One-shot Humanoid Whole-body Motion Learning

Anonymous CVPR submission

Paper ID *****

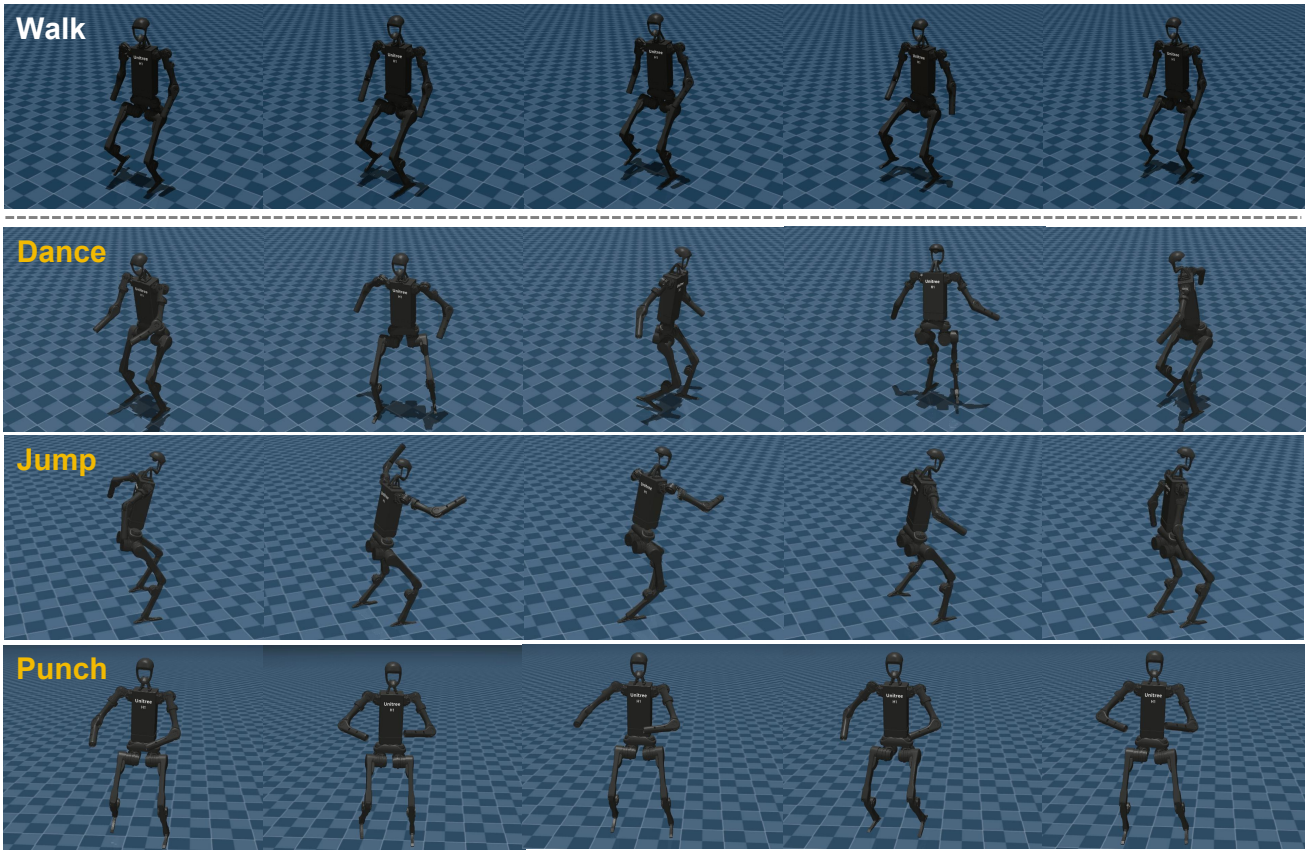

Figure 1. Sampled frames from motion sequences of a humanoid (Unitree H1) performing four distinct actions in sim-to-sim transfer settings, *i.e.*, trained in Isaac Gym and transferred to Mujoco. Walk is performed by the Base Model, while Dance, Jump, and Punch are performed by the model (Geo. FT Base Model) trained with our proposed method.

Abstract

Whole-body humanoid motion represents a cornerstone challenge in robotics, integrating balance, coordination, and adaptability to enable human-like behaviors. However, existing methods typically require multiple training samples per motion category, rendering the collection of high-quality human motion datasets both labor-intensive and costly. To address this, we propose a novel approach that trains effective humanoid motion policies using only a single non-walking target motion sample alongside readily available walking motions. The core idea lies in leveraging order-preserving optimal transport to compute distances between walking and non-walking sequences, followed by interpolation along geodesics to generate new intermediate pose skeletons, which are then optimized for collision-free configurations and retargeted to the humanoid before integration into a simulated environment for policy training via reinforcement learning. Experimental evaluations on the CMU MoCap dataset demonstrate that our method

consistently outperforms baselines, achieving superior performance across metrics.

1. Introduction

Humanoid whole-body motion is a fundamental challenge in robotics, encompassing the integration of balance, coordination, and adaptability to enable robots to act in diverse environments in a manner akin to humans. Advancements in this field have evolved from model-based control strategies, such as trajectory optimization and feedback controllers [8, 44], to data-driven approaches that leverage sensory feedback for real-time adjustments [13, 25, 32]. Recently, reinforcement learning (RL) pipelines for humanoid motion training typically involve the aggregation of massive datasets that incorporate a variety of different human motions, such as walking, running, and jumping, with each motion category represented by multiple training samples, *i.e.*, motion clips, to ensure generalization and robustness [2, 11, 15, 22]. These approaches retarget the collected motion data to humanoids' joints and train policy networks with historical proprioceptive observations, enabling the performance of agile behaviors [30].

Despite these advancements, a significant limitation in current training paradigms lies in the difficulty of collecting diverse human motion samples for a specific motion category with complex motion patterns, as motion capture processes are labor-intensive, require specialized equipment, and often suffer from inconsistencies due to environmental variations or human performer differences [47]. For instance, the CMU MoCap dataset[1] used in [2, 18] was generated by recording human subjects performing various motions in a lab equipped with 12 Vicon infrared MX-40 cameras, where subjects wore black jumpsuits with 41 reflective markers to enable infrared detection and triangulation of 3D marker positions. The raw data is then processed using ViconIQ software for skeleton template creation, marker labeling, and kinematic fitting, resulting in a collection of over 2600 motion trials. This challenge is exacerbated when scaling datasets for rare or complex motions, which demands extensive teleoperation or simulation, leading to inefficiencies and biases in policy learning [10]. Consequently, the question is whether a humanoid whole-body motion policy can be effectively trained with only *one-shot* human motion data, minimizing the need for exhaustive datasets while still achieving adaptable motion behaviors. Although one-shot learning paradigms have been extensively investigated in the domain of robotic arm manipulation, enabling manipulators to complete complex tasks from a single human demonstration [12, 45, 53], their application to humanoid motion—where dynamic balance and multi-joint coordination pose unique challenges—remains largely underexplored.

