# Adversarial Skill Embeddings applications in humanoid robotics

Philippe Gratias-Quiquandon

October 3, 2025



# Outline

# Generative Adversarial Architecture

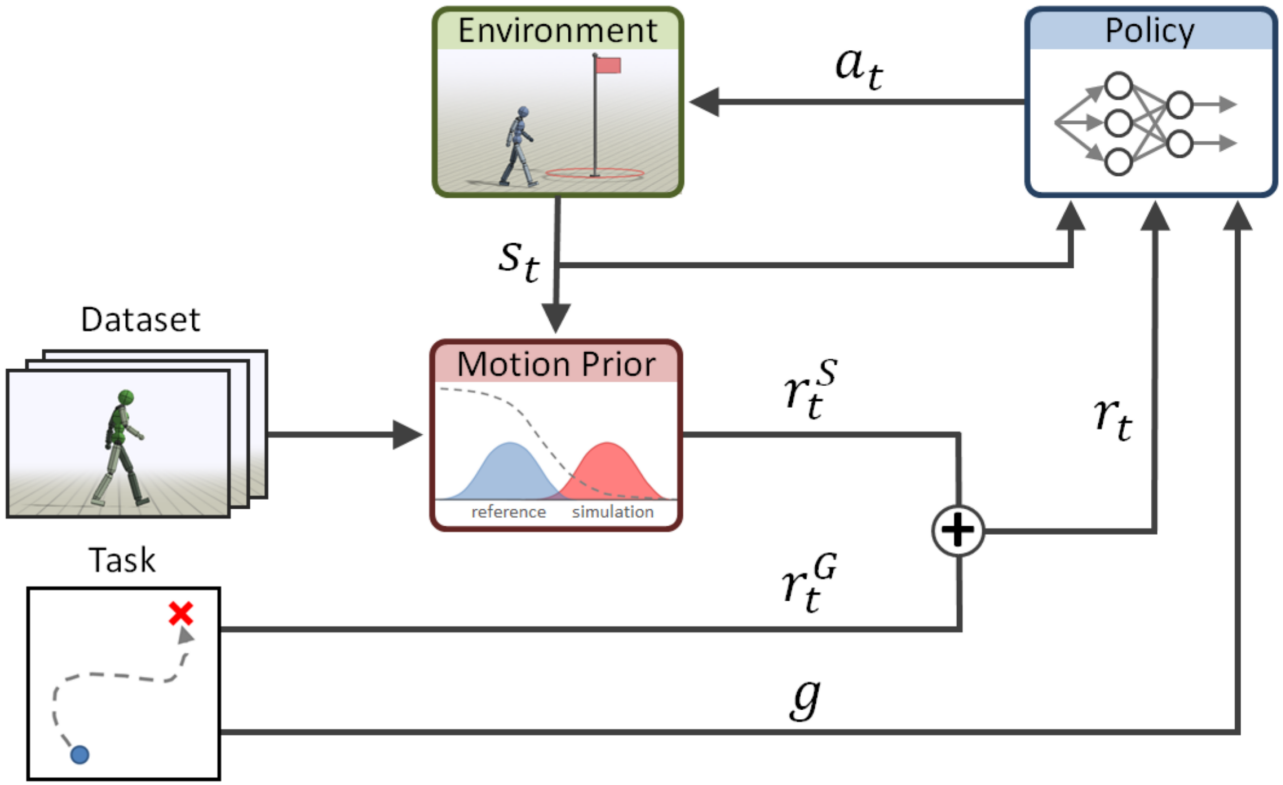



Figure: Adversarial Motion Prior architecture by Peng, Ma, et al. 2021

# Training and Rewards

We train simultaneously the discriminator and the policy with:

- **Discriminator**:

$$\min_D -\mathbb{E}_{d^{\mathcal{M}}}[\log D(s_t, s_{t+1})] - \mathbb{E}_{d^{\pi}}[\log(1 - D(s_t, s_{t+1}))]$$

- **Policy**: We define two rewards,

$$r_t^S = -\log(1 - D(s_t, s_{t+1}))$$
$$r_t^G = \text{any goal reward}$$

At the end, the goal is achieved by copying a style dataset.
New task $\Rightarrow$ New training

# Mixture of Residual Experts

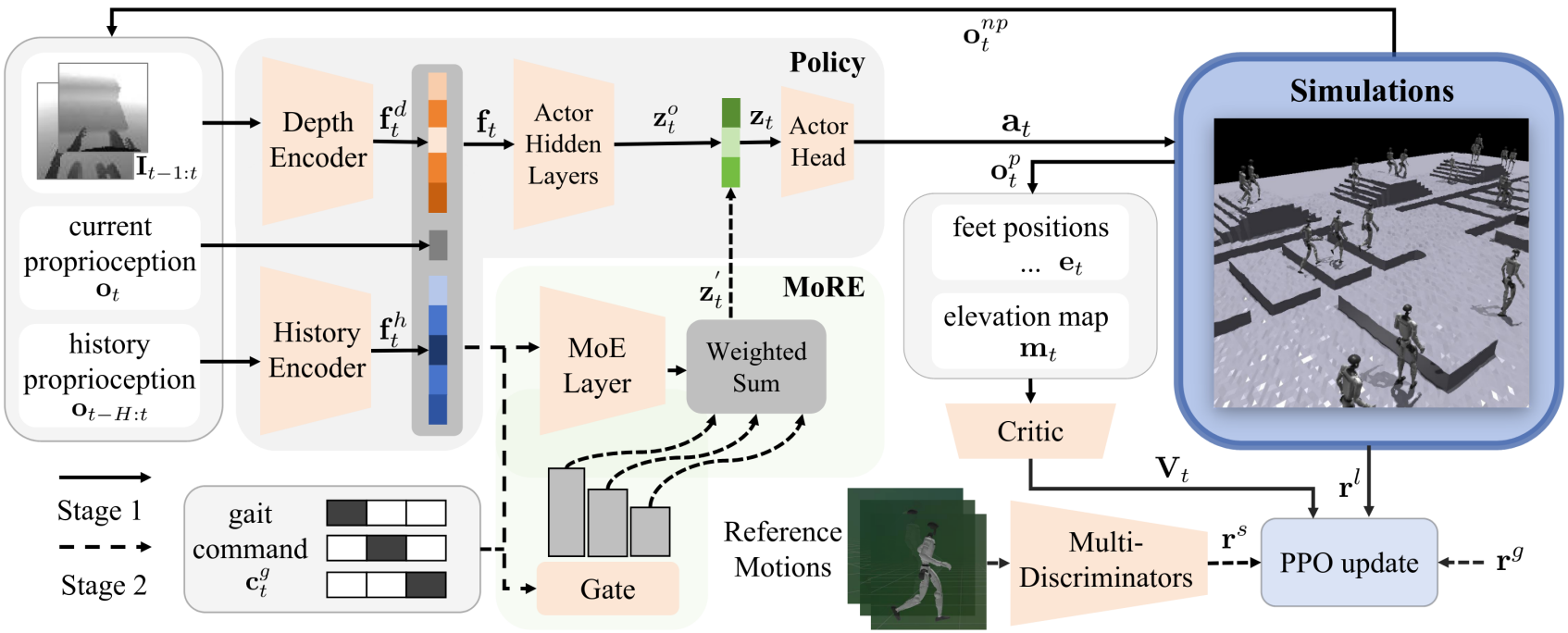



Figure: Mixture of Residual Experts (MoRE) Wang et al. 2025, the gait command chooses an expert policy to use given the states and the camera. The rewards needs to be specified for each gait: $R_t = r_l + \mathbf{1}_{[c_t^g = g]} \left( r_s^{(g)} + r_g^{(g)} \right)$

# The Framework

Adversarial Skill Embeddings (ASE) Peng, Guo, et al. 2022 consists of two main components:

- **Low-level policy**: Takes as input the state observations and a latent skill vector $\mathbf{z}$, and produces motions conditioned on $\mathbf{z}$. The reference motion clips are automatically organized in a latent space[1].
- **High-level policy**: Chooses latent skills $\mathbf{z}$ to pass to the low-level policy, learning to sequence and combine them in order to achieve task objectives.

$\Rightarrow$ Finds general transitions between motions

[1] A latent space is an embedding of a set of items where similar items are positioned closer together.

