

# AdaptiveON: Adaptive Outdoor Local Navigation Method for Stable and Reliable Actions

Jing Liang<sup>✉</sup>, Kasun Weerakoon, Tianrui Guan<sup>✉</sup>, *Graduate Student Member, IEEE*,  
Nare Karapetyan<sup>✉</sup>, *Member, IEEE*, and Dinesh Manocha<sup>✉</sup>, *Fellow, IEEE*

**Abstract**—We present a novel outdoor navigation algorithm to generate stable and efficient actions to navigate a robot to reach a goal. We use a multi-stage training pipeline and show that our approach produces policies that result in stable and reliable robot navigation on complex terrains. Based on the Proximal Policy Optimization (PPO) algorithm, we developed a novel method to achieve multiple capabilities for outdoor local navigation tasks, namely alleviating the robot's drifting, keeping the robot stable on bumpy terrains, avoiding climbing on hills with steep elevation changes, and avoiding collisions. Our training process mitigates the reality (sim-to-real) gap by introducing generalized environmental and robotic parameters and training with rich features captured from light detection and ranging (Lidar) sensor in a high-fidelity Unity simulator. We evaluate our method in both simulation and real-world environments using Clearpath Husky and Jackal robots. Further, we compare our method against the state-of-the-art approaches and observe that, in the real world, our method improves stability by at least 30.7% on uneven terrains, reduces drifting by 8.08%, and decreases the elevation changes by 14.75%.

**Index Terms**—Autonomous vehicle navigation, motion control, reinforcement learning.

## I. INTRODUCTION

THE mobile robot navigation problem in indoor and outdoor scenes has been studied extensively. Nevertheless, outdoor navigation remains an unresolved challenge when dealing with unstructured terrains and environmental uncertainties [1], [2].

Robot navigation in complex terrains is severely affected by different characteristics of the environment [3], [4]. The texture of the ground surface may affect the friction between the robot and the terrain. For example, mud, leaves, or water may cause robots to drift [5], [6]. For stable navigation, along with collision avoidance, reducing vibration and preventing high elevation changes in motion are important objectives that will help prevent motor damage and ensure accurate motions [3], [7]. Several



Fig. 1. Trajectories generated by AdaptiveON (ours) when navigating Clearpath Jackal and Husky robots on complex outdoor terrains: Our method ensures stable navigation by reducing vibrations of the Jackal robot on a slope with uneven texture (left); AdaptiveON provides reliable motion control along the trajectories to prevent the Husky from drifting on slippery terrains, including weathered leaves, small rocks, and grass (right).

works use segmentation methods to classify the uneven terrains based on traversability [4], [8], [9], [10]. However, segmentation suffers from the noise and complexity of real-world features; for different robots, the traversability of the same area could be different based on their dynamics.

Simulations provide the means to perform extensive and safe training for navigation tasks. Nevertheless, the policies learned in these environments are always bound by the characteristics of the simulations and hence suffer when generalized to the real world. The sim-to-real problem or so-called “reality gap” arises when many uncertainties are included: the physical properties of the robot, the world model differences, and the dynamics of the simulated world. This means that better navigation requires real-world training. However, training in the real world is not always feasible, especially when dealing with expensive robots with stability constraints [11], [12]. Some of the approaches for closing the reality gap train the RL system in simulation and then use sim-to-real methods [13]. Alternative strategies aim to develop adaptive systems that do not require real-world training [14], [15], [16].

**Main Contributions:** We present AdaptiveON, a novel approach with a multi-stage pipeline based on the PPO algorithm for training robots that avoid obstacles and perform stable and reliable navigation on hills, uneven and slippery areas in complex outdoor scenarios. Detailed contributions are as follows:

- 1) We introduce the “Adaptive Model” to embed the recent history of the robot’s status, which helps navigation on terrains with different properties. Incorporating it with perception models, we adapt the robot to different terrains and train the robot with reliable navigation. For instance, moving on slippery terrains with less drift, moving on hills

Manuscript received 7 August 2022; accepted 1 December 2022. Date of publication 16 December 2022; date of current version 23 December 2022. This letter was recommended for publication by Associate Editor J. D. Hernandez and Editor A. Bera upon evaluation of the reviewers’ comments. This work was supported in part by ARO under Grant, W911NF2110026, and in part by the U.S. Army Cooperative Agreement under Grant W911NF2120076. (Corresponding author: Nare Karapetyan.)

The authors are with the University of Maryland, College Park, MD 20742 USA (e-mail: jingl@umd.edu; kasunw@umd.edu; rayguan@umd.edu; knare@umd.edu; dmanocha@umd.edu).

Code: <https://github.com/jingGM/adaptiveON.git> Project Page: <https://gamma.umd.edu/adaptiveon>.

This letter has supplementary downloadable material available at <https://doi.org/10.1109/LRA.2022.3229907>, provided by the authors.

Digital Object Identifier 10.1109/LRA.2022.3229907

with low elevation changes on each action, and avoiding collisions with static obstacles. We show that our method has less drifting than other state-of-the-art approaches.

- 2) Along with the encoded information of the history of the robot's states and actions, we also introduce a module for stability, "Stable Model" to adaptively control the thresholds of the robot's velocity on different types of terrains. The model can keep the robot at a relatively low speed on bumpy surfaces and make the robot stable.
- 3) We alleviate the reality gap by designing training scenarios in the RaiSim simulator [17] with different environmental, mechanical, and control parameters that could affect the robots' motions, e.g., frictions, unevenness, PID parameters, and weights. Additionally, we fine-tune the robot's perception in the Unity simulator with rich features.

The rest of this paper is organized as follows. First, an overview of related work is presented in Section II. In Section III, we give a formal definition of the problem and present details of the method. We report the experimental results of our method compared to state-of-the-art approaches in Section IV. Finally, conclusions are drawn in Section V.

## II. RELATED WORK

Approaches for addressing the robot navigation problem in outdoor environments vary from those specific to indoor environments primarily due to the complexity of the outdoor terrains. When the state of the environment and the kinematics of the system are known, some classical methods [18], [19], [20] have proven to be effective in tackling navigation; nevertheless, when the environment is challenging and changes are stochastic, learning-based approaches are more practical [3], [21], [22], [23].

