V4W to collect elevation map data on a custom-built indoor testbed designed for vertically challenging terrain. This testbed includes hundreds of rocks and boulders, averaging 30cm in size (matching the scale of the V4W), which are randomly laid out and stacked on a 3.1×1.3 m test course. The highest elevation of the test course can reach up to 0.5m, more than twice the height of the vehicle (Fig. 1 top). We create a grayscale Bitmap image (BMP) with the collected data to represent terrain elevation [54]. In the *Chrono* multi-physics simulation engine, a triangular mesh is generated by assigning a vertex to each pixel of the BMP image. The mesh is then horizontally to match the given extents and expanded vertically to align with the specified range. This ensures that the darkest pixel aligns with the minimum height and the lightest pixel corresponds to the maximum height (Fig. 1 bottom).

To create vertically challenging environments with different difficulty levels as shown in Fig. 2, we create a sequence of elevation maps by linearly interpolating between a starting map I_0 (flat terrain) and an ending map I_N (rugged terrain)

using a weighted average. The intermediate image I_k at stage k out of N stages can be calculated using the following equation:

$$I_k = (1 - \frac{k}{N})I_0 + \frac{k}{N}I_N, \quad \forall k \in \{0, 1, \dots, N\}. \quad (1)$$

In Eqn. (1), the term $\frac{k}{N}$ is used to define the interpolation weight w in Fig. 2. This approach is based on the principle of curriculum learning, which posits that models can learn more effectively and efficiently when tasks are introduced in a structured, incremental manner, starting with simpler tasks and gradually moving to more complex ones.

B. RL Problem Formulation

We employ RL to train a policy that receives environmental inputs and generates actions to drive the robot through vertically challenging terrain, avoiding getting stuck and rolling over while moving toward a designated goal.

A regular Markov Decision Process (MDP) can be defined by a tuple $(\mathcal{S}, \mathcal{A}, \mathcal{T}, \gamma, \mathcal{R})$, including state, action, state transition, discount factor and reward. The goal is to learn a policy $\pi : \mathcal{S} \rightarrow \mathcal{A}$ to maximize the expected cumulative reward over the task horizon T , i.e.,

$$\max_{\pi} \mathbb{E}_{a_t \sim \pi(\cdot | s_t)} \left[\sum_{t=0}^T \gamma^t R_t \right], \quad (2)$$

where $a_t \in \mathcal{A}$ and $s_t \in \mathcal{S}$ are the action and state of the system at each step. To learn wheeled mobility for vertically challenging terrain, the following design choices are made:

1) *State Space*: The inputs to our RL policy include angular difference between the vehicle and goal heading (in radian), current vehicle velocity (in m/s), and cropped elevation map centered at and aligned with the vehicle. We use a Sliced-Wasserstein Autoencoder (SWAE) to reduce elevation map dimensionality and utilize the latent vector to preserve original elevation information. After SWAE pretraining, we freeze the parameters of the encoder during RL training.

2) *Action Space*: The RL policy's outputs include desired linear speed and steering angle, instead of raw throttle and steering commands, in order to improve learning efficiency. A PID controller controls the throttle and steering commands to achieve the desired linear speed and steering angle.

3) *Policy Model Architecture*: We choose PPO [3] as the RL algorithm considering our continuous action space. PPO iteratively collects data through interactions with the environment and updates the policy to maximize the expected cumulative reward (Eqn. (2)). Unlike traditional methods, PPO employs a clipped surrogate objective to constrain policy updates, preventing significant deviations that could lead to instability. By balancing the exploration-exploitation trade-off with a proximal threshold, PPO continually improves the policy while ensuring stability.

C. Sliced-Wasserstein Autoencoder (SWAE)

We use SWAE as a feature extractor to reduce the dimension of the elevation map around the robot while preserving the original elevation information. SWAE is a

scalable generative model that captures the rich and often nonlinear distribution of high-dimensional data (e.g., images, videos, and audio). Learning such generative models involves minimizing a dissimilarity measure between the data distribution and the output distribution of the generative model, which essentially constitutes an optimal transport problem.