# Latent space

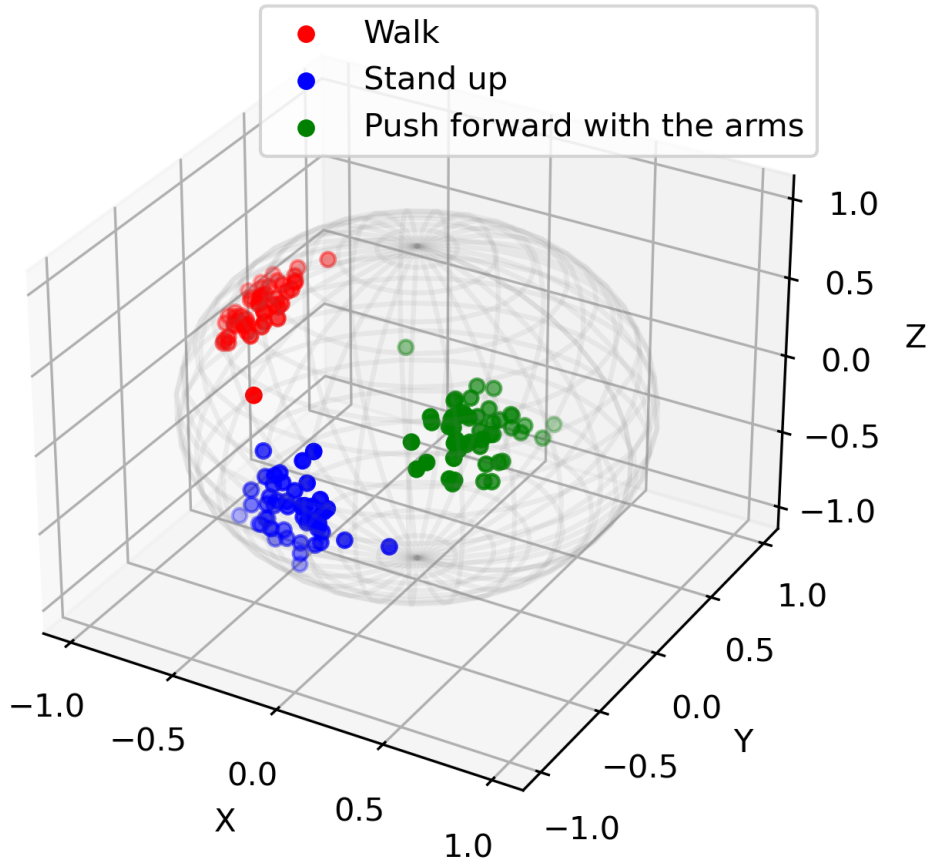



Figure: Illustration of a latent space on a sphere.

# Architecture

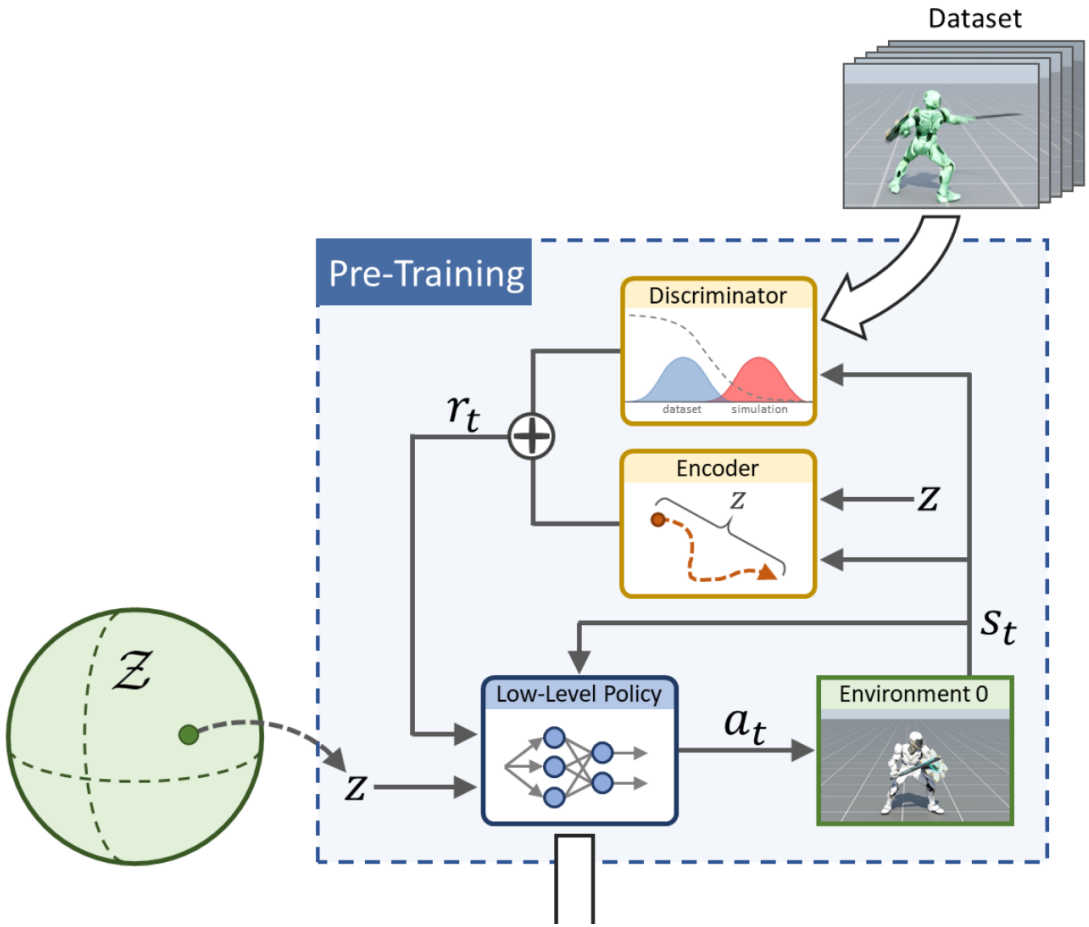



Figure: Architecture of the low-level policy

# Training objective

$$\max_{\pi} \mathbb{E}_{p(\mathbf{z})}\mathbb{E}_{p(\tau|\pi,\mathbf{z})}\left[\sum_{t=0}^{T-1}\gamma^t\left(r_t^S + \beta r_t^E\right)\right]$$
$$- w_{\text{div}}, \mathbb{E}_{d^\pi(\mathbf{s})}, \mathbb{E}_{\mathbf{z}_1,\mathbf{z}_2\sim p(\mathbf{z})}\left[\left(\frac{D_{KL}(\pi(\cdot|\mathbf{s},\mathbf{z}_1),\pi(\cdot|\mathbf{s},\mathbf{z}_2))}{D_{\mathsf{z}}(\mathbf{z}_1,\mathbf{z}_2)} - 1\right)^2\right]$$

The first term is easily recognizable as the cumulative reward and the second term is called the diversity term.

$$r_t^S = -\log(1 - D(s_t, s_{t+1}))$$
$$r_t^E = \kappa\mu_q(s_t, s_{t+1})^\top \mathbf{z}$$

# High-level Policy

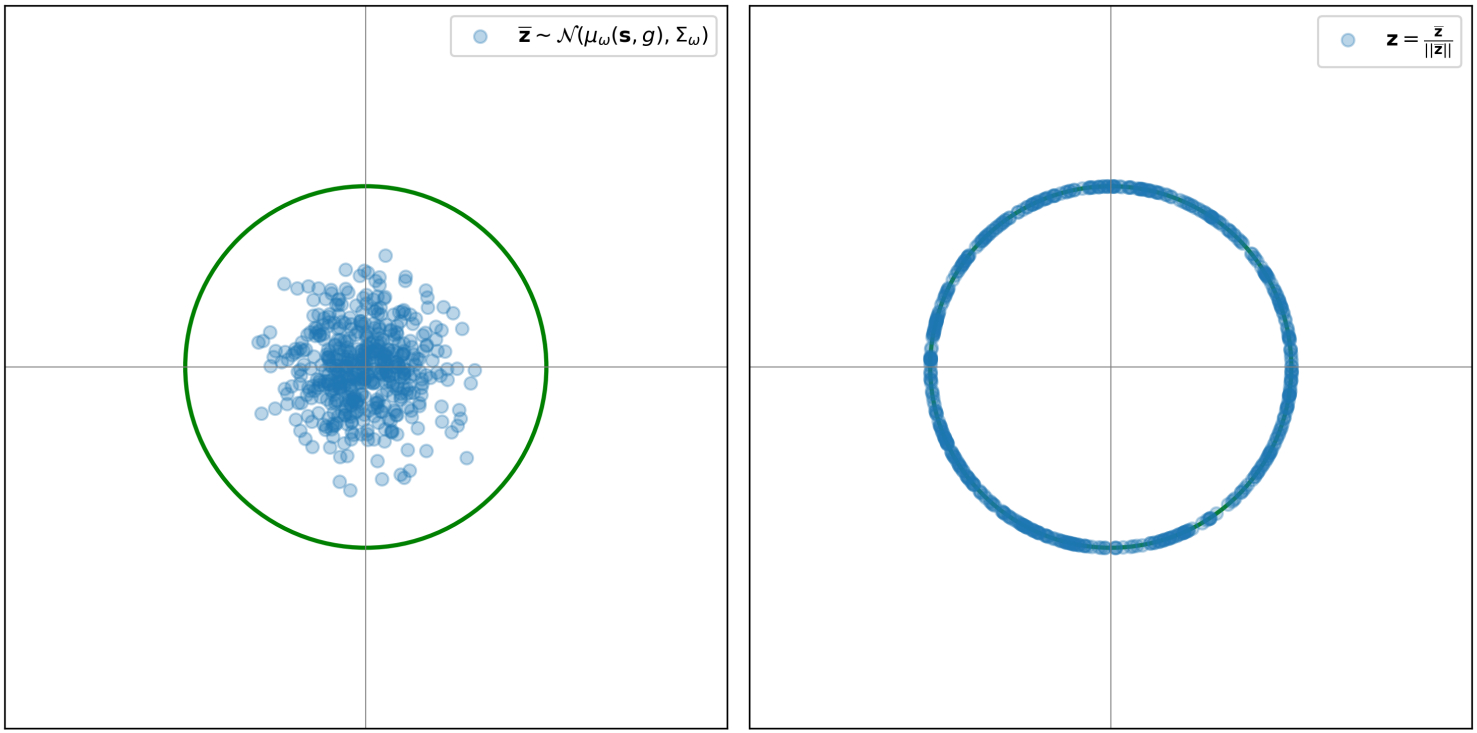



Figure: The high-level policy parameterizes the mean of a Gaussian distribution from which latent variables are sampled. By shifting this mean toward a particular region of the latent space, the policy effectively selects and activates a specific skill.

