Towards Adaptable Humanoid Control via Adaptive Motion Tracking

Tao Huang[2,1] Huayi Wang[1,2] Junli Ren[1] Kangning Yin[1,2] Zirui Wang[1] Xiao Chen[1]
Feiyu Jia[1] Wentao Zhang[1] Junfeng Long[1] Jingbo Wang[1,†] Jiangmiao Pang[1,†]
[1]Shanghai AI Laboratory [2]Shanghai Jiao Tong University [†]Equal Advising
Website: adamimic.github.io Code: https://github.com/InternRobotics/AdaMimic

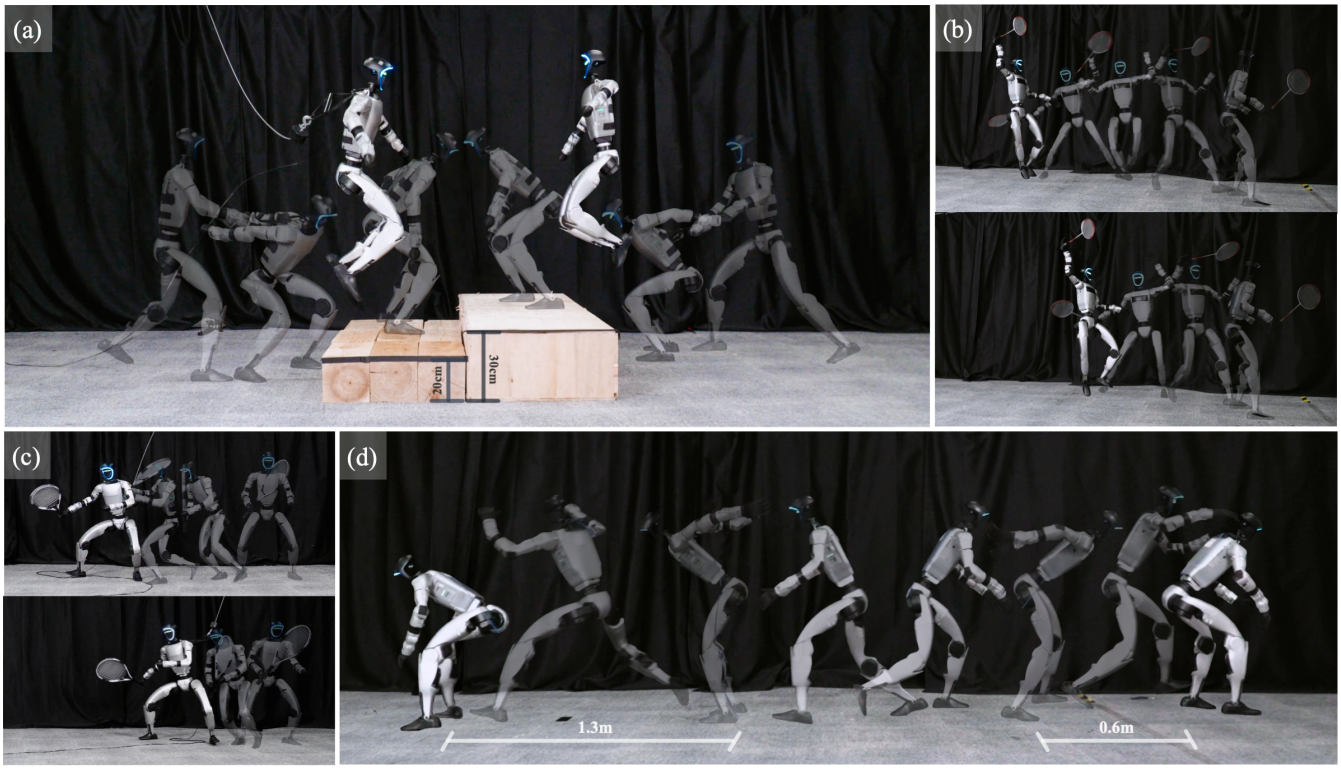

Fig. 1: Overview. Our method, **AdaMimic** (adaptive motion tracking), achieves agile humanoid whole-body adaptation from only a single reference motion while consistently preserving the underlying motion patterns. This enables adaptable control across diverse tasks, as illustrated by the varied outcomes: (a) higher jumping height, (b) extended movement in badminton hitting, (c) extended movement in tennis hitting, and (d) longer jumping distance.

***Abstract*— Humanoid robots are envisioned to adapt demonstrated motions to diverse real-world conditions while accurately preserving motion patterns. Existing motion prior approaches enable well adaptability with a few motions but often sacrifice imitation accuracy, whereas motion-tracking methods achieve accurate imitation yet require many training motions and a test-time target motion to adapt. To combine their strengths, we introduce AdaMimic, a novel motion tracking algorithm that enables adaptable humanoid control from a single reference motion. To reduce data dependence while ensuring adaptability, our method first creates an augmented dataset by sparsifying the single reference motion into keyframes and applying light editing with minimal physical assumptions. A policy is then initialized by tracking these sparse keyframes to generate dense intermediate motions, and adapters are subsequently trained to adjust tracking speed and refine low-level actions based on the adjustment, enabling flexible time warping that further improves imitation accuracy and adaptability. We validate these significant improvements in our approach in both simulation and the real-world Unitree G1 humanoid robot in multiple tasks across a wide range of adaptation conditions (Fig. 1). Videos and code are available on our project page.**

I. Introduction

Humans are good at acquiring whole-body skills by first mimicking experts and then adapting to new situations. For example, a tennis player reproduces the expert stroke pattern given different ball positions. Humanoid robots are expected to have a similar capability: adapt from reference motions to diverse real-world conditions, while accurately imitating motion patterns. We refer to this ability as adaptable humanoid control from reference motions.

Approaches to this motion-based adaptable control fall into two main paradigms. Methods that incorporate reference motions as prior into reinforcement learning (RL) enable adaptation beyond the data [1–5], but they often sacrifice imitation accuracy or require extensive reward tuning. Motion tracking methods, in contrast, can reproduce reference motions accurately with lower reward engineering burden [6–16]. However, their adaptability may be limited by the dependence on large-scale training motions to cover diverse

conditions and on possibly unavailable reference motions for each test-time deployment [17, 18]. How to combine the strengths of both paradigms—accurate imitation and broad adaptability—remains an open challenge.

In this work, we take an initial step towards this challenge by introducing AdaMimic, novel motion tracking algorithm that enables humanoid robots to adapt from a single reference motion while accurately preserving motion patterns. To reduce data dependence while supporting adaptation, we first generate an augmented dataset by sparsifying the reference motion into keyframes and editing a few keyframes with minimal physical assumptions, preserving the essential local pattern of the reference data. In the first stage, the policy is trained to track these sparse keyframes, producing dense intermediate motions that maintain local patterns to the reference. In a second stage, two adapters are jointly learned: a phase adapter that modulates motion speed, and a tracking adapter that compensates for the low-level actions based on the adjustment. This two-stage design enables flexible time warping [19, 20], enhancing both imitation accuracy and adaptability. The resulting policies can be deployed on hardware without requiring additional reference motions.

Extensive evaluations across diverse tasks and conditions, both in simulation and on the real-world Unitree G1 humanoid robot, demonstrate that AdaMimic outperforms existing methods. We overview of its real-world performance in Fig. 1 and summarize our core contributions as follows:

- We introduce AdaMimic, a novel motion tracking algorithm that enables humanoid robots to adapt from a single reference motion while accurately preserving key patterns.
- We validate the effectiveness of AdaMimic in extensive simulations, demonstrating significantly improved imitation accuracy and adaptability than existing methods.
- We deploy the trained policies on the real-world Unitree G1 humanoid robot, showing good performance across wide adaptation conditions in multiple tasks.

II. Related Work

A. Adaptable Humanoid Control

Adaptation to diverse real-world conditions is crucial in humanoid control, spanning both locomotion [22–29] and whole-body manipulation [30–37]. Approaches to these tasks often rely on sophisticated reward engineering and carefully designed sim-to-real transfer pipelines. In parallel, another line of work leverages human motion references to acquire adaptable whole-body skills [1–5, 15]. Their methods typically utilize motion data as priors, prioritizing adaptability over the strict preservation of original motion patterns. In contrast, our work emphasizes accurate tracking alongside adaptation.

B. Humanoid Motion Tracking

Motion tracking is proven effective for accurate imitation of a single reference motion [6–9]. However, its adaptability is often constrained by the scale of data. Recent advances address this limitation by training from massive public datasets [10–18, 38]. While such approaches demonstrate impressive generalization, they typically require substantial data curation, rely on teleoperation or pre-collected reference motions during deployment, and may face challenges in maintaining tracking accuracy for highly agile behaviors. In contrast, our work explores how adaptive tracking can be achieved in agile tasks from only a single reference motion, without the need for additional pre-collected trajectories.

C. Motion Adaptation

A classical approach to adapting humanoid motions is motion editing, which can be achieved either through space-time constraints [19, 39–41] or data-driven generation [42–44]. While both paradigms enable reusing a single motion as input, they often face limitations in physical plausibility caused by over-assumed constraints [41, 45] or under-assumed dynamics [46], limiting real-world deployment. This has motivated physics-based approaches that integrate simulation and reinforcement learning to improve motion plausibility [47–50]. However, such methods often depend on large-scale datasets for interpolation or rule-based plausibility heuristics. Inspired by these two lines of work, we explore an alternative strategy: selecting and editing keyframes from a single motion, and then leveraging RL-based tracking to generate physically plausible in-between motions.

