2026 IEEE International Conference on Robotics and Automation (ICRA 2026)
June 1-5, 2026. Vienna, Austria

SoftMimic: Learning Compliant Whole-body Control from Examples

Gabriel B. Margolis* Michelle Wang* Nolan Fey Pulkit Agrawal

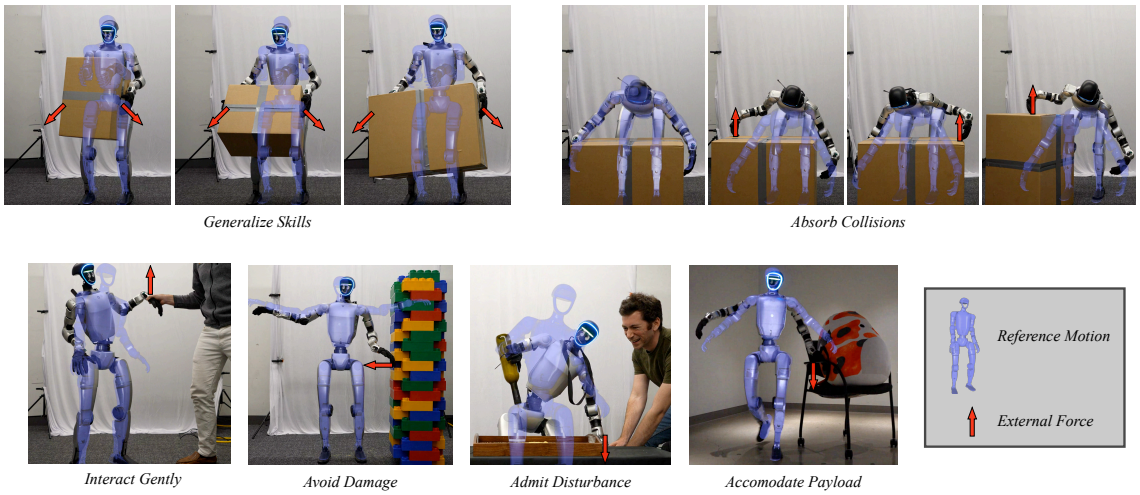


Fig. 1: **SoftMimic for Compliant Motion Tracking.** We train humanoid policies that compliantly respond to external forces while tracking a reference motion. The desired force-displacement relationship is modulated by a 'stiffness' input at deployment time, and a single policy learns to realize a wide range of stiffnesses. In diverse real-world experiments, SoftMimic benefits generalization and safety. In the images, the reference motion is visualized in blue, and the approximate external force on the robot is illustrated by the red arrows.

***Abstract*— We introduce *SoftMimic*, a framework for learning compliant whole-body control policies for humanoid robots from example motions. Imitating human motions with reinforcement learning allows humanoids to quickly learn new skills, but existing methods incentivize stiff control that aggressively corrects deviations from a reference motion, leading to brittle and unsafe behavior when the robot encounters unexpected contacts. In contrast, SoftMimic enables robots to respond compliantly to external forces while maintaining balance and posture. Our approach leverages an inverse kinematics solver to generate an augmented dataset of feasible compliant motions, which we use to train a reinforcement learning policy. By rewarding the policy for matching compliant responses rather than rigidly tracking the reference motion, SoftMimic learns to absorb disturbances and generalize to varied tasks from a single motion clip. We validate our method through simulations and real-world experiments, demonstrating safe and effective interaction with the environment.**

I. INTRODUCTION

A major goal in humanoid robotics is to build agents capable of performing a vast range of tasks humans execute in everyday environments. A promising avenue towards this goal is to leverage large-scale human motion capture data, enabling robots to learn human-like behaviors through imitation [1]. Recent work has successfully trained policies for tracking single motions, diverse motion datasets, and even real-time teleoperation on humanoid hardware [2]–[6]. These methods produce impressive, dynamic motions.

All authors are with the Improbable AI Lab, Massachusetts Institute of Technology, USA. Correspondence to: {gmargo, wangmj}@mit.edu
* indicates co-first authors.
Website: https://gmargo11.github.io/softmimic

However, motion tracking is usually insufficient for safe and effective deployment in the real world, where sensing uncertainty and frequent, unplanned physical (i.e., contact-rich) interactions are commonplace. Policies trained to rigidly track a reference motion treat any deviation in the robot's motion as an error to be corrected aggressively. Consequently, when the robot makes an unexpected contact, such as brushing against a table, misjudging an object's location, or interacting with a person, the controller attempts to correct the motion error caused by the contact with large, uncontrolled forces, resulting in brittle and potentially dangerous behavior. This lack of compliance is also a fundamental barrier to deploying humanoids alongside people, leading to the current state of humanoids operating in isolation.

To address these shortcomings and pave the path for real-world humanoid deployment, we propose a framework for *compliant* whole-body motion tracking called *SoftMimic*. The objective of *SoftMimic* is not to blindly minimize tracking error, but to controllably *depart* from the reference motion in response to external forces according to a user-specified stiffness. A lower stiffness setting allows the robot to comply more and thereby deviate more from the reference trajectory, given the same force disturbance. Achieving compliant behavior on a high-DoF humanoid is challenging because complying with a force on a single limb requires coordinated, full-body adjustments to maintain balance and preserve the overall posture and style of the motion.

