Agility Meets Stability: Versatile Humanoid Control with Heterogeneous Data

Yixuan Pan[1*] Ruoyi Qiao[4*] Li Chen[1] Kashyap Chitta[2] Liang Pan[1] Haoguang Mai[1]
Qingwen Bu[1] Hao Zhao[3] Cunyuan Zheng[4] Ping Luo[1] Hongyang Li[1]
[1]The University of Hong Kong [2]NVIDIA [3]Tsinghua University [4]Individual Contributor
*Equal contribution

https://opendrivelab.com/AMS/

Fig. 1: Introducing **AMS** (Agility Meets Stability), one single policy that performs diverse motions with stability and agility simultaneously on a humanoid robot. The robot can execute challenging balance motions such as (a) `Ip Man's Squat`, a Kung Fu-style single-leg squat, unseen during training (zero-shot); (b) `single-leg balance stances` which humans find hard to perform; (c) balanced stretching; as well as expressive motions and high-mobility movements with precise control, such as (d) dancing and (e) running. More examples are provided in the appended video.

***Abstract*— Humanoid robots are envisioned to perform a wide range of tasks in human-centered environments, requiring controllers that combine agility with robust balance. Recent advances in locomotion and whole-body tracking have enabled impressive progress in either agile dynamic skills or stability-critical behaviors, but existing methods remain specialized, focusing on one capability while compromising the other. In this work, we introduce AMS (Agility Meets Stability), the first framework that unifies both dynamic motion tracking and extreme balance maintenance in a single policy. Our key**

insight is to leverage heterogeneous data sources: human motion capture datasets that provide rich, agile behaviors, and physically constrained synthetic balance motions that capture stability configurations. To reconcile the divergent optimization goals of agility and stability, we design a hybrid reward scheme that applies general tracking objectives across all data while injecting balance-specific priors only into synthetic motions. Further, an adaptive learning strategy with performance-driven sampling and motion-specific reward shaping enables efficient training across diverse motion distributions. We validate AMS extensively in simulation and on a real Unitree G1 humanoid. Experiments demonstrate that a single policy can execute agile skills such as dancing and running, while also performing zero-shot extreme balance motions like Ip Man's Squat, highlighting AMS as a versatile control paradigm for future humanoid applications.

I. Introduction

Humanoid robots hold great promise for performing diverse tasks in human-centric environments, from household assistance to industrial applications [1]. Realizing this vision requires robots to emulate the remarkable capabilities that humans naturally master, *i.e.*, versatile, coordinated whole-body skills that seamlessly blend dynamic motion with precise balance. Recent progress in locomotion and whole-body tracking has enabled robust outdoor walking [2], [3], [4], multi-modal control [5], [6], [7], sequential movements [8], and challenging agile behaviors [9], [10], [11]. Despite these advances, humanoid robots still struggle to integrate dynamic motion with precise balance in a *unified* manner. Humans, by contrast, naturally demonstrate such capabilities, for example, by maintaining a stable single-leg stance while reaching for an object using a free limb as temporary support, or performing precise placement after dynamic walking. Endowing humanoids with such integrated versatility, however, remains a fundamental challenge. Current work typically adopts reinforcement learning (RL) to train whole-body tracking (WBT) policies with human motions as references to accumulate rewards. They focus on *single-sequence* policy training, either fitting dynamic movements such as ASAP [10] or balance motions like HuB [12], rather than achieving both capabilities in a unified and generalized manner.

The underlying reasons for this situation can be divided into two aspects: data limitation and divergent optimization objectives. Existing approaches [13], [14], [15] predominantly rely on human motion capture (MoCap) data for training. While such datasets provide rich dynamic behaviors, they suffer from long-tailed distributions in which extreme balance scenarios are underrepresented. Further, they inherently restrict the robot to motions that humans can perform, constraining the exploitation of the robot's unique mechanical capabilities. In addition, dynamic and balanced motions exhibit distinct distributions and thus require separate optimization objectives. In an RL-based paradigm, reward functions designed to guide one motion type can inadvertently hinder the learning of the other, leading to conflicts when combined within a unified learning framework. For instance, restricting the center of mass to remain above the support foot provides precise guidance for balance tasks but is overly restrictive for dynamic motions that rely on natural momentum transfer and coordinated whole-body movements. A desirable solution would allow a single policy to learn both dynamic agility and balance robustness without compromising either objective.

To address these challenges, we propose **AMS** (Agility Meets Stability), a unified framework that trains a single policy capable of both dynamic motion tracking and extreme balance maintenance through adaptive learning on heterogeneous data.

Our approach addresses the data limitations by generating constrained synthetic balance motions that complement existing human MoCap data [16]. Unlike MoCap data, which suffers from sensor noise and kinematic retargeting errors, these synthetic motions are sampled from the humanoid motion space directly while ensuring physical plausibility. By integrating these heterogeneous data sources, our approach alleviates the long-tailed distribution problem and broadens the range of physically achievable behaviors, complementing and going beyond what traditional human motion datasets can provide.

To resolve conflicting optimization objectives, we employ a hybrid reward scheme that combines two complementary components for policy training. General rewards encourage robust motion tracking across all data, while balance-specific rewards are applied exclusively to the controllable synthetic data, providing precise guidance for stability without inducing conflicts with dynamic tracking objectives.

We further introduce an adaptive learning strategy with two key components for effective learning from heterogeneous data. Adaptive sampling prioritizes challenging motions by automatically adjusting sampling probability for effective hard sample mining. In the meantime, adaptive reward shaping maintains motion-specific error tolerances based on individual performance rather than treating all motions uniformly.

