

GRoQ-LoCO: Generalist and Robot-agnostic Quadruped Locomotion Control using Offline Datasets

Narayanan PP Sarvesh Prasanth Venkatesan Srinivas Kantha Reddy Shishir Kolathaya
 Indian Institute of Science, Bangalore, India
 pnkrishnan27@gmail.com, sarveshv219@gmail.com,
 srinivas299792458@gmail.com, shishirk@iisc.ac.in

Abstract: Recent advancements in large-scale offline training have demonstrated the potential of generalist policy learning for complex robotic tasks. However, applying these principles to legged locomotion remains a challenge due to continuous dynamics and the need for real-time adaptation across diverse terrains and robot morphologies. In this work, we propose GRoQ-LoCO, a scalable, attention-based framework that learns a single generalist locomotion policy across multiple quadruped robots and terrains, relying solely on offline datasets. Our approach leverages expert demonstrations from two distinct locomotion behaviors - stair traversal (non-periodic gaits) and flat terrain traversal (periodic gaits) - collected across multiple quadruped robots, to train a generalist model that enables behavior fusion. Crucially, our framework operates solely on proprioceptive data from all robots without incorporating any robot-specific encodings. The policy is directly deployable on an Intel i7 nuc, producing low-latency control outputs without any test-time optimization. Our extensive experiments demonstrate zero-shot transfer across highly diverse quadruped robots and terrains, including hardware deployment on the Unitree Go1, a commercially available 12kg robot. Notably, we evaluate challenging cross-robot training setups where different locomotion skills are unevenly distributed across robots, yet observe successful transfer of both flat walking and stair traversal behaviors to all robots at test time. We also show preliminary walking on Stoch 5, a 70kg quadruped, on flat and outdoor terrains without requiring any fine tuning. These results demonstrate the potential of offline, data-driven learning to generalize locomotion across diverse quadruped morphologies and behaviors.

Keywords: Behavior cloning, Quadrupeds, Generalization, Zero-shot-transfer, Offline dataset

1 Introduction

Generalization is a central challenge in legged locomotion control. Robust controllers must not only produce stable and efficient motions but also adapt to new terrains, disturbances, and robot designs without the need for retraining. Achieving such generalization would enable legged robots to move out of controlled labs and operate reliably in the real world.

In robotic manipulation, *large-scale offline learning* has enabled this kind of generalization. Models like RT-1 and RT-2 [1, 2], and efforts like Open X-Embodiment [3] have demonstrated how diverse, pre-collected datasets can be used to train generalist policies capable of solving hundreds of tasks across different robotic arms without online interaction. These methods rely on *multi-task imitation learning*, foundation models, and scalable architectures that unify diverse behaviors across embodiments [3, 4].



Figure 1: Frames of Go1 robot traversing 15cm stairs. The policy showed zero shot transfer to Go1 for flat terrains and staircases.

Bringing these offline learning principles to *legged locomotion*, however, remains underexplored. Locomotion differs fundamentally from manipulation: it requires *continuous control*, real-time adaptation to dynamic environments, and often lacks clearly segmented tasks or episodic resets. Still, the benefits of offline learning—scalability, safety, and reusability—make it a compelling direction for legged robotics, where online exploration is risky and expensive.

Some early efforts have begun applying offline learning to locomotion [5, 6, 7, 8]. *DiffuseLoco* [6] showed that diffusion models trained on demonstration data can learn diverse gait patterns and enable zero-shot sim-to-real transfer. However, these were confined to a *single robot morphology*, and the scope of behaviors was limited. Other sequence modeling approaches like Decision Transformers [4, 9, 10] have shown promise in manipulation but have not been widely adopted for locomotion especially in multi-embodiment settings.

In contrast, *deep reinforcement learning (RL)* has been the dominant paradigm for learning legged locomotion policies. It has enabled *agile behaviors* such as trotting, jumping, and terrain traversal [11, 12, 13, 14, 15, 16, 17]. However, RL methods typically require *large-scale online interaction*, *task-specific reward engineering*, and *carefully tuned simulation environments*. As a result, they often produce *specialized policies* that generalize poorly to new robot morphologies or unseen environments without extensive fine-tuning.

Several RL-based approaches have tried to address this by *explicitly modeling morphology variation* [18, 19]. *GenLoco* [18] introduced morphology randomization during RL training, enabling a single policy to generalize across different quadrupeds such as A1, Mini Cheetah, and Sirius but only for *velocity tracking on flat terrain*. *MorAL* [20] added an adaptive module to infer robot dynamics implicitly, improving generalization, but still relied on *online RL* and lacked the ability to capture *multiple distinct locomotion behaviors* within a single model.

We introduce *GRoQ-LoCO*, a scalable offline learning framework that unifies locomotion control across different terrains and robot designs. Our core insight is that **dataset diversity** both in *robot morphologies and locomotion behaviors* is essential for generalization. GRoQ-LoCO is trained on expert demonstrations of *periodic gaits (flat terrain)* and *non-periodic traversal (stairs)* collected from multiple quadruped robots. It operates directly on *proprioceptive inputs*, without any morphology encoding or post-training optimization. GRoQ-LoCO demonstrates *strong generalization* across both behaviors and embodiments.

The key contributions of our work include:

- **A Generalist Locomotion Controller:** We develop a single policy that controls multiple distinct quadrupedal robots without requiring robot-specific information.
- **Offline Multi-Behavior Learning:** We demonstrate that purely offline training on diverse motion data produces a policy with periodic gaits and multi-terrain traversability.
- **Zero-Shot Transfer and Robustness:** Our framework achieves strong zero-shot transfer across diverse quadruped robots and terrains, including hardware deployment on commercial platforms like the Unitree Go1 (see Fig. 1) and the Stoch 5, without requiring fine-tuning.

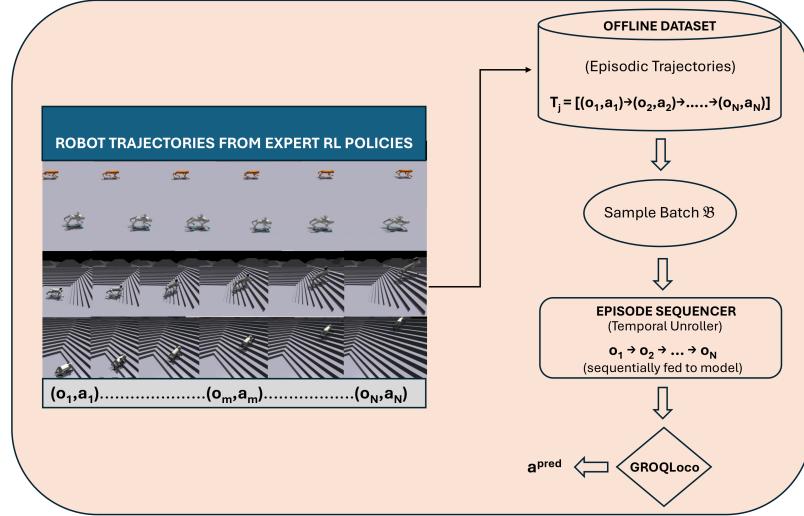


Figure 2: Offline data generation pipeline used in **GROQLoco**, illustrating trajectory collection from expert RL policies on diverse terrains and robot morphologies.

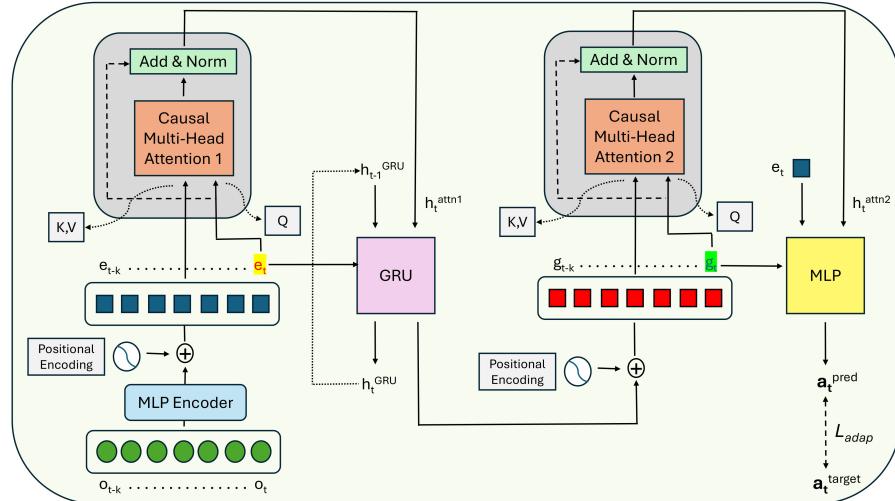


Figure 3: Model architecture of **GRoQ-LoCO**, showing the sequential processing pipeline with observation encoding, causal attention, GRU-based temporal modeling, and MLP action prediction.