To address this question, we propose a method for learning a whole-body humanoid motion that utilizes only one target motion sample. A native idea is to train a motion policy directly using this single sample. However, we empirically notice a failure of such a scheme for several motion categories. Instead, we begin by training a Base Model on multiple walking motion data (around 130 walking clips), which is relatively straightforward to collect from video data on the Internet.

Subsequently, given a single non-walking target motion data, we generate multiple synthetic training samples to bridge the gap between these walking motions and the target motion on a pose skeleton manifold, ensuring smooth interpolation across pose spaces and gradual adaptation of model parameters trained on the walking data to the target motion data. More specifically, order-preserving optimal transport (OPOT) [35, 36] is used to compute the Wasserstein distance between walking motions and the target motion, which maintains temporal coherence and minimizes distortion in the generated skeletons. To ensure feasibility—such as preventing collisions among body parts—we further optimize the generated motions on the pose skeleton manifold using manifold optimization. Unlike other motion generation approaches [46, 49], we do not need to train any neural network, which makes our approach lightweight. Then, we finetune the Base Model on the generated motion data using the same strategy as training the Base Model. Experimental evaluations demonstrate that our method consistently outperforms baseline approaches, achieving superior performance across different metrics. Furthermore, sim-to-sim transfer underscores the robustness of the model trained with our approach.

2. Related Work

Humanoid whole-body motion and control. Recent advances have significantly pushed the frontier of humanoid motion with deep RL. Transformer-based policies have proved effective: [31] frames control as autoregressive token prediction, enabling robust outdoor walking, while [26] builds a low-dimensional latent action space that speeds hierarchical RL and covers 99% of AMASS motions [27]. World-model techniques further improve stability and perception: HuWo [51] couples a Transformer-XL [5] estimator with policy learning to handle complex contact dynamics, and World Model Reconstruction [37] reconstructs implicit terrain state for blind locomotion across rough ground. Curriculum and hierarchical strategies extend task scope—Lin *et al*. [24] integrate temporal vision transformers inside hierarchical RL to master long-horizon trail hiking, whereas Cui *et al*. [4] blend privileged pre-training with curriculum refinement to mitigate joint oscillations. Skill-

[1] https://mocap.cs.cmu.edu/

diversity and exploration are also actively studied: Wan *et al*. [42] augment inverse RL with quality-diversity objectives to learn gait repertoires, and Chiappa *et al*. [3] show lattice-structured exploration yields 18% higher rewards on the MuJoCo Humanoid benchmark. Complementary work focuses on data efficiency and imitation learning: CROSS-LOCO [23] drives bipedal agents with cross-modal motion priors, while Dugar *et al*. [7] track arbitrary pose subsets for robust whole-body imitation. Further progress addresses expressive whole-body control: ExBody [2] and its successor Exbody2 [18] utilize CMU MoCap human motion repertoire for motion mimicking; HumanPlus [11] studies real-time mirroring of human subjects; ASAP [15] improves hardware transfer, and H2O [14] enables low-latency kinesthetic control. Humanoid-X [28] converts 20 million human-video pose–text pairs into a large humanoid model, enabling scalable text-driven whole-body control. In contrast to these recent studies, we propose a method for learning humanoid motion using only one-shot human motion data, alleviating the burden of data collection.

Human motion generation. Recent efforts on human motion generation have embraced VAE [21], diffusion [16], and hybrid models built atop large motion corpora. Tevet *et al*. [38] pioneer text- and action-conditioned denoising diffusion for long-range sequences, and Motion Latent Diffusion [1] accelerates this paradigm by operating in a learned VAE latent space. Physics awareness is introduced by PhysDiff [48], which projects each denoised step through a simulator to suppress foot-sliding and penetration, and by BioMoDiffuse [19] that embeds Euler–Lagrange dynamics to ensure biomechanical plausibility. Controllability advances include Guided Motion Diffusion [20], which optimizes motions toward goal functions during sampling, and SMooDi [52], enabling arbitrary motion-style transfer via multi-condition diffusion. Goal-directed generation is addressed by WANDR [6], employing a feedback conditional VAE to reach 3D targets without RL, whereas CrossDiff^3 [33] leverages cross-domain diffusion to couple 2D and 3D constraints. Broader coverage of environments is pursued by TRAM [43] for long-range path-plus-pose synthesis and by PC-MRL [29], which homogenizes diverse point-cloud motions for cross-dataset interpolation. Collectively, these works establish a rich foundation of data-driven priors, physics-based constraints, and controllable diffusion processes. Zhao *et al*. [50] propose a conditional diffusion model that leverages bidirectional Markov chains to generate realistic, diverse, and variable-length 3D skeleton motions for specified action classes. SALAD [17] introduces a model that explicitly captures inter-relationships between skeletal joints, temporal frames, and text descriptors. Unlike prior studies, our method generates novel human motions without requiring the training of any neural networks, rendering it lightweight and well-suited for whole-body robot motion policy learning.

3. Methods

A prevalent framework adopted by recent works [2, 14, 18] for training a humanoid to perform diverse and expressive whole-body motions starts by curating human motion clips from sources like the CMU MoCap or AMASS [27] datasets with video-to-pose estimation models or visual tracking systems. These motion clips are then retargeted to the robot's kinematic structure (*e.g*., Unitree H1 with 19 DoFs) by mapping 3D human character's joint rotations to the humanoid's joint angles. From the retargeted clips, behavior goals (upper body joint positions and keypoints) and root movement goals (linear velocity, roll/pitch/yaw, and height) are extracted. Next, a goal-conditioned reinforcement learning policy is trained with PPO [34] in a simulated environment (*i.e*., Isaac Gym[2]) with randomized terrains and carefully designed rewards, where the upper body imitates reference motions via tracking rewards, while the legs focus on robustly following root commands to keep body balance without strict imitation. The overview of this framework is sketched in the right part of Figure 2. However, almost all existing methods require a bunch of human motion clips of the same motion category to learn such a behavior successfully. It is well-known that collecting a large amount of these motion clips containing different complex motions, either from videos or with teleoperations, is laborious and tedious. At the same time, we also note that walking videos, the most common motions in our daily life, are easy to access on the Internet, but the other types of diverse motion videos are not easy to acquire. Therefore, we seek to explore a solution to train an expressive whole-body humanoid motion model with only a single non-walking target video clip and multiple easy-to-acquire walking videos. The core innovation resides in generating multiple intermediate motions that bridge the gap between walking and non-walking target motions, thereby facilitating seamless learning transfer during the policy learning process.

