Any2Any: Efficient Cross-Embodiment Transfer for Humanoid Whole-Body Tracking

Ming Yang, Tao Yu[†*], Feng Li, Hua Chen
LimX Dynamics
[†]Project Lead [*]Corresponding Author

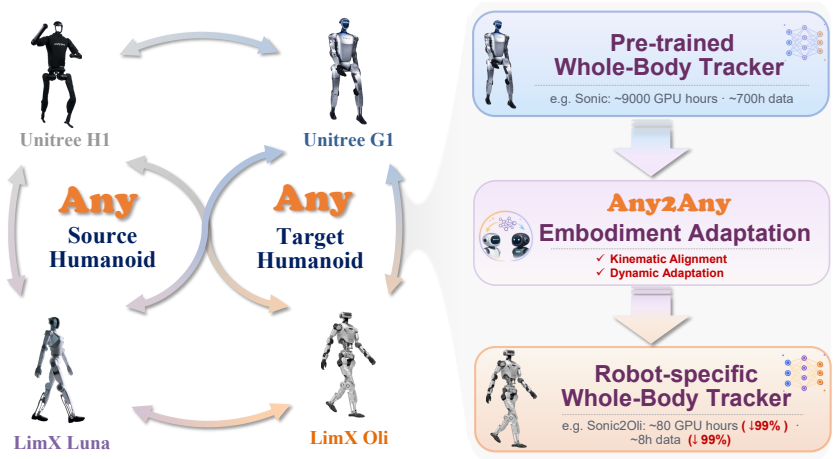


Figure 1: Illustration of ANY2ANY. A pretrained whole-body tracker (WBT) learned on specific humanoid can be efficiently transferred to another humanoid platform through the proposed ANY2ANY. For example, GEAR-SONIC [1], a large-scale pretrained WBT, can be adapted to a target robot LimX Oli using only a small fraction of the original training compute and data.

Abstract: Whole-body tracking (WBT) models have become a key foundation for humanoid robots, enabling them to imitate diverse motions with high fidelity. Training such models from scratch requires large-scale data and computation, making rapid deployment on new humanoid platforms costly. This raises a natural question: *Can pretrained WBT models transfer across embodiments with minimal adaptation?* To answer this question, we propose ANY2ANY, a paradigm that efficiently transfers an existing robot-specific WBT policy to a new humanoid embodiment with only a small amount of data and compute. Any2Any first performs *kinematic alignment* between source and target humanoids, aligning their input and output spaces so that the pretrained source policy can be meaningfully reused on the target embodiment. Any2Any then performs *dynamics adaptation*, inserting lightweight parameter-efficient fine-tuning (PEFT) components into only the most dynamics-sensitive modules to preserve the source policy's behavioral priors while correcting for the target robot. Extensive experiments on multiple humanoid platforms and pretrained backbones show that Any2Any substantially accelerates convergence and reduces training cost compared with training from scratch, while achieving competitive or superior tracking performance. Notably, using only 1% of the compute and data required for full training, Any2Any successfully transfers Sonic models pre-trained on Unitree G1 to LimX Oli and LimX Luna. These results suggest that pretrained robot-specific WBT policies can be efficiently reused across embodiments, providing a scalable path toward deploying humanoid whole-body control on new robots. More results and videos are available on our project page: https://yming2001.github.io/any2any-project-site/.

1 Introduction

Behavior foundation models (BFMs) [2, 3] are emerging as a promising paradigm for humanoid control. By pretraining on large-scale behavior data, BFMs aim to acquire reusable motor skills and broad behavioral priors that can be rapidly adapted to downstream tasks [4, 5]. For humanoid robots, whole-body tracking (WBT) [6, 7] is a natural BFM-style control interface: the policy learns to reproduce diverse full-body reference motions while coordinating legs, torso, arms, and head to maintain balance. A large-scale WBT policy can therefore serve as a reusable whole-body motor prior for teleoperation [8, 9], motion imitation [10, 1], and downstream loco-manipulation.

Recent humanoid WBT systems have advanced rapidly, scaling up in data, model size, and motion coverage. TWIST [9] and GMT [10] demonstrate that end-to-end WBT policies can cover diverse whole-body motions through reinforcement learning, behavior cloning, teacher-student training, adaptive sampling, and mixture-of-experts designs. SONIC [1] further frames motion tracking as a scalable foundation task, scaling the policy from 1.2M to 42M parameters, using over 100M motion frames, and reporting 9k GPU hours of training. HoloMotion [11] trains a large-scale whole-body control foundation model on extensive motion data, while OmniXtreme [12] scales tracking to high-dynamic, extreme motions without sacrificing fidelity as the motion library grows. These results mark important progress toward general humanoid motor priors, but they also expose a practical barrier: reproducing such natural and robust whole-body behaviors requires large-scale motion data, massively parallel simulation infrastructure, and substantial compute budgets.

Beyond training cost, embodiment dependence [13, 14] remains a fundamental limitation of current large-scale WBT policies. Such policies are tightly coupled to the source robot's morphology and actuator configuration. Consequently, even between similar humanoids, structural and dynamic differences make direct deployment unreliable, while full-policy fine-tuning tends to overwrite the source behavioral prior. Existing cross-embodiment methods mainly address this issue by training universal controllers over many embodiments with morphology randomization or multi-embodiment data [15, 16, 14], or by scaling robot foundation models with embodiment-aware representations, unified action spaces, post-training, or expert routing [17, 18, 19]. However, these approaches typically require large-scale multi-embodiment datasets, broad morphology randomization, or expensive large-scale pretraining from scratch, and they mostly address locomotion or manipulation rather than high-fidelity whole-body tracking. In contrast, we study a complementary post-training problem: given a pretrained robot-specific WBT policy, can we efficiently adapt it to another humanoid with limited data and compute?

A long-standing principle in humanoid control is to decouple kinematics from dynamics: classical pipelines typically plan a kinematic reference and track it with a hierarchical whole-body controller [20, 21], while learning-based whole-body tracking (WBT) policies often encode retargeted reference motions [22] separately from the dynamics-aware control stream [1]. This decoupling reveals the structure of cross-embodiment transfer. A source policy trained on one humanoid differs from a target humanoid along two distinct axes. *Kinematically*, the two robots may have different joint counts, link geometries, and observation/action layouts, making the pretrained policy structurally incompatible with the target embodiment. *Dynamically*, they differ in mass distributions, inertias, actuator responses, and contact behaviors, so even after interface alignment, the source policy may still produce suboptimal or incorrect actions on the target. Bridging this dynamics gap requires parameter updates, but directly fine-tuning the full policy risks overwriting the very behavioral prior that makes it valuable. This kinematics-dynamics separation is also reflected in modern WBT network designs, which commonly adopt a two-stream structure consisting of a reference motion encoder and an action decoder [9, 10, 1, 23, 24, 11]. The encoder primarily extracts motion-related, largely kinematic features that are more transferable across embodiments, whereas the decoder is more tightly coupled to embodiment-specific dynamics. We therefore hypothesize that embodiment changes affect different components of the WBT prior unevenly, i.e. global motor skills such as balance and inter-limb coordination may remain broadly reusable, while embodiment-sensitive modules require adaptation. This motivates a localized adaptation strategy, where lightweight PEFT components [25, 26] such as LoRA [27] and adapters [28] are inserted only into modules with the

largest source-target discrepancy, while the remaining parameters are frozen to preserve the source prior. This leads to three key questions: *Can localized fine-tuning adapt to the target embodiment without destroying the source prior? How should the kinematic mismatch be resolved? Which modules should be adapted to maximize cross-embodiment transfer?*

To answer these questions, we propose ANY2ANY, a parameter-efficient post-training framework that decomposes cross-embodiment WBT transfer into two complementary steps, *kinematic alignment* and *dynamics adaptation*. Our contributions are threefold:

1. We demonstrate that humanoid WBT policies pretrained on a single source robot can be effectively transferred to other humanoid embodiments through principled handling of the kinematic and dynamics mismatches. To the best of our knowledge, this constitutes the first systematic study of cross-embodiment transfer for humanoid WBT.
2. We propose ANY2ANY, a cross-embodiment post-training framework for humanoid WBT that decomposes transfer into two stages: *kinematic alignment*, which resolves the structural kinematic mismatch between source and target robots, and *dynamics adaptation*, which adapts only the dynamics-sensitive modules while freezing the rest of the pretrained backbone.
3. We validate Any2Any on five source-target transfer pairs spanning two pretrained WBT backbones and four target humanoid embodiments, achieving successful transfer with only $\sim$1% of the compute and data of training from scratch, and deploy the adapted policies on real hardware across multiple downstream tasks.

2 Related Works

Humanoid Whole-Body Tracking. Humanoid control has progressed from task-specific motion imitation to large-scale whole-body tracking (WBT). DeepMimic [29] demonstrates that physics-based reinforcement learning can imitate reference motions and produce natural full-body behaviors, while PHC [30] improves scalable humanoid motion tracking from large-scale unstructured human motion data. Recent studies further extend WBT to real humanoids, including teleoperation [9], general motion tracking [10], long-horizon closed-loop control [31], and balance-intensive motions [32]. SONIC scales this paradigm with larger policies and motion datasets, showing the potential of WBT as a foundation-level motor interface [1]. However, these policies are still trained and tuned for a specific robot embodiment [32, 33, 12]. Their observation layout, action space, reward design, and dynamics randomization are tightly coupled to one morphology. In contrast, ANY2ANY studies how to transfer an existing WBT policy and efficiently adapt it to a different humanoid embodiment.