2 Methodology

In this section, we present our data collection strategy, the architecture of the proposed behavior cloning policy, the adaptive loss formulation, and the training procedure. Our approach enables a single policy to learn different locomotion behaviors, namely, stable cyclic gaits on flat terrain and stair-climbing on rough terrain, across a diverse set of quadruped robots. We also outline the evaluation of zero-shot generalization to new robots unseen during training.

2.1 Data Collection and Expert Demonstrations

We collect expert demonstrations in simulation on select quadruped platforms, including Unitree B2, Go1, Aliengo, and Stoch3 (Refer Fig. 2). Additional platforms such as Unitree B1, X30, Lite3 and Stoch5 are included primarily for evaluating zero-shot generalization and cross-morphology transfer. Two specialized locomotion controllers are used to generate expert trajectories:

- **Flat-Ground Controller:** A periodic gait controller designed for stable and efficient locomotion on flat terrain.
- **Rough-Terrain Controller:** A non-periodic controller tailored for stair climbing and traversal of uneven terrain.

Each expert trajectory τ consists of a sequence of observations $\{o_0, o_1, \dots, o_T\}$ and corresponding expert actions $\{a_0, a_1, \dots, a_T\}$, where each observation o_t includes:

$$o_t = [q_t, \dot{q}_t, a_{t-1}, a_{t-2}, g_t, \omega_t, v_t^{\text{cmd}}]^T \quad (1)$$

Here, q_t denotes joint positions, \dot{q}_t joint velocities, a_{t-1}, a_{t-2} are previous actions, g_t the gravity-aligned vector, ω_t the angular velocity, and v_t^{cmd} the commanded linear and angular velocities (v_x, v_y, ω_z). No robot-specific identifiers are used; policies are trained purely from proprioceptive and command data to encourage generalization across robots.

2.2 Policy Architecture

Our behavior cloning policy is composed of four principal modules: an observation encoder, an attention-enhanced recurrent core, a secondary attention module over the recurrent history, and a final multi-layer perceptron (MLP) for action prediction (Fig. 3).

Observation Encoder: At each timestep t , the observation $o_t \in \mathbb{R}^{d_{\text{obs}}}$ is embedded into a latent space via:

$$\mathbf{e}_t = \text{LayerNorm}(\text{ELU}(\mathbf{W}_e \mathbf{o}_t + \mathbf{b}_e)),$$

where $\mathbf{W}_e, \mathbf{b}_e$ are learnable parameters.

Positional Embeddings To inject temporal ordering information without assuming a fixed history size, we employ fixed sinusoidal positional embeddings:

$$PE(t, 2i) = \sin\left(\frac{t}{10000^{2i/d_{\text{emb}}}}\right), \quad PE(t, 2i+1) = \cos\left(\frac{t}{10000^{2i/d_{\text{emb}}}}\right)$$

These embeddings are added to observation and GRU histories before attention operations.

Attention over Observation History: The encoded observations $\{e_{t-k}, \dots, e_t\}$ are stacked with positional encoding and processed by a multi-head attention layer:

$$\mathbf{h}_t^{\text{attn1}} = \text{MHA}_{\text{attn1}}([e_{t-k} + PE, \dots, e_t + PE])$$

where MHA denotes multi-head self-attention. The most recent k outputs are aggregated.

GRU Memory: The encoded observation \mathbf{e}_t is concatenated with the attended context $\mathbf{h}_t^{\text{attn1}}$ and passed into a GRU:

$$\mathbf{g}_t, \mathbf{h}_t^{\text{GRU}} = \text{GRU}([\mathbf{e}_t; \mathbf{h}_t^{\text{attn1}}], \mathbf{h}_{t-1}^{\text{GRU}}).$$

Attention over GRU History: The GRU outputs over time $\{\mathbf{g}_{t-k}, \dots, \mathbf{g}_t\}$ are again stacked, positional embeddings added, and processed through a second multi-head attention module:

$$\mathbf{h}_t^{\text{attn2}} = \text{MHA}([\mathbf{g}_{t-k} + PE, \dots, \mathbf{g}_t + PE])$$

Action Head: The final action is computed by feeding the concatenation of \mathbf{e}_t , \mathbf{g}_t , and $\mathbf{h}_t^{\text{attn2}}$ into an MLP:

$$\mathbf{a}_t = \text{MLP}([\mathbf{e}_t; \mathbf{g}_t; \mathbf{h}_t^{\text{attn2}}]).$$

2.3 Adaptive Loss for Behavior Cloning

An adaptive loss is employed instead of a standard MSE. Given the predicted action $\hat{\mathbf{a}}_t$ and expert action \mathbf{a}_t , the loss is defined as:

$$\mathcal{L}_{\text{adaptive}} = \frac{1}{T} \sum_{t=1}^T \left(\exp(-\log \sigma) \cdot \delta^2 \log \left(1 + \left(\frac{\hat{\mathbf{a}}_t - \mathbf{a}_t}{\delta} \right)^2 \right) + \log \sigma \right),$$

where σ is a learnable parameter per action dimension, and δ is a fixed scaling hyperparameter (e.g., $\delta = 0.5$). T denotes the number of timesteps within each truncated sequence (e.g., $T = 20$ during training with truncated BPTT). This loss behaves like a Huber loss with adaptive weighting, allowing important joints to have higher influence during training.

Training is performed using batches of size 400, with truncated BPTT over recurrent states. Hidden states are detached periodically to prevent backpropagation through arbitrarily long sequences. Additional details are present in Appendix C.

2.4 Training Setup and Details

Let $\mathcal{D} = \{\tau_i\}_{i=1}^M$ denote a dataset of M expert trajectories, where

$$\tau_i = \{(\mathbf{o}_{i,1}, \mathbf{a}_{i,1}), \dots, (\mathbf{o}_{i,N_i}, \mathbf{a}_{i,N_i})\}$$

and N_i is the length of episode i . Define $N_{\max} = \max_i N_i$. Each τ_i is padded to length N_{\max} with dummy zero vectors and a mask $\mathbf{m}_{i,t} \in \{0, 1\}$ marking valid timesteps.

Training runs for E epochs. In our setup, one epoch corresponds to sampling a single batch (not a full pass over \mathcal{D}). At epoch e , we sample a batch B and extract the padded sequence:

$$(\mathbf{o}_t^{(B)}, \mathbf{a}_t^{(B)}, \mathbf{m}_t^{(B)}), \quad t = 1, \dots, N_{\max}$$

We process each batch sequentially over time, passing $\mathbf{o}_t^{(B)}$ and previous hidden state $\mathbf{h}_{t-1}^{(B)}$ to the model, which outputs predicted action and updated hidden state:

$$\hat{\mathbf{a}}_t^{(B)}, \mathbf{h}_t^{(B)} = \text{Model}(\mathbf{o}_t^{(B)}, \mathbf{h}_{t-1}^{(B)})$$

The loss is computed over valid steps:

$$\mathcal{L}_B = \frac{1}{\sum_t m_t^{(B)}} \sum_{t=1}^{N_{\max}} m_t^{(B)} \ell_{\text{adaptive}}(\hat{\mathbf{a}}_t^{(B)}, \mathbf{a}_t^{(B)})$$

Every T_u steps (e.g., $T_u = 20$), we compute gradients:

$$\nabla_{\theta} \left(\frac{1}{b} \sum_B \mathcal{L}_B \right)$$

(where b is the batch size), update parameters using Adam, and truncate BPTT by detaching hidden states:

$$\mathbf{h}_t^{(b)} \leftarrow \text{stop_grad}(\mathbf{h}_t^{(b)}) \quad \forall t \bmod T_u = 0$$

To stabilize training, we apply a warmup of $E_w = 50$ epochs where, post-update, all hidden states are reset:

$$\mathbf{h}_t^{(b)} \leftarrow \mathbf{0} \quad \text{if } e \leq E_w$$

For $e > E_w$, hidden states are preserved, enabling continuity across (padded) episodes. This balances TBPTT efficiency with long-horizon memory retention.