3.1. Geodesic Distance for Pose Skeletons

First, we define a mathematical representation for human pose skeletons extracted from human motion clips. Formally, a *pose skeleton* at each timestep is represented as $(x, \{q_1, \ldots, q_J\})$ [39, 40], where:

- $x \in \mathbb{R}^3$ is a translation vector, specifying the position of the skeleton's root (pelvis) in 3D Euclidean space.
- $\{q_1, \ldots, q_J\}$ is a list of J unit quaternions, where each $q_j \in \mathbb{H}_1 \subset \mathbb{R}^4$ represents the rotation of the j-th joint. Here, $\mathbb{H}_1 = \{q \in \mathbb{R}^4 : \|q\|_2 = 1\}$ denotes the unit quaternion group, which is a double cover of SO(3).

One benefit of this representation is that it is invariant w.r.t.

[2] https://developer.nvidia.com/isaac-gym

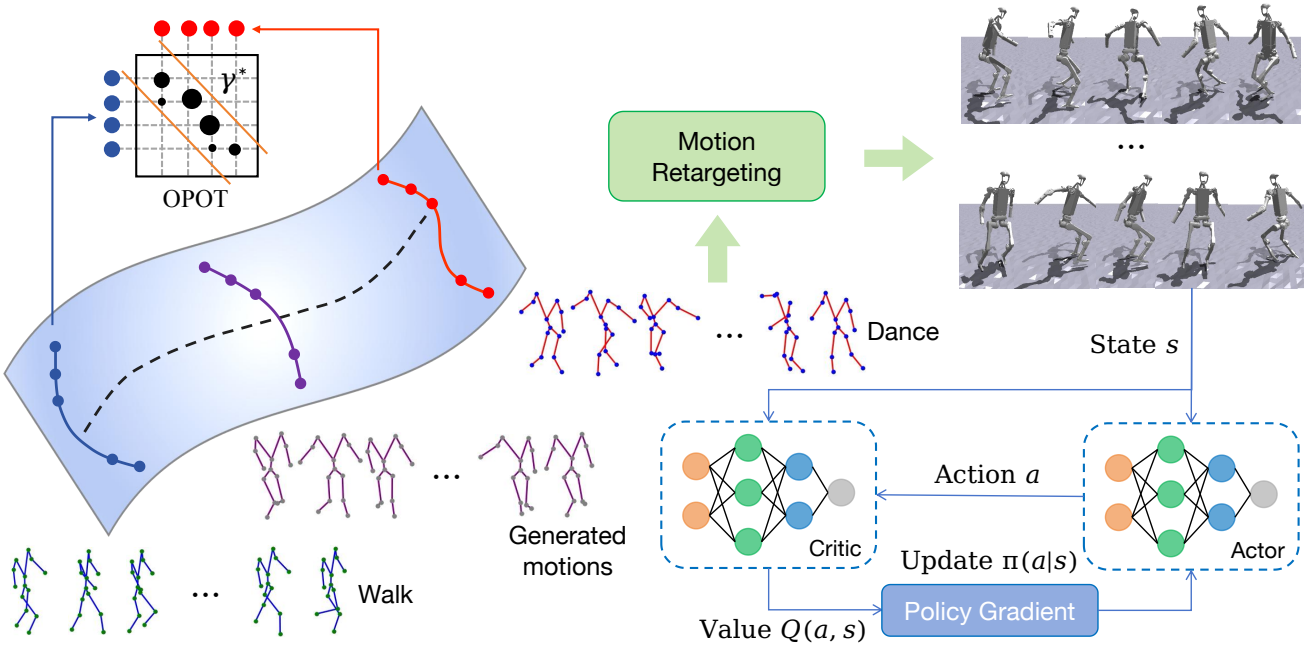

Figure 2. Given a sequence of walking motion pose skeletons and a target sequence comprising non-walking motions, we employ order-preserving optimal transport (OPOT) to compute the distance between these two sequences. Subsequently, we interpolate and sample novel pose skeletons along the geodesics connecting the walking and non-walking sequences. These sampled skeletons, together with the target sequence, are then retargeted to a humanoid robot and integrated into a simulated environment for training a whole-body motion policy via the Proximal Policy Optimization (PPO) algorithm.

skeleton scale. The *geodesic distance* between two poses, $(x_1, \{q_{1,1}, \dots, q_{1,J}\})$ and $(x_2, \{q_{2,1}, \dots, q_{2,J}\})$, is a metric that combines:

- The Euclidean distance between root translations in $\mathbb{R}^3$:

$$d_{\mathrm{t}}(x_1, x_2) = \|x_1 - x_2\|_2 \ . \tag{1}$$

- The sum of geodesic distances on $\mathrm{SO}(3)$ for all the corresponding joint rotations, each of which is computed via:

$$d_{\mathrm{r}}(q_{1,j}, q_{2,j}) = 2 \arccos(|\langle q_{1,j}, q_{2,j} \rangle|) \ , \tag{2}$$

  where $\langle q_{1,j}, q_{2,j} \rangle = q_{1,j} \cdot q_{2,j}$ is the Euclidean dot product in $\mathbb{R}^4$, and the absolute value ensures the shortest arc (since q and $-q$ represent the same rotation in $\mathrm{SO}(3)$).

Since the pose representation space is the direct product of Lie groups $\mathbb{R}^3 \times \mathrm{SO}(3)^J$, allowing the metric to be defined as the weighted sum of the independent geodesic distances on the translational component in $\mathbb{R}^3$ and the rotational components in each $\mathrm{SO}(3)$. Thus, the total geodesic distance is a weighted combination:

$$d((x_1, \{q_{1,j}\}), (x_2, \{q_{2,j}\})) = d_{\mathrm{t}}(x_1, x_2) + w \sum_{j=1}^{N} d_{\mathrm{r}}(q_{1,j}, q_{2,j}) \ , \tag{3}$$

where $w \in \mathbb{R}_{\geq 0}$ is a weight parameter balancing the contribution of translational and rotational components. In our settings, we set $w = 1$ for all cases. With this definition, we can measure the distance between two pose skeletons at different time steps across different motion clips. Then, we can further compute the discrepancy between the walking and non-walking motion sequences, as detailed below.

3.2. Order-Preserving Optimal Transport for Skeleton Sequence Alignment

Given a skeleton sequence from a walking clip, denoted as $S = \{s_1, \dots, s_N\}$ where $s_n = (x_n, \{q_{n,1}, \dots q_{n,J}\})$ is the pose skeleton at the nth timestep as defined above. Similarly, a skeleton sequence from a non-walking target motion clip is denoted as $T = \{t_1, \dots, t_M\}$ where $t_m = (x_m, \{q_{m,1}, \dots q_{m,J}\})$. We define a *distance* to measure the discrepancy between these two skeleton sequences. Optimal transport with Wasserstein distance [41] is a natural choice. However, classical optimal transport treats pose skeletons at each timestep as unordered sets, allowing non-monotonic mappings that violate the inherent chronological structure of motion sequences.