Directly learning compliant behavior with reinforcement learning (RL) poses significant exploration challenges, as a

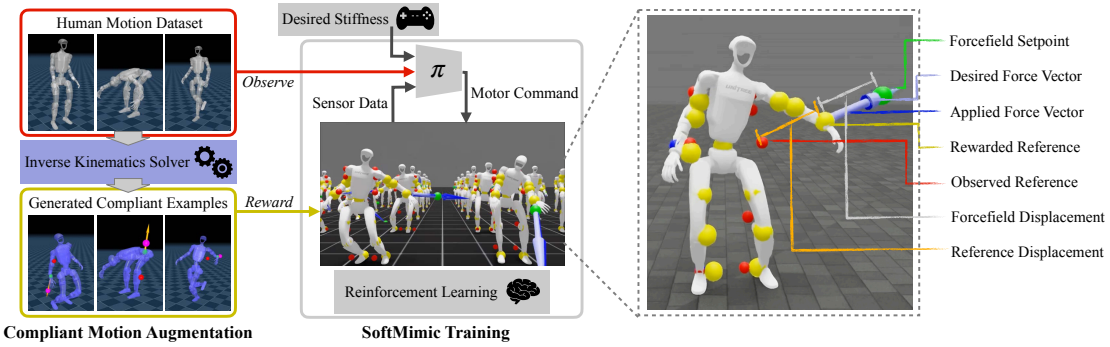


Fig. 2: **Soft Whole-body Control via Compliant Motion Augmentation.** ***Left***: Given an original reference motion (q_{ref}) and a specified interaction (external wrench W_{ext} and stiffness K_{robot}), our offline data generation stage uses an IK solver to generate a kinematically feasible and stylistically consistent compliant motion (q_{aug}). ***Right***: During online training, a policy learns to reproduce this behavior. It observes the robot's proprioceptive state and the original reference (q_{ref}), but is rewarded for matching the augmented target (q_{aug}). This forces the policy to implicitly infer the external wrench and react appropriately, resulting in a robot that can controllably comply with generalized unanticipated perturbations. In the graphic, we only annotate translational forces and displacements for ease of interpretation, but the analogous rotational quantities are also simulated.

stiff, non-compliant policy is often a strong local optimum. In scenarios where the robot must comply substantially, reward terms like tracking keypoints and joint angles, which typically reinforce each other, can come into conflict, making it difficult to balance them across a large potential solution space [7]. Furthermore, many desired compliant responses are kinematically or dynamically infeasible depending on the robot's configuration, resulting in training for impossible tasks, which hinders learning [8].

To overcome these challenges, we adopt a learning-from-examples strategy. Instead of tuning RL to discover compliant behaviors through artisanal reward shaping, we first generate a dataset of kinematic references for compliant behaviors. We use an inverse kinematics (IK) solver to author a large-scale dataset of feasible and stylistically-consistent compliant trajectories for a wide range of interaction scenarios. This offline process allows us to filter out impossible tasks and precisely define the desired whole-body coordination. We then train a policy using RL, where the agent observes the robot's state and the original, non-compliant reference motion, but is rewarded for tracking the corresponding pre-computed compliant trajectory from our augmented dataset. This formulation forces the policy to learn to infer the external forces from proprioceptive sensing and react with the demonstrated compliant behavior.

Our experiments demonstrate that this approach yields a policy capable of tracking reference motions while exhibiting predictable, controllable compliance. Our compliant controller is more robust to disturbances, can generalize a single motion clip to handle variations in a task (e.g., picking up boxes of different sizes), and safely manages unexpected collisions. Crucially, these benefits are achieved while preserving good motion tracking performance in the absence of external forces. We validate our framework in simulation and on a real Unitree G1 humanoid robot.

II. RELATED WORK

A. Learning Humanoid Whole-Body Control

The convergence of recent progress in articulated rigid-body simulation [9]–[11], sim-to-real transfer techniques [12]–[14], and advanced legged hardware [15]–[18], combined with reinforcement learning paradigms like DeepMimic [1], [19], has enabled impressive performance in humanoid motion imitation [2], [5], [6] and real-time teleoperation [3], [4], [20], [21], with some works further training visuomotor policies on the resulting demonstrations [22]–[25]. A key challenge arises when the robot contacts objects absent from the reference motion: light objects may be pushed aside, but heavier objects impede tracking and raise the question of what posture the robot should adopt and how much force it should exert. Modern quasi-direct-drive actuators support torque sensing and software-emulated stiffness, and a prevalent strategy within learning-based frameworks is for a neural network policy to send target setpoints to per-joint PD controllers, where position targets can be modulated to regulate applied forces [13] and PD gains can be tuned to shape exploration distributions [26]. However, as we show, the stiffness of a policy's interactions is dictated foremost by its high-level incentives, i.e., its reward function and training data, and neural network policies trained with constant low-level gains are capable of representing a wide range of stiff and compliant behaviors in task space. Indeed, works that combine motion-tracking rewards with random external pushes instruct the robot to follow the same trajectory regardless of interaction force [21], [27], [28], encouraging the policy to apply arbitrary resistive forces and act as stiff as possible.

B. Analytical Approaches to Compliant Control

Hybrid position/force control [29] and task-space impedance–admittance control [30] are longstanding formulations for compliant manipulation, prescribing a motion–wrench relationship such as a virtual

mass–spring–damper at the end effector; one line of work implements this relationship so the closed loop is *passive* at the interaction ports for robustness and safety, while another uses inverse kinematics/dynamics to synthesize joint trajectories realizing desired task motions and apparent stiffness. Extending these ideas to humanoids requires incorporating floating-base dynamics, intermittent contacts, and coordination of interaction objectives with posture and balance, which the operational-space formulation [31] and its whole-body extensions address through contact-consistent projections [7], [32]–[34], complemented by passivity-based methods that ensure compliant hand and foot interaction while balancing [35]–[41], notably demonstrated on the DLR TORO platform [42]. Drawing inspiration from this literature, our approach (SoftMimic) adopts the classical goal of making the robot behave like a spring in response to generalized disturbances, but replaces hand-engineered controllers with a learned policy trained on procedurally generated compliant trajectories based on simple kinematic heuristics, where the RL training stage accounts for the full dynamics model and the policy observes proprioception and the original reference to reproduce the authored compliant deviation at a user-specified stiffness.

C. Data-Driven Compliant Control

Reinforcement learning has recently been used to directly learn compliant behaviors: Deep Compliant Control [43] demonstrated success with simulated characters but relied on perfect state information, while Portela et al. [44] showed that an end-to-end policy trained in simulation can learn accurate task-space force application on a real legged manipulator using only proprioceptive sensing, facilitating impedance control of the end-effector. Other work has trained locomotion policies to mimic specific dynamic models such as spring-mass-damper templates [45], a concept extended by FACET [46] to various embodiments, and UniFP [47] demonstrated that explicit force information from such policies can benefit downstream imitation learning. However, these prior approaches focus on controlling force interaction while satisfying simple locomotion tasks, whereas humanoid robots learning from teleoperation or demonstrations must reconcile interaction objectives with whole-body motion tracking. Our work addresses the open challenge of unifying wide-range impedance control with high-fidelity motion mimicry on real hardware by training a single policy to imitate reference motions while achieving user-specified stiffness, enabling both soft compliance and stiff resistance (Figure 8).