3 Experiments

We conduct a series of experiments to evaluate how multi-robot and multi-terrain training enables generalist locomotion policies that scale across diverse quadruped embodiments and terrain types. Our setup includes cross-robot training configurations with locomotion skills unevenly distributed across robots. We first study zero-shot transfer and behavior fusion in stair-climbing scenarios, using gait visualizations and step-wise completion tables to capture detailed behavior. We then examine generalization to entirely novel terrains. Additional analyses, including comparisons with and without explicit robot encodings under identical multi-robot settings, will be provided in the supplementary material.

Setting	Robot	Mode	13 cm	17 cm	21 cm (OOD)	25 cm (OOD)	29 cm (OOD)
1	Go1	ZS	✓	✓	✗	✗	✗
	Stoch5	ZS	✓	✓	✓	✓	✓
	B1	ZS	✓	✓	✓	✓	✓
	B2	FO	✓	✓	✓	✓	✓
	Aliengo	SO	✓	✓	✓	✗	✗
	Stoch3	SO	✓	✓	✓	✓	✓
2	Go1	FO	✓	✓	✗	✗	✗
	Stoch5	ZS	✓	✓	✓	✗	✗
	B1	ZS	✓	✓	✓	✓	✓
	B2	SO	✓	✓	✓	✓	✓
	Aliengo	ZS	✓	✓	✗	✗	✗
	Stoch3	FO	✓	✓	✓	✓	✗
3	Go1	FO	✓	✓	✗	✗	✗
	Stoch5	ZS	✓	✓	✓	✓	✓
	B1	ZS	✓	✓	✓	✓	✓
	B2	ZS	✓	✓	✓	✓	✓
	Aliengo	SO	✓	✓	✓	✗	✗
	Stoch3	SO	✓	✓	✓	✓	✓
4	Go1	SO	✓	✓	✗	✗	✗
	Stoch5	ZS	✓	✓	✗	✗	✗
	B1	ZS	✓	✓	✗	✗	✗
	B2	ZS	✓	✗	✗	✗	✗
	Aliengo	ZS	✓	✓	✗	✗	✗
	Stoch3	FO	✓	✓	✗	✗	✗
5	Go1	ZS	✓	✓	✓	✗	✗
	Stoch5	ZS	✓	✓	✓	✓	✓
	B1	ZS	✓	✓	✓	✓	✓
	B2	SO	✓	✓	✓	✓	✓
	Aliengo	SO	✓	✓	✓	✓	✗
	Stoch3	SO	✓	✓	✓	✓	✓

Table 1: Evaluation on stair environments with increasing difficulty (13–29 cm step heights). A checkmark (✓) indicates successful climbing of 8 stairs, and a cross (✗) indicates failure.

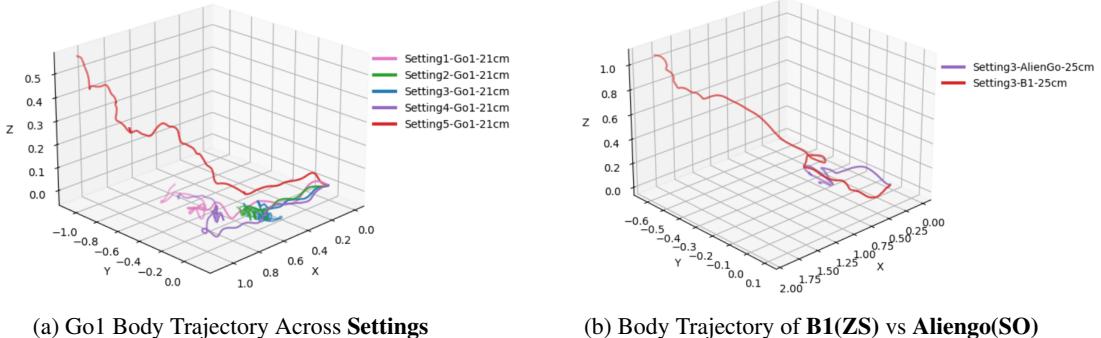


Figure 4

3.1 Cross-Robot and Cross-Terrain Generalization Analysis

We construct multiple training regimes using combinations of the two locomotion behaviors (periodic gaits and stair traversal) across different robot morphologies. Some configurations include data from both flat and stair policies, while others restrict terrain or robot access to examine generalization. Our goal is to understand how skill and morphology diversity in training data influences a policy’s ability to (a) Generalize to unseen robots (cross-morphology). (b) Transfer learned skills

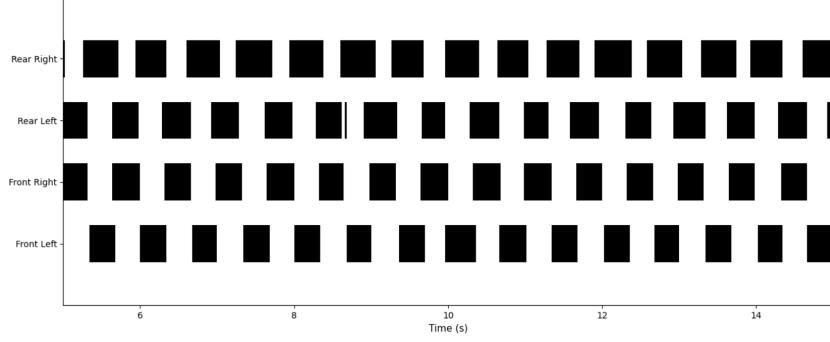


Figure 5: Go1(ZS) Foot contact sequence on Flat terrain- black regions are periods of foot contact

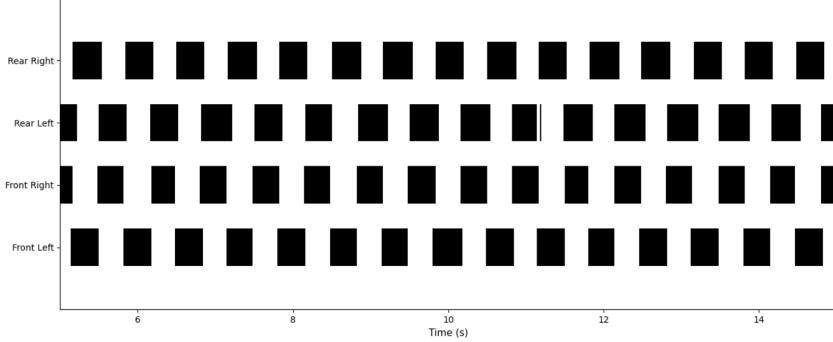


Figure 6: Stoch3(SO) Foot contact sequence on Flat terrain- black regions are periods of foot contact

(e.g., from flat to stair). **(c)** Acquire and retain multiple locomotion behaviors simultaneously. **(d)** Handle increasingly difficult Out-of-Distribution (OOD) terrain such as higher stairs.

Our training configurations, each involving a different subset of robots and terrain skills (flat and/or stair). For each configuration, robots are categorized as follows:

- **Zero-Shot (ZS):** The robot is entirely unseen during training (neither flat nor stair).
- **Flat Only (FO):** The robot contributed only *flat terrain* data during training.
- **Stair Only (SO):** The robot was included in training with *stair terrain* data.

Table 1 summarizes the results. Flat walking evaluations for the same policies are discussed later in this section to investigate transfer of cyclic motion. All experiments are conducted with a commanded forward velocity of 1 m/s along the x -axis.

