2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 14-18, 2024. Abu Dhabi, UAE

Terrain-Attentive Learning for Efficient 6-DoF Kinodynamic Modeling on Vertically Challenging Terrain

Aniket Datar, Chenhui Pan, Mohammad Nazeri, Anuj Pokhrel, and Xuesu Xiao

***Abstract*— Wheeled robots have recently demonstrated superior mechanical capability to traverse vertically challenging terrain (e.g., extremely rugged boulders comparable in size to the vehicles themselves). Negotiating such terrain introduces significant variations of vehicle pose in all six Degrees-of-Freedom (DoFs), leading to imbalanced contact forces, varying momentum, and chassis deformation due to non-rigid tires and suspensions. To autonomously navigate on vertically challenging terrain, all these factors need to be efficiently reasoned within limited onboard computation and strict real-time constraints. In this paper, we propose a 6-DoF kinodynamics learning approach that is attentive only to the specific underlying terrain critical to the current vehicle-terrain interaction, so that it can be efficiently queried in real-time motion planners onboard small robots. Physical experiment results show our Terrain-Attentive Learning (TAL) demonstrates on average 51.1% reduction in model prediction error among all 6 DoFs compared to a state-of-the-art model for vertically challenging terrain.**[1]

I. Introduction

Despite their wide availability, wheeled mobile robots are usually limited in terms of mobility, mostly moving in 2D flat environments. After dividing their planar workspaces into free spaces and obstacles, those robots are assumed to be rigid bodies and efficiently find collision-free paths to move from one point to another, using extremely simplified kinodynamic models. When facing *vertically challenging* terrain, e.g., spaces filled with large obstacles like boulders or tree trunks where a collision-free 2D path does not exist, roboticists have mostly sought help from more sophisticated mechanical design, such as legged, leg-wheeled, and articulated tracked vehicles or adding active suspension systems.

Recent advances in wheeled mobility have shown that even conventional wheeled vehicles without sophisticated hardware modification have unrealized mobility potential on vertically challenging terrain [1]. With a set of minimal hardware requirements, e.g., all-wheel drive, independent suspensions, and differential lock, those simple vehicles can also, at least with human teleoperation, venture into environments that would normally be deemed as non-traversable obstacles by state-of-the-art autonomous navigation systems.

To achieve such unrealized mobility autonomously, wheeled robots need to reason about the complex vehicle-terrain interaction, including imbalanced contact forces, varying momentum, and chassis deformation due to non-rigid tires and suspensions. All these factors are tightly dependent on the underlying terrain. In state-of-the-art motion planners, e.g., sampling-based or optimization-based, such vehicle-terrain interaction needs to be modeled and computed for a large number of future terrain patches beneath candidate vehicle poses. For highly articulated systems, efficient decomposition is possible to break down the modeling of the vehicle chassis and actuators (e.g., legs and active suspensions) so that the chassis trajectory can be computed separately in parallel and the low-level actuation solved using fast control and optimization techniques. Unfortunately, for under-actuated convectional wheeled robots, the whole system is fully coupled and such decomposition is not possible, requiring sequential, un-parallelizable computation along potential future robot trajectories.

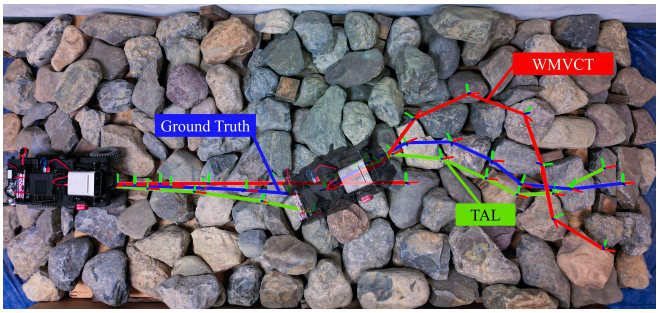

Fig. 1: Two Sets of 6-DoF Kinodynamic Trajectory Predictions by TAL and WMVCT [2] Compared to Ground Truth.

To this end, we present Terrain-Attentive Learning (TAL), a 6-DoF kinodynamics learning approach that is attentive (only) to the specific underlying terrain critical to the current vehicle-terrain interaction, so that it can be efficiently queried in real-time motion planners onboard small robots. TAL is combined with a state-of-the-art sampling-based motion planner and allows to sequentially rollout future trajectories in an efficient manner for downstream cost-based kinodynamic planning. Using TAL, we demonstrate on average 51.1% reduction in model prediction error among all 6 DoFs compared to another state-of-the-art kinodynamics modeling approach for vertically challenging terrain [2] (Fig. 1).

II. Related Work

This section discusses related work in terms of wheeled robot kinodynamic modeling, off-road navigation, and learning-based mobility.