III. Background: Humanoid Motion Tracking

We formulate humanoid motion tracking as a goal-conditioned reinforcement learning (RL) problem within the framework of a Markov decision process (MDP; [51]). The objective is to learn a tracking policy π_{track} that accurately reproduces a reference motion $\hat{\boldsymbol{q}}$.

1) Motion representation: The retargeted reference motion $\hat{\boldsymbol{q}}$ is sampled from a reference dataset $\mathcal{D}_{\text{ref}}^{\text{init}}$, which initially contains a single motion [6, 7]. Each motion trajectory consists of global poses $\hat{\boldsymbol{q}}^{\text{global}}$ (*e.g.*, position and orientation of robot bodies) and local poses $\hat{\boldsymbol{q}}^{\text{local}}$ (*e.g.*, joint angles). A normalized phase variable $\phi \in [0, 1]$ parameterizes the whole reference motion: $\hat{\boldsymbol{q}}_\phi = \left(\hat{\boldsymbol{q}}_\phi^{\text{global}}, \hat{\boldsymbol{q}}_\phi^{\text{local}}\right)$, which advances discretely at each timestep k as:

$$\phi_k = \phi_{k-1} + \Delta\phi_k, \quad \Delta\phi_k = \Delta t \,/\, T_{\hat{\boldsymbol{q}}}, \tag{1}$$

where Δt is typically the simulation timestep and $T_{\hat{\boldsymbol{q}}}$ is the duration of the reference motion.

2) Observations and actions: At timestep k, the observation is defined as $\boldsymbol{o}_k = \left[\hat{\boldsymbol{q}}_{\phi_k}, \boldsymbol{q}_{\phi_k}, \dot{\boldsymbol{\theta}}_k, \dot{\boldsymbol{\omega}}_k, \phi_k\right]$, where $\boldsymbol{q}_{\phi_k}$ denotes current global and local robot poses, $\dot{\boldsymbol{\theta}}$ the joint velocities, and $\dot{\boldsymbol{\omega}}_k$ the base angular velocity. The policy produces a tracking action $\boldsymbol{a}_k \sim \pi_{\text{track}}(\cdot|\boldsymbol{o}_k, \Delta\phi_k)$ conditioned on the current observation, the reference motion, and the fixed phase interval. The action is the target of a PD controller during a control period.

3) Rewards and objectives: At timestep k, the rewards $\boldsymbol{r}_k = (r_k^{\text{global}}, r_k^{\text{local}})$ are aggregated from multiple terms that evaluate different aspects of tracking performance. For clarity, we group them into two categories: global-level rewards r_k^{global} and local-level rewards r_k^{local}. The former encourages accurate tracking of global trajectories, while

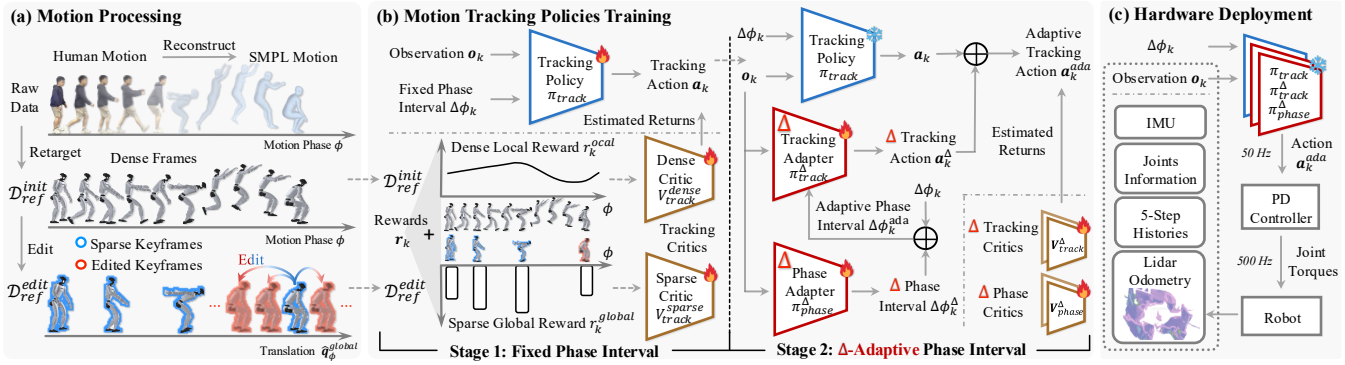

Fig. 2: Method overview. (a) Human motions are reconstructed into SMPL motions via GVHMR [21] and retargeted to the humanoid robot. Sparse keyframes are then selected and edited to form an augmented dataset for adaptive tracking. (b) Based on this dataset, AdaMimic first trains a tracking policy with fixed phase intervals and double critics for sparse global tracking and dense local tracking rewards, followed by phase and tracking adapters that enable effective time warping for improved tracking performance. (c) The resulting policies can be directly deployed on the real Unitree G1 robot.

the latter focuses on matching local motion patterns. The tracking policy π_{track} is then optimized to maximize the expected return:

$$\max_{\pi_{\text{track}}} \mathbb{E}_{\pi_{\text{track}}} \left[\sum_{k=0} \gamma^k \left(\boldsymbol{w} \cdot \boldsymbol{r}_k \right) \middle| \Delta\phi_k, \hat{\boldsymbol{q}} \sim \mathcal{D}_{\text{ref}}^{\text{init}} \right], \quad (2)$$

where $\boldsymbol{w}$ denotes the weighting between the two reward groups, and $\gamma \in [0, 1)$ is the discount factor.

IV. Method: Adaptive Motion Tracking

A. Problem Reformulation

The key insight of adaptive motion tracking is to extend the standard motion tracking problem by allowing variations in global trajectories while strictly preserving the local motion pattern of a given motion. Formally, starting from an initial reference dataset $\mathcal{D}_{\text{ref}}^{\text{init}} = \{\hat{\boldsymbol{q}}\}$ that contains a single motion trajectory, we assume access to an augmented dataset $\mathcal{D}_{\text{ref}}^{\text{edit}} = \{\hat{\boldsymbol{q}}_i^{\text{edit}}\}_{i=0}^{M}$. Each edited motion $\hat{\boldsymbol{q}}_i^{\text{edit}}$ shares the same local joint trajectories as the original motion:

$$\hat{\boldsymbol{q}}_{i,\phi_k}^{\text{edit, local}} = \hat{\boldsymbol{q}}_{\phi_k}^{\text{local}}, \quad \forall\, \hat{\boldsymbol{q}}_i^{\text{edit}} \in \mathcal{D}_{\text{ref}}^{\text{edit}}, \ \forall\, \phi_k \in [0, 1], \quad (3)$$

while the global component $\hat{\boldsymbol{q}}_{\phi_k}^{\text{edit, global}}$ may vary to reflect different displacements or base translations. For clarity, we interchangeably refer to $\hat{\boldsymbol{q}}_i^{\text{edit}}$ and $\hat{\boldsymbol{q}}$ without causing ambiguity in the rest of the paper.

The tracking policy π_{track} is then optimized to reproduce motions sampled from $\mathcal{D}_{\text{ref}}^{\text{edit}}$:

$$\max_{\pi_{\text{track}}} \mathbb{E}_{\pi_{\text{track}}} \left[\sum_{k=0} \gamma^k \left(\boldsymbol{w} \cdot \boldsymbol{r}_k \right) \middle| \Delta\phi_k, \hat{\boldsymbol{q}} \sim \mathcal{D}_{\text{ref}}^{\text{edit}} \right]. \quad (4)$$

This formulation abstracts away the specific mechanism for constructing $\mathcal{D}_{\text{ref}}^{\text{edit}}$, which will be discussed below. Adaptation is achieved by specifying a different reference motion.

B. Keyframing and Editing: $\mathcal{D}_{\text{ref}}^{\text{init}} \rightarrow \mathcal{D}_{\text{ref}}^{\text{edit}}$

A classical approach to construct $\mathcal{D}_{\text{ref}}^{\text{edit}}$ is to sparsely edit keyframes and interpolate the intermediate frames under trajectory-level consistency constraints [19, 39], denoted as $\mathcal{D}_{\text{ref}}^{\text{rule}}$. However, such optimization-based methods may yield physically implausible motions, which may impede

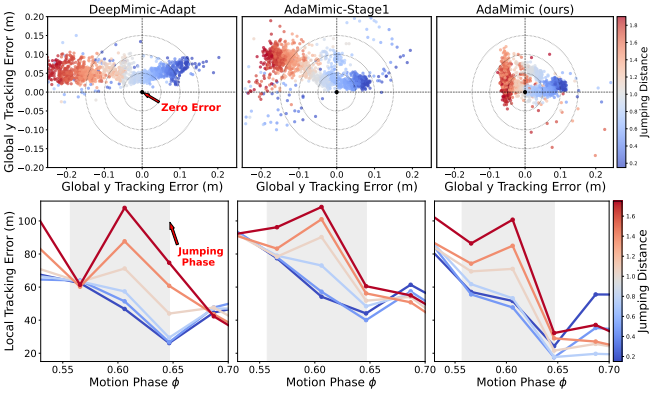

Fig. 3: Motivations of keyframing and adapters. (Top) Global tracking errors at the landing moment of far jumping indicate that AdaMimic outperforms baselines—DeepMimic-Adapt, which augments motions with rules, and AdaMimic-Stage1, which tracks at a fixed speed—by incorporating sparse keyframes and adaptive time warping to improve physical plausibility. (Bottom) Local tracking errors further verify the improvements.

hardware deployment (see Fig. 7), due to the absence of dynamics [41, 45]. Inspired by these works, we also adopt a keyframe-based editing scheme to preserve the global motion structure, but go beyond their limitations by using RL to generate more dynamically plausible intermediate frames, drawing from the ideas of [47, 48, 52].

