MorphDLoco: Morphology aware Diffusion policy for multi-robot multi-skill locomotion

Mohit Javale[*,1], Kartik Garg[*,1], Narayanan PP[*,1,‡], Shishir Kolathaya[1]

Abstract— **Learning from offline datasets poses scalability challenges for robotics due to their limited availability. Furthermore, current offline learning techniques in robotics either lack the ability to translate effectively to real-time deployment or fail to capture the inherent diversity within the datasets. In this work, we propose a diffusion learning framework as a unified policy capable of distilling multiple skills across a variety of robotic platforms into a single architecture. We leverage diffusion's multimodal capabilities to learn from diverse datasets, demonstrating that our single diffusion-based policy can learn multiple locomotion gaits—such as trot, pronk, bound, and pacing—across quadrupedal robots with varying morphologies and physical parameters. The policy is conditioned on the robot's characteristics and desired gait commands. Moreover, the policy generalizes across different terrains, successfully handling multiple slopes in simualtion and real-world hardware tests, with slopes varying between 13–16 degrees. Through comprehensive experimentation, we illustrate the superiority of our diffusion-based policy compared to reinforcement learning and non-diffusion behavior cloning baselines. We validate the robustness of our approach by deploying the policy on two quadrupedal robots with significantly different morphologies: the Unitree Go1, a commercially available `12kg` robot, and the Stoch 5, a `70kg` in-house developed quadruped.**

I. INTRODUCTION

Traditionally, learning from large-scale offline data was not possible in robotics, primarily due to the lack of methods for generating diverse and accurate data, as well as the computational demands required to process such data in a way that remains adaptable to real-time deployment on hardware. However, such methods of scaling offline data in conjuction with powerful architecture has shown great promises in the realms of natural language processing, seen by the rise of large language models in recent years. Following suit, recent advances in robotics, including improvements in the quality of simulation data and the availability of large-scale collaborative teleoperation datasets [34], along with increasingly powerful on-board computational resources, have contributed to a growing interest in the utilization of offline learning techniques for controls [41], [12] and decision-making applications.

In contrast, online reinforcement learning (RL) techniques [17], [45], [44], [5] have achieved significant success in real-time control tasks. However, these methods are typically specialized for particular tasks or robot morphologies on which they have been trained. Adapting RL policies to new morphologies presents considerable challenges due to variations in robot dynamics and actuator models, which necessitate complex reward structures and extensive reward tuning. Additionally, the development of new skills often requires the design of task-specific reward functions, further increasing the need for detailed reward engineering. As a result, existing approaches have explored two distinct strategies: morphology-aware methods [27], [4], [39], which aim to generalize a single skill across multiple robots, and multi-skill methods [36], [35], [23], which enable a single robot to acquire multiple skills. Despite these advancements, RL-based methods have yet to demonstrate a unified policy capable of learning and executing multiple skills across multiple robots simultaneously.

When it comes to demonstrating multiple skills across different robots, most approaches focus on applying imitation learning to large offline datasets, often generated using teleoperated or handcrafted base controllers [15], [11], [14]. A recent innovation in this area has explored the use of powerful goal- and state-conditioned diffusion architectures to clone behaviors from such datasets [7], [38]. However, these models often suffer from long inference times, which limits their effectiveness. As a result, they have primarily shown success in quasi-static environments, such as manipulation tasks, where the environment remains stable enough to accommodate the compute-heavy nature of these models.

To address these challenges, we propose a unified, lightweight conditional diffusion model capable of scaling across multiple skills, diverse robot morphologies, and varying slopes. Our approach focuses on quadrupedal locomotion—a time-sensitive control task with minimal passive stability. We collect simulation data from expert-trained models, capturing a variety of gaits—trot, bound, pronk, and pace—across robots with diverse morphologies and on different slopes. Using this rich, multidimensional dataset, we train a single diffusion model that learns a unified policy, enabling zero-shot deployment for real-time control on hardware across all three dimensions: multiple gaits, robots, and terrain slopes.

The primary contributions of our paper include -

1) **Multi-Robot**: We propose a novel morphology-aware controller that learns quadrupedal locomotion for robots of diverse morphologies in a single unified policy.

* Equal Contribution
[1] Stochastic Robotics Lab, Indian Institute of Science, Bangalore
‡ Corresponding Author: narayananp@iisc.ac.in
Code: https://github.com/StochLab/MorphDLoco

2) **Multi-Skill Distillation**: Using our novel framework, we show that data generated from multiple policies or streams can be distilled into a singular task-conditioned diffusion, without compromising heavily on the performance with respect to any task. Another emergent advantage of our method is the ability to interpolate across the learnt skills, as well as better skill transition.
3) **Hardware**: We validate our method through extensive hardware testing on two quadrupeds with vastly different morphologies. Our controller is deployed on the lightweight Unitree Go1, weighing around 12.5 kg, and the much bulkier Stoch 5, which weighs 70 kg—over five times the mass of the Go1. We deploy the controller at 50Hz without any hardware-specific tuning. To the best of our knowledge, this is one of the first works to demonstrate a diffusion-based policy operating in real-time on quadrupeds with such a significant difference in size and weight. We showcase that our novel controller can successfully climb slopes of varying magnitudes in simulation and hardware. Our controller also inherits properties of the expert policy, such as being robust over rough terrains, random disturbances, and changing friction values.

