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

CoTaP: Compliant Task Pipeline and Reinforcement Learning of Its Controller with Compliance Modulation

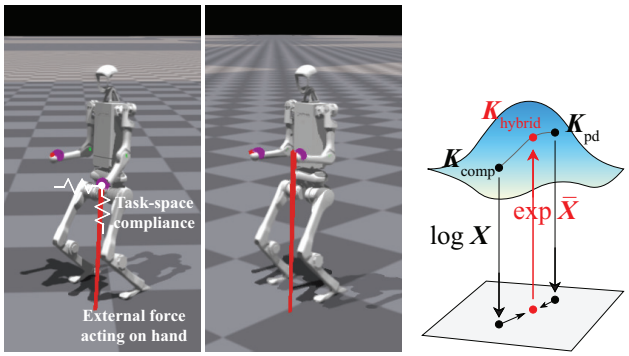
Zewen He, Chenyuan Chen, Dilshod Azizov, Yoshihiko Nakamura

***Abstract*— Humanoid whole-body locomotion control is a critical approach for humanoid robots to leverage their inherent advantages. Learning-based control methods derived from retargeted human motion data provide an effective means of addressing this issue. However, because most current human datasets lack measured force data, and learning-based robot control is largely position-based, achieving appropriate compliance during interaction with real environments remains challenging. This paper presents Compliant Task Pipeline (CoTaP): a pipeline that leverages compliance information in the learning-based structure of humanoid robots. A two-stage dual-agent reinforcement learning framework combined with model-based compliance control for humanoid robots is proposed. In the training process, first a base policy with a position-based controller is trained; then in the distillation, the upper-body policy is combined with model-based compliance control, and the lower-body agent is guided by the base policy. In the upper-body control, adjustable task-space compliance can be specified and integrated with other controllers through compliance modulation on the symmetric positive definite (SPD) manifold, ensuring system stability. We validated the feasibility of the proposed strategy in simulation and experiment, primarily comparing the responses to external disturbances under different compliance settings.**

I. Introduction

In recent decades, humanoid robot technology has made significant advancements. Particularly over the past five years, there has been a surge in the development of diverse humanoid robot body designs, including Atlas from Boston Dynamics, Optimus from Tesla, Figure's humanoid robots, and Unitree's humanoid robots like H1 and G1. Meanwhile, with the rapid progress in the field of artificial intelligence, reinforcement learning (RL) and imitation learning (IL) have been increasingly applied to the control of humanoid robots, leading to significant breakthroughs. In the first step, the imitation motion controller was applied in the simulation for humanoid character control, such as in [1], [2]. After that, this approach was extended into real humanoid robot whoel-body control (WBC) [3], [4]. Based on the learning method, the controller no longer requires accurate modeling of the robot and environment such as model predictive control (MPC), which demonstrates improved robustness and generalizability in complex environments.

Humanoid robots have the key features that can simultaneously perform locomotion and manipulation, which is abbreviated as *loco-manipulation* [5]. In many studies, the upper and lower bodies of humanoid robots are controlled independently. During manipulation tasks, the legs are typically kept stationary to maintain balance and ensure stability of the center of mass (CoM) [6]. In recent studies, IL based on human motion data has also been applied to loco-manipulation. The data sources include publicly available human motion datasets as well as data collected through teleoperation [7], [8].



(a) (b) (c)

Fig. 1. (a) Simulation of H1 under a vertical load in the low-stiffness condition. (b) Simulation of H1 under a vertical load in the high-stiffness condition. (c) Illustration of the stiffness matrices modulation on SPD manifold. Two different original stiffness matrices are first mapped to the Log-Euclidean space using the log mapping, then linearly interpolated, and finally mapped back using the exp mapping.

There are several key aspects in the current research receiving significant attention. First, most current research focuses on joint space PD control for humanoid motion control. This control approach may yield satisfactory results in current purely motion control scenarios; however, it fails to implement force control when interactions with the environment (such as manipulation and multi-contact motion) or even human-robot interactions (HRI) occur, thereby making it difficult to achieve desirable outcomes. Then, most human even humanoid robot data are only proprioception-based, which lacks sensory input and action output. Although some studies have already attempted to incorporate contact force data into human motion data collection [9], [10], the overall size and generality of such datasets remain insufficient. In addition, the commonly used physical simulators make sim-to-real transfer more difficult, as they lack accuracy and are computationally expensive when simulating complex contacts. Therefore, a key challenge lies in how to leverage the currently limited data resources to achieve force control for humanoid robot loco-manipulation.