Concretely, we select N keyframes, some of which are associated with semantic contexts (e.g., start, take-off, and landing in far jumping), and denote their phases as $\boldsymbol{\Phi}^{\text{key}} = \{\phi_0^{\text{key}}, \phi_1^{\text{key}}, \ldots, \phi_{N-1}^{\text{key}}\}$. A subset $\boldsymbol{\Phi}^{\text{edit}} \subset \boldsymbol{\Phi}^{\text{key}}$ is then chosen for editing, with the principle of introducing as few physical assumptions as possible. For each edited keyframe at phase $\phi_k \in \boldsymbol{\Phi}^{\text{edit}}$, we apply a transformation to its global pose based on a variable ψ_i (e.g., jumping distance):

$$\hat{\boldsymbol{q}}_{i,\phi_k}^{\text{edit, global}} = f_{\text{edit}}\left(\hat{\boldsymbol{q}}_{\phi_k}^{\text{global}}, \psi_i\right), \quad \forall\, \phi_k \in \boldsymbol{\Phi}^{\text{edit}}, \quad (5)$$

while keeping the local joint path unchanged. This editing function f_{edit} is task-dependent; for instance, in far jumping, it may translate post-landing frames forward or backward. The resulting set of edited keyframes forms $\mathcal{D}_{\text{ref}}^{\text{edit}}$, which serves as reference motions for adaptive motion tracking. The whole editing process is illustrated in Fig. 4.

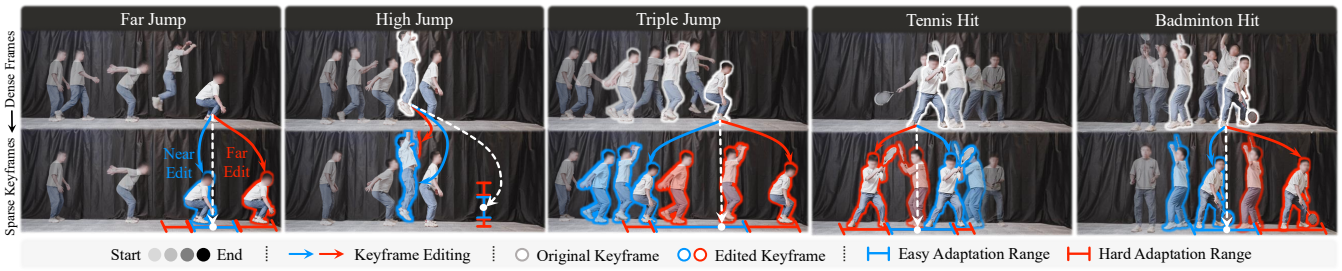

Fig. 4: (Top) Task visualization: Five representative motions used as input for humanoid retargeting. **(Bottom) Keyframing and editing:** Sparse keyframes are extracted from each motion, and few selected ones are further edited to enable adaptation. Colors denote adaptation difficulty relative to the original keyframe (gray): blue indicates easy adaptation cases, while red indicates hard adaptation cases. The adaptation ranges are presented in Table I.

TABLE I: Adaptation ranges of seven tasks during training and testing.

Task	Raw Data	Simulation Train/Test (m)		Hardware Test (m)	
		Easy	Hard	Easy	Hard.
Far Jump →	1.1m	[0.7, 1.6]	[0.2, 0.7] ∪	{0.7, 1.0}	{1.2, 0.4}
High Jump ↑	0.2m	[0.1, 0.35]	[0.05, 0.1] ∪ [0.35, 0.6]	{0.1, 0.2}	{0.3, 0.4}
Triple Jump →	2.4m	[1.65, 3.15]	[1.2, 1.65] ∪ [3.15, 4.2]	{2.1, 2.7}	{1.5, 3.3}
Step Jump ↕	0.2m	[0.05, 0.25]	[0.25, 0.5]	{0.2}	{0.3}
Tennis →	1.0m	[0.8, 1.2]	[1.2, 1.7]	{0.8, 1.2}	{1.4, 1.6}
Badminton ↘	1.3m	[1.3, 1.9]	[1.9, 2.5]	{1.3, 1.6}	{1.9, 2.2}

Remark: The current scope of tasks we consider is limited to those where the keyframes $\mathbf{\Phi}^{\text{key}}$ and editing function f_{edit} can be defined in a relatively straightforward and task-specific way. Extending to more tasks requires developing more general mechanisms for constructing $\mathbf{\Phi}^{\text{key}}$ and f_{edit}, which we leave as an important direction for future work.

C. Stage 1: Motion Tracking with Fixed Phase Interval

Given the edited motions, we first employ existing motion-tracking algorithms [6, 7] to train a tracking policy π_{track} using a fixed phase interval $\Delta\phi$. This stage aims to provide the agent with an initial capability of imitating the edited keyframe motions under a simple and stable training setup.

1) Sparse global reward: To ensure global-space motion alignment, we design a sparse global reward function:

$$r_k^{\text{global}} = \mathcal{R}_{\text{track}}^{\text{global}}(\boldsymbol{q}_{\phi_{k+1}}^{\text{global}}, \hat{\boldsymbol{q}}_{\phi_{k+1}}^{\text{global}}) \cdot \mathbb{1}(\phi_{k+1} \in \mathbf{\Phi}^{\text{key}}), \quad (6)$$

which is only activated when the current phase matches one of the keyframe phases $\mathbf{\Phi}^{\text{key}}$. This design avoids over-constraining the motion in global space and instead enforces accurate alignment only at sparse but crucial keyframes. To stabilize training under this sparse signal, we follow prior works [27, 47] and introduce a separate value function $V_{\text{track}}^{\text{sparse}}$ to better estimate the return from such sparse rewards.

2) Dense local reward: In addition, we design a dense local reward to preserve the local joint-space consistency with the reference motion. This reward encourages the reproduced motion to mimic fine-grained patterns from the references:

$$r_k^{\text{local}} = \mathcal{R}_{\text{track}}^{\text{local}}(\boldsymbol{q}_{\phi_{k+1}}^{\text{local}}, \hat{\boldsymbol{q}}_{\phi_{k+1}}^{\text{local}}), \quad (7)$$

A separate value function $V_{\text{track}}^{\text{local}}$ is employed to estimate the return induced by this dense signal, complementing the sparse global reward. The two value functions $\boldsymbol{V}_{\text{track}} = (V_{\text{track}}^{\text{sparse}}, V_{\text{track}}^{\text{dense}})$ are optimized separately following [53, 54]. All reward functions are listed in Table II.

TABLE II: Reward functions. The tracking rewards largely follow [7] and are categorized into sparse and dense reward groups, with additional regularization rewards to ensure hardware deployability.

Term	Weight	Term	Weight
Sparse Reward → $\boldsymbol{V}^{\text{sparse}}$			
Global body position	10	Global body rotation	5
Global feet position	10	Termination	-200
Dense Reward → $\boldsymbol{V}^{\text{dense}}$			
Local body position	0.75	Local body rotation	0.5
Local DoF position	0.75	Feet orientation	$-5e^{-2}$
DoF acceleration	$-2.5e^{-7}$	DoF velocity	$-5e^{-4}$
Action rate	$-5e^{-1}$	Smoothness	$-1e^{-2}$
Torques	$-1e^{-6}$	Torque limits	5
DoF position limits	-10	DoF velocity limits	-5

D. Stage 2: Adapters Learning with Adaptive Phase Interval

While stage 1 provides a good policy initialization, its adaptation ability to edited motions remains limited. We posit that the fixed phase interval is a key limitation. As illustrated in Fig. 3, this rigidity results in substantial global and local errors, especially when the required adaptation is large. In practice, such errors may manifest as unnatural pacing or artifacts such as unstable landings.

