CAHSOR: Competence-Aware High-Speed Off-Road Ground Navigation in $\mathbb{SE}(3)$

Anuj Pokhrel, Mohammad Nazeri, Aniket Datar, and Xuesu Xiao
George Mason University
{apokhre, mnazerir, adatar, xiao}@gmu.edu

Abstract—While the workspace of traditional ground vehicles is usually assumed to be in a 2D plane, i.e., $\mathbb{SE}(2)$, such an assumption may not hold when they drive at high speeds on unstructured off-road terrain: High-speed sharp turns on high-friction surfaces may lead to vehicle rollover; Turning aggressively on loose gravel or grass may violate the non-holonomic constraint and cause significant lateral sliding; Driving quickly on rugged terrain will produce extensive vibration along the vertical axis. Therefore, most offroad vehicles are currently limited to drive only at low speeds to assure vehicle stability and safety. In this work, we aim at empowering high-speed off-road vehicles with competence awareness in $\mathbb{SE}(3)$ so that they can reason about the consequences of taking aggressive maneuvers on different terrain with a 6-DoF forward kinodynamic model. The model is learned from visual and inertial Terrain Representation for Off-road Navigation (TRON) using multimodal, self-supervised vehicle-terrain interactions. We demonstrate the efficacy of our Competence-Aware High-Speed Off-Road (CAHSOR) navigation approach on a physical ground robot in both an autonomous navigation and a human shared-control setup and show that CAHSOR can efficiently reduce vehicle instability by 62% while only compromising 8.6% average speed with the help of TRON.

I. INTRODUCTION

Autonomous mobile robot navigation has been a research topic in the robotics community for decades [10, 33]. Being equipped with perception, planning, and control techniques, different types of ground robots, e.g., differential-drive or Ackermann-steering, are able to efficiently move toward their goals in their 2D workspaces considering their 3-DoF motion models (x, y, and yaw) without colliding with obstacles, mostly in structured and homogeneous environments [53, 55].

Bringing those robots into the unstructured real world, researchers have also investigated off-road navigation since the DARPA Grand Challenge [37] and LAGR (Learning Applied to Ground Vehicles) Program [14]. While significant research effort on off-road navigation focuses on the perception side [24, 46, 2, 23], researchers have also investigated off-road mobility, including inverse [16, 48] and forward [1, 22] kinodynamics, wheel slip modeling [35, 34], and end-to-end learning [29, 39]. Most off-road robots drive at slow speeds to assure vehicle stability and safety [7, 8]. Even when aiming at driving fast, they still assume a simplified 2D workspace and 3-DoF model in $\mathbb{SE}(2)$ despite the highly likely disturbances from the off-road terrain on other dimensions of the state space (e.g., drift along y, roll around x, or bumpiness along z). These realistic kinodynamic effects may be tolerable in some cases, but may lead to catastrophic consequences in others with increasing speed on unstructured terrain (Fig. 1).

Fig. 1. Challenges of High-Speed Off-Road Ground Navigation in $\mathbb{SE}(3)$.

To enable safe and robust off-road navigation, high-speed ground robots need to be competence-aware, i.e., knowing what is the consequence of taking an aggressive maneuver on different off-road terrain. For example, a sharp turn on high-friction pavement may lead to vehicle rollover (Fig. 1 top); Blasting through rugged surfaces can generate extensive vertical vibrations and damage onboard components (Fig. 1 bottom left); Aggressive swerving on loose grass or gravel will cause the vehicle to slide sideways and risk collision or falling off a cliff (Fig. 1 bottom right).

To this end, we propose a Competence-Aware High-Speed Off-Road (CAHSOR) ground navigation approach based on a 6-DoF forward kinodynamic model in $\mathbb{SE}(3)$. The model is learned as a downstream task of a new Terrain Representation for Off-road Navigation (TRON) approach with multimodal, self-supervised learning using viewpoint-invariant visual terrain patches and underlying Inertia Measurement Unit (IMU) responses during vehicle-terrain interactions. CAHSOR learns to predict potential next states according to different candidate actions and the current visual and/or inertial terrain representation to make competence-aware decisions in order to maximize speed while satisfying 6-DoF vehicle stability constraints in $\mathbb{SE}(3)$, e.g., without excessive sliding and rolling motions or bumpy vibrations. Our contributions can be summarized as:

- a TRON approach with multimodal self supervision that

allows onboard visual and inertial observations to augment each other and maximizes the information embedded in the representation of each perceptual modality;
- a comprehensive study of various end-to-end and representation learning techniques with different modalities for different off-road kinodynamic modeling tasks;
- a CAHSOR framework for high-speed off-road vehicles to take aggressive maneuvers with stability and safety; and
- a set of real-world, off-road robot experiments to demonstrate the effectiveness of CAHSOR based on TRON in both an autonomous navigation and a human shared-control setup, exhibiting 62% vehicle instability reduction while only compromising 8.6% average speed.

II. RELATED WORK

We review related robot navigation research focusing on off-road conditions and using machine learning approaches.

A. Off-Road Navigation

Since the DARPA Urban Challenge [37] and LAGR Program [14], robotics researchers have investigated autonomous navigation techniques for off-road conditions. Going beyond the simple delineation of the workspace into obstacles and free spaces, the first challenge comes from robot perception, e.g., identifying semantic information such as pavement, gravel, grass, pebble, and mud. Terrain classification methods perceive the underlying terrain and make navigation decisions tailored to the terrain class [2, 3, 4], while terrain segmentation approaches use terrain semantics and build traversability costmaps to inform subsequent planning [24, 46, 23, 18, 9, 41, 40]. Furthermore, high-speed off-road navigation has been investigated by robotics researchers from the mobility side. End-to-end learning approaches allow aggressive off-road driving in a closed circuit [29]. More structured approaches have been taken to learn inverse [16, 48] and forward [1, 22] kinodynamics and wheel slip models [35, 34]. In most aforementioned approaches, ground robots are treated as vehicles moving in a 2D workspace with 3 DoFs (x, y, and yaw). Recently, researchers have also started looking into autonomous off-road crawling, in which vehicles slowly drive on non-flat, extremely rugged rocks and boulders [7, 8], requiring the vehicle state space to be expanded into $\mathbb{SE}(3)$. Our CAHSOR also operates in $\mathbb{SE}(3)$, but our 6-DoF forward kinodynamic model aims at confidently navigating ground robots at the maximum possible speed on various off-road terrain while maintaining stability on other state dimensions, e.g., drift along y, roll around x, or bumpiness along z, which are often ignored by traditional ground robot models.