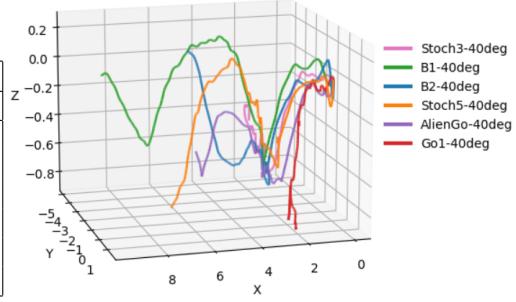
Experiment Settings. We consider five distinct data distribution settings for training:

- **Setting 1:** Flat-terrain data from B2; stair-climbing data from Aliengo and Stoch3.
- **Setting 2:** Flat-terrain data from Go1 and Stoch3; stair-climbing data from B2.
- **Setting 3:** Flat-terrain data from Go1; stair-climbing data from Aliengo and Stoch3.
- **Setting 4:** Flat-terrain data from Stoch3; stair-climbing data from Go1.
- **Setting 5:** Both flat terrain and stair-climbing data from Aliengo, B2, and Stoch3.

Stair Climbing Generalization Analysis. We draw four key insights from the results in Table 1, supported by base trajectories and gait visualizations across robots.

Robot	Smooth Slopes			Rough Slopes		
	25°	30°	40°	25°	30°	40°
Go1	✓	✗	✗	✓	✗	✗
Stoch5	✓	✓	✓	✓	✓	✓
Aliengo	✓	✓	✓	✓	✓	✓
Stoch3	✓	✓	✗	✓	✓	✗
B2	✓	✓	✓	✓	✓	✓
B1	✓	✓	✓	✓	✓	✓

(a) Zero-shot slope traversal results.



(b) Base trajectory on 40° rough slope.

Figure 7: Binary performance assessment and visual evaluation of zero-shot generalization to novel slopes.

1. **Full Diversity Training Enables Strong Zero-Shot Transfer.** Setting 5, which includes flat and stair data from three diverse robots, yields the best zero-shot (ZS) generalization to unseen embodiments and out-of-distribution (OOD) stairs. For example, Go1 in Setting 5 climbs 21 cm stairs ZS, whereas it fails at the same height in Setting 2 as depicted in the Figure 4a. Compared to Setting 4, where all ZS robots fail beyond 17 cm, Setting 5 shows clear cross-embodiment transfer.
2. **Stair-Trained Policies Generalize Beyond Training Range.** Stair-only (SO) robots generalize to OOD stairs beyond their training limit of 17 cm. In Setting 5, B2 and Stoch3 successfully traverse 25 cm and 29 cm stairs, showing strong terrain extrapolation.
3. **Zero-Shot Robots Can Outperform Stair Specialists.** In some cases, ZS robots surpass SO-trained ones. For instance, in Setting 3, B1 (ZS) climbs 29 cm stairs, while Aliengo (SO) fails. Figure 4b suggests more adaptive motions in ZS policies due to morphology-driven robustness.
4. **Flat-Only Policies Exhibit Stair Climbing Generalization.** Flat-only (FO) policies show generalization to stairs. In Setting 2, Stoch3 (FO) climbs 25 cm OOD stairs despite no elevation exposure during training, indicating transferable skills like stable gait and foot placement.
5. **Emergence of Cyclic Gaits Across Robots.** Gait plots (Figures 5, 6) show structured, cyclic patterns in both ZS (Go1) and SO (Stoch3) policies across terrains.

3.2 Generalization to Novel Terrains

We evaluate zero-shot generalization of our locomotion policies to smooth and rough inclined slopes at angles of 25°, 30°, and 40°, none of which appeared in training. All robots are deployed in a fully zero-shot setting (*unseen robot × terrain*). Table in Figure 7a summarizes binary success (✓) or failure (✗) on each slope.

Consistent High-Performers. Stoch5, Aliengo, B2, and B1 succeed on all smooth and rough slopes up to 40°, demonstrating exceptionally robust zero-shot slope traversal across morphologies and terrain irregularities.

Minimal Impact of Roughness for Robust Policies. For high-performing robots, undulating terrain (see Fig. 7b) does not degrade performance compared to smooth slopes. This suggests that our policy architecture and particularly the cross-robot training captures slope-invariant locomotion strategies.

Zero-Shot Emergence of Adaptive Behavior. Despite no slope data during training, policies exhibit adaptive base movement and posture control on inclines. Base trajectory plots (Figure 7b) reveal smooth and progressive elevation changes, indicating stable and coordinated climbing behavior in a zero-shot setting.

Hardware deployment. We deployed a cloned policy on the Unitree Go1 and Stoch5. Go1 policy (Setting 5 in Table 1) demonstrated a zero-shot transfer for both flat-ground and staircases (15cm height). Stoch 5 policy demonstrated robust walking on flat-ground and slopes. Detailed results will be provided in the supplementary material.

4 Conclusion and Limitations

Currently, GRoQ-LoCO is focused on robots with comparable kinematic setups. Extending the framework to quadrupeds with more diverse morphologies such as those with different leg proportions or joint arrangements remains an exciting direction for future work. Furthermore, while our current system is based on proprioceptive feedback, integrating exteroceptive inputs like vision could allow the policy to become more visually aware, enhancing its ability to navigate complex environments and adapt to dynamic terrain. Another area for exploration involves expanding the approach beyond quadruped robots to other types of legged robots, such as hexapods or bipeds, where the dynamics of locomotion may present new challenges. Addressing these aspects will further enhance the versatility and generalization capabilities of our locomotion policies.

Acknowledgments

This research is funded by AI & Robotics Technology Park (ARTPARK), India

References

- [1] A. Brohan et al. Rt-1: Robotics transformer for real-world control at scale. *arXiv preprint arXiv:2212.06817*, 2022. URL <https://arxiv.org/abs/2212.06817>.
- [2] A. Brohan et al. Rt-2: Vision-language-action models transfer web knowledge to robotic control. *arXiv preprint arXiv:2307.15818*, 2023. URL <https://arxiv.org/abs/2307.15818>.
- [3] Open X-Embodiment Collaboration. Open x-embodiment: Robotic learning datasets and rt-x models. In *Proceedings of the 2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 6892–6903. IEEE, 2024. doi:[10.1109/ICRA57147.2024.10611477](https://doi.org/10.1109/ICRA57147.2024.10611477). URL <https://arxiv.org/abs/2310.08864>.
- [4] K. Bousmalis, G. Vezzani, D. Rao, C. M. Devin, A. X. Lee, M. B. Villalonga, T. Davchev, Y. Zhou, A. Gupta, A. Raju, A. Laurens, C. Fantacci, V. Dalibard, M. Zambelli, M. F. Martins, R. Peceviciute, M. Blokzijl, M. Denil, N. Batchelor, T. Lampe, E. Parisotto, K. ona, S. Reed, S. G. Colmenarejo, J. Scholz, A. Abdolmaleki, O. Groth, J.-B. Regli, O. Sushkov, T. Rothr, J. E. Chen, Y. Aytar, D. Barker, J. Ortiz, M. Riedmiller, J. T. Springenberg, R. Hadsell, F. Nori, and N. Heess. Robocat: A self-improving generalist agent for robotic manipulation. *Transactions on Machine Learning Research*, 2024. URL <https://openreview.net/forum?id=vsCpILiWHu>. Accepted.
- [5] A. Reske, J. Carius, Y. Ma, F. Farshidian, and M. Hutter. Imitation learning from mpc for quadrupedal multi-gait control. In *Proceedings of the 2021 IEEE International Conference on Robotics and Automation (ICRA)*, pages 5014–5020. IEEE, 2021. doi:[10.1109/ICRA48506.2021.9561214](https://doi.org/10.1109/ICRA48506.2021.9561214). URL <https://arxiv.org/abs/2103.14331>.
- [6] X. Huang, Y. Chi, R. Wang, Z. Li, X. B. Peng, S. Shao, B. Nikolic, and K. Sreenath. Diffuse-loco: Real-time legged locomotion control with diffusion from offline datasets. In P. Agrawal, O. Kroemer, and W. Burgard, editors, *Proceedings of The 8th Conference on Robot Learning*, volume 270 of *Proceedings of Machine Learning Research*, pages 1567–1589. PMLR, 06–09 Nov 2025. URL <https://proceedings.mlr.press/v270/huang25a.html>.
- [7] G. Mothish, M. Tayal, and S. Kolathaya. Birodiff: Diffusion policies for bipedal robot locomotion on unseen terrains. In *Proceedings of the 10th Indian Control Conference (ICC)*, pages