### A. Learning-Based Navigation

Deep Reinforcement Learning (DRL) has been widely used in complex control problems under uncertainties. Brunner et al. [24] used DRL for indoor navigation to drive out of a random maze by learning to read a global map and finding the shortest path. However, retraining is required after changing targets when the navigation has been designed for specific tasks. Staroverov et al. [25] proposed hierarchical DRL with classical methods integrated into the model. The navigation model presented by [26] is an end-to-end neural network trained with a combination of expert demonstrations, imitation learning, and reinforcement learning.

DRL-based navigation methods show promise, but mostly in simulated environments. These methods are often difficult to execute in the real world due to underlying challenges such as sample efficiency or data synchronization [22], [27]. On the other hand, some of the approaches require real-world training or data collection, which might be non-trivial. BADGR [1] uses a self-supervised learning approach, which needs to experience real-world events such as colliding with obstacles to learn its policy.

In our prior work, we used learning-based approaches to navigate in dense crowds [28], [29], [30], in scenarios with

fast moving obstacles [31], or in outdoor terrains by selecting locations with low elevation changes to travel through [3]. We also used segmentation methods for identifying navigable regions [4]. These works are limited by the predefined characteristics of the traversability of a region associated with the specific type of robot. In this work, we introduce an adaptive model that will allow us to alleviate those dependencies.

### B. Sim-to-Real

The problem of transferring from simulation to the real world is called the reality gap. Learning in the real world is challenging because robots can be damaged when interacting with the world. Hwangbo et al. [11] proposed a method based on the Trust Region Optimization Policy (TRPO) that learns gaits of the robot. Their model produced the fastest moving and most agile actions at the time for legged robots. The model has been deployed on the ANYmal robot, which successfully completed a 2.2 km length trail up a mountain and back. CrowdSteer [30] used PPO to train the policy in different scenarios for collision avoidance and achieved relatively good performance.

Outdoor navigation for legged robots based on adaptation from simulation has been proposed by [12]. Their approach, Rapid Motor Adaptation (RMA), consists of two components: a base policy and an adaptation module. It is trained completely in simulation without using any domain knowledge and deployed on the robot without any fine tuning. This ensures that, for complex systems such as legged robots, no additional real-world training is required. However, the objective of this work is to generate only stable motions without considering target-approaching objectives.

End-to-end DRL strategy [32] incorporates an elevation map, depth images, and orientation details to train a DRL policy to navigate in uneven terrains. In [33], a DRL method for local planning in unknown rough terrain with zero-range to local-range sensing is proposed. The method utilizes ground elevation data, self-attention modules, the geometric transformation between two successive timesteps, and the corresponding action to navigate on surfaces with different levels of friction and elevation. However, these methods are trained and tested only in simulated environments.

## III. APPROACH

In this section, we formally define the problem, introduce the architecture of the neural networks, and discuss the training details.

### A. Problem Formulation

In this project we use Lidar, odometer and IMU as primary sensors to perceive the environment and robot's status. We assume that the system states can be estimated from the observations. Thus, we formulate the outdoor navigation problem as Markov Decision Process (MDP) problem:

$$\mathcal{P} = (\mathcal{S}, \mathcal{A}, \mathcal{T}, \mathcal{R}, \gamma),$$

where  $\mathcal{S}$  is the state space, and we have each state  $\mathbf{s} = (\mathbf{o}_l, \mathbf{o}_s) \in \mathcal{S}$ , which contains the lidar observation,  $\mathbf{o}_l$ , and the current status of the robot,  $\mathbf{o}_s$ , perceived by IMU and odometer.  $\mathcal{A} \in \mathbb{R}^2$  is the action space, and each action  $\mathbf{a} = (v, \omega) \in \mathcal{A}$  has linear and angular velocities,  $v \in [0, 1](\text{m/s})$  and  $\omega \in [-1, 1](\text{r/s})$ , respectively. Each action is chosen by the end-to-end network  $\mathbf{a} = \pi_{\theta}(\mathbf{s})$ .  $\mathcal{T}$  is the transitional probability from the current state and action to the next state, and  $P(\mathbf{s}_{t+1}|\mathbf{s}_t, \mathbf{a}_t)$  denotes its probability.  $\mathcal{R} : \mathcal{S} \times \mathcal{A} \rightarrow \mathbb{R}$  represents the mapping from states and actions to rewards. In each time step, we have the reward  $r(\mathbf{s}_t, \mathbf{a}_t) \in \mathcal{R}$ , where  $\mathbf{s}_t$  and  $\mathbf{a}_t$  are the state and action at the time step  $t$ , and the reward is observed from the environment/simulator. The discount factor  $\gamma$  is set to 0.97. Denote  $\pi_{\theta}(\mathbf{s}_t)$  as the policy, which represents the entire network.  $\theta$  represents the parameters of the networks.

We use the PPO reinforcement learning algorithm to solve this MDP problem, and the objective of the algorithm is to find the  $\pi_{\theta^*}(\mathbf{s}_t)$  that achieves the highest rewards. We use different reward component functions for different navigation functionalities, including collision avoidance, stability maintenance, drift prevention, and low elevation changes.

The **stability** objective in this work is to minimize the vibrations while moving. The vibration  $vb$  is defined as the sum of angular velocities of roll and pitch angles  $\omega_r$  and  $\omega_p$ , respectively.

$$vb = |\omega_r| + |\omega_p| \quad (1)$$

We call the navigation **reliable** if the robot can successfully reach the goal while avoiding obstacles, achieves lower elevation changes, and has less drifting. To be more specific, the drifting issue is always caused by low friction, i.e., given a certain velocity the robot may not move as predicted.

## B. Architecture

The architecture of the proposed approach consists of several components, each designed for one capability of the robot. In addition to collision avoidance, our objectives are to prevent drifting, maintain stable navigation on uneven terrain, and keep changes in the elevation of the robot small in each time-step. Accordingly, there are 4 models embedding the information required by the four functional targets and a single Policy Model, which takes this information and generates actions for the robot. The complete architecture of our proposed method consists of perception models (Lidar Model and Elevation-map Model), Adaptive Model, Stable Model and Policy Model (see Fig. 2 for the architecture and details in Supplement of [34]).