We plan to release the code for our method as open-source in the near future.

II. RELATED WORKS

A. Diffusion in Robotics

Recently, diffusion models [40] have gained prominence in control tasks due to their multimodal capabilities. They have demonstrated significant success in learning from offline datasets and human demonstrations. For example, [19] and [2] highlight the efficacy of diffusion models in planning and offline tasks, respectively. In the realm of online reinforcement learning, recent studies [3], [37], [24], [9] have explored various techniques for fine-tuning and training diffusion models through policy optimization methods. Specifically, [3] interprets the diffusion process as a Markov Decision Process (MDP) with a sparse reward at the end of the denoising phase, while [37] addresses two MDPs—one for diffusion and another for the environment. Despite these advances, most of these approaches have been confined to simulations and have not yet been applied to actual hardware.

Diffusion models can be conditioned on various inputs to generate unique outputs, creating opportunities for integrating multimodal states into robotic control tasks [7]. For manipulation tasks, these models have been employed to combine visual inputs with language commands [20], [13]. Hierarchical methods for skill chaining and planning have also been developed [32], [31], [30], though these approaches generally operate at a low action frequency of around 10Hz due to the stable dynamics involved. In contrast, legged locomotion tasks require high-frequency feedback control due to their inherently unstable dynamics. [18] demonstrates a real-time diffusion-based locomotion policy that showcases multiple skills across quadruped robots by optimizing the inference process of their transformer model. Our approach, however, extends this by employing a single diffusion policy to demonstrate multiple skills on morphologically diverse robots and across various slopes, using only MLPs as the backbone for real-time performance.

B. Learning-based Legged Locomotion

Recent studies in Deep Reinforcement Learning (DRL) for quadruped locomotion [22], [33], [25], [1], [26], [10] have garnered significant interest and have made substantial progress in enabling locomotion over diverse terrains such as slopes, uneven ground, stairs, etc. However, despite these advances, most of these approaches are tailored to specific robot morphologies, meaning these frameworks are designed to train locomotion policies for only a single quadruped morphology thereby limiting their generalizability to different quadruped designs. Developing a single learning policy for a range of morphologically diverse quadruped robots presents a challenging problem due to variations in mass, link lengths, dynamics, and actuator models, which result in complex locomotion requirements. This challenge is further compounded by the need for the policy to address multiple dimensions simultaneously: accommodating a variety of morphologies, including very large quadrupeds; learning versatile locomotion skills, such as executing different gaits; and adapting to varying terrains like slopes. Each of these dimensions adds complexity, making the development of such a versatile policy exceptionally demanding.

Bohlinger et al. [4] developed a Multi-Task Reinforcement Learning (MTRL) framework that enables a single policy to control various legged robots, including quadrupeds, humanoids, and hexapods. While this framework allows for policy learning across different legged morphologies, its hardware evaluations are mostly limited to flat terrain, such as pavements, grass, and plastic turf. The hardware tests were conducted on three small quadruped robots: Unitree Go1, MAB Honey Badger, and MAB Silver Badger, while the performance of the framework on larger quadrupeds, such as the AnyMal-C, which has complex actuator dynamics, remains untested in real-world scenarios.

In contrast, Zeren et al. [27] proposed the MorAL framework, which facilitates locomotion across diverse terrains, including stairs and slopes, for quadrupeds of varying morphologies. This framework generates a range of robot morphologies during simulation initialization to train generalizable strategies. Although the paper claims to handle up to 245% variations in size and mass, the hardware tests have been confined to the randomization range used in simulations. There is no demonstration of zero-shot transfer or results for larger quadrupeds, such as a 70kg robot, which exceeds their simulated range.

Both approaches face challenges in real-world validation on larger robots, as sim-to-real transfer becomes harder with increased mass and larger link dimensions. Additionally, in both of these approaches, the generalist RL policy is not tasked with ”skill learning”, i.e. learning multiple gaits, which is essential for versatile locomotion.

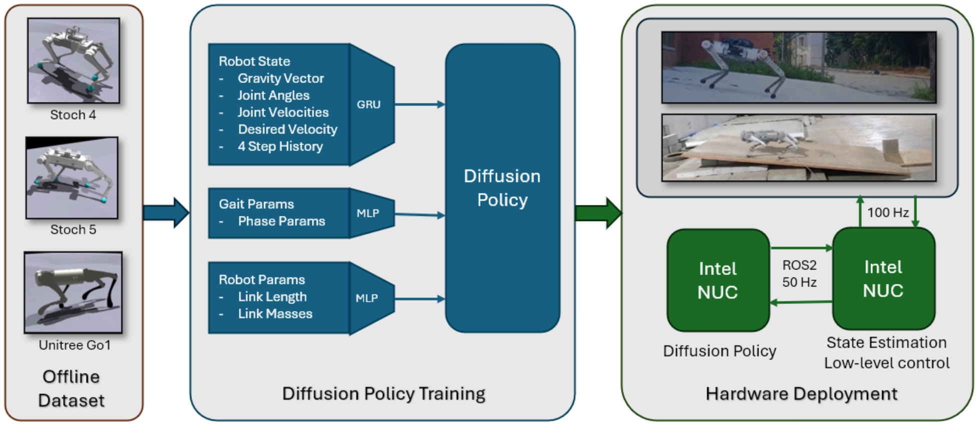


Fig. 1: Offline Diffusion framework data collection, training and deployment pipeline

A promising approach for learning diverse skills is through diffusion models, which have effectively demonstrated the acquisition of various locomotion skills, such as jumping, trotting, and bipedal walking, in quadruped robots [18]. This success highlights their advantage over AMP and Motion Imitation methods [43], [42], [23], [8], [6], [21]. However, while diffusion-based policies have shown effectiveness in skill learning, their ability to generalize across different quadruped morphologies and terrains remains unexplored. To address this, we propose **MorphDLoco**, a novel diffusion-based methodology designed to handle a range of quadruped morphologies, enabling skill learning through multiple gaits and effective locomotion on slopes.