All authors are with the Department of Computer Science, George Mason University {adatar, cpan7, mnazerir, apokhre, xiao}@gmu.edu. This work has taken place in the RobotiXX Laboratory at George Mason University. RobotiXX research is supported by National Science Foundation (NSF, 2350352), Army Research Office (ARO, W911NF2220242, W911NF2320004, W911NF2420027), US Air Forces Central (AFCENT), Google DeepMind (GDM), Clearpath Robotics, and Raytheon Technologies (RTX).

[1]Full version: https://arxiv.org/abs/2403.16419

A. Wheeled Robot Kinodynamic Modeling

Most existing wheeled robot kinodynamic models, despite their differences in fidelities, still assume the robot moves in a 2D space and its motion is constrained in $\mathbb{SE}(2)$. However, when facing off-road environments, especially vertically challenging terrain, such an assumption no longer holds and the workspace has to be extended to $\mathbb{SE}(3)$ [1], [2]. Modeling in $\mathbb{SE}(3)$ faces challenges in terms of both accuracy and efficiency: the significant variations of 6-DoF vehicle pose caused by the variety of underlying terrain needs to be precisely modeled, while such a model also needs to be queried efficiently in real-time motion planners. Our TAL approach aims at tackling both challenges simultaneously in a data-driven manner using representation learning.

B. Off-Road Navigation

A large percentage of off-road navigation research has focused on the perception side since the DARPA Urban Challenge [3] and LAGR Program [4]. Extending from the simple differentiation of obstacles and free spaces, off-road perception systems need to consider semantic information [5]–[8], such as gravel, grass, bushes, pebbles, and rocks, and then devise cost functions based on the semantic understanding for subsequent path and motion planning.

Recent research efforts have gradually moved towards the mobility side. Inverse [9], [10] and forward [11], [12] kinodynamic models have been created from real-world vehicle-terrain interactions to enable high-speed off-road navigation. End-to-end learned mobility [13] has eliminated the boundary between perception and mobility systems so the whole navigation system can be learned in a data-driven manner. Most existing off-road navigation work still assume the vehicles are moving in a 2D plane, while deliberately choosing which part of the 2D plane to drive on or modeling how different terrain would affect the 3-DoF vehicle motion.

When facing vertical protrusions from the ground, e.g., large boulders or fallen tree trunks, most existing off-road navigation systems still treat them as non-traversable obstacles, e.g., with a large cost assigned to the corresponding semantic class. In this work, we aim to allow vehicles to efficiently reason about the consequences of interacting with such vertically challenging terrain and autonomously plan feasible motions to traverse through.

C. Learning-Based Mobility

Recent advancement in machine learning has been utilized for robot mobility using imitation [14] or reinforcement learning [15]. Learning enhances robot adaptivity and agility, increases movement speed [9]–[11], [13], [16], enables visual-only navigation [17], [18], and creates socially compliant mobile robots [19].

While having the potential to learn from data, learning-based mobility also faces challenges from being data-hungry and computation-intensive, especially onboard a mobile robot. TAL aims at alleviating the need of large-scale real-world datasets from constraining the learning process only to a forward kinodynamic model, which will be combined in a Model Predictive Control (MPC) [20] setup. TAL also utilizes representation learning [16] so that the learned kinodynamic model can efficiently attend to the specific underlying terrain critical to the current vehicle-terrain interaction, without extensive computation required to pre-process input data.

III. Approach

We first formulate the problem of forward kinodynamic modeling for wheeled mobility on vertically challenging terrain. We then present how this problem is approached in a data-driven manner to avoid the need of analytical vehicle-terrain interaction models. Finally, we introduce our TAL method which allows the learned 6-DoF kinodynamic model to efficiently attend to the specific underlying terrain so that it can quickly predict the next vehicle state in a MPC setup for downstream kinodynamic planning.

A. Problem Formulation

While most traditional 2D navigation problems are defined in a 2D state space, i.e., $X \subset \mathbb{SE}(2)$, our vertically challenging terrain requires the state space to be extended to $X \subset \mathbb{SE}(3)$. Traditional motion planners only move robots in free space and avoid obstacles, as divisions of the whole state space: $X = X_{\text{free}} \cup X_{\text{obs}}$. In contrast, our wheeled robot needs to decide which obstacles should be avoided (as making contact with them will cause immobilization or damage, e.g., hitting a wall), while which ones it can drive on top of (use them as support underneath the chassis), considering a collision-free 2D path may not always exist in vertically challenging environments.

