TOYOTA RESEARCH INSTITUTE

NAVER LABS

NVIDIA

NSF

IROS KYOTO 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems

M | ROBOTICS

Energy-Based Legged Robots Terrain Traversability Modeling via Deep Inverse Reinforcement Learning

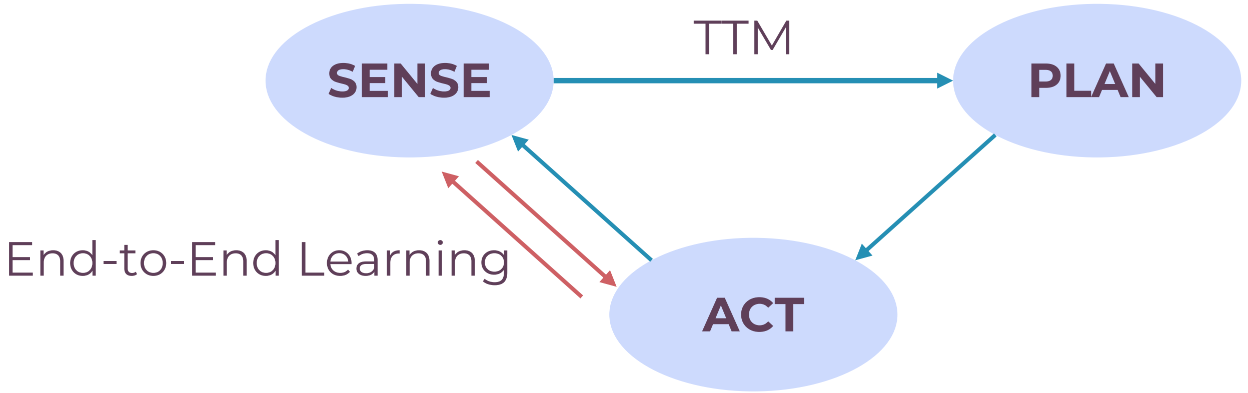
Lu Gan, Jessy W. Grizzle, Ryan M. Eustice, and Maani Ghaffari

Robotics Institute, University of Michigan

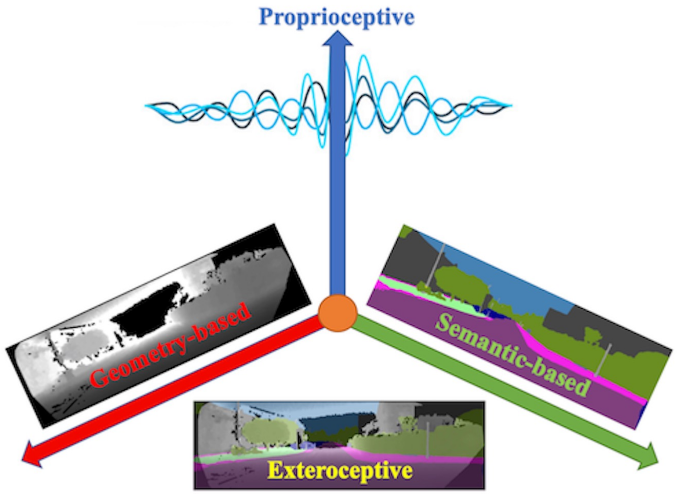
IEEE/RSJ International Conference on Intelligent Robots and Systems

Introduction

- Terrain modeling (TTM) vs. end-to-end learning for robot navigation and exploration