- 385–390. IEEE, 2024. doi:10.1109/icc64753.2024.10883743. URL <https://arxiv.org/abs/2407.05424>.
- [8] R. O’Mahoney, A. L. Mitchell, W. Yu, I. Posner, and I. Havoutis. Offline adaptation of quadruped locomotion using diffusion models. *arXiv preprint arXiv:2411.08832*, 2024. URL <https://arxiv.org/abs/2411.08832>.
 - [9] . Gajewski et al. Solving multi-goal robotic tasks with decision transformer. In *arXiv*, 2024. URL <https://arxiv.org/abs/2410.06347>.
 - [10] W. Dong et al. Optimizing robotic manipulation with decision-rwkv: A recurrent sequence modeling approach for lifelong learning. In *Journal of Computing and Information Science in Engineering*, 2025. URL <https://asmedigitalcollection.asme.org/computingengineering/article/25/3/031004/1210989>.
 - [11] J. Hwangbo, J. Lee, A. Dosovitskiy, D. Bellicoso, V. Tsounis, V. Koltun, and M. Hutter. Learning agile and dynamic motor skills for legged robots. *Science Robotics*, 4(26):eaau5872, 2019. doi:10.1126/scirobotics.aau5872. URL <https://www.science.org/doi/10.1126/scirobotics.aau5872>.
 - [12] A. Kumar, Z. Fu, D. Pathak, and J. Malik. Rma: Rapid motor adaptation for legged robots. In *Proceedings of Robotics: Science and Systems (RSS)*, 2021. doi:10.15607/RSS.2021.XVII.011. URL <https://arxiv.org/abs/2107.04034>.
 - [13] Y. Xue et al. Learning vision-guided quadrupedal locomotion end-to-end with a locomotion transformer. In *Proceedings of the 2021 Conference on Robot Learning*, 2021. URL <https://arxiv.org/abs/2107.03996>.
 - [14] D. Hoeller et al. Anymal parkour: Learning agile navigation for quadrupedal robots. In *Proceedings of the 7th Conference on Robot Learning*, 2023. URL <https://arxiv.org/abs/2306.14874>.
 - [15] Y. Wang et al. Amp in the wild: Learning robust, agile, natural legged locomotion skills. *arXiv preprint arXiv:2304.10888*, 2023. URL <https://arxiv.org/abs/2304.10888>.
 - [16] J. Long, Z. Wang, Q. Li, L. Cao, J. Gao, and J. Pang. Hybrid internal model: Learning agile legged locomotion with simulated robot response. In *Proceedings of the Twelfth International Conference on Learning Representations (ICLR)*, 2024. URL <https://openreview.net/forum?id=93LoCyww8o>.
 - [17] G. Kim, Y.-H. Lee, and H.-W. Park. A learning framework for diverse legged robot locomotion using barrier-based style rewards. *arXiv preprint arXiv:2409.15780*, 2024. URL <https://arxiv.org/abs/2409.15780>.
 - [18] G. Feng, H. Zhang, Z. Li, X. B. Peng, B. Basireddy, L. Yue, Z. Song, L. Yang, Y. Liu, K. Sreenath, and S. Levine. Genloco: Generalized locomotion controllers for quadrupedal robots. In L. P. Kaelbling, D. Kragic, and K. Fragkiadaki, editors, *Proceedings of the 6th Conference on Robot Learning (CoRL)*, volume 205 of *Proceedings of Machine Learning Research*, pages 1893–1903. PMLR, 2022. URL <https://proceedings.mlr.press/v205/feng23a.html>.
 - [19] N. Bohlinger, G. Czechmanowski, M. Krupka, P. Kicki, K. Walas, J. Peters, and D. Tateo. One policy to run them all: an end-to-end learning approach to multi-embodiment locomotion. In *Proceedings of the 8th Conference on Robot Learning (CoRL)*, 2024. URL <https://arxiv.org/abs/2409.06366>.
 - [20] Z. Luo, X. Li, R. Huang, Z. Shu, E. Xiao, and Y. Dong. Moral: Learning morphologically adaptive locomotion controller for quadrupedal robots on challenging terrains. *IEEE Robotics and Automation Letters*, 9(5):4019–4026, 2024. doi:10.1109/lra.2024.3375086. URL <https://ieeexplore.ieee.org/document/10463132>.

Appendix

A Extended Results section

We report additional zero-shot results beyond those presented in the main paper. The physical parameters of all quadruped robots, including those used in the main paper, are provided in Table 2. Table 3 reports zero-shot results on stair environments using two new quadruped robots: Lite3 and X30. Table 4 further shows zero-shot performance on slope terrains (smooth and rough) across robots.

All experiments are conducted with a commanded forward velocity of 1 m/s along the x -axis.

Parameter	A1	Go1	Aliengo	Stoch3	B1	B2	Stoch5	Lite	X30
Total weight (kg)	12	13	21	25	50	60	70	13	56
Base length (m)	0.40	0.38	0.65	0.54	0.92	0.80	0.67	0.53	0.90
Base width (m)	0.19	0.16	0.15	0.20	0.24	0.24	0.26	0.20	0.30
Height, fully standing (m)	0.40	0.40	0.48	0.50	0.63	0.64	0.55	0.40	0.47
Thigh Length (m)	0.20	0.22	0.26	0.30	0.35	0.35	0.35	0.2	0.3
Calf Length (m)	0.20	0.22	0.26	0.35	0.35	0.35	0.35	0.21	0.31

Table 2: Comparison of quadruped robot parameters

Setting	Robot	Mode	13 cm	17 cm	21 cm (OOD)	25 cm (OOD)	29 cm (OOD)
1	Lite3	ZS	✓	✓	✗	✗	✗
	X30	ZS	✓	✓	✓	✓	✗
2	Lite3	ZS	✗	✗	✗	✗	✗
	X30	ZS	✓	✓	✓	✗	✗
3	Lite3	ZS	✓	✗	✗	✗	✗
	X30	ZS	✓	✓	✓	✗	✗
4	Lite3	ZS	✓	✓	✓	✗	✗
	X30	ZS	✓	✓	✗	✗	✗
5	Lite3	ZS	✓	✓	✗	✗	✗
	X30	ZS	✓	✓	✓	✓	✗

Table 3: Zero-Shot evaluation on stair environments for Lite3 and X30 robots.

Robot	Smooth Slopes			Rough Slopes		
	25	30	40	25	30	40
Lite3	✓	✓	✗	✓	✓	✗
X30	✓	✓	✓	✓	✓	✓

Table 4: Zero-shot slope traversal results.

B Robot Encoding vs No Robot Encoding

The main paper presents five cross-robot training settings involving various combinations of flat-terrain and stair-climbing data. We provide extended analysis for **Setting 1** (flat data from B2; stair data from Aliengo and Stoch3), comparing models trained with explicit robot encodings under otherwise identical conditions. Table 5 shows the stair-climbing results using policies trained with explicit robot encodings.

We use robot encodings as part of observation input to the model, derived from predefined metadata as shown in the table 2. However, we observe that explicit robot encoding does not lead to reliable generalization. Specifically, robots like **Go1** and **Lite3**, whose embeddings are significantly different from those seen during training, fail to exhibit meaningful behavior. These robots attempt to walk but immediately collapse with erratic actions, indicating poor transfer to out-of-distribution embeddings.

Setting	Robot	Mode	13 cm	17 cm	21 cm (OOD)	25 cm (OOD)	29 cm (OOD)
1	Go1	ZS	✗	✗	✗	✗	✗
	Stoch5	ZS	✓	✓	✓	✗	✗
	B1	ZS	✓	✓	✓	✗	✗
	B2	FO	✓	✓	✓	✗	✗
	Aliengo	SO	✓	✓	✓	✗	✗
	Stoch3	SO	✓	✓	✓	✗	✗
	Lite3	ZS	✗	✗	✗	✗	✗
	X30	ZS	✓	✓	✓	✗	✗

Table 5: Evaluation on stair environments with increasing difficulty (1329 cm step heights). A checkmark (✓) indicates successful climbing of 8 stairs, and a cross (✗) indicates failure.

