Self-Supervised Terrain Representation Learning from Unconstrained Robot Experience

Haresh Karnan, Elvin Yang, Daniel Farkash, Garrett Warnell, Joydeep Biswas, Peter Stone

***Abstract*—*Terrain awareness*, i.e., the ability to sufficiently represent key differences in terrain, is a critical ability that robots must have in order to be able to succeed at autonomous off-road navigation. Current approaches that provide robots with this awareness are prohibitively expensive, requiring curated datasets with extensive human labeling effort or vast amounts of data gathered during expert-level driving. Towards endowing robots with terrain awareness without these expenses, we introduce *Self-supervised TErrain Representation LearnING* (STERLING), a novel approach for learning terrain representations that relies solely on easy-to-collect, unconstrained (e.g., non-expert), and unlabelled robot experience. STERLING employs a novel multi-modal self-supervision objective through non-contrastive representation learning to learn relevant terrain representations for terrain-aware navigation. Through physical robot experiments in off-road environments, we evaluate STERLING features on the task of operator-preference-aligned visual navigation and find that STERLING features perform on par with fully-supervised approaches and outperform other state-of-the-art methods with respect to preference alignment. Additionally, we perform a large-scale experiment of autonomously hiking a 3-mile long trail which STERLING completes successfully with only two manual interventions, demonstrating its robustness to real-world off-road conditions.**

***Index Terms*—Vision-Based Navigation, Representation Learning, Learning from Experience.**

I. INTRODUCTION

OFF-ROAD navigation is emerging as a crucial aspect of autonomous mobile robots, as they are utilized in a growing number of outdoor applications such as autonomous package delivery, search and rescue, and agricultural operations [1]. To enable off-road autonomy while ensuring robot safety and mission success, it is necessary for robots to be able to visually identify distinct terrain features that are relevant to the navigation task, also known as terrain awareness.

Endowing robots with terrain awareness is a challenging problem in autonomous off-road navigation [2, 3]. Prior work has typically relied on large curated datasets [4, 5, 6, 7], vast amounts of expert driving data collected in diverse environments [8, 9, 10], or task demonstrations collected with the supervision of an expert driver in the field [11, 12, 13]. While these approaches are effective, they rely on data that is expensive and labor-intensive to collect. It is typically cheaper and easier to collect unconstrained, unlabelled robot experience with minimal or no expert supervision, and algorithms that can learn from such easy-to-collect data can help scale robot learning in a self-supervised way. More recently, self-supervised approaches that can learn from such unconstrained robot experience [3, 14, 15] have shown promising results in off-road navigation. However, most self-supervised methods focus on learning specific behaviors, such as improving ride comfort [14] or collision avoidance [15], and as a result, the terrain features learned are relevant only for that specific task and may need to be retrained if the downstream task objectives change. To enable off-road autonomy at scale, new approaches are needed that address the perception problem of learning relevant terrain representations in a self-supervised way from easy-to-collect unconstrained robot experiences gathered without requiring an expert agent in the field. Learning such rich visual features of the terrain through self-supervision by leveraging multi-modal sensor data collected onboard a mobile robot can help tackle many off-road navigation problems such as high-speed navigation [2, 16], traversing challenging terrain [17, 18], and preference-aligned planning [12, 7].

Towards robot terrain awareness from unconstrained and unlabelled robot experience, we introduce *Self-supervised TErrain Representation LearnING* (STERLING), a novel approach to learning terrain representations for off-road navigation. STERLING utilizes easy-to-collect unconstrained robot experiences and employs a novel multi-modal self-supervision objective through non-contrastive, non-reconstructive representation learning to learn relevant terrain representations useful for visual terrain-aware navigation. We evaluate STERLING against baseline methods on the task of operator preference-aligned off-road navigation and find that STERLING performs on par with or better than existing methods with respect to preference alignment. We additionally perform a large-scale qualitative experiment of autonomously hiking a 3-mile long trail[1], demonstrating the effectiveness of STERLING-features.

II. RELATED WORK

In this section, we review related work on self-supervised learning for terrain-aware navigation. To alleviate the need for extensive human labeling, self-supervised learning methods have been proposed to either learn terrain representations or costs from data gathered onboard a mobile robot.

Brooks et al. [19] utilize contact vibrations and visual sensors to classify terrains via self-supervision. Zürn et al. [3] introduce SE-R which utilizes acoustic and visual sensors

Haresh Karnan is with the Walker Department of Mechanical Engineering, The University of Texas at Austin (email: haresh.miriyala@utexas.edu)

Elvin Yang, Joydeep Biswas, Peter Stone are with the Department of Computer Science, The University of Texas at Austin (email: {eyang, joydeepb, pstone}@cs.utexas.edu)

Daniel Farkash is with Cornell University (email: dmf248@cornell.edu)

Garrett Warnell is with Army Research Laboratory (email: garrett.a.warnell.civ@army.mil)

Peter Stone is also with Sony AI, North America

[1]Autonomously hiking the Ann & Roy Butler trail @ Austin, Texas, USA

on the robot to segment terrains using a self-supervised triplet-contrastive learning framework. Using triplet-based contrastive learning methods requires negative samples which may not be available when learning using unlabeled data. In STERLING, we use recently proposed non-contrastive unsupervised learning approaches such as VICReg [20] that do not require any negative samples and instead rely on correlations between data modalities to learn terrain representations.

Several methods have applied self-supervision to assign traversability costs for the downstream off-road navigation task [14, 21, 15, 22]. Specifically, these methods rely on inertial spectral features [14], future predictive models [15], inertial-odometry errors [21], or foothold positions [22] as self-supervision signals to learn a traversability cost map, used to evaluate candidate actions. Instead of inferring costs or rewards using self-supervision for a fixed task, in this work, we focus on learning relevant visual features from unconstrained robot experiences that could be used in downstream tasks. This framework allows a designer to reuse features across tasks without retraining entirely from scratch.