III. METHODOLOGY

In this work, we employ Denoising Diffusion Probabilistic Models (DDPM) [16] to train a single policy that learns the distribution of actions required for locomotion on both flat and sloped surfaces across different robot morphologies while utilizing various gaits. This section outlines our formulation of the problem using diffusion models. We also discuss our offline data collection process and the design choices that enable our method to operate in real time.

A. Denoising Diffusion Probabilistic Models

Denoising Diffusion Probabilistic Models (DDPM) have proven effective in modeling multimodal data distributions. In DDPM, the forward noising process progressively corrupts the data by adding Gaussian noise. This process begins with a data point sampled from the original distribution $x_{k-1} \sim q(x)$, and noise is added according to a variance schedule $\{\beta_k \in (0,1)\}_{k=1}^{K}$. The noising process is defined as:

$$q(x_k|x_{k-1}) = \mathcal{N}(x_k; \sqrt{1-\beta_k}x_{k-1}, \beta_k I) \tag{1}$$

Here, x_k represents the noisy data at step k, and $\mathcal{N}$ denotes a Gaussian distribution with mean $\sqrt{1-\beta_k}x_{k-1}$ and covariance $\beta_k I$.

The generative process follows an iterative denoising process similar to Langevin Dynamics. Starting from $x_k \sim \mathcal{N}(0,1)$, the denoising process to get x_{k-1} can be represented mathematically as follows:

$$x_{k-1} = \frac{1}{\sqrt{\alpha_k}}(x_k - \frac{1-\alpha_k}{\sqrt{1-\hat{\alpha}_k}}\epsilon_\theta) + \sigma_k \mathcal{N}(0,1) \tag{2}$$

where $\alpha_k = 1 - \beta_k, \hat{\alpha_k} = \prod_{i=1}^{k} \alpha_i$ and $\sigma_k = \sqrt{\beta_k}$. This process continues iteratively for timesteps K to reduce noise in the sample x_k until the desired noise-free output x_0 is generated.

B. Problem Formulation

In our application, the DDPM policy predicts target joint angles (a_0) to be applied to the quadruped, with respect to the default initial joint positions. The goal is to predict the appropriate joint angle actions by progressively denoising the predictions through each step. We apply equation 1 to add noise during the training process. The policy is conditioned on the history of states, previous actions, commands, and robot-specific parameters to predict $\epsilon_\theta(a_k, C_{inputs}, k)$ to sample the next denoised action a_{k-1} using eq 2 for any given denoising step k. We train our policy using the following objective function:

$$\mathcal{L} = MSE(\epsilon_k, \epsilon_\theta(a_k, C_{inputs}, k)) \tag{3}$$

where ϵ_k and a_k represent the added noise in the forward diffusion process, and the denoised action at step k respectively, and C_{inputs} represents the conditional inputs.

C. Conditional Inputs

We use the following robot data as conditional inputs:

1) **State History**: The state history includes the gravity vector, DOF velocities, changes in DOF positions from the initial state, and the last two actions, spanning a history length of 4. We found that this minimal set of robot states is sufficient for training an efficient and lightweight DDPM policy. The state history is processed through a **Temporal Dynamics Encoder**, which produces a compact and informative representation, $f_{hist} \in \mathbb{R}^{\mathcal{D}_{hist}}$, where $\mathcal{D}_{hist} = 64$.

Robot Paramters	Unitree Go1	Stoch4	Stoch5
Body Length (in m)	0.40	0.54	0.67
Body Width (in m)	0.13	0.20	0.26
Abduction Length (in m)	0.08	0.12	0.135
Thigh Length (in m)	0.23	0.30	0.35
Shank Length (in m)	0.24	0.35	0.35
Base Mass (in kg)	4.80	9.75	38.07
Abduction Mass (in kg)	0.51	0.94	2.61
Thigh Mass (in kg)	0.89	2.49	4.88
Shank Mass (in kg)	0.16	0.38	1.5

TABLE I: Description of robot parameters

Model and Encoder Details	
Activation Function	ELU
Temporal Dynamics Encoder	GRU (hidden dims: 64)
Diffusion Policy	6-Layer MLP (256 hidden units/layer)
Gait Encoder	1-Layer MLP (Input dim: 3, Output dim: 8)
Robot Profile Encoder	1-Layer MLP (Input dim: 9, Output dim: 16)
Training Details	
Learning rate	1e-3
Batch size	4096
Denoising timesteps	60
Training samples	600k (200k/robot)
Timer embeddings	Linear embeddings

TABLE II: Training and Hyperparameter Details

2) **Robot Parameters**: To discriminate between different robots, we condition our DDPM policy on the robot width, length, link lengths, and link masses. These parameters are further encoded to give $f_{rob} \in \mathbb{R}^{\mathcal{D}_{rob}}$, where $\mathcal{D}_{rob} = 16$. The actual values of the parameters for all the three robots can be found in Table I

3) **Commands**: We pass velocity commands (v_x, v_y, w_z) and gait commands θ_{cmd} to control the policy. We follow [29] to represent the gait commands. The gait commands are encoded into meaningful representation $f_{gait} \in \mathbb{R}^{\mathcal{D}_{gait}}$ where $\mathcal{D}_{gait} = 8$.