Order-Preserving Wasserstein (OPW) distance [35, 36] between two sequences handling local temporal distortions and periodic sequences, extends the classical Wasserstein distance by incorporating temporal regularizations to preserve order in sequential data. Specifically, the OPW distance views these sequences as empirical probability mea-

sures with uniform masses: $\alpha = \mathbf{1}_N/N \in \mathbb{R}^N$ and $\beta = \mathbf{1}_M/M \in \mathbb{R}^M$, where $\mathbf{1}_k$ denotes the all-ones vector of length k. The cost matrix $D \in \mathbb{R}^{N\times M}$ is computed for each pair of pose skeletons as defined in Eq. 3, *i.e.*, $D(n,m) = d((x_n, \{q_{n,j}\}), (x_m, \{q_{m,j}\}))$. The OPW additionally incorporates two regularizations:

- Inverse Difference Moment Regularization: Encourages local homogeneous transport by favoring couplings near the "diagonal" in normalized indices. Defined via the matrix $H \in \mathbb{R}^{N\times M}$ where:

$$H_{nm} = \frac{\lambda_1}{(n/N - m/M)^2 + 1} \quad , \tag{4}$$

  with $\lambda_1 > 0$ as the weight parameter.
- KL-Divergence Regularization with Gaussian Prior: Aligns the transport plan with a prior distribution $P \in \mathbb{R}^{N\times M}$ modeling expected couplings based on positional similarity. The prior is a scaled Gaussian:

$$P_{nm} = \frac{1}{\delta\sqrt{2\pi}} \exp\left(-\frac{d_{nm}^2}{2\delta^2}\right) \quad , \tag{5}$$

  where $d_{nm} = \left|\frac{n}{N} - \frac{m}{M}\right| / \sqrt{\frac{1}{N^2} + \frac{1}{M^2}}$, and $\delta > 0$ is the standard deviation parameter.

Then, the OPW distance is formulated as the minimum cost under these regularizations, approximated via entropic regularization with parameter $\lambda_2 > 0$:

$$d^{\lambda_1,\lambda_2}(S,T) = \min_{\Gamma\in U(\alpha,\beta)} \langle \Gamma, D\rangle - \lambda_1 \mathrm{I}(\Gamma) + \lambda_2 \mathrm{KL}(\Gamma \| P) \quad , \tag{6}$$

where $U(\alpha,\beta)$ is the transport polytope for alignment, $\langle\cdot,\cdot\rangle$ is the Frobenius inner product, $I(\Gamma) = \sum_{nm} \Gamma_{nm} H_{nm}/\lambda_1$ (adjusted for scaling), and $\mathrm{KL}(\Gamma\|P)$ is the Kullback-Leibler divergence. Eq. 6 is solved using Sinkhorn iterations [36]. In our settings, we set $\lambda_1 = 50$, $\lambda_2 = 0.1$, $\delta = 1$, and the maximum number of iterations to 20 for all cases. After solving Eq. 6, we can get 1. the optimal per-frame *alignment* (a.k.a transport plan) between the walking and non-walking motion sequences, and 2. the shortest *distance* between these two sequences. This process is shown in the left part of Figure 2. Then, we select $Q = 10$ walking motions with the smallest shortest distances w.r.t. the target non-walking motions to generate new intermediate motions.

3.3. Soft Transport Plans to Hard Assignments

The solution to Eq. 6 is a soft transport plan, *i.e.*, the pose skeleton at each timestep in the non-walking target sequence can be possibly aligned to more than one pose skeleton across multiple (usually two to three) adjacent timesteps in the walking sequence. However, our proposed motion generation approach necessitates a hard assignment of each target pose skeleton at every timestep to exactly one walking pose skeleton, permitting reuse of the latter. Thus, we propose a method for projecting a soft transport plan onto a hard assignment matrix, ensuring each target is matched to exactly one source, with sources potentially reused. The method leverages the linear sum assignment problem, solved using the Hungarian algorithm, to maximize the total assignment score derived from the input soft transport plan.

Let $\Gamma^* \in \mathbb{R}^{n\times m}_{\geq 0}$ be the optimal soft transport plan derived from Eq. 6. The matrix Γ^* represents a probabilistic coupling, where $\Gamma_{nm} \geq 0$ indicates the strength of association between source s_n and target t_m. The goal is to produce a hard assignment matrix $\gamma \in \{0,1\}^{n\times m}$, where $\gamma_{nm} = 1$ if source s_n is assigned to target t_m, and $\gamma_{nm} = 0$ otherwise, subject to the constraint that each target skeleton is assigned exactly one walking skeleton:

$$\sum_{n=1}^{N} \gamma_{nm} = 1, \quad \forall m \in \{1,\ldots,M\}, \tag{7}$$

while allowing sources to be reused (*i.e.*, $\sum_{m=1}^{M} \gamma_{nm} \geq 0$). The problem is formulated as maximizing the total assignment score:

$$\max_{\gamma\in\mathcal{T}} \sum_{n=1}^{N}\sum_{m=1}^{M} \gamma_{nm}\Gamma^*_{nm} \quad , \tag{8}$$

where $\mathcal{T} = \{\gamma \in \{0,1\}^{N\times M} \mid \sum_{n=1}^{N} \gamma_{nm} = 1, \forall m\}$ is the set of valid assignment matrices. To use standard optimization algorithms, we take $C = -\Gamma^*$, transforming the maximization into a minimization:

$$\min_{\gamma\in\mathcal{T}} \sum_{n=1}^{N}\sum_{m=1}^{M} C_{nm}\gamma_{nm} \quad . \tag{9}$$

Note that we need to take two cases into consideration based on the relative sizes of N and M:

Case 1: sufficient walking skeletons ($N \geq M$): When the walking clip is longer than or equal to the target motion clip, each target pose skeleton can be assigned a unique walking one. The cost matrix $C = -\Gamma^* \in \mathbb{R}^{N\times M}$ is transposed to shape (M,N) to align with the convention of assigning targets to sources. The Hungarian algorithm solves:

$$\begin{aligned} &\min \sum_{m=1}^{M}\sum_{n=1}^{N} C_{nm}\tilde{\gamma}_{mn}, \\ &\text{s.t.} \sum_{n=1}^{N} \tilde{\gamma}_{mn} = 1, \forall m, \quad \sum_{m=1}^{M} \tilde{\gamma}_{mn} \leq 1, \forall n, \end{aligned} \tag{10}$$

where $\tilde{\gamma} \in \{0,1\}^{m\times n}$ assigns each target skeleton to a walking skeleton. The algorithm returns indices (m,n) such that $\tilde{\gamma}_{mn} = 1$. These are mapped to $\gamma \in \{0,1\}^{n\times m}$ by setting $\gamma_{nm} = 1$ if $\tilde{\gamma}_{mn} = 1$, ensuring each target skeleton is assigned exactly one walking skeleton.