The approach closest to ours is VRL-PAP [12] which requires human expert teleoperated demonstrations of a particular trajectory pattern to both explicitly learn visual terrain representations as well as to infer terrain preference costs. However, in this work, we focus on learning terrain features from unconstrained robot experiences without requiring a human expert in the field for demonstrations, which is a more general problem than the one considered by VRL-PAP [12].

III. Approach

In this section, we introduce the self-supervised terrain representation learning approach, STERLING, proposed in this work. The offline pre-processing performed on the recorded data is detailed in Appendix VI-C. The mathematical formulation for preference-aligned off-road navigation, and integrating STERLING within a planner is explained in Appendix VI-A.

Terrain Representation Learning: It is desired for learned representations of terrains to be such that representations of similar terrain are close together and of different terrains are distinctly separated. Although we do not possess privileged information such as semantic labels of terrains for training since the data is unlabelled, the visual and kinodynamic observations experienced by the robot reflect similarities and differences between terrain samples. For instance, traversing a smooth terrain such as `cement sidewalk` may require relatively less effort on the robot's joints, whereas a rough terrain such as `marbled rocks` may require additional effort, measured as joint torque by the proprioceptive sensors. STERLING leverages this multi-modal experience observed by the robot and computes a correlation objective between visual and inertial-proprio-tactile signals to learn desired terrain representations. Additionally, STERLING uses viewpoint invariance as an objective unique to the visual component of the experience to learn viewpoint-invariant terrain representations.

Fig. 1 provides an overview of the self-supervised representation learning framework adopted in STERLING. A parameterized visual encoder (4-layer CNN with 0.25 million

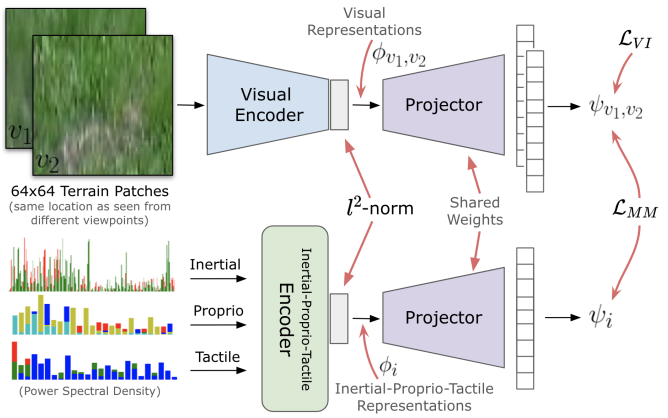

Fig. 1. Overview of the self-supervised training architecture used in STERLING. Two terrain patches v_1 and v_2 of the same location from different viewpoints are encoded as ϕ_{v_1} and ϕ_{v_2} respectively, and mapped into embeddings ψ_{v_1} and ψ_{v_2}. Similarly, inertial, proprio, tactile signals are encoded as ϕ_i, and mapped into ψ_i. Self-supervision objectives $\mathcal{L}_{VI}(\psi_{v_1}, \psi_{v_2})$ for viewpoint-invariance and $\mathcal{L}_{MM}(\psi_{v_{1,2}}, \psi_i)$ for multi-modal correlation are computed across a mini-batch of samples, and parameters of the two encoders and the shared projector network are updated together.

parameters) encodes terrain image patches of the same location v_1 and v_2 into visual representations ϕ_{v_1} and ϕ_{v_2}, respectively. Similarly, an inertial-proprio-tactile encoder (4-layer MLP with 0.25 million parameters) encodes frequency domain IPT observations of the robot at that location to an inertial-proprio-tactile representation ϕ_i. We follow the framework of prior self-supervised representation learning algorithms from the computer vision community such as VICReg [20], and utilize a parameterized projector network (2-layer MLP with 0.25 million parameters) that maps encoded visual and non-visual representations independently to a higher-dimensional feature space $\psi_{v_{1,2}}$ and ψ_i respectively, over which the self-supervision objectives are computed. The STERLING objective composed of the multi-modal correlation $\mathcal{L}_{MM}(\psi_{v_{1,2}}, \psi_i)$ and viewpoint-invariance $\mathcal{L}_{VI}(\psi_{v_1}, \psi_{v_2})$ objectives are defined as:

$$\begin{aligned} \mathcal{L}_{\text{STERLING}} &= \mathcal{L}_{VI}(\psi_{v_1}, \psi_{v_2}) + \mathcal{L}_{MM}(\psi_{v_{1,2}}, \psi_i) \\ \mathcal{L}_{VI}(\psi_{v_1}, \psi_{v_2}) &= \mathcal{L}_{\text{VICR}eg}(\psi_{v_1}, \psi_{v_2}) \\ \mathcal{L}_{MM}(\psi_{v_{1,2}}, \psi_i) &= [\mathcal{L}_{\text{VICR}eg}(\psi_{v_1}, \psi_i) + \mathcal{L}_{\text{VICR}eg}(\psi_{v_2}, \psi_i)]/2 \end{aligned} \tag{1}$$

$\mathcal{L}_{\text{VICR}eg}$ is the VICReg loss that is composed of variance-invariance-covariance representation learning objectives, as proposed by Bardes et al. [20]. Given two alternate projected representations Z and Z' of a data sample (in STERLING, Z and Z' are projected representations of the visual and non-visual sensor modalities), the VICReg loss is defined as $\mathcal{L}_{\text{VICR}eg}(Z, Z') = \lambda s(Z, Z') + \mu[v(Z) + v(Z')] + \nu[c(Z) + c(Z')]$. λ, μ, and ν are hyper-parameters and the functions v, s, and c are the variance, invariance, and covariance terms computed on a mini-batch of projected features. We refer the reader to Bardes et al. [20] for more details regarding the individual terms in the loss function. On a mini-batch of data containing paired terrain image patches and IPT observations, we compute the $\mathcal{L}_{\text{STERLING}}$ loss and update parameters of the two encoder networks and the shared projector network

together using Adam optimizer.