D. Data Collection

In this section, we outline the methodology for collecting the datasets used to train our diffusion model. The dataset consists of conditional diffusion inputs and target joint actions, gathered using expert RL policies trained in simulation. We leverage the NVIDIA IsaacGym [28] simulation engine to collect data concurrently from 400 parallel robots, allowing efficient large-scale data generation.

The data is collected from three morphologically distinct quadrupeds: Unitree Go1, Stoch 4, and Stoch 5, with expert RL policies trained to execute multiple quadrupedal gaits [29]. We gather data across both flat ground and slopes, splitting the terrain evenly. Each robot generates 200,000 samples across all gaits. To enhance the robustness of our diffusion policy, we incorporate domain randomization techniques during data collection. This includes varying ground friction (0 to 1), joint friction (0.1 to 3), restitution (0 to 0.4), and applying external forces to simulate real-world conditions. These variations help our model generalize and adapt to diverse environments during real-world deployment.

During data collection, each robot is commanded with a single velocity input—either forward velocity v_x, lateral velocity v_y, or yaw rate w_z. While testing, we aim to provide all three velocity commands simultaneously and evaluate whether the diffusion model can interpolate and handle combined motion commands effectively.

E. Architecture

Our diffusion policy architecture consists of a 6-layer MLP that predicts joint actions by conditioning on several encoded inputs. First, the observation history (a 4-step sequence) is processed by the **Temporal Dynamics Encoder**, a GRU-based module specifically designed to capture time-dependent features from the robot's sensory inputs. Naively conditioning the diffusion model directly on the raw observation history led to instability, where the robot would stumble after taking a few steps, making this encoder crucial for stable learning. Additionally, we employ a **Gait Encoder** to process the command for different locomotion gaits (trotting, pronking, bounding, and pacing), and a **Robot Profile Encoder** to encode the unique physical parameters of each robot, such as link lengths and mass. The diffusion model is then conditioned on these encoded representations—observation history, gait, and robot parameters—along with velocity commands, predicting denoised joint angles for control. This approach allows the model to generalize effectively across various gaits, different robot morphologies, and multiple terrains like slopes. Full training and hyperparameter details are provided in Table II.

IV. RESULTS

In this section, we first benchmark our method against existing baselines. Following this, we conduct in-depth experiments to evaluate and compare our policy across several scenarios: Multi-Robot Experiments, and Multi-Skill Learning. Additionally, we demonstrate the versatility of our approach through out-of-distribution gait transitions. Lastly, we explore zero-shot skill transfer to a robot morphology and velocity interpolation across morphologies. We provide videos of hardware results in the supplementary.

A. Robots and Hardware setup

We showcase the results on two robots (Go1 and Stoch 5) in the real world and three robots (Go1, Stoch 5 and Stoch 4) in simulation. More details about these robots can be found on in Table I. The low-level control and state estimation runs on an on-board Intel NUC at 100 Hz, while the diffusion policy runs on another separate on-board Intel NUC, communicating with the main NUC over ROS2 channels at 50 Hz, receiving the required states and responding with corresponding actions to be taken to be forwarded to the low-level controller. The diffusion policy was optimized for inference using just-in-time (JIT) compilation, and was able to achieve real-time inference solely based on CPU compute of Intel NUC.

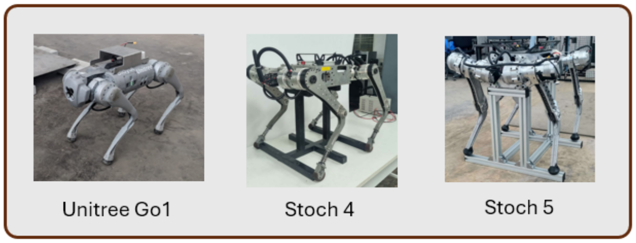


Fig. 2: Robots on which we have validated our policy. Unitree Go1 is a commercially available 12 kg robot made by Unitree. Stoch 4 and Stoch 5 are lab-built robots, with Stoch 4 weighing around 25 kg, and Stoch 5 exceeding that reaching up to 70 kg.

B. Baselines and Tasks

For our baselines, we compare against the source expert policy, which in our case is the Walk These Ways policy [29]. Additionally, we benchmark our diffusion policy against non-diffusion methods, specifically Behavior Cloning. This baseline uses a similar architecture to our diffusion policy, but with the loss function replaced by a reconstruction loss.

We evaluate our method on three different tasks:

1) **Flat Ground Walking:** We present velocity tracking comparisons of our method against the baselines, tracking a target velocity of 0.5 m/s in the x-direction. For simplicity, we focus on the trot gait in the main paper, as the results observed for the trot gait are representative of all gaits. Additional plots for other gaits can be found in the supplementary material.
2) **Slope walking:** For this experiment, we show velocity for all three robots on slopes of 15-degree inclination.
3) **Gait transitions:** We demonstrate our method's capability to follow gait commands and transition between different gaits.