B. Machine Learning for Navigation

Recently, machine learning approaches have been widely adopted to enable autonomous mobile robot navigation [51]. Those learning methods enable navigation behaviors in a data-driven manner without the need of manually designing systems and components which tend to fail to capture the complexities and intricacies in the real world. In particular,

$$\mathbf{s}_t = [x_t, y_t, z_t, r_t, p_t, \phi_t]^T \in \mathbb{SE}(3)$$
$$\mathbf{s}_{t+1} = f_\theta(\mathbf{s}_t, \mathbf{u}_t, \psi_t^{V, S, I})$$
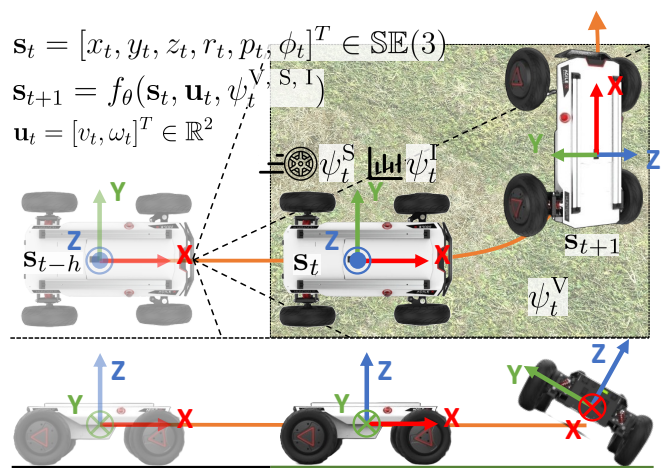$$\mathbf{u}_t = [v_t, \omega_t]^T \in \mathbb{R}^2$$

Fig. 2. Overview of the Competence-Aware High-Speed Off-Road (CAHSOR) Ground Navigation (Side and Top View).

researchers have used machine learning to learn end-to-end systems [17, 29, 32, 5], local planners [58, 21, 11, 43, 49, 50], planner parameterization [47, 42, 44, 56, 52, 57], kinodynamic models [16, 48, 1, 26], and cost functions [54, 38, 20, 19, 36, 6], for tasks like highly-constrained [53, 55, 43, 49, 50, 31, 27], social [30, 54, 15, 28, 12, 20, 5, 19, 13, 25], and off-road navigation [29, 16, 48, 1]. Considering the complexity of unstructured off-road terrain and their intricate effects on vehicle kinodynamics at high speeds, we also adopt a data-driven method for CAHSOR to learn a forward kinodynamic function conditioned on viewpoint-invariant visual and underlying inertial terrain representation learned by TRON using multimodal, self-supervised vehicle-terrain interaction experiences.

III. APPROACH

We formulate the problem of forward kinodynamics modeling in $\mathbb{SE}(3)$ for ground robots driving on unstructured off-road terrain at high speeds, present a multimodal self-supervised learning approach to represent off-road conditions using onboard visual and inertial observations, introduce a data-driven approach to learn the forward kinodynamic model from past vehicle-terrain interactions, and develop a competence-aware navigation framework that allows robots to drive at the maximum possible speed while maintaining vehicle stability in $\mathbb{SE}(3)$.

A. Forward Ground Kinodynamics in $\mathbb{SE}(3)$

We adopt a forward kinodynamics formulation for ground robots to reason about the consequences of taking different aggressive maneuvers on various off-road terrain. We denote vehicle state as $\mathbf{s}$, which includes 6-DoF vehicle pose in $\mathbb{SE}(3)$ (x, y, z, roll r, pitch p, and yaw ϕ, expressed in the global or robot frame, Fig 2) and their corresponding velocity components. For brevity, only the pose components are included in the following derivation. The vehicle control $\mathbf{u} = [v, \omega]^T$ contains linear velocity and angular velocity (for differential-driven vehicles, or steering curvature for Ackermann-steering

vehicles). We use a world state $\mathbf{w}$ to denote all necessary effects from the environment that will affect kinodynamics, in our case, from unstructured off-road terrain. Therefore, in a discrete setting, we have

$$\mathbf{s}_{t+1} = f(\mathbf{s}_t, \mathbf{u}_t, \mathbf{w}_t), \quad \mathbf{o}_t = g(\mathbf{s}_t, \mathbf{w}_t),$$
$$\mathbf{s}_t = [x_t, y_t, z_t, r_t, p_t, \phi_t]^T \in \mathbb{SE}(3), \ \mathbf{u}_t = [v_t, \omega_t]^T \in \mathbb{R}^2,$$

where $f(\cdot)$ is a forward kinodynamic function in $\mathbb{SE}(3)$, while $g(\cdot)$ is an observation function. For off-road driving, the forward kinodynamic function $f(\cdot)$ also takes in the world state $\mathbf{w}$ as input, in contrast to models derived for structured and homogeneous terrain that only need to consider vehicle state $\mathbf{s}_t$ and control $\mathbf{u}_t$ alone. For example, slippery and rugged terrain surfaces may cause extensive movement along the vehicle y and z directions respectively. However, world state $\mathbf{w}$ is usually not directly observable and cannot be easily modeled. Also notice that most kinodynamic models for ground robots only consider $\mathbf{s}_t = [x_t, y_t, \phi_t]^T \in \mathbb{SE}(2)$ and ignore all other state dimensions. Such a forward model can be used in rolling out candidate trajectories for sampling-based path and motion planners [45, 10], or its inverse form can be derived to achieve desired next state $\mathbf{s}_{t+1}^*$ [48, 16].

B. Visual and Inertial Representation of World State

Considering the difficulty in analytically modeling world state $\mathbf{w}$, we use a multimodal self-supervised learning approach to represent $\mathbf{w}$ and approximate the observation function $g(\cdot)$ with onboard visual and inertial sensors. In particular, an onboard visual camera can provide visual signature of the terrain patch λ_t the robot drives over and an IMU can sense the underlying kinodynamic responses i_t in terms of linear accelerations and angular velocities. We assume the current vehicle speed can also be observed by, e.g., odometry or GPS, denoted as s_t. While IMU readings i_t can be directly sensed when in contact with the underlying terrain, for high-speed off-road navigation, the vehicle may need to reason about future kinodynamic consequences up to a certain planning horizon, for which only visual observations from a forward-looking camera are available, not the IMU readings. So in this work, we use multimodal self-supervised learning to allow both visual and inertial observations to augment each other by correlating them in effective representation spaces, thus either (or both) can be used to enable competence awareness when available (e.g., manual shared-control using current underlying inertia and autonomous planning with vision of future terrain).

We posit that the visual and inertial observations can provide multimodal self-supervised learning signals to represent different terrain kinodynamics. To achieve such self-supervision, we use a non-contrastive approach to maximize the correlation between visual and inertial embeddings. In this way, we avoid the need to use privileged information such as terrain labels that require manual annotation. However, a key difference of high-speed off-road navigation compared to existing terrain representation learning approaches [38, 18] is that the correlation between vision and inertia is also dependent on the (high) vehicle speed: driving quickly vs. slowly on the

same visual patch of grass may produce completely different inertial responses, while different speeds on grass vs. gravel may coincidentally lead to similar IMU readings. Therefore, CAHSOR extends the vision–inertia correlation to vision & speed–inertia correlation to account for the effect caused by various speeds during high-speed off-road navigation.