III. Method

Our goal is to train a policy that enables a humanoid robot to track a whole-body reference motion while compliantly responding to external forces with a user-specified stiffness. A naive approach could involve a standard motion imitation setup [1] with an additional task reward for compliant responses. Directly optimizing this objective with RL is not ideal for several reasons. First, exploration is brittle: We hypothesize that a purely stiff tracker is a strong local optimum that suppresses compliant responses when these rewards are in conflict. Second, the humanoid's large postural null-space makes reward design balancing interaction forces with whole-body style and stability nontrivial. Third, the robot's feasible compliance is highly dependent on its configuration due to kinematic constraints and sensing limitations, yielding many tasks infeasible. Fourth, the desired deviation from reference motions is incompatible with the use of *early termination* and *reference state initialization* strategies commonly used to stabilize and accelerate training.

Our solution to these problems is to generate an *augmented* dataset with reference motions that specify how the robot should comply to different external forces. This sets up a motion tracking problem where the robot observes the original motion target but is rewarded for inferring the force interaction and matching the applicable augmented target. We show that this approach enables fine-grained control of the compliant response. A key challenge is how to generate a dataset of feasible complying motions that preserve desired components of the original motion style. In this work, we use differential inverse kinematics to ensure kinematically feasible and stylistically desirable reference motions, and an analysis of force and position sensing noise to specify feasible force/compliance tasks.

A. Compliant Motion Tracking (CMT)

Given an original reference configuration $q_{\text{ref}}(t)$ and an external wrench on link i, $\mathbf{w}_i(t) = (\mathbf{F}_i(t), \boldsymbol{\tau}_i(t))$, with a commanded robot stiffness $\mathbf{K}_{\text{cmd}} = \text{diag}(K^t_{\text{cmd}}\mathbf{I}_3,\ K^r_{\text{cmd}}\mathbf{I}_3)$, we define the *desired compliant target pose* for link i relative to the reference:

$$\mathbf{p}_{i,\text{des}} = \mathbf{p}_{i,\text{ref}} + \frac{1}{K^t_{\text{cmd}}}\,\mathbf{F}_i, \quad R_{i,\text{des}} = R_{i,\text{ref}}\exp\big([\boldsymbol{\tau}_i/K^r_{\text{cmd}}]_\times\big).$$

Let $T_i(q) = (R_i(q), \mathbf{p}_i(q))$ denote the pose of link i. The instantaneous IK objective is

$$\hat{q} = \arg\min_q d(q, q_{\text{ref}}) \quad \text{s.t.}\ \ \mathbf{p}_i(q) \approx \mathbf{p}_{i,\text{des}},\ \ R_i(q) \approx R_{i,\text{des}}.$$

This dictates that link i behaves like a spring with stiffness $(K^t_{\text{cmd}}, K^r_{\text{cmd}})$ around the reference, while the rest of the posture stays as close as possible to q_{ref} under the distance metric d. When the robot's stiffness is low or the external force is large, the optimal configuration $\hat{q}$ can deviate significantly from the reference q_{ref}. In such cases, the choice of the metric d (e.g., a simple joint-space error like $\|q - q_{\text{ref}}\|^2$ versus a task-space error on other keypoints) has a large impact on the resulting behavior. This contrasts with typical motion tracking scenarios where the optimal solution remains close to the reference, and all errors are near zero. In this work, we choose to define d using a mixture of keypoint error, joint position error, foot placement consistency, and center-of-pressure maintenance as described in Section III-C.

B. SoftMimic: Reinforcement Learning for CMT

Observation, Reward, Action Space. We formulate compliant whole-body control as a reinforcement learning

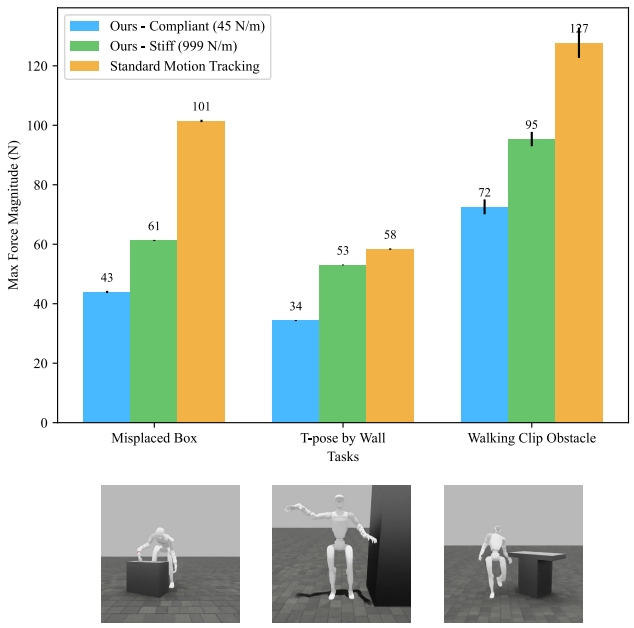


Fig. 3: **SoftMimic reduces collision forces across various motions in unseen environments.** The bar chart compares the maximum contact force generated by our policy (at low and high stiffness) and the stiff baseline across three challenging scenarios involving unexpected contact. In all cases, the compliant policy operating at a low stiffness significantly reduces interaction forces, enhancing safety.