We also show zero-shot skill transfer of learned skill to a robot of new morphology as a downstream application of our method. Further we observed that although our training data only included single velocity commands for each robot, our diffusion policy successfully learns to interpolate and generalize across multiple velocity commands during testing.

C. Mutli Robot Experiments

Our method learns useful locomotion skills on different morphology robots. To test the capabilities of our policy, we compare its command tracking against baselines. Figure 3 present the comparative performance of our method against the RL and Behavior Cloning baselines for both flat ground and sloped environments. As shown, our method exhibits strong velocity tracking capabilities across all three robots. The RL baseline performs well on Go1, as it was specifically trained for this robot. It performs reasonably on Stoch 4 due to its morphological similarity to Go1, but fails completely on Stoch 5, which is expected since RL policies struggle to generalize to out-of-distribution scenarios. Admittedly, our diffusion method does not outperform the RL baseline for Go1 and is, at best, on par with it. However, the strength of our approach lies in its ability to maintain the same level of proficiency across robots with significantly different morphologies, using a single unified policy.

It is evident from Figure 3 that our method significantly surpasses the behavior cloning baseline in performance. Behavior cloning fails to capture the inherent dynamics of each robot from the offline dataset. As a result, it fails catastrophically to follow velocity commands. This shortcoming likely arises because the policy is trained solely using reconstruction loss, which averages out the diverse morphological behaviors, resulting in suboptimal performance. Moreover, these experiments indicate that the behavior cloning baseline is not receptive to the robot-specific conditional inputs. Our conditional diffusion policy can learn morphologically aware behaviors owing to its multimodal capacity and thus perform better than Behaviour cloning in all scenarios.

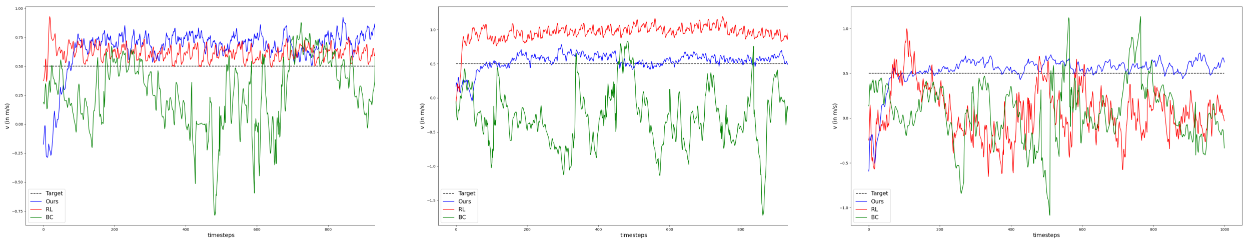
D. Multi Skill Learning

We demonstrate that our method successfully follows gait commands for trot, bound, pace, and pronk gaits by conditioning on gait parameters. We first collect evenly distributed data from each robot for each gait using robot-specific expert policies to train the diffusion model. For evaluation, we change the gait every 250 timesteps following the sequence: bound, pronk, trot, and pace. Figure 4 shows the transition between different gaits performed by our method on go1. The model effectively learns a single policy that can execute all these gaits. Notably, the dataset did not include any skill transition data points, highlighting the capability of our method to transition between different gaits, even without explicit transition data during training.

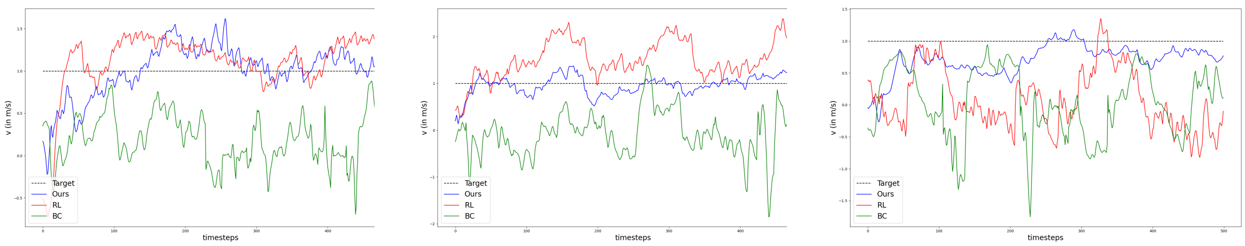
E. Zero shot skill transfer and Velocity Interpolation

This section demonstrates a downstream use case of our trained conditional diffusion policy. The goal of this experiment is to show that our policy is capable of transferring learned skills to a robot with a new morphology in a zero-shot manner. We train the diffusion policy using gait data from Go1 and Stoch 5, while for Stoch 4, we provide data for all gaits except trot. The trained policy successfully learns the trot gait for Stoch 4. These results demonstrate that conditioning on robot-specific parameters can lead to a degree of generalization, even in behavior cloning methods. Our method captures the dynamics of a skill across different morphologies and is able to interpolate that skill to new morphologies.