However, two challenges still exist for visual representation: 1. Visual perception is very sensitive to environment conditions, such as changes in viewpoints and lighting, as well as occlusion and motion blur; 2. Unlike current IMU readings, the visual signature of the terrain underneath (or right in front of) the current robot state $\mathbf{s}_t$ can not be directly captured by the onboard camera due to limited onboard field-of-view. The robot needs to seek help from previous camera images captured before t. Therefore, we design a viewpoint-invariant visual patch extraction technique to overcome both challenges: Denote the camera image captured h time steps ahead as c_{t-h} and the transformation from time step $t - h$ to t extracted from vehicle odometry as d_{t-h}^t. By projecting c_{t-h} to an overhead Bird-Eye View (BEV) using the camera homography $h_{t-h} = H(i_{t-h})$, which is dependent on the vehicle roll and pitch angles determined by the IMU readings i_{t-h} due to aggressive off-road driving, we can extract the terrain patch currently underneath the robot, $\lambda_t = P(c_{t-h}, h_{t-h}, d_{t-h}^t)$. λ_t is designed to be slightly larger than the vehicle footprint to consider actuation latency. By varying the history length h, it is possible to generate a set of different instantiations of λ_t from different viewpoints with different lighting conditions, $\Lambda_t = \{\lambda_t^j\}_{j=1}^J$, in which each λ_t^j has at least a certain amount of visible pixels of the terrain patch underneath $\mathbf{s}_t$, considering the homography projection may cause invisible BEV pixels.

C. Terrain Representation for Off-road Navigation

The set of viewpoint-invariant visual terrain patches $\Lambda_t = \{\lambda_t^j\}_{j=1}^J$, the IMU readings i_t, and the current vehicle speed s_t correspond to multimodal perception of the robot at time t and provide self-supervised learning signals for our vision & speed–inertia correlation (Fig. 3 left). To be specific, a vision, speed, and inertia encoder embeds any terrain patch $\lambda_t \in \Lambda_t$, current vehicle speed s_t, and underlying inertial readings i_t into a visual, speed, and inertial representation, ψ_t^V, ψ_t^S, and ψ_t^I, respectively. Considering the causal relation from driving at a particular speed on a certain visual terrain patch to corresponding IMU readings, we concatenate the visual and speed representations, ψ_t^V and ψ_t^S, and further encode them into a joint vision & speed embedding ψ_t^{VS}. To correlate ψ_t^{VS} and ψ_t^I, we project them independently into a higher dimensional feature space, ρ_t^{VS} and ρ_t^I. We then maximize the correlation between ρ_t^{VS} and ρ_t^I while considering viewpoint invariance using Barlow Twins [60]:

$$\mathcal{L}_{\text{TRON}} = \mathcal{BL}_V(\rho_t^{V^1S}, \rho_t^{V^2S}) + \tag{1}$$
$$(0.5 \times \mathcal{BL}_{V^1I}(\rho_t^{V^1S}, \rho_t^I) + 0.5 \times \mathcal{BL}_{V^2I}(\rho_t^{V^2S}, \rho_t^I)),$$

where V^1 and V^2 correspond to two views of the same terrain patch to encourage viewpoint invariance, i.e., $\lambda_t^1, \lambda_t^2 \sim \Lambda_t$. $\mathcal{BL}$
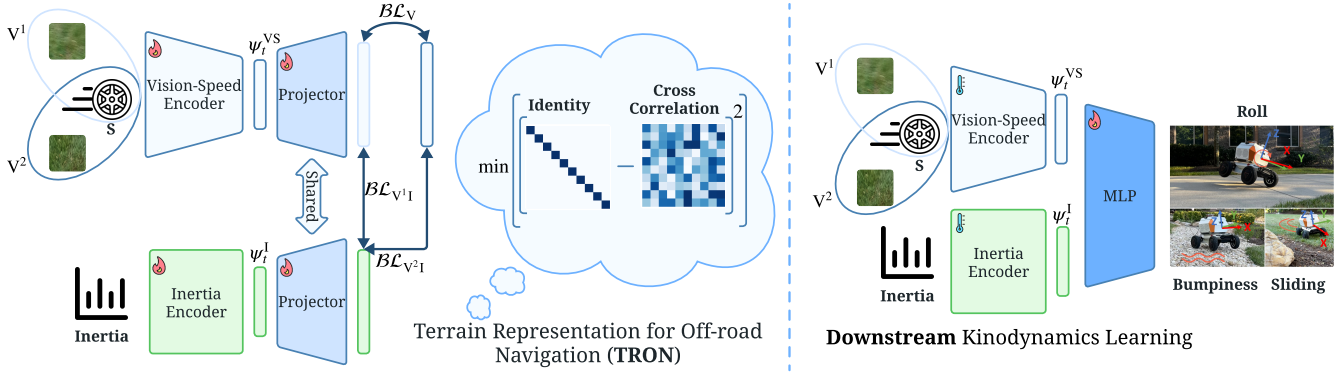
Fig. 3. TRON (Left) and Downstream Kinodynamics Learning (Right) Architecture: Flame and temperature denote training and frozen parameters respectively.

is defined as:

$$\mathcal{BL} = \sum_i (1 - \mathcal{C}_{ii})^2 + \gamma \sum_i \sum_{j \neq i} \mathcal{C}_{ij}^2, \quad (2)$$

where γ is a weight term to trade off the importance between invariance and redundancy reduction. $\mathcal{C}$ is the cross-correlation matrix computed between ρ^1 and ρ^2:

$$\mathcal{C}_{ij} = \frac{\sum_b \rho_{b,i}^1 \rho_{b,j}^2}{\sqrt{\sum_b (\rho_{b,i}^1)^2} \sqrt{\sum_b (\rho_{b,j}^2)^2}},$$

where $\rho_{b,i}^{1(2)}$ denotes the ith dimension of the bth sample in a data batch of $\rho^{1(2)}$, which can be one of $\rho_t^{V^1 S}$, $\rho_t^{V^2 S}$, or ρ_t^I.

Trained with multimodal self-supervision, the visual, speed, and inertial representation, ψ_t^V, ψ_t^S, and ψ_t^I, can be used to enable downstream kinodynamic modeling tasks. Depending on the scenario, either ψ_t^V (predicting multiple future states without terrain interactions to induce inertial responses) or ψ_t^I (directly predicting the immediate next state from the induced inertial responses from the underlying terrain), or both, may be available. For simplicity, we denote our visual-speed, inertial, or visual-speed-inertial (by concatenating the first two) representation as $\psi_t^{V, S, I}$.

D. Downstream Kinodynamic Model Learning

After learning the terrain representation and freezing the learned parameters, we also adopt a self-supervised approach to learn the forward kinodynamics due to the difficulty in analytically modeling $f(\cdot)$. We represent the unknown world state $\mathbf{w}_t$ using $\psi_t^{V, S, I}$ and learn an approximate forward kinodynamic function $f_\theta(\cdot)$ as a downstream task of TRON:

$$\mathbf{s}_{t+1} = f_\theta(\mathbf{s}_t, \mathbf{u}_t, \psi_t^{V, S, I}). \quad (3)$$

With a self-supervised vehicle-terrain interaction dataset,

$$\mathcal{D} = \{\mathbf{s}_{j+1}, \mathbf{s}_j, \mathbf{u}_j, \psi_j^{V, S, I}\}_{j=0}^{N-1},$$

of N data points, the optimal parameters θ^* can then be learned by minimizing a supervised loss function (Fig. 3 right),

$$\theta^* = \arg\min_\theta \sum_{\substack{(\mathbf{s}_{j+1}, \mathbf{s}_j, \mathbf{u}_j, \psi_j^{V, S, I}) \\ \in \mathcal{D}}} ||\mathbf{s}_{j+1} - f_\theta(\mathbf{s}_j, \mathbf{u}_j, \psi_j^{V, S, I})||. \quad (4)$$