D. Reward Design

Our RL agent is trained using a reward function composed of three key terms. These terms are designed to incentivize the agent's movement toward the goal and prevent immobilization. The components of the reward function are:

$$R_t := R_{\text{progress}} + R_{\text{rollover}} + R_{\text{timeout}} \quad (3)$$

1) *Progress Reward*: This term promotes the agent's advancement toward the goal by providing positive rewards for progress made. Additionally, if the agent has not moved at least 1cm within 0.1 seconds, a penalty is applied:

$$R_{\text{progress}} = w_1 \cdot \Delta d - w_2 \cdot \mathbb{I}(\Delta d < 0.01),$$

where Δd is the distance moved towards the goal between the previous timestamp and the current timestamp, $\mathbb{I}()$ is an indicator function, and all different w_i are weight terms.

2) *Rollover Penalty*: To prevent the agent from rolling over, we penalize excessive roll and pitch angles:

$$R_{\text{rollover}} = -w_3 \cdot \sum_{i \in \{\text{roll}, \text{pitch}\}} \max(0, |\theta_i| - \alpha), \quad (4)$$

where θ_{roll} and θ_{pitch} are the roll and pitch angles, respectively, w_3 is a weight term and α is a constant threshold angle.

3) *Timeout Penalty*: For each episode, if a time limit T is reached before the robot reaching the goal, a fixed penalty c is applied for the timeout, along with an additional penalty based on the remaining distance to the goal:

$$R_{\text{timeout}} = -(w_4 \cdot d_{\text{remaining}} + c) \cdot \mathbb{I}(t \geq T),$$

where $d_{\text{remaining}}$ is the remaining distance to the goal.

Table II shows all hyper-parameters of our reward function.

IV. RESULTS

In this section, we present the experimental results of our RL system. compared against two baselines designed for vertically challenging terrain.

A. Baselines

We design two baselines for our VW-Chrono simulation environment.

1) *Optimistic Planner with flat-terrain assumption*: The primary input for this controller is the angular difference between the vehicle's current heading and the desired heading towards the goal. By minimizing this angle difference, the planner guides the vehicle towards its target.

TABLE I: Experiment Results of RL and Baselines

Approach	Stage 1	Stage 2	Stage 3	Stage 4
RL	25/25 , 5.75s, 1.19°/1.26°	25/25 , 5.23s, 2.59°/2.30°	20/25 , 5.35s , 3.55°/2.23°	15/25 , 6.21s, 5.58°/3.82°
Optimistic Planner	25/25 , 4.65s , 1.36°/1.50°	25/25 , 4.82s , 2.63°/2.00°	17/25, 5.46s, 4.32°/2.97°	10/25, 5.68s , 7.11°/4.11°
Naive Planner	25/25 , 5.39s, 1.31°/1.37°	25/25 , 5.18s, 2.80°/ 1.99°	20/25 , 6.20s, 4.90°/2.76°	12/25, 6.73s, 5.67°/4.07°
Best Reward Mean (RL)	2860.9	2415.4	1393.3	739.5

TABLE II: Reward Weights

w_1	w_2	w_3	w_4	α	c	T
50	10	20	10	30	100	15

2) *Naive Planner with elevation heuristic*: The Optimistic Planner with flat-terrain assumption often struggles with steep slopes and rugged boulders, leading to the vehicle getting stuck. To enhance the planner’s performance on challenging rock terrain, we employ a 64×64 cropped elevation map centered on the vehicle. From the front part of the vehicle, we evenly split the map into five regions and choose the most traversable direction: At each time step, we calculate the mean and variance of the elevation values of these five regions and select the region with the most similar mean and lowest variance as the driving direction, compared to the region of the same size centered at the vehicle.

We utilize three metrics to compare results in Table I:

- 1) Number of successful trials (out of 25).
- 2) Mean traversal time (of successful trails in seconds).
- 3) Average roll/pitch angles (in degrees).

B. Simulation Results

In VW-Chrono, we randomly set vehicle start and goal position on the testbed every time and test baselines against our RL system. We present our experiment results in Table I, where best results are shown in bold. The four stages correspond to four increasing difficulty levels, 25 trials each. The RL method consistently achieves a high number of successful trials, particularly excelling in the earlier stages with a perfect success rate in Stages 1 and 2 and maintaining reasonable success rates in Stages 3, 4. However, in Stage 4, while the RL method achieves a success rate of 15 out of 25, it maintains the best roll/pitch stability compared to the Optimistic Planner and Naive Planner, indicating its effectiveness in handling complex terrain with slower and more cautious navigation.