Moreover, we observe that our method can interpolate to velocities not present in the training dataset. The policy was trained with only one velocity command at a time. However, during testing, we provide v_x, v_y, and w_z, simultaneously, each set to 0.5 m/s, to assess the policy's interpolation capabilities. Our method successfully learns to interpolate these velocity commands. In contrast, Behavior Cloning fails to effectively learn these interpolation skills. RL baselines trained on Go1 excel for Go1 but perform poorly on Stoch 4 and Stoch 5, highlighting their struggle with velocity interpolation. More detailed results are present



(a) Go1 (Flat ground) (b) Stoch 4 (Flat ground) (c) Stoch 5 (Flat ground)



(d) Go1 (Slopes) (e) Stoch 4 (Slopes) (f) Stoch 5 (Slopes)

Fig. 3: Velocity tracking comparison MorphDLoco (Blue), RL Expert Policy trained on Go1 (Red), Behaviour Cloning (Green) baselines. MorphDLoco consistently tracks the desired velocity (Black) well for all morphologies while being at par with RL Expert Policy for Go1.

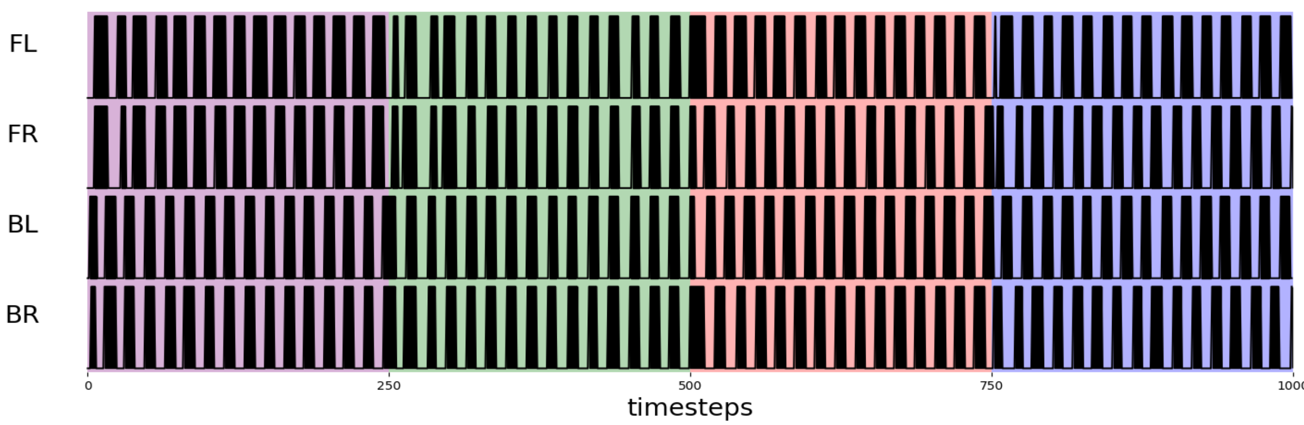


Fig. 4: Foot contact sequence - black regions are periods of foot contact. We commanded the robot to change gaits amidst walking: bound, pronk, trot and pace in that order respectively, represented by purple, green, red and blue regions to observe gait transitions. The transition between gaits is smooth and quick, with the policy adjusting the contact phases within 1 gait cycle to adapt to the commanded velocity.

in the supplementary.

V. CONCLUSIONS

We present MorphDLoco, a unified Morphology-Aware Diffusion policy designed to learn multiple locomotion skills for quadruped robots with diverse morphologies from offline datasets, achieving transfer to hardware. By leveraging the multimodal capabilities of diffusion models, MorphDLoco trains a single policy that effectively handles various skills by conditioning on gait commands and robot-specific parameters. Our results demonstrate that MorphDLoco excels in velocity tracking and gait transitions, even without transition data in the training sets. Additionally, we highlight the ability of MorphDLoco to transfer learned skills to new robot morphologies, showcasing its adaptability and versatility.
In future work, MorphDLoco can be extended to accommodate a broader range of morphologies and more extreme terrains by incorporating vision inputs into the diffusion policy. Additionally, a deeper investigation into the zero-shot skill transfer is warranted, focusing on how data and architecture impact the ability to transfer skills to new robot morphologies. MorphDLoco offers a scalable approach to learning from offline data, potentially paving the way for developing robot-agnostic generalist locomotion policies.

VI. ACKNOWLEDGMENT

This research is funded by Core research grant, ANRF. We would like to thank K Srinivas, Sarvesh Prasanth V and Shaligram Dewangan for their help with the hardware setup.

REFERENCES

[1] A Agarwal, A Kumar, J Malik, and D Pathak. Legged locomotion in challenging terrains 352 using egocentric vision. In *Conference on Robot Learning*, pages 403–415.

[2] Anurag Ajay, Yilun Du, Abhi Gupta, Joshua Tenenbaum, Tommi Jaakkola, and Pulkit Agrawal. Is conditional generative modeling all you need for decision-making? *arXiv preprint arXiv:2211.15657*, 2022.

[3] Kevin Black, Michael Janner, Yilun Du, Ilya Kostrikov, and Sergey Levine. Training diffusion models with reinforcement learning. *arXiv preprint arXiv:2305.13301*, 2023.

[4] Nico Bohlinger, Grzegorz Czechmanowski, Maciej Krupka, Piotr Kicki, Krzysztof Walas, Jan Peters, and Davide Tateo. One policy to run them all: an end-to-end learning approach to multi-embodiment locomotion. *arXiv preprint arXiv:2409.06366*, 2024.