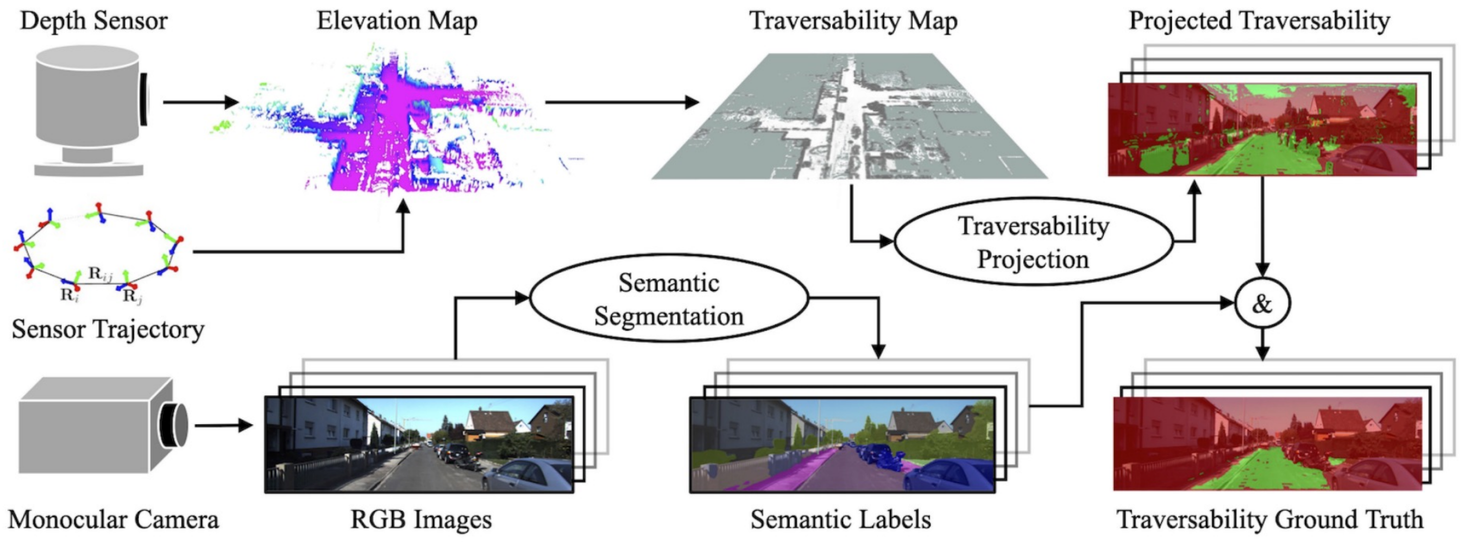


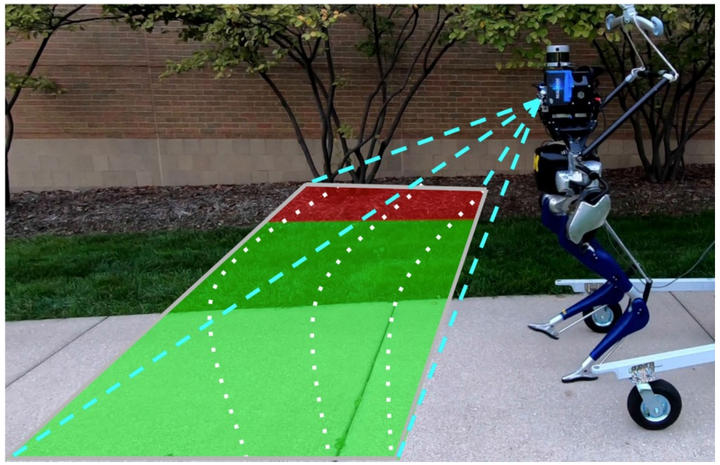
- Classical methods for TTM:
  - Classification: semantic-based
  - Regression: geometry-based, proprioceptive



Current TTM for Legged Robots

- Our previous work (2022)
  - Geometry-based + Semantic-based



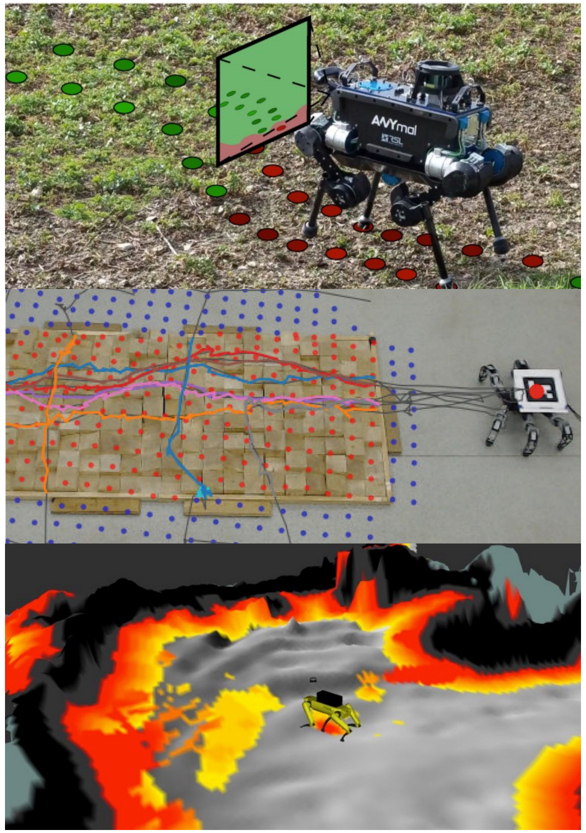


[1] Gan et al. "Multitask Learning for Scalable and Dense Multilayer Bayesian Map Inference." *IEEE TRO* (2022).

Current TTM for Legged Robots

- Wellhausen *et al.* (2019)
  - Ground reaction score
- Faigl *et al.* (2019)
  - Maximum forward velocity, attitude stability
- Fan *et al.* (2021)
  - Risks from collision, step, slippage, etc.
- Manually defined traversability

➡ learn from demonstration



[1] Wellhausen et al. "Where should I walk? predicting terrain properties from images via self-supervised learning." *IEEE RAL* 4.2 (2019): 1509-1516.
[2] Faigl et al. "On unsupervised learning of traversal cost and terrain types identification using self-organizing maps." *ICANN*. Springer, Cham, 2019.
[3] Fan et al. " STEP: Stochastic traversability evaluation and planning for risk-aware off-road navigation" *RSS* (2021).

IRL-Based TTM for Autonomous Vehicles

- Wulfmeier *et al.* (2017)
  - Maximum Entropy Deep Inverse Reinforcement Learning (MEDIRL)
- Zhang *et al.* (2018), Jung *et al.* (2021)
  - MEDIRL + handcrafted kinematics/route plan
- when applied on legged robots can suffer from:
  - Low model fidelity due to the incapability of modeling more agile legged robot motion using handcrafted features
  - Suboptimality of demonstrations due to insufficient feedback

$$\hat{r}_\theta(s) = f(\phi(s);\theta)$$

$$\frac{\partial \mathcal{L}_\mathcal{D}}{\partial \boldsymbol{\theta}} = \frac{\partial \mathcal{L}_\mathcal{D}}{\partial r}\frac{\partial r}{\partial \boldsymbol{\theta}} = \underbrace{(\boldsymbol{\mu}_\mathcal{D} - \mathbb{E}[\boldsymbol{\mu}])}_{\text{State Visitation Matching}} \underbrace{\partial r(\boldsymbol{\theta})/\partial \boldsymbol{\theta}}_{\text{Backpropagation}}$$