1) Phase adapter $\pi_{\text{phase}}^{\Delta}$: This motivates our design of the phase adapter $\pi_{\text{phase}}^{\Delta}$. It is inspired by the time-warping mechanism in classical motion path editing works, where motions are temporally re-parameterized to preserve naturalness and pacing under spatial edits [20, 48, 52]. Formally, the adapter takes the observation as input and outputs a delta phase interval $\Delta\phi_k^{\Delta}$, which is added to the base interval $\Delta\phi_k$ to obtain an adaptive phase interval $\Delta\phi_k^{\text{ada}}$:

$$\Delta\phi_k^{\text{ada}} = \Delta\phi_k + \Delta\phi_k^{\Delta}, \quad \Delta\phi_k^{\Delta} \sim \pi_{\text{phase}}^{\Delta}(\cdot|\boldsymbol{o}_k). \quad (8)$$

2) Tracking adapter $\pi_{\text{track}}^{\Delta}$: Given the adaptive phase interval, we need a tracking adapter $\pi_{\text{track}}^{\Delta}$ to compensate for the tracking action to track the next-step reference motion:

$$\boldsymbol{a}_k^{\text{ada}} = \boldsymbol{a}_k + \Delta\phi_k^{\Delta} \cdot \boldsymbol{a}_k^{\Delta}, \quad \boldsymbol{a}_k^{\Delta} \sim \pi_{\text{track}}^{\Delta}(\cdot|\boldsymbol{o}_k, \Delta\phi_k^{\text{ada}}), \quad (9)$$

where the scaling by $\Delta\phi_k^{\Delta}$ ensures that when the delta phase interval is zero, the adaptive action degenerates to the original tracking action. This design eases optimization empirically.

3) Optimization: During training, the two adapters $\boldsymbol{\pi}^{\Delta} = (\pi_{\text{phase}}^{\Delta}, \pi_{\text{track}}^{\Delta})$ are paired with separate double critics $\boldsymbol{V}^{\Delta} = (\boldsymbol{V}_{\text{phase}}^{\Delta}, \boldsymbol{V}_{\text{track}}^{\Delta})$ to estimate the sparse and dense rewards described in Section IV-C. The overall objective is

$$\max_{\boldsymbol{\pi}^{\Delta}} \mathbb{E}_{\boldsymbol{\pi}^{\Delta}} \left[\sum_{k=0} \gamma^k (\boldsymbol{w} \cdot \boldsymbol{r}_k) \,\middle|\, \Delta\phi_k, \hat{\boldsymbol{q}} \sim \mathcal{D}_{\text{ref}}^{\text{edit}} \right]. \quad (10)$$

TABLE III: Main simulation results. AdaMimic is compared against multiple baselines adapted to our problem setting and its ablated versions in a fair setup. Results report mean and standard deviation over three evaluations, each comprising more than ten thousand simulation episodes.

Comparison Methods	Components			Easy Adaptation				Hard Adaptation				Overall			
(a) Baselines	r^{global}	$\mathcal{D}_{\text{ref}}$	$\Delta\phi^{\text{ada}}$	Success ↑	$E^{\text{dense}}_{\text{l-bpe}}$ ↓	$E^{\text{sparse}}_{\text{g-bpe}}$ ↓	$E^{\text{dense}}_{\text{smth}}$ ↓	Success ↑	$E^{\text{dense}}_{\text{l-bpe}}$ ↓	$E^{\text{sparse}}_{\text{g-bpe}}$ ↓	$E^{\text{dense}}_{\text{smth}}$ ↓	Success ↑	$E^{\text{dense}}_{\text{l-bpe}}$ ↓	$E^{\text{sparse}}_{\text{g-bpe}}$ ↓	$E^{\text{dense}}_{\text{smth}}$ ↓
AMP-Style	Sparse	$\mathcal{D}^{\text{init}}_{\text{ref}}$	•	95.5%$_{\pm 0.1\%}$	44.5$_{\pm 0.0}$	211.2$_{\pm 0.1}$	19.1$_{\pm 0.3}$	70.3%$_{\pm 0.0\%}$	44.5$_{\pm 0.0}$	247.9$_{\pm 0.1}$	22.3$_{\pm 0.3}$	82.7%$_{\pm 0.0\%}$	44.5$_{\pm 0.0}$	229.8$_{\pm 0.2}$	20.7$_{\pm 0.3}$
AMP-Mimic	Sparse	$\mathcal{D}^{\text{init}}_{\text{ref}}$	•	96.8%$_{\pm 0.0\%}$	35.8$_{\pm 0.0}$	164.4$_{\pm 0.2}$	19.9$_{\pm 0.4}$	62.3%$_{\pm 0.0\%}$	35.7$_{\pm 0.0}$	190.3$_{\pm 0.1}$	21.3$_{\pm 0.4}$	79.1%$_{\pm 0.0\%}$	35.7$_{\pm 0.0}$	177.9$_{\pm 0.1}$	20.6$_{\pm 0.4}$
DeepMimic-NoAdapt	Dense	$\mathcal{D}^{\text{init}}_{\text{ref}}$	•	92.6%$_{\pm 0.0\%}$	36.6$_{\pm 0.0}$	205.1$_{\pm 0.1}$	17.6$_{\pm 0.6}$	**81.0%**$_{\pm 0.0\%}$	36.7$_{\pm 0.0}$	484.6$_{\pm 0.2}$	19.4$_{\pm 0.4}$	86.8%$_{\pm 0.0\%}$	36.6$_{\pm 0.0}$	351.8$_{\pm 0.2}$	18.6$_{\pm 0.5}$
DeepMimic-Adapt	Dense	$\mathcal{D}^{\text{rule}}_{\text{ref}}$	•	95.8%$_{\pm 0.0\%}$	33.3$_{\pm 0.0}$	123.6$_{\pm 0.0}$	**15.1**$_{\pm 0.3}$	74.8%$_{\pm 0.0\%}$	33.3$_{\pm 0.0}$	142.8$_{\pm 0.1}$	**16.7**$_{\pm 0.3}$	85.1%$_{\pm 0.0\%}$	33.3$_{\pm 0.0}$	133.5$_{\pm 0.1}$	**15.9**$_{\pm 0.3}$
DeepMimic-Adapt-$\Delta\phi^{\text{ada}}$	Dense	$\mathcal{D}^{\text{rule}}_{\text{ref}}$	✓	93.7%$_{\pm 0.0\%}$	38.7$_{\pm 0.0}$	170.9$_{\pm 0.0}$	17.2$_{\pm 0.5}$	70.0%$_{\pm 0.0\%}$	38.7$_{\pm 0.0}$	194.2$_{\pm 0.0}$	17.8$_{\pm 0.8}$	81.4%$_{\pm 0.1\%}$	38.8$_{\pm 0.0}$	182.8$_{\pm 0.0}$	17.4$_{\pm 0.5}$
AdaMimic-Dense	Dense	$\mathcal{D}^{\text{rule}}_{\text{ref}}$	✓	98.4%$_{\pm 0.0\%}$	36.3$_{\pm 0.0}$	107.6$_{\pm 0.0}$	17.6$_{\pm 0.1}$	78.0%$_{\pm 0.0\%}$	36.2$_{\pm 0.0}$	124.7$_{\pm 0.0}$	20.0$_{\pm 0.1}$	**88.0%**$_{\pm 0.0\%}$	36.2$_{\pm 0.0}$	115.0$_{\pm 0.0}$	18.8$_{\pm 0.1}$
AdaMimic (ours)	Sparse	$\mathcal{D}^{\text{edit}}_{\text{ref}}$	✓	**99.6%**$_{\pm 0.0\%}$	**30.3**$_{\pm 0.0}$	**87.9**$_{\pm 0.0}$	16.0$_{\pm 0.2}$	74.2%$_{\pm 0.0\%}$	**30.3**$_{\pm 0.0}$	**99.8**$_{\pm 0.1}$	17.3$_{\pm 0.2}$	86.8%$_{\pm 0.0\%}$	**30.3**$_{\pm 0.0}$	**94.8**$_{\pm 0.0}$	16.6$_{\pm 0.2}$
(b) Ablations	Stage 2	Freeze	$\Delta\phi^{\text{ada}}$	Success ↑	$E^{\text{dense}}_{\text{l-bpe}}$ ↓	$E^{\text{sparse}}_{\text{g-bpe}}$ ↓	$E^{\text{dense}}_{\text{smth}}$ ↓	Success ↑	$E^{\text{dense}}_{\text{l-bpe}}$ ↓	$E^{\text{sparse}}_{\text{g-bpe}}$ ↓	$E^{\text{dense}}_{\text{smth}}$ ↓	Success ↑	$E^{\text{dense}}_{\text{l-bpe}}$ ↓	$E^{\text{sparse}}_{\text{g-bpe}}$ ↓	$E^{\text{dense}}_{\text{smth}}$ ↓
AdaMimic-Stage1	•	•	•	96.7%$_{\pm 0.0\%}$	43.4$_{\pm 0.0}$	188.1$_{\pm 0.0}$	17.8$_{\pm 0.4}$	75.2%$_{\pm 0.0\%}$	43.4$_{\pm 0.0}$	211.8$_{\pm 0.1}$	18.7$_{\pm 0.4}$	85.7%$_{\pm 0.0\%}$	43.4$_{\pm 0.0}$	200.4$_{\pm 0.1}$	18.2$_{\pm 0.4}$
AdaMimic-Stage1-$\Delta\phi^{\text{ada}}$	•	•	✓	92.7%$_{\pm 0.0\%}$	45.3$_{\pm 0.0}$	195.0$_{\pm 0.1}$	19.8$_{\pm 0.3}$	**83.2%**$_{\pm 0.0\%}$	45.3$_{\pm 0.0}$	219.8$_{\pm 0.0}$	20.5$_{\pm 0.2}$	**88.0%**$_{\pm 0.0\%}$	45.3$_{\pm 0.0}$	208.1$_{\pm 0.0}$	20.2$_{\pm 0.2}$
AdaMimic-NoFreeze	✓	•	✓	85.0%$_{\pm 0.0\%}$	44.8$_{\pm 0.0}$	203.4$_{\pm 0.0}$	25.8$_{\pm 0.1}$	71.2%$_{\pm 0.0\%}$	43.6$_{\pm 0.0}$	224.2$_{\pm 0.0}$	26.9$_{\pm 0.1}$	77.8%$_{\pm 0.0\%}$	44.1$_{\pm 0.1}$	214.1$_{\pm 0.1}$	26.4$_{\pm 0.1}$
AdaMimic (default)	✓	✓	✓	**99.6%**$_{\pm 0.0\%}$	**30.3**$_{\pm 0.0}$	**87.9**$_{\pm 0.0}$	**16.0**$_{\pm 0.2}$	74.2%$_{\pm 0.0\%}$	**30.3**$_{\pm 0.0}$	**99.8**$_{\pm 0.1}$	17.3$_{\pm 0.2}$	86.8%$_{\pm 0.0\%}$	**30.3**$_{\pm 0.0}$	**94.8**$_{\pm 0.0}$	**16.6**$_{\pm 0.2}$