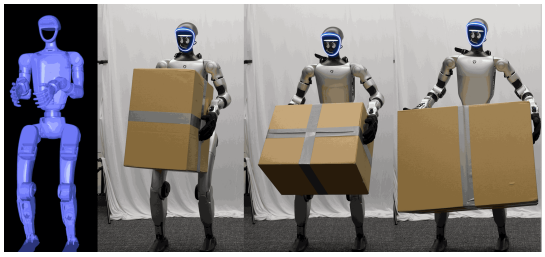
problem. The policy observes a state history containing the robot's proprioceptive information $[q_t, \dot{q}_t]$, partial base state $[g_t^b, \omega_t^b]$, previous action a_{t-1}, and reference posture q_t^{ref}. The agent is rewarded with a sum of a DeepMimic-style reference motion tracking reward, $r_{\text{ref}} + r_{\text{smooth}}$, and a spring-like compliance reward, $r_{\text{spring}} = r_{\text{force}} + r_{\text{torque}} + r_{\text{pos}} + r_{\text{rot}}$, which depends on the current external wrench W_i. The policy outputs joint-space position targets for a PD controller with moderate gains, enabling torque control by modulating the position error.

Observation Content. The policy can implicitly learn admittance-style (estimate wrench, command pose), impedance-style (estimate pose, command wrench), or hybrid strategies depending on the desired stiffness and external force profile. Note that the policy directly observes neither the wrench nor displacement information, but can make inferences about them based on proprioceptive sensing. For an impedance strategy, the end-effector pose can be inferred from the joint positions q_t and root orientation g_t^b via forward kinematics. For an admittance strategy, the external wrench can be inferred from the robot's dynamics, using observations of previous joint position q_{t-1}, joint position target a_{t-1}, joint velocity $\dot{q}_{t-1}$, and joint accelerations (derived from $\dot{q}_t, \dot{q}_{t-1}, \omega_t^b, \omega_{t-1}^b$). To ensure this temporal information is available, the policy observes a history of the past 3 observation steps. As is standard in legged systems, the full root state and contact states are not directly observed; instead, the policy may partially infer them as needed, leveraging the associations between historical observations, commands, and simulation outcomes.

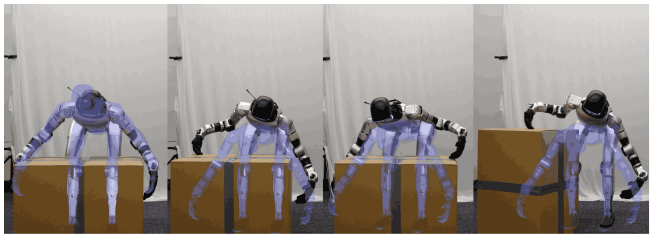
Command Sampling and Force Field Dynamics. Training episodes consist of sampling a motion clip, a desired robot stiffness, an environment/contact stiffness and an applied wrench. The external force and torque are implemented as "fields" [44] which pull a selected link of the robot towards a moving setpoint with a distance-proportional force and rotation-proportional torque according to the environment stiffness K_{env} (Figure 2). $K_{\text{env}} \to 0$ corresponds to a constant-force source (an admittance-like environment) and $K_{\text{env}} \to \infty$ corresponds to an immovable object (an impedance-like environment) [48]. Additional details of the force sampling parameters are provided in the appendix.

- *Stiffness Bounds:* When sensing and dynamics are noisy and the robot's state is only partially observable, inferences about pose and wrench also become noisy. This noise makes realizing highly sensitive responses, including extremely low or high stiffnesses, infeasible. To address this, we first train a state estimator to establish the approximate noise floor of the pose and wrench estimates. We then use these noise values in an idealized analysis to guide our stiffness sampling range. We define requirements of $10\,\text{N}$ force accuracy and $10\,\text{cm}$ position accuracy, and empirically observe that the learned force estimator has average noise of $4\,\text{N}$. Thus, an admittance control strategy should be able to achieve the positioning target only for $K > \frac{4\,\text{N}}{0.10\,\text{m}} = 40\,\text{N/m}$. Likewise, with a position estimation noise of $1\,\text{cm}$, an impedance controller can achieve the desired force accuracy only if $K < \frac{10\,\text{N}}{0.01\,\text{m}} = 1000\,\text{N/m}$. This analysis establishes approximate upper and lower feasible bounds for training.
- *Log-Uniform Sampling:* We aim to realize behaviors across a wide range of stiffness and compliance values. Since compliance is the inverse of stiffness, uniformly sampling stiffness would heavily bias the dataset towards high-stiffness, low-compliance behaviors, and vice versa. A change in stiffness from 1040 to $1080\,\text{N/m}$ is a minor tweak to a stiff behavior, while a change from 40 to $80\,\text{N/m}$ is a significant change for a compliant one. To ensure we explore these different regimes equally, we sample both the robot and environment stiffness from a log-uniform distribution.
- *Velocity-based Event Sampling:* Suppose that every point in space has the same probability of containing a stationary collision surface. Then if the robot is moving with no information about its surroundings, its probability of some point on the robot colliding is proportional to the point's velocity. Therefore, we sample force event onsets for each link with probability proportional to its velocity, with a small constant additional probability for colliding while stationary.

Early Termination and Reference State Initialization. It



(a) **Generalization of a single reference:** Real-world deployment with the same reference motion and stiffness: successful picks across multiple box widths with a gentle, consistent squeeze.



(b) **Zero-shot robustness to misalignment:** picking a large box with nominal alignment (left) and under lateral/rotational misinitializations (middle–right). All behavior is achieved without simulating boxes or defining a prior over their location; robustness comes from SoftMimic training with generalized external forces.

Fig. 4: **SoftMimic enables generalization to unseen objects and disturbance scenarios.** Using a single motion reference designed for a 20cm box, our policy can successfully pick up boxes of increasing width. By commanding the same low stiffness, the robot not only successfully picks up different sized boxes with a consistent, gentle squeezing force, but it also is able to safely handle collisions with misaligned boxes. In contrast, the stiff baseline generates large and uncontrolled force spikes, risking damage to the object or robot.

is common practice to exploit early termination and reference state initialization to accelerate and stabilize motion imitation [1]. A key advantage of our data augmentation approach (Section III-C) is that we can appropriately terminate and initialize episodes while the robot is under load by using the augmented compliant posture q_{aug} as the reference. Without this augmented data, there would be no way to initialize the robot consistently with active wrenches as complying with the load alters the desired posture.

C. Compliant Motion Augmentation (CMA)

A key challenge with the RL problem posed in Section III-B is that the final compliant posture arises from a competition between the motion tracking and spring-behavior rewards. This complicates exploration and makes the resulting behavior difficult to tune and predict.