Cross-Embodiment Robot Learning. Cross-embodiment learning aims to share control knowledge across robots with different morphologies. Existing methods commonly train generalist controllers over multiple embodiments using morphology randomization [15], topology-aware architectures [34], unified observation-action spaces [16], or residual adaptation modules [14]. In robot foundation models, cross-embodiment generalization is often addressed through large-scale pretraining [35, 36], embodiment-aware architectures [17], and unified action/state representations with expert routing [18, 19]. These methods are powerful, but they typically require multi-embodiment datasets, broad robot coverage, and expensive training from scratch. This is difficult for WBT, where high-quality specialist policies already require substantial data and compute. ANY2ANY addresses a complementary setting: given one pretrained WBT specialist, we transfer it to a target humanoid through kinematic alignment and lightweight policy adaptation, without rebuilding a multi-robot generalist model.

Parameter-Efficient Fine-tuning for Robotics. A direct way to adapt a pretrained model is full fine-tuning, which updates all model parameters for the target domain [37]. Parameter-efficient fine-tuning instead freezes most pretrained weights and updates only a small set of task-specific parameters. Adapter tuning inserts trainable bottleneck modules [28]; Prefix-Tuning optimizes contin-

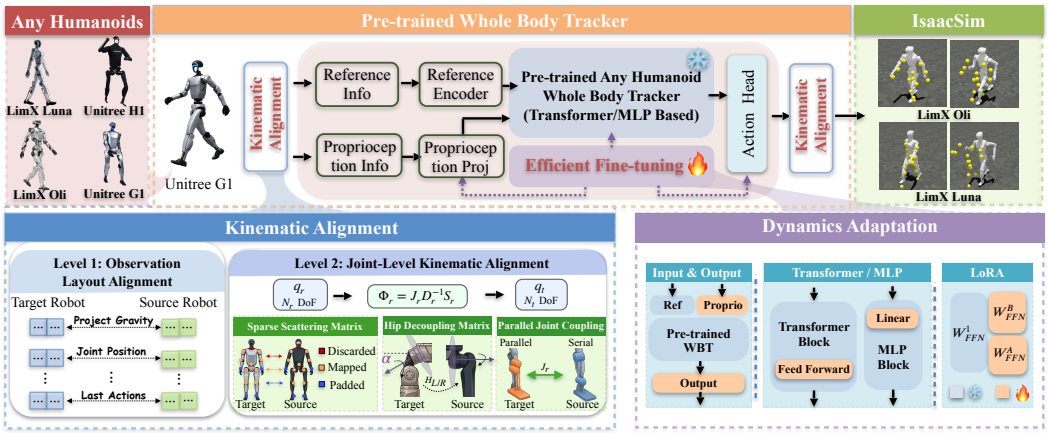


Figure 2: **Architecture of Any2Any.** The proposed framework adapts a pretrained whole-body tracker to arbitrary humanoid embodiments by combining **Kinematic Alignment**, which maps observations and actions across different robot morphologies, with **Dynamics Adaptation**, which efficiently fine-tunes lightweight modules to account for target-specific dynamics.

uous prefix tokens while keeping the backbone frozen [26]; LoRA represents weight updates with low-rank factors [27]. Recent works have begun to adapt PEFT techniques to robotic foundation models, enabling efficient specialization to downstream tasks and embodiments without retraining the entire policy [38, 39]. However, adapting a closed-loop humanoid WBT policy is fundamentally different from adapting a static prediction model: small parameter changes can propagate through contact dynamics, balance regulation, and action feedback, leading to non-trivial shifts in the resulting state distribution. It therefore remains unclear which PEFT mechanism and where PEFT components should be inserted, and to what extent the motion priors encoded in an existing WBT policy can be reused on a structurally different humanoid. In this work, ANY2ANY aims to fill this gap by introducing a systematic paradigm tailored for humanoid whole-body tracking, which transfers pretrained WBT policy to new humanoids with only a small amount of adaptation data and compute.

3 Method

3.1 Problem Formulation

We formulate humanoid whole-body tracking (WBT) as a Markov Decision Process (MDP) $\mathcal{M} = (\mathcal{X}, \mathcal{A}, \mathcal{P}, r, \gamma)$, with state space $\mathcal{X}$, action space $\mathcal{A}$, transition kernel $\mathcal{P}$, reward r, and discount γ. For a given robot embodiment, $\mathcal{X}$ and $\mathcal{A}$ are determined by its morphological properties (number of actuated joints, kinematic structure, link masses, inertias, actuator gains) and by the specific tracking task.

Let $\mathcal{E}_\mathcal{S}$ denote a source robot equipped with a pretrained WBT policy $\pi_{\theta_\mathcal{S}} : \mathcal{X}_\mathcal{S} \to \mathcal{A}_\mathcal{S}$, where $\theta_\mathcal{S}$ is its full parameter set. For a target robot $\mathcal{E}_\mathcal{T}$, $\mathcal{X}_\mathcal{S} \neq \mathcal{X}_\mathcal{T}$ and $\mathcal{A}_\mathcal{S} \neq \mathcal{A}_\mathcal{T}$ due to differences in joint configuration and degrees of freedom. Re-training the full parameter set $\theta_\mathcal{S}$ on $\mathcal{E}_\mathcal{T}$ is computationally expensive and risks catastrophic forgetting of the transferable motor priors encoded in $\theta_\mathcal{S}$.

We instead freeze $\theta_\mathcal{S}$ and learn a small target-specific adaptation $\Delta\theta_\mathcal{T}$ with $|\Delta\theta_\mathcal{T}| \ll |\theta_\mathcal{S}|$. The resulting target policy is

$$\pi_{\theta_\mathcal{T}}(a_t \mid s_t) \;=\; \pi_{\theta_\mathcal{S} \oplus \Delta\theta_\mathcal{T}}(a_t \mid s_t), \tag{1}$$

where $\oplus$ denotes the parameter-efficient injection of $\Delta\theta_\mathcal{T}$ into selected modules of the frozen backbone. The adaptation is trained under the standard expected-return objective

$$\max_{\Delta\theta_\mathcal{T}} \; \mathbb{E}_{\pi_{\theta_\mathcal{T}}}\left[\sum_{t=0}^{T} \gamma^t \, r(s_t, a_t)\right]. \tag{2}$$

Different PEFT mechanisms such as LoRA, Adapter, and Prefix Tuning realize the operator $\oplus$ at different locations within the network, yielding distinct trade-offs in parameter efficiency, training stability, and tracking quality. Identifying which positions and mechanisms yield the most effective cross-embodiment WBT adaptation under limited data and compute is the focus of this work.

3.2 Pretrained Whole Body Tracking Policy Structure

We adopt an actor–critic framework trained with Proximal Policy Optimization (PPO) for motion imitation. The actor π_θ outputs target joint position offsets, while the critic V_ϕ estimates state values using privileged observations (e.g., base velocity, contact forces, body mass distribution) that are available only during training. To demonstrate the generalizability of the proposed method, we instantiate the actor with three representative policy backbones that share a unified observation formulation but differ in how they encode and aggregate the inputs. The first two backbones are trained under the Oli whole-body tracking setting and are collectively referred to as **Oli-WBT**.

Unified Observation Formulation. At each control step t, the policy input is organized into two components: proprioceptive feedback and reference motion information.

- **Proprioception** ($\phi_t \in \mathbb{R}^{d_p}$): the robot-centric state, including base angular velocity, projected gravity, joint positions, joint velocities, and the previous action. These observations provide feedback about the current physical state and recent control history of the robot.
- **Reference** ($g_t \in \mathbb{R}^{d_r}$): the target motion state extracted from the reference trajectory, which provides task-level guidance for motion imitation.

MLP Backbone. The MLP policy flattens multi-modal observations over $H+1$ timesteps into a vector of dimension $(H+1)(d_p+d_r)$, which is then processed by an L-layer MLP with GELU activations to predict actions. The critic uses the same architecture with privileged observations as input. This simple yet effective design is widely adopted in recent humanoid whole-body tracking works [8, 7].

Transformer Backbone. The Transformer policy models temporal dependencies by projecting each modality into a shared d-dimensional embedding space and processing the resulting sequence with N causal Transformer Decoder blocks. The action is decoded from the current-timestep representation, while the critic follows the same architecture with privileged observations.

Sonic Backbone. We further adopt Sonic [1] as a representative large-scale humanoid motion tracking architecture. In our experiments, we employ its Robot Motion Encoder, FSQ bottleneck, and dynamics decoder modules for training and evaluation. More architectural details can be found in [1].

3.3 Cross-Embodiment Adaptation

Since the source and target robots may differ in their degrees of freedom ($n_j^S \neq n_j^T$), their observation and action spaces cannot be directly shared. Before applying any PEFT method, we introduce a kinematic alignment step to construct a unified joint-level representation across embodiments.