E. Competence-Aware High-Speed Off-Road Navigation

The approximate forward kinodynamic function (Eqn. (3)) learned with the self-supervised loss (Eqn. (4)) can be combined with subsequent planners, e.g., sampling-based model predictive motion planners [45, 10], or used in human shared-control settings to enable competence-aware off-road navigation at high speeds. By rolling out the forward kinodynamic model, the robot can pick the optimal control command(s) that produces minimal cost or is most similar to human control, without violating vehicle stability constraints. While a motion planner or a human controller needs to consider a variety of costs including obstacle avoidance, goal distance, execution accuracy, etc., for simplicity, we combine all these costs into one general cost term $C(\mathbf{s}_t, \mathbf{s}_{t+1})$ and use only one time-step rollout in our presentation in order to explicitly showcase the high speed and competence awareness aspect of the navigation problem. Otherwise, the robot is solely maximizing navigation speed or following human control. Notice that it is easy to combine it with any other costs when necessary and extend to multiple time steps (see examples in Sec. V). Expressing the robot $\mathbb{SE}(3)$ state in the current robot frame (i.e., x forward, y left, and z up), the competence-aware navigation can be formulated as a constrained optimization problem:

$$\mathbf{u}_t^* = \arg\max_{\mathbf{u}_t} \left[||x_{t+1} - x_t|| - C(\mathbf{s}_t, \mathbf{s}_{t+1}) \right],$$
$$\text{s.t. all } \mathbb{SE}(3) \text{ constraints are satisfied}, \quad (5)$$
$$\mathbf{s}_{t+1} = f_\theta(\mathbf{s}_t, \mathbf{u}_t, \psi_t^{V, S, I}).$$

Notice that the objective function in Eqn. (5) can be formulated in other ways when necessary. For example, maximizing the displacement along x can be replaced by minimizing the difference between the control u and a desired manual command (see Sec. IV for two possible instantiations of Eqn. (5)). The navigation planner then finds the best control $\mathbf{u}_t^*$ to maximize speed along x (and considers other costs in $C(\cdot, \cdot)$), while in a human shared-control setup $\mathbf{u}_t^*$ aims to minimize the difference compared to human command, both without violating $\mathbb{SE}(3)$ vehicle kinodynamic constraints.

IV. IMPLEMENTATIONS

We implement CAHSOR ground navigation on a 1/6-scale autonomous vehicle, an AgileX Hunter SE, with a top speed of 4.8m/s on different off-road terrain at high speeds to demonstrate the proposed competence awareness. We collect a dataset of 30-minute vehicle-terrain interactions. The collected GPS-RTK, onboard IMU, front-facing camera, and vehicle control data are synchronized and processed into training data. We integrate the learned TRON and downstream kinodynamic models and the CAHSOR framework with an autonomous navigation planner and a human shared-control setup.

A. CAHSOR *Implementations*

1) TRON: The terrain vision encoder is a 4-layer Convolutional Neural Network (CNN) to produce a 512-dimensional viewpoint-invariant visual representation. The speed encoder is a 2-layer neural network, whose 512-dimensional output is combined with the visual representation to construct our vision-speed representation ψ_t^{VS}. The last 2-second accelerometer and gyroscope data are converted into the frequency domain using Power-Spectral Density (PSD) representation [59] before being fed into the 2-layer inertia encoder and producing a 512-dimensional inertial representation ψ_t^{I}. All encoders are trained to minimize $\mathcal{L}_{\text{TRON}}$ (Eqn. (1)).

2) Kinodynamics: Our $\mathbb{SE}(3)$ vehicle state is instantiated in the current robot frame, i.e., $[x_t, y_t, z_t, r_t, p_t, \phi_t]^T = \mathbf{0}$, and therefore omitted from the input of our forward kinodynamic model (Eqn. 3). To explicitly showcase the efficacy of the learned kinodynamic model on state dimensions beyond $\mathbb{SE}(2)$, we limit the model output to three metrics to reflect sliding along y, roll around x, and bumpiness along z, i.e., $[\texttt{sliding}_{t+1}, \texttt{roll}_{t+1}, \texttt{bumpiness}_{t+1}]^T$. While it is not necessary for the human shared-control setting, for autonomous navigation planning, other state dimensions are produced using a simple Ackermann-steering model, whose predicted 3-DoF trajectories are evaluated for competence awareness with the learned kinodynamic model. Such a practice also avoids the computation overhead of sequentially rolling out a large set of multi-step, 6-DoF candidate trajectories, which cannot be efficiently parallelized on GPUs. Using a learned kinodynamic model for both competence awareness and trajectory rollout can be done with more onboard computation. To be specific, $\texttt{sliding}_{t+1}$ is captured by the ground speed sensed by GPS-RTK projected onto the robot y axis (left); We compute the absolute angular acceleration around the x axis (front) from the gyroscope averaged over 0.1s as $\texttt{roll}_{t+1}$; $\texttt{bumpiness}_{t+1}$ is computed as the absolute jerk along the z axis (up) from the accelerometer averaged over 0.1s. As a downstream task of TRON, the kinodynamic model (Eqn. 3) is learned with three 256-64-1 neural network heads, which take as input the pretrained visual, speed, and/or inertial representation $\psi_t^{\text{V, S, I}}$ and candidate control actions $\mathbf{u}_t = (v, \omega)$ (omitted in Fig. 3 right for simplicity), to produce $\texttt{sliding}_{t+1}$, $\texttt{roll}_{t+1}$, and $\texttt{bumpiness}_{t+1}$.

B. Autonomous Navigation Planning with CAHSOR

We integrate our CAHSOR model with a Model Predictive Path Integral (MPPI) planner [45]. Our MPPI planner rolls out a set of candidate 3-DoF state trajectories using sampled action sequences and then combines those samples based on a predefined cost function. The cost function is informed by the prediction of the learned 6-DoF kinodynamic model, assigning infinitely large costs to candidate trajectories that involve sliding, roll, and bumpiness. MPPI then updates the sampling distribution to sample actions that are more likely to lead to low cost trajectories, i.e., moving the robot toward a goal at the fastest possible speed. MPPI rolls out 550 trajectories, each with 8 vehicle states. We select six goals in an outdoor off-road environment for the robot to drive to in a loop (Fig. 5). For MPPI rollouts, future terrain inertial responses are not available to the TRON model. Therefore, we only use the visual and speed representation ψ_t^{VS} as $\psi_t^{\text{V, S, I}}$ to represent the world state $\mathbf{w}_t$ associated with each future vehicle state $\mathbf{s}_t$ on the candidate trajectories. For computation efficiency, we divide the current BEV into a 15×51 grid and pick the terrain patch that is closest to $\mathbf{s}_t$ on the candidate trajectories for parallelized model query on GPU during one MPPI cycle.