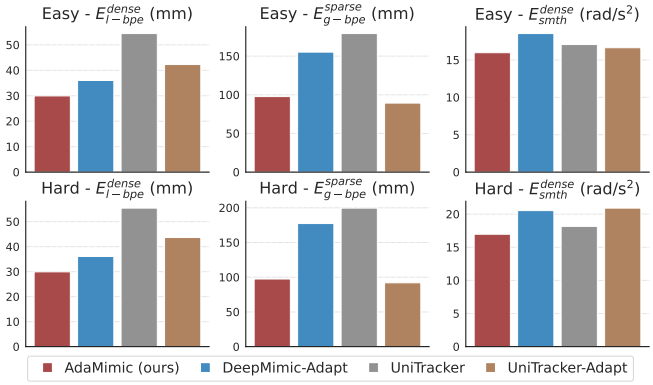

Fig. 5: Baselines trained with large-scale motion data. Across five tasks, AdaMimic achieves better performance than UniTracker [18] and its variant adapted with the rule-based motions from DeepMimic-Adapt. Besides, the results indicate that additional motion data does not provide UniTracker with obvious gains in specialization over DeepMimic-Adapt.

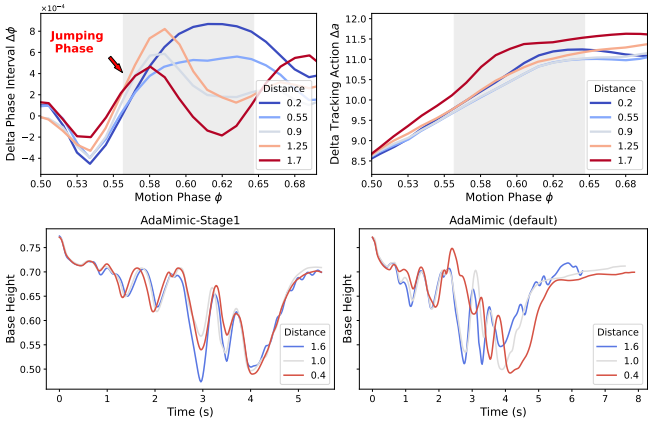

Fig. 6: Effectiveness of adapters. (Top) The phase interval and tracking action adapt to different far-jump distances, where longer distances correspond to extended airtime and larger action compensation, while shorter distances lead to reduced adjustments. (Bottom) With adapters, the policy performs effective time warping, adjusting motion speed to improve adaptation.

Finally, the resulting adaptive action $\boldsymbol{a}_k^{\text{ada}}$ is then applied to control the robot to perform adaptive motion tracking.

E. Real-World Deployment

We directly deploy the trained policies on the Unitree 29-DoFs G1 humanoid robot. For hardware stability, the waist pitch and roll joints are locked. To enhance robustness, the observation is augmented with a 5-step history. Global localization is realized by lidar odometry through FastLIO [55]. The policy, low-level control, and odometry modules operate at 50Hz, 500Hz, and 10Hz, respectively.

F. Implementation Details

The human videos are translated into SMPL motions using GVHMR [21] for retargeting. Training is conducted in the Isaac Gym simulator [56] with 4096 parallel environments, employing PPO [57] as the RL algorithm. Both the policy and value functions are parameterized as 3-layer MLPs. Each episode is initialized from a randomly sampled keyframe [6]. To improve stability for real-world deployment, we adopt L2C2 regularization [54, 58]. The reward weight vector $\boldsymbol{w}$ is set to $(1, 0.5)$ for the sparse and dense reward groups, respectively. The adaptive phase interval is defined as $\Delta\phi_k^{\Delta} \in [-0.75\Delta\phi_k, \phi_k]$. Finally, PD controllers are configured following [54] for improved simulation performance, while employing configurations from [3] during hardware deployment to improve safety and smoothness.

V. Experiments

A. Experimental Setup

1) Tasks: We record seven human videos as the evaluation task set, including five various jumping skills and two ball-hitting skills. These tasks are considered agile and difficult enough. Some representative tasks are visualized in Fig. 4 and their adaptation ranges are presented in Table I.

2) Comparison methods: We adapt the following baselines to our problem formulation:

- **Motion as priors:** Adversarial motion priors (AMP-Style; [1]), which use motions as a style regularizer combined with sparse keyframe tracking as the task reward. Its variant, AMP-Mimic, additionally conditions on phase information to better mimic the reference motion.
- **Motion tracking from target data:** DeepMimic [6], including (i) DeepMimic-NoAdapt, which tracks the original motion $\mathcal{D}^{\text{init}}_{\text{ref}}$, and (ii) DeepMimic-Adapt(-$\Delta\phi^{\text{ada}}$), which tracks the reference motions $\mathcal{D}^{\text{rule}}_{\text{ref}}$ generated with a classical linear motion editing method [52].
- **Motion tracking from prior motions:** UniTracker [18], trained on large-scale motion datasets, and its adapted variant (UniTracker-Adapt), fine-tuned on $\mathcal{D}^{\text{rule}}_{\text{ref}}$.

TABLE IV: Main hardware results. In the absence of accurate odometry, we report success rate, local joint tracking error, and smoothness to quantitatively compare AdaMimic with three representative baselines. The results indicate that AdaMimic achieves strong hardware deployability, particularly in challenging adaptation cases, benefiting from the proposed tracking objectives, motion editing, and adapters. '/' indicates a complete failure in that case.

Easy Adaptation	Far Jump			High Jump			Triple Jump			Tennis Hit			Badminton Hit		
	Succ.↑	E^{dense}_{l-dof}↓	E^{dense}_{smth}↓	Succ.↑	E^{dense}_{l-dof}↓	E^{dense}_{smth}↓	Succ.↑	E^{dense}_{l-dof}↓	E^{dense}_{smth}↓	Succ.↑	E^{dense}_{l-dof}↓	E^{dense}_{smth}↓	Succ.↑	E^{dense}_{l-dof}↓	E^{dense}_{smth}↓
AMP-Style	4/6	35.2±0.2	42.8±2.2	5/6	34.8±0.7	37.4±7.9	4/6	32.6±0.9	41.7±3.5	5/6	31.7±0.5	52.9±4.5	0/6	/	/
DeepMimic-Adapt	4/6	34.4±0.1	41.5±1.0	1/6	32.8±0.0	65.6±0.0	**6/6**	34.0±0.9	49.0±3.3	**6/6**	32.1±0.1	36.5±1.0	**6/6**	30.4±0.2	50.2±2.2
AdaMimic-Stage1	**6/6**	**33.8**±0.2	43.7±2.8	**6/6**	32.7±0.1	35.7±1.5	**6/6**	32.6±0.6	54.4±6.3	**6/6**	31.6±0.1	33.9±0.9	5/6	**28.6**±0.2	44.5±3.5
AdaMimic (our)	5/6	35.2±0.7	**38.7**±1.4	**6/6**	**30.5**±0.3	**28.7**±3.4	**6/6**	**30.7**±0.3	**31.3**±6.7	**6/6**	**30.7**±0.2	**31.9**±1.8	**6/6**	28.7±0.1	**36.4**±2.0
Hard Adaptation	**Far Jump**			**High Jump**			**Triple Jump**			**Tennis Hit**			**Badminton Hit**		
	Succ.↑	E^{dense}_{l-dof}↓	E^{dense}_{smth}↓	Succ.↑	E^{dense}_{l-dof}↓	E^{dense}_{smth}↓	Succ.↑	E^{dense}_{l-dof}↓	E^{dense}_{smth}↓	Succ.↑	E^{dense}_{l-dof}↓	E^{dense}_{smth}↓	Succ.↑	E^{dense}_{l-dof}↓	E^{dense}_{smth}↓
AMP-Style	3/6	35.0±0.0	**31.8**±1.1	0/6	/	/	3/6	33.7±0.2	42.7±8.9	4/6	31.6±0.5	66.4±7.3	0/6	/	/
DeepMimic-Adapt	3/6	34.3±0.1	46.6±4.6	0/6	/	/	**6/6**	33.8±0.1	47.7±2.1	5/6	31.1±0.1	49.5±4.3	**6/6**	30.8±0.1	55.3±1.7
AdaMimic-Stage1	2/6	**34.1**±0.3	34.3±5.4	3/6	32.5±0.1	37.1±3.9	**6/6**	32.3±0.1	38.3±2.3	**6/6**	31.9±0.1	**40.9**±2.2	5/6	29.1±0.1	50.9±2.7
AdaMimic (our)	**5/6**	35.3±0.5	46.2±7.7	**5/6**	**31.3**±0.5	**28.6**±4.2	**6/6**	**31.5**±1.9	**31.7**±4.6	**6/6**	**30.8**±0.1	42.9±1.7	**6/6**	**28.9**±0.1	**44.5**±2.0