In contrast, we observe some degree of skill fusion in the **FO (Flat-only)** robots. For instance, **B2**, trained solely on flat terrain, can climb stairs up to 21 cm. Robots with similar embeddings **B1**, **Stoch5**, and **X30** also demonstrate similar stair-climbing behavior in zero-shot settings. This suggests that proximity in the embedding space can enable generalization.

Gait Analysis: We observe clear periodic gaits for **ZS (Zero-shot)** robots (Stoch5) in the no-robot-encoding case (Figure 8), indicating successful skill fusion. In contrast, the robot-encoding variant (Figure 9) shows disrupted periodicity, suggesting weaker generalization.

Hypothesis: We hypothesize that more diverse training data spanning a broader range of robot morphologies could improve the learned embedding space. Additionally, future work could explore varying the embedding dimensionality or structure to enhance generalization and avoid overfitting to specific robot identities.

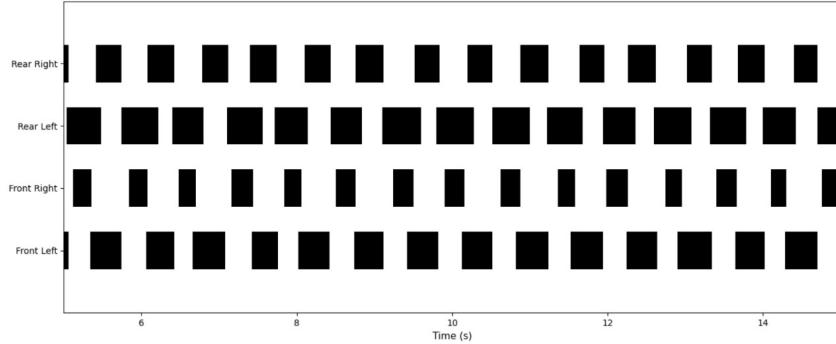


Figure 8: Stoch5(ZO) Foot contact sequence on Flat terrain - No Robot Encoding

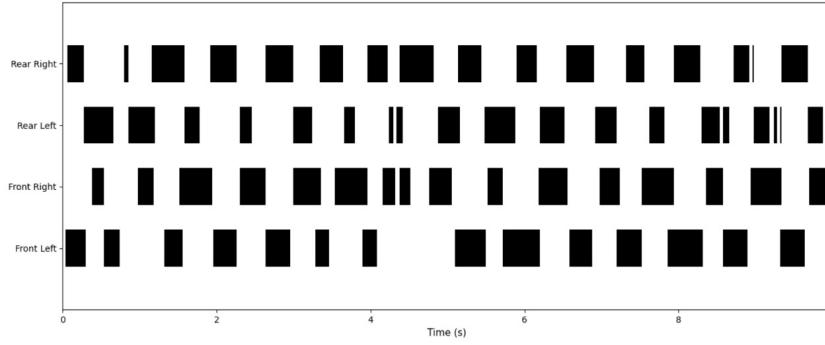


Figure 9: Stoch5(ZO) Foot contact sequence on Flat terrain - Robot Encoding

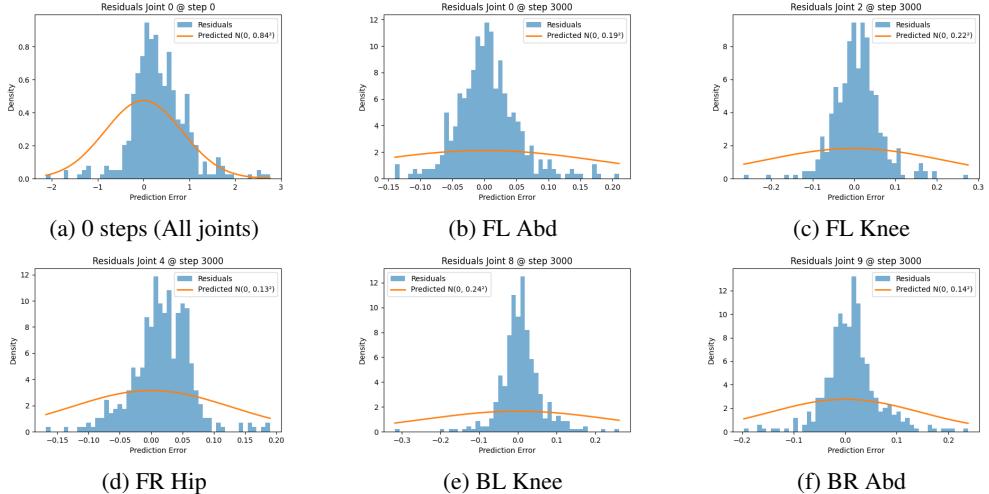


Figure 10: Kernel density plots of residual errors across representative joints. Subfigure (a) shows the initial high-variance residual distribution across all joints at step 0, while (bf) show joint-specific residuals after 3000 training steps, revealing progressive variance adaptation.

C Adaptive Loss and Residual Variance Dynamics

To analyze the effectiveness of the adaptive loss in modeling complex locomotion behaviors across diverse terrains, we investigate the learned per-joint residual distributions after 3000 training steps. Our loss function dynamically adjusts the contribution of each joint prediction by learning a joint-specific variance σ^2 , which controls the weighting of the corresponding residuals during optimization.

Per-Joint Residual Distribution Analysis

The residuals approximately follow zero-mean Gaussian distributions, with their learned variances reflecting the relative modeling difficulty of each joint.

We visualize representative residual distributions for selected joints after 3000 training steps under the adaptive loss framework. The plots show per-joint residuals fitted to zero-mean Gaussians, with variance values learned through joint-wise adaptive weighting. The selected joints span different joint types (abdomen, hip, knee) and leg locations (front/back, left/right), ensuring a balanced representation across the body.

Visualization

Figure 10 shows kernel density plots of the residuals after 3000 training iterations. The shift in density and narrowing of variance highlights the transition from cautious early-stage learning to confident, fine-tuned predictions in later training. These plots further validate the utility of per-joint adaptive weighting for both stable and high-variance locomotion regimes.

D Architectural Details

D.1 Model Architecture

Table 6 displays the hyperparameters of the model. Our architecture follows a modular and interpretable design intended for generalist quadruped locomotion. The architecture is composed of the following blocks:

Component	Value	Remarks
Embedding Dimension (<code>emb_dim</code>)	64	Used in the observation encoder module
Observation Dimensionality	$54 + 3$	Observation + commanded velocity
GRU Hidden Size	64	Matching the embedding dim
Attention Heads	4	In both attention blocks
Attention Window Size	100	For both obs and GRU attention
MLP Hidden Size	256	2-layer ELU MLP head
Optimizer	Adam	Standard
Learning Rate	$1e-3$	Fixed across experiments
Batch Size	400	Per forward pass (batch of 400 trajectories)

Table 6: Model Hyperparameters

- **Observation Encoder:** A linear layer with ELU activation followed by LayerNorm that encodes raw observations and estimated velocity into an embedding space of dimension `emb_dim` = 128.
- **Observation Attention Block:** Multi-head self-attention with 4 heads is applied over a temporal window of past observations using fixed sinusoidal positional embeddings.
- **GRU Block:** A GRU module with input size $2 \times \text{emb_dim}$ and hidden size `emb_dim` encodes the temporal evolution of attended observation features.
- **GRU Attention Block:** Similar to the observation attention block, this layer captures temporal dependencies over the GRU outputs.
- **MLP Head:** A 3-layer MLP maps the concatenated context vector (raw observation encoding, GRU output, and GRU attention output) into action space.

A schematic diagram of the model structure is given in the main paper.

D.2 Attention Design: Single Query Approach

In our model, two types of history are maintained: the observation history and the GRU history. These histories allow the model to capture temporal dependencies across observations and GRU outputs, respectively.

Observation History: The observation history consists of the most recent observations, stored over a window of size $W = 100$. For each timestep, the encoded observation is appended to the history. When the window exceeds its size, the earliest observation is removed to make room for the new one.