We adopt a discrete vehicle forward kinodynamic model:

$$\mathbf{x}_{t+1} = f_\theta(\mathbf{x}_t, \mathbf{u}_t, \mathbf{m}_t), \tag{1}$$

where $\mathbf{x}_t \in X$, $\mathbf{u}_t \in U$, and $\mathbf{m}_t \in M$ denote the vehicle state, control input, and environment state respectively. $\mathbf{x}_t$ includes the translations along the $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$ axis (x, y, and z) and the rotations around them (roll, pitch, and yaw) in a coordinate system, as well as their velocity components when necessary. For control input, $\mathbf{u}_t = (v_t, \omega_t) \in U \subset \mathbb{R}^2$, where v_t and ω_t are the linear and angular velocity or throttle and steering command. The environment state $\mathbf{m}_t$ includes all necessary information in the environment to determine the next vehicle state $\mathbf{x}_{t+1}$, given $\mathbf{x}_t$ and $\mathbf{u}_t$. Such information can include terrain geometry and semantics. Considering the fact that ground vehicle dynamics depends primarily on the terrain topology and the high computational overhead of using a full 3D map, we use a 2.5D terrain elevation map to construct $\mathbf{m}_t$ underneath the current vehicle state $\mathbf{x}_t$ to represent terrain topology and leave semantics (e.g., slipperiness, deformability, and elasticity) as future work. The motion planning problem is to find a control function $u : \{t\}_{t=0}^{T-1} \to U$ that produces an optimal path $\mathbf{x}_t \in X_{\text{free}}, \forall t \in \{t\}_{t=0}^{T}$ from an initial state $\mathbf{x}_0 = \mathbf{x}_{\text{init}}$ to a goal region $X_{\text{goal}} \subset X$, i.e., $\mathbf{x}_T \in X_{\text{goal}}$. The path needs to observe the system dynamics $f_\theta(\cdot, \cdot, \cdot)$, parameterized by θ, and minimize a given cost function $c(x)$, which maps from a state trajectory $x : \{t\}_{t=0}^{T} \to X$ to a positive real number.

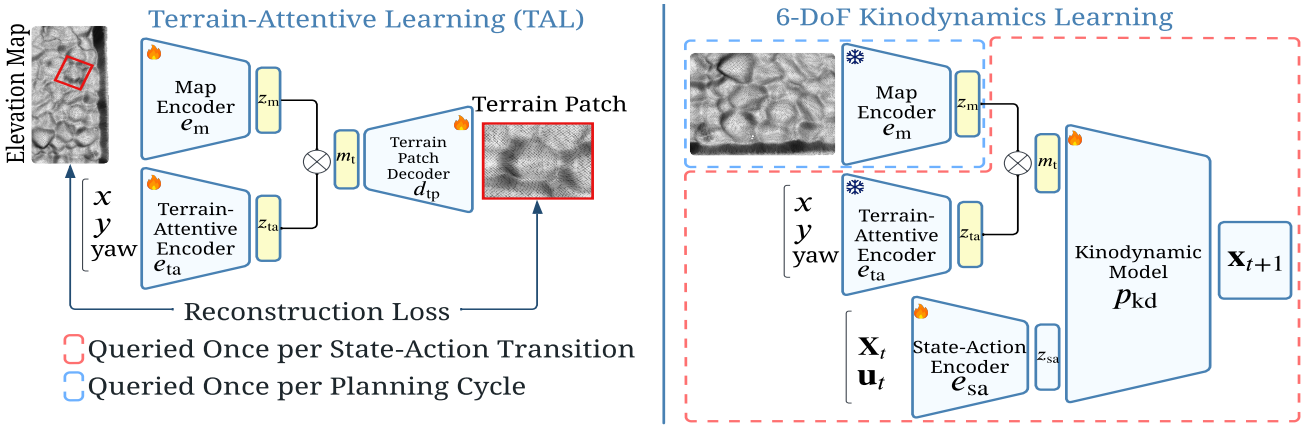

Fig. 2: Terrain-Attentive Learning (TAL, Left) and 6-DoF Kinodynamics Learning (Right) Architecture: Flame and temperature denote training and frozen parameters respectively.

B. Data-Driven Kinodynamics

Most existing 2D vehicle kinodynamic models only condition next state $\mathbf{x}_{t+1}$ on current state $\mathbf{x}_t$ and input $\mathbf{u}_t$ and are significantly simplified using, e.g., differential-drive, unicycle, bicycle, or Ackermann-steering mechanisms. However, the inclusion of $\mathbf{m}_t$ when moving in off-road environments, especially on vertically challenging terrain, substantially complicates the model. For example, driving the vehicle toward a wall or steep slope can get the vehicle stuck; speeding on uneven terrain may cause it to go airborne; and navigating slanted terrain can lead to rollovers. Modeling these vehicle-terrain interactions analytically is challenging.

To avoid the difficulty in analytically modeling f_θ, we adopt a data-driven approach. We assume a training dataset of size N is available: $\mathcal{D} = \{\langle \mathbf{x}_t, \mathbf{x}_{t+1}, \mathbf{m}_t, \mathbf{u}_t \rangle_{t=1}^{N}\}$. θ can then be learned by minimizing a supervised loss function:

$$\theta^* = \underset{\theta}{\operatorname{argmin}} \sum_{(\mathbf{x}_t, \mathbf{x}_{t+1}, \mathbf{m}_t, \mathbf{u}_t) \in \mathcal{D}} \| f_\theta(\mathbf{x}_t, \mathbf{u}_t, \mathbf{m}_t) - \mathbf{x}_{t+1} \|, \quad (2)$$

The learned vehicle-terrain forward kinodynamic model $f_\theta(\cdot, \cdot, \cdot)$, e.g., instantiated as a deep neural network, can be used to rollout future trajectories for minimal-cost planning.