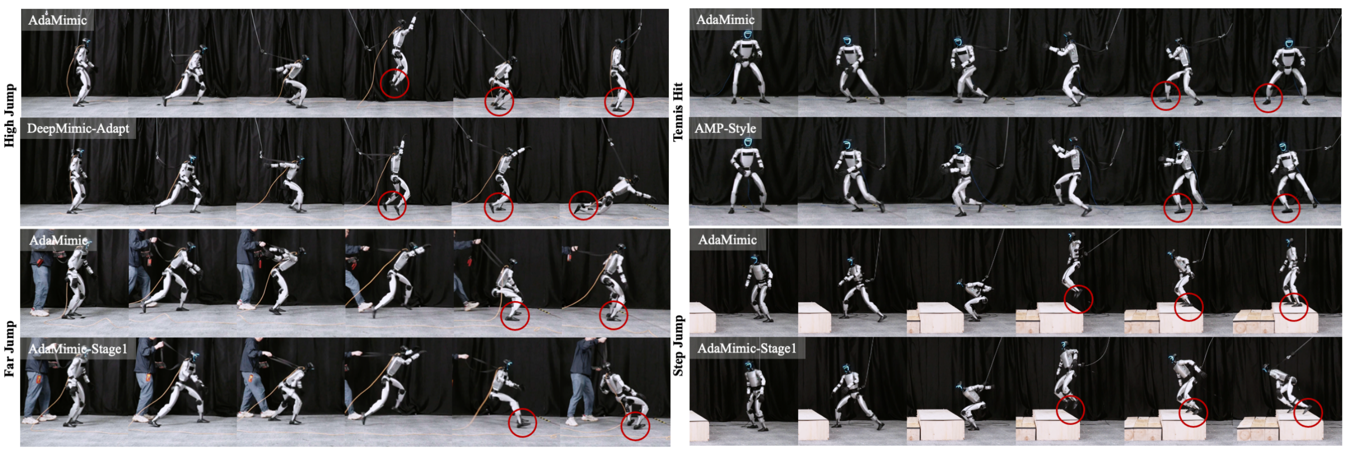

Fig. 7: Snapshots of real robot motions. Four representative groups of hardware demonstrations highlight the differences between AdaMimic and baseline methods: (1) DeepMimic-Adapt is constrained by physically implausible edited motions used for dense per-frame tracking, (2) AMP-Style exhibits jerky motions and limited imitation accuracy, and (3) AdaMimic-Stage1 yields hardware-unstable policies without the proposed adapters.

For fair comparison, we use the same reward functions and observation spaces across all methods, except for UniTracker, which cannot be directly transferred due to its special observations. For this baseline, we use its official implementation.

3) Evaluation metrics: We evaluate each method using four metrics. (1) **Success rate** considers an episode failed if the average body distance error exceeds 1m at any keyframe phase; unlike prior work, we emphasize keyframes and adopt a looser criterion under the sparse tracking setup. (2) **Tracking error** measures body distance error in two forms: the *sparse global error* ($E^{\text{sparse}}_{\text{g-bpe}}$, mm), averaged in the world frame at keyframe phases to assess global mimicking accuracy, and the *dense local error* ($E^{\text{dense}}_{\text{l-bpe}}$, mm), averaged in the root frame over all timesteps to assess local accuracy. (3) **Smoothness** ($E^{\text{dense}}_{\text{smth}}$, rad/s^2) is defined as the average joint acceleration over the entire episode. Results are reported as averages over three evaluation runs, each covering more than ten thousand episodes in IsaacGym.

B. Main Simulation Results

AdaMimic exhibit both great adaptation and tracking precision across all adaptation cases, as shown in Table III.
Comparison with AMP. The AMP-Style baseline demonstrates moderate adaptability but limited imitation accuracy, primarily due to the absence of phase information in its original formulation. Incorporating phase information alleviates this issue and yields improvements in both adaptability and precision. Nonetheless, the resulting motions remain jerky because of weak motion constraints, and overall performance still falls short of other adaptable methods and AdaMimic.
Comparison with DeepMimic. As expected, DeepMimic-noAdapt fails to adapt due to relying on a single reference motion. Its adaptable variants, DeepMimic-Adapt(-$\Delta\phi^{\text{ada}}$), achieve notable performance gains, highlighting the effectiveness and generality of our formulation. However, their tracking performance degrades substantially from easy to hard adaptation scenarios (see also example in Fig. 7), reflecting limitations imposed by the underlying motion path editing. In contrast, our method avoids such restrictive editing assumptions and instead leverages RL with keyframing to produce more physically plausible motions.
Comparison with UniTracker. As shown in Fig. 5, we observe that: (1) the pre-trained model, despite being trained on massive motion data, struggles to perform agile motions; (2) finetuning with $\mathcal{D}^{\text{rule}}_{\text{ref}}$ yields some gains but does not surpass DeepMimic-Adapt trained from scratch; and (3) both versions overall underperform AdaMimic. These findings further suggest that universal trackers could be made less dependent on pre-defined data during deployment by incorporating a more physically plausible motion editing module.

C. Analyses of Key Designs

Effectiveness of adapters. Table III and Fig. 3 quantify the benefits of adapters, while Fig. 6 reveals their mechanism. The top panel shows that phase intervals and delta action scale with jump distance: longer jumps lead to extended air-time and larger compensation, whereas shorter jumps require smaller adjustments, mostly concentrated around landing. Together, this decoupling of timing and correction enables motion speed adaptation without altering the motion pattern, resulting in lower tracking error, smoother landings, and higher success rates, especially in hard adaptation cases.

Keyframe editing outperforms per-frame editing. We observe that keyframe-based AdaMimic consistently outperforms its per-frame counterpart, AdaMimic-Dense, indicating the advantage of sparse keyframe selection for adaptation. Interestingly, the trend is reversed when comparing AdaMimic-Stage1 with DeepMimic-Adapt, suggesting that the combination of keyframing and adapters is crucial for achieving both precise tracking and great adaptation.

Two-stage training is necessary. Our motivation for the two-stage design arises from the observation that training with adaptive phase intervals (AdaMimic-Stage1-$\Delta\phi^{\text{ada}}$) but inference with fixed intervals can even degrade performance compared to training with fixed intervals (AdaMimic-Stage1). We hypothesize that this is due to increased optimization difficulty incurred by the varied phase interval, which can be substantially mitigated by our two-stage design.

Freezing the tracking policy is important. Finetuning the tracking policy during the second stage (AdaMimic-NoFreeze) noticeably degrades both smoothness and tracking accuracy. We attribute this degradation to increased optimization difficulty when updating the policy alongside adapters.

D. Main Hardware Results

We present snapshots of real robot motions to compare AdaMimic with representative baselines in Fig. 7. DeepMimic-Adapt fails in extreme cases, such as high jumps, because the rule-based augmented motions are physically inconsistent and cannot be reliably executed on hardware. AMP-Style shows noticeable instability in dynamic tasks like tennis hitting: motions are jerky, and forceful strikes lead to poor balance and uncoordinated execution. AdaMimic-Stage1, lacking the proposed adapters, exhibits large sim-to-real gaps; in particular, the ankle joints become highly unstable during forceful motions, reflecting inadequate temporal and action adaptation.

In contrast, AdaMimic combines sparse keyframes with phase and tracking adapters, enabling the policy to adjust motion timing and per-step actions while maintaining original motion patterns. This results in physically plausible and robust execution across all tasks. Quantitatively, as shown in Table IV, AdaMimic consistently achieves high success rates, reduced local joint errors, and smoother motions in both easy and hard adaptation scenarios, highlighting the effectiveness and generality of our method on the real robot.

VI. Conclusion

We have presented AdaMimic, a novel motion tracking framework for adaptable humanoid control from a single reference motion. Our framework addresses the limitations of prior methods, which either sacrifice tracking accuracy for adaptation or require large-scale reference motions for each adaptation condition. By leveraging augmented sparse keyframes and a two-stage training strategy with phase and tracking adapters, AdaMimic enables accurate imitation while extending adaptability to diverse tasks and conditions. Experimental results on the Unitree G1 humanoid robot demonstrate that our controllers achieve adaptable motions across many tasks, validating both the effectiveness and generality of our framework. Looking forward, AdaMimic holds promise for extending beyond the demonstrated scenarios to more complex and interactive whole-body skills, such as perceptive and professional ball games.

VII. Limitations and Future Directions

We acknowledge several key limitations that are considered valuable to be addressed in the future.

Wide task scope. Our method currently focuses on tasks with clear parametric forms (e.g., jump or strike distances), while many skills lack such representations. Extending to less structured tasks could broaden its applicability.