3.3.1 Kinematic Alignment

To bridge the embodiment gap between the source and target robots before any parameter adaptation, we introduce a two-level kinematic alignment module that operates on the policy's input and output streams (Fig. 2). The first level aligns the *global observation layout*, while the second level aligns the *joint-level kinematic representation*. Together, they guarantee that the frozen source policy always receives observations with a consistent semantic layout and a compatible joint ordering, while the target robot still executes actions in its own control convention.

Level 1: Observation Layout Alignment. Different humanoid platforms often organize their observation vectors in different orders, even when the underlying entries are semantically equivalent.

At this level, we rearrange the target robot's observation components, such as base states, reference motion features, proprioceptive states (e.g., projected gravity, joint positions), and action history, to match the input convention of the pretrained source policy, so that each term lands at the position expected by the frozen backbone.

Level 2: Joint-Level Kinematic Alignment. The second level aligns the internal ordering and geometry of joint-related variables. Consider a family of robots $\mathcal{R} = \{r_1, r_2, \dots\}$ that share the same humanoid topology but differ in joint count, where each robot r has N_r degrees of freedom. We treat the joint layout of the pretrained source robot r_0, with $N_{r_0} = T$, as a unified kinematic semantic space, and for every $r \in \mathcal{R}$ we construct a pair of mappings $\Phi_r : \mathbb{R}^{N_r} \to \mathbb{R}^T$ and $\Phi_r^+ : \mathbb{R}^T \to \mathbb{R}^{N_r}$ between the target and source joint spaces. On the joints shared by both robots, Φ_r^+ inverts Φ_r, so that $\Phi_r^+ \Phi_r$ is the identity on these matched target joints and reduces to I_{N_r} when every target joint has a source counterpart. The mapping Φ_r is composed of three structured components that progressively account for joint-level discrepancies between the two embodiments.

(i) Sparse Scattering Matrix. Joint correspondence between the target and source robots is specified by a partial injection $\pi_r : \{0, \dots, N_r - 1\} \rightharpoonup \{0, \dots, T - 1\}$ defined on the target joints that have a source counterpart, which induces a sparse scattering matrix $S_r \in \{0, 1\}^{T \times N_r}$ with $(S_r)_{ij} = \mathbf{1}[\pi_r(j) = i]$. Each matched target joint yields a column with a single nonzero entry, so S_r scatters target values into the source layout: source positions without a target counterpart become zero rows (zero-padding), while surplus target joints without a source counterpart lie outside the domain of π_r and become zero columns (dropped).

(ii) Hip Decoupling Matrix. The source embodiment contains an inclined hip-pitch axis, which induces a coupling between the corresponding hip coordinates. To compensate for this structural difference, we introduce a hip decoupling matrix $D_r \in \mathbb{R}^{T \times T}$. Specifically, D_r is initialized as a $T \times T$ identity matrix, and its left and right hip submatrices are replaced by the following decoupling blocks:

$$H_{\mathrm{L/R}} \;=\; \begin{pmatrix} \cos\alpha & 0 \\ \mp \sin\alpha & 1 \end{pmatrix}, \tag{3}$$

where the opposite signs account for the mirrored hip-axis orientations on the left and right legs. In this way, D_r acts only on the hip-related coordinates and leaves all other joints unchanged.

Combining the scattering matrix S_r and the hip decoupling matrix D_r gives the aligned forward and inverse mappings

$$\Phi_r \;=\; D_r^{-1}\, S_r, \qquad \Phi_r^+ \;=\; S_r^\top\, D_r. \tag{4}$$

Here, S_r first scatters target-robot joint quantities into the source joint layout, and D_r^{-1} further converts the hip coordinates into the source-aligned convention. Conversely, D_r maps the policy output back from the source-aligned convention, and $S_r^\top$ gathers the corresponding entries for the target robot.

Structured joint-related terms in the observation, including reference motion, joint position observations, and action history, are mapped to the source space via $\widetilde{\mathbf{q}}_r = \Phi_r\, \mathbf{q}_r$. The same rule is applied to the action space: the policy output $\widetilde{\mathbf{a}}$ produced under the source-aligned joint convention is converted back to the target robot's actuated joints through $\mathbf{a}_r = \Phi_r^+\, \widetilde{\mathbf{a}}$ for execution.

(iii) Parallel Joint Coupling. Beyond serial-chain joint correspondence, some target robots contain closed-chain mechanisms, such as parallelogram-driven ankles or closed-loop waists, whose actuated joint values do not coincide with their kinematic counterparts in the source serial chain. For these joints, we incorporate a closed-chain Jacobian correction $J_r \in \mathbb{R}^{T \times T}$ into the mapping, so that the effective joint-level alignment becomes

$$\Phi_r \;=\; J_r\, D_r^{-1}\, S_r, \qquad \Phi_r^+ \;=\; S_r^\top\, D_r\, J_r^{-1}, \tag{5}$$

which preserves this relation on the matched joints. This term explicitly handles parallel-to-serial discrepancies that purely permutation-based alignment cannot capture.

Through these two levels of alignment, the frozen source policy operates on a stable kinematic semantic space, while embodiment-specific details, such as joint ordering, hip-axis inclination, and

closed-chain coupling, are absorbed by the alignment module itself. This consistently reduces the embodiment gap before parameter adaptation, and provides a more reliable basis on which subsequent PEFT methods can specialize the policy to the target humanoid.

3.3.2 Dynamics Adaptation

Once kinematic alignment is in place, the remaining gap between the source embodiment $\mathcal{S}$ and the target embodiment $\mathcal{T}$ is no longer representational but *dynamical*. Here, $\mathcal{S}$ denotes the humanoid embodiment on which the whole-body tracker is originally pretrained, while $\mathcal{T}$ denotes the new target humanoid embodiment to which the tracker is transferred. After kinematic alignment, we use q to denote the joint state expressed in the shared source-aligned joint convention. The rigid-body dynamics of a humanoid admits the standard manipulator-equation form

$$M(q)\,\ddot{q} + C(q,\dot{q})\,\dot{q} + G(q) \;=\; \tau + \tau_{\text{ext}}, \tag{6}$$

which makes explicit that an embodiment's physical identity is carried by three state-dependent terms: the mass matrix $M(q)$, the Coriolis/centrifugal coupling $C(q,\dot{q})$, and the gravity loading $G(q)$. Under the same motion target $(q,\dot{q},\ddot{q})$, two humanoids with different dynamics $(M_{\mathcal{S}}, C_{\mathcal{S}}, G_{\mathcal{S}})$ and $(M_{\mathcal{T}}, C_{\mathcal{T}}, G_{\mathcal{T}})$ require different generalized forces to realize that motion, yielding the rigid-body residual

$$\Delta\tau(q,\dot{q},\ddot{q}) \;=\; \Delta M(q)\,\ddot{q} \;+\; \Delta C(q,\dot{q})\,\dot{q} \;+\; \Delta G(q), \tag{7}$$

where $\Delta M = M_{\mathcal{T}} - M_{\mathcal{S}}$, $\Delta C = C_{\mathcal{T}} - C_{\mathcal{S}}$, and $\Delta G = G_{\mathcal{T}} - G_{\mathcal{S}}$. Beyond these rigid-body terms, the realized force-motion relation also depends on unmodeled effects such as actuator characteristics, joint friction, and contact behavior, which contribute additional residual components on top of $\Delta\tau$. Collecting all embodiment-specific physical quantities into a single descriptor η_e, the full cross-embodiment dynamics gap can be characterized by

$$\Delta\eta \;=\; \eta_{\mathcal{T}} - \eta_{\mathcal{S}}. \tag{8}$$

For humanoids of similar topology and scale, $\Delta\eta$ is structurally low-dimensional: only a limited subset of physical quantities differs between the two robots, and the rigid-body part $(\Delta M, \Delta C, \Delta G)$ is fully parameterized by per-link inertial differences, whose number scales with the link count rather than with the size of the full backbone $\theta_{\mathcal{S}}$. Once kinematic alignment has removed the dominant structural mismatch, $\Delta\eta$ becomes the primary remaining source of embodiment-dependent policy error. The pretrained policy has already captured the task structure (reference tracking, balance, inter-limb coupling) in a source-aligned representation; what remains is a compact, embodiment-specific correction whose capacity, we hypothesize, should scale with $\Delta\eta$ rather than with $\theta_{\mathcal{S}}$. We therefore adapt the target robot through a low-rank correction rather than full-parameter fine-tuning, understanding it as a learned representation of the source–target dynamics residual rather than a generic perturbation of the source weights.

This motivates a strongly asymmetric adaptation: we freeze the pretrained parameters θ and learn an embodiment-specific low-rank correction. Concretely, for each adapted linear projection $W \in \mathbb{R}^{d_{\text{out}} \times d_{\text{in}}}$ in the policy, we apply a LoRA decomposition

$$W' \;=\; W \;+\; BA, \qquad A \in \mathbb{R}^{k \times d_{\text{in}}},\; B \in \mathbb{R}^{d_{\text{out}} \times k}, \tag{9}$$

with rank $k \ll \min(d_{\text{in}}, d_{\text{out}})$. Only $\{A, B\}$ are trained per target embodiment, while W remains shared with the source. The rank k controls how much capacity is allocated to absorbing the cross-embodiment dynamics gap, providing a tunable trade-off between transfer fidelity and adaptation cost.