C. Human-Autonomy Shared-Control with CAHSOR

We also demonstrate the use case of CAHSOR in a human-autonomy shared-control setup, in which a human driver aims at driving the robot as fast as possible, while CAHSOR takes care of satisfying all vehicle $\mathbb{SE}(3)$ constraints with the closet possible vehicle control to the human command. In this case, the objective function in Eqn. (5) becomes

$$\mathbf{u}_t^* = \arg\min_{\mathbf{u}_t} ||\mathbf{u}_t - \mathbf{u}_t^H||. \tag{6}$$

$\mathbf{u}_t^H$ is the desired human control input, which will potentially violate the $\mathbb{SE}(3)$ constraints. In this shared-control setup, both inertia and vision (from past camera images) are available, so TRON takes in visual, speed, and inertial representation as $\psi_t^{\text{V, S, I}}$ to represent the current world state $\mathbf{w}_t$.

V. EXPERIMENTS

We deploy CAHSOR navigation on the Hunter SE in locations different from where the training data is collected, but with similar terrain types. We only expect the models to generalize to similar terrain with different characteristics, e.g., grassy area of different densities, rocky patches with different rock sizes, and cement pavement under different lighting conditions, but not to completely unseen terrain, e.g., mud or gravel, for which more training data may be necessary.

A. TRON *Learning Results*

To demonstrate the effectiveness of the multimodal, i.e., visual, speed, and inertial, and self-supervised TRON learning, we present TRON's downstream kinodynamics learning results compared against a set of baselines. To be specific, we implement END-TO-END kinodynamics learning from vision (V), inertia (I), vision & inertia (VI), vision &

Loss	END-TO-END						STERLING		TRON		
	V	I	VI	VS	IS	VSI	V	VI	I	VS*	VSI*
Roll	0.32	0.10	0.07	0.23	0.11	0.09	0.27	0.05	0.03	0.08	**0.02**
Sliding	0.69	0.61	0.54	0.46	0.45	0.42	0.62	0.47	0.44	0.35	**0.29**
Bumpiness	0.24	0.05	0.04	0.17	0.06	0.05	0.17	0.04	0.02	0.06	**0.02**
Combined	0.72	0.24	0.21	0.30	0.21	0.18	0.36	0.18	0.15	0.16	**0.10**

TABLE I

KINODYNAMICS LOSS WITH DIFFERENT REPRESENTATIONS OF VISION (V), INERTIA (I), AND SPEED (S) USING END-TO-END, STERLING, AND TRON. * DENOTES THE MODELS DEPLOYED ON THE PHYSICAL ROBOT.

speed (VS), inertia & speed (IS), and all three (VSI), both as an ablation study for the proposed TRON learning and to analyze the information contained in each (and different combinations of) perceptual modality. We also implement a representation learning approach to correlate only vision and inertia and deploy with vision only, without considering speed, similar to STERLING [18]. TRON is experimented with inertia, vision & speed, and vision & speed & inertia representation. Table I shows the MSE loss on downstream kinodynamics learning tasks of predicting roll, sliding, bumpiness, and all three combined.

As END-TO-END results show, vision contains the least amount of information for kinodynamic model learning, producing the highest losses on all dimensions. inertia contains much more information than vision and produces lower losses. Combining vision and inertia keeps reducing the losses. On the other hand, adding speed to vision significantly boosts performance, while inertia combined with speed only shows marginal improvement, possibly due to the fact that inertia already contains a large amount of information from speed. Using all three modalities results in the lowest combined loss overall for END-TO-END. One important observation is that inertia significantly outperforms vision & speed in most dimensions, except for sliding, in the END-TO-END setup. However, TRON is able to reduce such a performance gap by allowing them to augment each other during representation learning and maximize the information contained in their representation spaces, as they achieve comparably low losses. Combining all three after TRON learning expectedly achieves the best results overall. STERLING is designed to use inertia to augment vision so that information contained in vision can be maximized when only vision is available during deployment. While STERLING works well on downstream terrain preference learning (e.g., grass is better than pebble) [18], not considering speed during representation learning introduces ambiguity in the representation spaces and therefore leads to bad performance on our kinodynamics learning task (e.g., what will happen when driving quickly/slowly on grass/pebble). For a fair comparison, we also show the results of STERLING with both vision and inertia (which is not available for planning on future states). Despite improved performance compared to using vision alone, due to the missing speed information during representation learning, STERLING still suffers from worse performance compared to TRON.

B. Human-CAHSOR Shared Autonomy

We conduct the human-CAHSOR shared autonomy experiments in three separate locations with different terrain types to provide a detailed view of how CAHSOR affects vehicle states in $\mathbb{SE}(3)$. The fully manual teleoperation mode gives the human the full range of vehicle control commands, i.e., -4.8m/s to 4.8ms throttle and -0.5rad to 0.5rad steering. CAHSOR selects actions based on Eqn. (5) with the objective function in Eqn. (6). The human is asked to drive the robot through a rocky patch and maneuver aggressively on grass and pavement, all at the fastest speed.

We show the human commanded throttle and steering and the CAHSOR constrained commands in Fig. 4 top, with resulted bumpiness, sliding, and roll (Fig. 4 bottom green). For comparison, we drive the robot without CAHSOR and overlay the resulted bumpiness, sliding, and roll (Fig. 4 bottom red) on the figure. CAHSOR knows to reduce speed on grass to avoid sliding and on pavement to prevent roll during high-speed turns, and to slow down on rocks to reduce bumpiness even without turning. However, due to the higher requirement of instantaneous accuracy of roll prediction compared to bumpiness and sliding, we observe that during on-robot experiments, CAHSOR is not able to prevent roll in advance all the time.

C. Autonomous CAHSOR Navigation

The autonomous CAHSOR navigation using MPPI is compared against two baselines: vanilla MPPI without CAHSOR with the same maximum speed of 4.8m/s and a slow MPPI with a maximum speed of 3.0m/s, which aims to show that conventional approaches need to slow down much more than CAHSOR in order to achieve similar vehicle stability in $\mathbb{SE}(3)$.