Lidar is widely used in navigation tasks [28], [30], [31] due to its accurate distance estimation capabilities. In this work, we use a 16-Channel Velodyne Lidar as the primary perception sensor. Our perception models include the Lidar and the Elevation-map Models. These perception models are used to extract the features of the nearby environment of the robot. We use a pointcloud-based elevation map [35] to detect the terrains in the robot's vicinity. However, the elevation map cannot identify the obstacles accurately. Therefore, we use the middle channel of the raw Lidar data to extract the nearby obstacles, as

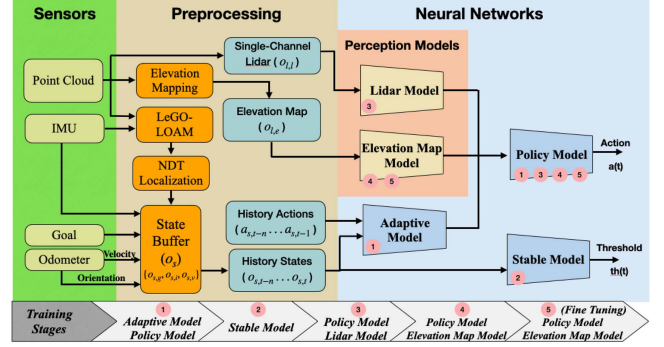


Fig. 2. System Architecture consisting of 5 models: Perception Models (Lidar Model and Elevation Map Model); Adaptive Model; Policy Model that takes the output from all the other models and generates actions resulting in reliable navigation; and Stable Model, which generates an action threshold to ensure stability. The training pipeline, on the bottom of the figure, is explained in Section III-C3.

shown in Fig. 2. The Lidar Model is composed of a sequence of convolutional layers and fully connected layers, and the Elevation-Map Model is composed of 2-D convolutional layers and fully connected layers. The output of the Lidar Model is  $\mathbf{o}_{l,l}$ , and the output of the Elevation-Map Model is  $\mathbf{o}_{l,e}$ ; hence, the Lidar observation is  $\mathbf{o}_l = (\mathbf{o}_{l,l}, \mathbf{o}_{l,e})$ . The input and network details are shown in the supplement [34].

Inspired by RMA [12], we also use the previous consecutive frames of the robot's observed states and actions to estimate the robot's current dynamic status in motion. Those states and actions are states  $\mathbf{o}_{s,t-n} \dots \mathbf{o}_{s,t}$  and actions  $\mathbf{a}_{t-n} \dots \mathbf{a}_{t-1}$ . Each state  $\mathbf{o}_s = \{\mathbf{o}_{s,g}, \mathbf{o}_{s,i}, \mathbf{o}_{s,v}\}$  contains the information of the relative goal position  $\mathbf{o}_{s,g}$  (detected from the positioning model in Section IV), filtered IMU data  $\mathbf{o}_{s,i} \in \mathbb{R}^6$  from the IMU sensor, and the robot's current velocity  $\mathbf{o}_{s,v}$  from the odometry sensor. The goal observation  $\mathbf{o}_{s,g} = (d_g, a_g, h_g)$  comprises the distance, angle, and height from the robot's current position to the goal position.  $\mathbf{o}_{s,i}$  contains the linear accelerations  $(a_x, a_y, a_z)$ , angular velocities  $(\omega_r, \omega_p, \omega_y)$  of roll, pitch, and yaw rotations, and the roll, pitch, and yaw angles  $(r, p, y)$ . Further,  $\mathbf{o}_{s,v}$  includes the linear and angular velocities  $(v, \omega)$  of the robot obtained from the odometer (see Fig. 2). Finally, we concatenate relative goal position, IMU, and Odometer data together as a state. The details of the Adaptive Model are in the Supplement section of [34].

As shown in Fig. 2, there are two output models: the Policy Model and the Stable Model. The Policy Model takes the embedded values from perception models and the Adaptive Model to generate actions,  $\mathbf{a}_p$ , to move the robot, and the Stable Model takes the historic state information,  $\mathbf{o}_{s,t-n} \dots \mathbf{o}_{s,t}$ , to generate the threshold,  $\mathbf{th}$ , for the actions of the robots. These two models are all composed of fully connected layers. We use the final clipped actions,  $\mathbf{a} = \min(\mathbf{th}, \mathbf{a}_p)$ , to drive the robot. The details of the Output Models are in the Supplement section in [34].

## C. Training

1) *Training Objectives*: By training different models in different scenarios, the policy can obtain different functionalities: preventing drifts, running slower on uneven terrain, avoiding



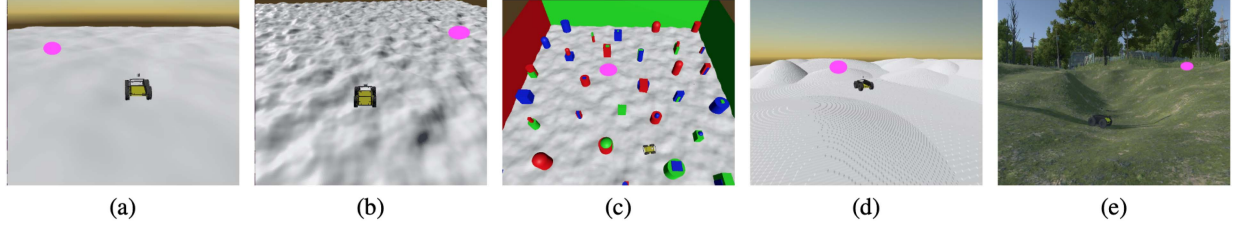


Fig. 3. Different Scenarios in simulation used for training different functionalities of robots. The purple circles denote the given goals. (a) has different environmental, mechanical, and controlling parameters for the robot. This type of scenario trains the policy to adapt to different outdoor conditions and prevent drift. (b) has high unevenness and is used to train the Stable Model to reduce vibration. (c) is used to train the robot to avoid collisions with obstacles. (d) contains hills with different heights and gradients, and it is used to train the robot to move in lower elevation changes in each step. (e) includes most of the properties of the previous scenarios and has rich features to fine-tune the perception networks.

collisions, and having small elevation changes when it runs on hills. Since the Raisim simulator [17] provides better tools to customize terrain features, we train all the basic functionalities in Raisim using the scenarios shown in Fig. 3.

The other objective of the training is to alleviate the differences between the simulation and the real world. To this end, we focus on two aspects: First, in real-world outdoor scenarios, terrains can have different levels of frictions, unevenness and restitution, and the robots can have different weights, PID parameters, and joint numbers. We designed Raisim simulation environments with different values of these parameters (discussed in Section III-C3) and trained the Adaptive Model to take the robot's observed states in the scenarios with these different properties. Second, the real world has richer features than the Raisim simulator w.r.t. Lidar data. Hence, we use a high-fidelity Unity simulator, which contains significantly rich features, to fine-tune the parameters of the perception models.