Together with the kinematic alignment, this design factorizes the cross-embodiment gap into a fixed structural component absorbed by the alignment, and a learned dynamical component absorbed by LoRA, keeping the source policy's behavioral prior intact while allowing each new embodiment to be specified by a small set of low-rank factors. Concretely, we inject LoRA into the proprioceptive input projection and the actor and critic linear layers, while keeping the reference-motion encoder frozen.

Table 1: **Cross-embodiment benchmark robots.** DoF is reported as leg/arm/waist/head; structural gaps and their role in the kinematic alignment are discussed in the text.

Robot	DoF (l/a/w/h)	Joint conv.	Incl. hip $\rightarrow D_r$	Closed-ch. $\rightarrow J_r$	Height	Mass	Transfer role
LimX Oli	31 (12/14/3/2)	Serial†	✓ 25°	–†	1.65 m	55 kg	Src → G1, H1, Luna; Tgt ← Sonic
LimX Luna	27 (12/10/3/2)	Parallel	✓ 30°	✓	1.60 m	54 kg	Tgt ← Oli-WBT, Sonic
Unitree G1	29‡ (12/14/3/0)	Serial	–	–	1.32 m	35 kg	Src → Oli, Luna; Tgt ← Oli-WBT
Unitree H1	19 (10/8/1/0)	Serial	–	–	1.80 m	47 kg	Tgt ← Oli-WBT

†Oli is physically parallel (achilles ankle and parallel waist), but the WBT policy operates on its serialized model with the parallel mapping resolved at the low level, so J_r is not exercised for Oli. ‡29-DoF body of the `g1_29dof` configuration on which Sonic is pretrained; hand DoF are excluded from WBT. Heights and masses are nominal manufacturer values.

3.4 Training Procedure

All policies are trained with PPO in Isaac Lab. To isolate the effect of cross-embodiment adaptation, we keep the action space, observations, reward formulation, PPO hyperparameters, reference-motion sampling, and domain randomization protocol identical to those of the corresponding source pretraining setup, only the components introduced by our kinematic alignment and dynamics adaptation differ.

4 Experimental Results

Our experiments are designed to test the central hypothesis of Any2Any: a robot-specific pretrained whole-body tracking policy can be efficiently reused on a new humanoid if the cross-embodiment gap is decomposed into a structural kinematic mismatch and a compact dynamics residual.

- **Q1:** Can the ANY2ANY successfully transfer diverse pretrained whole-body tracking policies to novel robotic platforms with varying morphologies?
- **Q2:** How much target-side data and compute does Any2Any save compared with training a target specialist from scratch?
- **Q3:** What are the specific contributions and roles of the individual algorithmic components within the ANY2ANY paradigm ?

4.1 Experimental Protocol

4.1.1 Cross-Embodiment Transfer Benchmark

We build our cross-embodiment benchmark on four humanoid platforms spanning two morphological families and 19–31 degrees of freedom (DoF): LimX Oli [40] (31-DoF), LimX Luna [41] (27-DoF), Unitree G1 [42] (29-DoF), and Unitree H1 [43] (19-DoF), summarized in Table 1. Beyond differences in DoF and body scale, the two families also differ structurally: the LimX robots use an inclined hip-pitch axis (Oli 25°, Luna 30°) and closed-chain parallelogram ankle and waist actuation, whereas the Unitree robots use orthogonal hip axes and fully serial chains. These structural gaps are exactly what the kinematic alignment of ANY2ANY is designed to absorb (S_r for the DoF mismatch, D_r for the inclined hip-pitch, and J_r for the closed-chain actuation), making the benchmark a direct stress test of cross-embodiment transfer.

We use two pretrained WBT policies as source backbones. The first is our self-pretrained **Oli-WBT**, trained on LimX Oli with over 500 hours of motion data using both Transformer and MLP backbones. The second is **Sonic** [1], the open-source Gear-Soni, a large-scale WBT policy pretrained on Unitree G1, of which we reuse the Robot Motion Encoder, FSQ bottleneck, and dynamics decoder. From these backbones we evaluate five transfer instances across two directions: adapting Oli-WBT to Unitree G1, Unitree H1, and LimX Luna yields OLIWBT2G1, OLIWBT2H1, and OLIWBT2LUNA, while adapting Sonic to LimX Oli , LimX Luna, Unitree H1 yields SONIC2OLI,

SONIC2LUNA and SONIC2H1. Each adaptation uses limited target-side compute: 4 NVIDIA A100 GPUs, $\approx$22.5 h of wall-clock on average, which shows that pretrained WBT priors can be reused across embodiments at low cost.

Adaptation and evaluation data. All adaptation is driven by the AMASS [44] motion dataset, retargeted to each target robot through General Motion Retargeting [22]; sharing a single adaptation corpus isolates the effect of the adaptation method across source policies and target embodiments. For evaluation, we hold out two motion benchmarks chosen to cover complementary regimes: **OMOMO** [45], a whole-body object-manipulation dataset (500 selected clips), and **LAFAN** [46], a high-dynamic locomotion dataset (after removing near-ground, crawling, and retargeting-corrupted clips). As neither is seen during adaptation, all deployment metrics are reported on these two held-out sets and thus jointly measure out-of-distribution tracking of manipulation (OMOMO) and dynamic locomotion (LAFAN).

4.1.2 Baselines Design

To evaluate the proposed cross-embodiment adaptation framework, we compare against several representative training and adaptation strategies built on the same PPO pipeline. All methods share identical network architectures, training environments, reward functions, and motion-retargeting pipelines, and crucially are run under the *same compute budget and the same number of training iterations*, so that observed differences reflect the adaptation mechanism rather than the amount of training.

Scratch. Our primary baseline trains the target policy from scratch: it is randomly initialized and optimized with PPO on the retargeted AMASS data, without any pretrained prior. This reflects the conventional per-robot humanoid RL pipeline and is the reference point for adaptation efficiency.

Full fine-tuning. To isolate the value of parameter-efficient adaptation, we also fine-tune the *entire* pretrained backbone in two variants: without kinematic alignment (loading the source weights directly) and with our kinematic alignment; both update all parameters.

PEFT mechanisms. Finally, under the same kinematic alignment, we compare the low-rank adaptation adopted by ANY2ANY (**LoRA** [27]) against two alternative parameter-efficient mechanisms, **Adapter** [28] and **Prefix-Tuning** [26], each inserting lightweight trainable modules while freezing the backbone. LoRA is the mechanism used by ANY2ANY; Adapter and Prefix-Tuning serve as PEFT baselines.

4.1.3 Evaluation Metrics

We evaluate from two complementary perspectives. During training, we monitor the core WBT reward (the *tracking joint position* or *tracking body position* reward) and compare methods by both converged reward and convergence speed, which reflect asymptotic tracking quality and sample efficiency. All such comparisons use the same compute budget and number of iterations.

During deployment in MuJoCo, all metrics are reported on the held-out OMOMO and LAFAN benchmarks. The *success rate* is the fraction of evaluation clips tracked to completion without early termination; following the training termination criteria, a rollout fails once the root height error exceeds 0.25 m, the root orientation error exceeds 0.8 (squared quaternion error, $\approx 46^\circ$ tilt), or any tracked body-link height error exceeds 0.3 m. We further report the Mean Per-Joint Position Error (MPJPE, mm) for joint-level fidelity, the world-frame base position (cm) and orientation (deg) errors for global accuracy, and the mean per-frame velocity (mm/frame) magnitudes of the policy output as smoothness indicator (lower is smoother, a proxy for real-world deployability rather than tracking accuracy).