[1] Wulfmeier et al. "Large-scale cost function learning for path planning using deep inverse reinforcement learning." *IJRR* 36.10 (2017): 1073-1087.
[2] Zhang et al. "Integrating kinematics and environment context into deep inverse reinforcement learning for predicting off-road vehicle trajectories." *CoRL* (2018).
[3] Jung et al. "Incorporating multi-context into the traversability map for urban autonomous driving using deep inverse reinforcement learning." *RAL* 6.2 (2021): 1662-1669.

Contributions

- We propose to incorporate robot proprioceptive (inertial) feature learning in an IRL framework for legged robots terrain traversability modeling
- We extend the MEDIRL framework into a Trajectory-ranked MEDIRL framework and use locomotion energy as trajectory preference label to alleviate suboptimality

Problem Formulation

- We model the process of legged robot walking on local terrain as an agent following a Markov Decision Process (MDP): $\mathcal{M} = \{\mathcal{S}, \mathcal{A}, \mathcal{T}, \gamma, r\}$
- A trajectory is defined as a sequence of state-action pairs followed by the agent:

$$\tau := \{(s_0, a_0), (s_1, a_1), ..., s_{|\tau|})\}$$

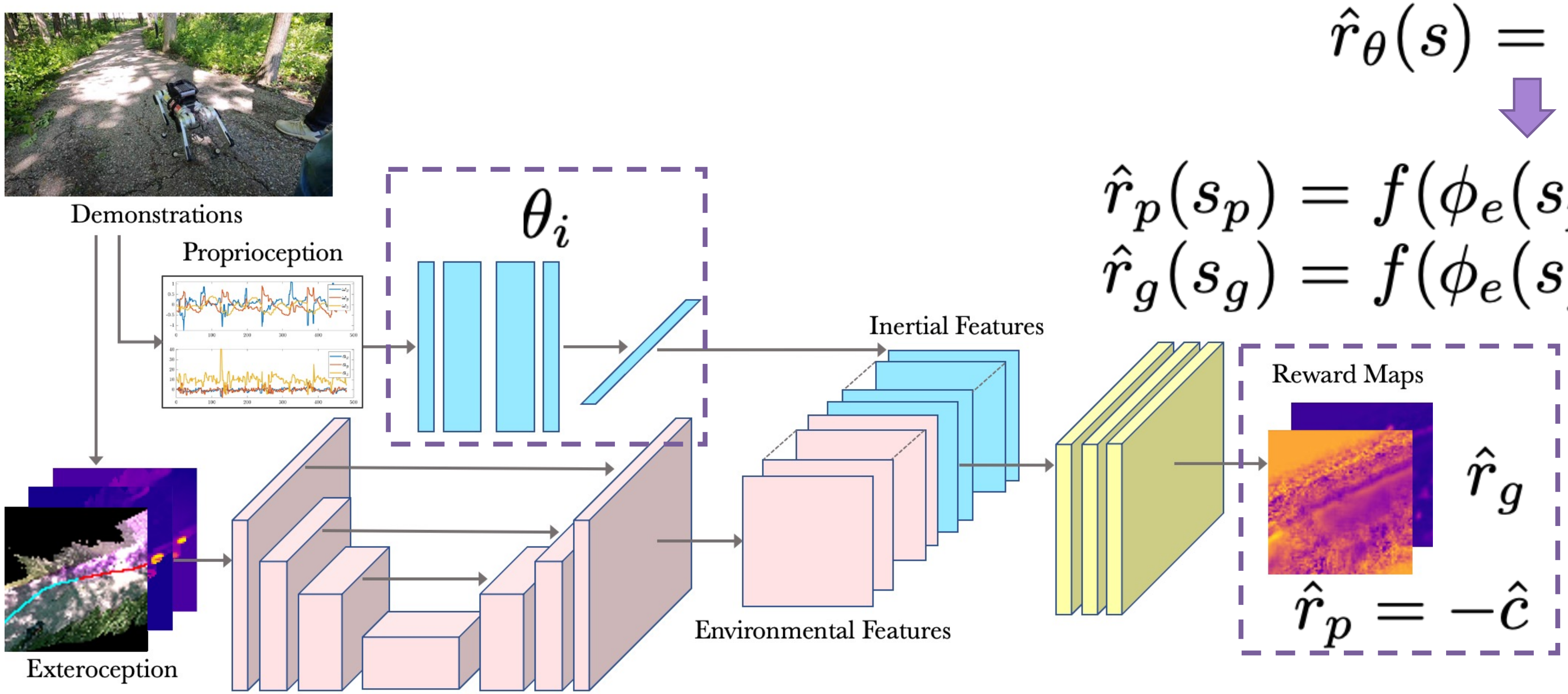
- A demonstration $\mathcal{D}$ is a set of τ collected from expert operation.
- IRL: Given $\mathcal{M}/r$, to recover the underlying r explaining $\mathcal{D}$.