# High-level Policy

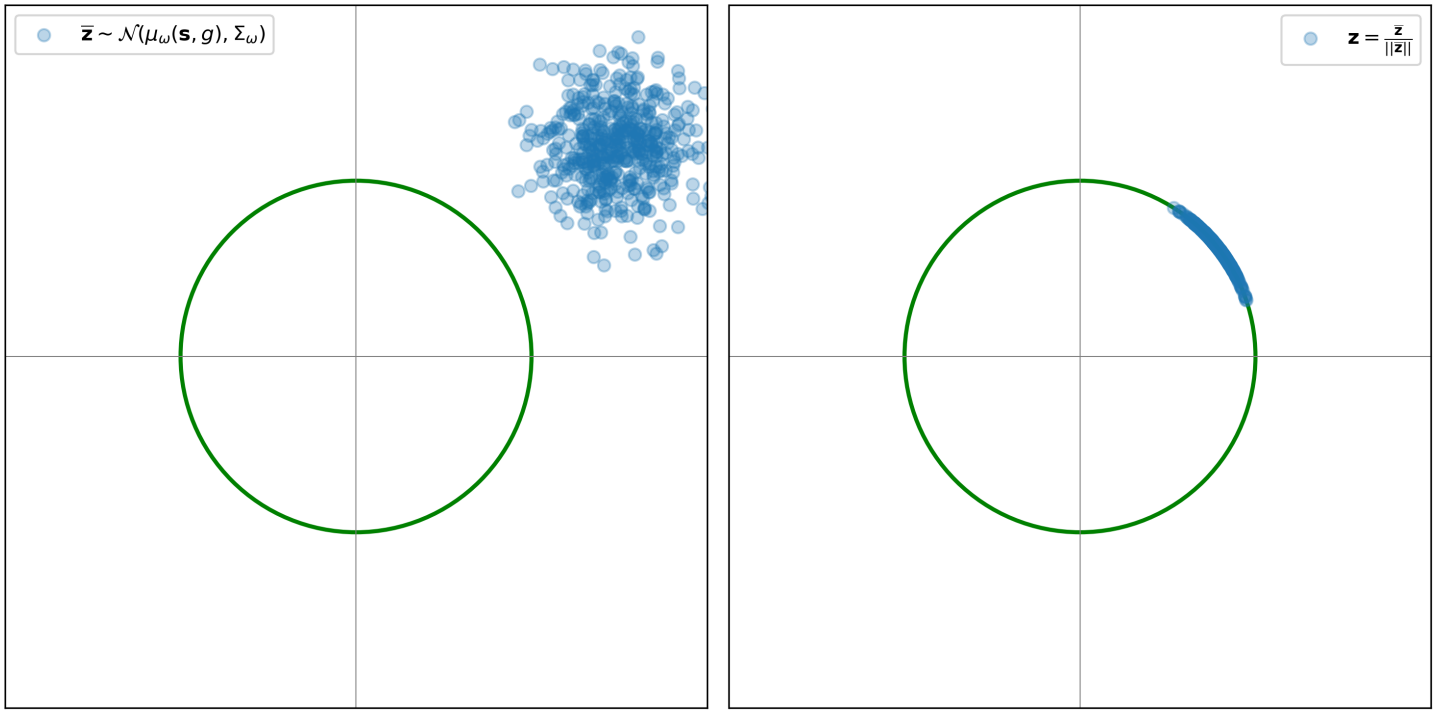



Figure: The high-level policy parameterizes the mean of a Gaussian distribution from which latent variables are sampled. By shifting this mean toward a particular region of the latent space, the policy effectively selects and activates a specific skill.

# Designing a Retargeting Pipeline

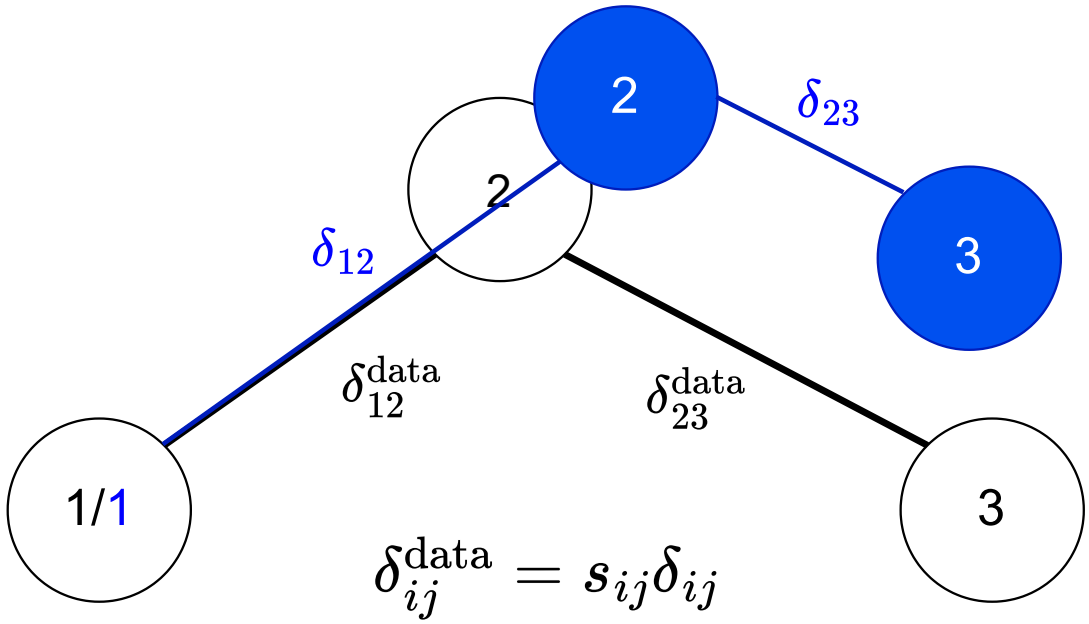


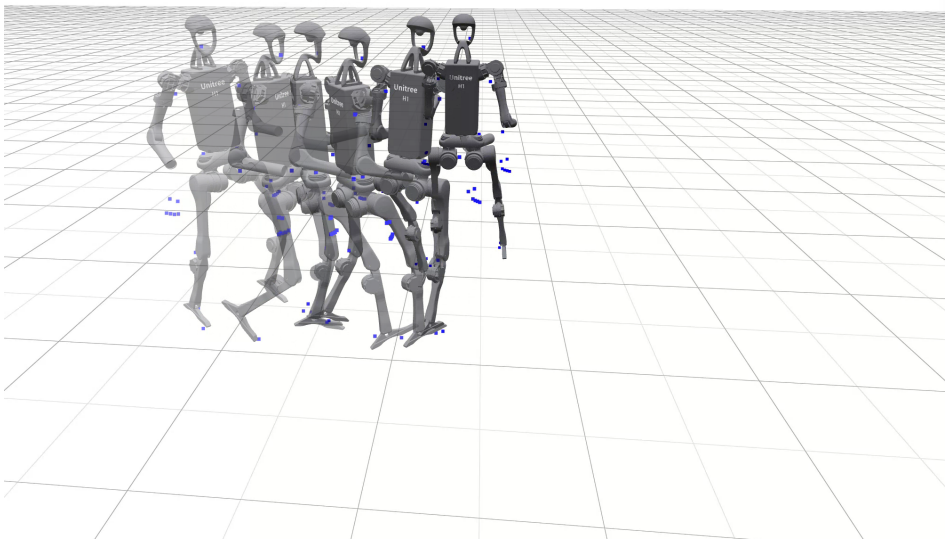

Figure: Alignment between the motion capture skeleton (black) and the robot model (blue). Due to structural differences, an offset exists between corresponding nodes. To achieve optimal retargeting, scaling factors must be computed for each pair of joints to minimize these discrepancies.

# Table of Residuals

| **Residual Name** | **Formula** |
|---|---|
| Global Alignment | $\sum_i w_i - w_i^{\text{data}}$ |
| Local Alignment Position | $\sum_{(i,j)} \left( \delta_{ij}^{\text{data}} - s_{ij}\delta_{ij} \right)$ |
| Local Alignment Angle | $\sum_{(i,j)} 1 - \cos(\theta_{ij})$ |
| Joint Limit | $\sum_i \max(0, c_i - c_{i,\max}) + \max(0, c_{i,\min} - c_i)$ |
| Rest | $\sum_i c_i - c_{i,\text{rest}}$ |
| Smoothness | $\sum_i c_{i,t} - c_{i,t-1}$, where $c_{i,t}$ and $c_{i,t-1}$ are the $i$-th joint angle at time $t$ and $t-1$ |

Table: Residual terms and their mathematical formulations.