4.2 Any2Any Transfer Performance

To answer **Q1**, we evaluate ANY2ANY on six cross-embodiment transfer instances and compare it with the training-from-scratch baseline under the same training budget. We analyze the results from

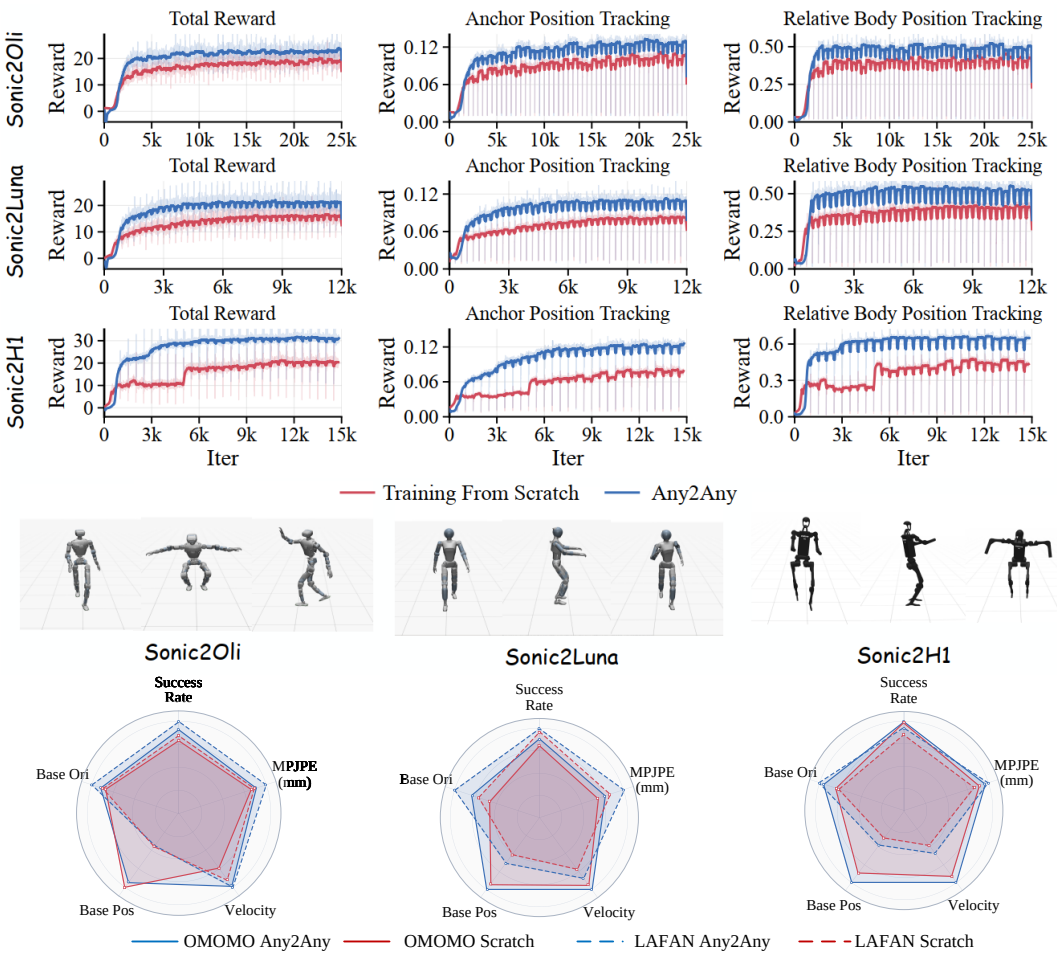


Figure 3: ANY2ANY transfer from Sonic to LimX humanoids, including SONIC2OLI and SONIC2LUNA. The curves compare ANY2ANY with the baseline, and the snapshots show stable rollout motions after adaptation.

two complementary perspectives: training time optimization, which measures convergence speed and converged reward, and held-out deployment performance, which measures whether the learned policy generalizes to unseen motions rather than merely optimizing the training reward.

Sonic as source. For the external Sonic source (SONIC2OLI, SONIC2LUNA, SONIC2H1), LoRA modules are inserted into the actor dynamics decoder and the critic network, while the FSQ module and other pretrained components remain frozen.

As shown in Figure 3, ANY2ANY consistently transfers the external Sonic WBT prior to three representative target humanoids. These targets cover different types of embodiment gaps: SONIC2OLI involves transferring from the Unitree G1 source to a larger Oli humanoid with an inclined hip structure; SONIC2LUNA further introduces parallel/closed-chain mechanisms; and SONIC2H1 transfers to a lower-DoF serial-chain humanoid with a different body scale. Despite these different kinematic and dynamic gaps, ANY2ANY shows the same trend across all three targets: the total reward rises faster and converges to a higher value than Scratch, and the gains are also consistent in both anchor-position tracking and relative-body-position tracking rewards.

These results directly support the core design of ANY2ANY. Training from scratch must rediscover whole-body tracking, balance regulation, and embodiment-specific coordination entirely from target-side interaction. In contrast, ANY2ANY first uses kinematic alignment to place the target observations and actions into the source policy's semantic space, so that the pretrained Sonic WBT

prior can be reused. LoRA then adapts only the remaining target-specific dynamics residual. Therefore, the faster convergence and higher converged rewards across Oli, Luna, and H1 show that ANY2ANY provides an efficient post-training paradigm for reusing robot-specific WBT policies across different humanoid embodiments.

The radar plots further evaluate whether the training-time gains transfer to unseen motion distributions. We report results on two held-out benchmarks with complementary properties: OMOMO and LAFAN. For SONIC2OLI, ANY2ANY consistently improves over Scratch on both benchmarks. On LAFAN, the improvement is even more pronounced: the success rate increases from 83.1 % to 90.6 % , MPJPE decreases from 69.7 mm to 47.2 mm, and the base-position and base-orientation errors decrease from 10.48 cm to 8.49 cm and from 7.3° to 5.6°. These results indicate that the transferred Sonic prior improves both manipulation-style tracking and dynamic locomotion tracking on Oli. For SONIC2LUNA, ANY2ANY also provides consistent gains despite the larger structural mismatch introduced by Luna's inclined hip and closed-chain/parallel mechanisms. For SONIC2H1, ANY2ANY mainly improves stability, joint-level fidelity, orientation tracking, and smoothness. Overall, the held-out radar results show that ANY2ANY does not merely optimize the training reward. Across OMOMO and LAFAN, it generally improves completion rate, joint-level tracking, root-orientation tracking, and action smoothness.

The rollout snapshots provide qualitative evidence that the adapted policies remain stable across different target humanoids and diverse motions. Together with the faster reward convergence and improved held-out tracking metrics, these results indicate that ANY2ANY does not learn whole-body control from scratch. Instead, it preserves the reusable WBT prior from the source policy through kinematic alignment and only learns a compact target-specific dynamics correction. Beyond simulation, we further deploy the adapted SONIC2OLI policy on the physical LimX Oli, where it reliably reproduces a range of generalized behaviors such as walking, running, turning, and dancing (Figure 8), confirming that the transferred prior remains effective under real-world dynamics.

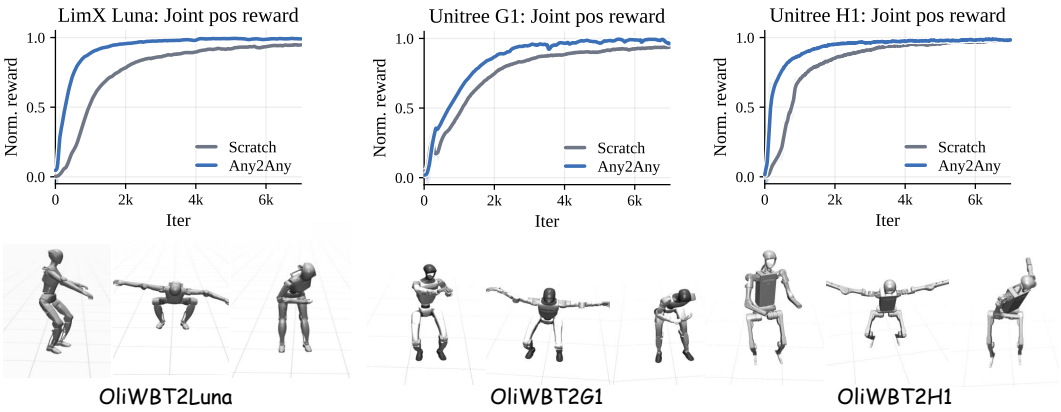


Figure 4: ANY2ANY transfer from Oli-WBT to three target humanoids: OLIWBT2LUNA, OLIWBT2G1, and OLIWBT2H1. ANY2ANY is compared with the baseline trained from scratch. (a) Tracking-error radar plots. (b) Training curves of normalized tracking reward. (c) Sim-to-sim rollouts on diverse motions.

Oli-WBT as source. We further evaluate whether the same transfer paradigm also holds for our self-pretrained Oli-WBT source policy. As shown in Figure 4, ANY2ANY transfers Oli-WBT to three representative target humanoids: OLIWBT2LUNA, OLIWBT2G1, and OLIWBT2H1.

The training curves show a consistent optimization advantage. On all three targets, ANY2ANY reaches the high-reward regime much earlier than Scratch. For OLIWBT2LUNA, ANY2ANY rapidly increases the normalized joint-position reward. The same trend appears on OLIWBT2G1 and OLIWBT2H1, ANY2ANY enters the high-reward region earlier and converges to a higher or comparable final reward. This indicates that the pretrained Oli-WBT policy already contains

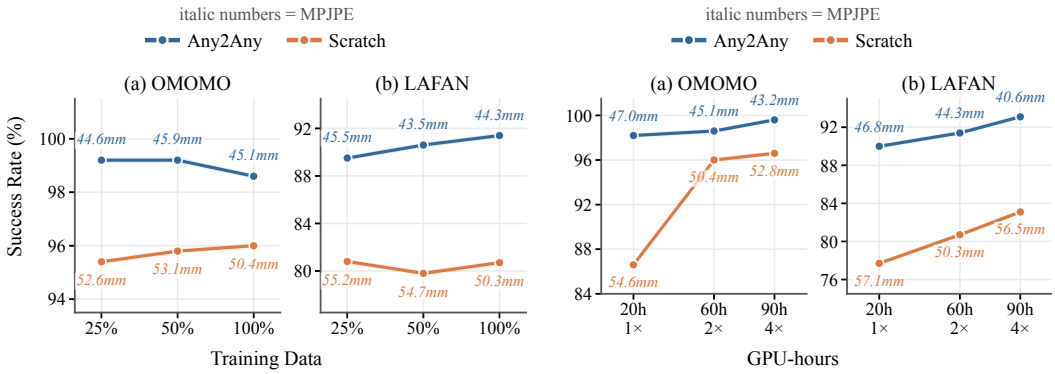


Figure 5: **Data and compute efficiency on SONIC2OLI.** (a) Held-out success rate (MPJPE annotated) on OMOMO and LAFAN as the adaptation data is scaled over $25/50/100\%$ of AMASS under a fixed iteration budget. (b) Training reward versus parallel-sampling budget (one/two/four A100 GPUs at 8,000 iterations). ANY2ANY stays above training-from-scratch across all data and compute budgets.

reusable joint-level tracking and whole-body coordination priors, and the target policy does not need to rediscover them from scratch.