The Optimistic Planner, while achieving the fastest traversal times, shows a decline in performance as the terrain difficulty increases, with a significant drop in the number of successful trials and increasing roll/pitch angles in Stages 3 and 4. This indicates that the Optimistic Planner, although

efficient on less challenging terrain, struggles with stability and success in more complex environments.

The Naive Planner strikes a balance between speed and stability, with a high success rate and relatively low roll/pitch angles across all stages. It demonstrates superior performance over the Optimistic Planner in maintaining lower roll/pitch angles, particularly in the most difficult Stages 4. However, it still does not surpass the RL approach in terms of overall stability in those complex stages.

C. Physical Demonstration

We also deploy the RL policy learned in simulation on a physical V4W platform on a real-world rock testbed (Fig. 1 top). The robot is a four-wheeled platform based on an off-the-self, two-axle, four-wheel-drive, off-road vehicle from Traxxas. The onboard computation platform is a NVIDIA Jetson Xavier NX module. First, we place the V4W on flat terrain and specify a direction for it to follow. The RL policy successfully guides the V4W in the intended direction. Next, we introduce a large obstacle to assess the RL policy’s performance. Finally, we test the V4W on the rock testbed and observe that the RL policy effectively enables the V4W to move toward its goal across the rocky terrain as shown in Fig. 3.

V. CONCLUSION

This paper presents a comprehensive RL system to unlock the previously unrealized potential of wheeled mobility on vertically challenging terrain. The VW-Chrono simulator can generate challenging terrain for future off-road navigation research with adjustable mobility difficulty levels. We utilize PPO as our RL algorithm based on a carefully designed reward structure. The experimental results confirm our hypothesis that conventional wheeled robots possess the mechanical capability to navigate vertically challenging terrain, which are normally considered as non-traversable obstacles, especially with the help of data-driven approaches. Furthermore, we demonstrate the feasibility of transferring RL-learned mobility from simulation to a physical robot, enabling it to navigate real-world vertically challenging terrain.

This paper opens up a new research direction aimed at achieving extreme off-road robot mobility using RL methods. One promising future research direction is to employ a teacher-student structure to automatically create different

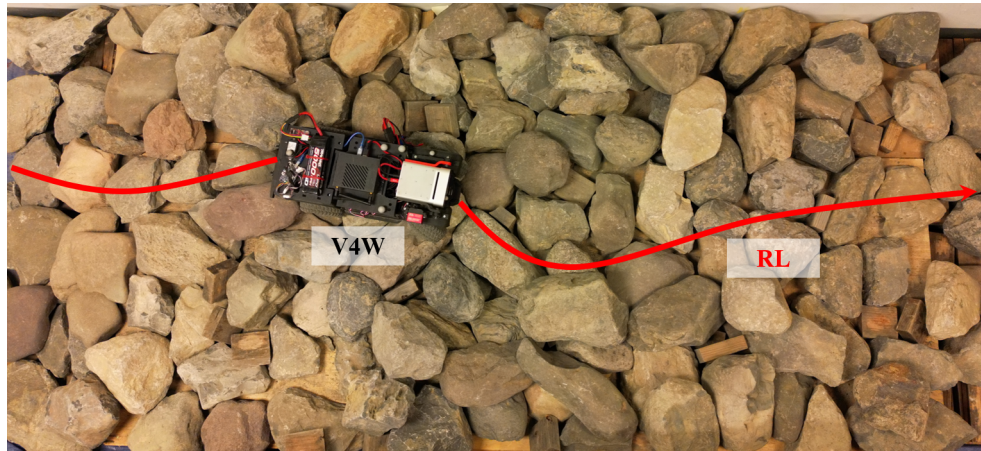


Fig. 3: Custom-Built Testbed with V4W and an Example Trajectory by the RL Algorithm.

levels of terrain in an automatic curriculum learning setting to improve learning efficiency.