We evaluate AMS through extensive experiments on a Unitree G1 humanoid robot. As demonstrated in Fig. 1, our unified policy effectively executes both dynamic motions such as dancing and challenging balance motions like `Ip Man's Squat` in a zero-shot manner. The versatile framework also enables real-time teleoperation for various motions, highlighting its potential as a foundational control model for autonomous humanoid applications.

To summarize, our contributions are threefold: **(1)** We introduce AMS, the first framework that successfully unifies dynamic motion tracking and extreme balance maintenance in a single policy. **(2)** We develop a learning approach that leverages both human-captured motion data and controllable synthetic balance motions, coupled with hybrid rewards and adaptive learning for effective policy training. **(3)** We showcase that a single policy can execute both dynamic motions and robust balance control on a humanoid in the real world, outperforming baseline methods and enabling interactive teleoperation.

II. Related Work

A. Learning-based Humanoid Whole-body Tracking

Learning-based whole-body tracking has enabled humanoid robots to achieve increasingly versatile behaviors and is

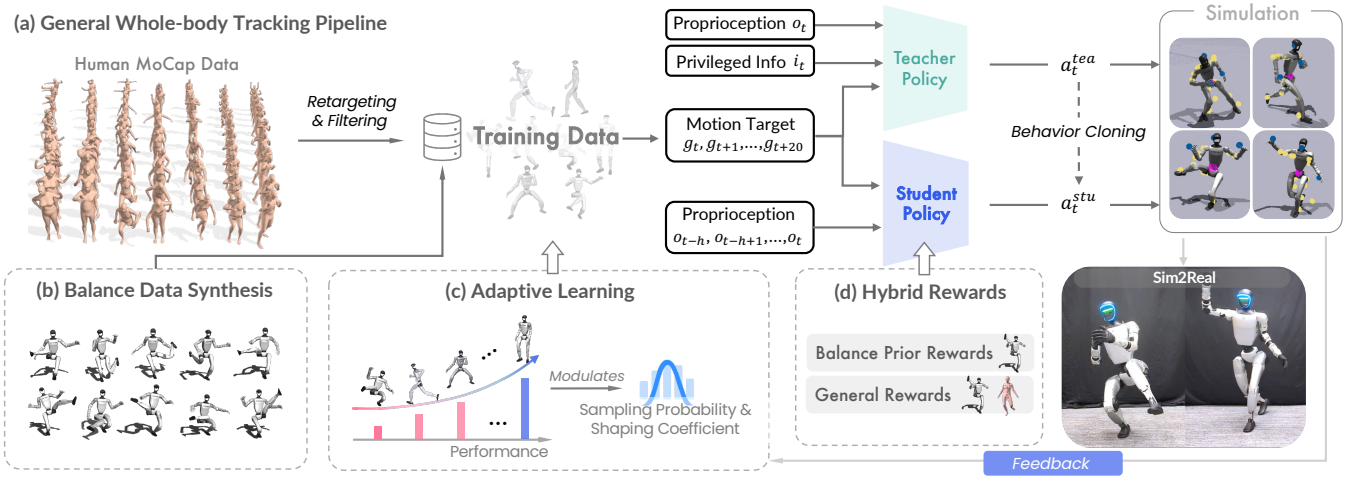

Fig. 2: **Overview of AMS. (a)** The general whole-body tracking pipeline retargets human MoCap data to reference motions and adopts a teacher-student-based strategy for reinforcement learning (Sec. III-A). To address data limitations and conflicting optimization objectives, AMS introduces three key components as follows. **(b)** Synthetic balance data is generated to complement human MoCap data and address data limitations (Sec. III-B). **(c)** Adaptive learning is employed with adaptive sampling and reward shaping based on individual motion performance (Sec. III-D). **(d)** Hybrid rewards are designed with general rewards for all motions and balance prior rewards exclusively for synthetic motions (Sec. III-C).

promised as a method to collect humanoid data for embodied policy training. Building upon DeepMimic [17], recent work has demonstrated agile controllers capable of expressive skills such as dancing [15], [18], martial arts [19], and general athletic maneuvers [13], [20], [21]. Other approaches scale to large motion libraries, where universal policies trained on MoCap datasets [22], [14], [13] provide broad coverage of human-like movements.

In parallel, another research thrust targets robust balance control, focusing on quasi-static stability rather than agility. HuB [12], for example, introduces motion filtering and task-specific rewards to train policies for extreme balancing poses that are typically absent from human datasets. While effective for maintaining stability, such methods often constrain dynamic motions that inherently require momentum and transient instability.

These two directions, agility and stability, have so far been pursued largely in isolation. AMS aims to bridge this gap by training a single policy on heterogeneous motion data that integrates both dynamic and balance-critical examples, thereby achieving high-fidelity tracking and robust stability within a unified framework.

B. Motion Targets for Policy Learning

The capabilities of a learned controller are fundamentally shaped by the reference motions it is trained to imitate. Most works rely on human demonstrations, either from large-scale motion capture datasets [16], [23] or from monocular video through pose estimation [24], [25], which provide diverse and natural kinematics for general-purpose policies [13], [20], [26]. However, such data inevitably reflects the bias of human movement and exhibits a long-tail distribution that undersamples balance-critical or robot-specific behaviors [27]. Complementary to this, optimization- and sampling-based approaches generate feasible trajectories directly in the robot's configuration space, thereby expanding policy coverage toward versatile locomotion [28] and loco-manipulation [29], [30]. Recent advances further leverage generative models conditioned on high-level commands, enabling motion synthesis from language or vision-language prompts [31], [32]. Motivated by these prior efforts, our work adopts a heterogeneous data strategy. We combine the natural movement of MoCap-derived motions with a controllable generator that produces physically verified, balance-critical behaviors, providing broad supervision for training a single policy that can handle both agility and stability.