# Simple Pipeline and Good Results



# Back to the Diversity Term

$$\max_{\pi} \mathbb{E}_{p(\mathbf{Z})}\mathbb{E}_{p(\tau|\pi,\mathbf{Z})}\left[\sum_{t=0}^{T-1}\gamma^t\left(r_t^S+\beta r_t^E\right)\right]$$
$$-w_{\text{div}},\mathbb{E}_{d^\pi(\mathbf{s})},\mathbb{E}_{\mathbf{z}_1,\mathbf{z}_2\sim p(\mathbf{z})}\left[\left(\frac{D_{KL}(\pi(\cdot|\mathbf{s},\mathbf{z}_1),\pi(\cdot|\mathbf{s},\mathbf{z}_2))}{D_{\mathbf{z}}(\mathbf{z}_1,\mathbf{z}_2)}-1\right)^2\right]$$

To accelerate training, we have to reduce the number of latent dimensions. When we do so, the diversity term explodes during training $\Rightarrow$ remove it !

# Why does the Diversity Term explode?

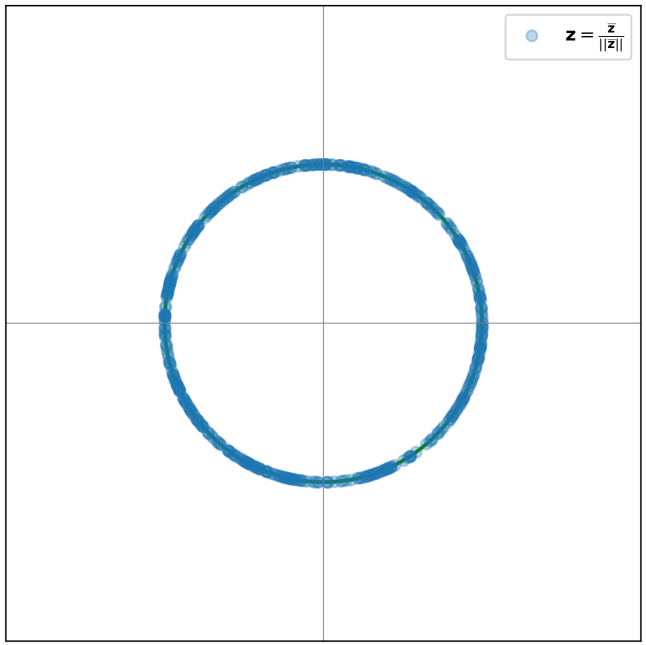



Figure: During the training, latent variables are sampled uniformly on the latent space. The probability of getting two close latent variables decreases as the number of latent dimensions increases. $\mathbb{E}[\langle \mathbf{z}_1, \mathbf{z}_2 \rangle^2] = \frac{1}{d}$.

# A Simple Experiment: Forward and Backward

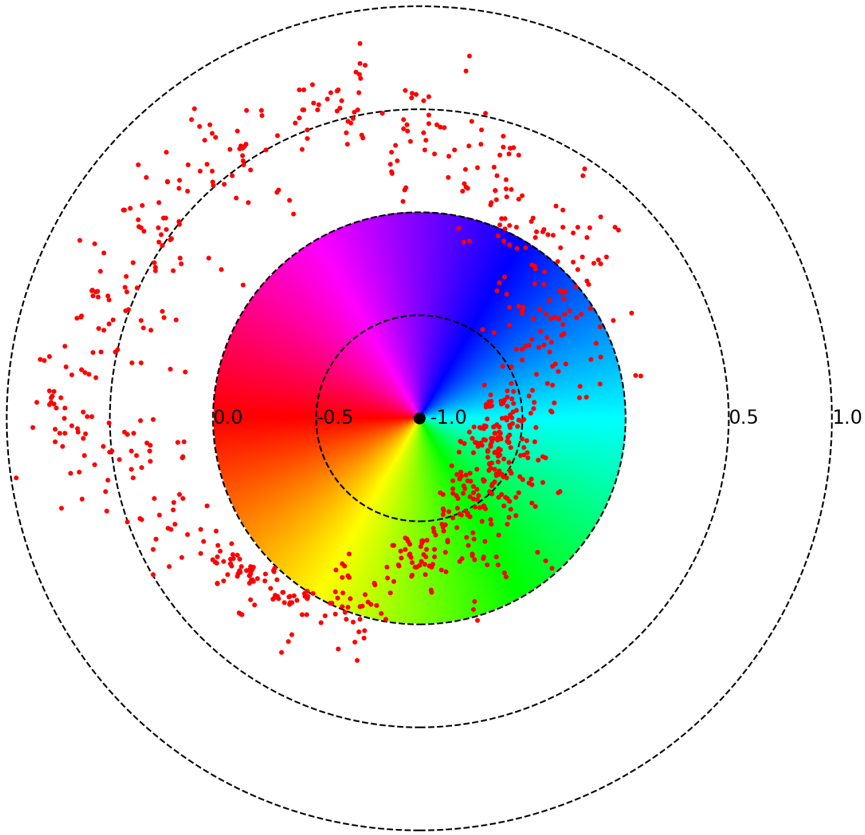
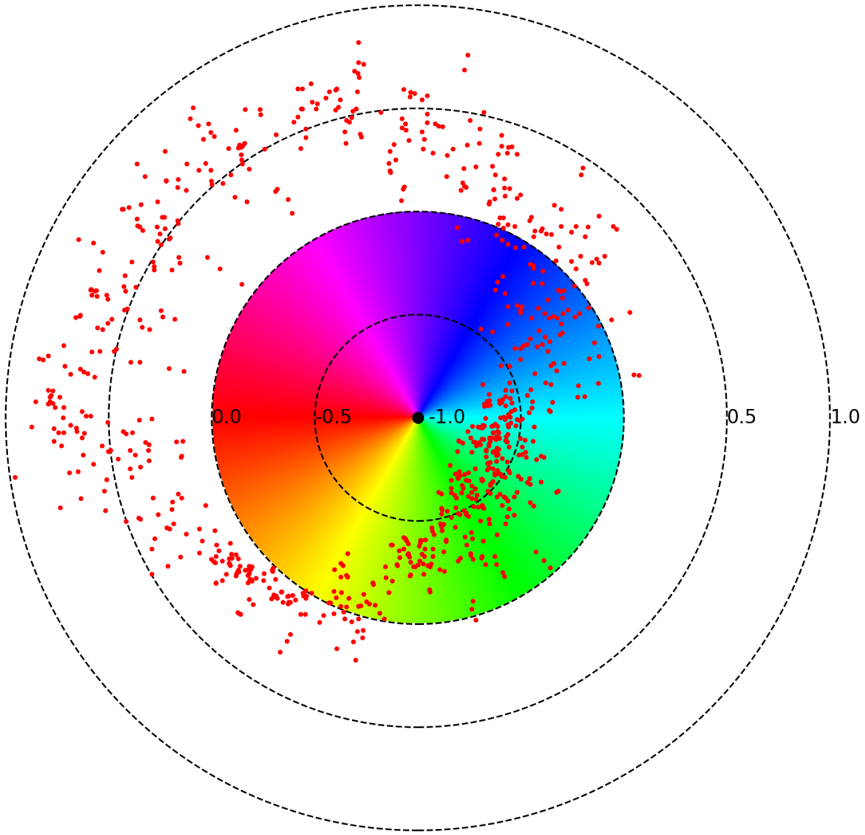
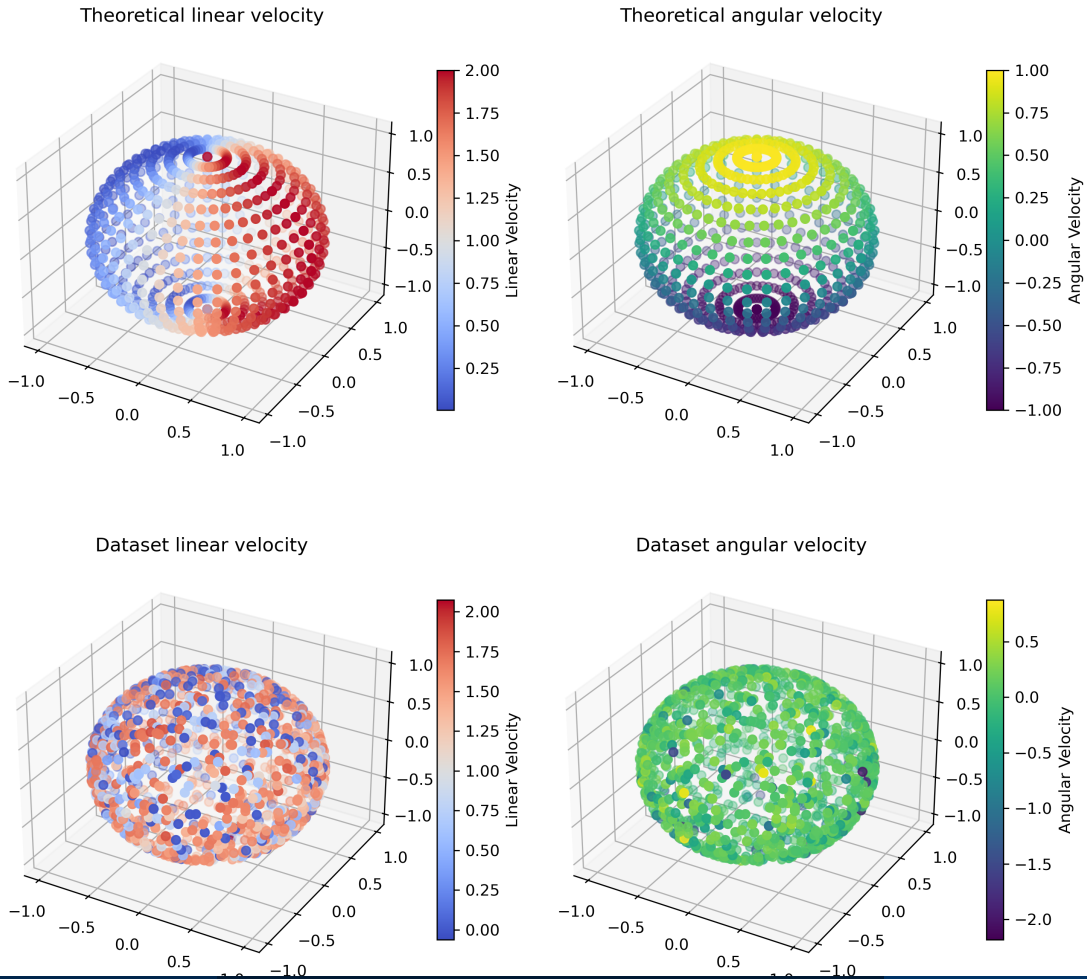



Figure: Color wheel representation of the latent space. The red points indicate measured velocities for sampled latent vectors.

