EmergentScaleMotion (ESM) Proposal

Yuer Tang, Jiayuan Mao

February 14, 2026

Motivation

Existing imitation learning methods primarily focus on coarse-grained policies (e.g., "open the door"), while fine-grained control (e.g., "how much to open the door") is often overlooked. However, controllability is crucial for two reasons:

1. **Improved task performance.** Control over the scale of an action can directly affect task success. In the door-opening example, if an obstacle behind the door prevents it from opening fully, the policy should be able to adjust the opening angle accordingly.
2. **Sequential task composition.** Fine-grained control enables flexible composition of subtasks. For instance, consider a composite manipulation task such as open the door halfway, place an object inside, and then close it. Each stage requires continuous control over the degree of opening to ensure smooth and coordinated behavior.

Problem Formulation

We aim to construct a compact, smooth, and interpretable parameterization of policies such that a small number of parameters can control the overall scale of policies (e.g., door-opening angle or motion speed). Specifically:

- **Compactness** A low-dimensional representation proviees easy meaning ful specificvtions for humans and efficient sampling for machines.
- **Smoothness** Each parameter is a clear, continuous factor of variation. The learned representation should allow continuous control over policy scales, supporting gradual adjustments rather than discrete switches.
- **Interpretability** The representation should be disentangled, with scale-related parameters explicitly represented so that humans can understand and manipulate them in downstream tasks.

Then for downstream tasks, enables policy control by simply adjusting parameters.

Methods

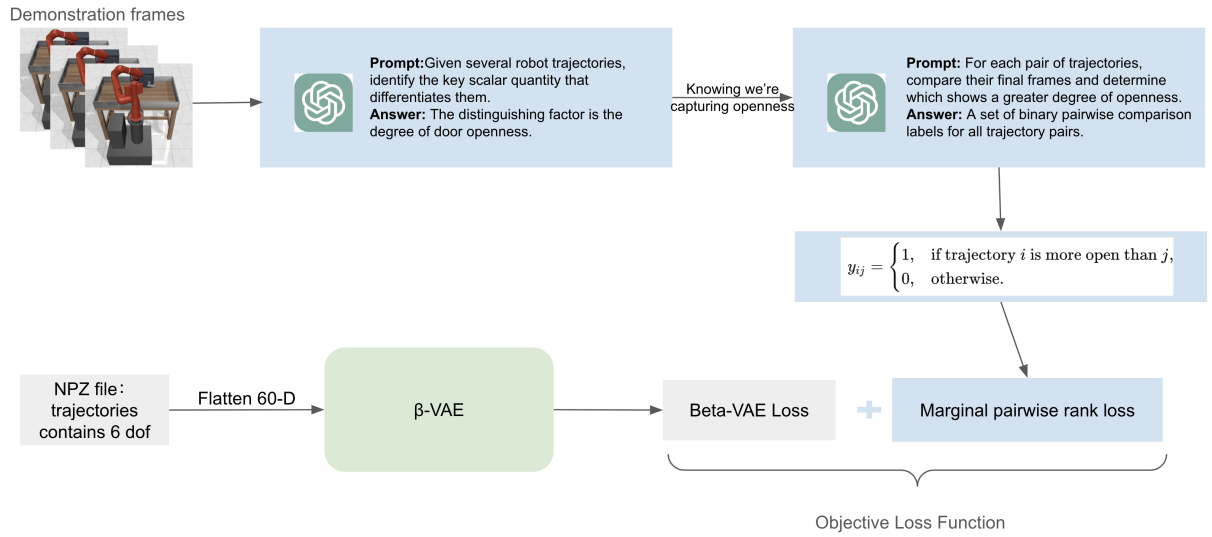

Figure 1: Architecture

We develop a two-part framework for learning compact, controllable policy representations.

Part A: β-RankVAE (demo-ready). A β-VAE with a pairwise ranking objective where beta value will suggest disentangled of the paramaters while learns a latent axis z_r aligned with a task scale (e.g., door openness or motion speed). This yields a compact, smooth, and interpretable control parameter.

The total loss is defined as:

$$\mathcal{L} = \mathcal{L}_{\text{recon}} + \beta \mathcal{L}_{\text{KL}} + \lambda_{\text{rank}} \mathcal{L}_{\text{rank}}. \tag{1}$$

While the marginal pairwise rank loss is defined as:

$$\mathcal{L}_{\text{rank}} = \frac{1}{|\mathcal{P}|} \sum_{(i,j)\in\mathcal{P}} \max\big(0,\, \alpha \cdot \max(|y_i - y_j|, \text{min\_gap}) - (z_i - z_j)\big), \qquad \mathcal{P} = \{(i,j) : y_i > y_j\}. \tag{2}$$

Part B: LLM-Assisted Scale Perception (will start soon). A vision-enabled LLM module serves two roles: (i) *Scale recognizer*—identify which scale is present in a dataset (openness vs. speed vs. distance, etc.); (ii) *Pairwise comparator*—given two trajectory snippets, judge which exhibits the greater value of the detected scale. The resulting pairwise labels are aggregated into scalar targets and used to supervise z_r.