III. Methodology

A. Problem Setup

We formulate humanoid whole-body tracking as a goal-conditioned reinforcement learning (RL) task, where a policy π is optimized to track a reference motion sequence in real time. The pipeline is illustrated in Fig. 2(a). Human MoCap data is initially retargeted to the humanoid motion space and erroneous or infeasible motions for humanoids are filtered out [14]. At timestep t, the system state s_t contains the agent's proprioceptive observations o_t, while g_t denotes the target motion state from reference motions. The reward is defined as $r_t = R(s_t, a_t, g_t)$, encouraging alignment between executed and reference motions. The action $a_t \in \mathbb{R}^{23}$ specifies desired joint positions, applied through a PD controller. We train a teacher policy with privileged information i_t using Proximal Policy Optimization (PPO) [33], and distill it into a student policy that depends only on deployable sensory inputs with supervision from the teacher policy [14], [34].

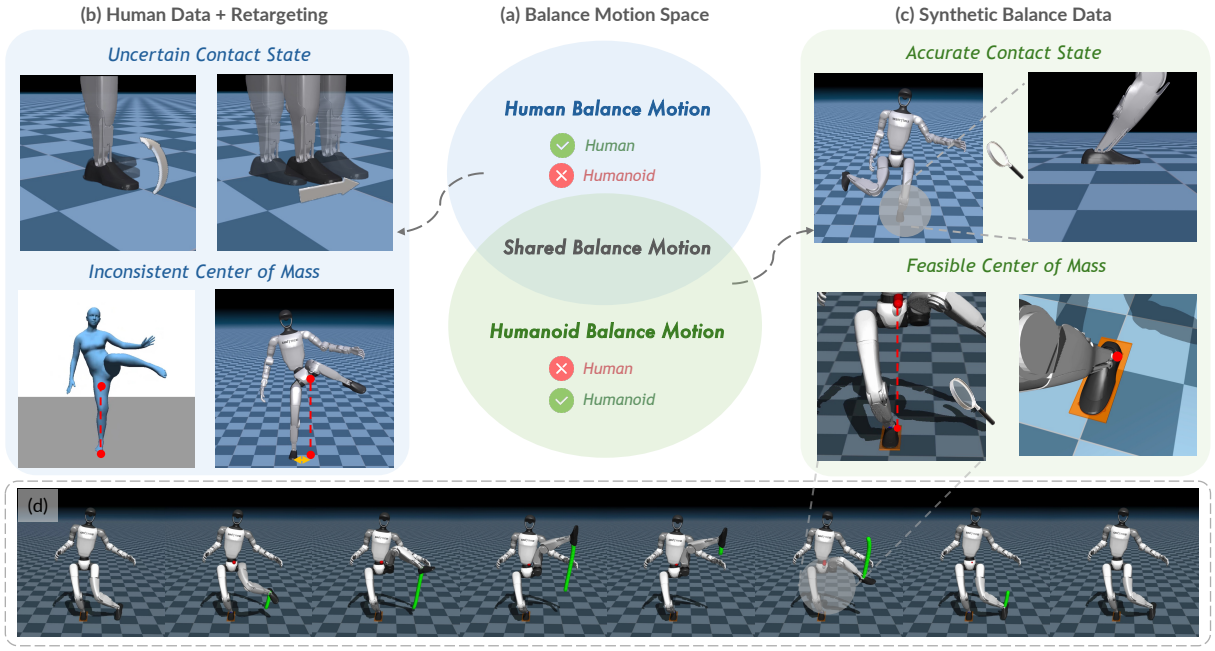

Fig. 3: **Motion space analysis of human data and generated balance data. (a)** Humans and humanoids feature distinctive balance motion spaces, leading to limited reference motions for training whole-body balancing skills. **(b)** Sensor noise and kinematic retargeting errors greatly affect the reference motion quality from human MoCap data. **(c)** Constrained synthetic balance data guarantees physical realism, such as the foot contact state and center of mass. **(d)** Example of a generated synthetic balance motion, with the swinging foot trajectory shown in green.

B. Synthetic Balance Motion Generation

Analysis of Balance Motion References. Due to kinematic and morphological differences between humans and humanoid robots, their balance motion spaces only partially overlap, as illustrated in Fig. 3(a). Prior works [10], [13], [12], [14] predominantly rely on human motion data, which inherently constrains the policy's capabilities to this shared space. However, humanoid robots possess unique mechanical features—different joint limits, actuator capabilities, and mass distributions—that enable balance configurations different from human physical constraints. Additionally, as shown in Fig. 3(b), human data combined with retargeting introduces noise from sensor measurements and the retargeting process, further limiting the quality of training data. To address these limitations, we propose generating synthetic balance data by directly sampling from the humanoid balance motion space, as shown in Fig. 3(c) and (d), complementing human-centric datasets with a broader range of feasible behaviors.

Motion Generation. To enhance the training dataset with physically plausible and diverse whole-body motion sequences, we propose a motion generation framework that synthesizes balanced whole-body trajectories for single-support maneuvers, as shown in Fig. 2(b). The method synthesizes trajectories that transition the ungrounded swinging foot to a target pose while maintaining the center of mass (CoM) within a valid support region, ensuring kinematic feasibility and smoothness.

Given a robot model, a designated support foot, and a time horizon N, we first sample a target pose for the swinging foot, a target pelvis height, and an initial joint configuration biased toward natural lower-limb postures with randomized upper-body joints. These samples induce diversity in end-effector goals and whole-body configurations.

