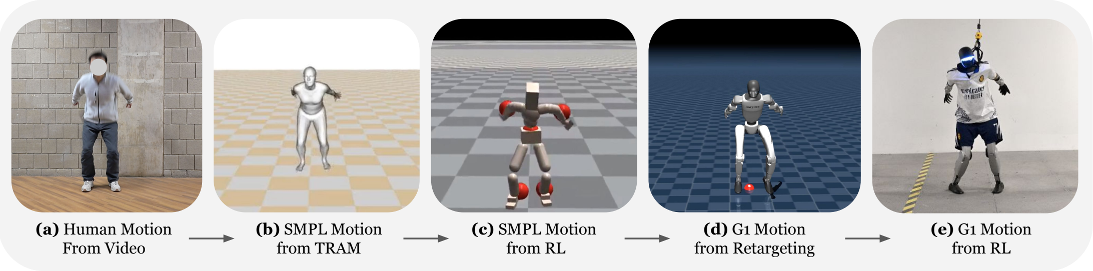
(a) Human Motion From Video
(b) SMPL Motion from TRAM
(c) SMPL Motion from RL
(d) G1 Motion from Retargeting
(e) G1 Motion from RL

Fig. 3. Retargeting Human Video Motions to Robot Motions: (a) Human motions are captured from video. (b) Using TRAM [93], 3D human motion is reconstructed in the SMPL parameter format. (c) A reinforcement learning (RL) policy is trained in simulation to track the SMPL motion. (d) The learned SMPL motion is retargeted to the Unitree G1 humanoid robot in simulation. (e) The trained RL policy is deployed on the real robot, executing the final motion in the physical world. This pipeline ensures the retargeted motions remain physically feasible and suitable for real-world deployment.

system. The collected data are then replayed in simulation, where the dynamics mismatch manifests as tracking errors. We then train a delta action model that learns to compensate for these discrepancies by minimizing the difference between real-world and simulated states. This model effectively serves as a residual correction term for the dynamics gap. Finally, we fine-tune the pre-trained policy using the delta action model, allowing it to adapt effectively to real-world physics.

We validate ASAP on diverse agile motions and successfully achieve whole-body agility on the Unitree G1 humanoid robot [77]. Our approach significantly reduces motion tracking error compared to prior SysID, DR, and delta dynamics learning baselines in both sim-to-sim (IsaacGym to IsaacSim, IsaacGym to Genesis) and sim-to-real (IsaacGym to Real) transfer scenarios. Our contributions are summarized below.

1) We introduce ASAP, a framework that bridges the sim-to-real gap by leveraging a delta action model trained via reinforcement learning (RL) with real-world data.
2) We successfully deploy RL-based whole-body control policies in the real world, achieving previously difficult-to-achieve humanoid motions.
3) Extensive experiments in both simulation and real-world settings demonstrate that ASAP effectively reduces dynamics mismatch, enabling highly agile motions on robots and significantly reducing motion tracking errors.
4) To facilitate smooth transfer between simulators, we develop and open-source a multi-simulator training and evaluation codebase for help accelerate further research.

II. Pre-training: Learning Agile Humanoid Skills

A. Data Generation: Retargeting Human Video Data

To track expressive and agile motions, we collect a video dataset of human movements and retarget it to robot motions, creating imitation goals for motion-tracking policies, as shown in Figure 3 and Figure 2 (a).

a) Transforming Human Video to SMPL Motions: We begin by recording videos (see Figure 3 (a) and Figure 12) of humans performing expressive and agile motions. Using TRAM [93], we reconstruct 3D motions from videos. TRAM estimates the global trajectory of the human motions in SMPL parameter format [52], which includes global root translation, orientation, body poses, and shape parameters, as shown in Figure 3 (b). The resulting motions are denoted as $\mathcal{D}_{\text{SMPL}}$.

b) Simulation-based Data Cleaning: Since the reconstruction process can introduce noise and errors [25], some estimated motions may not be physically feasible, making them unsuitable for motion tracking in the real world. To address this, we employ a "sim-to-data" cleaning procedure. Specifically, we use MaskedMimic [86], a physics-based motion tracker, to imitate the SMPL motions from TRAM in IsaacGym simulator [58]. The motions (Figure 3 (c)) that pass this simulation-based validation are saved as the cleaned dataset $\mathcal{D}_{\text{SMPL}}^{\text{Cleaned}}$.

c) Retargeting SMPL Motions to Robot Motions: With the cleaned dataset $\mathcal{D}_{\text{SMPL}}^{\text{Cleaned}}$ in SMPL format, we retarget the motions into robot motions following the shape-and-motion two-stage retargeting process [25]. Since the SMPL parameters estimated by TRAM represent various human body shapes, we first optimize the shape parameter $\boldsymbol{\beta}'$ to approximate a humanoid shape. By selecting 12 body links with correspondences between humans and humanoids, we perform gradient descent on $\boldsymbol{\beta}'$ to minimize joint distances in the rest pose. Using the optimized shape $\boldsymbol{\beta}'$ along with the original translation $\boldsymbol{p}$ and pose $\boldsymbol{\theta}$, we apply gradient descent to further minimize the distances of the body links. This process ensures accurate motion retargeting and produces the cleaned robot trajectory dataset $\mathcal{D}_{\text{Robot}}^{\text{Cleaned}}$, as shown in Figure 3 (d).

B. Phase-based Motion Tracking Policy Training

We formulate the motion-tracking problem as a goal-conditioned reinforcement learning (RL) task, where the policy π is trained to track the retargeted robot movement trajectories in the dataset $\mathcal{D}_{\text{Robot}}^{\text{Cleaned}}$. Inspired by [67], the state s_t includes the robot's proprioception s_t^{p} and a time phase variable $\phi \in [0, 1]$, where $\phi = 0$ represents the start of a motion and $\phi = 1$ represents the end. This time phase variable alone ϕ is proven to be sufficient to serve as the goal state $\boldsymbol{s}_t^{\text{g}}$ for single-motion tracking [67]. The proprioception s_t^{p} is defined as $s_t^{\text{p}} \triangleq \left[\boldsymbol{q}_{t-4:t}, \dot{\boldsymbol{q}}_{t-4:t}, \boldsymbol{\omega}^{root}_{t-4:t}, \boldsymbol{g}_{t-4:t}, \boldsymbol{a}_{t-5:t-1}\right]$, with 5-step history of joint position $\boldsymbol{q}_t \in \mathbb{R}^{23}$, joint velocity $\dot{\boldsymbol{q}}_t \in \mathbb{R}^{23}$, root angular velocity $\boldsymbol{\omega}^{root}_t \in \mathbb{R}^3$, root projected gravity $\boldsymbol{g}_t \in \mathbb{R}^3$, and last action $\boldsymbol{a}_{t-1} \in \mathbb{R}^{23}$. Using the agent's proprioception s_t^{p} and the goal state $\boldsymbol{s}_t^{\text{g}}$, we define the reward as $r_t = \mathcal{R}\left(s_t^{\text{p}}, s_t^{\text{g}}\right)$, which is used for policy optimization. The specific reward terms can be found in Table I. The action $\boldsymbol{a}_t \in \mathbb{R}^{23}$ corresponds to the target joint positions and is passed to a