IV. Experiments

In this section, we describe the experiments performed to evaluate STERLING. Specifically, the experiments presented in this section are tailored to address the following questions:

(Q_1) How effective are STERLING features in comparison to baseline approaches at enabling terrain awareness in off-road navigation?

(Q_2) In a large-scale real-world off-road setting, how reliable is STERLING at autonomous off-road navigation?

Question Q_1 is evaluated through physical robot experiments on the task of operator-preference-aligned off-road navigation. We then investigate Q_2 by conducting a large-scale qualitative evaluation by autonomously hiking a 3-mile long off-road trail in Austin, Texas, USA using preference costs learned using STERLING features. In Appendix Sec VI-D, we present an additional quantitative study comparing STERLING with other unsupervised representation learning methods and perform an ablation on the proposed self-supervision objectives. To quantitatively compare various methods, we use the success rate of preference alignment as the metric. A trial is considered preference aligned if the robot follows the shortest path with the most preferred terrain. However, if there are other shorter trajectories with even more preferred terrains, the trial is deemed a failure in achieving preference alignment. Details on the data collection procedure are provided in Appendix Sec. VI-B.

Baselines: To perform quantitative evaluations for Q_2, we compare STERLING with SE-R [3], RCA [14], GANav [7], geometric-only planning [23], and a fully-supervised baseline. SE-R and RCA perform self-supervised learning from unconstrained robot experience to learn terrain representations and traversability costs respectively. The geometric-only approach ignores terrain costs ($\mathcal{L}_{terrain}$) and plans with geometric cost ($\mathcal{L}_{geom}$) only. GANav is an image segmentation-based navigation framework trained on the RUGD [4] dataset. We use the open-source author-provided implementation of GANav[2]. Since an open-source implementation of RCA is unavailable, we replicate it to the best of our efforts. We additionally train the fully-supervised baseline in which the terrain cost function is learned end-to-end using supervised learning from linear extrapolation of operator preferences.

A. Evaluating Terrain-Awareness via Robot Experiments

In this subsection, we report on experiments to investigate the effectiveness of STERLING features in enabling terrain awareness during off-road navigation. We quantitatively compare the performance of STERLING with baselines RCA [14], GANav [7], SE-R [3] and the fully-supervised baseline, on the task of operator-preference-aligned navigation. We identify six environments within the UT Austin campus, with eight different terrain types, as shown in Fig. 2. For this study, we use the same data collected on the robot to train RCA, SE-R, fully-supervised baseline, and STERLING, and the operator provides the same rankings for all methods during training.

[2]https://github.com/rayguan97/GANav-offroad

Fig. 2 shows the operator's (first author) terrain preferences for all Envs. 1 to 5, and the performance of baseline approaches, including a human operator demonstrated trajectory for reference. In all environments, we see that STERLING navigates in a terrain-aware manner while adhering to operator-provided preferences. Note that although Fully-Supervised also completes the task successfully, it requires privileged information such as terrain labels during training, whereas STERLING does not require such supervision, and can potentially be used on large datasets containing unlabelled, unconstrained robot experiences. RCA uses inertial spectral features to learn terrain traversability costs and hence does not adhere to operator preference. SE-R does not address viewpoint invariance which is a significant problem in vision-based off-road navigation and hence performs poorly in Envs. 1 and 2. We additionally study adhering to operator preferences when the preference changes in the same environment (Env. 6). The results of this study are detailed in Sec. VI-E. Table I shows the success rate of preference alignment for all approaches in all environments, over five different trials. We see that STERLING outperforms other self-supervised baselines and performs on par with the fully-supervised approach.

B. Large-Scale Qualitative Evaluation

In this subsection, we address Q_2 by performing a large-scale evaluation of STERLING by autonomously hiking a 3-mile-long off-road trail[3]. We train STERLING using unconstrained robot experience collected within the UT Austin campus and train the preference utility function using operator-provided preferences: `marble rocks < grass < dirt = cement`. The task is to navigate the trail without a global map, adhering to operator preferences at all times. Since we do not use a global map, visual terrain awareness is necessary to navigate within the trail and avoid catastrophic events such as falling into the river next to the trail. While the robot navigates autonomously, the operator walks behind the robot and takes manual control only to correct the robot's path during forks, or to yield to incoming pedestrians and pets. The attached video [4] shows the robot navigating the trail successfully while avoiding less preferred terrains. The robot needed two manual interventions while traversing along the trail. Fig. 3 shows the 3-mile trajectory traced by the robot and the two failure cases that required manual intervention. This large-scale qualitative experiment addresses Q_2 by demonstrating the reliability of STERLING during real-world off-road deployments.