General motion editing mechanism. Keyframe selection and editing are manually specified, which restricts scalability. Developing automatic editing mechanisms could make the framework more general and efficient.

Utilization of massive motion data. Universal trackers trained on large datasets still underperform, even with finetuning. Better strategies to exploit massive motion data are valuable for stronger adaptation.

Adaptation for interactive tasks. Our framework executes motions without environment feedback, limiting interactivity. Incorporating perception will enable adaptive and responsive behaviors in interactive tasks such as ball games.

References

[1] X. B. Peng, Z. Ma, P. Abbeel, S. Levine, and A. Kanazawa, "Amp: Adversarial motion priors for stylized physics-based character control," *Transactions on Graphics (TOG)*, 2021.

[2] L. Ma, Z. Meng, T. Liu, Y. Li, R. Song, W. Zhang, and S. Huang, "Styleloco: Generative adversarial distillation for natural humanoid robot locomotion," *ArXiv*, vol. abs/2503.15082, 2025.

[3] Q. Liao, T. E. Truong, X. Huang, G. Tevet, K. Sreenath, and C. K. Liu, "Beyondmimic: From motion tracking to versatile humanoid control via guided diffusion," *ArXiv*, vol. abs/2508.08241, 2025.

[4] H. Xue, X. Huang, D. Niu, Q. Liao, T. Kragerud, J. T. Gravdahl, X. B. Peng, G. Shi, T. Darrell, K. Screenath, and S. S. Sastry, "Leverb: Humanoid whole-body control with latent vision-language instruction," *ArXiv*, vol. abs/2506.13751, 2025.

[5] A. Allshire, H. Choi, J. Zhang, D. McAllister, A. Zhang, C. M. Kim, T. Darrell, P. Abbeel, J. Malik, and A. Kanazawa, "Visual imitation enables contextual humanoid control," in *Conference on Robot Learning (CoRL)*, 2025.

[6] X. B. Peng, P. Abbeel, S. Levine, and M. Van de Panne, "Deepmimic: Example-guided deep reinforcement learning of physics-based character skills," *Transactions on Graphics (TOG)*, 2018.

[7] T. He, J. Gao, W. Xiao, Y. Zhang, Z. Wang, J. Wang, Z. Luo, G. He, N. Sobanbabu, C. Pan, Z. Yi, G. Qu, K. Kitani, J. Hodgins, L. J. Fan, Y. Zhu, C. Liu, and G. Shi, "Asap: Aligning simulation and real-world

physics for learning agile humanoid whole-body skills," in *Robotics Science and Systems (RSS)*, 2025.
[8] W. Xie, J. Han, J. Zheng, H. Li, X. Liu, J. Shi, W. Zhang, C. Bai, and X. Li, "Kungfubot: Physics-based humanoid whole-body control for learning highly-dynamic skills," *ArXiv*, vol. abs/2506.12851, 2025.
[9] T. Zhang, B. Zheng, R. Nai, Y. Hu, Y.-J. Wang, G. Chen, F. Lin, J. Li, C. Hong, K. Sreenath, and Y. Gao, "Hub: Learning extreme humanoid balance," in *Conference on Robot Learning (CoRL)*, 2025.
[10] 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," in *Conference on Robot Learning (CoRL)*, 2024.
[11] T. He, W. Xiao, T. Lin, Z. Luo, Z. Xu, Z. Jiang, J. Kautz, C. Liu, G. Shi, X. Wang, L. Fan, and Y. Zhu, "Hover: Versatile neural whole-body controller for humanoid robots," in *International Conference on Robotics and Automation (ICRA)*, 2025.
[12] J. Shi, X. Liu, D. Wang, O. Lu, S. Schwertfeger, F. Sun, C. Bai, and X. Li, "Adversarial locomotion and motion imitation for humanoid policy learning," vol. abs/2504.14305, 2025.
[13] Y. Xue, W. Dong, M. Liu, W. Zhang, and J. Pang, "A unified and general humanoid whole-body controller for versatile locomotion," in *Robotics Science and Systems (RSS)*, 2025.
[14] X. Cheng, Y. Ji, J. Chen, R. Yang, G. Yang, and X. Wang, "Expressive whole-body control for humanoid robots," in *Robotics Science and Systems (RSS)*, 2024.
[15] Z. Zhuang and H. Zhao, "Embrace collisions: Humanoid shadowing for deployable contact-agnostics motions," in *Conference on Robot Learning (CoRL)*, 2025.
[16] 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 *Conference on Robot Learning (CoRL)*, 2025.
[17] Z. Chen, M. Ji, X. Cheng, X. Peng, X. B. Peng, and X. Wang, "Gmt: General motion tracking for humanoid whole-body control," *ArXiv*, vol. abs/2506.14770, 2025.
[18] K. Yin, W. Zeng, K. Fan, Z. Wang, Q. Zhang, Z. Tian, J. Wang, J. Pang, and W. Zhang, "Unitracker: Learning universal whole-body motion tracker for humanoid robots," *ArXiv*, vol. abs/2507.07356, 2025.
[19] A. Witkin and Z. Popović, "Motion warping," in *SIGGRAPH*. ACM, 1995.
[20] E. Hsu, M. da Silva, and J. Popović, "Guided time warping for motion editing," in *SIGGRAPH*, 2007.
[21] Z. Shen, H. Pi, Y. Xia, Z. Cen, S. Peng, Z. Hu, H. Bao, R. Hu, and X. Zhou, "World-grounded human motion recovery via gravity-view coordinates," in *SIGGRAPH Asia*, 2024.
[22] I. Radosavovic, T. Xiao, B. Zhang, T. Darrell, J. Malik, and K. Sreenath, "Real-world humanoid locomotion with reinforcement learning," *Science Robotics*, 2024.
[23] Z. Li, X. B. Peng, P. Abbeel, S. Levine, G. Berseth, and K. Sreenath, "Reinforcement learning for versatile, dynamic, and robust bipedal locomotion control," *The International Journal of Robotics Research (IJRR)*, 2024.
[24] Z. Zhuang, S. Yao, and H. Zhao, "Humanoid parkour learning," in *Conference on Robot Learning (CoRL)*, 2024.
[25] J. Long, J. Ren, M. Shi, Z. Wang, T. Huang, P. Luo, and J. Pang, "Learning humanoid locomotion with perceptive internal model," in *International Conference on Robotics and Automation (ICRA)*, 2025.
[26] Y. Li, Y. Zhang, W. Xiao, C. Pan, H. Weng, G. He, T. He, and G. Shi, "Hold my beer: Learning gentle humanoid locomotion and end-effector stabilization control," in *Conference on Robot Learning (CoRL)*, 2025.
[27] H. Wang, Z. Wang, J. Ren, Q. Ben, T. Huang, W. Zhang, and J. Pang, "Beamdojo: Learning agile humanoid locomotion on sparse footholds," in *Robotics Science and Systems (RSS)*, 2025.
[28] J. He, C. Zhang, F. Jenelten, R. Grandia, M. Bächer, and M. Hutter, "Attention-based map encoding for learning generalized legged locomotion," *Science Robotics*, 2025.
[29] X. Gu, Y.-J. Wang, X. Zhu, C. Shi, Y. Guo, Y. Liu, and J. Chen, "Advancing humanoid locomotion: Mastering challenging terrains with denoising world model learning," in *Robotics Science and Systems (RSS)*, 2024.
[30] Z. Fu, Q. Zhao, Q. Wu, G. Wetzstein, and C. Finn, "Humanplus: Humanoid shadowing and imitation from humans," in *Conference on Robot Learning (CoRL)*, 2024.
[31] T. Lin, K. Sachdev, L. Fan, J. Malik, and Y. Zhu, "Sim-to-real reinforcement learning for vision-based dexterous manipulation on humanoids," in *Conference on Robot Learning (CoRL)*, 2025.
[32] C. Lu, X. Cheng, J. Li, S. Yang, M. Ji, C. Yuan, G. Yang, S. Yi, and X. Wang, "Mobile-television: Predictive motion priors for humanoid whole-body control," in *International Conference on Robotics and Automation (ICRA)*, 2025.
[33] J. Li, X. Cheng, T. Huang, S. Yang, R.-Z. Qiu, and X. Wang, "Amo: Adaptive motion optimization for hyper-dexterous humanoid whole-body control," in *Robotics Science and Systems (RSS)*, 2025.
[34] Y. Zhang, Y. Yuan, P. Gurunath, T. He, S. Omidshafiei, A. akbar Aghamohammadi, M. Vazquez-Chanlatte, L. Pedersen, and G. Shi, "Falcon: Learning force-adaptive humanoid loco-manipulation," in *Conference on Robot Learning (CoRL)*, 2025.
[35] Q. Ben, F. Jia, J. Zeng, J. Dong, D. Lin, and J. Pang, "Homie: Humanoid loco-manipulation with isomorphic exoskeleton cockpit," in *Robotics Science and Systems (RSS)*, 2025.
[36] Y. Ze, Z. Chen, J. P. Ara'ujo, Z. ang Cao, X. B. Peng, J. Wu, and C. K. Liu, "Twist: Teleoperated whole-body imitation system," in *Conference on Robot Learning (CoRL)*, 2025.
[37] F. Liu, Z. Gu, Y. Cai, Z. Zhou, S. Zhao, H. Jung, S. Ha, Y. Chen, D. Xu, and Y. Zhao, "Opt2skill: Imitating dynamically-feasible whole-body trajectories for versatile humanoid loco-manipulation," *arXiv preprint arXiv:2409.20514*, 2024.
[38] M. Ji, X. Peng, F. Liu, J. Li, G. Yang, X. Cheng, and X. Wang, "Exbody2: Advanced expressive humanoid whole-body control," *ArXiv*, vol. abs/2412.13196, 2024.
[39] M. Gleicher, "Motion editing with spacetime constraints," in *Proceedings of the 24th annual conference on Computer graphics and interactive techniques (SIGGRAPH)*. ACM, 1997, pp. 139–148.
[40] J. Lee and S. yong Shin, "A hierarchical approach to interactive motion editing for human-like figures," in *SIGGRAPH*, 1999.
[41] Z. Popović and A. Witkin, "Physically based motion transformation," in *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, 1999, pp. 11–20.
[42] F. G. Harvey, M. Yurick, D. Nowrouzezahrai, and C. J. Pal, "Robust motion in-betweening," *ACM Transactions on Graphics (TOG)*, 2020.
[43] G. Delmas, P. Weinzaepfel, F. Moreno-Noguer, and G. Rogez, "Posefix: Correcting 3d human poses with natural language," in *International Conference on Computer Vision (ICCV)*, 2023.
[44] J. Qin, Y. Zheng, and K. Zhou, "Motion in-betweening via two-stage transformers," *ACM Transactions on Graphics (TOG)*, 2022.
[45] C. K. Liu, A. Hertzmann, and Z. Popović, "Learning physics-based motion style with nonlinear inverse optimization," *ACM Transactions on Graphics (TOG)*, vol. 24, no. 3, pp. 1071–1081, 2005.
[46] Y. Yuan, J. Song, U. Iqbal, A. Vahdat, and J. Kautz, "Physdiff: Physics-guided human motion diffusion model," *International Conference on Computer Vision (ICCV)*, 2022.
[47] F. Zargarbashi, J. Cheng, D. Kang, R. Sumner, and S. Coros, "Robotkeyframing: Learning locomotion with high-level objectives via mixture of dense and sparse rewards," in *Conference on Robot Learning (CoRL)*, 2024.
[48] S. Lee, S. Lee, Y. Lee, and J. Lee, "Learning a family of motor skills from a single motion clip," *ACM Transactions on Graphics (TOG)*, vol. 40, no. 4, pp. 1–13, 2021.
[49] X. Huang, Y. Chi, R. Wang, Z. Li, X. B. Peng, S. Shao, B. Nikolic, and K. Sreenath, "Diffuseloco: Real-time legged locomotion control with diffusion from offline datasets," in *Conference on Robot Learning (CoRL)*, 2025.
[50] X. Huang, T. E. Truong, Y. Zhang, F. Yu, J.-P. Sleiman, J. Hodgins, K. Sreenath, and F. Farshidian, "Diffuse-cloc: Guided diffusion for physics-based character look-ahead control," in *SIGGRAPH*, 2025.
[51] M. L. Puterman, *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons, 2014.
[52] M. Gleicher, "Motion path editing," in *SIGGRAPH*, 2001.
[53] S. Mysore, G. Cheng, Y. Zhao, K. Saenko, and M. Wu, "Multi-critic actor learning: Teaching rl policies to act with style," in *International Conference on Learning Representations (ICLR)*, 2022.
[54] T. Huang, J. Ren, H. Wang, Z. Wang, Q. Ben, M. Wen, X. Chen, J. Li, and J. Pang, "Learning humanoid standing-up control across diverse postures," in *Robotics Science and Systems (RSS)*, 2025.
[55] W. Xu and F. Zhang, "Fast-lio: A fast, robust lidar-inertial odometry package by tightly-coupled iterated kalman filter," *Robotics and Automation Letters (RAL)*, 2021.
[56] V. Makoviychuk, L. Wawrzyniak, Y. Guo, M. Lu, K. Storey, M. Macklin, D. Hoeller, N. Rudin, A. Allshire, A. Handa *et al.*, "Isaac gym: High performance gpu-based physics simulation for robot learning,"