2) *Reward Functions*: The general reward function is:

$$r = \begin{cases} r_t & \text{if terminated,} \\ \beta_g r_g + \beta_d r_d + \beta_s r_s + \beta_c r_c + \beta_e r_e & \text{otherwise,} \end{cases} \quad (2)$$

where the values  $\beta_g, \beta_d, \beta_s, \beta_c, \beta_e$  are coefficients of the rewards. During different training stages, we use higher weights to the corresponding rewards and lower otherwise (even 0). In this work, we only consider (0,1) for the weights, which are discussed in Training Pipeline Section III-C3. To efficiently train the different models with different functionalities and training stages, the rewards are weighted according to those functionalities.

1) *Distance to Goal  $r_g$* : This reward encourages the robot to move towards the target and penalizes it for deviating.  $d_{g,\text{last}}$  and  $d_{g,\text{now}}$  are the distances between the robot and the goal in the last and the current time step, respectively:

$$r_g = d_{g,\text{last}} - d_{g,\text{now}}. \quad (3)$$

2) *Drift  $r_d$* : This novel reward function penalizes drifting on slippery terrains.  $\mathbf{p}_a$  is the position of the robot in the next time step calculated by the given action, and  $\mathbf{p}_r$  is the real position in the next time step:

$$r_d = -\|\mathbf{p}_a - \mathbf{p}_r\|_2. \quad (4)$$

3) *Stability  $r_s$* : This reward encourages the robot to run with more stability, which is  $vb$  as defined in (1). The  $vb_{th}$  is

the stability threshold, which we consider as 0.5 during the experiments:

$$r_s = \begin{cases} 0, & vb \leq vb_{th} \\ vb_{th} - vb, & vb > vb_{th}, \end{cases} \quad (5)$$

4) *Obstacles  $r_c$* : This reward penalizes the robot when it runs close to obstacles.  $c_{th} = 1.5 \times r_{\text{rob}}$  is the distance threshold to obstacles, where  $r_{\text{rob}}$  is the radius of the robot.  $d_c = \min(\mathbf{d}_o)$  is the nearest distance to all the nearby obstacles. Here,  $\mathbf{d}_o$  is the single-channel Lidar data, which is a vector of sampled distances:

$$r_c = \begin{cases} 0, & d_c > c_{th} \\ \exp(d_c - c_{th}), & d_c \leq c_{th}, \end{cases} \quad (6)$$

5) *Elevation changes  $r_e$* : This reward encourages the robot to move to areas with lower elevation changes. The elevation change is represented by the gradient of the elevation map.  $\mathbf{g}_e$  represents the gradient vector along the heading direction of the robot:

$$r_e = \|\mathbf{g}_e\|_2, \quad (7)$$

6) *Termination rewards  $r_t$* : This reward is triggered during specific events such as: robots reaching the goals  $r_{\text{success}}$ , collisions  $r_{\text{collision}}$ , or robot flip-overs  $r_{\text{flips}}$ :

$$r_t = \begin{cases} r_{\text{success}} = 100 & \text{If reaches the goal,} \\ r_{\text{collision}} = -100 & \text{If in collision,} \\ r_{\text{flips}} = -100 & \text{If flip-over,} \end{cases} \quad (8)$$

3) *Training Pipeline*: We train the functionalities consecutively in the scenarios listed below. While training a specific functionality, we only make the parameters of relative models derivable. Other models are fixed, and the untrained models are initialized with 0 to prevent disturbance to the current training stage. Before the Uneven Terrain, we use the maximum threshold  $th = [1 \text{ m/s}, 1 \text{ r/s}]$  to substitute the output of the Stable Model.

1) *Adaptive Scenarios* (Fig. 3(a)): This scenario has different environment parameters, including different levels of friction (coefficient in  $[0.3-5.0]$ ) and unevenness (different densities of bumps and height changes in the range  $[-1 \text{ m}, 1 \text{ m}]$  of the terrain). We also employ robots with different joints (2 and 4 joints) and control (PD) parameters to joints, where the PD values are the same for all joints in one robot.

This stage trains the policy with two functionalities: 1. To make the policy adaptive to different environmental, robotic, and controlling properties; 2. To make the robot able to reach the target. In this stage we only train the Policy and the Adaptive Models and keep other models fixed. For the reward function, we only use  $r_g$  and  $r_d$  in this stage, and others are weighted by 0.

- 2) *Uneven Terrain* (Fig. 3(b)): The scenarios have different levels of unevenness generated by Perlin noise, and the heights of the peaks vary from 0 meters to 1 m. The unevenness causes vibrations when robots navigate on the terrain. The training in this type of scenario makes the Stable Model generate appropriate velocity thresholds to keep the robot stable while it is running on the uneven terrain. During this training stage, we fix the perception models, the Adaptive Model, and the Policy Model. Hence, we use  $r_s$  and  $r_g$  in the reward function only to train the threshold of the actions; other rewards are weighted by 0.
- 3) *Scenarios with Obstacles* (Fig. 3(c)): This scenario has dense static obstacles (two to three meters apart). The obstacles consist of general features (corners, curves, lines) w.r.t. Lidar data. Hence, training in this stage encodes the distance information of the nearby obstacles by the Lidar Model. The input of the Lidar Model is the 1-channel Lidar data. In this stage, we only train the Lidar Model, while the Policy Model and other models are fixed. The reward function depends on  $r_g$  and  $r_c$ ; other rewards are weighted by 0.
- 4) *Scenarios with different hills* (Fig. 3(d)): The hills have different heights and gradients, which are generated randomly within the ranges  $[0.1 \text{ m}, 2 \text{ m}]$  and  $[10^\circ, 50^\circ]$ , respectively. In some regions with large gradients, the robot will flip over if the pitch or roll angles are significantly high. Hence, this stage is used to train the Elevation-map Model to encode the elevation information and produce a policy that can run the robot with a smaller change in elevation in each step. To this end, we use  $r_g$  and  $r_h$  in the reward function, and other reward items are weighted by 0.
- 5) *Unity Scenarios* (Fig. 3(e)): The unity simulator consists with a feature-rich outdoor environment. Lidar and elevation data, specifically, are more complex and accurate than the Raisim simulator. We fine-tune the perception models in the Unity simulator to equip them with better encoding of the environment, i.e., we keep non-zero weights only for  $r_g$  and  $r_h$ .