Case 2: insufficient walking skeletons ($N < M$): When the walking clip is shorter than the target motion clip, walking skeletons must be reused. To achieve this, the cost matrix C is tiled to create virtual skeletons. The number of repetitions is computed as $r = \lceil \frac{M}{N} \rceil$, and the cost matrix is replicated r times along the row dimension, yielding a tiled matrix $C_{\text{tiled}} \in \mathbb{R}^{rN \times M}$. The first M rows are selected to form a square matrix of shape (M, M). The Hungarian algorithm is applied to the transposition of this matrix to solve a one-to-one assignment problem:

$$\begin{aligned} &\min \sum_{m=1}^{M} \sum_{n=1}^{M} C_{\text{tiled},nm} \tilde{\gamma}_{mn}, \\ &\text{s.t.} \sum_{n=1}^{M} \tilde{\gamma}_{mn} = 1, \forall m, \quad \sum_{m=1}^{M} \tilde{\gamma}_{mn} = 1, \forall n \ . \end{aligned} \tag{11}$$

The resulting assignments map each target skeleton t_m to a virtual walking skeleton $s_{n_{\text{virtual}}}$. The real walking skeleton index is computed as $n_{\text{real}} = (n_{\text{virtual}} \mod N)$, and the assignment is recorded in γ by setting $\gamma_{n_{\text{real}},m} = 1$.

The above process returns $\gamma \in \{0,1\}^{N \times M}$, where each column sums to 1, ensuring every target non-walking skeleton is assigned exactly one walking skeleton. Then, we can generate new pose skeletons along the geodesics connecting each paired walking and non-walking target motions at every timestep.

3.4. Pose Skeletons Sampling along Geodesics

When each non-walking target pose skeleton is matched to a walking skeleton, we can interpolate and sample poses along the geodesic path in the product manifold $\mathbb{R}^3 \times \text{SO}(3)^J$ to generate new pose skeletons. On this manifold, $\mathbb{R}^3$ is equipped with the Euclidean metric, and each SO(3) component uses the geodesic metric induced by the unit quaternion representation. The *geodesic path* between two paired poses $s_n = (x_n, \{q_{n,j}\}_{j=1}^J)$ and $t_m = (x_m, \{q_{m,j}\}_{j=1}^J)$ is the shortest path in this manifold, parameterized by $\tau \in [0,1]$. The interpolated pose at τ is:

$$p(\tau) = (x(\tau), \{q_j(\tau)\}_{j=1}^J) \ , \tag{12}$$

where:

- $x(\tau) = (1-\tau)x_n + \tau x_m$ is the linear interpolation in $\mathbb{R}^3$ for the root translation.
- $q_j(\tau) = \text{SLERP}(q_{n,j}, q_{m,j}, \tau)$ is the spherical linear interpolation on $\mathbb{H}_1$ for the jth joints' rotation, defined for unit quaternions $q_{n,j}, q_{m,j}$ as:

$$q_j(\tau) = \frac{\sin((1-\tau)\theta_j)}{\sin(\theta_j)} q_{n,j} + \frac{\sin(\tau\theta_j)}{\sin(\theta_j)} q_{m,j} \ , \tag{13}$$

  where $\theta_j = \arccos(\langle q_{n,j}, q_{m,j} \rangle)$ is the angle between quaternions, and $\langle \cdot, \cdot \rangle$ is the Euclidean dot product in $\mathbb{R}^4$. For antipodal quaternions ($q_{m,j} \approx -q_{n,j}$), a shortest-path adjustment is typically applied.

Through interpolation and sampling, we generated new pose skeletons that exhibit motion intermediate between walking and the target motions. For each of the selected $Q = 10$ walking motions, we generate six new pose skeletons along geodesics.

3.5. Optimization for Collision-Free Skeletal Poses

Note that the generated new poses above may lead to bone collisions, and thus, we introduce an optimization routine to make the generated skeletal poses collision-free for robot retargeting. Given an interpolated pose $p(\tau)$, we first apply forward kinematics to compute world-space transforms and positions for each joint by iteratively composing local transformations along the kinematic tree, resulting in a $J \times 3$ matrix of joint positions in the world coordinate frame.

For collision checking, each joint is modeled as a sphere, and each bone as a capsule. A fast batched segment–segment distance kernel implements a Collision Detection Algorithm proposed in [9]. Using those distances, we build two differentiable energies: a sphere energy that penalizes interpenetration of non-adjacent joint spheres by masking out parent/child and grand-parent relations, and a capsule energy penalizes overlaps between non-adjacent bones by modeling each bone as a capsule (with radius inherited from its parent joint) and imposing a hinge-style penalty when two such capsules intrude into each other's space. The sphere energy function penalizes overlaps between joint-approximating spheres by summing squared penetration depths for non-adjacent joint pairs, excluding parents, children, and grandparents to focus on true self-collisions. The capsule energy function extends this to bone approximations, gathering bone segments, computing all-pairs segment distances, and summing squared penetrations for non-adjacent bone pairs after masking those sharing joints. Since forward kinematics, sphere and capsule energies are differentiable w.r.t. to $p(\tau)$, we apply a pose optimization process, as sketched in Algorithm 1, to minimize the combined sphere and capsule energies using Riemannian gradient descent on the product manifold $\text{SO}(3)^J$ for joint rotations. This process ensures descent along geodesic directions in the rotation manifold, with a learning rate lr controlling the optimization step size.

4. Experiments

Human motion data curation. The CMU MoCap dataset contains 780 human motion clips, containing not only walking, but also other expressive motion behaviors. Unlike previous works [2, 18] which train a model with a large portion of MoCap data, we formulate our training scheme in two phases. First, we train a Base Model on multiple walking motions, and then fine-tune the Base Model with our generated motions along the geodesics connecting the walking motions and the single target motion. Note that the motion

Algorithm 1 Pose Collision-Free Optimization Algorithm

$\mathbf{q} \leftarrow \text{clone}(p(\tau))$ with gradients enabled
for each step up to maximum number **do**
  $\mathbf{x} \leftarrow \text{FK}(\mathbf{q})$ {forward kinematics}
  $E \leftarrow \text{sphere_energy}(\mathbf{x}) + \cdot\text{capsule_energy}(\mathbf{x})$
  if $E < 10^{-6}$ **then**
    Break {terminate early if energy is small}
  end if
  $\nabla_{\mathbf{q}} E \leftarrow \text{backward}(E)$ {back-propagate}
  $\mathbf{g} \leftarrow \nabla_{\mathbf{q}} E$
  $\mathbf{g}_{\text{proj}} = \mathbf{g} - (\mathbf{q} \cdot \mathbf{g})\mathbf{q}$ {project gradient; $\cdot$ is dot product}
  $\mathbf{q} \leftarrow \mathbf{q} - lr \cdot \mathbf{g}_{\text{proj}}$ {lr is learning rate}
  $\mathbf{q} \leftarrow \text{quat_normalize}(\mathbf{q})$ {project back to unit sphere}
  Zero $\mathbf{q}$.grad
end for
$\mathbf{x}_{\text{final}} \leftarrow \text{FK}(\text{quat_normalize}(\mathbf{q}))$ {$\mathbf{x}_{\text{final}}$ is used for motion policy training}

clips for training and evaluation are performed by different actors to increase the challenges of motion variability. The select motions and the number of evaluation clips are listed in Table 1.