Problem Formulation

- State space: $\mathcal{S} = \{\mathcal{S}_p, \mathcal{S}_g\}$
- Action space: $\mathcal{A} = \{up, down, left, right, end\}$
- Transition function: $\mathcal{T} : \mathcal{S}_p \times \mathcal{A} \to \mathcal{S}$
- Rewards
  - Path reward: $r_p : \mathcal{S}_p \to \mathbb{R}$
  - Goal reward: $r_g : \mathcal{S}_g \to \mathbb{R}$
- Traversability cost: $c = -r_p : \mathcal{S}_p \to \mathbb{R}$

Proposed Method

- Inertial Feature Learning with an Inertial Branch



$$\hat{r}_\theta(s) = f(\phi(s); \theta)$$

$$\hat{r}_p(s_p) = f(\phi_e(s_p), \phi_i(s_p; \theta_i); \theta_p)$$
$$\hat{r}_g(s_g) = f(\phi_e(s_g), \phi_i(s_g; \theta_i); \theta_g)$$

θ_p, θ_g share parameters

Proposed Method

- Locomotion Energy ranked Reward Extrapolation
  - Trajectory ranking loss [1]:

$$\mathcal{L}_{i,j} = -\sum_{\tau_i \prec \tau_j} \log \frac{\exp \sum_{s\in\tau_j} r_{\theta,j}(s)}{\exp \sum_{s\in\tau_i} r_{\theta,i}(s) + \exp \sum_{s\in\tau_j} r_{\theta,j}(s)},$$

  - $\tau_i \prec \tau_j$, if $e_{\tau_i} < e_{\tau_j}$, where e_τ is Average Energy Consumption (AEC), defined using joint torque $\boldsymbol{u}$ and joint displacement $\boldsymbol{q}$:

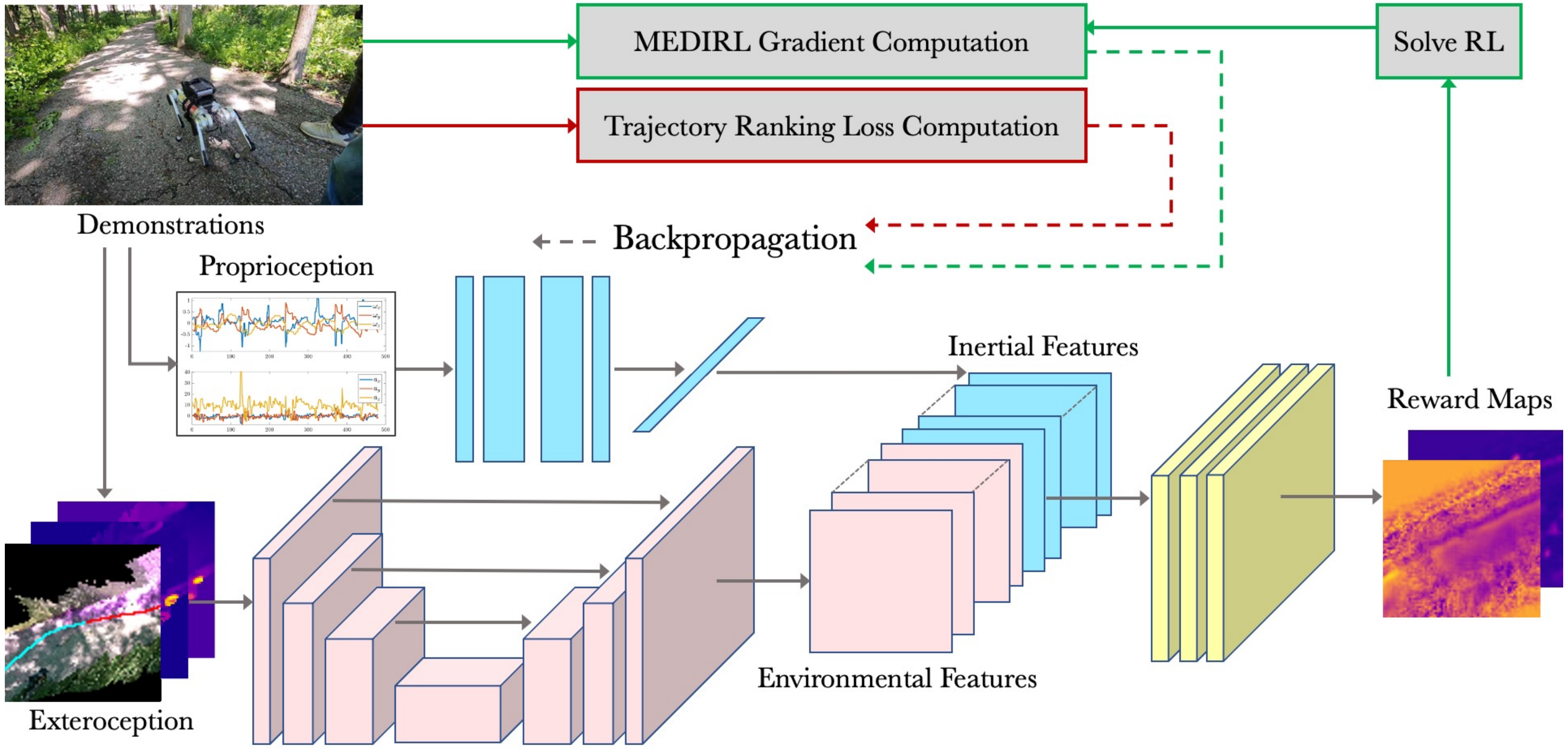
$$e_\tau = \sum_{i=1}^{n} \langle |\boldsymbol{u}_i|, |\Delta \boldsymbol{q}_i| \rangle,$$

  - Lower AEC ➡ higher trajectory rank ➡ higher return

[1] Brown et al. "Better-than-demonstrator imitation learning via automatically-ranked demonstrations." *CoRL*, 2020.