During training, except for the elevation map, all other inputs are directly from sensors. For elevation maps, we use the elevation mapping package [35]. However, the generated elevation map cannot cover all the areas near the robot without performing a complete exploration of the environment. Further, the elevation map is processed at a relatively slower rate. We alleviate the issues with the following two steps:

- 1) We use the linear interpolation method to fill the unknown regions in the elevation map and solve the sparsity issue of the generated elevation map.

TABLE I

SIMULATION RESULTS: THE LOWER THE VIBRATION, ELEVATION CHANGES, AND DRIFT VALUES, THE BETTER (MORE RELIABLE) THE PERFORMANCE. ADAPTIVEON SHOWS BETTER STABILITY ON UNEVEN TERRAIN, ENSURES LESS DRIFTING AND SMALLER ELEVATION CHANGES, AND IS MORE RELIABLE THAN OTHER METHODS. (NOTE EGO-GRAPH AND THE POLICY WITHOUT TRAINING IN SCENARIO C (STAGE 3), IN FIG. 3, DO NOT AVOID OBSTACLES, AND THUS THEY ARE NOT INCLUDED IN THE TABLE).

Scenarios	Methods	Trajectory Length (m)	Time (s)	Success Rate	Vibration (r/step)	Elevation Changes (cm/step)	Drift (r/step) $\times 100$
Uneven Terrain	AdaptiveON	7.25	35.5	1.0	<b>0.79</b>	0.19	<b>4.90</b>
	AdaptiveON Before Stage 2	6.83	7.80	1.0	0.85	0.47	6.94
	Base Policy	6.30	36.7	1.0	0.93	0.19	5.12
	Ego-graph	7.31	9.05	1.0	0.84	0.53	11.09
	CrowdSteer	7.11	8.00	1.0	0.81	0.45	5.46
	Terp	7.21	14.2	1.0	0.85	0.30	8.23
Static Obstacles	AdaptiveON	9.5	23	0.9	0.9	0.34	3.8
	AdaptiveON Before Stage 3	-	-	0	-	-	-
	Ego-graph	-	-	-	-	-	-
	CrowdSteer	7.82	9.2	<b>1.0</b>	0.93	0.45	8.19
Hills	Terp	8.73	14.9	0.7	0.93	0.31	3.99
	AdaptiveON	11.50	40.7	1.0	0.7	<b>0.12</b>	1.5
	AdaptiveON Before Stage 4	8.51	43.3	1.0	0.81	0.15	1.34
	Ego-graph	8.37	13.2	1.0	1.5	0.20	7.73
	CrowdSteer	10.9	12.3	0.9	1.4	0.21	3.22
	Terp	7.68	18.1	1.0	0.7	0.18	9.75

- 2) To mitigate the slower processing rate, we first use the ground truth elevation map during the training in the Raisim simulator to let the robot learn how to achieve the goal by running on the areas with lower elevation changes. Then, we train the model in the Unity simulator with richer environmental features and a real elevation map.

For each stage of training, we use the trained model from its last training stage, and the training curves are shown in the Supplement in [34].

#### IV. BENCHMARKS

To show the benefits of our approach in terms of stability (less vibration) and reliability (low elevation changes, less drifting, fewer collisions), we perform experiments in simulation and in real-world scenarios to compare against Terp [3], CrowdSteer [30], and Ego-graph [20] algorithms, which are discussed in Sections I and II, by measuring the following metrics: trajectory length, vibration, success rate, elevation changes, and drifts. Additionally, we compare the versions before/after different training stages in simulation to show the benefits of each model, summarized in Table I.

Ego-graph uses an elevation map as the cost map to choose the waypoints and moves the robot to follow the waypoints. Similar to Ego-Graph, Terp also generates waypoints in the elevation map, but it uses the DWA [19] motion planning algorithm for navigation. CrowdSteer is an end-to-end motion planning method with Lidar and IMU data as its inputs.

##### A. Simulation

We trained and tested the policy on workstations with an Intel Xeon 3.6 GHz CPU and an Nvidia Titan GPU. We evaluated the approaches in simulation using the Raisim simulator, and the results are summarized in Table I. The Base Policy in Table I is the policy trained only in one scenario, which means the policy cannot adapt to different terrain and robot configurations. In Table I, the final fine-tuned AdaptiveON has 15% improvement

in stability and drift reduction than the base policy in uneven terrain and reduces elevation changes in average 20% at each step in hills scenarios. We generated three different scenarios: 1. Uneven Terrain with different frictions and high bumpiness; 2. Static Obstacles with some small variations in friction and bumpiness; 3. Hills with high elevation differences from bottom to top. Each of the methods was executed 100 times in each of these scenarios to calculate the following metrics.

The **vibration** metric is measured as the average values of vibrations accumulated along the trajectory:

$$\text{vibration} = \frac{1}{N} \sum_{t=0}^N (vb_t), \quad (9)$$

Each vibration at time step  $t$  is defined in (1), and  $N$  is the final time step. Our approach is approximately 5 times more stable on uneven terrains than the other methods. The Stable Model ensures lower vibrations by running the robot slower on the bumpy areas.

From observation, the linear velocity doesn't change much in locomotion on slippery terrain and drifts are mostly caused by turning. Therefore, the **drift** metric is measured by the heading error, which is the difference between the actual direction and the calculated direction, which is calculated by the last direction and the given action.

$$\text{drift} = \frac{1}{N} \sum_{t=0}^N \|d_{a,t+1} - d_{c,t+1}(\mathbf{a}_t, d_{a,t})\|_2, \quad (10)$$

where  $p_{a,t+1}$  is the actual position of the next time step  $t+1$ , and  $p_{c,t+1}$  is the calculated next position at  $t+1$  based on the velocity and current position. The Adaptive Model can alleviate the drifting issues. From observation, it can drive the robot with lower angular velocity when the robot runs on slippery terrains.

The elevation changes metric is defined as follows:

$$\text{elevation\_change} = \frac{1}{N} \sum_{t=0}^N |e_{t+1} - e_t|, \quad (11)$$

where  $e_t$  is the elevation of the robot in the time step  $t$ . To highlight the performance for climbing hills, we consider scenarios with significantly low bumpiness and high friction. Then, we compare the algorithms with different start and goal locations. From Table I, the 0.9 success rate of CrowdSteer is caused by the fact that the algorithm cannot detect elevation changes, and the robot flips over when it runs on higher elevation gradients. Compared to the methods, our algorithm results in relatively lower elevation changes and a higher success rate, although the trajectory is longer than other algorithms.