The radar plots further show that the training-time gains transfer to unseen motion distributions. On OMOMO, which emphasizes whole-body object-manipulation motions, ANY2ANY generally improves joint-level and root tracking. On LAFAN, which contains more dynamic locomotion motions, ANY2ANY also maintains clear advantages in stability and tracking accuracy. These results indicate that the transferred Oli-WBT prior generalizes to both manipulation-style and high-dynamic locomotion motions.

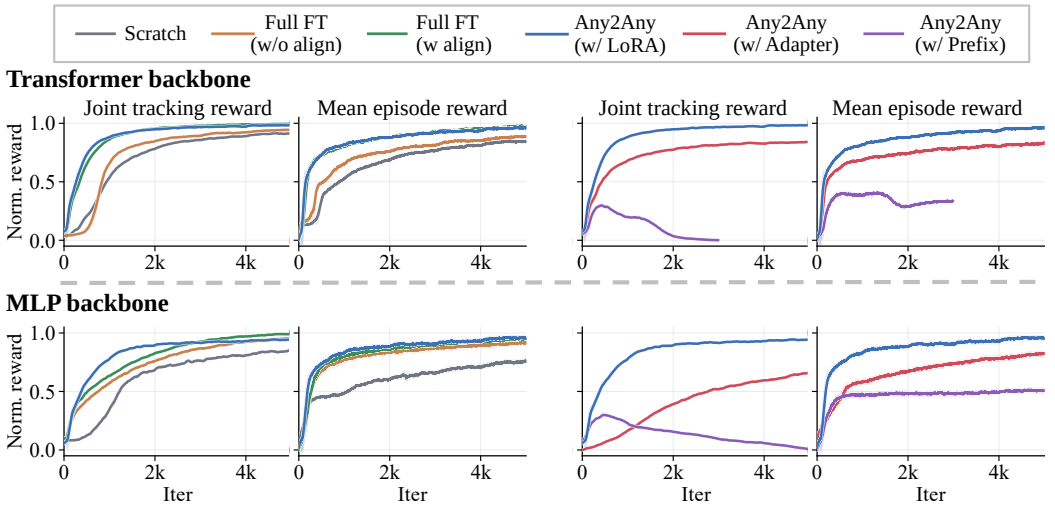
Overall, the Oli-WBT results are consistent with the Sonic-transfer results and further support the core hypothesis of ANY2ANY. Across Luna, G1, and H1, ANY2ANY achieves faster reward convergence and stronger held-out tracking performance than training from scratch. This shows that the proposed post-training paradigm is not tied to a specific pretrained backbone. By first aligning the target robot with the source policy's kinematic semantic space and then adapting only a compact dynamics residual, ANY2ANY efficiently reuses robot-specific WBT priors across different humanoid embodiments.

4.3 Data and Compute Efficiency

To address **Q2**, we study how strongly ANY2ANY depends on the amount of adaptation data and on the parallel-simulation budget. All experiments in this section use the unified SONIC2OLI task and compare ANY2ANY against training from scratch under matched settings (Figure 5).

Data efficiency. We vary the amount of adaptation data, using 25%, 50%, and 100% of the AMASS corpus, train every variant for the same number of iterations, and evaluate on both held-out benchmarks. On both OMOMO and LAFAN, ANY2ANY achieves a higher success rate and better joint-level tracking than scratch at every data scale. Interestingly, on OMOMO the success rate and tracking accuracy of ANY2ANY slightly *decrease* as more data is added. We attribute this to the near-static nature of OMOMO manipulation motions: a small amount of in-place data is already sufficient to learn the relatively simple quasi-static tracking, so additional data yields little benefit and mostly adds variance. On LAFAN, whose motions are highly dynamic, the expected trend emerges: ANY2ANY improves markedly as more data becomes available, whereas scratch stays at a consistently low level. Overall, the transferred WBT prior lets the target policy reach broad whole-body tracking from far less data than training from scratch.

Method	Trainable Params↓	FPS↑	Collection Cost↓	Learning Cost↓	Joint Reward↑	Total Reward↑
Full FT (w/ align)	100.00%	173.0 k	**3.90**	4.48	0.43	19.04
ANY2ANY (LoRA)	**5.26%**	**196.0 k**	4.53	**3.07**	**0.54**	**23.86**



(a) Kinematics alignment ablation (b) PEFT method ablation

Figure 6: Ablation of ANY2ANY architectural components on OLIWBT2LUNA. The top table compares aligned full fine-tuning and ANY2ANY with LoRA. (a) Kinematic alignment ablation. (b) PEFT method ablation under kinematic alignment. ANY2ANY-LoRA achieves comparable rewards to full fine-tuning while using fewer trainable parameters and lower training cost.

Compute efficiency. We then fix the data and vary the parallel-sampling budget, using one, two, and four A100 GPUs with the same number of environments per GPU, and compare all runs at 8,000 iterations. As the GPU budget grows, ANY2ANY improves only slightly, while scratch fluctuates substantially yet stays below ANY2ANY throughout. That is, scratch relies heavily on large-scale parallel sampling to discover whole-body control, whereas ANY2ANY already reaches a near-complete result under a small compute budget.

Together, these two studies show that, by reusing the WBT motion prior, ANY2ANY is far less dependent on large adaptation datasets and large-scale parallel simulation than training from scratch, attaining strong tracking under limited data and compute.

4.4 Ablation Analysis

To address **Q3**, we conduct architectural ablations on the representative OLIWBT2LUNA transfer instance using both Transformer and MLP backbones. These ablations examine whether the two stages of ANY2ANY are both necessary: kinematic alignment for resolving structural mismatch, and localized LoRA adaptation for absorbing the remaining dynamics residual.

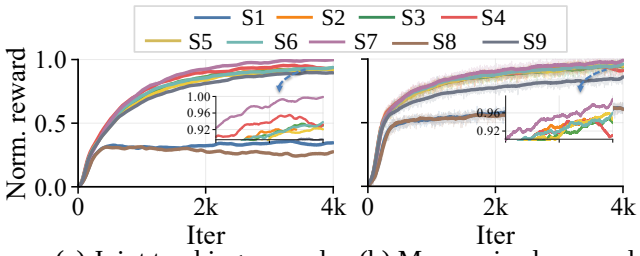
Figure 6(a) studies the effect of kinematic alignment. We compare four settings: training from scratch, full fine-tuning without alignment, full fine-tuning with alignment, and ANY2ANY with LoRA. Across both Transformer and MLP backbones, training from scratch converges slowly and reaches a lower final reward, indicating that learning whole-body tracking directly on the target embodiment remains data- and compute-intensive. Full fine-tuning without alignment improves over scratch, but its convergence and final reward are still limited. This suggests that the pretrained WBT prior cannot be effectively reused when the target robot's observations and actions are interpreted under inconsistent joint semantics. After kinematic alignment is introduced, full fine-tuning converges faster and achieves higher tracking rewards, confirming that aligning the observation, reference, and action spaces is necessary before transferring a pretrained policy across embodiments.

Notably, ANY2ANY with LoRA achieves performance comparable to aligned full fine-tuning while updating only a small fraction of the policy parameters. As shown in the table above Figure 6, LoRA reduces the trainable parameters from 100% to 5.26%, improves FPS from 34.8 k to 111.7 k. Although its collection cost is slightly higher, the overall adaptation cost is lower, and the final joint and total rewards increase from 0.43 and 19.04 to 0.54 and 23.86. These results show that, once kinematic semantics are aligned, adapting the entire pretrained policy is unnecessary: a lightweight low-rank update is sufficient to specialize the source WBT prior to the target embodiment.

Figure 6(b) further compares different PEFT mechanisms under the same aligned setting. LoRA consistently provides the best balance between training stability, convergence speed, and final performance for both Transformer and MLP backbones. Adapter tuning also improves over training from scratch, but converges more slowly and reaches lower rewards, suggesting that its additional bottleneck modules are harder to optimize in closed-loop whole-body control. Prefix tuning is the least stable: its reward either saturates at a low value or degrades during training, indicating that modifying only the input conditioning is insufficient to compensate for embodiment-level dynamical mismatch. Overall, these results support the design choice of ANY2ANY: kinematic alignment first establishes a shared semantic space, and LoRA then provides an efficient residual adaptation mechanism for the remaining embodiment-specific dynamics.