V. Conclusion

In this paper, we introduce *Self-supervised TErrain Representation LearnING* (STERLING), a novel framework for learning terrain representations that relies solely on easy-to-collect, unconstrained (e.g., non-expert), and unlabelled robot experience. STERLING utilizes non-contrastive multi-modal self-supervised learning through two objectives—viewpoint invariance and multi-modal correlation—to learn relevant representations for terrain-aware visual navigation. We show how

[3]Ann and Roy Butler Hike and Bike Trail, Austin, TX, USA

[4]https://youtu.be/dQb1XzocdtE

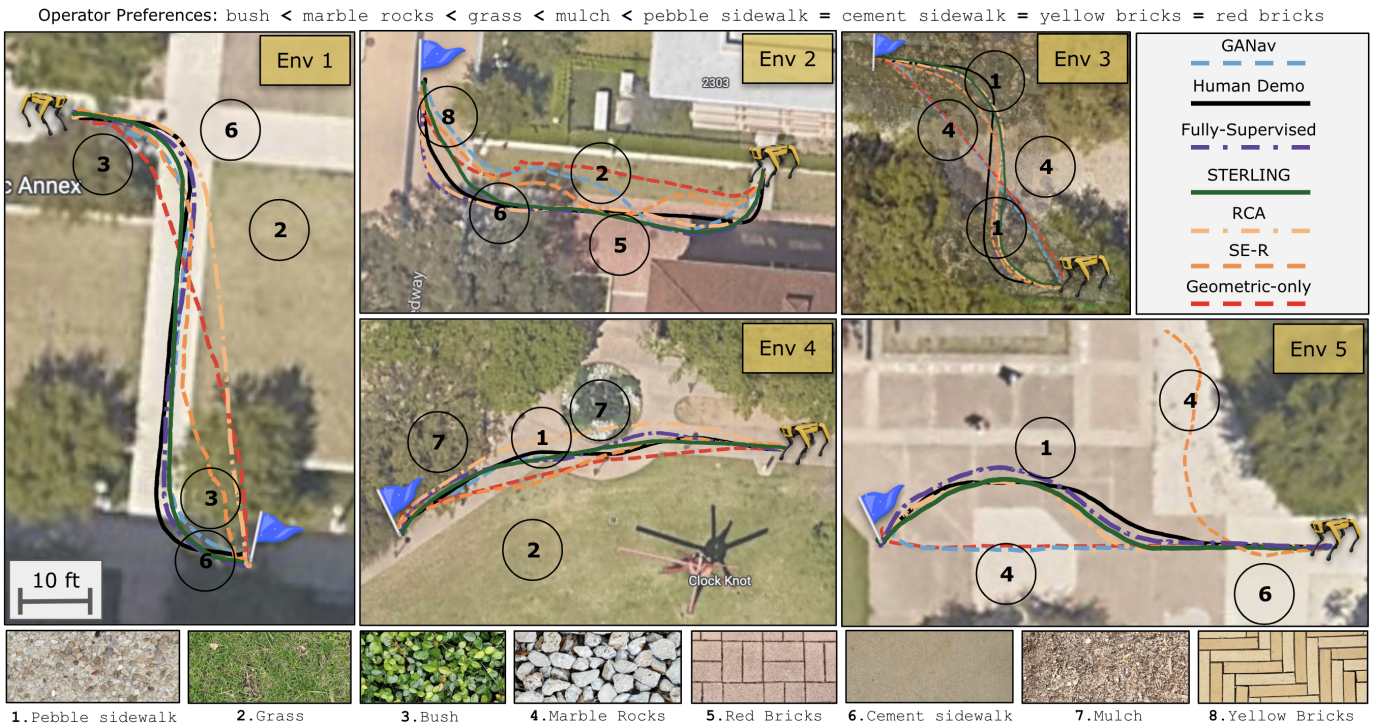

Fig. 2. Trajectories traced by different approaches in 5 environments within the UT Austin University campus, containing 8 different terrains. The operator preferences for the terrains in all environments are shown above. We see that STERLING navigates in an operator-preference aligned manner, by preferring `cement sidewalk`, `red bricks`, `pebble sidewalk`, and `yellow bricks` over `mulch`, `grass`, `marble rocks`, and `bush`, outperforming other baselines and performing on-par with the Fully-Supervised approach.

Fig. 3. A large-scale qualitative evaluation of STERLING on the 3-mile Ann and Roy Butler trail in Austin, Texas, USA. Without a global map, preference cost learned using STERLING features complete the hike successfully with only two manual interventions (shown in red).

features learned through STERLING can be utilized to learn operator preferences over terrains and integrated within a planner for preference-aligned navigation. We evaluate STERLING against state-of-the-art off-road navigation and unsupervised representation learning methods on the task of operator-preference-aligned visual navigation on a Spot robot and find that STERLING outperforms other methods and performs on par with a fully-supervised baseline. We additionally perform a qualitative large-scale experiment by successfully hiking a 3-mile-long trail using STERLING, demonstrating its robustness to off-road conditions in the real world.

TABLE I
SUCCESS RATES OF DIFFERENT ALGORITHMS ON THE TASK OF OPERATOR-PREFERENCE-ALIGNED OFF-ROAD NAVIGATION

Approach	Environment						
	1	2	3	4	5	6 (a)	6 (b)
Geometric-only	0/5	0/5	0/5	0/5	0/5	0/5	5/5
RCA[14]	5/5	4/5	2/5	0/5	1/5	5/5	0/5
GANav[7]	0/5	0/5	0/5	5/5	5/5	4/5	5/5
SE-R[3]	1/5	0/5	5/5	5/5	3/5	5/5	4/5
Fully-Supervised	5/5	5/5	5/5	5/5	5/5	5/5	5/5
STERLING (Ours)	5/5	5/5	5/5	5/5	5/5	5/5	5/5

ACKNOWLEDGMENT

This work has taken place in the Learning Agents Research Group (LARG) and Autonomous Mobile Robotics Laboratory (AMRL) at UT Austin. LARG research is supported in part by NSF (CPS-1739964, IIS-1724157, NRI-1925082), ONR (N00014-18-2243), FLI (RFP2-000), ARO (W911NF19-2-0333), DARPA, Lockheed Martin, GM, and Bosch. AMRL research is supported in part by NSF (CAREER2046955, IIS-1954778, SHF-2006404), ARO (W911NF-19-2- 0333, W911NF-21-20217), DARPA (HR001120C0031), Amazon, JP Morgan, and Northrop Grumman Mission Systems. Peter Stone serves as the Executive Director of Sony AI America and receives financial compensation for this work. The terms of this arrangement have been reviewed and approved by the University of Texas at Austin in accordance with its policy on objectivity in research.