To address these challenges in a comprehensive manner,

The authors are with the Department of Robotics, Mohamed bin Zayed University of Artificial Intelligence, Masdar City, Abu Dhabi, United Arab Emirates. {zewen.he, yoshihiko.nakamura}@mbzuai.ac.ae

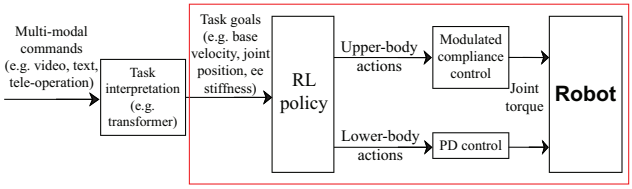


Fig. 2. Overview of CoTaP pipeline. The red frame highlights the method proposed in this paper, and our objective is to implement the entire pipeline on humanoid robots.

we turn to a classical topic in traditional robotics: compliance control. Compliance control, by adopting a relatively passive mechanism, is capable of adapting to unknown or inaccurate contacts. Such a property is exactly what is required to overcome the training challenges arising from the lack of sufficient contact information; moreover, compliance control inherently provides force control capabilities, which makes it particularly suitable for addressing the challenges of real-world robot–environment interactions in loco-manipulation tasks. In model-based robot control theory, compliance control is a practical method to guarantee robot interaction safety and stability with the environment. [11], [12] proposed the hierarchical compliance control method for humanoid robot. In the study [13], [14], the authors optimized the joint-space viscoelasticity matrices by an analytical way. This work has enabled compliance control to achieve promising results in maintaining balance in humanoid robots. In the field of robotic manipulation, compliance control is even more critical. It plays a key role in ensuring the robot's safety, enhancing HRI, and enabling adaptive manipulation capabilities [15], [16].

Compared with model-based control, the greatest advantage of RL lies in its robustness and generalization in complex and dynamic tasks. On this basis, RL method has also been integrated into compliance control. In [17], the controller achieves compliant behavior by modulating the target position, which is fundamentally akin to admittance control rather than direct force control. In other related studies [18], [19], [20], the RL-based controller typically focuses on optimizing the PD gains of individual joints or a single limb, without considering whole-body compliance. Recently, some methods based on model-free RL for compliance control and even force control have been proposed [21], [22], [23], but their accuracy has not been guaranteed.

Accordingly, the immediate task is to establish how parameter adjustability and stability of compliance control can be ensured within the RL framework, and to further verify the compliance effect under external perturbations. Therefore, we propose **Co**mpliant **Ta**sk **P**ipeline (CoTaP), a pipeline leveraging the compliance information in the humanoid robot loco-manipulation control. As illustrated in Fig. 2, in this paper we mainly focus on the compliance modulation on the RL framework of the pipeline. In this study, we present a compliance control approach that integrates the RL control framework with the robot's kinematic model. By performing stiffness matrix modulation on the symmetric positive definite (SPD) manifold, the stability of the joint-space control is guaranteed. For humanoid whole-body control, a two-stage dual-agent policy training framework is applied in our work.

The main contributions of this study are as follows:

1. Combined model-based compliance control with humanoid robot reinforcement learning control framework, designing a learning-based compliance control strategy including dual-agent policy and compliance modulation on SPD manifold for upper-body control;
2. Validated effectiveness of the proposed compliance control, achieving compliance modulation performance of a humanoid robot in simulation and experiment.

II. Related Works

A. Reinforcement Learning Based Humanoid Locomotion Control

In recent years, there has been a surge of research on humanoid robot control that leverages human motion data retargeting in combination with deep reinforcement learning and imitation learning. The main prevailing trends include the use of adversarial learning such as GAIL [24] and AMP [2], or directly applying imitation learning to the reference motion trajectories [1], [7], [25]. Moreover, such imitation learning methods have been extended to enable whole-body tele-operation in humanoid robots [7], [26]. In addition, for humanoid robot loco-manipulation, the dual-agent system with separated upper and lower body control has also achieved good results in researches [22], [27].

B. Humanoid Whole-body Compliance Control

According to the model-based compliance control method, there is a feedback controller in the task space:

$$\boldsymbol{f} = \boldsymbol{K}\Delta\boldsymbol{p} + \boldsymbol{D}\Delta\boldsymbol{v} \tag{1}$$

Correspondingly, the joint-space feedback controller can be expressed as follows:

$$\boldsymbol{\tau} = \boldsymbol{\tau}_{grav} + \boldsymbol{K}_\theta\Delta\boldsymbol{\theta} + \boldsymbol{D}_\theta\Delta\dot{\boldsymbol{\theta}} \tag{2}$$

where $\boldsymbol{\tau}_{grav}$ is the gravity compensation term. Under this setting, especially for redundant robots, a key research challenge in compliance control lies in establishing the mapping between task space and joint space while simultaneously considering null-space control. Studies such as [28], [13], [14] have addressed this issue from both kinematic and dynamic perspectives. It should also be noted that compliance and stiffness form a dual relationship; therefore, in this paper, we use the two terms interchangeably in derivations depending on the context.

C. Comparison with Related Works

Drawing on existing studies, we provide a comparison in Table I with three dimensions: adjustable, model-aware, and stability-accounted. Adjustable denotes the capability to adapt task parameters online without the need for retraining; model-aware denotes whether control explicitly incorporates

TABLE I

COMPARISON BETWEEN DIFFERENT LEARNING-BASED COMPLIANCE CONTROL METHOD

Method	Adjustable	Model-aware[2]	Stability-accounted
DCC [17]	No	Yes	No
FALCON [1]	No	No	No
FACET	Yes	No	No
HMC	Yes	Yes	No
Ours	Yes	Yes	Yes

[1] FALCON achieves adaptive force control, but can be considered as passive compliance.