[5] Yuanpei Chen, Tianhao Wu, Shengjie Wang, Xidong Feng, Jiechuan Jiang, Zongqing Lu, Stephen McAleer, Hao Dong, Song-Chun Zhu, and Yaodong Yang. Towards human-level bimanual dexterous manipulation with reinforcement learning. *Advances in Neural Information Processing Systems*, 35:5150–5163, 2022.

[6] Xuxin Cheng, Kexin Shi, Ananye Agarwal, and Deepak Pathak. Extreme parkour with legged robots. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 11443–11450. IEEE, 2024.

[7] Cheng Chi, Siyuan Feng, Yilun Du, Zhenjia Xu, Eric Cousineau, Benjamin Burchfiel, and Shuran Song. Diffusion policy: Visuomotor policy learning via action diffusion. *arXiv preprint arXiv:2303.04137*, 2023.

[8] Alejandro Escontrela, Xue Bin Peng, Wenhao Yu, Tingnan Zhang, Atil Iscen, Ken Goldberg, and Pieter Abbeel. Adversarial motion priors make good substitutes for complex reward functions. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 25–32. IEEE, 2022.

[9] Ying Fan, Olivia Watkins, Yuqing Du, Hao Liu, Moonkyung Ryu, Craig Boutilier, Pieter Abbeel, Mohammad Ghavamzadeh, Kangwook Lee, and Kimin Lee. Dpok: reinforcement learning for fine-tuning text-to-image diffusion models. In *Proceedings of the 37th International Conference on Neural Information Processing Systems*, pages 79858–79885, 2023.

[10] Zipeng Fu, Ashish Kumar, Jitendra Malik, and Deepak Pathak. Minimizing energy consumption leads to the emergence of gaits in legged robots. *arXiv preprint arXiv:2111.01674*, 2021.

[11] Zipeng Fu, Qingqing Zhao, Qi Wu, Gordon Wetzstein, and Chelsea Finn. Humanplus: Humanoid shadowing and imitation from humans. *arXiv preprint arXiv:2406.10454*, 2024.

[12] Zipeng Fu, Tony Z Zhao, and Chelsea Finn. Mobile aloha: Learning bimanual mobile manipulation with low-cost whole-body teleoperation. *arXiv preprint arXiv:2401.02117*, 2024.

[13] Gunshi Gupta, Karmesh Yadav, Yarin Gal, Dhruv Batra, Zsolt Kira, Cong Lu, and Tim GJ Rudner. Pre-trained text-to-image diffusion models are versatile representation learners for control. *arXiv preprint arXiv:2405.05852*, 2024.

[14] Tairan He, Zhengyi Luo, Xialin He, Wenli Xiao, Chong Zhang, Weinan Zhang, Kris Kitani, Changliu Liu, and Guanya Shi. Omnih2o: Universal and dexterous human-to-humanoid whole-body teleoperation and learning. *arXiv preprint arXiv:2406.08858*, 2024.

[15] Tairan He, Zhengyi Luo, Wenli Xiao, Chong Zhang, Kris Kitani, Changliu Liu, and Guanya Shi. Learning human-to-humanoid real-time whole-body teleoperation. *arXiv preprint arXiv:2403.04436*, 2024.

[16] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

[17] David Hoeller, Nikita Rudin, Dhionis Sako, and Marco Hutter. Anymal parkour: Learning agile navigation for quadrupedal robots. *Science Robotics*, 9(88):eadi7566, 2024.

[18] Xiaoyu Huang, Yufeng Chi, Ruofeng Wang, Zhongyu Li, Xue Bin Peng, Sophia Shao, Borivoje Nikolic, and Koushil Sreenath. Diffuseloco: Real-time legged locomotion control with diffusion from offline datasets. *arXiv preprint arXiv:2404.19264*, 2024.

[19] Michael Janner, Yilun Du, Joshua B Tenenbaum, and Sergey Levine. Planning with diffusion for flexible behavior synthesis. *arXiv preprint arXiv:2205.09991*, 2022.

[20] Tsung-Wei Ke, Nikolaos Gkanatsios, and Katerina Fragkiadaki. 3d diffuser actor: Policy diffusion with 3d scene representations. *arXiv preprint arXiv:2402.10885*, 2024.

[21] Arnaud Klipfel, Nitish Sontakke, Ren Liu, and Sehoon Ha. Learning a single policy for diverse behaviors on a quadrupedal robot using scalable motion imitation. In *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 2768–2775. IEEE, 2023.

[22] Ashish Kumar, Zipeng Fu, Deepak Pathak, and Jitendra Malik. Rma: Rapid motor adaptation for legged robots. *arXiv preprint arXiv:2107.04034*, 2021.

[23] Chenhao Li, Sebastian Blaes, Pavel Kolev, Marin Vlastelica, Jonas Frey, and Georg Martius. Versatile skill control via self-supervised adversarial imitation of unlabeled mixed motions. In *2023 IEEE international conference on robotics and automation (ICRA)*, pages 2944–2950. IEEE, 2023.

[24] Zechu Li, Rickmer Krohn, Tao Chen, Anurag Ajay, Pulkit Agrawal, and Georgia Chalvatzaki. Learning multimodal behaviors from scratch with diffusion policy gradient. *arXiv preprint arXiv:2406.00681*, 2024.