Proposed Method

- Overall framework



Experiments

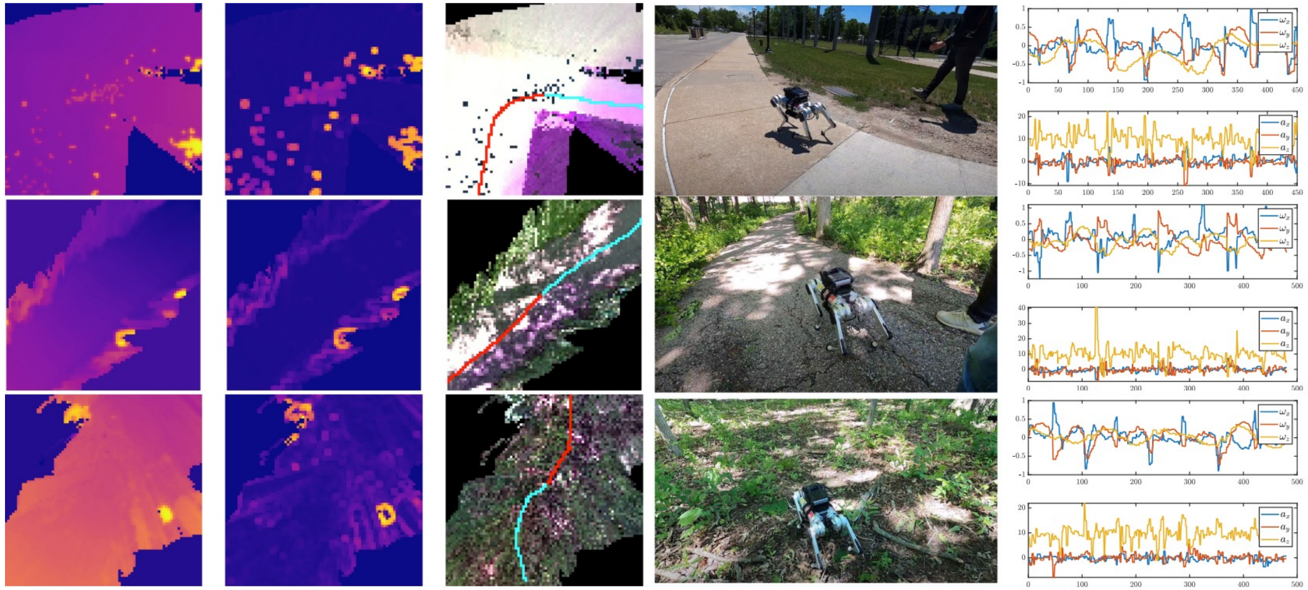
- Dataset Collection
  - Dataset is collected by expert operating a quadruped robot platform equipped with Intel RealSense depth camera, IMU and Nvidia Jetson Xavier on different types of terrains on campus

Mini-Cheetah Robot with Customized Sensor Suite

North Campus at University of Michigan

Experiments

- Dataset Generation
  - Elevation Map using Elevation Mapping [1]
  - Elevation Variance
  - Color Map by RGB averaging
  - Trajectory using ORBSLAM2 [2]
  - IMU raw signals



Elevation Elevation Var. Color 3rd Person View IMU Signals

[1] Fankhauser et al. "Probabilistic terrain mapping for mobile robots with uncertain localization." *IEEE RAL* 3.4 (2018): 3019-3026.
[2] Mur-Artal et al. "ORB-SLAM2: An open-source slam system for monocular, stereo, and RGB-D cameras." *IEEE TRO* 33.5 (2017): 1255-1262.