As a baseline of navigation, we also tested collision avoidance in dense scenarios with objects 2 m apart. AdaptiveOn has a relatively high success rate compared with other approaches (Table I). Note that we did not include a comparison for Ego-Graph for the Static Obstacles scenario as it does not provide the ability to specifically handle obstacles.

TABLE II

REAL-WORLD DEPLOYMENT RESULTS: THE LOWER THE VIBRATION, ELEVATION CHANGES, AND DRIFT VALUES, THE BETTER (MORE RELIABLE) THE PERFORMANCE. COMPARED TO OTHER METHODS, OUR APPROACH HAS AT LEAST 30.7% IMPROVEMENT IN STABILITY ON UNEVEN TERRAIN, AT LEAST 8.08% REDUCTION OF DRIFTS ON SLIPPERY TERRAIN, AND AT LEAST 14.75% IMPROVEMENT ON THE ELEVATION CHANGE IN EACH STEP ON THE HILLS.

Scenarios	Methods	Trajectory Length (m)	Time (s)	Vibration (r/step) $\times 100$	Elevation Changes (cm/step) $\times 100$	Drift (r/step) $\times 100$
Uneven Terrain	AdaptiveON	9.56	10.8	17.39	0.05	2.16
	CrowdSteer	9.16	12.0	27.19	0.01	2.10
	Terp	9.81	11.1	29.11	0.20	6.50
	Ego-graph	9.48	12.0	25.10	0.12	5.33
Slippery Terrain	AdaptiveON	5.05	10.2	24.81	0.03	3.69
	CrowdSteer	6.79	11.9	45.05	0.02	15.90
	Terp	7.69	12.1	38.70	0.03	4.01
	Ego-graph	8.10	8.4	32.31	0.04	12.32
Hills	AdaptiveON	13.24	24.2	10.23	0.52	0.76
	CrowdSteer	11.38	14.4	22.71	0.78	1.95
	Terp	11.13	14.5	30.34	0.61	1.84
	Ego-graph	10.30	10.9	23.15	0.70	0.97

### B. Field Experiments

To demonstrate that our approach can adapt to other differential driving robots, we implement the approach on both Jackal and Husky robots (Fig. 1). Both robots are equipped with Velodyne VLP16 3D Lidar mounted on top of the robot and IMU (we use a Zed camera to get IMU data because it is more accurate and less noisy than the robot's). The computational platforms are composed of an Intel i9 CPU and an Nvidia RTX 2080 GPU. The running speed of the algorithm is around 5 Hz.

The Normal Distributions Transform (NDT) [36] method supports a larger range of initial pose estimates and is faster than the Iterative Closest Points algorithm (ICP) [37]. Thus, for localization and mapping, we use NDT and the Lightweight and Ground-Optimized Lidar Odometry and Mapping (LeGO-LOAM) [38] method, respectively.

Collision avoidance of static obstacles is the baseline for real-world testing, so we only evaluate the performance of AdaptiveON with other methods (CrowdSteer, Ego-Graph, and Terp) on different challenging terrains with the same starting and goal positions, as shown in Fig. 4. We compared these approaches by taking the average values of 4 runs of each approach. The comparison is summarized in Table II, and the results correspond to the trajectory curves in Fig. 4. All the metrics are measured as they are in simulation. The real-world scenarios mostly merged different properties of terrain, e.g., unevenness, different elevation changes, obstacles, slippery ground, etc. However, each scenario has a different major navigation challenge for the robot.

*Uneven Terrain:* Scenario (a) tests the stability on uneven terrains, where the grass and the edge of the pavement cause vibrations. The curb of the grassy area is significantly bumpy due to the sudden elevation changes. From the trajectories, our approach and Terp tend to prevent running straight across the edge, while CrowdSteer and Ego-Graph navigate on the concrete area with larger pitch angles. Fig. 5 shows the stability changes on this terrain, where the vibration is measured by (1). We observe that our approach results in less vibration than other methods. The last part of the trajectory is on the concrete area, where all the methods have relatively small vibration values. The



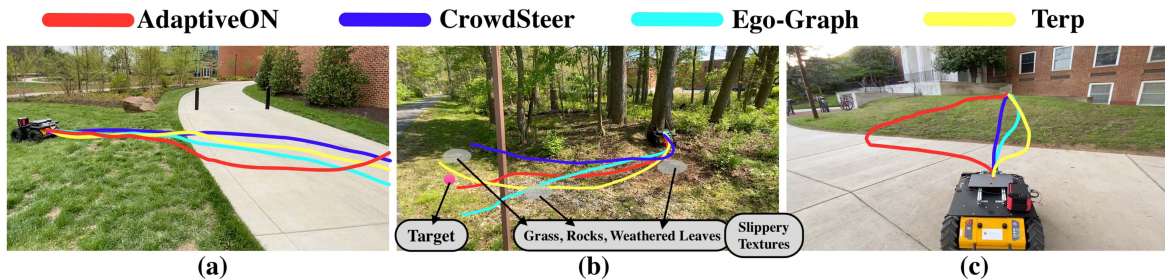


Fig. 4. Real-world deployment of the Husky robot on different terrains: (a) bumpy terrain with small slope gradient; (b) unstructured slippery terrain (grass, rocks, or weathered leaves), which can cause significant drifts when robots take sharp turns; (c) a hill with a high elevation difference between the bottom and the top. In the last scenario, we can see our method can deviate to the area with a smaller gradient. (Note the trajectories of different methods are marked with different colors.).