Our proposed solution is to pre-generate an augmented motion dataset, D_{aug}, that explicitly contains whole-body responses that desirably balance the tradeoff between compliantly responding to force events and maintaining accurate motion tracking. This offline process enables two key advantages: 1) we can reject infeasible commands before RL training using simple kinematic and dynamic checks, and 2) we can precisely specify the desired compliant behavior through a structured optimization. We generate D_{aug} using a differential inverse kinematics (IK) solver.

Task Hierarchy: The IK solver optimizes for the following objectives:

1) **Compliant Interaction (high priority; $w = 5.0$).** For the interacting link i, we define the desired pose via the commanded stiffness $(K^t_{\text{cmd}}, K^r_{\text{cmd}})$ and the wrench $\mathbf{w}_i = (\mathbf{F}_i, \boldsymbol{\tau}_i)$:

$$\mathbf{p}_{i,\text{des}} = \mathbf{p}_{i,\text{ref}} + \frac{1}{K^t_{\text{cmd}}}\mathbf{F}_i,$$

$$R_{i,\text{des}} = R_{i,\text{ref}} \exp\big([\boldsymbol{\tau}_i / K^r_{\text{cmd}}]_\times\big).$$

penalizing $\|\mathbf{p}_i(q) - \mathbf{p}_{i,\text{des}}\|^2$ and $\|\log(R^\top_{i,\text{des}} R_i(q))^\vee\|^2$. We perturb a single link at a time (hands in this work).
2) **Foot Placement (high priority; $w = 2.5$).** High-weight link pose tasks ensure that stance feet remain consistent with the reference contact schedule.
3) **CoM Stabilization (medium priority; $w = 0.1$).** A Center of Pressure (CoP)-aware Center of Mass (CoM) task provides moment compensation while allowing necessary body shifts.
4) **Keypoint Posture (low priority; $w = 0.01$).** Moderate-weight pose tasks on key links (e.g., elbows, shoulders, torso) preserve the original motion's style.
5) **Joint Posture (very low priority; $w = 10^{-4}$).** A regularization task tethers all degrees of freedom towards the reference configuration q_{ref} to resolve redundancy.

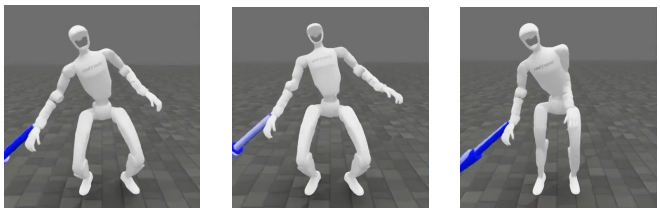
This hierarchy yields a continuous and feasible adapted joint trajectory, $q_{\text{aug}}(t)$, that embodies the desired compliant response across various interaction scenarios. When the IK solver fails to find a solution for a given wrench, we rewind the motion clip and iteratively scale down the wrench, rejecting the event entirely if the wrench falls below the sensing noise floor.

D. Motion Data, Training Details, Baselines

Motion Data. We trained and deployed compliant whole-body control policies on a Unitree G1 humanoid—one policy for each motion clip: standing, T-pose-move, walk, box-pick, pour, and dance, using identical hyperparameters. The motion data comes from the AMASS [49] and LAFAN1 [50] datasets, retargeted using methods from prior work [2], [4].

For each motion clip, we generate augmented data by solving the aforementioned inverse kinematics problem using Mink [51], [52] and MuJoCo [10]. The offline process is highly efficient, allowing us to generate 40 minutes of augmented data for a one-minute clip in approximately one minute of wall-clock time when parallelized. This produces a dataset of tuples $\big(q_{\text{ref}},\ \mathbf{w}_i,\ \mathbf{K}_{\text{cmd}},\ q_{\text{aug}}\big)$ that defines all interaction events for training.

Training Hyperparameters. Linear stiffness commands ranged from $40\,\text{N}\,\text{m}^{-1}$ to $1000\,\text{N}\,\text{m}^{-1}$; angular stiffness from $0.1\,\text{N}\,\text{m}\,\text{rad}^{-1}$ to $10\,\text{N}\,\text{m}\,\text{rad}^{-1}$. We train using PPO [19]



(a) IK Style 1 (b) IK Style 2 (c) Ablation: `no-aug`

Fig. 5: **Compliant Motion Augmentation provides fine-grained control over compliant style.** Three different compliant policies are shown receiving the same external force in simulation with the same commanded stiffness. By adjusting cost terms in the offline IK solver, such as adding a pelvis orientation cost in (b) compared to (a), we can specify distinct whole-body coordination strategies. The learned policies successfully reproduce the authored styles. In contrast, the policy trained without augmented data (`no-aug`, c) adopts an unpredictable emergent posture that also performs worse.

with the default hyperparameters from the IsaacLab and `rsl_rl` libraries [53].

Baselines. To rigorously evaluate our method, we compare it against two carefully designed baselines that aim to isolate the different components of our framework.

1) **Stiff Baseline:** To demonstrate the value of explicit compliance, we first compare against a high-performance baseline analogous to standard motion imitation methods [1]. This **Stiff Baseline** is trained with a reward function that only incentivizes rigid tracking of the original reference motion, q_{ref}. Crucially, to ensure a fair and direct comparison, this baseline is exposed to the exact same distribution of external force perturbations during training as our compliant policy. This setup tests the emergent behavior of a state-of-the-art tracking controller when faced with physical interactions it is not explicitly rewarded to handle.
2) **`no-aug` Ablation:** To specifically isolate the contribution of our learning-from-example data generation strategy, we design an ablation called **`no-aug`**. This policy is trained with the same spring-like compliance reward, r_{spring}, as our full method, but it does not have access to the augmented dataset D_{aug}. Consequently, its motion tracking reward, reference state initializations, and termination conditions are all based on the original non-compliant reference, q_{ref}. This creates a significant learning challenge: successful compliance generates a large tracking error relative to q_{ref}, which would normally trigger an early termination and thus penalize the desired behavior. To create a meaningful and learnable task, we modify the termination condition for this ablation: an episode only terminates if the robot's state deviates significantly from q_{ref} *without* satisfying the compliant displacement objective. This necessary adjustment allows the policy to explore compliant behaviors without being immediately punished, enabling a fair evaluation of learning with reward shaping alone.