REFERENCES

- [1] M. D. Teji, T. Zou, and D. S. Zeleke, "A survey of off-road mobile robots: Slippage estimation, robot control, and sensing technology," *Journal of Intelligent & Robotic Systems*, vol. 109, no. 2, p. 38, 2023.
- [2] A. Datar, C. Pan, M. Nazeri, and X. Xiao, "Toward wheeled mobility on vertically challenging terrain: Platforms, datasets, and algorithms," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2024.
- [3] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," *arXiv preprint arXiv:1707.06347*, 2017.
- [4] A. Tasora, R. Serban, H. Mazhar, A. Pazouki, D. Melanz, J. Fleischmann, M. Taylor, H. Sugiyama, and D. Negrut, "Chrono: An open source multi-physics dynamics engine," in *High Performance Computing in Science and Engineering: Second International Conference, HPCSE 2015, Solán, Czech Republic, May 25-28, 2015, Revised Selected Papers 2*. Springer, 2016, pp. 19–49.
- [5] S. Narvekar, B. Peng, M. Leonetti, J. Sinapov, M. E. Taylor, and P. Stone, "Curriculum learning for reinforcement learning domains: A framework and survey," *Journal of Machine Learning Research*, vol. 21, no. 181, pp. 1–50, 2020.
- [6] S. Kolouri, P. E. Pope, C. E. Martin, and G. K. Rohde, "Sliced-wasserstein autoencoder: An embarrassingly simple generative model," *arXiv preprint arXiv:1804.01947*, 2018.
- [7] D. V. Lu, D. Hershberger, and W. D. Smart, "Layered costmaps for context-sensitive navigation," in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2014, pp. 709–715.
- [8] H. Rastgoftar, B. Zhang, and E. M. Atkins, "A data-driven approach for autonomous motion planning and control in off-road driving scenarios," in *2018 Annual American control conference (ACC)*. IEEE, 2018, pp. 5876–5883.
- [9] R. He, C. Sandu, A. K. Khan, A. G. Guthrie, P. S. Els, and H. A. Hamersma, "Review of terramechanics models and their applicability to real-time applications," *Journal of Terramechanics*, vol. 81, pp. 3–22, 2019.
- [10] G. Williams, P. Drews, B. Goldfain, J. M. Rehg, and E. A. Theodorou, "Aggressive driving with model predictive path integral control," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2016.
- [11] S. Thrun, M. Montemerlo, H. Dahlkamp, D. Stavens, A. Aron, J. Diebel, P. Fong, J. Gale, M. Halpenny, G. Hoffmann *et al.*, "Stanley: The robot that won the darpa grand challenge," *Journal of field Robotics*, vol. 23, no. 9, pp. 661–692, 2006.
- [12] X. Xiao, B. Liu, G. Warnell, and P. Stone, "Motion planning and control for mobile robot navigation using machine learning: a survey," *Autonomous Robots*, vol. 46, no. 5, pp. 569–597, 2022.
- [13] Y. Pan, C.-A. Cheng, K. Saigol, K. Lee, X. Yan, E. A. Theodorou, and B. Boots, "Imitation learning for agile autonomous driving," *The International Journal of Robotics Research*, vol. 39, no. 2-3, pp. 286–302, 2020.
- [14] R. Manduchi, A. Castano, A. Talukder, and L. Matthies, "Obstacle detection and terrain classification for autonomous off-road navigation," *Autonomous robots*, vol. 18, pp. 81–102, 2005.
- [15] D. Maturana, P.-W. Chou, M. Uenoyama, and S. Scherer, "Real-time semantic mapping for autonomous off-road navigation," in *Field and Service Robotics*. Springer, 2018, pp. 335–350.
- [16] A. Shaban, X. Meng, J. Lee, B. Boots, and D. Fox, "Semantic terrain classification for off-road autonomous driving," in *Conference on Robot Learning*. PMLR, 2022, pp. 619–629.
- [17] X. Meng, N. Hatch, A. Lambert, A. Li, N. Wagener, M. Schmittle, J. Lee, W. Yuan, Z. Chen, S. Deng *et al.*, "Terrainnet: Visual modeling of complex terrain for high-speed, off-road navigation," *arXiv preprint arXiv:2303.15771*, 2023.
- [18] K. Viswanath, K. Singh, P. Jiang, P. Sujit, and S. Saripalli, "Offseg: A semantic segmentation framework for off-road driving," in *2021 IEEE 17th International Conference on Automation Science and Engineering (CASE)*. IEEE, 2021, pp. 354–359.
- [19] K. S. Sikand, S. Rabiee, A. Uccello, X. Xiao, G. Warnell, and J. Biswas, "Visual representation learning for preference-aware path planning," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 11 303–11 309.
- [20] X. Xiao, J. Biswas, and P. Stone, "Learning inverse kinodynamics for accurate high-speed off-road navigation on unstructured terrain," *IEEE Robotics and Automation Letters*, vol. 6, no. 3, pp. 6054–6060, 2021.
- [21] H. Karnan, K. S. Sikand, P. Atreya, S. Rabiee, X. Xiao, G. Warnell, P. Stone, and J. Biswas, "Vi-ikd: High-speed accurate off-road navigation using learned visual-inertial inverse kinodynamics," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022, pp. 3294–3301.
- [22] P. Atreya, H. Karnan, K. S. Sikand, X. Xiao, S. Rabiee, and J. Biswas, "High-speed accurate robot control using learned forward kinodynamics and non-linear least squares optimization," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022, pp. 11 789–11 795.
- [23] A. Datar, C. Pan, M. Nazeri, A. Pokhrel, and X. Xiao, "Terrain-attentive learning for efficient 6-dof kinodynamic modeling on vertically challenging terrain," in *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2024.
- [24] A. Datar, C. Pan, and X. Xiao, "Learning to model and plan for wheeled mobility on vertically challenging terrain," *arXiv preprint arXiv:2306.11611*, 2023.
- [25] A. Pokhrel, A. Datar, M. Nazeri, and X. Xiao, "CAHSOR: Competence-aware high-speed off-road ground navigation in SE (3)," *IEEE Robotics and Automation Letters*, 2024.
- [26] P. Maheshwari, W. Wang, S. Triest, M. Sivaprakasam, S. Aich, J. G. Rogers III, J. M. Gregory, and S. Scherer, "Piaug—physics informed augmentation for learning vehicle dynamics for off-road navigation," *arXiv preprint arXiv:2311.00815*, 2023.