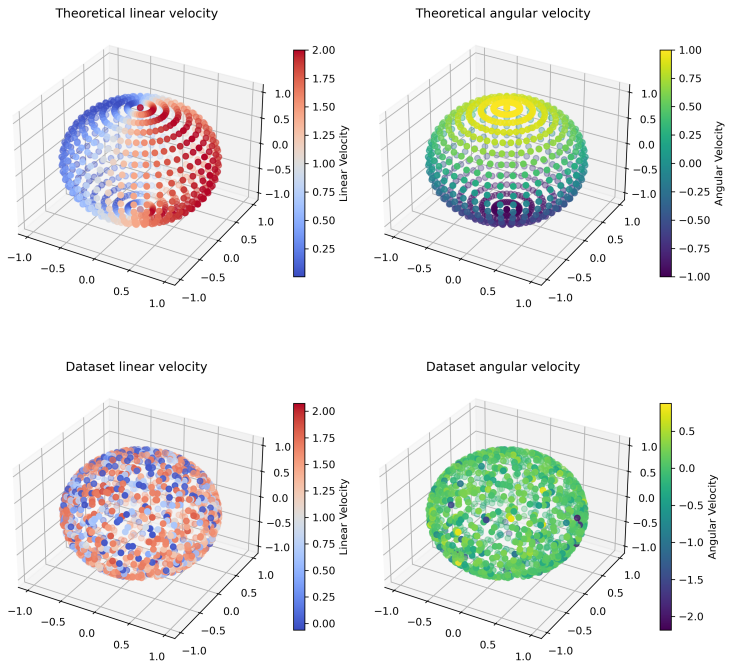
# A Simple Experiment: Forward and Backward



# Walking and Turning Policy



# Walking and Turning Policy



# Influence of the Latent Dimension

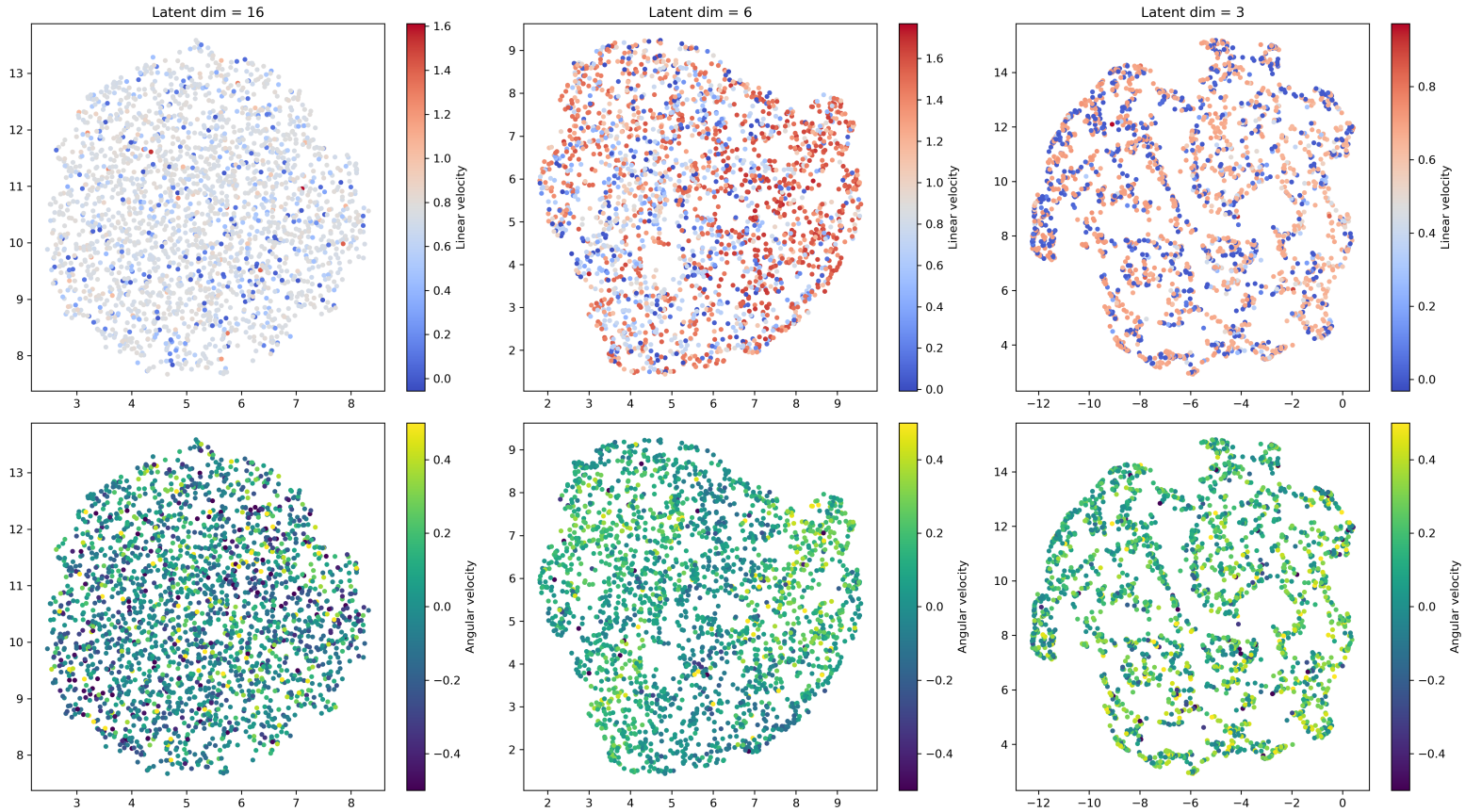
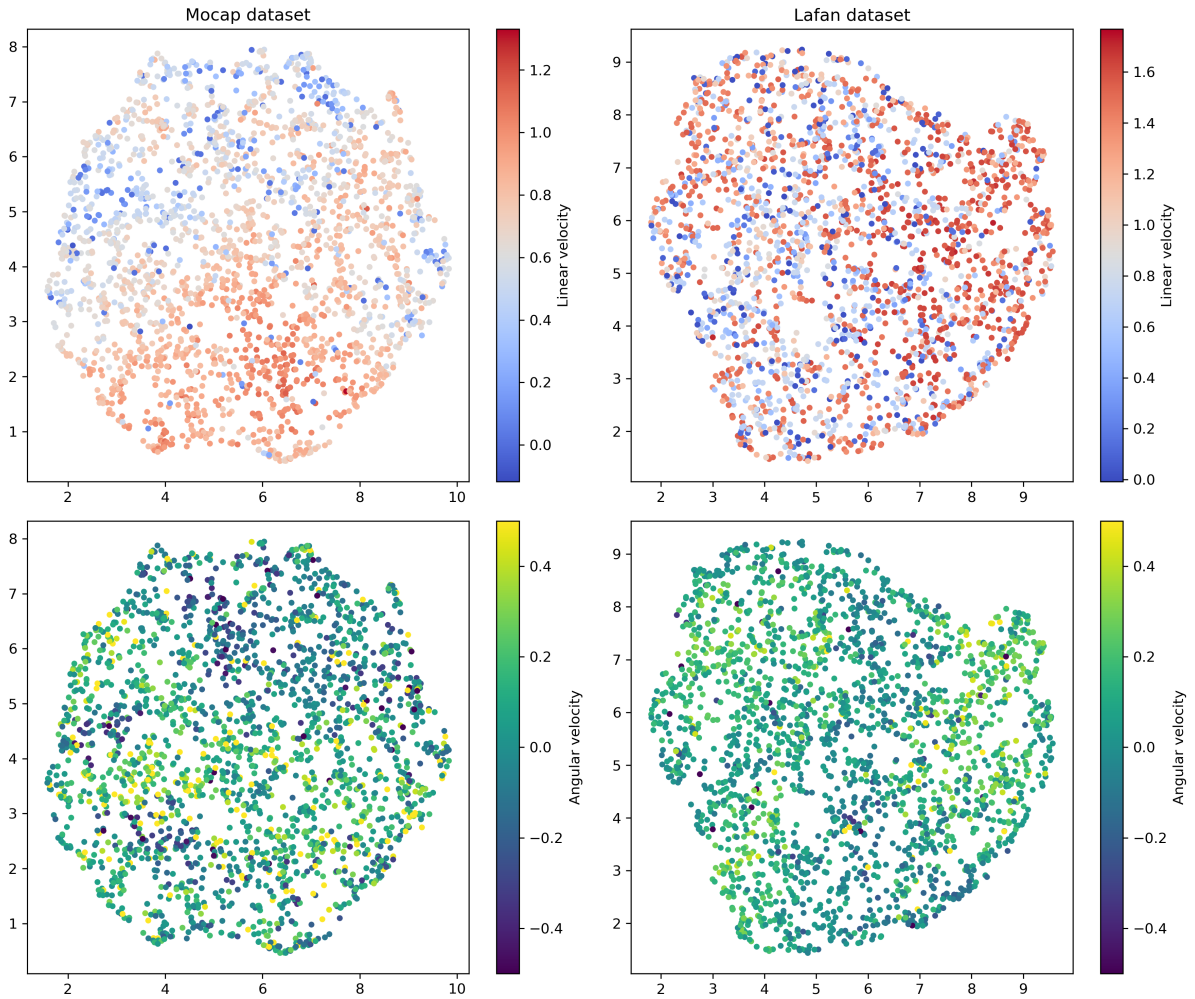
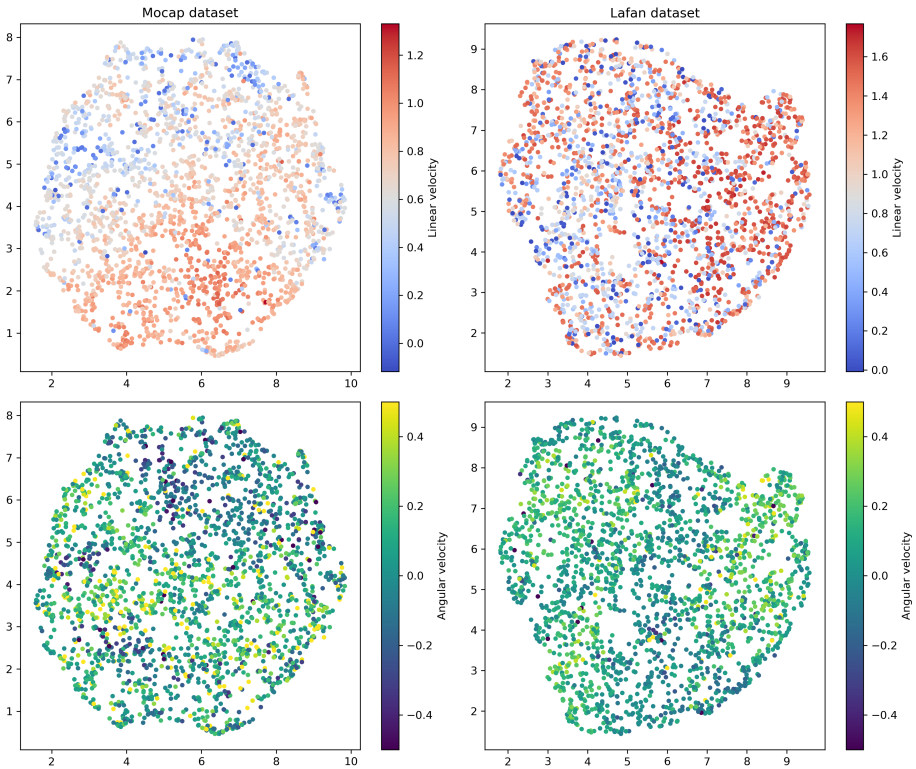