C. Terrain-Attentive Learning (TAL)

In addition to the difficulty in precisely deriving analytical models for f_θ, another difficulty brought by the inclusion of $\mathbf{m}_t$ is the increased computation cost and reduced efficiency during model query. The state-action transitions of simplified 2D kinodynamic models, when depending only on the current state and input, can therefore be very quickly computed. They can even be pre-computed and saved in advance, e.g., as state lattices [21] or pre-processed maps [22]. Conversely, even given the same current state $\mathbf{x}_t$ and input $\mathbf{u}_t$, different $\mathbf{m}_t$ as input will produce a variety of next state $\mathbf{x}_{t+1}$, which will further affect the transition into $\mathbf{x}_{t+2}$, and so on. In a MPC setup, such a sequential dependence of the next state-action transition on the current one precludes the possibility of processing the sequence of $\{\mathbf{m}_t\}_{t=0}^{H-1}$ for one single trajectory (with H as the planning horizon) in parallel and therefore incurs extensive computation overhead during sequential rollouts, especially when a large amount of potential state-action transitions must be computed for iterative, sampling-based motion planners [20]. Furthermore, how to efficiently extract $\mathbf{m}_t$ from raw perception within limited onboard computation is also a challenging task.

Therefore, TAL utilizes self-supervised representation learning to efficiently process robot perception into $\mathbf{m}_t$ (Fig. 2 left) and query the learned model f_θ (Fig. 2 right) to rollout and evaluate future candidate trajectories. Within a MPC planning cycle, the kinodynamic model needs to quickly retrieve relevant environment state from the space of all possible environment states, i.e., $\mathbf{m}_t \in M$. In our wheeled mobility on vertically challenging terrain problem, M is the terrain information space of all possible terrain patches that can be extracted from an elevation map built by an online mapping system [23]. Given a full 2.5D elevation map E of the vertically challenging terrain in the gravity-aligned world frame, the terrain patch underneath the robot state $\mathbf{x}_t$ is independent of the robot's roll$_t$, pitch$_t$, and z_t, and can thus be extracted using only x_t, y_t, and yaw$_t$. Notice that such terrain extraction requires translation, cropping, and rotation operations of the original full elevation map and therefore incurs an extensive amount of computation when repeated many times in a sampling-based MPC setting. Furthermore, consuming the terrain patch as kinodynamic model input during every state-action transition is also extremely computationally extensive. To use representation learning to alleviate the computation overhead during deployment, we generate a terrain patch dataset using many full elevation maps $\{E_i\}_{i=1}^{I}$ and terrain patches extracted from each of them based on randomly sampled $\langle x, y, \text{yaw} \rangle$ tuples, denoted as $\{E_i, \{p_i^j, \langle x_i^j, y_i^j, \text{yaw}_i^j \rangle\}_{j=1}^{J}\}_{i=1}^{I}$. As shown in Fig. 2 left, a map encoder e_m and a terrain-attentive encoder e_{ta} embed the full elevation map E and $\langle x, y, \text{yaw} \rangle$ into their latent spaces, before being concatenated and decoded using a terrain patch decoder d_{tp}. The map and terrain-attentive encoders and the terrain patch decoder are trained in an end-