Algorithm 1 Controllable Balance Motion Generation

Input: Robot model $\mathscr{R}$, support foot index s, target foot pose $\mathbf{T}_f$, pelvis height h_p, horizon N, cost weights λ
Output: Motion sequence $\mathscr{M} = \{(\mathbf{X}_t, \mathbf{q}_t)\}_{t=0}^{N-1}$

Reference construction:
$\mathbf{T}_s(t) \leftarrow \text{constant}(\mathbf{T}_s^0)$ {Support foot trajectory}
$\mathbf{T}_{\bar{s}}(t) \leftarrow \text{interp}(\mathbf{T}_{\bar{s}}^0, \mathbf{T}_f, t/N)$ {Swinging foot trajectory}
$\mathbf{T}_P(t) \leftarrow \text{interp}(\mathbf{T}_P^0, \mathbf{T}_P^{\text{target}}, t/N)$ {Pelvis trajectory}
Stage-1 optimization:
Solve $\min J_1(\mathbf{X}, \mathbf{q})$ where:
$J_1 = \lambda_{\text{track}} \underbrace{\|\mathbf{T} - \mathbf{T}_{\text{ref}}\|_{\mathbf{W}}^2}_{\text{tracking}} + \lambda_{\text{lim}} \underbrace{\text{clip}(\mathbf{q}, \mathscr{Q}_{\text{lim}})^2}_{\text{limits}}$
$+\lambda_{\text{rest}} \underbrace{\|\mathbf{q} - \mathbf{q}^{\text{init}}\|^2}_{\text{rest}} + \lambda_{\text{smooth}} \underbrace{\text{smooth}(\mathbf{X}, \mathbf{q})}_{\text{smoothness}}$
Stage-2 optimization:
Solve $\min J_2(\mathbf{X}, \mathbf{q})$ where:
$J_2 = J_1 + \lambda_{\text{bal}} \sum_{t=0}^{N-1} \max(0, d_t - \varepsilon)$
$d_t = \| \max(\mathbf{0}, |\mathbf{p}_t - \mathbf{c}_t| - \mathbf{s}) |_2$
$\mathbf{p}_t = \Pi_{xy} \text{CoM}(\mathbf{X}_t, \mathbf{q}_t; \mathscr{R})$, $\mathbf{c}_t = \Pi_{xy} \text{Trans}(\mathbf{T}_s(t))$
Validation:
return $\mathscr{M}$ if $\max_t d_t \leq \varepsilon$ **else fail**

We then construct reference trajectories for three key links, including the support foot, swinging foot, and pelvis, using SE(3) interpolation. To compute the motion, we employ a two-stage batch trajectory optimization as outlined in Algorithm 1.

The first stage minimizes a composite cost J_1 that includes pose tracking, soft joint limits, rest-pose regularization, and temporal smoothness. This yields a kinematically consistent and smooth trajectory. In the second stage, we augment the cost with a balance-enforcing term:

$$J_2 = J_1 + \lambda_{bal} \sum_{t=0}^{N-1} \max(\mathbf{0}, \|\mathbf{p}_t - \mathbf{c}_t - \mathbf{s}\|_2 - \varepsilon), \tag{1}$$

where $\mathbf{p}_t$ and $\mathbf{c}_t$ are the 2D projections of the CoM and support foot center, $\mathbf{s} = (s_x, s_y)$ defines the support rectangle, and ε is a small tolerance. This penalty encourages the CoM to remain within a valid support area. The optimization is solved using Levenberg–Marquardt solver [35], [36], [37], [38]. Only trajectories that satisfy $\max_t d_t <= \varepsilon$ are accepted, ensuring physical feasibility.

The two-stage approach hierarchically separates kinematic feasibility from balance constraints, where the first stage establishes a robust and smooth trajectory, while the second stage safely refines it for balance, enabling stable convergence.

C. Hybrid Rewards

A central challenge in training a single policy for both dynamic motion tracking and balance-critical behaviors is the conflicting objectives: rewards emphasizing balance could restrict dynamic motions, while rewards for agility may compromise stability. To address this, we introduce a hybrid reward scheme that distinguishes between general motion tracking and balance-specific guidance based on the motion source, as shown in Fig. 2(d).

For human motion capture data [16], [39], we rely on general motion-tracking terms solely, such as joint positions, velocities, and root orientation, which encourage natural, human-like movements while maintaining coarse stability. In contrast, for synthetic balance-critical motions, we augment the supervision with balance-specific priors, including center-of-mass alignment and foot contact consistency [12]. These priors provide physically grounded guidance, ensuring a feasible balance without overly constraining the agility captured from human MoCap data.

By selectively applying balance priors rewards only to synthetic data, the hybrid reward design enables the policy to capture agile behaviors from human motions while maintaining reliable stability in challenging postures.

D. Adaptive Learning

To further address both data limitations and conflicting objectives, we introduce an adaptive learning strategy comprising two key components, *i.e.*, adaptive sampling and adaptive reward shaping, as shown in Fig. 2(c).

Adaptive Sampling. We propose a performance-driven adaptive sampling strategy that dynamically adjusts motion sequence sampling probabilities based on tracking performance assessment. Unlike uniform sampling that treats all motion data equally, our approach implements a multi-dimensional performance evaluation mechanism to prioritize poorly-tracked samples while reducing emphasis on well-tracked motions.

The adaptive sampling strategy evaluates tracking performance across three key dimensions: (1) motion execution failure, (2) mean per-joint position error (MPJPE), and (3) maximum joint position error. For each motion sequence i, we maintain a sampling probability p_i that is dynamically updated based on periodic evaluation results.