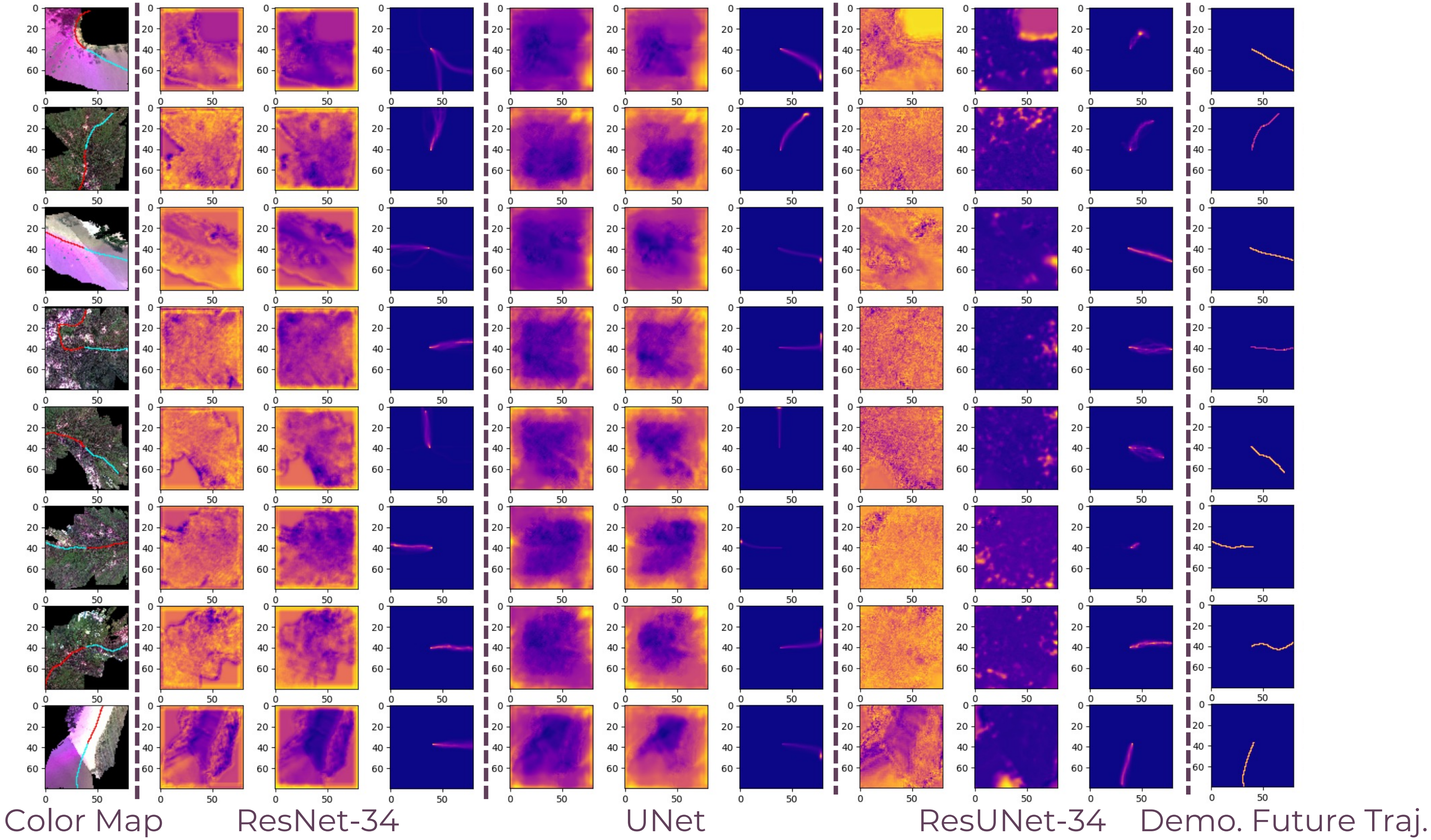
Environmental Branch Ablation

- IRL Metrics
  - Negative Log-Likelihood (NLL)
  - Hausdorff Distance (HD)

	NLL ↓	HD ↓
ResNet-34 [1]	0.9490	13.4957
UNet [2]	0.9016	12.0014
ResUNet-34 [3]	**0.8419**	**9.8219**

[1] Deo et al. "Trajectory forecasts in unknown environments conditioned on grid-based plans." arXiv preprint arXiv:2001.00735 (2020).
[2] Ronneberger et al. "U-Net: Convolutional networks for biomedical image segmentation." MICCAI. Springer, Cham, 2015.
[3] Zhang et al. "Road extraction by deep residual U-Net." IEEE GRSL 15.5 (2018): 749-753.

Environmental Branch Ablation



Inertial Feature Learning Evaluation

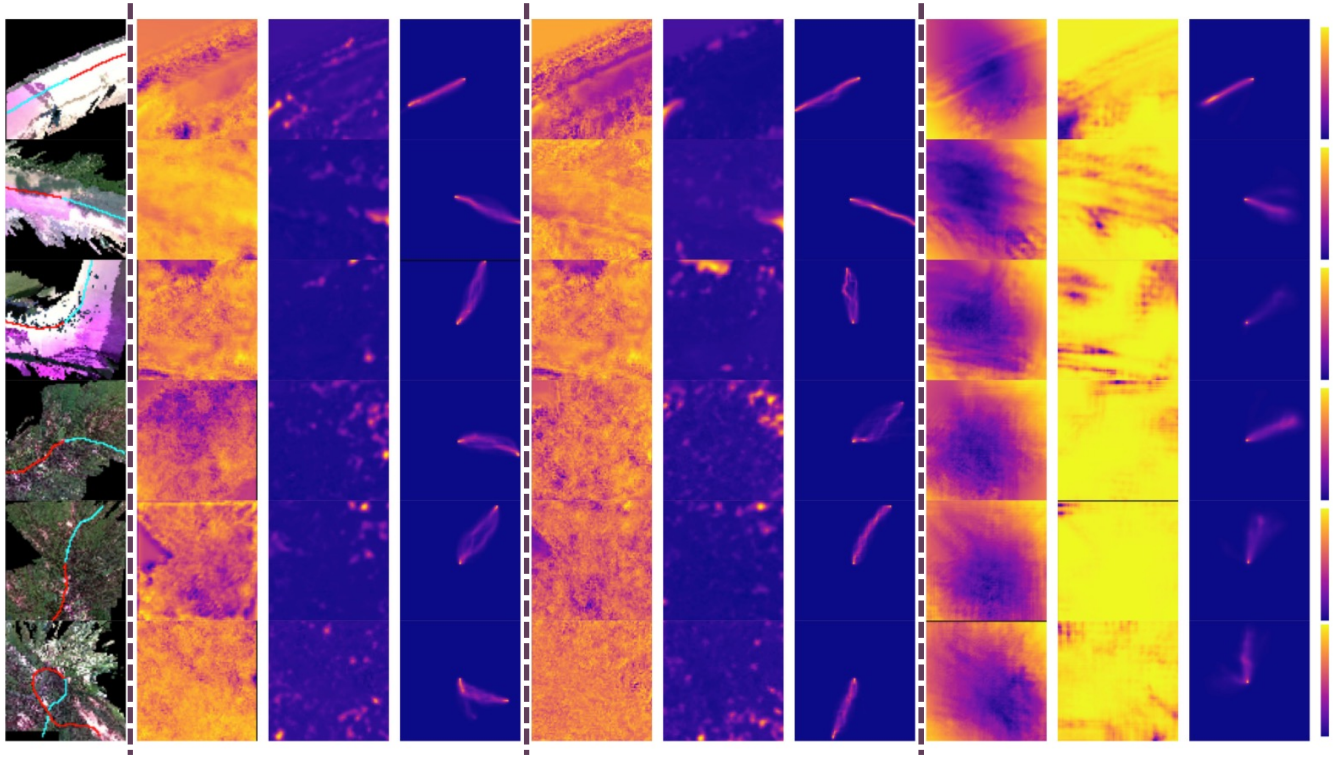
- Quantitative Results
  - Better performance shows better model fidelity