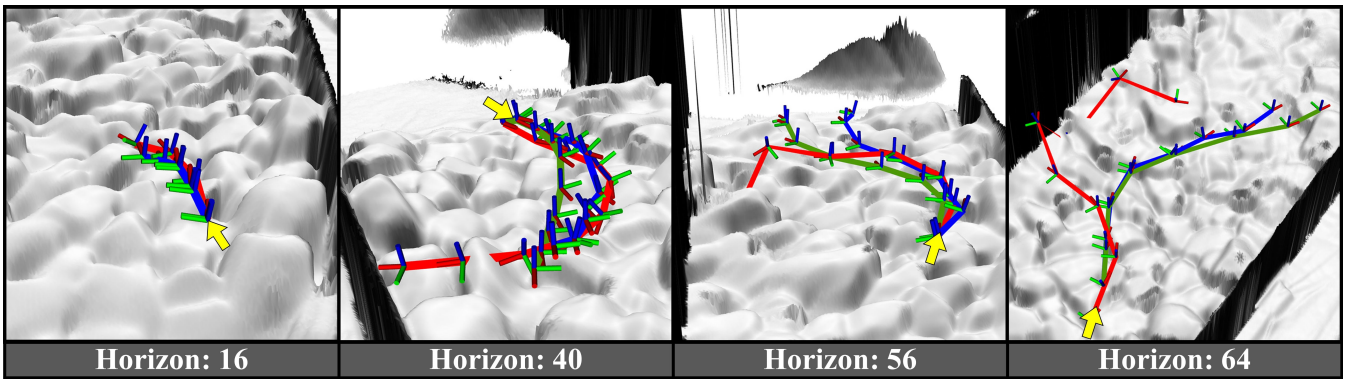

Fig. 3: 6-DoF Vehicle Trajectories of TAL, WMVCT, and Ground Truth with Increasing Horizon: TAL closely matches Ground Truth even with a long horizon, while WMVCT significantly diverges.

to-end fashion using self-supervised representation loss:

$$\mathcal{L}_{\text{TAL}} = \sum_{i=1}^{I}\sum_{j=1}^{J} \|p_i^j - d_{\text{tp}}(e_{\text{m}}(E_i), e_{\text{ta}}(\langle x_i^j, y_i^j, \text{yaw}_i^j\rangle))\|. \quad (3)$$

The latent embeddings of the full elevation map and $\langle x, y, \text{yaw}\rangle$ contain sufficient information to reconstruct the terrain patch, and therefore can be used as $\mathbf{m}_t$.

The parameters for the learned map and terrain-attentive encoders, e_{m} and e_{ta}, are then frozen during downstream 6-DoF kinodynamics learning (Fig. 2 right). The optimal kinodynamics parameters θ^*, in the form of a state-action encoder e_{sa} and kinodynamics predictor p_{kd}, are learned using the kinodynamics loss defined in Eqn. (2). During a single deployment planning cycle, the large map encoder will only need to be queried once and produce one elevation map embedding, while the small terrain-attentive encoder, state-action encoder, and kinodynamics predictor will be queried for every state-action transition. The learned kinodynamic model can then be efficiently queried for subsequent sampling-based MPC planning.

D. Implementations

1) Terrain Attentive Learning: TAL leverages a 3-layer Convolutional Neural Network (CNN) as the map encoder (e_{m}) that produces a latent embedding $\mathbf{z}_{\text{m}} \in \mathbb{R}^{160\times6\times6}$. In parallel, the terrain-attentive encoder (e_{ta}), a 2-layer Multi-Layer Perceptron (MLP), produces a latent embedding $\mathbf{z}_{\text{ta}} \in \mathbb{R}^{160\times6\times6}$, the same size as $\mathbf{z}_{\text{m}}$. The second embedding $\mathbf{z}_{\text{ta}}$ serves as attention weights, which are subsequently multiplied with $\mathbf{z}_{\text{m}}$ and passed through one linear layer producing a latent embedding $\mathbf{m}_t \in \mathbb{R}^{64\times6\times6}$ as the final terrain representation. The terrain patch decoder d_{tp} is a 4-layer Convolutional Transpose Network to reconstruct the patch corresponding to the robot footprint with a 0.24m^2 area in the real world. We use Mean Squared Error as the loss function to guide the reconstruction process.

2) 6-DoF Kinodynamics Learning: The kinodynamics learning consists of the pre-trained TAL model with the addition of the state-action encoder e_{sa} and the kinodynamics predictor p_{kd}. The state-action encoder e_{sa} incorporates two MLPs each with two layers to encode state ($\mathbf{x}_t$) into $\mathbf{z}_{\text{s}} \in \mathbb{R}^{16}$ and action ($\mathbf{u}_t$) into $\mathbf{z}_{\text{a}} \in \mathbb{R}^{16}$. Then we concatenate $\mathbf{z}_{\text{s}}$ and $\mathbf{z}_{\text{a}}$ into $\mathbf{z}_{\text{sa}}$. This $\mathbf{z}_{\text{sa}}$ is then further concatenated with the terrain representation, $\mathbf{m}_t$, obtained from the TAL model.

The concatenated vector, consisting of $\mathbf{z}_{\text{sa}}$ and $\mathbf{m}_t$, is subsequently fed into the kinodynamics predictor p_{kd}, a 2-layer MLP, to predict the next state $\mathbf{x}_{t+1}$. During this stage, the weights of e_{m} and e_{ta} are frozen, and only the weights of e_{sa} and p_{kd}, i.e., f_θ, are updated through training.

IV. Experiments

To verify that TAL model can produce accurate future vehicle state prediction based on the current state $\mathbf{x}_t$, current action $\mathbf{u}_t$, and underlying terrain $\mathbf{m}_t$, we compare the predictions from TAL against WMVCT planner [2]. We also deploy the TAL model in a sampling-based planner and show it can be used to generate feasible motion plans to navigate through vertically challenging terrain.