[2] Model means explicit robot kinematic or dynamic model in controller.

the robot's kinematic or dynamic model; and stability-accounted refers to whether the stability of the control method is taken into account or formally proven.

As shown in the table, the approach proposed in this work fulfills all three criteria. We provide a detailed comparison here with several studies that share objectives closely aligned with ours. First, FACET [21] achieves force (also impedance) control by using a simplified task-space model together with model-free RL-based tracking control, which is similar to an admittance control approach. However, unlike traditional robotic manipulators, current quadruped and humanoid robots find it difficult to achieve precise position tracking in the world frame using RL with low-level motor control, and consequently, the final force control objective is also affected. On the other hand, HMC [23] performs weighted linear interpolation of different model-based control laws at the torque level, while the control objectives and inputs of each controller are also different. However, due to the heterogeneity of these meta-controllers, it is hard to assess the stability of the control during blending at torque level, making it difficult to guarantee that the controller will not diverge at any given moment. Besides, the null-space stiffness in its impedance control was not considered.

To address the limitations of the above-mentioned studies, this paper proposes a method that integrates model-based control within the RL framework to achieve compliance control with guaranteed stability.

III. TASK-SPACE COMPLIANCE CONTROL IN REINFORCEMENT LEARNING STRUCTURE

A. Dual-agent Learning Strategy

In this paper, we take [22] as the primary baseline, and the basic learning network settings are largely based on this study. Similar as [22], [27], [23], we divide the whole body into two parts: upper- and lower-body. The torso is defined as the base link, which simultaneously functions as the separation link between the upper and lower body. Each part shares the same observation but have separate policies and actions. Fig. 3 shows the overview of the training framework of our work. The actions are the whole-body target joint angle $\boldsymbol{a}_t = \boldsymbol{q}^*$ ($*$ means reference). The robot state is defined as $\boldsymbol{s}_t := [\boldsymbol{q}_{t-4:t}, \dot{\boldsymbol{q}}_{t-4:t}, \boldsymbol{\omega}^{\text{torso}}_{t-4:t}, \boldsymbol{g}_{t-4:t}, \boldsymbol{\tau}^{\text{upper}}_{t-4:t}, \boldsymbol{a}_{t-5:t-1}]$, which contains five-step histories of joint positions, joint velocities, root angular velocity, projected gravity, upper-joint control torques and previous actions. The goal space $\mathcal{G}_t = [\mathcal{G}^l_t, \mathcal{G}^u_t]$ consists of locomotion goals $\mathcal{G}^l_t := [\boldsymbol{v}^{\text{torso}*}_t, h^{\text{torso}*}_t, w^{\text{yaw}*}_t]$, specifying desired torso linear velocities, torso heights, and torso yaw angles, and manipulation goals $\mathcal{G}^u_t := [\boldsymbol{q}^{\text{upper}*}_t, \boldsymbol{K}^{ee*}_t, k^{\text{null}*}_t, \alpha]$, specifying target joint configurations for the upper body, target task- and null-space stiffness of end-effector (ee means end-effector), and compliance modulation ratio.

In practical manipulation, the hands act as the task end-effector, while the lower-body mainly maintains balance, so we focus on upper-body compliance. Additionally, bipedal contact states change frequently, and including the lower-body in hand task-space compliance can make the solved joint compliance discontinuous. Therefore, we decouple the upper and lower-body, define upper-body compliance in a torso-fixed frame, and use PD control for the lower-body.

B. Decoupled Upper-body Compliance

For upper-body compliance control, we consider the arms are based on the torso link, therefore the upper-body compliance is directly related to the torso-link compliance. The joint-space stiffness matrix including upper and torso can be expressed as block matrices:

$$\boldsymbol{K}_q := \begin{bmatrix} \boldsymbol{K}_{torso} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{K}_u \end{bmatrix} \tag{3}$$

According to virtual work principle in quasi-static assumption, we can derive the compliance relationship between joint and task space (end-effector in the world frame):

$$\boldsymbol{C}_e = \boldsymbol{J}_e \boldsymbol{K}_q^{-1} \boldsymbol{J}_e^\top \tag{4}$$

where $\boldsymbol{J}_e$ is the Jacobian matrix of the end-effector velocity. Applying the block division on the Jacobian matrix $\boldsymbol{J}_e = \begin{bmatrix} \boldsymbol{J}_{eb} & \boldsymbol{J}_{eu} \end{bmatrix}$, we have

$$\boldsymbol{C}_e = \boldsymbol{J}_{eb} \boldsymbol{K}^{-1}_{torso} \boldsymbol{J}^\top_{eb} + \boldsymbol{J}_{eu} \boldsymbol{K}^{-1}_u \boldsymbol{J}^\top_{eu} \tag{5}$$

According to [13], the solution of upper-body joint compliance matrix is

$$\boldsymbol{K}^{-1}_u = \boldsymbol{J}^\sharp_{eu} \widehat{\boldsymbol{C}}_e \boldsymbol{J}^{\sharp\top}_{eu} + \boldsymbol{Y} - \boldsymbol{J}^\sharp_{eu} \boldsymbol{J}_{eu} \boldsymbol{Y} \boldsymbol{J}^\top_{eu} \boldsymbol{J}^{\sharp\top}_{eu} \tag{6}$$

where $\widehat{\boldsymbol{C}}_e := \boldsymbol{C}_e - \boldsymbol{J}_{eb} \boldsymbol{K}^{-1}_{torso} \boldsymbol{J}^\top_{eb}$, null-space compliance $\boldsymbol{Y} := 1/k^{\text{null}} \boldsymbol{I}$.