Preliminary Results

At this stage, the LLM-assisted module has not yet been integrated. We first evaluate the β-RankVAE component alone on the *Meta-World SawyerDoorOpen-v3* environment to establish a baseline for compact, interpretable motion representation.

Dataset and Setup

We collected 2,000 door-opening trajectories using continuous openness labels as supervision. Each trajectory records 500 steps of 6D end-effector poses (x, y, z, quat), resampled to $T = 50$ timesteps via nearest-neighbor

interpolation. The openness scalar label $o \in [0,1]$ is computed as the normalized angular displacement of the door hinge. The encoder is a 128-hidden-layer MLP with an 8-dimensional latent space.

Training uses $\beta = 4$ and the total loss:

$$\mathcal{L} = \mathcal{L}_{\text{recon}} + \beta \mathcal{L}_{\text{KL}} + \lambda_{\text{rank}} \mathcal{L}_{\text{rank}}.$$

This section summarizes the model's training dynamics, examines the relationship between the latent variable z_r and the openness scale to verify whether z_r captures the intended motion factor, evaluates reconstruction performance to confirm the model's ability to reproduce trajectories, analyzes generative interpolation to test whether z_r continuously controls motion scale, and finally assesses physical feasibility through simulation of both reconstructed and generated trajectories.

Training Dynamics

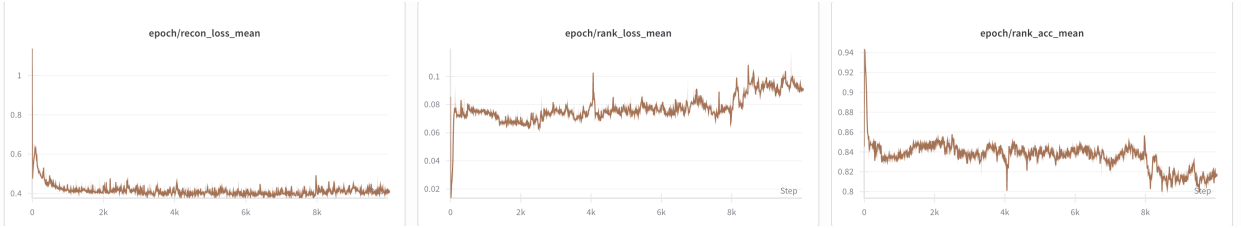

Figure 2: Training curves for reconstruction loss, rank loss, and rank accuracy.

As shown in Figure 2, the reconstruction loss steadily decreases and stabilizes around 0.4, suggesting that the model effectively encodes the door-opening motion. In contrast, the rank loss increases while the mean rank accuracy decreases over training, indicating that the ranking supervision has not yet successfully aligned the latent dimension with the openness scale.

Latent Representation Analysis

In training, we designate z_0 as the latent dimension corresponding to door openness. We visualize z_0 against the openness labels (Figure 3) and observe a weak positive correlation—larger z_0 values generally correspond to greater opening angles, but with considerable noise. This incomplete alignment is consistent with the observed high rank loss and low rank accuracy, indicating that the ranking supervision has not yet effectively disentangled the openness factor. Adjusting β alone did not significantly improve this correlation.

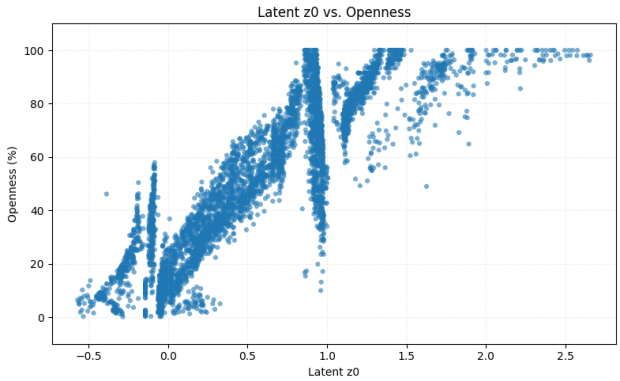

Figure 3: Correlation between latent variable z_0 and openness label o.

Reconstruction Performance

Before evaluating controllability, we first test whether the model can reliably reconstruct observed trajectories. This establishes a baseline for assessing the representation quality. We compare reconstructed and ground-truth trajectories in Cartesian space (Figure 4, 5). Most reconstructions capture the correct hinge motion with small deviations; however, slight endpoint drift can occasionally cause task failure during simulation. The average reconstruction error is MSE ≈ 0.40 on training data and 0.59 on the test set.