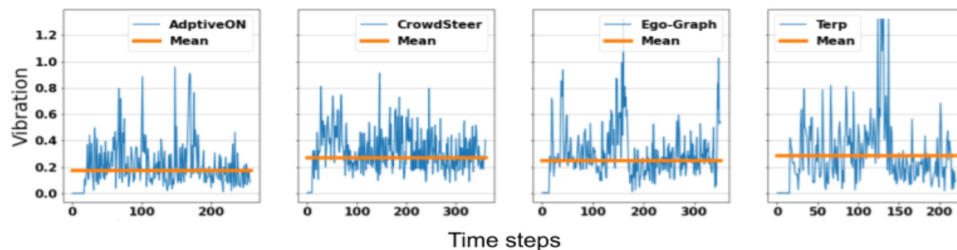


Fig. 5. Stability plots of the scenario in Fig. 4(a). The lower the value, the less vibration the robot experiences. For all the methods, vibration changes from 0 to 1.4, and our approach has a relatively low vibration without large fluctuations. The peaks occur when the robot runs onto pavement from the grass region.

peaks of the graphs are when the robot runs across the edge of the grass terrain. We can see our method has lower values for and less fluctuation of vibration. In contrast, CrowdSteer operates at high speeds, which makes the average vibration amplitude and variation significantly high. Ego-Graph and Terp tend to choose areas with lower elevation changes, but they depend heavily on waypoints instead of actions, and the vibration values are very high on the grass regions. We observe that the peak values for Terp and Ego-Graph are lower than CrowdSteer; on the edge of the grassy area, CrowdSteer rushes the robot onto the concrete, which causes significantly high pitch angle changes, whereas other methods navigate from a smaller angle with the curb. As shown in Table II, our method has at least 30.7% improvement in stability.

*Slippery and Unstructured Terrain:* Scenario (b) is a slippery terrain with unstructured features. The rocks are loose and withered leaves are soft. Combined with bumpy grass textures, the scenario can easily lead to drifting when robots move on the terrain. This scenario also has trees that the robot must avoid. The purple point is the target for all the methods. From the trajectories, we can observe that CrowdSteer deviates most because it moves the robot very fast with sharp turns. AdaptiveON and Terp have similar deviations. As shown in Table II, our approach reduces the drifting at least by 8.08%.

*Hills:* Scenario (c) contains a hilly region with high elevation changes and inconsistent elevation gradients. Further, the grass surface is bumpy and slightly slippery for the robot. We observe that AdaptiveON can navigate the robot along trajectories with lower elevation changes and lower vibrations. Although Terp can also choose lower elevations, its trajectory optimization could lead to relatively higher elevations when the waypoints

are significantly far away from the robot. Hence, as presented in Table II, we can see that our method provides at least a 14.75% decrease in the elevation changes when navigating in hilly regions.

## V. CONCLUSION

We presented AdaptiveON - a reinforcement learning-based method with a multi-stage training pipeline for stable and reliable outdoor navigation. We showed the performance on complex outdoor terrains through both simulation and real-world deployments. Our method is modular and can be used as a baseline for adding other functionalities. It is possible to add new models using similar architecture and training strategies to achieve different functions.

Nevertheless, there are still some limitations. The stable model depends on the vibration of the last three steps, which is latent and causes pick values of vibrations when the robot runs on a bumpy area in the current step. In addition, there are still many other complex scenarios our approach could not solve. e.g., crossing the area with tall grass, staying away from cliffs, or detecting fast-moving obstacles.

In the future, this work can be improved in several respects: To prevent the interference between preventing drifting and making robots run stably, those two functions can be decoupled, and the robot should prioritize using functions in different situations. More perceptive features can be used to predict the bumpiness of the terrain to reduce the extreme values. It can be integrated with a high level path planning algorithms that generates waypoints, to ensure more stable and reliable navigation between waypoints.