	NLL ↓	HD ↓
Ours w/o inertial feature learning	0.8419	9.8219
Baseline method using handcrafted kinematics [1]	0.8821	10.1953
Ours w/ inertial feature learning	**0.8419**	**9.8219**

[1] Zhang et al. "Integrating kinematics and environment context into deep inverse reinforcement learning for predicting off-road vehicle trajectories." *CoRL* (2018).

Inertial Feature Learning Evaluation

- Qualitative Results



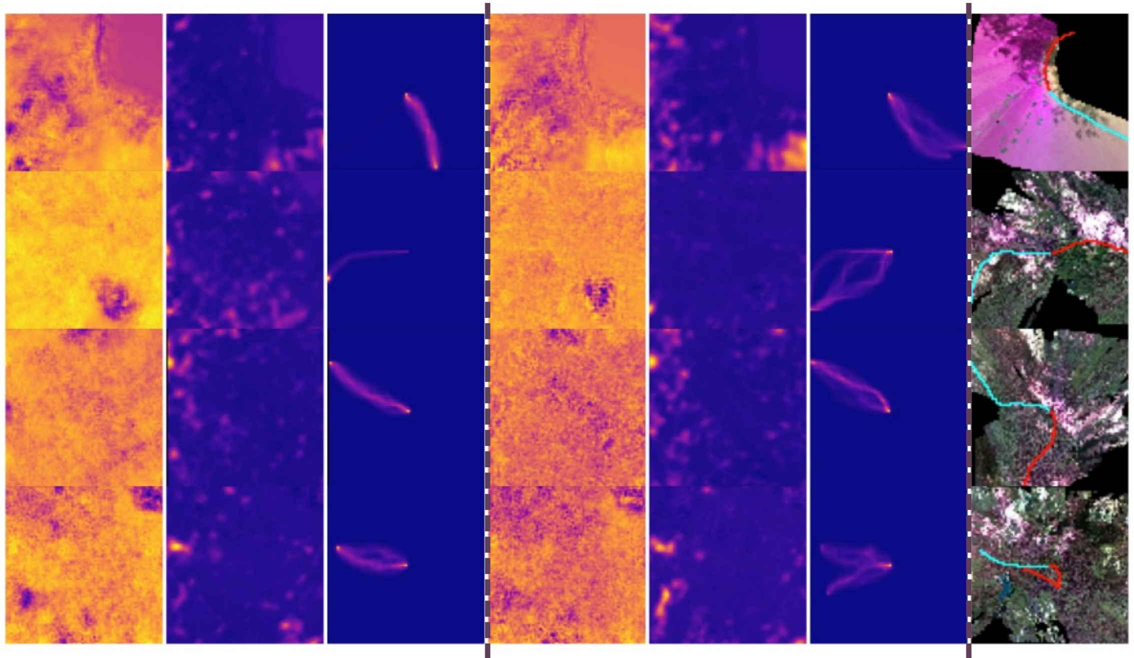
Color Map Ours w/ Inertial Ours w/o Inertial Baseline

Energy-Based Reward Extrapolating Evaluation

- We simulate Mini-Cheetah to follow the optimal trajectories from both methods on the input elevation map for evaluation.