Figure 7 further studies where the LoRA residual should be injected after kinematic alignment. The table groups the candidate injection sites into actor and critic components, including the actor backbone, reference input projection, proprioception input projection, action output projection, critic backbone, and critic input/output projections. This component-level ablation allows us to examine which parts of the pretrained WBT policy carry embodiment-agnostic motion priors and which parts require target-specific dynamics correction.

Setting	Actor				Critic		
	Backbone	Ref. In.	Prop. In.	Out.	Backbone	In.	Out.
S1	✓						
S2	✓				✓		
S3	✓	✓			✓		
S4	✓		✓		✓		
S5	✓			✓	✓		
S6	✓				✓	✓	
S7	✓		✓	✓	✓		
S8	✓	✓	✓	✓			
S9	✓	✓	✓	✓	✓	✓	✓



(a) Joint tracking reward (b) Mean episode reward

Figure 7: Ablation of LoRA injection scopes on OLI-WBT2LUNA. The table summarizes the component-level injection locations across actor and critic modules, while the curves show the resulting joint tracking reward and mean episode reward.

The best performance is achieved by S7, which applies LoRA to the actor backbone, actor proprioception input projection, actor output projection, and critic backbone. This indicates that the dominant residual mismatch after kinematic alignment lies in the dynamics-aware control pathway: the policy needs to reinterpret the target robot's proprioceptive state and recent action history, and also adjust how the shared motion representation is decoded into target-specific joint commands.

This finding is consistent with the design principle of ANY2ANY. After the observation, reference, and action spaces are kinematically aligned, the reference-motion stream becomes relatively embodiment-agnostic and should be largely preserved. In contrast, the proprioception stream, action output head, and critic backbone are more directly coupled to the target robot's mass distribution, actuator response, contact behavior, and closed-loop stability. Adapting these modules allows the policy to absorb the remaining dynamics residual while keeping the high-level WBT prior intact. Therefore, S7 provides the best trade-off between target-specific dynamics adaptation and preservation of reusable whole-body motion priors.

Figure 8: Snapshots of the SONIC2OLI real-world experiments. Trained on the AMASS dataset for 90 GPU-hours, the policy performs a variety of behaviors on the LimX Oli, including walking, running, turning, and dancing.

Overall, the ablations validate the design of ANY2ANY. Kinematic alignment is necessary to establish a shared semantic space; LoRA is more effective than alternative PEFT mechanisms for closed-loop humanoid WBT adaptation; and the best injection scope is concentrated on dynamics-sensitive modules rather than the entire policy. These results support the central decomposition of ANY2ANY: structural mismatch is handled by alignment, while the remaining dynamics gap is handled by localized low-rank adaptation.

5 Discussion and Conclusion

We presented ANY2ANY, a post-training paradigm that transfers a pretrained robot-specific whole-body tracking (WBT) policy to a new humanoid using only about 1% of the data and compute of training from scratch. ANY2ANY rests on a single decomposition of the cross-embodiment gap: a structural *kinematic* mismatch, resolved by an alignment that renders the target robot's observations and actions interpretable to the frozen source policy, and a residual *dynamics* gap, absorbed by a compact low-rank correction injected only into the dynamics-sensitive modules. Across six transfers spanning two pretrained backbones (Oli-WBT and Sonic) and four humanoids (LimX Oli, LimX Luna, Unitree G1, and H1), this recipe yields faster convergence, stronger held-out tracking, and reliable real-world deployment, while modifying only about 5% of the policy parameters.

We read this result as more than an efficient fine-tuning trick. To the best of our knowledge, it is the first evidence that a large-scale WBT prior is *not* intrinsically bound to the embodiment on which it was trained. We attribute this to a deeper structure: despite their diverse morphologies, humanoids are implicitly anchored to a common, human-centric motor manifold—their motion data is sourced from humans, their movements imitate human behavior, and their mechanical design itself references the human body. A pretrained WBT policy therefore encodes reusable structure such as balance regulation, inter-limb coordination, motion timing, and contact-aware tracking, which is largely shared across embodiments rather than specific to one robot. Once the embodiment-specific observation–action mismatch is removed, much of this prior can be preserved and reused, leaving only a low-dimensional, embodiment-specific dynamics residual to be learned. This is precisely why such aggressive data and compute savings are attainable, and we regard it as an encouraging signal that future humanoid policies can be explored far more boldly than the per-robot, train-from-scratch convention suggests.

This view also suggests a concrete design principle for cross-embodiment WBT: treat a compute-heavy WBT backbone as a reusable building block for motion understanding and whole-body coordination, and equip each new humanoid with only a lightweight kinematic mapping and a small

dynamics adapter. Rather than retraining a controller for every robot, or rebuilding a multi-robot generalist from scratch, one starts from an existing expert and pays only a small adaptation cost, accelerating deployment on new platforms.

Looking forward, our findings point to two complementary directions for making cross-embodiment transfer the rule rather than the exception. *First*, the breadth of our experiments leads us to advocate for a community *Reference Humanoid*: a standardized platform designed to be simultaneously human-like and easy to transfer to, serving as a common hub and benchmark for cross-embodiment transfer, and accelerating research not only on basic whole-body motion but also on higher-level skills such as loco-manipulation and dexterous manipulation. *Second*, and complementarily, transferability can be built in at pretraining time: training across a diverse set of robots together with structural randomization should yield priors that are inherently easier to migrate to unseen embodiments. These two directions reinforce each other: a shared reference platform gives multi-robot pretraining a natural anchor, while priors engineered for transfer make each new humanoid cheaper to bring online.

In summary, for the first time, we show that a large-scale humanoid WBT policy can be transferred to a new humanoid with only a small fraction of the original data and compute. ANY2ANY is only the beginning of this journey, but it suggests that reusable, cross-embodiment whole-body intelligence is within reach.

Acknowledgments

References

[1] Z. Luo, Y. Yuan, T. Wang, C. Li, S. Chen, F. Castaneda, Z.-A. Cao, J. Li, D. Minor, Q. Ben, et al. Sonic: Supersizing motion tracking for natural humanoid whole-body control. *arXiv preprint arXiv:2511.07820*, 2025.

[2] M. Pirotta, A. Tirinzoni, A. Touati, A. Lazaric, and Y. Ollivier. Fast imitation via behavior foundation models. In *International Conference on Learning Representations*, volume 2024, pages 12685–12724, 2024.

[3] M. Yuan, T. Yu, W. Ge, X. Yao, D. Li, H. Wang, J. Chen, B. Li, W. Zhang, W. Zeng, et al. A survey of behavior foundation model: Next-generation whole-body control system of humanoid robots. *IEEE transactions on pattern analysis and machine intelligence*, 2025.

[4] E. Cetin, A. Touati, and Y. Ollivier. Finer behavioral foundation models via auto-regressive features and advantage weighting. *arXiv preprint arXiv:2412.04368*, 2024.

[5] Y. Li, Z. Luo, T. Zhang, C. Dai, A. Kanervisto, A. Tirinzoni, H. Weng, K. Kitani, M. Guzek, A. Touati, et al. Bfm-zero: A promptable behavioral foundation model for humanoid control using unsupervised reinforcement learning. *arXiv preprint arXiv:2511.04131*, 2025.

[6] Z. Gu, J. Li, W. Shen, W. Yu, Z. Xie, S. McCrory, X. Cheng, A. Shamsah, R. Griffin, C. K. Liu, et al. Humanoid locomotion and manipulation: Current progress and challenges in control, planning, and learning. *IEEE/ASME Transactions on Mechatronics*, 31(2):2300–2330, 2026.

[7] X. Cheng, Y. Ji, J. Chen, R. Yang, G. Yang, and X. Wang. Expressive whole-body control for humanoid robots. *arXiv preprint arXiv:2402.16796*, 2024.

[8] T. He, Z. Luo, X. He, W. Xiao, C. Zhang, W. Zhang, K. Kitani, C. Liu, and G. Shi. Omnih2o: Universal and dexterous human-to-humanoid whole-body teleoperation and learning. *arXiv preprint arXiv:2406.08858*, 2024.

[9] Y. Ze, Z. Chen, J. P. Araújo, Z.-a. Cao, X. B. Peng, J. Wu, and C. K. Liu. Twist: Teleoperated whole-body imitation system. *arXiv preprint arXiv:2505.02833*, 2025.

[10] Z. Chen, M. Ji, X. Cheng, X. Peng, X. B. Peng, and X. Wang. Gmt: General motion tracking for humanoid whole-body control. *arXiv preprint arXiv:2506.14770*, 2025.

[11] M. Chen, K. Wang, B. Zhang, X. Ma, Z. Yang, Y. Ren, Q. Huang, Z. Zhu, Y. Wang, and Z. Su. Holomotion-1 technical report. *arXiv preprint arXiv:2605.15336*, 2026.

[12] Y. Wang, S. Zhu, P. Zhi, Y. Li, J. Li, Y.-L. Li, Y. Xiao, X. Wang, B. Jia, and S. Huang. Omnixtreme: Breaking the generality barrier in high-dynamic humanoid control. *arXiv preprint arXiv:2602.23843*, 2026.