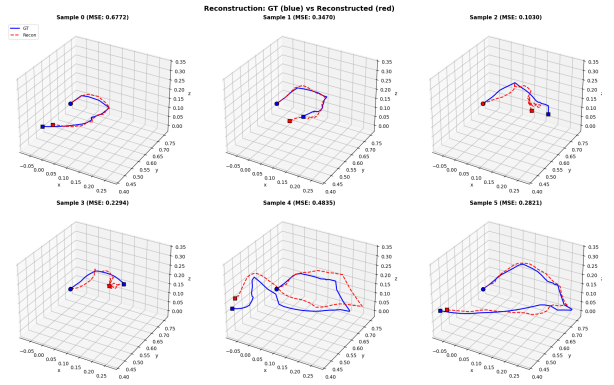

Figure 4: Training reconstructions. The solid blue line shows the ground-truth trajectory, and the dashed red line shows the model's reconstruction for the first six trajectories.

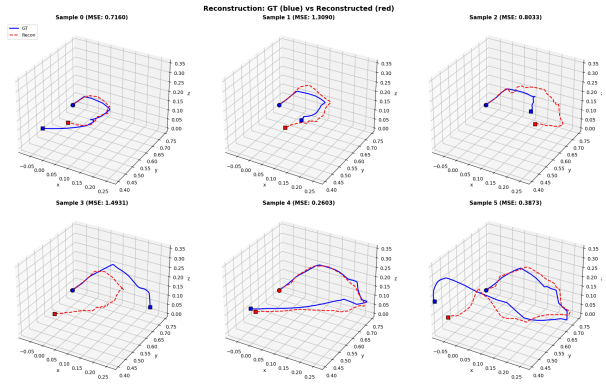

Figure 5: Test reconstructions. The solid blue line shows the ground-truth trajectory, and the dashed red line shows the model's reconstruction for the first six trajectories.

While several reconstructions successfully reproduce door-opening motions, others fail to make proper contact with the handle. This indicates that even when global trajectory shapes appear accurate, small spatial errors—especially near critical contact regions—can disrupt physical feasibility in simulation.

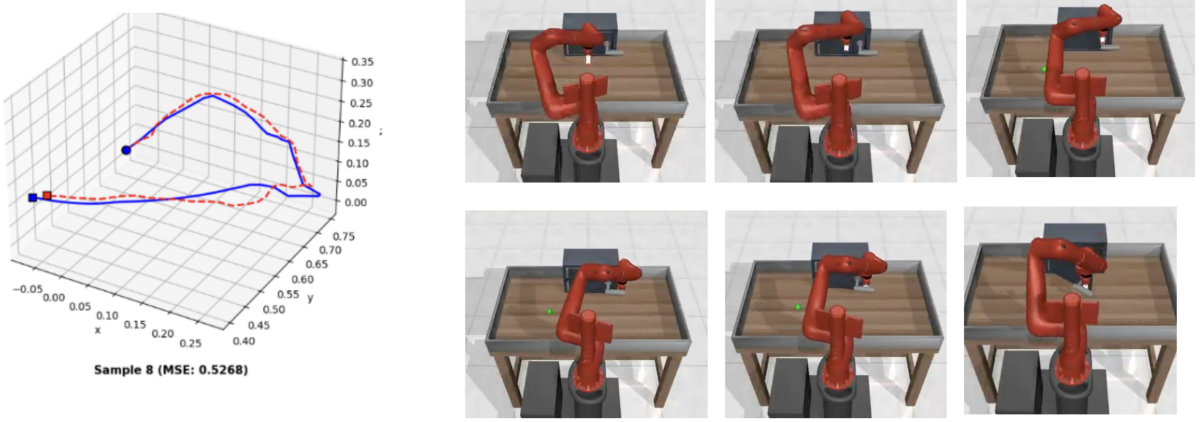

Figure 6: Simulation verification of reconstructed trajectories. The left panel shows the reconstructed 3D trajectory, and the right sequence displays six key snapshots during execution.

Generative Interpolation

To verify whether z_0 truly controls the degree of door openness, we interpolated z_0 across quantiles $[0.1, 0.2, \ldots, 0.9]$ and decoded the corresponding trajectories (Figure 7, 8). The generated motions exhibit a smooth and monotonic

increase in door-opening angle, indicating that z_0 encodes a physically meaningful scale of motion.