For the torso-link stiffness, due to the lower-body is totally PD-based, we can use kinematic relationship as in (4) to calculate. In this case, it is necessary to distinguish the different support case of the humanoid, such as single-support and double-support (and in rare cases, flight case). Nonetheless, for controller simplification, the torso stiffness can be treated as constant. Upper-body gravity compensation is also considered in our compliance control, which is expressed as $\boldsymbol{\tau}^{\text{upper}}_{grav}$.

C. Compliance Modulation on Symmetric Positive Definite (SPD) Manifold

For different control task, different compliance performance is required. In this study, we set resolved compliance and simple PD control as example, using a ratio variable α

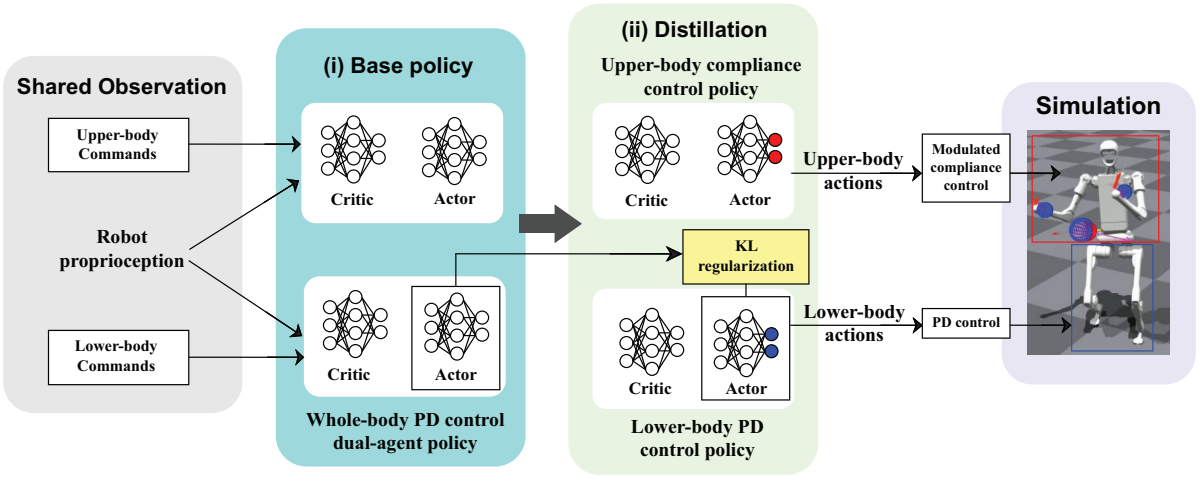


Fig. 3. Overview of the training framework in this study.

to balance the two controllers. Rather than simply applying scaled summation, we adopt proportional combination of the two (or more) stiffness matrices on the SPD manifold, which also named as Log-Euclidean Interpolation [29]:

$$\boldsymbol{K}^{\text{upper}}_{\text{modulated}} = \exp\left(\alpha \log \boldsymbol{K}^{\text{upper}}_{\text{comp}} + (1-\alpha)\log \boldsymbol{K}^{\text{upper}}_{\text{pd}}\right) \quad (7)$$

where $\boldsymbol{K}^{\text{upper}}_{\text{comp}}$ is the compliance control stiffness matrix $\boldsymbol{K}_u$ obtained in Section III-B; $\boldsymbol{K}^{\text{upper}}_{\text{pd}}$ is the original PD control stiffness, and α is the ratio to modulate the stiffness between compliance control and PD control. Based on this approach, the obtained joint stiffness after modulation is always positive definite. According to the proof in [14], the stability of the system at the current time can be guaranteed in the case of positive definite stiffness. Fig. 1(c) illustrates the mapping and modulation process of the stiffness matrices. In contrast to linear interpolation performed in Euclidean space, Log-Euclidean interpolation offers the following advantage [29]: it guarantees that the interpolated stiffness matrix lies on the SPD manifold, thus avoiding outcomes that violate the underlying manifold geometry; it avoids the swelling effect that may occur in linear interpolation, which can lead to physically unreasonable results. The above advantages are not discussed extensively in this paper; rather, we treat this method merely as a more rational approach to modulating compliance.

In our strategy, the ratio α can be set as command but also depends on the kinematic posture of the robot. To avoid solution problem at the near-singularity posture of the arms, we process the original α taking into account the condition number of task Jacobian matrix as follows:

$$\widehat{\alpha} = \frac{\alpha}{1 + \max(0, \texttt{cond\_num} - 10)} \quad (8)$$

where `cond_num` is the condition number of the upper body Jacobian matrix, and $\widehat{\alpha}$ is the processed modulation ratio. In the control based on (7), we replace α by $\widehat{\alpha}$ taking into account of the condition number. This method achieves an effect comparable to SR-inverse, while avoiding the need for extra handling during the computation of the pseudo-inverse matrix.