- [27] M. Nazeri, A. Datar, A. Pokhrel, C. Pan, G. Warnell, and X. Xiao, "Verticencoder: Self-supervised kinodynamic representation learning on vertically challenging terrain," *arXiv preprint arXiv:2409.11570*, 2024.
- [28] X. Xiao, B. Liu, G. Warnell, J. Fink, and P. Stone, "Appld: Adaptive planner parameter learning from demonstration," *IEEE Robotics and Automation Letters*, vol. 5, no. 3, pp. 4541–4547, 2020.
- [29] Z. Wang, X. Xiao, B. Liu, G. Warnell, and P. Stone, "Appli: Adaptive planner parameter learning from interventions," in *2021 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2021, pp. 6079–6085.
- [30] Z. Wang, X. Xiao, G. Warnell, and P. Stone, "Apple: Adaptive planner parameter learning from evaluative feedback," *IEEE Robotics and Automation Letters*, vol. 6, no. 4, pp. 7744–7749, 2021.
- [31] Z. Xu, G. Dhamankar, A. Nair, X. Xiao, G. Warnell, B. Liu, Z. Wang, and P. Stone, "Applr: Adaptive planner parameter learning from reinforcement," in *2021 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2021, pp. 6086–6092.
- [32] X. Xiao, Z. Wang, Z. Xu, B. Liu, G. Warnell, G. Dhamankar, A. Nair, and P. Stone, "Appl: Adaptive planner parameter learning," *Robotics and Autonomous Systems*, vol. 154, p. 104132, 2022.
- [33] M. Sivaprakasam, S. Triest, W. Wang, P. Yin, and S. Scherer, "Improving off-road planning techniques with learned costs from physical interactions," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 4844–4850.
- [34] N. Dashora, D. Shin, D. Shah, H. Leopold, D. Fan, A. Agha-Mohammadi, N. Rhinehart, and S. Levine, "Hybrid imitative planning with geometric and predictive costs in off-road environments," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 4452–4458.
- [35] X. Cai, M. Everett, J. Fink, and J. P. How, "Risk-aware off-road navigation via a learned speed distribution map," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022, pp. 2931–2937.
- [36] M. G. Castro, S. Triest, W. Wang, J. M. Gregory, F. Sanchez, J. G. Rogers, and S. Scherer, "How does it feel? self-supervised costmap learning for off-road vehicle traversability," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, 2023, pp. 931–938.
- [37] X. Cai, S. Ancha, L. Sharma, P. R. Osteen, B. Bucher, S. Phillips, J. Wang, M. Everett, N. Roy, and J. P. How, "EVORA: Deep evidential traversability learning for risk-aware off-road autonomy," *IEEE Transactions on Robotics*, 2024.
- [38] X. Cai, J. Queeney, T. Xu, A. Datar, C. Pan, M. Miller, A. Flather, P. R. Osteen, N. Roy, X. Xiao *et al.*, "Pietra: Physics-informed evidential learning for traversing out-of-distribution terrain," *arXiv preprint arXiv:2409.03005*, 2024.
- [39] J. Seo, S. Sim, and I. Shim, "Learning off-road terrain traversability with self-supervisions only," *IEEE Robotics and Automation Letters*, vol. 8, no. 8, pp. 4617–4624, 2023.
- [40] S. Jung, J. Lee, X. Meng, B. Boots, and A. Lambert, "V-STRONG: Visual self-supervised traversability learning for off-road navigation," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2024, pp. 1766–1773.