IV. Results

A. Motion Tracking Should Be Compliant

Compliance Improves Task Generalization. Compliant imitation of a single motion can enable its generalization to different task variations. We demonstrate this through a box-picking motion. We apply SoftMimic with a single motion reference of a person picking a box of fixed size. During deployment, a natural approach compatible with non-compliant WBC would be to perceive the size and location of the box visually and map this to a reference motion – either through an explicit perception module or via a learned high-level policy. We are interested in the scenario where the perception module is noisy and erroneously estimates the size or location of the box. In Figure 4, we compare the force exerted in simulation on differently sized boxes by our compliant whole-body controller vs. the standard motion imitation approach; both are tracking a single original motion reference. Our method is able to maintain a lower squeezing force while successfully picking up differently sized boxes, while the standard method exhibits larger and unpredictable forces as it faces larger box sizes outside the scope of the original motion reference.

Compliance Improves Disturbance Handling. We evaluated the response of compliant policies to unseen environmental circumstances common during deployment of humanoid robots.

- *T-Pose by Wall:* The robot attempts to raise its arm while standing next to a wall.
- *Walking Clip Obstacle:* The robot walks past a table and its hand clips the corner.
- *Misplaced Box:* The robot attempts to execute a bending pick while the target box is not centered, and hits its hand on the top of the box.

Figure 3 reports the maximum force the robot exerts on the environment for each task, evaluated at two different stiffness levels of our method as well as with the standard motion tracking approach. The very compliant policy exerts significantly lower forces on the environment, showing that compliant whole body control can more safely handle forceful disturbances compared to existing baselines.

Compliance Improves Safety. Figure 9 shows how modulating the commanded stiffness results in drastically different environment interactions. A small stiffness results in the robot gently pushing on the tower and deviating significantly from the T-pose reference, while a large stiffness causes the robot to strongly resist deviations to the original reference motion, consequently exerting a large force on and toppling the blocks.

Sim-to-Real Validation. Figure 1 shows how the robot complies in the real world during various interactions, demonstrating generalization, disturbance handling, and safety. The useful behaviors associated with compliance all transfer to the real robot.

B. Evaluating Stiffness Adherence

We apply external forces to the standing robot in simulation and measure the resulting displacement across a range

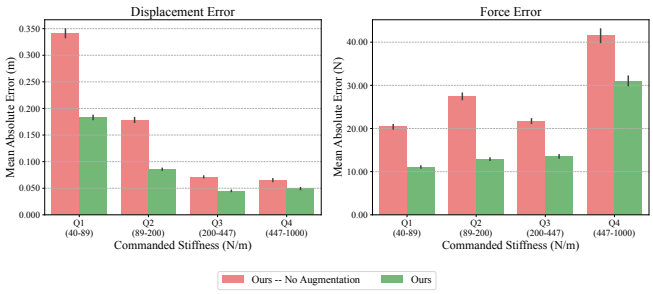


Fig. 6: **Effect of Compliant Motion Augmentation on compliance accuracy.** Policies trained with augmented compliant trajectories (green) attain lower position and force error than the `no-aug` ablation (red), with the largest gains at low stiffness where coordinated whole-body deviations are substantial.

of stiffnesses. Figure 8 shows the median effective stiffness (computed as the ratio between force and displacement) evaluated at various stiffness levels on a log-log plot. The standard motion tracking baseline, which is not conditioned on a stiffness command, yields an effective stiffness of about 500. As can be seen in the supplementary videos, the stiff policy preserves its posture when externally forced but tends to shuffle its feet, which registers as compliance in this evaluation conducted in the global reference frame. Our method displays a consistent sensitivity to the stiffness command across the entire range used in training. Figure 6 shows the displacement is often regulated below 10 cm and force error below 15 N with exceptions at the lowest stiffnesses (elevated displacement error) and highest stiffnesses (elevated force error). Figure 8 and 6 also show that training with augmented references boosts performance compared to the ablation `no-aug`, particularly at low stiffnesses where it results in a 50% reduction in displacement error.

C. Data Shaping Controls Behavior

A key benefit of our framework is the ability to resolve task specification ambiguities in the data augmentation stage. To illustrate this, we train compliant standing policies with two different IK-generated compliant datasets, one with a relatively higher pelvis orientation cost term that encourages the robot to squat and one with a relatively lower term that encourages the robot to bend. Figure 5 shows how the resulting policies respond to perturbations in different styles depending on the behavior designed during the IK dataset generation. It also compares the behavior of the best `no-aug` policy, which displays an emergent postural response resulting from the balance of rewards, which cannot be predicted before performing the expensive RL training.

D. Compliant Control Preserves Motion Quality

Our proposed method achieves compliance when interacting with forces and preserves competitive motion tracking accuracy in the non-perturbed case, even for long and dynamic motion clips. Under no perturbations, we compare the joint position and keypoint tracking error of our method and standard motion tracking for skills used in our demonstrations, as well as a long, challenging dance clip (`dance1_subject2` from LAFAN1 [50]) which has recently been used to demonstrate the high performance of motion tracking systems. Table I shows both our compliant policy and the standard motion tracking baseline achieve small tracking errors. This minor increase in tracking error is an expected trade-off for learning a much richer and more versatile behavioral repertoire. Figure 7 (Appendix) shows the training progression of total reward for SoftMimic vs. baseline and the convergence of compliance objectives. It takes a bit more time to train SoftMimic to convergence compared to stiff motion tracking. The policy must learn not only to track a single motion but also to embed a wide range of compliant responses.

TABLE I: **Motion tracking quality comparison.** Comparison of tracking error under no-perturbation conditions (free space) for our compliant policy and a stiff baseline on various skills. Errors are reported as mean joint position error (degrees) and keypoint Cartesian error (cm), with standard error of the mean over 36 episodes.