In the training of policy, we applied domain randomization on the original ratio. Therefore, this value can be flexibly modulated to balance the contribution of the two (or more) control laws in joint-space stiffness. Moreover, inspired by [23], it may also be defined as the output of the high-level controller, enabling adaptation to specific task requirements.

D. Two-stage Policy Distillation

In the training we apply a two-stage policy distillation. As shown in Fig. 3, (i) first, we train a base policy as whole-body PD position controller with two agents $\pi^{\text{lower}}_{base}$ and $\pi^{\text{upper}}_{base}$ to satisfy lower-body commands tracking and upper-body motion imitation task. (ii) Then, we use the base policy to guide the training of a new policy with upper body compliance control, which corresponds to policy distillation. Specifically, we impose a KL regularization on the lower-body policy as in (10), and apply it in the actor loss calculation (9):

$$\mathcal{L}_{\text{distill}} = \mathcal{L}_{\text{PPO}} + \mathcal{L}^{\text{lower}}_{\text{KL}}, \quad (9)$$

$$\mathcal{L}^{\text{lower}}_{\text{KL}} = \beta_{KL}\, D_{\text{KL}}\left(\pi^{\text{lower}}_{distill}(\cdot \mid \boldsymbol{s}_t) \,\|\, \pi^{\text{lower}}_{base}(\cdot \mid \boldsymbol{s}_t)\right), \quad (10)$$

where β_{KL} is a weighting coefficient decreasing with time.

In both training stages, PPO [30] is applied to maximize the cumulative rewards. In the reward setting, we refer to FALCON work [22]. But beyond the basic reward terms, we add an extra reward term for reducing the gap between the policy's output target joint angles and the original reference trajectory joint angles: $\exp(-\sigma_{\text{ref}}\|\pi^{\text{upper}}_{fine}(\boldsymbol{s}_t) - \boldsymbol{q}^{\text{upper}*}_t\|_2)$.

Since the new control goals are included, the domain randomization extends the following terms: end-effector stiffness matrix, null-space stiffness matrix, modulation ratio α. The arrangement of the domain randomization is shown in Table II.

IV. Results

In this section, we describe the setup for training RL policies on the Unitree H1 humanoid robot in Isaac Gym

TABLE II

THE RANGE OF RANDOMIZATION ADDED ON DEFAULT VALUES

Term	Value
Friction coefficient	$\mathcal{U}(0.5,\ 1.25)$
Link mass	$\mathcal{U}(0.9,\ 1.2) \times$ default [kg]
Base mass	$\mathcal{U}(-1.0,\ 3.0)$ [kg]
Control delay	$\mathcal{U}(0,\ 20)$ [ms]
P Gains (base)	$\mathcal{U}(0.9,\ 1.1) \times$ default [Nm/rad]
D Gains (base)	$\mathcal{U}(0.9,\ 1.1) \times$ default [Nms/rad]
K^{ee} Gains	$\mathcal{U}(0.5,\ 1.5) \times 300.0$ [N/m]
k_{null} Value	$\mathcal{U}(0.6,\ 1.4) \times 40.0$ [Nm/rad]
α Value	$\mathcal{U}(0,\ 1)$

environment. For upper-body motion priors, we employ the AMASS dataset [31] filtered by Perpetual Humanoid Control (PHC) method [32]. These processed datasets serve as demonstrations and references for initializing and guiding the policy. The basic network architecture and hyperparameter settings are kept consistent with those of FALCON [22], ensuring comparability and leveraging prior work on scalable humanoid locomotion control. All experiments are conducted on a workstation running Ubuntu 22.04, equipped with an Intel Core i9-14900K CPU and an NVIDIA RTX 4080 GPU.

In the specific configuration, torso link is adopted as the dividing boundary between the upper and lower body. 3-dimension position is set as the task space for both hands. In current simplified situation, we ignore the effect of lower-body configuration on torso stiffness, therefore $\boldsymbol{K}_{torso}^{-1}$ is considered as constant. The upper-body compliance controller calculates the required torque and sends it directly to the joint actuator; while the lower-body policy directly sends the target joint position to the actuator. The frequency of this step is 50 Hz. For a feasible policy training, it requires about 20,000 iterations.

In the following simulation results, different stiffness matrices were applied in the task-space, while null-space k_{null} is set as 25. For lower-body, all joint PD used the default values provided by Unitree (all upper-body joint K_p = 100). As mentioned above, our method is built on the FALCON framework; therefore, we adopt FALCON method as the baseline, namely pure joint-level PD control. It should be noted that, because our goal is to achieve compliant control within an RL framework rather than to realize compliance solely as a specific control task, all ablation studies are conducted within the RL setting; we therefore do not include comparisons against purely model-based compliant control.