REFERENCES

[1] T. Wang, B. Chen, Z. Zhang, H. Li, and M. Zhang, "Applications of machine vision in agricultural robot navigation: A review," *Computers and Electronics in Agriculture*, vol. 198, p. 107085, 2022.

[2] H. Karnan, K. S. Sikand, P. Atreya, S. Rabiee, X. Xiao, G. Warnell, P. Stone, and J. Biswas, "Vi-ikd: High-speed accurate off-road navigation using learned visual-inertial inverse kinodynamics," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022, pp. 3294–3301.

[3] J. Zürn, W. Burgard, and A. Valada, "Self-supervised visual terrain classification from unsupervised acoustic feature learning," *IEEE Transactions on Robotics*, vol. 37, no. 2, pp. 466–481, 2020.

[4] M. Wigness, S. Eum, J. G. Rogers, D. Han, and H. Kwon, "A rugd dataset for autonomous navigation and visual perception in unstructured outdoor environments," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2019, pp. 5000–5007.

[5] P. Jiang, P. Osteen, M. Wigness, and S. Saripalli, "Rellis-3d dataset: Data, benchmarks and analysis," 2020.

[6] S. Sharma, L. Dabbiru, T. Hannis, G. Mason, D. W. Carruth, M. Doude, C. Goodin, C. Hudson, S. Ozier, J. E. Ball *et al.*, "Cat: Cavs traversability dataset for off-road autonomous driving," *IEEE Access*, vol. 10, pp. 24 759–24 768, 2022.

[7] T. Guan, D. Kothandaraman, R. Chandra, A. J. Sathyamoorthy, K. Weerakoon, and D. Manocha, "Ga-nav: Efficient terrain segmentation for robot navigation in unstructured outdoor environments," *IEEE Robotics and Automation Letters*, vol. 7, no. 3, pp. 8138–8145, 2022.

[8] M. Bojarski, D. Del Testa, D. Dworakowski, B. Firner, B. Flepp, P. Goyal, L. D. Jackel, M. Monfort, U. Muller, J. Zhang, X. Zhang, J. Zhao, and K. Zieba, "End to end learning for self-driving cars," 2016.

[9] Y. LeCun, U. Muller, J. Ben, E. Cosatto, and B. Flepp, "Off-road obstacle avoidance through end-to-end learning," in *Proceedings of the 18th International Conference on Neural Information Processing Systems*, ser. NIPS'05. Cambridge, MA, USA: MIT Press, 2005, p. 739–746.

[10] M. Wulfmeier, P. Ondruska, and I. Posner, "Maximum entropy deep inverse reinforcement learning," *arXiv preprint arXiv:1507.04888*, 2015.

[11] G. Kahn, P. Abbeel, and S. Levine, "Land: Learning to navigate from disengagements," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 1872–1879, 2021.

[12] K. S. Sikand, S. Rabiee, A. Uccello, X. Xiao, G. Warnell, and J. Biswas, "Visual representation learning for preference-aware path planning," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 11 303–11 309.

[13] Y. Pan, C.-A. Cheng, K. Saigol, K. Lee, X. Yan, E. A. Theodorou, and B. Boots, "Imitation learning for agile autonomous driving," *The International Journal of Robotics Research*, vol. 39, no. 2-3, pp. 286–302, 2020.

[14] X. Yao, J. Zhang, and J. Oh, "Rca: Ride comfort-aware visual navigation via self-supervised learning," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022, pp. 7847–7852.

[15] G. Kahn, P. Abbeel, and S. Levine, "Badgr: An autonomous self-supervised learning-based navigation system," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 1312–1319, 2021.

[16] X. Meng, N. Hatch, A. Lambert, A. Li, N. Wagener, M. Schmittle, J. Lee, W. Yuan, Z. Chen, S. Deng, G. Okopal, D. Fox, B. Boots, and A. Shaban, "Terrainnet: Visual modeling of complex terrain for high-speed, off-road navigation," 2023.

[17] A. Datar, C. Pan, M. Nazeri, and X. Xiao, "Toward wheeled mobility on vertically challenging terrain: Platforms, datasets, and algorithms," 2023.

[18] T. Miki, J. Lee, J. Hwangbo, L. Wellhausen, V. Koltun, and M. Hutter, "Learning robust perceptive locomotion for quadrupedal robots in the wild," *Science Robotics*, vol. 7, no. 62, p. eabk2822, 2022.

[19] C. A. Brooks and K. D. Iagnemma, "Self-supervised classification for planetary rover terrain sensing," in *2007 IEEE aerospace conference*. IEEE, 2007, pp. 1–9.

[20] A. Bardes, J. Ponce, and Y. LeCun, "Vicreg: Variance-invariance-covariance regularization for self-supervised learning," *arXiv preprint arXiv:2105.04906*, 2021.

[21] A. J. Sathyamoorthy, K. Weerakoon, T. Guan, J. Liang, and D. Manocha, "Terrapn: Unstructured terrain navigation using online self-supervised learning," 2022.

[22] 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 Robotics and Automation Letters*, vol. 4, no. 2, pp. 1509–1516, 2019.

[23] J. Biswas, "Amrl autonomy stack," https://github.com/ut-amrl/graph_navigation, 2013.