Skill	Ours (Compliant)		Stiff Baseline	
	Joint (°)	Keypoint (cm)	Joint (°)	Keypoint (cm)
Box Pick	5.04 ± 0.01	2.65 ± 0.01	2.04 ± 0.00	1.36 ± 0.00
Walk	6.39 ± 0.00	3.44 ± 0.00	6.09 ± 0.00	3.50 ± 0.00
Dance	11.10 ± 0.01	6.05 ± 0.01	5.16 ± 0.01	3.01 ± 0.00

V. Conclusion

This work introduced a formulation for learning compliant whole-body motion tracking for humanoid robots. We demonstrate that our compliant policy outperforms the standard motion tracking baseline in generalization to unseen manipulation scenarios and in safety when handling disturbances, applying nearly half the force of the baseline during collisions while preserving comparable tracking performance under unperturbed conditions. Qualitatively, user-commanded stiffness values effectively modulate the robot's interactions; quantitatively, the policy shows good stiffness adherence across a wide range of values and a broad workspace. Future work might adjust stiffness based on context, improve the data augmentation pipeline by incorporating dynamics or a data-driven metric, and extend the method to handle multiple simultaneous forces applied to any link.

VI. Acknowledgment

We thank the members of the Improbable AI lab for helpful discussions and feedback. We acknowledge Unitree Robotics for hardware support provided for their robots. We are grateful to MIT Supercloud and the Lincoln Laboratory Supercomputing Center for providing HPC resources. This research was financially supported by the Ministry of Trade, Industry, and Energy (MOTIE), Korea, under the "Global Industrial Technology Cooperation Center program" supervised by the Korea Institute for Advancement of Technology (KIAT). (Grant No. P0028435)

REFERENCES

[1] 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. Graph.*, vol. 37, no. 4, pp. 1–14, 2018.

[2] 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 preprint arXiv:2506.12851*, 2025.

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

[4] 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 *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, 2024, pp. 8944–8951.

[5] 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.

[6] 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 preprint arXiv:2508.00225*, 2025.

[7] L. Sentis and O. Khatib, "Task-oriented control of humanoid robots through prioritization," in *Proc. IEEE Int. Conf. Humanoid Robots*, 2004, pp. 1–16.

[8] G. B. Margolis, G. Yang, K. Paigwar, T. Chen, and P. Agrawal, "Rapid locomotion via reinforcement learning," *Robot.: Sci. Syst.*, 2022.

[9] V. Makoviychuk, L. Wawrzyniak, Y. Guo, M. Lu, K. Storey, M. Macklin, D. Hoeller, N. Rudin, A. Allshire, A. Handa, and G. State, "Isaac Gym: High performance GPU-based physics simulation for robot learning," *arXiv preprint arXiv:2108.10470*, 2021.

[10] E. Todorov, T. Erez, and Y. Tassa, "MuJoCo: A physics engine for model-based control," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, 2012, pp. 5026–5033.

[11] K. Zakka, B. Tabanpour, Q. Liao, M. Haiderbhai, S. Holt, J. Y. Luo, A. Allshire, E. Frey, K. Sreenath, L. A. Kahrs *et al.*, "MuJoCo playground," *arXiv preprint arXiv:2502.08844*, 2025.

[12] X. B. Peng, M. Andrychowicz, W. Zaremba, and P. Abbeel, "Sim-to-real transfer of robotic control with dynamics randomization," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, 2018, pp. 3803–3810.

[13] J. Hwangbo, J. Lee, A. Dosovitskiy, D. Bellicoso, V. Tsounis, V. Koltun, and M. Hutter, "Learning agile and dynamic motor skills for legged robots," *Sci. Robot.*, vol. 4, no. 26, p. eaau5872, 2019.

[14] G. B. Margolis and P. Agrawal, "Walk these ways: Tuning robot control for generalization with multiplicity of behavior," in *Proc. Conf. Robot Learn. (CoRL)*. PMLR, 2023, pp. 22–31.

[15] M. Hutter, C. Gehring, D. Jud, A. Lauber, C. D. Bellicoso, V. Tsounis, J. Hwangbo, K. Bodie, P. Fankhauser, M. Bloesch *et al.*, "ANYmal—a highly mobile and dynamic quadrupedal robot," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, 2016, pp. 38–44.

[16] P. M. Wensing, A. Wang, S. Seok, D. Otten, J. Lang, and S. Kim, "Proprioceptive actuator design in the MIT cheetah: Impact mitigation and high-bandwidth physical interaction for dynamic legged robots," *IEEE Trans. Robot.*, vol. 33, no. 3, pp. 509–522, 2017.

[17] B. Katz, J. Di Carlo, and S. Kim, "Mini Cheetah: A platform for pushing the limits of dynamic quadruped control," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, 2019, pp. 6295–6301.

[18] Unitree Robotics, "G1," https://www.unitree.com/g1, 2025, accessed: Sep. 2025.

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

[20] 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," *arXiv preprint arXiv:2505.03738*, 2025.

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

[22] Z. Fu, Q. Zhao, Q. Wu, G. Wetzstein, and C. Finn, "HumanPlus: Humanoid shadowing and imitation from humans," *arXiv preprint arXiv:2406.10454*, 2024.

[23] Q. Ben, F. Jia, J. Zeng, J. Dong, D. Lin, and J. Pang, "HOMIE: Humanoid loco-manipulation with isomorphic exoskeleton cockpit," *arXiv preprint arXiv:2502.13013*, 2025.

[24] Z. Fu, T. Z. Zhao, and C. Finn, "Mobile ALOHA: Learning bimanual mobile manipulation with low-cost whole-body teleoperation," *arXiv preprint arXiv:2401.02117*, 2024.

[25] H.-S. Fang, B. Romero, Y. Xie, A. Hu, B.-R. Huang, J. Alvarez, M. Kim, G. Margolis, K. Anbarasu, M. Tomizuka, E. Adelson, and P. Agrawal, "DEXOP: A device for robotic transfer of dexterous human manipulation," *arXiv preprint arXiv:2509.04441*, 2025.

[26] J. Eßer, G. B. Margolis, O. Urbann, S. Kerner, and P. Agrawal, "Action space design in reinforcement learning for robot motor skills," in *Proc. Conf. Robot Learn. (CoRL)*, 2024.