A. Evaluation Criterion

For a numerical comparison of the baseline and the proposed CoTaP, the following evaluation criteria are established:

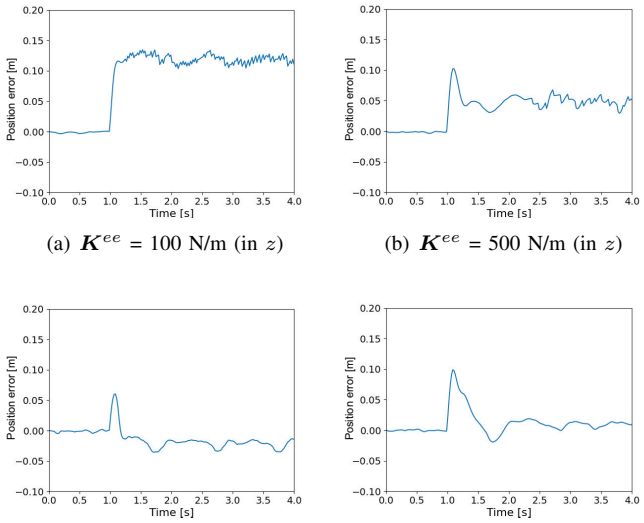
1) Torso velocity tracking error:
$e^{\text{torso}} = \frac{1}{T}\sum_{t=0}^{T}\left\|\boldsymbol{v}_t^{\text{torso}*} - \boldsymbol{v}_t^{\text{torso}}\right\|_2$
2) End-effector tracking error:
$e^{\text{ee}} = \frac{1}{T}\sum_{t=0}^{T}\left\|\boldsymbol{p}_t^{\text{ee}*} - \boldsymbol{p}_t^{\text{ee}}\right\|_2$
3) Average of upper-body joint torques:
$J^{\text{upper}} = \frac{1}{T}\sum_{t=0}^{T}\left\|\boldsymbol{\tau}_t^{\text{upper}}\right\|_2$
4) Upper-body tracking error:
$e^{\text{upper}} = \frac{1}{T}\sum_{t=0}^{T}\left\|\boldsymbol{q}_t^{\text{upper}*} - \boldsymbol{q}_t^{\text{upper}}\right\|_2$

B. Simulation Results

1) Constant Load Test on End-effector: As in FALCON, an external command is employed in our controller to configure the stance and stepping (walking) modes. In the standing mode, the robot maintains its lower body stationary on the ground through double-support, while the arms are set to maintain their default L-shape. At this point, a constant vertical downward external force (-50 N in z-axis) is applied to the robot's left hand to simulate the action of lifting a heavy object. Fig. 1(a) and Fig. 1(b) are the robot in low ($\boldsymbol{K}^{ee}$ = 100 N/m) and high ($\boldsymbol{K}^{ee}$ = 1000 N/m) task-space stiffness settings, respectively. We can see the displacement of left hand are different in two cases. Fig. 4 shows the results of hand position error in different stiffness settings of the proposed method. From these figures, we can observe that under small hand stiffness (i.e., high compliance, such as $\boldsymbol{K}^{ee}$ = 100 N/m), the hand error is relatively large and sustained (about 0.12 m); whereas under high stiffness ($\boldsymbol{K}^{ee}$ = 500 N/m), the hand error exhibits small oscillations and quickly decays to a constant value (about 0.05 m, respectively). Furthermore, as stiffness increases, the residual error at steady state decreases. This is consistent with the ideal task-space control objective established beforehand. Nevertheless, the error arises between the displacement values under different stiffness settings and the expected outcomes. This can primarily be attributed to two factors: (i) large displacements are restricted by the structural limits of the arm, and (ii) the RL action inherently compensates for hand errors to some extent. In addition, we compared the hand error when applying PD and setting α = 0.5, $\boldsymbol{K}^{ee}$ = 100 N/m. The results indicate that while PD control achieves a lower peak error, it subsequently exhibits reverse errors (about -0.03 m), whereas the compliance modulation method provides a favorable oscillation-damping effect. In other words, due to the modulation of stiffness, it combines the characteristics of both approaches and improves performance.

2) Impact Force at End-effector in Stance Mode: In the stance mode, we added an external impact (500 N in 0.05 s, $+x$) on H1 left hand in different compliance cases. After the impact, the robot's left arm was driven into large swings, while the body developed velocity errors. For the experimental evaluation, we measured several error metrics in Section IV-A across 4096 environments with randomly sampled reference motions. The results are presented in Table III(a). For the proposed compliance controller (central 3 rows), task-space $\boldsymbol{K}^{ee}$ in x, y axis is set as in the table, while in z is 300 N/m. Modulation ratio α modulates the compliance between PD control and compliance control in the case of $\boldsymbol{K}^{ee}$ = 100 N/m.

This table allows us to derive a wealth of information as follows. First, when comparing the three central rows corresponding to the proposed method under varying task stiffness, it can be observed that increasing stiffness leads to smaller tracking errors in the torso and hand, while



(a) $\boldsymbol{K}^{ee}$ = 100 N/m (in z) (b) $\boldsymbol{K}^{ee}$ = 500 N/m (in z)

(c) PD (d) α = 0.5 (PD and 100 N/m)

Fig. 4. Left hand position error in z-axis under a constant -50 N payload (applied after 1.0 s). The values of stiffness are in z-axis, in x and y we set as 300 N/m. The unit of position error: [m].