Table II shows the top speed, average speed, average bumpiness, maximum sliding, and maximum roll achieved by the three approaches. Autonomous navigation has the highest top speed of 4.8m/s and achieves an average of 4.29m/s around all three loops, only decelerating to reach each pre-defined GPS waypoint. However, it experiences significant bumpiness, sliding, and roll motions along the way. The high bumpiness may damage onboard components. In fact, we have to stop the experiments twice in order to fix a loose USB connection due to extensive vibration on the rocks (Fig. 5 top left and Fig. 1 lower left). Large roll angles also appear many times when executing the sharp turn from cement pavement
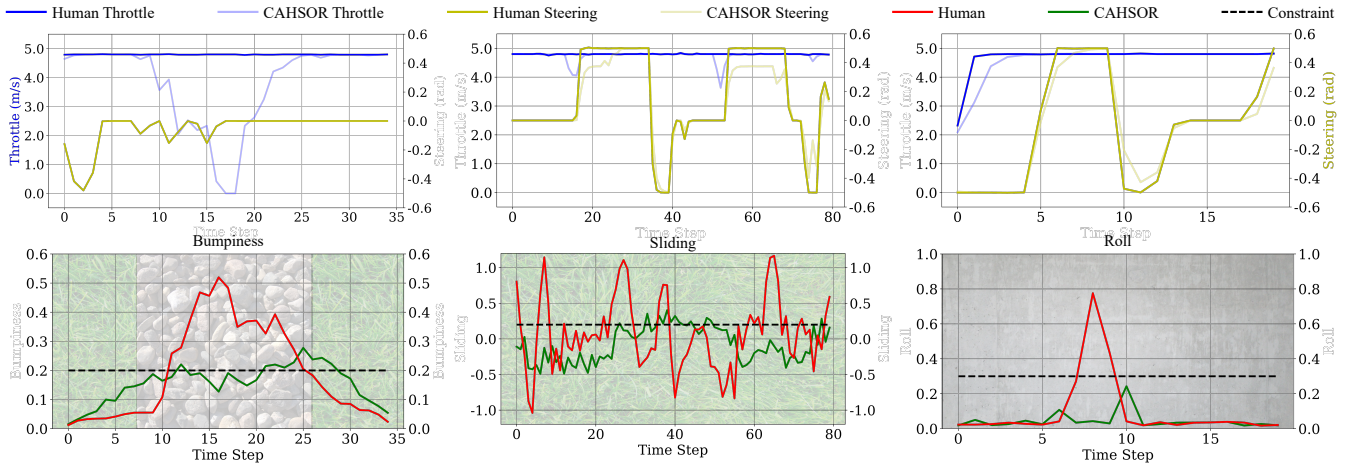
Fig. 4.   Example of Human-CAHSOR Shared Autonomy on Rocks, Grass, and Pavement to Limit Bumpiness, Sliding, and Roll.

	Top Speed (m/s)	Average Speed (m/s)	Average `bumpiness`	Maximum `sliding`	Maximum `roll`
4.8m/s Autonomous	**4.8**	**4.29±0.1**	0.11±0.007	1.11	1.2
4.8m/s Autonomous+CAHSOR	4.6	3.92±0.12	**0.054±0.007**	**0.62**	**0.11**
3.0m/s Slow Autonomous	2.75	2.45±0.02	0.065±0.01	0.69	0.12

TABLE II
SPEED AND $\mathbb{SE}(3)$ COMPETENCE AWARENESS OF AUTONOMOUS, AUTONOMOUS+CAHSOR, AND SLOW AUTONOMOUS NAVIGATION.
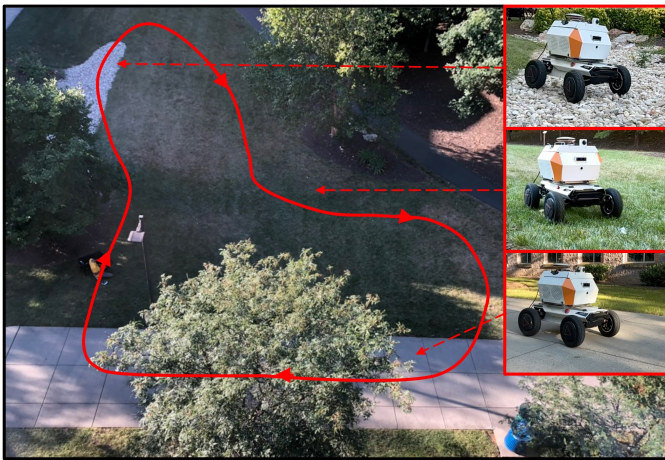

Fig. 5.   Autonomous Navigation in a Loop (Three Loops each Method).

to grass (Fig. 5 lower left and Fig. 1 top). When turning around obstacles, excessive sliding that is not considered by normal $\mathbb{SE}(2)$ models may risk collision (Fig. 5 middle and Fig. 1 lower right). On the other hand, autonomous navigation assisted by CAHSOR achieves a 4.6m/s top speed and slows down more, not only in order to reach GPS waypoints, but also due to competence awareness. We observe the vehicle significantly slows down when about to enter the rock patch to reduce bumpiness and executes less sharp or slower turns on grass and on cement to avoid too much sliding and roll. The slow autonomous navigation approach achieves a top speed of 2.75m/s and an average speed of 2.45m/s.

This slow autonomous navigation system clocks much longer lap time compared to its full speed counterpart without or with CAHSOR. Even with a low speed, the slow system still cannot achieve better average bumpiness, maximum sliding, and maximum roll as Autonomous+CAHSOR does, because it does not know when it is necessary to slow down to reduce such undesired movement and when it is possible to accelerate to achieve high speeds. The GPS waypoint loop experiments show that CAHSOR is able to maintain a very high average speed and only slows down when the $\mathbb{SE}(3)$ constraints would be violated.

VI. CONCLUSIONS

Our CAHSOR ground navigation approach is able to utilize multimodal, self-supervised terrain representation, i.e., TRON, to reason about the consequences of taking aggressive maneuvers on different off-road terrain, i.e., being competence-aware. Inertial observations contain the most information to enable efficient kinodynamics learning, but may not be available during planning. Augmenting easily available vision combined with speed using inertia with TRON, similar kinodynamics learning performance can be achieved. Extensive physical experiments in both an autonomous navigation planning and human shared-control setup demonstrate CAHSOR's superior competence awareness during high-speed off-road navigation.

REFERENCES

[1] Pranav Atreya, Haresh Karnan, Kavan Singh Sikand, Xuesu Xiao, Sadegh Rabiee, and Joydeep 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)*, pages 11789–11795. IEEE, 2022.

[2] Chengchao Bai, Jifeng Guo, and Hongxing Zheng. Three-dimensional vibration-based terrain classification for mobile robots. *IEEE Access*, 7:63485–63492, 2019.

[3] Xiaoyi Cai, Michael Everett, Jonathan Fink, and Jonathan 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)*, pages 2931–2937. IEEE, 2022.

[4] Xiaoyi Cai, Michael Everett, Lakshay Sharma, Philip R Osteen, and Jonathan P How. Probabilistic traversability model for risk-aware motion planning in off-road environments. In *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 11297–11304. IEEE, 2023.

[5] Yu Fan Chen, Miao Liu, Michael Everett, and Jonathan P How. Decentralized non-communicating multiagent collision avoidance with deep reinforcement learning. In *2017 IEEE international conference on robotics and automation (ICRA)*, pages 285–292. IEEE, 2017.

[6] Nitish Dashora, Daniel Shin, Dhruv Shah, Henry Leopold, David Fan, Ali Agha-Mohammadi, Nicholas Rhinehart, and Sergey Levine. Hybrid imitative planning with geometric and predictive costs in off-road environments. In *2022 International Conference on Robotics and Automation (ICRA)*, pages 4452–4458. IEEE, 2022.

[7] Aniket Datar, Chenhui Pan, Mohammad Nazeri, and Xuesu Xiao. Toward wheeled mobility on vertically challenging terrain: Platforms, datasets, and algorithms. *arXiv preprint arXiv:2303.00998*, 2023.

[8] Aniket Datar, Chenhui Pan, and Xuesu Xiao. Learning to model and plan for wheeled mobility on vertically challenging terrain. *arXiv preprint arXiv:2306.11611*, 2023.