[25] Junfeng Long, Zirui Wang, Quanyi Li, Liu Cao, Jiawei Gao, and Jiangmiao Pang. Hybrid internal model: Learning agile legged locomotion with simulated robot response. In *The Twelfth International Conference on Learning Representations*, 2024.

[26] Shixin Luo, Songbo Li, Ruiqi Yu, Zhicheng Wang, Jun Wu, and Qiuguo Zhu. Pie: Parkour with implicit-explicit learning framework for legged robots. *IEEE Robotics and Automation Letters*, 2024.

[27] Zeren Luo, Yinzhao Dong, Xinqi Li, Rui Huang, Zhengjie Shu, Erdong Xiao, and Peng Lu. Moral: Learning morphologically adaptive locomotion controller for quadrupedal robots on challenging terrains. *IEEE Robotics and Automation Letters*, 9(5):4019–4026, 2024.

[28] Viktor Makoviychuk, Lukasz Wawrzyniak, Yunrong Guo, Michelle Lu, Kier Storey, Miles Macklin, David Hoeller, Nikita Rudin, Arthur Allshire, Ankur Handa, et al. Isaac gym: High performance gpu-based physics simulation for robot learning. *arXiv preprint arXiv:2108.10470*, 2021.

[29] Gabriel B Margolis and Pulkit Agrawal. Walk these ways: Tuning robot control for generalization with multiplicity of behavior. In *Conference on Robot Learning*, pages 22–31. PMLR, 2023.

[30] Utkarsh A Mishra and Yongxin Chen. Reorientdiff: Diffusion model based reorientation for object manipulation. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 10867–10873. IEEE, 2024.

[31] Utkarsh Aashu Mishra, Yongxin Chen, and Danfei Xu. Generative factor chaining: Coordinated manipulation with diffusion-based factor graph. In *ICRA 2024 Workshop {\textemdash} Back to the Future: Robot Learning Going Probabilistic*.

[32] Utkarsh Aashu Mishra, Shangjie Xue, Yongxin Chen, and Danfei Xu. Generative skill chaining: Long-horizon skill planning with diffusion models. In *Conference on Robot Learning*, pages 2905–2925. PMLR, 2023.

[33] I Made Aswin Nahrendra, Byeongho Yu, and Hyun Myung. Dreamwaq: Learning robust quadrupedal locomotion with implicit terrain imagination via deep reinforcement learning. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 5078–5084. IEEE, 2023.

[34] Abhishek Padalkar, Acorn Pooley, Ajinkya Jain, Alex Bewley, Alex Herzog, Alex Irpan, Alexander Khazatsky, Anant Rai, Anikait Singh, Anthony Brohan, et al. Open x-embodiment: Robotic learning datasets and rt-x models. *arXiv preprint arXiv:2310.08864*, 2023.

[35] Xue Bin Peng, Erwin Coumans, Tingnan Zhang, Tsang-Wei Lee, Jie Tan, and Sergey Levine. Learning agile robotic locomotion skills by imitating animals. *arXiv preprint arXiv:2004.00784*, 2020.

[36] Xue Bin Peng, Ze Ma, Pieter Abbeel, Sergey Levine, and Angjoo Kanazawa. Amp: Adversarial motion priors for stylized physics-based character control. *ACM Transactions on Graphics (ToG)*, 40(4):1–20, 2021.

[37] Allen Z Ren, Justin Lidard, Lars L Ankile, Anthony Simeonov, Pulkit Agrawal, Anirudha Majumdar, Benjamin Burchfiel, Hongkai Dai, and Max Simchowitz. Diffusion policy policy optimization. *arXiv preprint arXiv:2409.00588*, 2024.

[38] Moritz Reuss, Maximilian Li, Xiaogang Jia, and Rudolf Lioutikov. Goal-conditioned imitation learning using score-based diffusion policies. *arXiv preprint arXiv:2304.02532*, 2023.

[39] Milad Shafiee, Guillaume Bellegarda, and Auke Ijspeert. Manyquadrupeds: Learning a single locomotion policy for diverse

quadruped robots. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3471–3477. IEEE, 2024.

[40] Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International conference on machine learning*, pages 2256–2265. PMLR, 2015.

[41] Octo Model Team, Dibya Ghosh, Homer Walke, Karl Pertsch, Kevin Black, Oier Mees, Sudeep Dasari, Joey Hejna, Tobias Kreiman, Charles Xu, et al. Octo: An open-source generalist robot policy. *arXiv preprint arXiv:2405.12213*, 2024.

[42] Jinze Wu, Yufei Xue, and Chenkun Qi. Learning multiple gaits within latent space for quadruped robots. *arXiv preprint arXiv:2308.03014*, 2023.

[43] Ruihan Yang, Zhuoqun Chen, Jianhan Ma, Chongyi Zheng, Yiyu Chen, Quan Nguyen, and Xiaolong Wang. Generalized animal imitator: Agile locomotion with versatile motion prior. *arXiv preprint arXiv:2310.01408*, 2023.

[44] Chong Zhang, Wenli Xiao, Tairan He, and Guanya Shi. Wococo: Learning whole-body humanoid control with sequential contacts. *arXiv preprint arXiv:2406.06005*, 2024.

[45] Ziwen Zhuang, Shenzhe Yao, and Hang Zhao. Humanoid parkour learning. *arXiv preprint arXiv:2406.10759*, 2024.