## REFERENCES

- [1] G. Kahn, P. Abbeel, and S. Levine, "BADGR: An autonomous self-supervised learning-based navigation system," *IEEE Robot. Automat. Lett.*, vol. 6, no. 2, pp. 1312–1319, Apr. 2021.
- [2] A. Nagariya and S. Saripalli, "An iterative LQR controller for off-road and on-road vehicles using a neural network dynamics model," in *Proc. IEEE Intell. Veh. Symp.*, 2020, pp. 1740–1745.
- [3] K. Weerakoon, A. J. Sathiamoorthy, U. Patel, and D. Manocha, "TERP: Reliable planning in uneven outdoor environments using deep reinforcement learning," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2022, pp. 9447–9453.
- [4] T. Guan, D. Kothandaraman, R. Chandra, A. J. Sathiamoorthy, and D. Manocha, "GANav: Group-wise attention network for classifying navigable regions in unstructured outdoor environments," *IEEE Robot. Automat. Lett.*, vol. 7, no. 3, pp. 8138–8145, 2022.
- [5] L. Wellhausen, A. Dosovitskiy, R. Ranftl, K. Walas, C. Cadena, and M. Hutter, "Where should I walk? predicting terrain properties from images via self-supervised learning," *IEEE Robot. Automat. Lett.*, vol. 4, no. 2, pp. 1509–1516, Apr. 2019.
- [6] S. Wapnick, T. Manderson, D. Meger, and G. Dudek, "Trajectory-constrained deep latent visual attention for improved local planning in presence of heterogeneous terrain," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2021, pp. 460–467.
- [7] L. Yihuan, L. Daokui, and T. Guojin, "Motion planning for vibration reducing of free-floating redundant manipulators based on hybrid optimization approach," *Chin. J. Aeronaut.*, vol. 24, no. 4, pp. 533–540, 2011.
- [8] N. Hirose, A. Sadeghian, M. Vázquez, P. Goebel, and S. Savarese, "GONet: A semi-supervised deep learning approach for traversability estimation," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2018, pp. 3044–3051.
- [9] K. Viswanath, K. Singh, P. Jiang, P. Sujit, and S. Saripalli, "Offseg: A semantic segmentation framework for off-road driving," in *Proc. IEEE 17th Int. Conf. Automat. Sci. Eng.*, 2021, pp. 354–359.
- [10] M. J. Eder, R. Halatschek, and G. Steinbauer-Wagner, "Autonomous robot navigation in alpine terrain," 2021.
- [11] J. Hwangbo et al., "Learning agile and dynamic motor skills for legged robots," *Sci. Robot.*, vol. 4, no. 26, 2019, Art. no. eaa5872.
- [12] A. Kumar, Z. Fu, D. Pathak, and J. Malik, "RMA: Rapid motor adaptation for legged robots," in *Proc. Robot.: Sci. Syst., Virtual*, 2021.
- [13] J. Tobin, R. Fong, A. Ray, J. Schneider, W. Zaremba, and P. Abbeel, "Domain randomization for transferring deep neural networks from simulation to the real world," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2017, pp. 23–30.
- [14] X. B. Peng, E. Coumans, T. Zhang, T.-W. Lee, J. Tan, and S. Levine, "Learning agile robotic locomotion skills by imitating animals," in *Robot.: Sci. Syst.*, Aug. 2020.
- [15] J. Lee, J. Hwangbo, L. Wellhausen, V. Koltun, and M. Hutter, "Learning quadrupedal locomotion over challenging terrain," *Sci. Robot.*, vol. 5, no. 47, 2020, Art. no. eabc5986.
- [16] H. Hu, K. Zhang, A. H. Tan, M. Ruan, C. Agia, and G. Nejat, "A sim-to-real pipeline for deep reinforcement learning for autonomous robot navigation in cluttered rough terrain," *IEEE Robot. Automat. Lett.*, vol. 6, no. 4, pp. 6569–6576, Oct. 2021.
- [17] J. Hwangbo, J. Lee, and M. Hutter, "Per-contact iteration method for solving contact dynamics," *IEEE Robot. Automat. Lett.*, vol. 3, no. 2, pp. 895–902, Apr. 2018.
- [18] J. V. D. Berg, J. Snape, S. J. Guy, and D. Manocha, "Reciprocal collision avoidance with acceleration-velocity obstacles," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2011, pp. 3475–3482.
- [19] D. Fox, W. Burgard, and S. Thrun, "The dynamic window approach to collision avoidance," *IEEE Robot. Automat. Mag.*, vol. 4, no. 1, pp. 23–33, Mar. 1997.
- [20] A. Lacaze, Y. Moscovitz, N. DeClariss, and K. Murphy, "Path planning for autonomous vehicles driving over rough terrain," in *Proc. IEEE Int. Symp. Intell. Control Held Jointly IEEE Int. Symp. Comput. Intell. Robot. Automat. Intell.*, 1998, pp. 50–55.
- [21] G. Kahn, A. Villafior, V. Pong, P. Abbeel, and S. Levine, "Uncertainty-aware reinforcement learning for collision avoidance," 2017, *arXiv:1702.01182*.
- [22] J. Ibarz, J. Tan, C. Finn, M. Kalakrishnan, P. Pastor, and S. Levine, "How to train your robot with deep reinforcement learning: Lessons we have learned," *Int. J. Robot. Res.*, vol. 40, no. 4–5, pp. 698–721, 2021.
- [23] K. Weerakoon, A. J. Sathiamoorthy, J. Liang, T. Guan, U. Patel, and D. Manocha, "GrASPE: Graph based multimodal fusion for robot navigation in unstructured outdoor environments," 2022, *arXiv:2209.05722*.
- [24] G. Brunner, O. Richter, Y. Wang, and R. Wattenhofer, "Teaching a machine to read maps with deep reinforcement learning," in *Proc. AAAI Conf. Artif. Intell.*, 2018, pp. 2763–2770.
- [25] A. Staroverov, D. A. Yudin, I. Belkin, V. Adeshkin, Y. K. Solomentsev, and A. I. Panov, "Real-time object navigation with deep neural networks and hierarchical reinforcement learning," *IEEE Access*, vol. 8, pp. 195608–195621, 2020.
- [26] M. Pfeiffer et al., "Reinforced imitation: Sample efficient deep reinforcement learning for mapless navigation by leveraging prior demonstrations," *IEEE Robot. Automat. Lett.*, vol. 3, no. 4, pp. 4423–4430, Oct. 2018.
- [27] K. Weerakoon, S. Chakraborty, N. Karapetyan, A. J. Sathiamoorthy, A. S. Bedi, and D. Manocha, "HTRON: Efficient outdoor navigation with sparse rewards via heavy tailed adaptive reinforcement algorithm," in *Proc. Conf. Robot. Learn.*, Jun. 2022.
- [28] A. J. Sathiamoorthy, J. Liang, U. Patel, T. Guan, R. Chandra, and D. Manocha, "Densecavoid: Real-time navigation in dense crowds using anticipatory behaviors," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2020, pp. 11345–11352.
- [29] A. M. Roth, J. Liang, and D. Manocha, "XAI-N: Sensor-based robot navigation using expert policies and decision trees," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2021, pp. 2053–2060.
- [30] J. Liang, U. Patel, A. J. Sathiamoorthy, and D. Manocha, "Crowd-steer: Realtime smooth and collision-free robot navigation in densely crowded scenarios trained using high-fidelity simulation," in *Proc. 29th Int. Conf. Int. Joint Conf. Artif. Intell.*, 2021, pp. 4221–4228.
- [31] J. Liang, Y.-L. Qiao, T. Guan, and D. Manocha, "OF-VO: Efficient navigation among pedestrians using commodity sensors," *IEEE Robot. Automat. Lett.*, vol. 6, no. 4, pp. 6148–6155, Oct. 2021.
- [32] K. Zhang, F. Niroui, M. Ficocelli, and G. Nejat, "Robot navigation of environments with unknown rough terrain using deep reinforcement learning," in *Proc. IEEE Int. Symp. Safety, Secur., Rescue Robot.*, 2018, pp. 1–7.
- [33] S. Josef and A. Degani, "Deep reinforcement learning for safe local planning of a ground vehicle in unknown rough terrain," *IEEE Robot. Automat. Lett.*, vol. 5, no. 4, pp. 6748–6755, Oct. 2020.
- [34] J. Liang, K. Weerakoon, T. Guan, N. Karapetyan, and D. Manocha, "AdaptiveON: Adaptive outdoor navigation method for stable and reliable motions," 2022. [Online]. Available: <https://arxiv.org/abs/2205.03517>
- [35] P. Fankhauser, M. Bloesch, and M. Hutter, "Probabilistic terrain mapping for mobile robots with uncertain localization," *IEEE Robot. Automat. Lett.*, vol. 3, no. 4, pp. 3019–3026, Oct. 2018.
- [36] M. Magnusson, A. Lilienthal, and T. Duckett, "Scan registration for autonomous mining vehicles using 3D-NDT," *J. Field Robot.*, vol. 24, no. 10, pp. 803–827, 2007.
- [37] M. Magnusson, A. Nuchter, C. Lorken, A. J. Lilienthal, and J. Hertzberg, "Evaluation of 3D registration reliability and speed—a comparison of ICP and NDT," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2009, pp. 3907–3912.
- [38] T. Shan and B. Englot, "LeGO-LOAM: Lightweight and ground-optimized lidar odometry and mapping on variable terrain," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2018, pp. 4758–4765.