[24] M. Zucker, N. Ratliff, M. Stolle, J. Chestnutt, J. A. Bagnell, C. G. Atkeson, and J. Kuffner, "Optimization and learning for rough terrain legged locomotion," *The International Journal of Robotics Research*, vol. 30, no. 2, pp. 175–191, 2011.

VI. Appendix

A. Preference-Aligned Off-Road Navigation

In this subsection, we describe the problem formulation of terrain-aware navigation and how STERLING features can be utilized within a planner to navigate in an operator-preference-aligned manner. While the main novelty of our work is on learning terrain representations from unconstrained robot experience, we use operator-preference-aligned terrain-aware navigation as a real-world task to evaluate STERLING features.

Preliminaries: We formulate the task of operator-preference-aligned terrain-aware navigation as a local path-planning problem, where the robot operates within a state space $\mathcal{S}$, action space $\mathcal{A}$, and a deterministic transition function $\mathcal{T} : \mathcal{S} \times \mathcal{A} \longrightarrow \mathcal{S}$ in the environment. The state space consists of $s = [x, y, \theta, \phi_v]$, where $[x, y, \theta]$ denote the robot's position in $SE(2)$ space, and ϕ_v denotes the visual features of the terrain at this location. Given a goal location G, the preference-aligned navigation task is to reach this goal while adhering to operator preferences over terrains.

Sampling-based planning: We assume access to a receding horizon sampling-based motion planner with a fixed set of constant-curvature arcs $\{\Gamma_0, \Gamma_1, \ldots, \Gamma_{ns}\}$, $\Gamma \in \mathcal{S}^N$ which solves for the optimal arc $\Gamma^* = \arg\min_{\Gamma}[\mathcal{J}(\Gamma, G)]$, minimizing the objective function $\mathcal{J}(\Gamma, G), \mathcal{J} : (\Gamma, G) \longrightarrow \mathbb{R}^+$. For the task of preference-aligned off-road navigation, we assume the objective function is composed of two components $\mathcal{J}_{geom}(\Gamma, G)$ and $\mathcal{J}_{terrain}(\Gamma)$, and can be defined as $\mathcal{J}(\Gamma, G) = \alpha\mathcal{J}_{geom}(\Gamma, G) + (1-\alpha)\mathcal{J}_{terrain}(\Gamma)$. $\mathcal{J}_{geom}(\Gamma, G)$ is the geometric cost that deals with progress towards the goal G and avoiding geometric obstacles, whereas $\mathcal{J}_{terrain}(\Gamma)$ is the terrain cost associated with preference-alignment. We utilize the geometric cost as defined in AMRL's graph navigation stack [5]. The multiplier $\alpha \in [0, 1]$ trades off relative contributions of the geometric and terrain preference components of the path planning objective. A 1D time-optimal controller translates the sequence of states in the optimal trajectory Γ^* to a sequence of receding horizon actions $(a_0, a_1, \ldots, a_N)$. For a given arc $\Gamma = \{s_0, s_1, \ldots, s_N\}$, such that state s_0 is closest to the robot, the terrain-preference cost can be computed as follows.

$$\mathcal{J}_{terrain}(\Gamma) = \sum_{v_i \sim \Gamma, i=0}^{N} \frac{\gamma^i C(u(f_v(v_i)))}{N} \tag{2}$$

The function $f_v(.)$ maps from RGB space of a visual patch of terrain v_i at a specific state s_i, to its visual representation $\phi_v \in \Phi_v$. For instance, f_v can be the visual encoder learned using STERLING, as described in Section III. The utility function $u(.)$ maps the visual representation ϕ_v of a patch of terrain to a real-valued utility of preferences. We follow the utility function formulation of Zucker et al. [24] and assume the terrain preference cost follows a multiplicative formulation such that given a utility value $x \in \mathbb{R}^+$, the traversability cost is $C(x) = e^{-x}$. The discount factor γ weighs the terrain cost proportional to its proximity to the robot. We set γ to 0.8, which we find to work well in practice.

Learning the preference utility: Following Zucker et al. [24], we learn the utility function $u : \Phi_v \rightarrow \mathbb{R}^+$ using human queries. From the predicted terrain features on data samples in our training set, we cluster the terrain representations using k-means with silhouette-score elbow criterion, and sample candidate terrain patches from each cluster, which is presented to the human operator using a GUI. The human operator then provides a full-order ranking of terrain preferences over clusters, which is utilized to learn the utility function $u(.)$, represented by a 2-layer MLP. While this method cannot recover absolute preferences between terrains and can only preserve their preference rankings, we find that this approximation provided by Zucker et al. [24] works well in practice.

Planning at deployment: Fig. 4 provides an overview of the cost inference process for local planning at deployment. To evaluate the terrain cost $\mathcal{J}_{terrain}(\Gamma)$ for the constant-curvature arcs, we overlay the arcs on the bird's eye view image, extract terrain patches at states along the arc, and compute the cost according to Eq. 2. We compute the visual representation, utility value, and terrain cost of all images at once as a single batch inference. Since the visual encoder and the utility function are relatively lightweight neural networks with about 0.5 million parameters, we are able to achieve real-time planning rates of 40 Hz using a laptop-grade Nvidia GPU.