Let the observation history be denoted as:

$$\mathbf{o} = [o_1, o_2, \dots, o_W]$$

Where o_t represents the encoded observation at timestep t , and W is the maximum number of timesteps the history holds.

GRU History: Similarly, the GRU history stores the output of the GRU at each timestep. The GRU processes the concatenated observation and its associated attention context, and the resulting output is stored in a history buffer of size $W = 100$.

Let the GRU history be denoted as:

$$\mathbf{h} = [h_1, h_2, \dots, h_W]$$

Where h_t represents the GRU output at timestep t , and W is the size of the GRU history window.

Single Query Attention: After storing these histories, the attention mechanism uses the most recent timestep from the observation history and GRU history for the query. This reduces computational complexity by using only the latest timestep for attention calculations, rather than the entire history.

Let $q_o = o_W$ and $q_h = h_W$ be the queries from the observation and GRU histories, respectively. The attention computation is performed as follows:

$$\text{Attention}(q_o, \mathbf{o}) = \text{Softmax} \left(\frac{q_o K_o^T}{\sqrt{d_k}} \right) V_o$$

$$\text{Attention}(q_h, \mathbf{h}) = \text{Softmax} \left(\frac{q_h K_h^T}{\sqrt{d_k}} \right) V_h$$

Where: - K_o and K_h are the key matrices for observation and GRU histories, respectively, - V_o and V_h are the value matrices for observation and GRU histories, and - d_k is the dimension of the key vectors.

This single query approach is computationally lighter compared to using full attention, as it minimizes both memory and computation costs.

D.3 Attention Pattern Analysis Across Terrains

To understand how the policy attends to information during decision-making, we analyze the mean attention scores per head of both GRU-based temporal attention and observation (OBS) attention layers. Figures 11, 12, and 13 show attention patterns for GRU embeddings across three terrains: flat, stairs, and slope. Corresponding observation attention patterns are shown in Figures 14, 15, and 16.

Key Insight: Each attention head exhibits consistent patterns across different robots when conditioned on the same terrain, indicating that the policy has learned to focus on terrain-specific dynamics rather than robot-specific features.

GRU Attention: GRU attention heads show rhythmic and alternating patterns over time, with the attention weights oscillating between recent and earlier GRU embeddings. This reflects temporal reasoning and the use of history to maintain gait periodicity, especially in flat and slope terrains. The average attention maps confirm that previous time steps are actively attended to, suggesting the network is leveraging temporal memory to drive locomotion.

Observation Attention: In contrast, observation attention heads primarily focus on the most recent observation, regardless of terrain type, which could indicate that immediate sensory feedback is critical for terrain-reactive behavior.

Robots Used for Evaluation:

- **Go1 (small)** Agile and lightweight
- **Stoch3 (medium)** Mid-weight, versatile platform
- **B1 (large)** Heavy-duty robot for large terrain disturbances

Overall, the shared attention behavior across robot morphologies reinforces the terrain-conditioned generalization capability of the policy, validating the design's robustness and scalability.

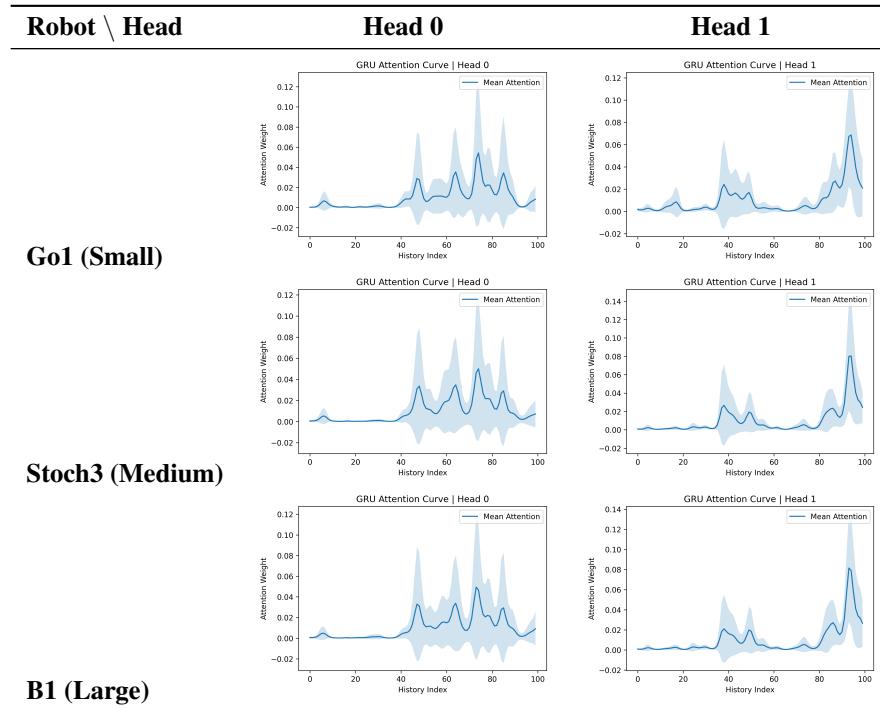


Figure 11: Mean GRU Attention Curves for **Flat Terrain** (17cm) Across Robots and Selected Heads

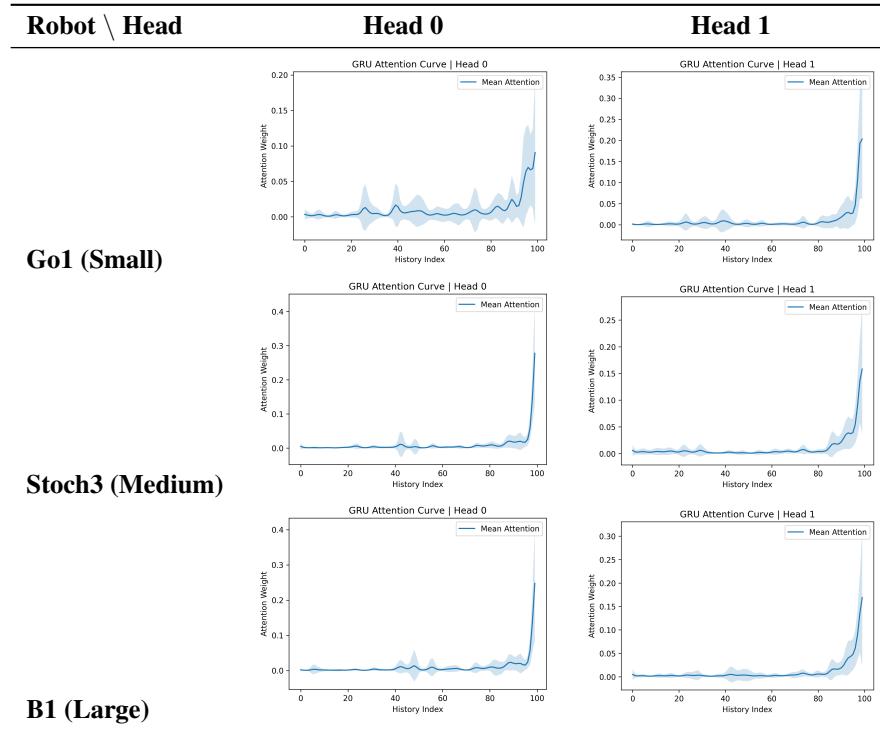


Figure 12: Mean GRU Attention Curves for **Stair Terrain** (17cm) Across Robots and Selected Heads