Let $\mathscr{F}$ denote the set of failed motions during evaluation, and e^i_{mean}, e^i_{max} represent the mean and maximum joint position errors for motion i, respectively. For successful motions, we define performance thresholds using percentiles of the error distribution: $\tau_{poor} = P_{75}(e)$ and $\tau_{good} = P_{25}(e)$, where P_k denotes the k-th percentile.

The probability update mechanism operates as follows:

$$p_i^{t+1} = \begin{cases} p_i^t \cdot \gamma_{fail}, & \text{if } i \in \mathscr{F}, \\ p_i^t \cdot g_i, & \text{otherwise}, \end{cases} \tag{2}$$

where p_i^t is the current sampling probability for motion i at training iteration t, γ_{fail} is the failure boost factor, and the adjustment factor g_i is computed as:

$$g_i = 1 + w_{mean}(f_{mean}(e^i_{mean}) - 1) + w_{max}(f_{max}(e^i_{max}) - 1), \tag{3}$$

where w_{mean} and w_{max} are weighting coefficients that control the relative importance of mean and maximum error adjustments. The error-specific adjustment functions $f_{mean}(\cdot)$ and $f_{max}(\cdot)$ are defined identically as:

$$f(e) = \begin{cases} \beta_{min} + (\beta_{max} - \beta_{min}) \cdot r_{poor}, & \text{if } e > \tau_{poor}, \\ \alpha_{min} + (\alpha_{max} - \alpha_{min}) \cdot (1 - r_{good}), & \text{if } e < \tau_{good}, \\ 1, & \text{otherwise}, \end{cases} \tag{4}$$

where $\beta_{min}, \beta_{max} > 1$ are the minimum and maximum boost factors for poor-performing motions, $\alpha_{min}, \alpha_{max} < 1$ are the minimum and maximum reduction factors for well-performing motions. The normalized ratios are computed as:

$$r_{poor} = \frac{e - \tau_{poor}}{e_{max} - \tau_{poor}}, \quad r_{good} = \frac{\tau_{good} - e}{\tau_{good} - e_{min}}, \tag{5}$$

with e_{max} and e_{min} being the maximum and minimum errors observed in the current evaluation.

To ensure exploration and prevent any motion from being completely ignored, we enforce a minimum sampling probability constraint. After updating all probabilities, they are first normalized, then clamped to a minimum threshold:

$$p_i^{final} = \max\left(\frac{p_i^{t+1}}{\sum_{j=1}^{N} p_j^{t+1}}, p_{min} \right), \tag{6}$$

where $p_{min} = \lambda \cdot \frac{1}{N}$ with λ being the minimum probability factor and N the total number of motions. The probabilities are then re-normalized to ensure they sum to unity.

This adaptive sampling mechanism enables AMS to automatically focus on poorly-tracked motion patterns by continuously adjusting the training data distribution based

on tracking performance, thereby improving both sample efficiency and generalization performance.

Adaptive Reward Shaping. Existing universal WBT methods [20], [14], [40], [41] typically employ uniform and fixed shaping coefficients to modulate reward functions for all motions. Typically, the reward is defined as:

$$r = \exp\left(-\frac{err}{\sigma}\right), \tag{7}$$

where *err* represents the tracking error for a given motion and σ serves as the shaping coefficient controlling error tolerance. However, this uniform treatment presents two challenges: (1) fixed tolerance does not adapt to improving tracking performance; (2) identical parameters create conflicting objectives between dynamic and balance motions that require different shaping strategies.

Inspired by PBHC [19], we extend their adaptive strategy from single-motion tracking to general multi-motion tracking scenarios. Specifically, we maintain motion-specific σ parameter sets, with separate adjustments for different body parts. For stable and responsive adaptation, we employ Exponential Moving Average (EMA) to update these parameters:

$$\sigma_{\text{new}} = (1-\alpha)\cdot\sigma_{\text{current}} + \alpha\cdot \text{err}_{\text{current}}, \tag{8}$$

where α is the update rate controlling adaptation responsiveness, and $\text{err}_{\text{current}}$ represents the current tracking error.

This motion-specific adaptive reward shaping mechanism enables AMS to simultaneously adapt to training progress and motion diversity, significantly improving learning efficiency in general motion tracking scenarios.

IV. Experiment

Our experiments aim to answer the following questions:

- **Q1:** How well does AMS perform on both dynamic and balance motions compared to existing approaches?
- **Q2:** How do the synthetic data and training strategies contribute to the overall performance?
- **Q3:** Can AMS generalize to unseen scenarios and real-world deployment?

A. Experimental Setup

We evaluate AMS in both simulation and real-robot experiments. In simulation, we use IsaacGym [42] as our physics simulator. Our training dataset comprises a filtered subset of the AMASS [16] and LAFAN1 [39] datasets, containing over 8,000 motion sequences and 10,000 synthetic balance motion sequences generated by our methods. For real-world experiments, we deploy our policy on Unitree G1 [43], a humanoid robot with 23 DoFs and a height of 1.3 meters, weighing about 35kg.

Metrics. We evaluate the motion tracking performance using five metrics [12], [14]. (1) **Success rate** (Succ., %). Imitation fails if the average deviation from reference exceeds 0.5m at any point, measuring whether the robot can maintain tracking without losing balance. (2) **Global MPJPE** ($E_{\text{g-mpjpe}}$, *mm*) measures global position tracking accuracy. (3) **Root-relative MPJPE** (E_{mpjpe}, *mm*) evaluates local joint position tracking performance. To assess policy stability and fidelity on balance motions, we additionally employ (4) **Contact mismatch** (Cont., %), measuring the percentage of frames where foot contact states differ from the reference motion; and (5) **Slippage** (Slip., m/s), which quantifies the ground-relative velocity of the support foot, where higher values indicate unstable foot contact.