Figure: UMAP projections of latent spaces extracted from three different policies trained on the LAFAN1 dataset

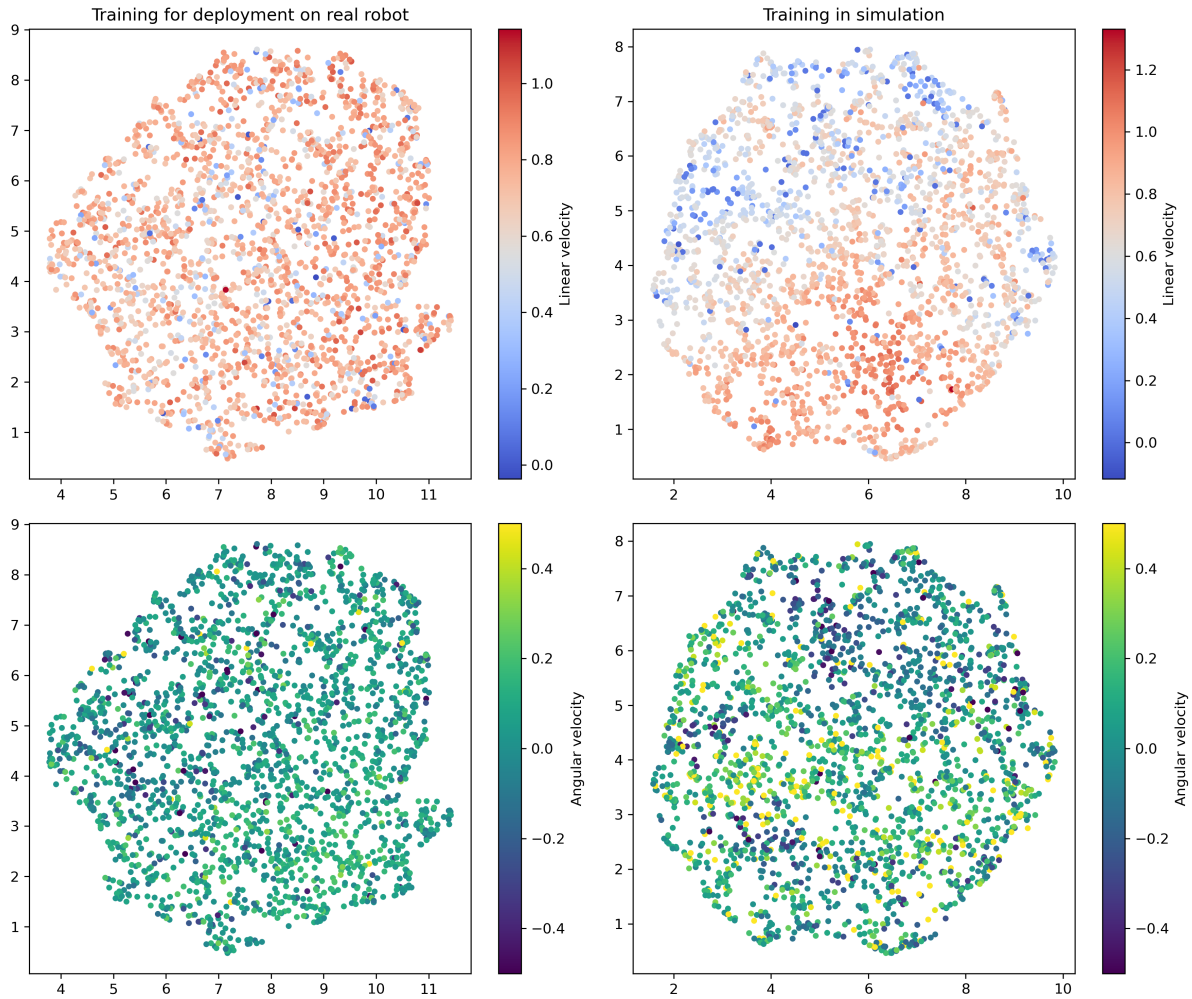
# Influence of the Dataset



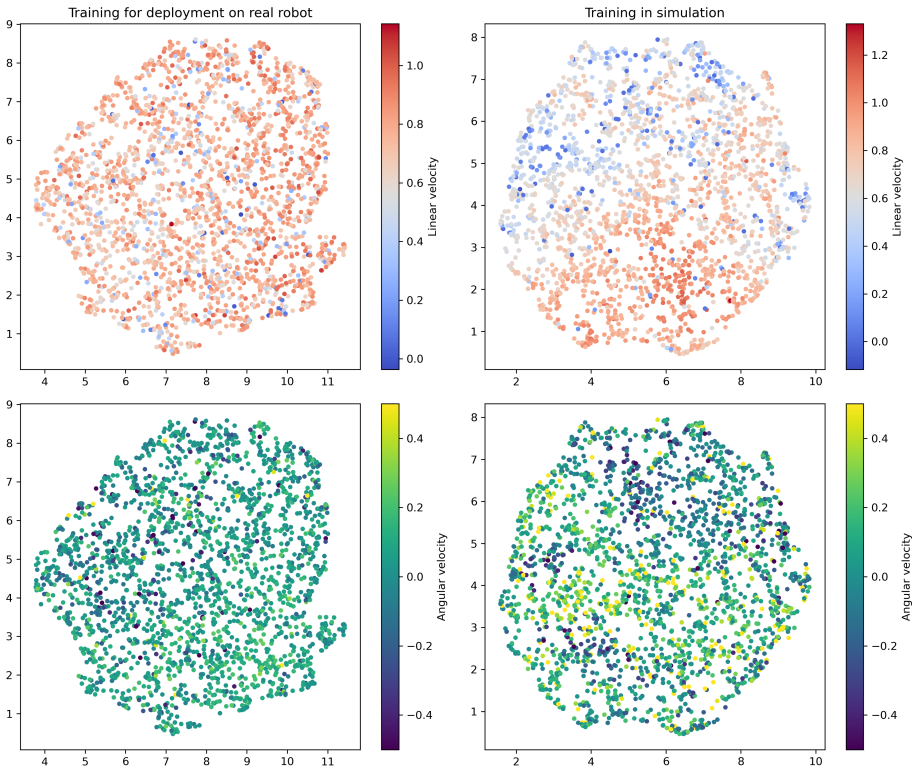
# Influence of the Dataset



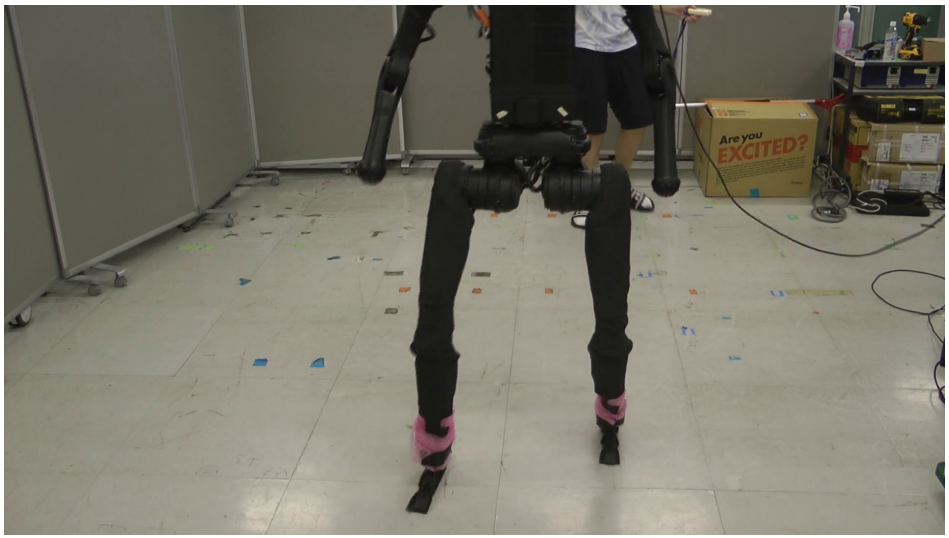