Baselines and evaluation metrics. Our baselines include:

- Base Model: We train this model using the 135 Walk clips from scratch.
- Single Motion Motion: For each category, we train a model using only one training clip from scratch.
- Single Finetune (FT) Base Model: For each category, we finetune the Base Model using only one training clip.
- Geodesic (Geo.) Motion Base Model: For each category, we train a model using the motions sampled along the geodesics from scratch.
- Geodesic (Geo.) Finetune (FT) Base Model: For each category, we finetune the Base Model using the motions sampled along the geodesics.

Following [2], we use the metrics below:

- Mean Episode Linear Velocity Tracking Reward (MELV).
- Mean episode roll pitch tracking reward (MERP).
- Mean episode key body tracking reward (MEK).
- Mean episode lengths (MEL).

4.1. Results and Analysis

The results of five target non-walking motions are listed in Table 2 to Table 4. Across all five target motions, two clear trends emerge. First, relying on a single non-walking clip without any prior knowledge (Single Motion Model) is risky: the robot falls in three of the five tasks (marks), and even when it achieves higher scores in some cases. Starting from the Base Model and simply fine-tuning on that same single clip (Single FT Base Model) already remedies stability and often produces competitive second-tier results. However, the most consistent gains come from training with motions generated along geodesics. Training only on those synthetic sequences (Geo. Motion Model) gives the best or second-best numbers on every metric for Basketball and yields solid improvements for Jump, Punch, and Wash. When geodesic data are used for fine-tuning the Base Model (Geo. FT Base Model) the method dominates: it attains the top or the second top scores in 17 of the 20 metrics. Second, the pattern is relatively stable across metrics. Geo. FT Base Model drives the highest MELV and MEKB values for four motions, and records the largest MERP for Dance, Punch, and Wash, while achieving competitive results with the best performance for Basketball and Jump. In short, training from scratch or fine-tuning Base Model with our interpolated motions provides the most reliable route to high quantitative performance and robust execution, whereas naively training from scratch with a single motion clip is prone to failure.

Categories	Training		Evaluation	
	Clips	Duration (s)	Clips	Duration (s)
Walk	135	1752.5	-	-
Basketball	1	50.5	13	135.1
Dance	1	7.7	74	1455.9
Jump	1	3.5	23	84.9
Punch	1	15.4	12	316.6
Wash	1	14.2	8	250.1

Table 1. Statistics for training and evaluation motion clips.

5. Ablation Study

Effect of the transport formulation. We replace order-preserving optimal transport with classical, permutation-free optimal transport on the Jump motion, and the results are reported in Table 5. We observe a consistent performance degradation across all four metrics, indicating that the matching strategy plays a non-trivial role in downstream policy learning. In particular, because the order-preserving formulation explicitly respects the temporal structure of the walking sequence, it yields correspondences that remain coherent over time and avoid unrealistic frame-to-frame reassignments. As a result, it produces geodesic trajectories whose kinematic statistics (e.g., phase progression and velocity profiles) align more closely with the target jump sequence, leading to smoother intermediate motions.

Effect of the number of geodesic samples. We hold the order-preserving optimal transport fixed, but vary the number of interpolated poses used for training. Performance

Methods	Basketball				Dance			
	MELV	MERP	MEKB	MLE	MELV	MERP	MEKB	MLE
Base Model	17.525	43.452	37.542	**1501.990**	4.286	19.956	15.146	789.447
Single Motion Model	17.657	41.408	40.220	**1501.990**	3.803	29.389	22.132	1220.065
Single FT Base Model	**20.143**	43.247	41.347	**1501.990**	6.323	25.625	19.792	960.378
Geo. Motion Model	14.206	**45.458**	**41.618**	**1501.990**	7.209	22.539	16.318	862.318
Geo. FT Base Model	15.670	45.449	39.218	**1501.990**	**14.133**	**36.744**	**26.563**	**1421.610**

Table 2. Quantitative evaluation of motion models on Basketball and Dance using MELV, MERP, MEKB, and MLE metrics.

Methods	Jump				Punch			
	MELV	MERP	MEKB	MLE	MELV	MERP	MEKB	MLE
Base Model	8.464	**42.959**	**33.048**	**1487.360**	8.405	38.967	31.275	1256.202
Single Motion Model	$1.202_{(\times)}$	$4.657_{(\times)}$	$2.500_{(\times)}$	$163.030_{(\times)}$	$36.122_{(\times)}$	$36.752_{(\times)}$	$16.159_{(\times)}$	$1459.245_{(\times)}$
Single FT Base Model	6.686	37.217	27.179	1252.151	**22.559**	44.435	35.164	**1501.990**
Geo. Motion Model	10.869	41.927	30.062	1456.650	15.218	44.658	35.795	**1501.990**
Geo. FT Base Model	**14.294**	42.063	32.865	1458.304	17.904	**46.369**	**36.375**	**1501.990**

Table 3. Quantitative evaluation of motion models on Jump and Punch using MELV, MERP, MEKB, and MLE metrics.

Methods	Wash			
	MELV	MERP	MEKB	MLE
Base Model	5.864	44.144	29.704	**1501.990**
Single Motion Model	$16.862_{(\times)}$	$18.915_{(\times)}$	$9.684_{(\times)}$	$720.329_{(\times)}$
Single FT Base Model	7.735	44.651	**34.709**	**1501.990**
Geo. Motion Model	12.336	44.192	32.591	**1501.990**
Geo. FT Base Model	**13.187**	**44.902**	32.190	**1501.990**

Table 4. Quantitative evaluation of motion models on Wash using MELV, MERP, MEKB, and MLE metrics.

Methods	Jump			
	MELV	MERP	MEKB	MLE
Classicial OT	12.577	41.678	31.922	1458.255
Order-preserving OT	**14.294**	**42.063**	**32.865**	**1458.304**
1 generated samples	12.960	41.716	31.897	1458.275
3 generated samples	13.376	41.703	31.700	1458.245
6 generated samples	**14.294**	**42.063**	**32.865**	**1458.304**

Table 5. Ablation study results on Jump, evaluating with MELV, MERP, MEKB, and MLE metrics.

improves monotonically when increasing the samples from one to three to six, with six matching the best overall scores. This trend indicates that the quantity of synthesized supervision is an important factor under the one-shot setting, and that too sparse interpolation can leave sizable gaps between the source and target motion distributions. Denser sampling along the geodesic supplies the policy with richer coverage of intermediate poses and transition dynamics. Consequently, the learned behaviors exhibit smoother trajectories and more accurate end-point kinematics.