- [41] X. Xiao, T. Zhang, K. M. Choromanski, T.-W. E. Lee, A. Francis, J. Varley, S. Tu, S. Singh, P. Xu, F. Xia, S. M. Persson, L. Takayama, R. Frostig, J. Tan, C. Parada, and V. Sindhwani, "Learning model predictive controllers with real-time attention for real-world navigation," in *Conference on robot learning*. PMLR, 2022.
- [42] C. Pan, A. Datar, A. Pokhrel, M. Choulas, M. Nazeri, and X. Xiao, "Traverse the non-traversable: Estimating traversability for wheeled mobility on vertically challenging terrain," *arXiv preprint arXiv:2409.17479*, 2024.
- [43] Z. Xu, B. Liu, X. Xiao, A. Nair, and P. Stone, "Benchmarking reinforcement learning techniques for autonomous navigation," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2023, pp. 9224–9230.
- [44] Z. Xu, X. Xiao, G. Warnell, A. Nair, and P. Stone, "Machine learning methods for local motion planning: A study of end-to-end vs. parameter learning," in *2021 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*. IEEE, 2021, pp. 217–222.
- [45] Z. Xu, A. H. Raj, X. Xiao, and P. Stone, "Dexterous legged locomotion in confined 3d spaces with reinforcement learning," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2024.
- [46] H. Karnan, A. Nair, X. Xiao, G. Warnell, S. Pirk, A. Toshev, J. Hart, J. Biswas, and P. Stone, "Socially compliant navigation dataset (scand): A large-scale dataset of demonstrations for social navigation," *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 11 807–11 814, 2022.
- [47] D. M. Nguyen, M. Nazeri, A. Payandeh, A. Datar, and X. Xiao, "Toward human-like social robot navigation: A large-scale, multi-modal, social human navigation dataset," in *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2023, pp. 7442–7447.
- [48] H. Karnan, G. Warnell, X. Xiao, and P. Stone, "Voila: Visual-observation-only imitation learning for autonomous navigation," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 2497–2503.
- [49] X. Xiao, B. Liu, G. Warnell, and P. Stone, "Toward agile maneuvers in highly constrained spaces: Learning from hallucination," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 1503–1510, 2021.
- [50] X. Xiao, B. Liu, and P. Stone, "Agile robot navigation through hallucinated learning and sober deployment," in *2021 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2021, pp. 7316–7322.
- [51] Z. Wang, X. Xiao, A. J. Nettekoven, K. Umasankar, A. Singh, S. Bommakanti, U. Topcu, and P. Stone, "From agile ground to aerial navigation: Learning from learned hallucination," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2021, pp. 148–153.
- [52] S. A. Ghani, Z. Wang, P. Stone, and X. Xiao, "Dyna-lflh: Learning agile navigation in dynamic environments from learned hallucination," *arXiv preprint arXiv:2403.17231*, 2024.
- [53] L. Wang, Z. Xu, P. Stone, and X. Xiao, "Grounded curriculum learning," *arXiv preprint arXiv:2409.19816*, 2024.
- [54] T. Miki, L. Wellhausen, R. Grandia, F. Jenelten, T. Homberger, and M. Hutter, "Elevation mapping for locomotion and navigation using gpu," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022, pp. 2273–2280.