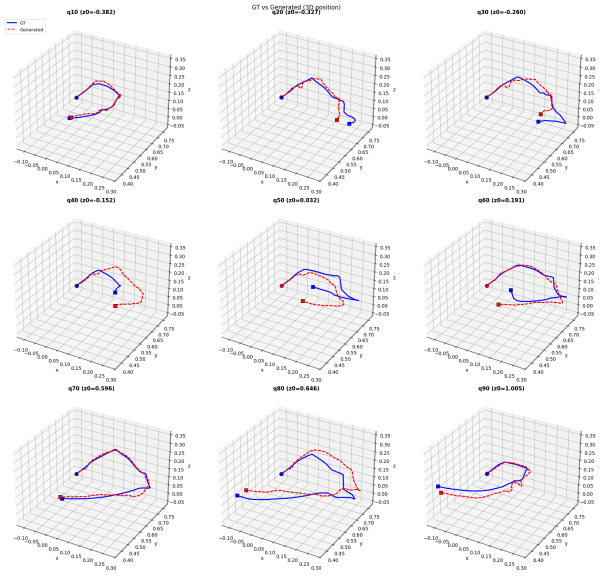

Figure 7: Generated trajectories from latent interpolation on the training set.

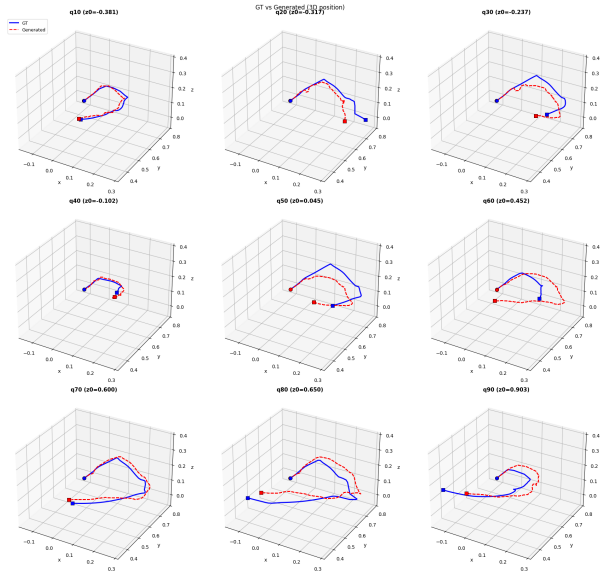

Figure 8: Generated trajectories from latent interpolation on the test set.

To illustrate the effect in simulation, Figure 9 and 10 show two representative generations at $z_0 = 0.1$ and $z_0 = 0.6$. Each example pairs a 3D visualization of the decoded trajectory (left) with its physical execution in the simulator (right). As z_0 increases, the door-opening motion scales smoothly from a slight to a wide opening, confirming interpretable latent control.

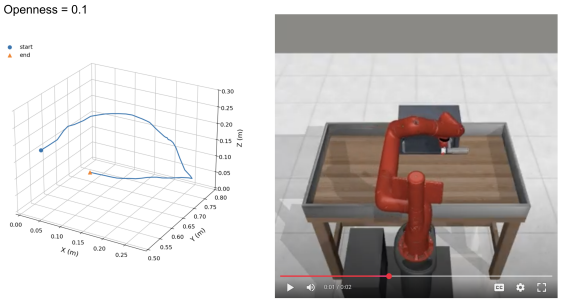

Figure 9: Controlled generation at $z_0 = 0.1$. **Left:** Generated trajectory in Cartesian space. **Right:** Simulation playback showing a small door-opening motion, demonstrating fine-grained control.

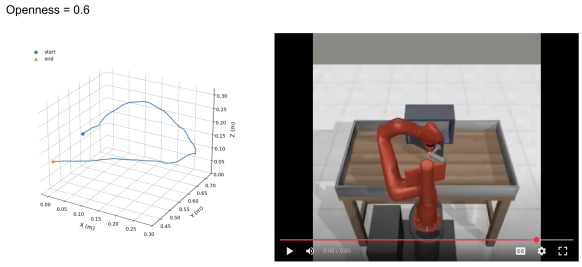

Figure 10: Controlled generation at $z_0 = 0.6$. **Left:** Generated trajectory with larger spatial displacement. **Right:** Simulation playback showing a wider door-opening motion, confirming smooth scaling with z_0.

Not all generated trajectories successfully open the door; some fail to reach or rotate the handle even with low reconstruction error. These cases are under investigation to identify systematic failure patterns.

Summary and Next Steps

Overall, these preliminary results validate the feasibility of the β-VAE pipeline for capturing the latent structure of door-opening trajectories. While ranking alignment remains underdeveloped, the current model achieves stable reconstruction and interpretable latent transitions. Future work will focus on:

- Tuning the ranking objective (λ_{rank}, α, min_gap) to improve monotonic separation;
- Exploring alternative generative models, such as diffusion-based architectures, if β-VAE limitations persist.
- Integrating LLM-based pairwise comparison modules for scale recognition;