# Influence of the Sim-to-Real Gap



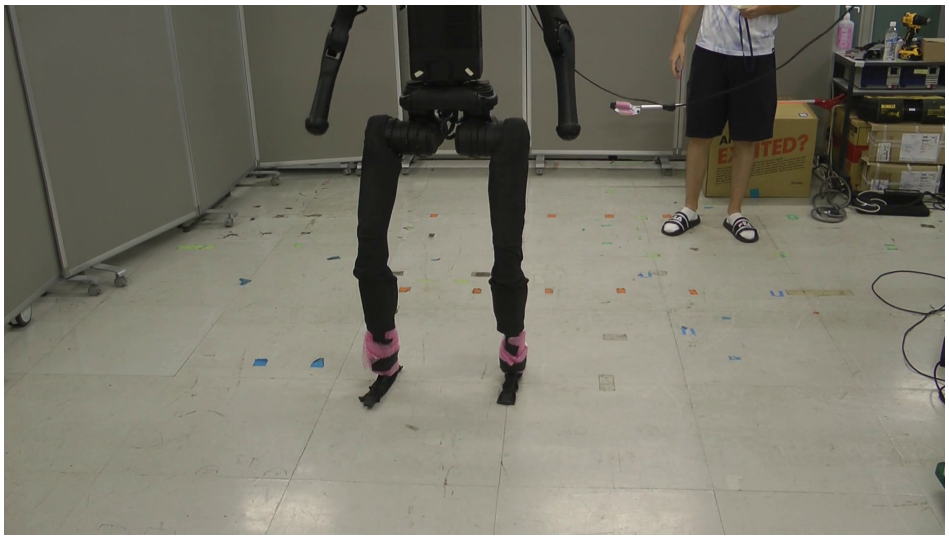
# Influence of the Sim-to-Real Gap



# Human-like Walking on the Real H1 Robot



# Human-like Walking on the Real H1 Robot



# Applying ASE to a Door Opening Task

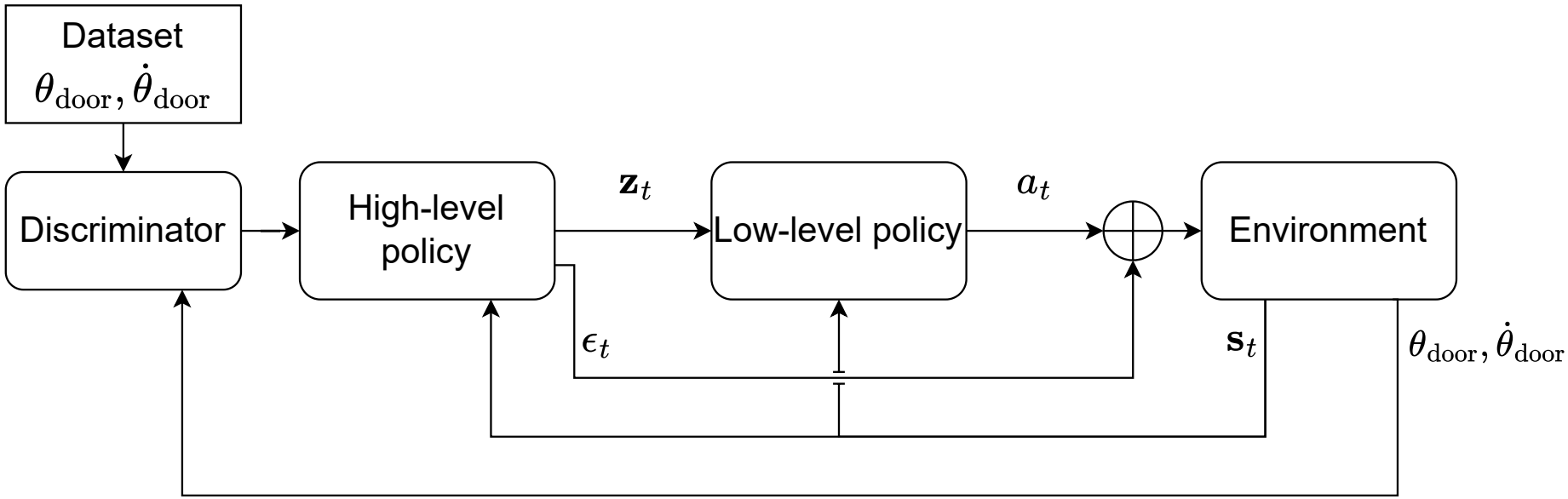
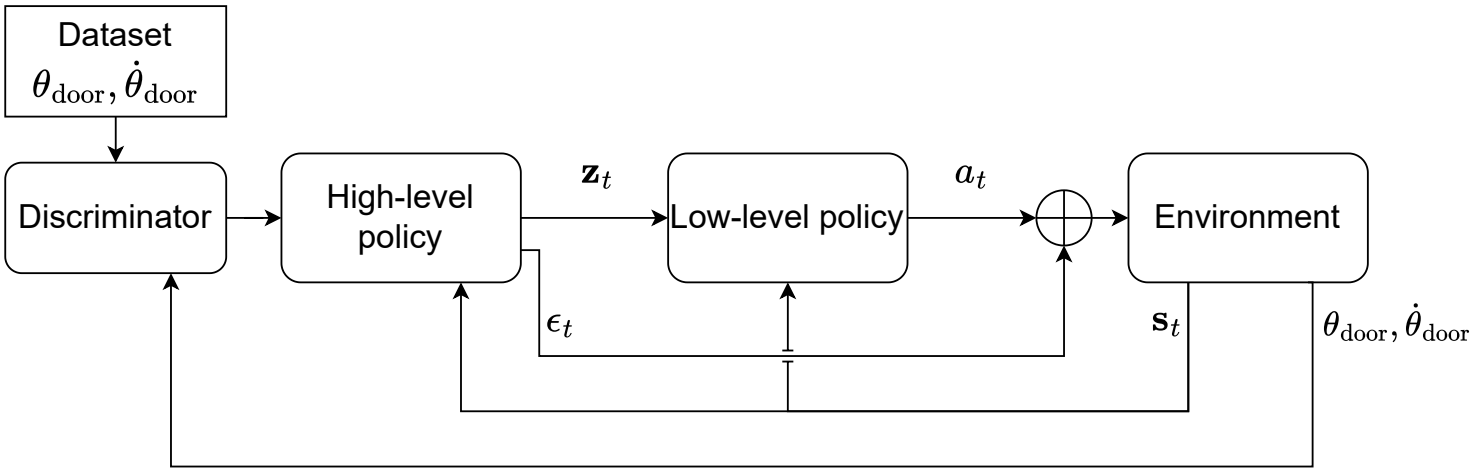
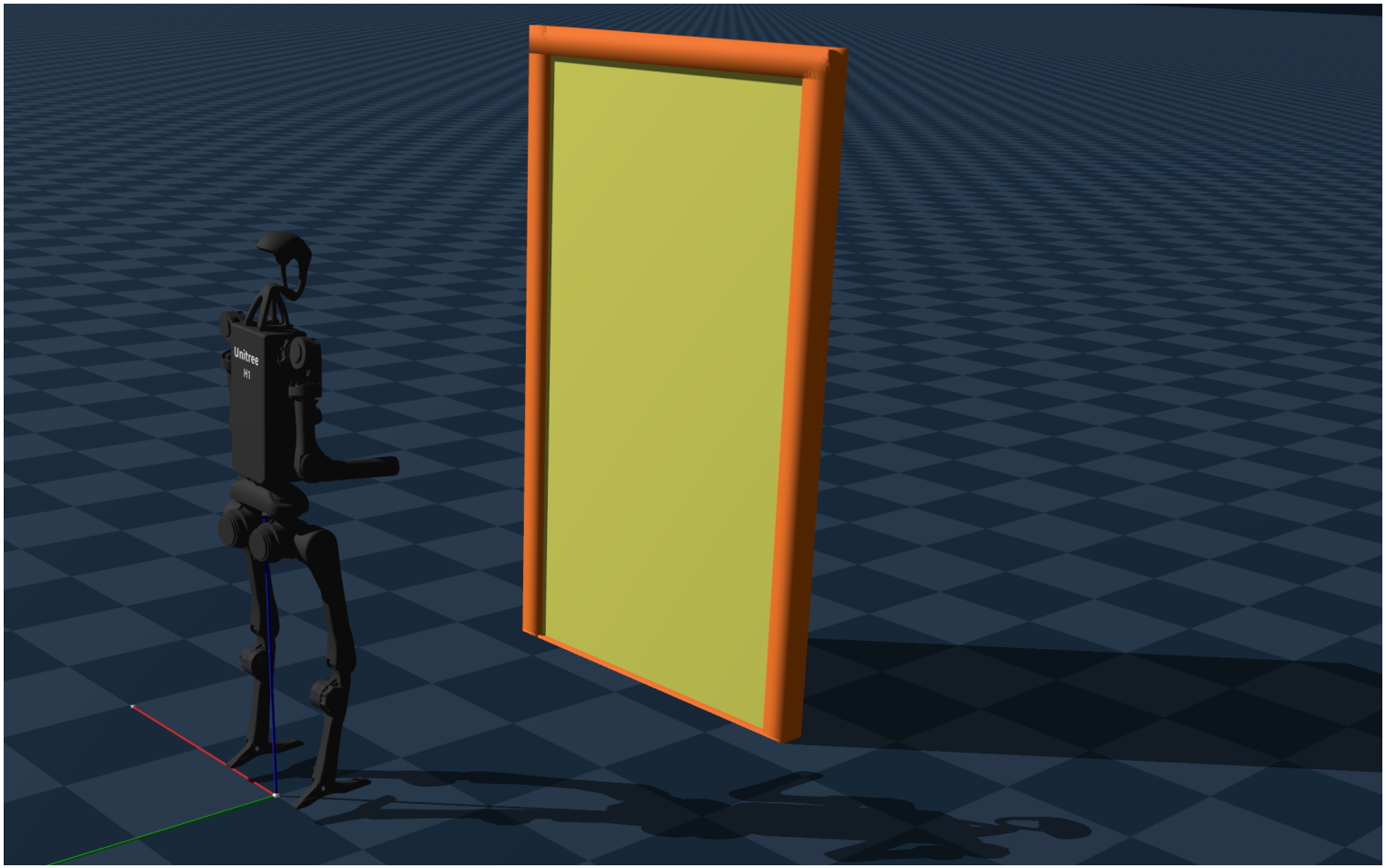