[27] Z. Fu, X. Cheng, and D. Pathak, "Deep whole-body control: Learning a unified policy for manipulation and locomotion," in *Proc. Conf. Robot Learn. (CoRL)*. PMLR, 2023, pp. 138–149.

[28] Y. Zhang, Y. Yuan, P. Gurunath, T. He, S. Omidshafiei, A.-a. Aghamohammadi, M. Vazquez-Chanlatte, L. Pedersen, and G. Shi, "FALCON: Learning force-adaptive humanoid loco-manipulation," *arXiv preprint arXiv:2505.06776*, 2025.

[29] M. H. Raibert and J. J. Craig, "Hybrid position/force control of manipulators," *ASME J. Dyn. Syst., Meas., Control*, 1981.

[30] N. Hogan, "Impedance control: An approach to manipulation: Part II—implementation," *ASME J. Dyn. Syst., Meas., Control*, 1985.

[31] O. Khatib, "A unified approach for motion and force control of robot manipulators: The operational space formulation," *IEEE J. Robot. Autom.*, vol. 3, no. 1, pp. 43–53, 1987.

[32] L. Sentis and O. Khatib, "Synthesis of whole-body behaviors through hierarchical control of behavioral primitives," *Int. J. Humanoid Robot.*, vol. 2, no. 04, pp. 505–518, 2005.

[33] ——, "A whole-body control framework for humanoids operating in human environments," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, 2006, pp. 2641–2648.

[34] L. Sentis, J. Park, and O. Khatib, "Compliant control of multicontact and center-of-mass behaviors in humanoid robots," *IEEE Trans. Robot.*, vol. 26, no. 3, pp. 483–501, 2010.

[35] A. Albu-Schäffer, C. Ott, and G. Hirzinger, "A unified passivity-based control framework for position, torque and impedance control of flexible joint robots," *Int. J. Robot. Res.*, vol. 26, no. 1, pp. 23–39, 2007.

[36] S.-H. Hyon, J. G. Hale, and G. Cheng, "Full-body compliant human–humanoid interaction: Balancing in the presence of unknown external forces," *IEEE Trans. Robot.*, vol. 23, no. 5, pp. 884–898, 2007.

[37] C. Ott, O. Eiberger, W. Friedl, B. Bauml, U. Hillenbrand, C. Borst, A. Albu-Schaffer, B. Brunner, H. Hirschmuller, S. Kielhofer *et al.*, "A humanoid two-arm system for dexterous manipulation," in *Proc. IEEE-RAS Int. Conf. Humanoid Robots*, 2006, pp. 276–283.

[38] C. Ott, B. Henze, and D. Lee, "Kinesthetic teaching of humanoid motion based on whole-body compliance control with interaction-aware balancing," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, 2013, pp. 4615–4621.

[39] B. Henze, A. Dietrich, and C. Ott, "An approach to combine balancing with hierarchical whole-body control for legged humanoid robots," *IEEE Robot. Autom. Lett.*, vol. 1, no. 2, pp. 700–707, 2015.

[40] J. M. Garcia-Haro, B. Henze, G. Mesesan, S. Martinez, and C. Ott, "Integration of dual-arm manipulation in a passivity based whole-body controller for torque-controlled humanoid robots," in *Proc. IEEE-RAS Int. Conf. Humanoid Robots*, 2019, pp. 644–650.

[41] F. Abi-Farraj, B. Henze, C. Ott, P. R. Giordano, and M. A. Roa, "Torque-based balancing for a humanoid robot performing high-force interaction tasks," *IEEE Robot. Autom. Lett.*, vol. 4, no. 2, pp. 2023–2030, 2019.

[42] J. Englsberger, A. Werner, C. Ott, B. Henze, M. A. Roa, G. Garofalo, R. Burger, A. Beyer, O. Eiberger, K. Schmid, and A. Albu-Schäffer, "Overview of the torque-controlled humanoid robot TORO," in *Proc. IEEE Int. Conf. Humanoid Robots*, 2014, pp. 916–923.

[43] S. Lee, P. S. Chang, and J. Lee, "Deep compliant control," in *Proc. ACM SIGGRAPH Conf.*, 2022, pp. 1–9.

[44] T. Portela, G. B. Margolis, Y. Ji, and P. Agrawal, "Learning force control for legged manipulation," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, 2024, pp. 15 366–15 372.

[45] A. Hartmann, D. Kang, F. Zargarbashi, M. Zamora, and S. Coros, "Deep compliant control for legged robots," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, 2024, pp. 11 421–11 427.

[46] B. Xu, H. Weng, Q. Lu, Y. Gao, and H. Xu, "FACET: Force-adaptive control via impedance reference tracking for legged robots," *arXiv preprint arXiv:2505.06883*, 2025.

[47] P. Zhi, P. Li, J. Yin, B. Jia, and S. Huang, "Learning unified force and position control for legged loco-manipulation," *arXiv preprint arXiv:2505.20829*, 2025.

[48] C. Ott, R. Mukherjee, and Y. Nakamura, "Unified impedance and admittance control," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, 2010, pp. 554–561.

[49] N. Mahmood, N. Ghorbani, N. F. Troje, G. Pons-Moll, and M. J. Black, "AMASS: Archive of motion capture as surface shapes," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, 2019, pp. 5442–5451.

[50] F. G. Harvey, M. Yurick, D. Nowrouzezahrai, and C. Pal, "Robust motion in-betweening," *ACM Trans. Graph.*, vol. 39, no. 4, pp. 60:1–60:12, 2020.

[51] K. Zakka, "Mink: Python inverse kinematics based on MuJoCo," https://github.com/kevinzakka/mink, 2024.

[52] J. Carpentier, G. Saurel, G. Buondonno, J. Mirabel, F. Lamiraux, O. Stasse, and N. Mansard, "The Pinocchio C++ library—a fast and flexible implementation of rigid body dynamics algorithms and their analytical derivatives," in *Proc. IEEE Int. Symp. Syst. Integr. (SII)*, 2019.

[53] C. Schwarke, M. Mittal, N. Rudin, D. Hoeller, and M. Hutter, "rsl_rl: A learning library for robotics research," *arXiv preprint arXiv:2509.10771*, 2025.