TABLE III
IMPACT TEST RESULT

(a) Stance Mode

Method	e^{torso}	e^{ee}	J^{upper}
PD (FALCON)	$0.701_{\pm 0.221}$	$0.092_{\pm 0.055}$	49.723
Soft P (K_p = 30)	$0.670_{\pm 0.192}$	$0.110_{\pm 0.055}$	43.627
$\boldsymbol{K}^{ee}$=100 N/m (CoTaP)	$0.861_{\pm 0.398}$	$0.101_{\pm 0.055}$	46.664
$\boldsymbol{K}^{ee}$=500 N/m (CoTaP)	$0.759_{\pm 0.273}$	$0.095_{\pm 0.055}$	48.648
$\boldsymbol{K}^{ee}$=800 N/m (CoTaP)	$0.771_{\pm 0.274}$	$0.096_{\pm 0.055}$	50.812
α=0.3 (CoTaP)	$0.711_{\pm 0.224}$	$0.093_{\pm 0.055}$	48.478
α=0.7 (CoTaP)	$0.758_{\pm 0.261}$	$0.096_{\pm 0.055}$	47.507

(b) Walking Mode

Method	e^{torso}	e^{ee}	J^{upper}
PD (FALCON)	$2.126_{\pm 0.734}$	$0.090_{\pm 0.055}$	50.811
Soft P (K_p = 30)	$2.071_{\pm 0.720}$	$0.112_{\pm 0.055}$	44.440
$\boldsymbol{K}^{ee}$=100 N/m (CoTaP)	$2.194_{\pm 0.755}$	$0.100_{\pm 0.054}$	46.541
$\boldsymbol{K}^{ee}$=500 N/m (CoTaP)	$2.153_{\pm 0.737}$	$0.094_{\pm 0.054}$	49.169
$\boldsymbol{K}^{ee}$=800 N/m (CoTaP)	$2.154_{\pm 0.727}$	$0.094_{\pm 0.054}$	51.565
α=0.3 (CoTaP)	$2.106_{\pm 0.716}$	$0.092_{\pm 0.057}$	48.944
α=0.7 (CoTaP)	$2.123_{\pm 0.740}$	$0.095_{\pm 0.054}$	47.524

* In the table, the larger values denote the means, and the smaller values following the ± symbol represent the standard deviations.

simultaneously increasing the torque demands on the arms. Moreover, while the basic method yields reduced torso and hand tracking error, it comes at the cost of increased joint energy expenditure (larger upper-body torques). For further comparison, we employed a soft joint P gain (K_p = 30). The results indicate that while this setting leads to smaller torso tracking errors (meaning the task-space impact exerts less influence on the body), and lower torque demands, it also produces the largest hand tracking mean error (0.11 m). Hence, simply reducing the joint P gain is impractical in the context of robot loco-manipulation. In the final category presented in the table, different ratio α are employed to evaluate the compliance modulation performance of the two control laws. These results show that the modulation integrates the strengths of both controllers: smaller tracking error and lower torque consumption. In practical robotic applications, the ratio can be tuned to accommodate different task requirements.

3) Impact Force at End-effector in Walking Mode: When the robot is walking at a given velocity, a random external impact is applied, and the average performance across different compliance parameter settings is subsequently evaluated over multiple metrics in Section IV-A. All the results are presented in Table III(a). The analysis of the walking state results is largely consistent with the static case, with the main difference being a substantial increase in torso tracking error. Meanwhile, incorporating the ratio α for modulated joint-space stiffness highlights more clearly the advantages of reducing tracking error while balancing joint torque consumption.

Fig. 5 illustrates the screenshots of the robot's reaction under an impact during walking in the case of α = 0.7. In the figure, the robot is depicted walking forward, with the upper body simultaneously tracking a punching motion. After the left hand experiences an external impact (red line in the figure), it is pulled straight; subsequently, due to task-space stiffness, the left hand returns to its original trajectory, while the body, in order to maintain the target velocity after the impact, shows a backward and rotate tendency. An external force of this magnitude was not accounted for in the policy training, as the objective of compliance control is not to achieve exact trajectory tracking at every instant in task space, but to preserve compliance and stability in the presence of unforeseen large disturbances. As indicated in the title, the primary aim of the proposed method is to realize adjustable task-space compliance. Accordingly, these results can be regarded as meeting the requirements of our research objectives.

4) Ablation Study: For ablation study, we first tested the compliance controller directly on the base policy without distillation, but effective control could not be achieved, with the robot failing to maintain stability. The main issue arose from the lower-body motion failing to adapt to the upper-body control law. The results are presented in Table IV. Then, for the upper body we directly employed the original reference trajectory as the control target at each timestep. Although the control was generally realizable, the errors remained considerable, with the upper-body tracking error (0.252) exceeding that of the proposed method (0.107) by more than twice. The reason is still that the upper-body motion was not trained in coordination with the lower-body policy. Moreover, to demonstrate the importance of the modulation ratio α considering Jacobian condition number, we conducted an ablation study of training the policy without setting α. Fig. 6 illustrates a frequent phenomenon observed after training: due to the influence of singular postures, the policy deliberately avoids joint singularity in order to maintain control stability, which in turn leads to larger tracking errors (0.517 for upper-body position).