[13] A. Gupta, L. Fan, S. Ganguli, and L. Fei-Fei. Metamorph: Learning universal controllers with transformers. *arXiv preprint arXiv:2203.11931*, 2022.

[14] S. Yang, Z. Fu, Z. Cao, J. Guo, P. Wensing, W. Zhang, and H. Chen. Multi-loco: Unifying multi-embodiment legged locomotion via reinforcement learning augmented diffusion. *arXiv preprint arXiv:2506.11470*, 2025.

[15] Y. Xue, Y. Lin, W. Dong, Y. Tang, J. Wang, J. Pang, M. Zhou, M. Liu, and W. Zhang. Scalable and general whole-body control for cross-humanoid locomotion. *arXiv preprint arXiv:2602.05791*, 2026.

[16] M. Liu, D. Pathak, and A. Agarwal. Locoformer: Generalist locomotion via long-context adaptation. In *Proceedings of The 9th Conference on Robot Learning*, 2025.

[17] J. Bjorck, F. Castañeda, N. Cherniadev, X. Da, R. Ding, L. Fan, Y. Fang, D. Fox, F. Hu, S. Huang, et al. Gr00t n1: An open foundation model for generalist humanoid robots. *arXiv preprint arXiv:2503.14734*, 2025.

[18] H. Luo, Y. Wang, W. Zhang, S. Zheng, Z. Xi, C. Xu, H. Xu, H. Yuan, C. Zhang, Y. Wang, et al. Being-h0.5: Scaling human-centric robot learning for cross-embodiment generalization. *arXiv preprint arXiv:2601.12993*, 2026.

[19] S. Bai, M. Li, X. Lv, J. Wang, X. Wang, F. Liao, C. Hou, L. Gu, W. Zhou, K. Wu, et al. Hex: Humanoid-aligned experts for cross-embodiment whole-body manipulation. *arXiv preprint arXiv:2604.07993*, 2026.

[20] D. Kim, J. Lee, J. Ahn, O. Campbell, H. Hwang, and L. Sentis. Computationally-robust and efficient prioritized whole-body controller with contact constraints. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1–8. IEEE, 2018. doi:10.1109/IROS.2018.8593767.

[21] M. Chignoli, D. Kim, E. Stanger-Jones, and S. Kim. The mit humanoid robot: Design, motion planning, and control for acrobatic behaviors. In *2020 IEEE-RAS 20th International Conference on Humanoid Robots (Humanoids)*, pages 1–8. IEEE, 2021. doi:10.1109/HUMANOIDS47582.2021.9555782.

[22] J. P. Araujo, Y. Ze, P. Xu, J. Wu, and C. K. Liu. Retargeting matters: General motion retargeting for humanoid motion tracking. *arXiv preprint arXiv:2510.02252*, 2025.

[23] W. Zeng, S. Lu, K. Yin, X. Niu, M. Dai, J. Wang, and J. Pang. Behavior foundation model for humanoid robots. *arXiv preprint arXiv:2509.13780*, 2025.

[24] T. Zhu, G. Cai, Y. Zhaohui, G. Ren, H. Xie, Z. Wang, J. Wu, J. Wang, X. Yang, Y. Mu, et al. Clot: Closed-loop global motion tracking for whole-body humanoid teleoperation. *arXiv preprint arXiv:2602.15060*, 2026.

[25] N. Ding, Y. Qin, G. Yang, F. Wei, Z. Yang, Y. Su, S. Hu, Y. Chen, C.-M. Chan, W. Chen, et al. Parameter-efficient fine-tuning of large-scale pre-trained language models. *Nature Machine Intelligence*, 5(3):220–235, 2023. doi:10.1038/s42256-023-00626-4. URL https://www.nature.com/articles/s42256-023-00626-4.

[26] X. L. Li and P. Liang. Prefix-tuning: Optimizing continuous prompts for generation. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics*, pages 4582–4597, 2021. doi:10.18653/v1/2021.acl-long.353. URL https://aclanthology.org/2021.acl-long.353/.

[27] E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, and W. Chen. Lora: Low-rank adaptation of large language models. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=nZeVKeeFYf9.

[28] N. Houlsby, A. Giurgiu, S. Jastrzebski, B. Morrone, Q. de Laroussilhe, A. Gesmundo, M. Attariyan, and S. Gelly. Parameter-efficient transfer learning for nlp. In *International Conference on Machine Learning*, pages 2790–2799, 2019. URL https://arxiv.org/abs/1902.00751.

[29] X. B. Peng, P. Abbeel, S. Levine, and M. van de Panne. Deepmimic: Example-guided deep reinforcement learning of physics-based character skills. *ACM Transactions on Graphics*, 37(4):1–14, 2018. doi:10.1145/3197517.3201311. URL https://arxiv.org/abs/1804.02717.

[30] Z. Luo, J. Cao, A. Winkler, K. Kitani, and W. Xu. Perpetual humanoid control for real-time simulated avatars. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023. URL https://openaccess.thecvf.com/content/ICCV2023/

html/Luo_Perpetual_Humanoid_Control_for_Real-time_Simulated_Avatars_ICCV_2023_paper.html.

[31] Y. Li, Y. Lin, J. Cui, T. Liu, W. Liang, Y. Zhu, and S. Huang. Clone: Closed-loop whole-body humanoid teleoperation for long-horizon tasks. In *Proceedings of The 9th Conference on Robot Learning*, 2025. URL https://arxiv.org/abs/2506.08931.

[32] Y. Pan, R. Qiao, L. Chen, K. Chitta, L. Pan, H. Mai, Q. Bu, H. Zhao, C. Zheng, P. Luo, et al. Agility meets stability: Versatile humanoid control with heterogeneous data. *arXiv preprint arXiv:2511.17373*, 2025.

[33] Z. Sun, B.-S. Huang, Y. Peng, X. Li, J. Ma, Y. Sun, Z. Li, H. Jiang, B. Gao, Z. Bing, et al. Mosaic: Bridging the sim-to-real gap in generalist humanoid motion tracking and teleoperation with rapid residual adaptation. *arXiv preprint arXiv:2602.08594*, 2026.

[34] Y. Lin, M. Liu, Y. Xue, M. Zhou, Y. Yu, J. Pang, and W. Zhang. H-zero: Cross-humanoid locomotion pretraining enables few-shot novel embodiment transfer. *arXiv preprint arXiv:2512.00971*, 2025. URL https://arxiv.org/abs/2512.00971.

[35] Open X-Embodiment Collaboration, A. O'Neill, A. Rehman, A. Gupta, A. Maddukuri, A. Gupta, A. Padalkar, A. Lee, A. Pooley, A. Gupta, et al. Open x-embodiment: Robotic learning datasets and rt-x models. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pages 6892–6903, 2024. doi:10.1109/ICRA57147.2024.10611477. URL https://arxiv.org/abs/2310.08864.

[36] Octo Model Team, D. Ghosh, H. Walke, K. Pertsch, K. Black, O. Mees, S. Dasari, J. Hejna, T. Kreiman, C. Xu, J. Luo, Y. L. Tan, L. Y. Chen, P. Sanketi, Q. Vuong, T. Xiao, D. Sadigh, C. Finn, and S. Levine. Octo: An open-source generalist robot policy. In *Proceedings of Robotics: Science and Systems (RSS)*, 2024. doi:10.15607/RSS.2024.XX.090. URL https://arxiv.org/abs/2405.12213.

[37] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4171–4186. Association for Computational Linguistics, 2019. doi:10.18653/v1/N19-1423. URL https://aclanthology.org/N19-1423/.

[38] M. J. Kim, C. Finn, and P. Liang. Fine-tuning vision-language-action models: Optimizing speed and success, 2025.

[39] Y. Wang, P. Ding, L. Li, C. Cui, Z. Ge, X. Tong, W. Song, H. Zhao, W. Zhao, P. Hou, S. Huang, Y. Tang, W. Wang, R. Zhang, J. Liu, and D. Wang. Vla-adapter: An effective paradigm for tiny-scale vision-language-action model, 2025.

[40] LimX Dynamics. LimX Oli: Full-Size General-Purpose Humanoid Robot. https://www.limxdynamics.com/en/products/oli, 2025. Accessed: 2026-05-22.

[41] LimX Dynamics. LimX Luna Humanoid Robot. https://x.com/LimX_Dynamics, 2026. Official product page not yet publicly available at the time of access; accessed: 2026-05-22.

[42] Unitree Robotics. Unitree G1 Humanoid Robot. https://www.unitree.com/g1, 2024. Accessed: 2026-05-22.

[43] Unitree Robotics. Unitree H1 Universal Humanoid Robot. https://www.unitree.com/h1, 2023. Accessed: 2026-05-22.

[44] N. Mahmood, N. Ghorbani, N. F. Troje, G. Pons-Moll, and M. 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.

[45] J. Li, J. Wu, and C. K. Liu. Object motion guided human motion synthesis. In *ACM Transactions on Graphics (SIGGRAPH Asia)*, 2023.

[46] F. G. Harvey, M. Yurick, D. Nowrouzezahrai, and C. Pal. Robust motion in-betweening. *ACM Transactions on Graphics (SIGGRAPH)*, 39(4), 2020.