[9] Anushri Dixit, David D Fan, Kyohei Otsu, Sharmita Dey, Ali-Akbar Agha-Mohammadi, and Joel W Burdick. Step: Stochastic traversability evaluation and planning for risk-aware off-road navigation; results from the darpa subterranean challenge. *arXiv preprint arXiv:2303.01614*, 2023.

[10] Dieter Fox, Wolfram Burgard, and Sebastian Thrun. The dynamic window approach to collision avoidance. *IEEE Robotics & Automation Magazine*, 4(1):23–33, 1997.

[11] Anthony Francis, Aleksandra Faust, Hao-Tien Lewis Chiang, Jasmine Hsu, J Chase Kew, Marek Fiser, and Tsang-Wei Edward Lee. Long-range indoor navigation with prm-rl. *IEEE Transactions on Robotics*, 36(4):1115–1134, 2020.

[12] Anthony Francis, Claudia Pérez-d'Arpino, Chengshu Li, Fei Xia, Alexandre Alahi, Rachid Alami, Aniket Bera, Abhijat Biswas, Joydeep Biswas, Rohan Chandra, et al. Principles and guidelines for evaluating social robot navigation algorithms. *arXiv preprint arXiv:2306.16740*, 2023.

[13] Justin Hart, Reuth Mirsky, Xuesu Xiao, Stone Tejeda, Bonny Mahajan, Jamin Goo, Kathryn Baldauf, Sydney Owen, and Peter Stone. Using human-inspired signals to disambiguate navigational intentions. In *International Conference on Social Robotics*, pages 320–331. Springer, 2020.

[14] Larry D Jackel, Eric Krotkov, Michael Perschbacher, Jim Pippine, and Chad Sullivan. The darpa lagr program: Goals, challenges, methodology, and phase i results. *Journal of Field robotics*, 23(11-12):945–973, 2006.

[15] Haresh Karnan, Anirudh Nair, Xuesu Xiao, Garrett Warnell, Sören Pirk, Alexander Toshev, Justin Hart, Joydeep Biswas, and Peter Stone. Socially compliant navigation dataset (scand): A large-scale dataset of demonstrations for social navigation. *IEEE Robotics and Automation Letters*, 7(4):11807–11814, 2022.

[16] Haresh Karnan, Kavan Singh Sikand, Pranav Atreya, Sadegh Rabiee, Xuesu Xiao, Garrett Warnell, Peter Stone, and Joydeep 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)*, pages 3294–3301. IEEE, 2022.

[17] Haresh Karnan, Garrett Warnell, Xuesu Xiao, and Peter Stone. Voila: Visual-observation-only imitation learning for autonomous navigation. In *2022 International Conference on Robotics and Automation (ICRA)*, pages 2497–2503. IEEE, 2022.

[18] Haresh Karnan, Elvin Yang, Daniel Farkash, Garrett Warnell, Joydeep Biswas, and Peter Stone. Sterling: Self-supervised terrain representation learning from unconstrained robot experience. In *Conference on Robot Learning*, pages 2393–2413. PMLR, 2023.

[19] Beomjoon Kim and Joelle Pineau. Socially adaptive path planning in human environments using inverse reinforcement learning. *International Journal of Social Robotics*, 8:51–66, 2016.

[20] Henrik Kretzschmar, Markus Spies, Christoph Sprunk, and Wolfram Burgard. Socially compliant mobile robot navigation via inverse reinforcement learning. *The International Journal of Robotics Research*, 35(11):1289–1307, 2016.

[21] Bo Liu, Xuesu Xiao, and Peter Stone. A lifelong learning approach to mobile robot navigation. *IEEE Robotics and Automation Letters*, 6(2):1090–1096, 2021.

[22] Parv Maheshwari, Wenshan Wang, Samuel Triest, Matthew Sivaprakasam, Shubhra Aich, John G Rogers III, Jason M Gregory, and Sebastian Scherer. Piaug–physics informed augmentation for learning vehicle dynamics for off-road navigation. *arXiv preprint*

arXiv:2311.00815, 2023.

[23] Daniel Maturana, Po-Wei Chou, Masashi Uenoyama, and Sebastian Scherer. Real-time semantic mapping for autonomous off-road navigation. In *Field and Service Robotics: Results of the 11th International Conference*, pages 335–350. Springer, 2018.

[24] Xiangyun Meng, Nathan Hatch, Alexander Lambert, Anqi Li, Nolan Wagener, Matthew Schmittle, JoonHo Lee, Wentao Yuan, Zoey Chen, Samuel Deng, et al. Terrainnet: Visual modeling of complex terrain for high-speed, off-road navigation. *arXiv preprint arXiv:2303.15771*, 2023.

[25] Reuth Mirsky, Xuesu Xiao, Justin Hart, and Peter Stone. Conflict avoidance in social navigation–a survey. *arXiv preprint arXiv:2106.12113*, 2021.

[26] Akhil Nagariya and Srikanth Saripalli. An iterative lqr controller for off-road and on-road vehicles using a neural network dynamics model. In *2020 IEEE Intelligent Vehicles Symposium (IV)*, pages 1740–1745. IEEE, 2020.

[27] Anirudh Nair, Fulin Jiang, Kang Hou, Zifan Xu, Shuozhe Li, Xuesu Xiao, and Peter Stone. Dynabarn: Benchmarking metric ground navigation in dynamic environments. In *2022 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*, pages 347–352. IEEE, 2022.

[28] Duc M Nguyen, Mohammad Nazeri, Amirreza Payandeh, Aniket Datar, and Xuesu 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.

[29] Yunpeng Pan, Ching-An Cheng, Kamil Saigol, Keuntaek Lee, Xinyan Yan, Evangelos A Theodorou, and Byron Boots. Imitation learning for agile autonomous driving. *The International Journal of Robotics Research*, 39(2-3):286–302, 2020.

[30] Jin-Soo Park, Xuesu Xiao, Garrett Warnell, Harel Yedidsion, and Peter Stone. Learning perceptual hallucination for multi-robot navigation in narrow hallways. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 10033–10039. IEEE, 2023.

[31] Daniel Perille, Abigail Truong, Xuesu Xiao, and Peter Stone. Benchmarking metric ground navigation. In *Proceedings of IEEE International Symposium on Safety, Security, and Rescue Robotics*, pages 116–121. IEEE, 2020.

[32] Mark Pfeiffer, Michael Schaeuble, Juan Nieto, Roland Siegwart, and Cesar Cadena. From perception to decision: A data-driven approach to end-to-end motion planning for autonomous ground robots. In *2017 ieee international conference on robotics and automation (icra)*, pages 1527–1533. IEEE, 2017.

[33] Sean Quinlan and Oussama Khatib. Elastic bands: Connecting path planning and control. In *[1993] Proceedings IEEE International Conference on Robotics and Automation*, pages 802–807. IEEE, 1993.

[34] Sadegh Rabiee and Joydeep Biswas. A friction-based kinematic model for skid-steer wheeled mobile robots. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 8563–8569. IEEE, 2019.

[35] Forrest Rogers-Marcovitz, Michael George, Neal Seegmiller, and Alonzo Kelly. Aiding off-road inertial navigation with high performance models of wheel slip. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 215–222. IEEE, 2012.