B. Comparison with Existing Methods

To address **Q1**, we compare our method against two representative baselines:

- **OmniH2O [14]** is a general humanoid whole-body motion tracking framework that employs a teacher-student learning paradigm. We adapt OmniH2O to the G1 robot and optimize its curriculum parameters for our experimental setup.
- **HuB [12].** Building upon the OmniH2O framework, we re-implement HuB by replacing the reward function with HuB's stability-focused reward design, which emphasizes balance motions and contact-aware tracking.

For fair comparison, all baselines are trained from scratch with consistent domain randomization. We evaluate the teacher policies in simulation experiments, and the student policies are derived through direct imitation learning from their respective teachers. Table I(a) shows that the proposed method significantly outperforms both OmniH2O and HuB. The approach achieves improvements in tracking performance ($E_{\text{g-mpjpe}}$ and E_{mpjpe}), while simultaneously maintaining high stability (Cont. and Slip.).

C. Ablation Study

To address **Q2**, we conduct comprehensive ablation studies on each key component of AMS.

Ablation on Synthetic Balance Data.

As shown in Table I(b), the variant *w/o Synthetic Balance Data* demonstrates that training exclusively on MoCap data results in poor performance on balance motions. Incorporating synthetic balance data maintains comparable performance on MoCap data while significantly improving tracking performance and stability on balance-critical motions. To further answer **Q3** regarding generalization capability, we collect 1000 unseen motions as out-of-distribution (OOD) test data, including self-recorded random motions and single-leg motions generated by our proposed method. As shown in Table II, adding synthetic balance data to the training set effectively improves OOD performance.

Ablation on Hybrid Rewards. As shown in Table I(c), removing balance prior rewards and using only general rewards (*w/ General Rewards Only*) leads to degraded performance on balance motions, with higher tracking errors and more frequent contact mismatches. Conversely, applying balance prior rewards uniformly across all motions (*w/ All Rewards for All Data*) shows certain improvement on balance tasks, but creates conflicting objectives that harm overall policy performance. As evidenced in the table, this approach causes performance degradation on MoCap data, which

TABLE I: **Simulation performance comparison on different datasets and ablation study.** Our method consistently achieves lower tracking errors and higher success rates across both agile motion and challenging balance motions, demonstrating strong generalization and robustness.

Method	MoCap Data (AMASS+LAFAN1)			Synthetic Balance Data					All		
	Tracking Error		Completion	Stability		Tracking Error		Completion	Tracking Error		Completion
	$E_{\text{g-mpjpe}} \downarrow$	$E_{\text{mpjpe}} \downarrow$	Succ. ↑	Cont. ↓	Slip. ↓	$E_{\text{g-mpjpe}} \downarrow$	$E_{\text{mpjpe}} \downarrow$	Succ. ↑	$E_{\text{g-mpjpe}} \downarrow$	$E_{\text{mpjpe}} \downarrow$	Succ. ↑
(a) Main Results											
OmniH2O [14]	68.31	37.23	98.49%	0.24	0.038	72.51	56.88	99.76%	69.84	44.18	98.93%
HuB [12]	158.80	82.13	67.03%	**0.10**	**0.030**	137.44	128.22	99.82%	151.26	98.42	77.23%
AMS (Ours)	**48.60**	**24.48**	**99.69%**	0.12	**0.030**	**64.03**	**37.30**	**99.95%**	**54.06**	**29.02**	**99.78%**
(b) Ablation on Synthetic Balance Data											
AMS w/o Synthetic Balance Data	50.25	**24.10**	99.64%	0.69	0.047	112.20	71.89	94.54%	72.20	40.99	98.09%
AMS (Ours)	**48.60**	24.48	**99.69%**	**0.12**	**0.030**	**64.03**	**37.30**	**99.95%**	**54.06**	**29.02**	**99.78%**
(c) Ablation on Hybrid Rewards											
AMS w/ General Rewards Only	49.70	25.41	**99.72%**	0.39	0.036	65.39	45.98	99.46%	55.31	32.75	99.65%
AMS w/ All Rewards for All Data	54.09	27.30	99.60%	0.31	0.095	71.62	40.56	99.89%	60.32	31.99	99.70%
AMS (Ours)	**48.60**	**24.48**	99.69%	**0.12**	**0.030**	**64.03**	**37.30**	**99.95%**	**54.06**	**29.02**	**99.78%**
(d) Ablation on Adaptive Learning											
AMS w/o Adaptive Learning (AS+ARS)	78.88	27.74	98.21%	**0.09**	**0.029**	87.86	43.21	**99.95%**	82.11	33.22	98.75%
AMS w/o Adaptive Sampling (AS)	52.92	24.60	98.85%	**0.09**	0.030	66.51	39.15	99.69%	57.74	29.74	99.14%
AMS w/o Adaptive Reward Shaping (ARS)	74.45	26.86	99.49%	0.13	0.030	89.03	47.27	99.90%	79.76	34.11	99.61%
AMS (Ours)	**48.60**	**24.48**	**99.69%**	0.12	0.030	**64.03**	**37.30**	**99.95%**	**54.06**	**29.02**	**99.78%**

TABLE II: **Out-of-distribution (OOD) performance comparison.** Our method achieves the lowest tracking errors and highest completion rate, showing better generalization to unseen motions.