A. Robot, Testbed, and Data

We implement TAL on an open-source, 1/10th-scale, unmanned ground vehicle, the Verti-4-Wheeler (V4W) platform [1]. The robot is equipped with a low-high gear and lockable front and rear differentials enhancing its mobility on vertically challenging terrain. For simplicity, we only use low-gear with locked differentials in our dataset and experiments and leave the investigation of their effects on kinodynamics to future work. A Microsoft Azure Kinect RGB-D camera provides both visual and depth information, which we use to run Visual Inertia Odometry (VIO) [24] and generate real-time elevation maps [23]. The entire system operates on an onboard NVIDIA Jetson Orin NX computer.

We construct a $3.1\text{m}\times1.3\text{m}$ rock testbed with a maximum height of 0.6m (Fig. 1). For comparison, the V4W has a height of 0.2m, width of 0.249m, and length of 0.523m with a 0.312m wheel base. The numerous rocks on this rock testbed can be easily reconfigured to facilitate data collection and mobility experiments in various configurations. Apart from the rock testbed, we also create a foam testbed of $1\text{m}\times0.5\text{m}$ with a height of 0.4m to test the generalizability of the model in unseen environments. Note that in addition

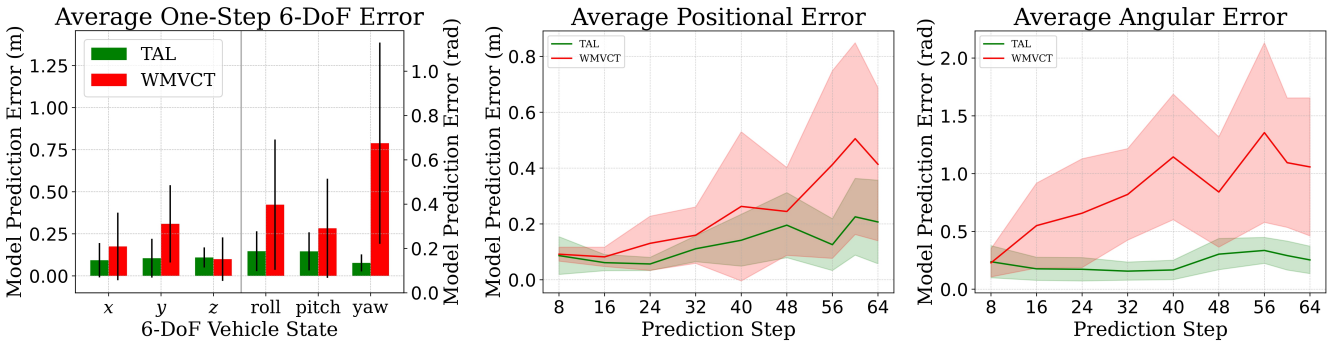

Fig. 4: Model Prediction Error of TAL and WMVCT: Average One-Step 6-DoF Positional and Angular Error (Left); Prediction Error vs. Prediction Step (Middle and Right). TAL achieves lower prediction error and variance than WMVCT in all cases.

to unseen topology, this testbed also has different friction dynamics compared to the rock testbed.

We collect 30 minutes of data on the rock testbed and 30 minutes of data on a planar surface. We use a 9:1 ratio to split train and test data and report all results on unseen test data. The dataset contains VIO for vehicle state estimation, elevation maps built from depth images, and teleoperated vehicle controls including throttle and steering commands. A variety of 6-DoF vehicle states are included in the rock testbed data, including vehicle rollover and getting stuck.

B. Trajectory Prediction Visualization

Fig. 3 visualizes predicted trajectory examples by TAL and WMVCT compared to the ground truth at different horizon steps. At horizon 16, all three trajectories are close to each other. At horizon 40, WMVCT fails to account for a large rock and diverges significantly from TAL, which closely matches the ground truth. By horizon 56, TAL continues to follow the ground truth direction, while WMVCT deviates left. At horizon 64, WMVCT accumulates significant error, causing the red trajectory to extend beyond the elevation map and penetrate rocks, whereas TAL remains aligned with the ground truth.

C. 6-DoF Prediction Accuracy

We compare the accuracy of the TAL model in predicting the next 6-DoF vehicle state with the WMVCT model [2]. For efficiency, the WMVCT model decomposes the 6 DoFs into three parts: x, y, and yaw are determined by a simple planar Ackermann-steering model; z is based on the elevation map value at (x, y); roll and pitch are computed using a neural network which takes as input a terrain patch located at (x, y) and aligned with yaw. Fig. 4 left shows the average error with standard deviation in predicting the 6-DoF vehicle state. Except the negligible difference in z-position of the robot, TAL outperforms WMVCT for all other DoFs by a wide margin. Averaged among all DoFs, TAL achieves 51.1% reduction in modeling error and 62.5% reduction in error standard deviation. Fig. 4 middle and right shows the 6-DoF prediction error of the models with respect to different prediction steps. As steps increase, WMVCT accumulates significant error and shows greater variance, indicating higher uncertainty. Conversely, TAL maintains more accurate position and angle predictions with lower variance.