Figure: Modified ASE pipeline for the door-opening task. The high-level policy generates latent variables $\mathbf{z}_t$ for the low-level controller, which outputs actions $a_t$ along with a residual correction $\epsilon_t$. The final executed action is $a_t + \epsilon_t$. A door-specific discriminator provides imitation rewards by comparing the simulated motion with Mocap demonstrations of door opening.

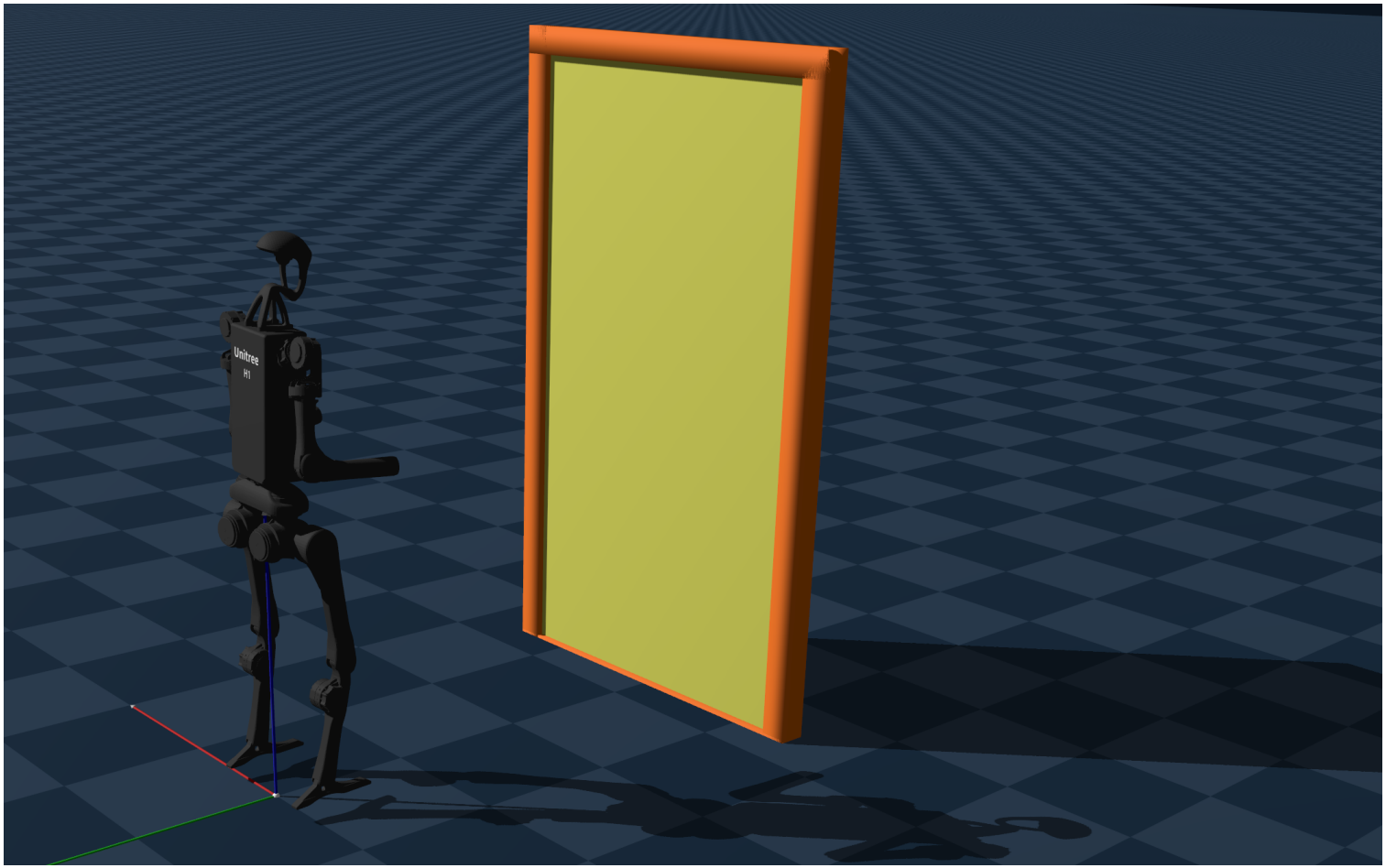
# Applying ASE to a Door Opening Task



# Classical Imitation Pipeline



# Classical Imitation Pipeline



# Conclusion and Perspectives

- Applied **Adversarial Skill Embeddings** to humanoid locomotion with the Unitree H1 robot.
- Investigated:
  - Latent dimensionality and skill representation
  - Role of diversity loss
  - Retargeting pipeline for motion capture datasets
- ASE produces smooth, human-like walking transitions, but is sensitive to noise and lacks scalability.
- **Diffusion-based policies** Liao et al. 2025 surpass ASE in robustness and multi-skill generalization.
- Promising directions: **Forward-Backward Representations** Touati and Ollivier 2021 for generalizable multi-skill locomotion policies.

# References I


- Liao, Qiayuan et al. (2025). *BeyondMimic: From Motion Tracking to Versatile Humanoid Control via Guided Diffusion*. arXiv: 2508.08241 [cs.RO]. URL: https://arxiv.org/abs/2508.08241.
- Peng, Xue Bin, Yunrong Guo, et al. (July 2022). "ASE: large-scale reusable adversarial skill embeddings for physically simulated characters". In: *ACM Transactions on Graphics* 41.4, pp. 1–17. ISSN: 1557-7368. DOI: 10.1145/3528223.3530110. URL: http://dx.doi.org/10.1145/3528223.3530110.
- Peng, Xue Bin, Ze Ma, et al. (July 2021). "AMP: adversarial motion priors for stylized physics-based character control". In: *ACM Transactions on Graphics* 40.4, pp. 1–20. ISSN: 1557-7368. DOI: 10.1145/3450626.3459670. URL: http://dx.doi.org/10.1145/3450626.3459670.

# References II


- Touati, Ahmed and Yann Ollivier (2021). *Learning One Representation to Optimize All Rewards*. arXiv: 2103.07945 [cs.LG]. URL: https://arxiv.org/abs/2103.07945.
- Wang, Dewei et al. (2025). *MoRE: Mixture of Residual Experts for Humanoid Lifelike Gaits Learning on Complex Terrains*. arXiv: 2506.08840 [cs.RO]. URL: https://arxiv.org/abs/2506.08840.