[36] Fabio Ruetz, Nicholas Lawrancel, Emili Hernández, Paulo Borges, and Thierry Peynot. Foresttrav: Accurate, efficient and deployable forest traversability estimation for autonomous ground vehicles. *arXiv preprint arXiv:2305.12705*, 2023.

[37] Guna Seetharaman, Arun Lakhotia, and Erik Philip Blasch. Unmanned vehicles come of age: The darpa grand challenge. *Computer*, 39(12):26–29, 2006.

[38] Kavan Singh Sikand, Sadegh Rabiee, Adam Uccello, Xuesu Xiao, Garrett Warnell, and Joydeep Biswas. Visual representation learning for preference-aware path planning. In *2022 International Conference on Robotics and Automation (ICRA)*, pages 11303–11309. IEEE, 2022.

[39] Sriram Siva, Maggie Wigness, John Rogers, and Hao Zhang. Robot adaptation to unstructured terrains by joint representation and apprenticeship learning. In *Robotics: science and systems*, 2019.

[40] Samuel Triest, Matthew Sivaprakasam, Sean J Wang, Wenshan Wang, Aaron M Johnson, and Sebastian Scherer. Tartandrive: A large-scale dataset for learning off-road dynamics models. In *2022 International Conference on Robotics and Automation (ICRA)*, pages 2546–2552. IEEE, 2022.

[41] Kasi Viswanath, Kartikeya Singh, Peng Jiang, PB Sujit, and Srikanth Saripalli. Offseg: A semantic segmentation framework for off-road driving. In *2021 IEEE 17th International Conference on Automation Science and Engineering (CASE)*, pages 354–359. IEEE, 2021.

[42] Zizhao Wang, Xuesu Xiao, Bo Liu, Garrett Warnell, and Peter Stone. Appli: Adaptive planner parameter learning from interventions. In *2021 IEEE international conference on robotics and automation (ICRA)*, pages 6079–6085. IEEE, 2021.

[43] Zizhao Wang, Xuesu Xiao, Alexander J Nettekoven, Kadhiravan Umasankar, Anika Singh, Sriram Bommakanti, Ufuk Topcu, and Peter Stone. From agile ground to aerial navigation: Learning from learned hallucination. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 148–153. IEEE, 2021.

[44] Zizhao Wang, Xuesu Xiao, Garrett Warnell, and Peter Stone. Apple: Adaptive planner parameter learning from evaluative feedback. *IEEE Robotics and Automation Letters*, 6(4):7744–7749, 2021.

[45] Grady Williams, Andrew Aldrich, and Evangelos A Theodorou. Model predictive path integral control: From theory to parallel computation. *Journal of Guidance,*

Control, and Dynamics, 40(2):344–357, 2017.

[46] Patrick Wolf, Axel Vierling, Thorsten Ropertz, and Karsten Berns. Advanced scene aware navigation for the heavy duty off-road vehicle unimog. In *IOP Conference Series: Materials Science and Engineering*, volume 997-1, page 012093. IOP Publishing, 2020.

[47] Xuesu Xiao, Bo Liu, Garrett Warnell, Jonathan Fink, and Peter Stone. Appld: Adaptive planner parameter learning from demonstration. *IEEE Robotics and Automation Letters*, 5(3):4541–4547, 2020.

[48] Xuesu Xiao, Joydeep Biswas, and Peter Stone. Learning inverse kinodynamics for accurate high-speed off-road navigation on unstructured terrain. *IEEE Robotics and Automation Letters*, 6(3):6054–6060, 2021.

[49] Xuesu Xiao, Bo Liu, and Peter Stone. Agile robot navigation through hallucinated learning and sober deployment. In *2021 IEEE international conference on robotics and automation (ICRA)*, pages 7316–7322. IEEE, 2021.

[50] Xuesu Xiao, Bo Liu, Garrett Warnell, and Peter Stone. Toward agile maneuvers in highly constrained spaces: Learning from hallucination. *IEEE Robotics and Automation Letters*, 6(2):1503–1510, 2021.

[51] Xuesu Xiao, Bo Liu, Garrett Warnell, and Peter Stone. Motion planning and control for mobile robot navigation using machine learning: a survey. *Autonomous Robots*, 46(5):569–597, 2022.

[52] Xuesu Xiao, Zizhao Wang, Zifan Xu, Bo Liu, Garrett Warnell, Gauraang Dhamankar, Anirudh Nair, and Peter Stone. Appl: Adaptive planner parameter learning. *Robotics and Autonomous Systems*, 154:104132, 2022.

[53] Xuesu Xiao, Zifan Xu, Zizhao Wang, Yunlong Song, Garrett Warnell, Peter Stone, Tingnan Zhang, Shravan Ravi, Gary Wang, Haresh Karnan, Joydeep Biswas, Nicholas Mohammad, Lauren Bramblett, Rahul Peddi, Nicola Bezzo, Zhanteng Xie, and Philip Dames. Autonomous ground navigation in highly constrained spaces: Lessons learned from the benchmark autonomous robot navigation challenge at icra 2022 [competitions]. *IEEE Robotics & Automation Magazine*, 29(4):148–156, 2022.

[54] Xuesu Xiao, Tingnan Zhang, Krzysztof Marcin Choromanski, Tsang-Wei Edward Lee, Anthony Francis, Jake Varley, Stephen Tu, Sumeet Singh, Peng Xu, Fei Xia, Sven Mikael Persson, Leila Takayama, Roy Frostig, Jie Tan, Carolina Parada, and Vikas Sindhwani. Learning model predictive controllers with real-time attention for real-world navigation. In *Conference on robot learning*. PMLR, 2022.

[55] Xuesu Xiao, Zifan Xu, Garrett Warnell, Peter Stone, Ferran Gebelli Guinjoan, Romulo T Rodrigues, Herman Bruyninckx, Hanjaya Mandala, Guilherme Christmann, Jose Luis Blanco-Claraco, et al. Autonomous ground navigation in highly constrained spaces: Lessons learned from the 2nd barn challenge at icra 2023. *arXiv preprint arXiv:2308.03205*, 2023.

[56] Zifan Xu, Gauraang Dhamankar, Anirudh Nair, Xuesu Xiao, Garrett Warnell, Bo Liu, Zizhao Wang, and Peter Stone. Applr: Adaptive planner parameter learning from reinforcement. In *2021 IEEE international conference on robotics and automation (ICRA)*, pages 6086–6092. IEEE, 2021.

[57] Zifan Xu, Xuesu Xiao, Garrett Warnell, Anirudh Nair, and Peter 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)*, pages 217–222. IEEE, 2021.

[58] Zifan Xu, Bo Liu, Xuesu Xiao, Anirudh Nair, and Peter Stone. Benchmarking reinforcement learning techniques for autonomous navigation. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 9224–9230. IEEE, 2023.

[59] Richard N Youngworth, Benjamin B Gallagher, and Brian L Stamper. An overview of power spectral density (psd) calculations. *Optical manufacturing and testing VI*, 5869:206–216, 2005.

[60] Jure Zbontar, Li Jing, Ishan Misra, Yann LeCun, and Stéphane Deny. Barlow twins: Self-supervised learning via redundancy reduction. In *International Conference on Machine Learning*, pages 12310–12320. PMLR, 2021.