TABLE I: Comparison of Success Rate and Average Time.

	OL	RB	BC	WMVCT	TAL
Success Rate	0/10	0/10	7/10	**10/10**	**10/10**
Average Time	-	-	12.28±2.69	16.76±1.44	**16.53±1.08**

D. On-Robot Deployment

We deploy TAL with the Model Predictive Path Integral (MPPI) planner [25] on V4W. The MPPI planner operates by rolling out 400 candidate trajectories at each time step, extending its planning horizon to 20 steps. It samples diverse control sequences using a normal distribution centered around the last executed control sequence. This set of candidate sequences, along with the elevation map, is fed into the TAL model. For each time step within each trajectory, TAL predicts the resulting 6-DoF state of the robot based on the initial or last predicted state. These resultant states are evaluated using a custom cost function that considers the Euclidean distance from the last state to the goal and penalizes high roll and pitch values, encouraging stable poses. The performance of MPPI with TAL is tested on unseen rock configurations after shuffling the rock testbed with increased difficulty, introducing "tricky corners" for the robot to avoid. We conduct 10 trials each for MPPI with TAL, the WMVCT planner [2], Behavior Cloning (BC) [26]–[28], and two baselines from the Verti-Wheelers project: Rule-Based (RB) and Open-Loop (OL) [1]. The WMVCT planner uses a fixed set of state trajectories, the same MPPI cost function for trajectory evaluation, and a PID controller for trajectory tracking. The goal is consistently set across the rock testbed for all trials.

Table I presents the success rate and average traversal time for all five methods. The "tricky corners" consistently cause OL and RB to fail, with the V4W getting stuck or rolling over in every trial. BC fails three times for similar reasons. WMVCT with the decomposed 6-DoF model performs similarly to MPPI with TAL, indicating that the significantly higher accuracy of TAL does not directly translate to much better navigation performance.

E. Discussions

In our experiments, the TAL model achieves significantly better prediction accuracy compared to the WMVCT model in all six DoFs and avoids extensive error during in long-horizon predictions. However, this superior model accuracy does not lead to a higher success rate in MPPI planner. We posit that the lack of direct correlation between model accuracy and navigation performance stems from the MPPI planner's computational demands. While the WMVCT planner quickly updates plans with an efficient but less accurate model, the MPPI planner requires more time to converge using the high-accuracy TAL model. This increased computational cost reduces planning frequency, limiting the robot's ability to react and avoid obstacles in time. Such an observation motivates future investigation into the tradeoff between high model fidelity and planning frequency.

V. Conclusions

This work introduces Terrain Attentive Learning (TAL) for 6-DoF kinodynamics learning, focusing on extracting important features that influence robot-terrain interaction. Specifically, we pre-train neural networks to use robot poses as attention weights. These attention weights guide the extraction of important underlying features from the elevation map, utilizing patch reconstruction as a form of self-supervision. With the pre-trained networks, TAL predicts the next vehicle state based on the current pose, control input, and elevation map. This approach enables efficient deployment in real-time planners for small robots with limited computational resources. We quantitatively and qualitatively show that TAL can accurately predict the next robot state, which helps to plan feasible, stable, and efficient paths through vertically challenging terrain in a sampling-based motion planner.

References

[1] A. Datar, C. Pan, M. Nazeri, and X. Xiao, "Toward wheeled mobility on vertically challenging terrain: Platforms, datasets, and algorithms," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 2024.

[2] A. Datar, C. Pan, and X. Xiao, "Learning to model and plan for wheeled mobility on vertically challenging terrain," *arXiv preprint arXiv:2306.11611*, 2023.

[3] G. Seetharaman, A. Lakhotia, and E. P. Blasch, "Unmanned vehicles come of age: The darpa grand challenge," *Computer*, 2006.

[4] L. D. Jackel, E. Krotkov, M. Perschbacher, J. Pippine, and C. Sullivan, "The darpa lagr program: Goals, challenges, methodology, and phase i results," *Journal of Field robotics*, vol. 23, no. 11-12, pp. 945–973, 2006.

[5] X. Meng, N. Hatch, A. Lambert, A. Li, N. Wagener, M. Schmittle, J. Lee, W. Yuan, Z. Chen, S. Deng *et al.*, "Terrainnet: Visual modeling of complex terrain for high-speed, off-road navigation," *arXiv preprint arXiv:2303.15771*, 2023.

[6] K. Viswanath, K. Singh, P. Jiang, P. Sujit, and S. Saripalli, "Offseg: A semantic segmentation framework for off-road driving," in *2021 IEEE 17th International Conference on Automation Science and Engineering (CASE)*. IEEE, 2021.

[7] D. Maturana, P.-W. Chou, M. Uenoyama, and S. Scherer, "Real-time semantic mapping for autonomous off-road navigation," in *Field and Service Robotics*. Springer, 2018.