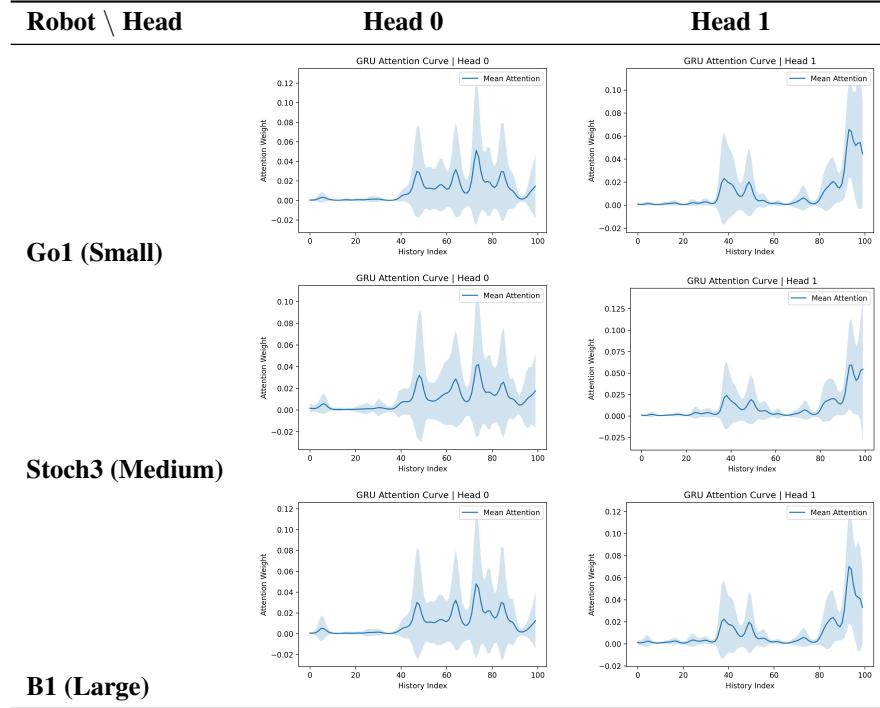


Figure 13: Mean GRU Attention Curves for **Slope Terrain** (Rough Slopes - 25°) Across Robots and Selected Heads

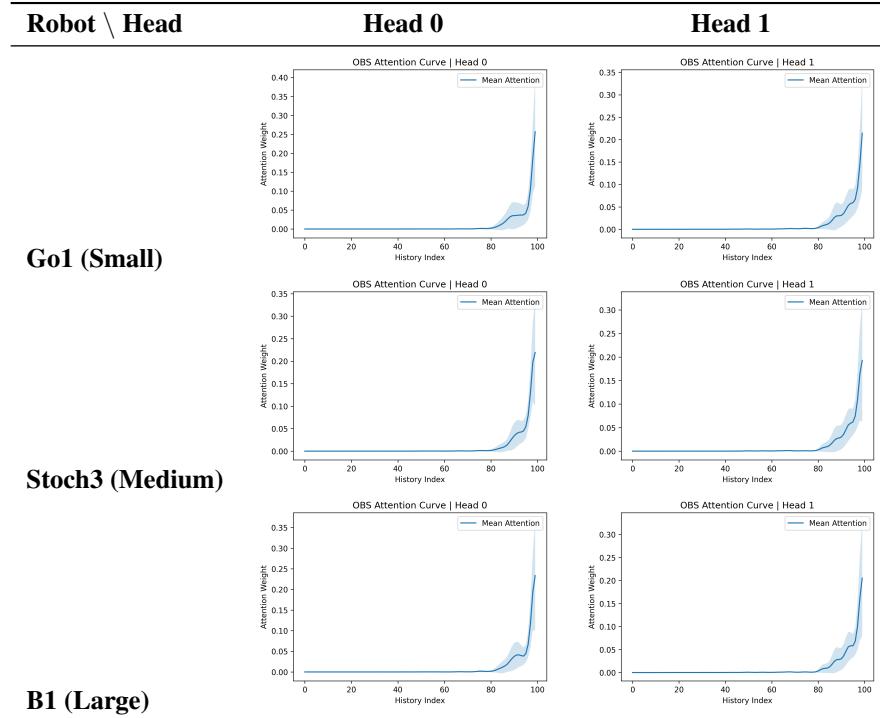


Figure 14: Mean OBS Attention Curves for **Flat Terrain** Across Robots and Selected Heads

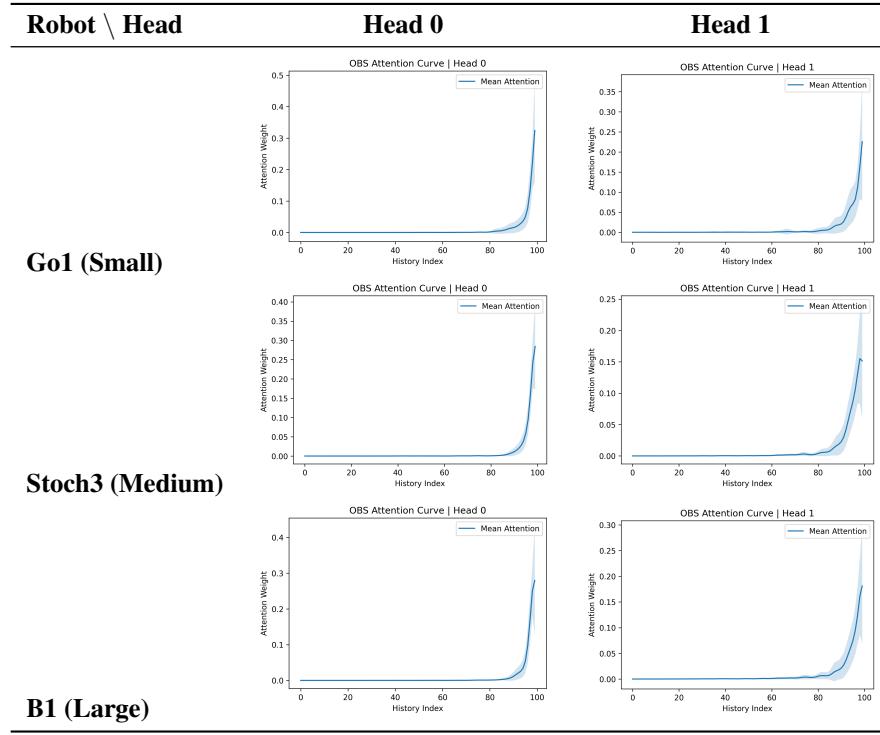


Figure 15: Mean OBS Attention Curves for **Stair Terrain** Across Robots and Selected Heads

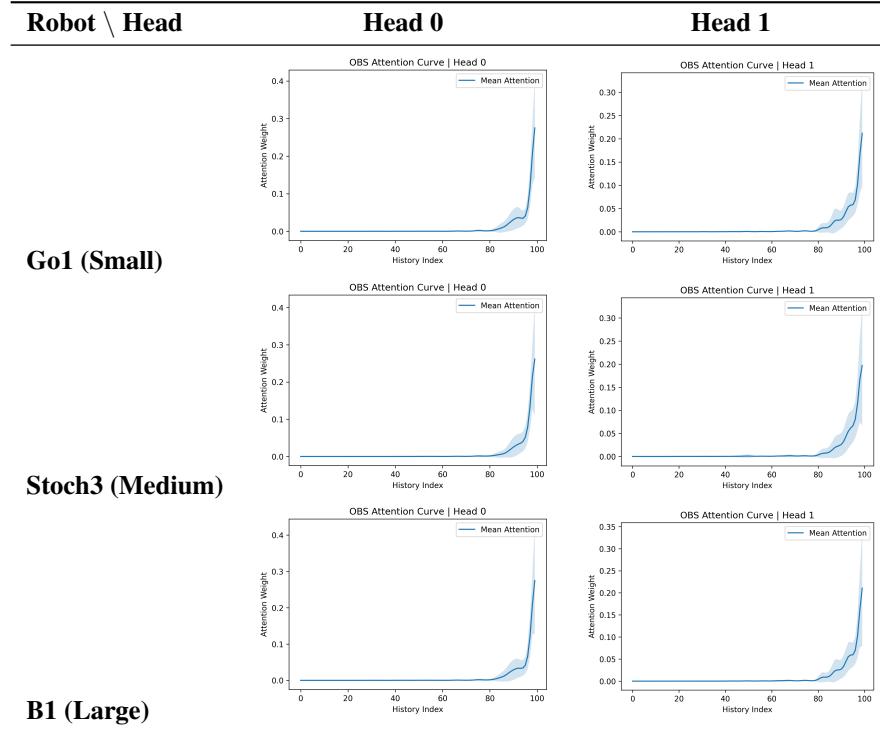


Figure 16: Mean OBS Attention Curves for **Slope Terrain** (Rough Slopes - 25°) Across Robots and Selected Heads