6. Conclusion

In this paper, we propose a solution to address the challenge of training robust humanoid motion policies by leveraging readily available walking motions alongside a single non-walking target sample. Our approach uses order-preserving optimal transport to synthesize meaningful intermediate poses along geodesic paths that smoothly bridge the source and target motions, and then refines these poses through collision-free optimization and retargeting to ensure physical plausibility and compatibility with the humanoid model. The resulting motion set serves as an effective curriculum for reinforcement learning, enabling stable training despite the extreme data scarcity. Experiments show that our method consistently surpasses strong baselines across tasks and evaluation settings, validating its ability to improve both generalization and final performance while keeping data requirements minimal.

References

[1] Xin Chen, Biao Jiang, Wen Liu, Zilong Huang, Bin Fu, Tao Chen, and Gang Yu. Executing your commands via motion diffusion in latent space. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18000–18010, 2023. 3

[2] Xuxin Cheng, Yandong Ji, Junming Chen, Ruihan Yang, Ge Yang, and Xiaolong Wang. Expressive whole-body control for humanoid robots. In *Robotics: Science and Systems*, 2024. 2, 3, 6, 7

[3] Alberto Silvio Chiappa, Alessandro Marin Vargas, Ann Huang, and Alexander Mathis. Latent exploration for reinforcement learning. *Advances in Neural Information Processing Systems*, 36:56508–56530, 2023. 3

[4] Wenhao Cui, Shengtao Li, Huaxing Huang, Bangyu Qin, Tianchu Zhang, Liang Zheng, Ziyang Tang, Chenxu Hu, NING Yan, Jiahao Chen, et al. Adapting humanoid locomotion over challenging terrain via two-phase training. In *Annual Conference on Robot Learning*, 2024. 2

[5] Zihang Dai, Zhilin Yang, Yiming Yang, Jaime G Carbonell, Quoc Le, and Ruslan Salakhutdinov. Transformer-xl: Attentive language models beyond a fixed-length context. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pages 2978–2988, 2019. 2

[6] Markos Diomataris, Nikos Athanasiou, Omid Taheri, Xi Wang, Otmar Hilliges, and Michael J Black. Wandr: Intention-guided human motion generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 927–936, 2024. 3

[7] Pranay Dugar, Aayam Shrestha, Fangzhou Yu, Bart van Marum, and Alan Fern. Learning multi-modal whole-body control for real-world humanoid robots. *arXiv preprint arXiv:2408.07295*, 2024. 3

[8] Mohamed Elobaid, Giulio Romualdi, Gabriele Nava, Lorenzo Rapetti, Hosameldin Awadalla Omer Mohamed, and Daniele Pucci. Online non-linear centroidal mpc for humanoid robots payload carrying with contact-stable force parametrization. In *IEEE International Conference on Robotics and Automation*, pages 12233–12239. IEEE, 2023. 2

[9] Christer Ericson. *Real-time collision detection*. CRC Press, 2004. 6

[10] Hao-Shu Fang, Hongjie Fang, Zhenyu Tang, Jirong Liu, Chenxi Wang, Junbo Wang, Haoyi Zhu, and Cewu Lu. Rh20t: A comprehensive robotic dataset for learning diverse skills in one-shot. In *IEEE International Conference on Robotics and Automation*, pages 653–660. IEEE, 2024. 2

[11] Zipeng Fu, Qingqing Zhao, Qi Wu, Gordon Wetzstein, and Chelsea Finn. Humanplus: Humanoid shadowing and imitation from humans. In *Annual Conference on Robot Learning*, 2024. 2, 3

[12] Abraham George and Amir Barati Farimani. One act play: Single demonstration behavior cloning with action chunking transformers. *arXiv preprint arXiv:2309.10175*, 2023. 2

[13] Zhaoyuan Gu, Junheng Li, Wenlan Shen, Wenhao Yu, Zhaoming Xie, Stephen McCrory, Xianyi Cheng, Abdulaziz Shamsah, Robert Griffin, C Karen Liu, et al. Humanoid locomotion and manipulation: Current progress and challenges in control, planning, and learning. *IEEE/ASME Transactions on Mechatronics*, 2025. 2

[14] Tairan He, Zhengyi Luo, Wenli Xiao, Chong Zhang, Kris Kitani, Changliu Liu, and Guanya Shi. Learning human-to-humanoid real-time whole-body teleoperation. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 8944–8951. IEEE, 2024. 3

[15] Tairan He, Jiawei Gao, Wenli Xiao, Yuanhang Zhang, Zi Wang, Jiashun Wang, Zhengyi Luo, Guanqi He, Nikhil Sobanbab, Chaoyi Pan, et al. Asap: Aligning simulation and real-world physics for learning agile humanoid whole-body skills. *arXiv preprint arXiv:2502.01143*, 2025. 2, 3

[16] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33:6840–6851, 2020. 3

[17] Seokhyeon Hong, Chaelin Kim, Serin Yoon, Junghyun Nam, Sihun Cha, and Junyong Noh. Salad: Skeleton-aware latent diffusion for text-driven motion generation and editing. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 7158–7168, 2025. 3

[18] Mazeyu Ji, Xuanbin Peng, Fangchen Liu, Jialong Li, Ge Yang, Xuxin Cheng, and Xiaolong Wang. Exbody2: Advanced expressive humanoid whole-body control. In *RSS 2025 Workshop on Whole-body Control and Bimanual Manipulation: Applications in Humanoids and Beyond*, 2025. 2, 3, 6

[19] Zixi Kang, Xinghan Wang, and Yadong Mu. Biomodiffuse: Physics-guided biomechanical diffusion for controllable and authentic human motion synthesis. *arXiv preprint arXiv:2503.06151*, 2025. 3

[20] Korrawe Karunratanakul, Konpat Preechakul, Supasorn Suwajanakorn, and Siyu Tang. Guided motion diffusion for controllable human motion synthesis. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2151–2162, 2023. 3

[21] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. In *International Conference on Learning Representations*, 2014. 3

[22] Jialong Li, Xuxin Cheng, Tianshu Huang, Shiqi Yang, Ri-Zhao Qiu, and Xiaolong Wang. Amo: Adaptive motion optimization for hyper-dexterous humanoid whole-body control. *arXiv preprint arXiv:2505.03738*, 2025. 2

[23] Tianyu Li, Hyunyoung Jung, Matthew Gombolay, Yong Kwon Cho, and Sehoon Ha. Crossloco: Human motion driven control of legged robots via guided unsupervised reinforcement learning. *arXiv preprint arXiv:2309.17046*, 2023. 3

[24] Kwan-Yee Lin and Stella X Yu. Let humanoids hike! integrative skill development on complex trails. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 22498–22507, 2025. 2

[25] Junfeng Long, Junli Ren, Moji Shi, Zirui Wang, Tao Huang, Ping Luo, and Jiangmiao Pang. Learning humanoid locomotion with perceptive internal model. *arXiv preprint arXiv:2411.14386*, 2024. 2