[8] A. Shaban, X. Meng, J. Lee, B. Boots, and D. Fox, "Semantic terrain classification for off-road autonomous driving," in *Conference on Robot Learning*. PMLR, 2022.

[9] X. Xiao, J. Biswas, and P. Stone, "Learning inverse kinodynamics for accurate high-speed off-road navigation on unstructured terrain," *IEEE Robotics and Automation Letters*, 2021.

[10] H. Karnan, K. S. Sikand, P. Atreya, S. Rabiee, X. Xiao, G. Warnell, P. Stone, and J. Biswas, "Vi-ikd: High-speed accurate off-road navigation using learned visual-inertial inverse kinodynamics," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022.

[11] P. Atreya, H. Karnan, K. S. Sikand, X. Xiao, S. Rabiee, and J. Biswas, "High-speed accurate robot control using learned forward kinodynamics and non-linear least squares optimization," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022.

[12] P. Maheshwari, W. Wang, S. Triest, M. Sivaprakasam, S. Aich, J. G. Rogers III, J. M. Gregory, and S. Scherer, "Piaug–physics informed augmentation for learning vehicle dynamics for off-road navigation," *arXiv preprint arXiv:2311.00815*, 2023.

[13] Y. Pan, C.-A. Cheng, K. Saigol, K. Lee, X. Yan, E. A. Theodorou, and B. Boots, "Imitation learning for agile autonomous driving," *The International Journal of Robotics Research*, 2020.

[14] M. Pfeiffer, M. Schaeuble, J. Nieto, R. Siegwart, and C. Cadena, "From perception to decision: A data-driven approach to end-to-end motion planning for autonomous ground robots," in *IEEE International Conference on Robotics and Automation*. IEEE, 2017.

[15] A. Faust, K. Oslund, O. Ramirez, A. Francis, L. Tapia, M. Fiser, and J. Davidson, "Prm-rl: Long-range robotic navigation tasks by combining reinforcement learning and sampling-based planning," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018.

[16] A. Pokhrel, A. Datar, M. Nazeri, and X. Xiao, "Cahsor: Competence-aware high-speed off-road ground navigation in SE (3)," *arXiv preprint arXiv:2402.07065*, 2024.

[17] G. Kahn, P. Abbeel, and S. Levine, "Badgr: An autonomous self-supervised learning-based navigation system," *arXiv preprint arXiv:2002.05700*, 2020.

[18] K. Stachowicz, D. Shah, A. Bhorkar, I. Kostrikov, and S. Levine, "Fastrlap: A system for learning high-speed driving via deep rl and autonomous practicing," *arXiv preprint arXiv:2304.09831*, 2023.

[19] X. Xiao, T. Zhang, K. M. Choromanski, T.-W. E. Lee, A. Francis, J. Varley, S. Tu, S. Singh, P. Xu, F. Xia, S. M. Persson, L. Takayama, R. Frostig, J. Tan, C. Parada, and V. Sindhwani, "Learning model predictive controllers with real-time attention for real-world navigation," in *Conference on robot learning*. PMLR, 2022.

[20] G. Williams, P. Drews, B. Goldfain, J. M. Rehg, and E. A. Theodorou, "Aggressive driving with model predictive path integral control," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2016.

[21] M. Pivtoraiko, R. A. Knepper, and A. Kelly, "Differentially constrained mobile robot motion planning in state lattices," *Journal of Field Robotics*, 2009.

[22] X. Cai, M. Everett, J. Fink, and J. P. How, "Risk-aware off-road navigation via a learned speed distribution map," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022.

[23] T. Miki, L. Wellhausen, R. Grandia, F. Jenelten, T. Homberger, and M. Hutter, "Elevation mapping for locomotion and navigation using gpu," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022.

[24] K. Chen, R. Nemiroff, and B. T. Lopez, "Direct lidar-inertial odometry: Lightweight lio with continuous-time motion correction," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2023.

[25] G. Williams, A. Aldrich, and E. A. Theodorou, "Model predictive path integral control: From theory to parallel computation," *Journal of Guidance, Control, and Dynamics*, 2017.

[26] D. A. Pomerleau, "Alvinn: An autonomous land vehicle in a neural network," in *Advances in neural information processing systems*, 1989.

[27] M. Bojarski, D. Del Testa, D. Dworakowski, B. Firner, B. Flepp, P. Goyal, L. D. Jackel, M. Monfort, U. Muller, J. Zhang *et al.*, "End to end learning for self-driving cars," *arXiv preprint arXiv:1604.07316*, 2016.

[28] M. H. Nazeri and M. Bohlouli, "Exploring reflective limitation of behavior cloning in autonomous vehicles," in *2021 IEEE International Conference on Data Mining (ICDM)*, 2021, pp. 1252–1257.