[5]https://github.com/ut-amrl/graph_navigation

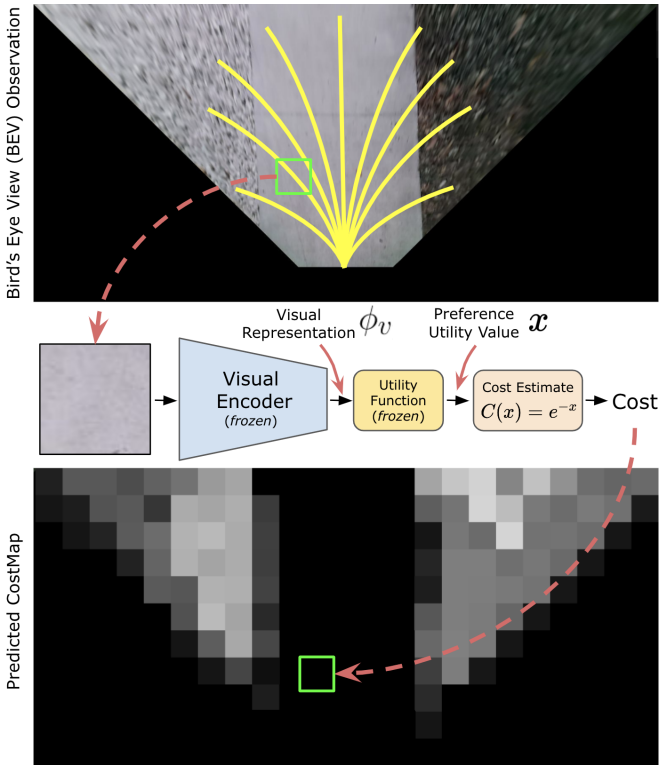

Fig. 4. An overview of the cost inference process for local planning at deployment. The constant-curvature arcs (yellow) are overlayed on the BEV image, and the terrain cost $\mathcal{J}_{terrain}(\Gamma)$ is computed on patches extracted along all arcs. White is high cost and black is low cost.

B. Data Collection

In all experiments, we use a legged Boston Dynamics Spot robot and collect robot experiences on eight different terrains—`mulch`, `pebble sidewalk`, `cement sidewalk`, `grass`, `bushes`, `marbled rock`, `yellow bricks`, `and red bricks`—around the UT Austin University campus. The data is collected through human teleoperation (by the first and second authors) such that each trajectory contains a unique terrain throughout, with random trajectory shapes. Note that STERLING does not require a human expert to teleoperate the robot to collect robot experience nor does it require the experience to be gathered on a unique terrain per trajectory. We follow this data collection approach since it is easier to label the terrain, for evaluation purposes. STERLING can also work with random trajectory lengths, with multiple terrains encountered along the same trajectory, without any semantic labels such as terrain names, and any navigation policy can be used for data collection. We record 8 trajectories per terrain, each five minutes long, and use 4 trajectories for training and the remaining for validation.

C. Data-Collection and Pre-Processing

STERLING learns terrain representations from unconstrained, unlabeled robot experiences collected using any navigation policy, for

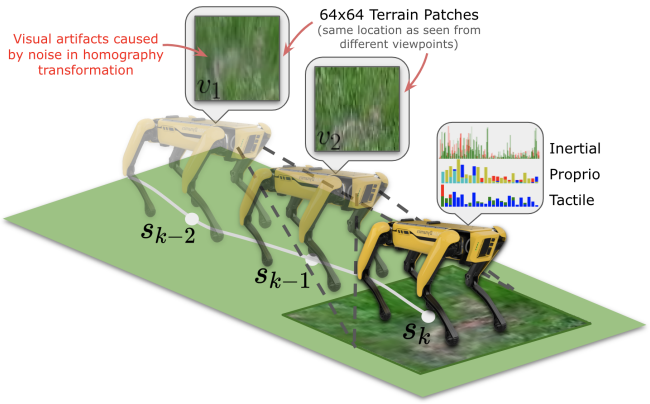

Fig. 5. An illustration of the offline data preprocessing performed on unconstrained robot experience. Image patches of traversed terrain at location s_k are extracted from bird's eye view observations at prior locations s_{k-1}, s_{k-2} along the trajectory. Note the visual artifacts in patches extracted further away from the scene due to homography measurement errors. The corresponding inertial, proprioception, and tactile observations at s_k are transformed from time series to power spectral density signals.

instance, human teleoperation, curiosity-driven exploration, or point-to-point navigation using any underlying planner. Collecting such robot experience is cheap and easy, in comparison to requiring a human expert to provide demonstrations and labels. We additionally assume that the robot is equipped with multiple sensors such as an egocentric RGB camera, odometry sensors, onboard IMU, proprioceptive, and tactile sensors that together provide rich multi-modal observations as the robot traverses over different terrains collecting experience. STERLING leverages this multi-modal data, and correlation between different modes of data to learn terrain representations. In order to learn terrain representations using STERLING, we begin by pre-processing the visual and non-visual observations, which are explained in detail below.

Visual Patch Extraction: The egocentric camera observations are homography-projected into a virtual bird's eye view (BEV) frame using the intrinsic and extrinsic camera matrices. As shown in Fig. 5, we project the robot's trajectory onto the bird's eye view frame and extract 64-by-64 pixels (equivalent to the robot's footprint of 0.5-by-0.5 meters) square visual patches of the terrain along with the corresponding inertial, proprioceptive, and tactile measurements at the same location, along the trajectory. To extract visual patches of terrain at a particular location, say s_k, since the terrain underneath is visually unobservable from s_k, we extract the terrain patches from BEV observations of the robot at previous locations $s_{k-1}, s_{k-2}, \ldots$ along the trajectory. Fig. 5 illustrates the offline patch extraction process from two previous viewpoints, however, we extract patches from up to 20 previous viewpoints within 2 meters. Although just one viewpoint is sufficient to learn the correlation between visual and other sensor observations, when planning to navigate, one may need to visually evaluate terrains at future locations ahead of the robot. As shown in Fig. 5, errors in homography transformation due to noisy camera extrinsic measurements may cause visual artifacts such as stretching and blur at locations farther away from the robot in the BEV image, and may potentially affect planning. To address this problem, in addition to the multi-modal correlation objective on visual and non-visual data at a location, STERLING utilizes a viewpoint invariance objective, which enforces invariance to homography artifacts in the learned visual representations.