Method	Tracking Error		Completion
	$E_{\text{g-mpjpe}} \downarrow$	$E_{\text{mpjpe}} \downarrow$	Succ.↑
AMS w/o Synthetic Balance Data	86.61	46.43	96.0
OmniH2O [14] w/ All Data	76.26	49.57	99.1
AMS (Ours)	**63.48**	**32.06**	**99.7**

contains substantial dynamic motions. Our hybrid reward approach provides strong balance guidance while avoiding these conflicts, delivering improved overall performance by applying balance-specific rewards exclusively to synthetic data while preserving dynamic motion agility through general rewards.

Ablation on Adaptive Learning. We separately evaluate the two main components of our adaptive learning strategy: Adaptive Sampling (AS) and Adaptive Reward Shaping (ARS), as shown in Table I(d). AS adaptively mines hard samples by automatically prioritizing difficult motions, leading to improved success rates and tracking performance on challenging and underrepresented samples. ARS provides targeted reward adjustments for each individual motion, significantly reducing tracking errors. When both components are removed (*w/o Adaptive Learning*), the performance degradation is most pronounced, demonstrating that uniform treatment of all motions fails to address the inherent data diversity and difficulty distribution.

D. Real-World Deployment

To further address **Q3**, we deploy our unified policy on the Unitree G1 humanoid robot, demonstrating execution of a wide range of motions that span both dynamic and balance-critical behaviors. As illustrated in Fig. 1, the robot can execute challenging balancing motions unseen during training, such as `Ip Man's Squat` and single-leg balancing stances, as well as high-mobility movements and expressive motions like running and dancing. To further validate generalizability, we conduct real-time teleoperation with an off-the-shelf human pose estimation model [44], as shown in Fig. 4. The poses' keypoints captured by the RGB camera are scaled to humanoid sizes for tracking. Though not strictly optimized like the complex retargeting process, this simple teleportation still shows robust adaptation to diverse motions.

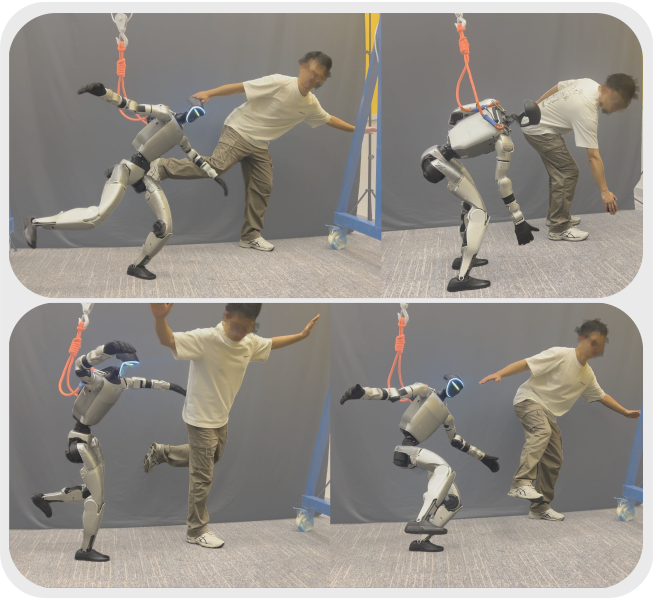

Fig. 4: **RGB camera-based real-time teleoperation.**

V. Conclusion

In this work, we introduce AMS, the first framework that successfully unifies dynamic motion tracking and extreme balance maintenance in a single policy. Through leveraging heterogeneous data sources, hybrid rewards, and adaptive learning, our approach enables effective policy training across diverse motion distributions. Our real-world demonstrations on the Unitree G1 robot showcase that a single policy can execute both dynamic motions and robust balance control, outperforming baseline methods and enabling interactive teleoperation.

Limitations and Future Work. While our approach shows promising results, it lacks precise end-effector control, limiting its applicability to manipulation and contact-rich tasks. Additionally, our RGB-based pose estimation teleoperation system introduces significant noise in global motion estimation, making agile locomotion operations challenging. Future work will explore adopting more precise teleoperation systems, incorporating online retargeting algorithms.

Acknowledgement

This study is supported by National Natural Science Foundation of China (62206172). We are grateful to Jingbo Wang, Shenyuan Gao, and Tairan He for their valuable discussions. We would like to acknowledge Jiacheng Qiu, Shijia Peng, Haoran Jiang, and Zherui Qiu for their assistance and support throughout this project. We also sincerely thank Kinetix AI for supporting the real-world experiments and assisting with demo filming.

References

[1] 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," *arXiv preprint arXiv:2501.02116*, 2025.

[2] I. Radosavovic, T. Xiao, B. Zhang, T. Darrell, J. Malik, and K. Sreenath, "Real-world humanoid locomotion with reinforcement learning," *Science Robotics*, 2023.

[3] I. Radosavovic, B. Zhang, B. Shi, J. Rajasegaran, S. Kamat, T. Darrell, K. Sreenath, and J. Malik, "Humanoid locomotion as next token prediction," *NIPS*, 2024.

[4] 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 *RSS*, 2024.

[5] P. Dugar, A. Shrestha, F. Yu, B. van Marum, and A. Fern, "Learning multi-modal whole-body control for real-world humanoid robots," *arXiv preprint arXiv:2408.07295*, 2024.

[6] T. He, W. Xiao, T. Lin, Z. Luo, Z. Xu, Z. Jiang, J. Kautz, C. Liu, G. Shi, X. Wang *et al.*, "Hover: Versatile neural whole-body controller for humanoid robots," in *ICRA*, 2025.

[7] C. Tessler, Y. Guo, O. Nabati, G. Chechik, and X. B. Peng, "Maskedmimic: Unified physics-based character control through masked motion inpainting," *ACM Trans. on Graphics*, 2024.