arXiv preprint arXiv:2108.10470, 2021.
[57] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," *arXiv preprint arXiv:1707.06347*, 2017.
[58] T. Kobayashi, "L2c2: Locally lipschitz continuous constraint towards stable and smooth reinforcement learning," in *International Conference on Intelligent Robots and Systems (IROS)*, 2022.

APPENDIX

A. Task Details

Motion editing. For each task, we selected two types of keyframes: (1) Keyframes with semantics. For example, the frame before and after jumping. (2) Keyframes without semantics. These keyframes are uniformly sampled from a clip of motion to ensure smoothness and tracking accuracy. The global poses of the former keyframes are further edited in one dimension. For motion details, please refer to the code.

Observations are composed of the following parts:

- Base angular velocity, 3 dimensions.
- Projected gravity $\dot{\boldsymbol{\omega}}_k$, 3 dimension.
- Joint positions $\boldsymbol{\omega}_k$, 27 dimension.
- Joint velocities $\dot{\boldsymbol{\theta}}$, 27 dimension.
- Previous actions $\boldsymbol{a}_{k-1}$, 27 dimension.
- Current phase ϕ, 1 dimension.
- Task variable ψ, 1 dimension.
- Lidar odometry, 3 dimensions. The relative position based on the initial lidar coordinate.
- Base linear velocity, 3 dimension, critic-only.
- Reference joint positions, 27 dimensions, critic-only.

B. Baseline implementations.

- *AMP-style* and *AMP-Mimic* use 5-step DoF positions as the discriminator observation. The weight of the style reward is 0.1, while the weight of the task reward is 0.9. The style reward is classified into the sparse reward group.
- *DeepMimic*-based baselines use the same rewards as ours. Differently, *DeepMimic-NoAdapt* uses the original motion and therefore has no adaptation ability. *DeepMimic-Adapt* uses the rule-based augmented motions with a fixed phase interval. *DeepMimic-Adapt-*$\Delta\phi^{\text{ada}}$ additionally trains a flexible phase interval.
- *UniTracker* follows its official implementation without modification. It uses the rule-based augmented motions during inference. *UniTracker-Adapt* additionally introduces a fine-tuning stage on the augmented motions. Since training UniTracker in terrains is not straightforward, we only test UniTracker in five tasks that do not depend on terrains.

The method of rule-based motion augmentation mainly follow [52]. Specifically, we first determine the keyframes in the reference motion, which are the same as those in the motion processing stage. Then, we linearly adjust the global position of in-between frames without adjusting their local pose. Each task variable ψ results in a corresponding motion, which is collected into $\mathcal{D}_{\text{ref}}^{\text{rule}}$.

C. Hardware Setup

We use the Unitree 29-DoF G1 humanoid robot (6 per leg, 7 per arm, and 1 in the waist). Waist roll and pitch joints are locked for stability and safety. We used the MID360 lidar on board with FastLIO [55] to obtain odometry. The PD controller follows the designs in BeyondMimic [3], which significantly improves the hardware deployment in terms of stability, smoothness, and safety. Due to the drift issue of the odometry and the randomly initialized robot's pose, we observe randomness in policy performance. Therefore, we conduct multiple tests of each policy in each task for statistical significance.

D. Training Details

Domain randomization. We apply domain randomization during training in the simulation. The randomization terms largely follow HoST [54], which are sufficient for overcoming the sim-to-real gap. Table V below lists each term:

TABLE V: Domain randomization for adaptive motion tracking.

Term	Value
Trunk Mass	$\mathcal{U}(-2, 5)$kg
Base CoM offset	$\mathcal{U}(-0.1, 0.1)$m (XYZ)
Link mass	$\mathcal{U}(0.9, 1.1)\times$ default kg
Fiction	$\mathcal{U}(0.1, 1.1)$
Restitution	$\mathcal{U}(0, 0.1)$
P Gain	$\mathcal{U}(0.85, 1.15)$
D Gain	$\mathcal{U}(0.85, 1.15)$
Motor Strength	$\mathcal{U}(0.9, 1.1)$
Control delay	$\mathcal{U}(0, 100)$ms

Termination conditions. We follow the termination conditions in ASAP [7], including termination curriculum and termination terms. Beyond that, given the nature of keyframe tracking, we observe that very strict termination conditions will impede the exploration of highly dynamic parts of the motion (*e.g.*, jumping). We therefore relax the conditions after 6000 training epochs to ensure successful learning of the whole motion.

Reward scales. We observe that the identical reward scales of each keyframe in the sparse reward group will lead to conservative policies, *e.g.*, the policy learns to avoid jumping to collect more local rewards. To overcome this problem, we assign some keyframes with semantics with a higher reward scale to encourage exploration of more dynamic behaviors.

PPO hyperparameters are listed in Table VI:

TABLE VI: Hyperparameters of PPO.

Hyperparameter	**Value**
Number of envs	4096
Number of steps per iteration	75
Number of learning epochs	5
Clip range	0.2
Entropy coefficient	0.01
GAE balancing factor λ	0.95
Desired KL-divergence	0.01
Actor and double critic MLP	[512, 256, 128]
Discount factor	Sparse 1; Dense 0.99
Initial learning rate	$1e^{-3}$