IPT Preprocessing: For the non-visual modes of data such as inertial, proprioceptive, and tactile (IPT) observations, we retain up to 2-second history and convert the time-series signals into power-spectral density (PSD) representation in the frequency domain. This ensures the IPT time-series data is phase independent, which is more conducive to the architecture of our representation learning system.

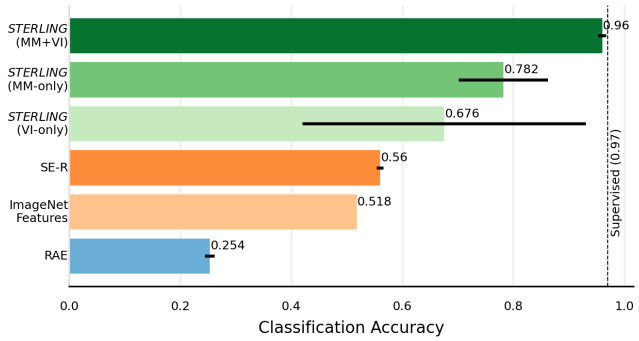

Fig. 6. Ablation study depicting unsupervised k-means-classification accuracy (value closer to 1.0 is better) from terrain representations learned using different approaches and objectives. We see that the combined objective (VI + MM) proposed in STERLING achieves the highest classification accuracy, indicating that the learned representations are sufficiently discriminative of terrains.

D. Evaluating Self-Supervision Objectives

In this subsection, we investigate the effectiveness of STERLING at learning discriminative terrain features and compare with baseline unsupervised terrain representation learning methods such as Regularized Auto-Encoder (RAE) and SE-R [3]. STERLING uses multi-modal correlation ($\mathcal{L}_{MM}$) and viewpoint invariance ($\mathcal{L}_{VI}$) objectives for self-supervised representation learning. SE-R and RAE use soft-triplet-contrastive loss and pixel-wise reconstruction loss respectively to learn visual representations of terrains. Additionally, we also perform an ablation study on the two objectives in STERLING to understand their contributions to learning discriminative terrain features. To evaluate different visual representations, we perform unsupervised classification using k-means clustering (using silhouette score to compute the elbow) and compare their relative classification accuracies with true terrain labels. For this experiment, we train STERLING, SE-R, and RAE on our training set and evaluate on a held-out validation set. Fig. 6 shows the results of this study. We see that STERLING features using both the self-supervision objectives perform the best among all methods. Additionally, we see that using a non-contrastive representation learning approach such as VICReg [20] within STERLING performs better than contrastive learning methods such as SE-R, and reconstruction-based methods such as RAE. This study helps address Q_1 and shows that the proposed self-supervision objectives in STERLING indeed help learn discriminative terrain features.

E. Preference Alignment Evaluation

We utilize Env. 6 to further study adherence to operator preferences. We hypothesize that the discriminative features learned using STERLING is sufficient to learn the preference cost for local planning. To test this hypothesis, in Env. 6 containing three terrains as shown in Fig. 7, the operator provides two different preferences 6(a) and 6(b). While `bush` is the least preferred in both cases, in 6(a), `sidewalk` is more preferred than `grass` and in 6(b), both `grass` and `sidewalk` are equally preferred. We see in Fig. 7 that using STERLING features, the planner is able to sufficiently distinguish the terrains and reach the goal while adhering to operator preferences. Although SE-R [3] adheres to operator preference in 6(b), it incorrectly maps `grass` to `bush`, assigning a higher cost and taking a longer route to reach the goal. Table I shows the success rate of preference alignment for all approaches in all environments, over five different trials. We see that STERLING outperforms other self-supervised baselines and performs on par with the fully-supervised approach.

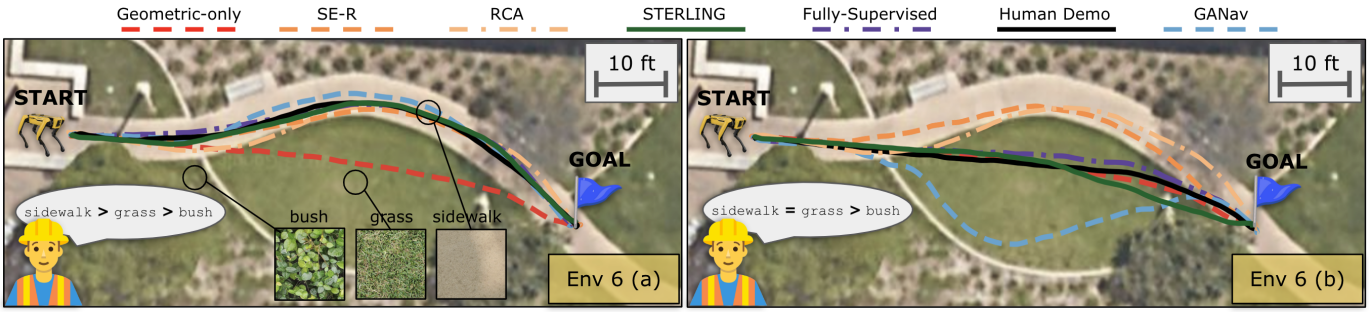

Fig. 7. Trajectories traced by different approaches for the task of operator-preference-aligned off-road navigation. Shown here are two different preferences expressed by the operator in the same environment—in 6 (a), `sidewalk` is more preferred than `grass` which is more preferred that `bush`, and in 6 (b), `grass` and `sidewalk` are equally preferred and `bush` is least preferred. We see that without retraining the terrain features, in both cases (a) and (b), STERLING optimally navigates to the goal while adhering to operator preferences.