[8] C. Zhang, W. Xiao, T. He, and G. Shi, "Wococo: Learning whole-body humanoid control with sequential contacts," in *CoRL*, 2024.

[9] Z. Zhuang, S. Yao, and H. Zhao, "Humanoid parkour learning," *arXiv preprint arXiv:2406.10759*, 2024.

[10] T. He, J. Gao, W. Xiao, Y. Zhang, Z. Wang, J. Wang, Z. Luo, G. He, N. Sobanbab, C. Pan *et al.*, "ASAP: Aligning simulation and real-world physics for learning agile humanoid whole-body skills," in *RSS*, 2025.

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

[12] 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 *CoRL*, 2025.

[13] 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.

[14] T. He, Z. Luo, X. He, W. Xiao, C. Zhang, W. Zhang, K. M. Kitani, C. Liu, and G. Shi, "OmniH2O: Universal and dexterous human-to-humanoid whole-body teleoperation and learning," in *CoRL*, 2024.

[15] X. Cheng, Y. Ji, J. Chen, R. Yang, G. Yang, and X. Wang, "Expressive whole-body control for humanoid robots," in *RSS*, 2024.

[16] N. Mahmood, N. Ghorbani, N. F. Troje, G. Pons-Moll, and M. J. Black, "AMASS: Archive of motion capture as surface shapes," in *ICCV*, 2019.

[17] X. B. Peng, P. Abbeel, S. Levine, and M. Van de Panne, "DeepMimic: Example-guided deep reinforcement learning of physics-based character skills," *ACM Trans. on Graphics*, 2018.

[18] M. Ji, X. Peng, F. Liu, J. Li, G. Yang, X. Cheng, and X. Wang, "Exbody2: Advanced expressive humanoid whole-body control," *arXiv preprint arXiv:2412.13196*, 2024.

[19] 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," in *IROS*, 2025.

[20] Y. Ze, Z. Chen, J. P. Araújo, Z. ang Cao, X. B. Peng, J. Wu, and C. K. Liu, "TWIST: Teleoperated whole-body imitation system," in *CoRL*, 2025.

[21] Z. Zhuang and H. Zhao, "Embrace Collisions: Humanoid shadowing for deployable contact-agnostics motions," in *CoRL*, 2025.

[22] T. He, Z. Luo, W. Xiao, C. Zhang, K. Kitani, C. Liu, and G. Shi, "Learning human-to-humanoid real-time whole-body teleoperation," in *IROS*, 2024.

[23] C. Ionescu, D. Papava, V. Olaru, and C. Sminchisescu, "Human3. 6m: Large scale datasets and predictive methods for 3d human sensing in natural environments," *IEEE TPAMI*, 2013.

[24] S. Shin, J. Kim, E. Halilaj, and M. J. Black, "WHAM: Reconstructing world-grounded humans with accurate 3d motion," in *CVPR*, 2024.

[25] 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 Conference Proceedings*, 2024.

[26] 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," *arXiv preprint arXiv:2505.03729*, 2025.

[27] H. Li, Y. Cui, and D. Sadigh, "How to train your robots? the impact of demonstration modality on imitation learning," *arXiv preprint arXiv:2503.07017*, 2025.

[28] Y. Xue, W. Dong, M. Liu, W. Zhang, and J. Pang, "A unified and general humanoid whole-body controller for fine-grained locomotion," in *RSS*, 2025.

[29] Q. Ben, F. Jia, J. Zeng, J. Dong, D. Lin, and J. Pang, "HOMIE: Humanoid loco-manipulation with isomorphic exoskeleton cockpit," in *RSS*, 2025.

[30] J. Li, X. Cheng, T. Huang, S. Yang, R. Qiu, and X. Wang, "AMO: Adaptive motion optimization for hyper-dexterous humanoid whole-body control," in *RSS*, 2025.

[31] Y. Shao, X. Huang, B. Zhang, Q. Liao, Y. Gao, Y. Chi, Z. Li, S. Shao, and K. Sreenath, "LangWBC: Language-directed humanoid whole-body control via end-to-end learning," in *RSS*, 2025.

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

[33] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," *arXiv preprint arXiv:1707.06347*, 2017.

[34] J. Lee, J. Hwangbo, L. Wellhausen, V. Koltun, and M. Hutter, "Learning quadrupedal locomotion over challenging terrain," *Science Robotics*, 2020.

[35] K. Levenberg, "A method for the solution of certain non-linear problems in least squares," *Quarterly of applied mathematics*, 1944.

[36] D. W. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," *Journal of the society for Industrial and Applied Mathematics*, 1963.

[37] Y. Nakamura and H. Hanafusa, "Inverse kinematic solutions with singularity robustness for robot manipulator control," *J. Dyn. Sys., Meas., Control.*, 1986.

[38] C. M. Kim, B. Yi, H. Choi, Y. Ma, K. Goldberg, and A. Kanazawa, "PyRoki: A modular toolkit for robot kinematic optimization," *arXiv preprint arXiv:2505.03728*, 2025.
[39] F. G. Harvey, M. Yurick, D. Nowrouzezahrai, and C. Pal, "Robust motion in-betweening," *ACM Trans. on Graphics*, 2020.
[40] 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 *CoRL*, 2025.
[41] 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 preprint arXiv:2507.07356*, 2025.
[42] 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," in *NeurIPS*, 2021.
[43] Unitree Robotics, "Unitree g1 humanoid robot," 2024, available at: https://www.unitree.com/g1.
[44] I. Sárándi, T. Linder, K. O. Arras, and B. Leibe, "MeTRAbs: metric-scale truncation-robust heatmaps for absolute 3D human pose estimation," *IEEE T-BIOM*, 2021.