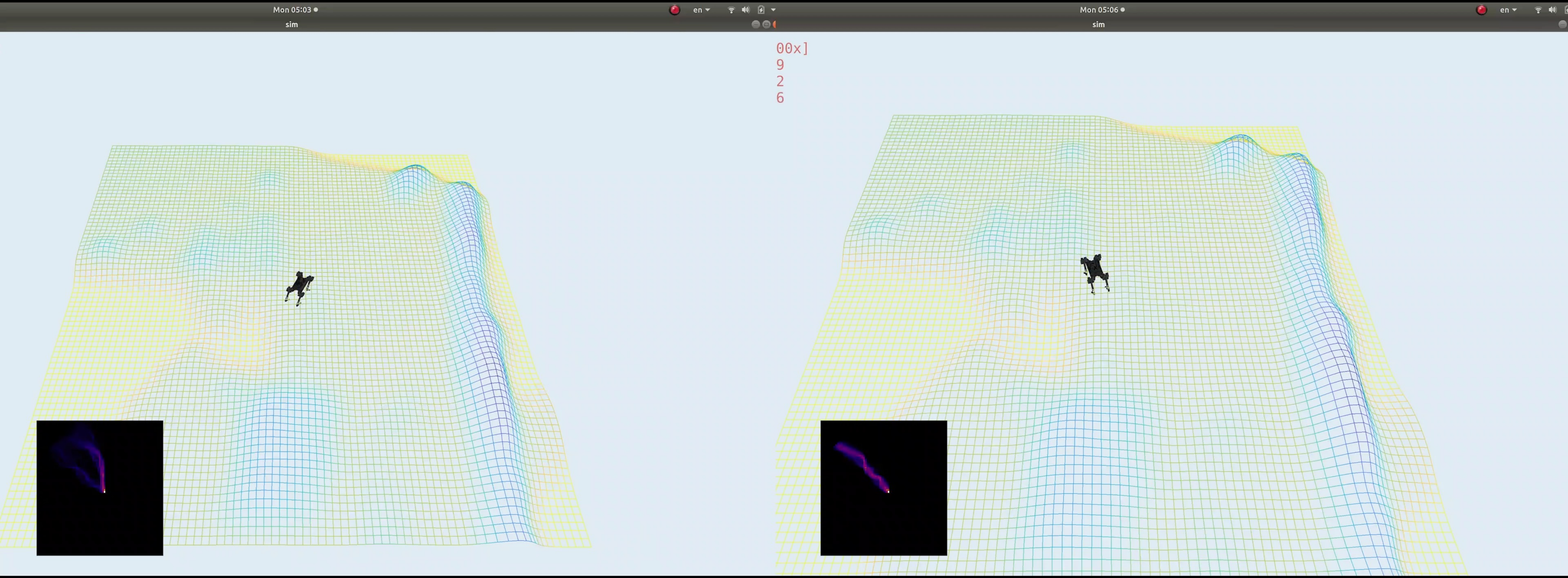
	MEDIRL	T-MEDIRL
NLL↓	**0.7734**	0.8132
HD↓	**8.1460**	9.9126
Accuracy↑	0.4001	**0.6412**
AEC↓	4.634e-2J	**4.179e-2J**

The difference in the simulated AEC corresponds to about 7 minutes extra operation.



MEDIRL T-MEDIRL Color Map

Experiment 2: Mini-Cheetah Simulator (Speed x2)

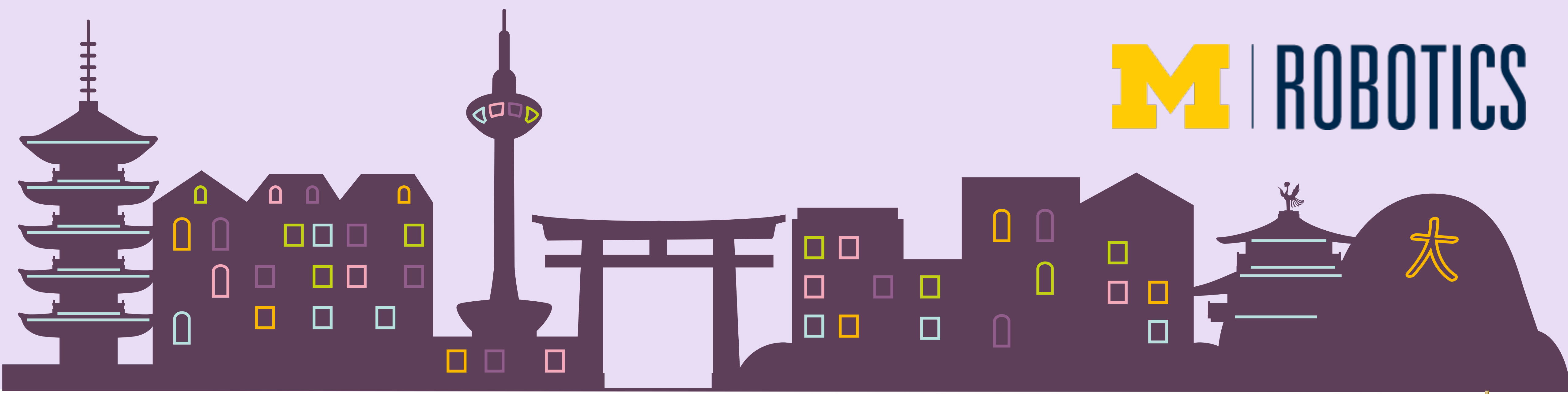


T-MEDIRL

MEDIRL

Limitations and Future Work

- The simplified discrete states and actions are unable to fully capture the agile motion capability of a legged robot.
- However, adding more dimensions comes with an exponential increase in computation complexity.
- Designing a hierarchical action and state space is an interesting future research direction.



TOYOTA RESEARCH INSTITUTE

NAVER LABS

NVIDIA

NSF

Q&A

ganlu@umich.edu

https://github.com/ganlumomo/minicheetah-traversability-irl

IEEE/RSJ International Conference on Intelligent Robots and Systems