[26] Zhengyi Luo, Jinkun Cao, Josh Merel, Alexander Winkler, Jing Huang, Kris M Kitani, and Weipeng Xu. Universal humanoid motion representations for physics-based control. In *International Conference on Learning Representations*, 2024. 2

[27] Naureen Mahmood, Nima Ghorbani, Nikolaus F Troje, Gerard Pons-Moll, and Michael J Black. Amass: Archive of motion capture as surface shapes. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5442–5451, 2019. 2, 3

[28] Jiageng Mao, Siheng Zhao, Siqi Song, Tianheng Shi, Junjie Ye, Mingtong Zhang, Haoran Geng, Jitendra Malik, Vitor Campagnolo Guizilini, and Yue Wang. Universal humanoid robot pose learning from internet human videos. In *ICRA Workshop: Human-Centered Robot Learning in the Era of Big Data and Large Models*, 2025. 3

[29] Clinton Mo, Kun Hu, Chengjiang Long, Dong Yuan, and Zhiyong Wang. Motion keyframe interpolation for any human skeleton via temporally consistent point cloud sampling and reconstruction. In *European Conference on Computer Vision*, pages 159–175. Springer, 2024. 3

[30] Ilija Radosavovic, Tete Xiao, Bike Zhang, Trevor Darrell, Jitendra Malik, and Koushil Sreenath. Real-world humanoid locomotion with reinforcement learning. *Science Robotics*, 9(89):eadi9579, 2024. 2

[31] Ilija Radosavovic, Bike Zhang, Baifeng Shi, Jathushan Rajasegaran, Sarthak Kamat, Trevor Darrell, Koushil Sreenath, and Jitendra Malik. Humanoid locomotion as next token prediction. *Advances in Neural Information Processing Systems*, 37:79307–79324, 2024. 2

[32] Junli Ren, Tao Huang, Huayi Wang, Zirui Wang, Qingwei Ben, Junfeng Long, Yanchao Yang, Jiangmiao Pang, and Ping Luo. Vb-com: Learning vision-blind composite humanoid locomotion against deficient perception. *arXiv preprint arXiv:2502.14814*, 2025. 2

[33] Zeping Ren, Shaoli Huang, and Xiu Li. Realistic human motion generation with cross-diffusion models. In *European Conference on Computer Vision*, pages 345–362. Springer, 2024. 3

[34] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017. 3

[35] Bing Su and Gang Hua. Order-preserving wasserstein distance for sequence matching. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1049–1057, 2017. 2, 4

[36] Bing Su and Gang Hua. Order-preserving optimal transport for distances between sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(12):2961–2974, 2018. 2, 4, 5

[37] Wandong Sun, Long Chen, Yongbo Su, Baoshi Cao, Yang Liu, and Zongwu Xie. Learning humanoid locomotion with world model reconstruction. *arXiv preprint arXiv:2502.16230*, 2025. 2

[38] Guy Tevet, Sigal Raab, Brian Gordon, Yoni Shafir, Daniel Cohen-or, and Amit Haim Bermano. Mdm: Human motion diffusion model. In *International Conference on Learning Representations*, 2023. 3

[39] Raviteja Vemulapalli and Rama Chellapa. Rolling rotations for recognizing human actions from 3d skeletal data. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4471–4479, 2016. 3

[40] Raviteja Vemulapalli, Felipe Arrate, and Rama Chellappa. Human action recognition by representing 3d skeletons as points in a lie group. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 588–595, 2014. 3

[41] Cédric Villani et al. *Optimal transport: old and new*. Springer, 2008. 4

[42] Zhenglin Wan, Xingrui Yu, David Mark Bossens, Yueming Lyu, Qing Guo, Flint Xiaofeng Fan, Yew-Soon Ong, and Ivor Tsang. Diversifying robot locomotion behaviors with extrinsic behavioral curiosity. In *International Conference on Machine Learning*, 2025. 3

[43] Yufu Wang, Ziyun Wang, Lingjie Liu, and Kostas Daniilidis. Tram: Global trajectory and motion of 3d humans from in-the-wild videos. In *European Conference on Computer Vision*, pages 467–487. Springer, 2024. 3

[44] Patrick M Wensing, Michael Posa, Yue Hu, Adrien Escande, Nicolas Mansard, and Andrea Del Prete. Optimization-based control for dynamic legged robots. *IEEE Transactions on Robotics*, 40:43–63, 2023. 2

[45] Albert Wu, Ruocheng Wang, Sirui Chen, Clemens Eppner, and C Karen Liu. One-shot transfer of long-horizon extrinsic manipulation through contact retargeting. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 13891–13898. IEEE, 2024. 2

[46] Hua Yu, Weiming Liu, Jiapeng Bai, Xu Gui, Yaqing Hou, YewSoon Ong, and Qiang Zhang. Towards efficient and diverse generative model for unconditional human motion synthesis. In *Proceedings of the ACM International Conference on Multimedia*, pages 2535–2544, 2024. 2

[47] Mingqi Yuan, Tao Yu, Wenqi Ge, Xiuyong Yao, Dapeng Li, Huijiang Wang, Jiayu Chen, Xin Jin, Bo Li, Hua Chen, et al. Behavior foundation model: Towards next-generation whole-body control system of humanoid robots. *arXiv preprint arXiv:2506.20487*, 2025. 2

[48] Ye Yuan, Jiaming Song, Umar Iqbal, Arash Vahdat, and Jan Kautz. Physdiff: Physics-guided human motion diffusion model. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 16010–16021, 2023. 3

[49] Mingyuan Zhang, Zhongang Cai, Liang Pan, Fangzhou Hong, Xinying Guo, Lei Yang, and Ziwei Liu. Motiondiffuse: Text-driven human motion generation with diffusion model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(6):4115–4128, 2024. 2

[50] Mengyi Zhao, Mengyuan Liu, Bin Ren, Shuling Dai, and Nicu Sebe. Denoising diffusion probabilistic models for action-conditioned 3d motion generation. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 4225–4229. IEEE, 2024. 3

[51] Han Zheng, Yi Cheng, Hang Liu, Linqi Ye, and Houde Liu. Huwo: Building physical interaction world models for humanoid robot locomotion. In *ICLR Workshop on World Models: Understanding, Modelling and Scaling*, 2025. 2

[52] Lei Zhong, Yiming Xie, Varun Jampani, Deqing Sun, and Huaizu Jiang. Smoodi: Stylized motion diffusion model. In *European Conference on Computer Vision*, pages 405–421. Springer, 2024. 3

[53] Huayi Zhou, Ruixiang Wang, Yunxin Tai, Yueci Deng, Guiliang Liu, and Kui Jia. You only teach once: Learn one-shot bimanual robotic manipulation from video demonstrations. *arXiv preprint arXiv:2501.14208*, 2025. 2