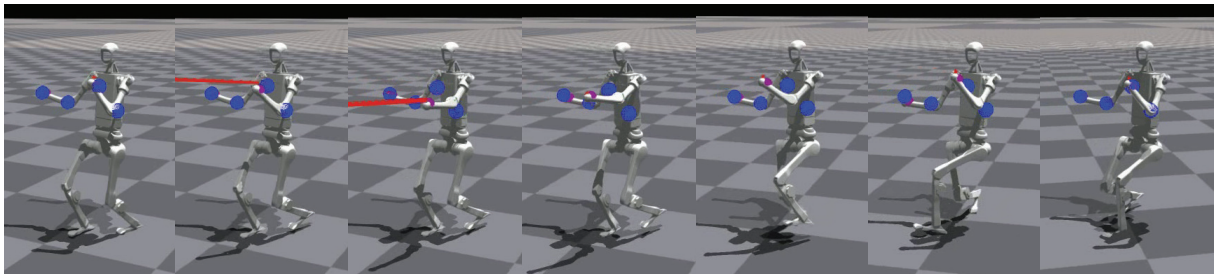


Fig. 5. Screenshots of walking control under an external impact on left hand. The upper-body reference motion is punching. The red line represents the external impact (500 N in 0.05 s). The blue balls are reference points of hands and elbows.

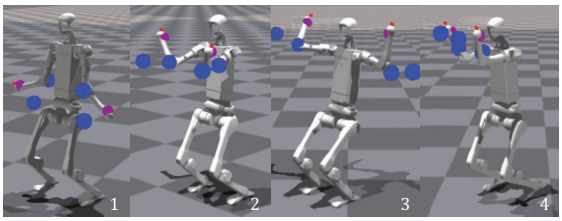


Fig. 6. Policy distillation results without considering Jacobian condition number. At the singular postures (e.g., when the arms are fully extended), the compliance control yields unstable joint stiffness. Consequently, the RL training tends to avoid these postures, leading to larger upper-body tracking errors. The blue balls are reference points of hands and elbows.

TABLE IV

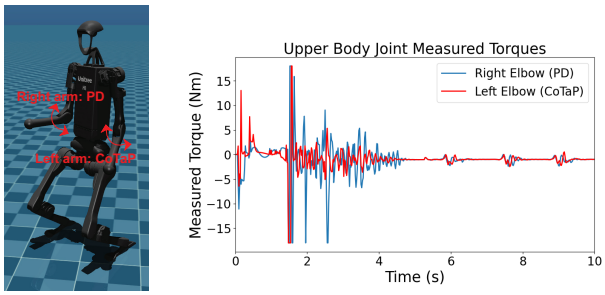
ABLATION STUDY RESULT (STANCE MODE, $\boldsymbol{K}^{ee}$ = 100 N/M)

Method	No-train	w/o α	CoTaP
e^{torso}	$0.950_{\pm 0.621}$	$1.008_{\pm 0.370}$	$0.861_{\pm 0.398}$
e^{upper}	$0.252_{\pm 0.173}$	$0.517_{\pm 1.298}$	$0.107_{\pm 0.105}$

C. Experimental Validation

We applied the policy trained in Isaac Gym together with the overall control method to the MuJoCo simulation [33], thereby achieving sim-to-sim transfer. In this simulation, beyond basic velocity-tracking locomotion and upper-body control, a periodic load experiment was performed: with the robot's arms maintained in their initial configuration, a sinusoidal external force in the z-direction, with a period of 4 s and an amplitude of 30 N, was applied to both hands (as shown in Fig. 7(a)). Meanwhile, different control methods were applied to the robot's arms: the right arm was controlled using pure PD, whereas the left arm employed compliance modulation control with α = 0.7 (between PD and $\boldsymbol{K}^{ee}$ = 500 N/m). Fig. 7(b) shows the results of the measured elbow torques. In the figure, we can observe that the joint torques under modulated compliance control (CoTaP) are much smaller than those under PD control. These results demonstrate that, in practical applications, the task-space compliance value together with the modulation ratio α can be varied to quantitatively regulate the robot's compliance, enabling adaptation to specific objectives such as precision control and energy conservation.

For sim-to-real validation, after training the proposed



(a) (b)

Fig. 7. (a) Simulation of H1 in MuJoCo under a periodic load applied on both hands. In this simulation, right arm is using PD control and left arm is using modulated compliance control (α = 0.7). The left hand swings with a larger amplitude than the right hand. (b) Torque curves of both elbow joints under periodic loading. It can be seen that the torque of the left arm is overall much smaller than that of the right arm.

method on the Unitree G1 robot, we deploy the resulting policy on the physical robot for evaluation. In deployment, we compute the corresponding joint torques (including lower-body joints) directly using (2) and send them to the robot's low-level actuators for execution. Fig. 8 compares the arm compliance behaviors under the baseline and the proposed CoTaP method. The robot is set to a standing mode, with the arms maintained in a nominal L-shaped posture. When the operator pushes or pulls the robot's hand, the baseline remains nearly stationary, whereas under the CoTaP method, the arm exhibits pronounced compliance. Moreover, during force application, we observe that even near kinematic singular configurations, the robot joints do not exhibit jitter.

V. CONCLUSIONS

In this study, we proposed CoTaP, a pipeline applying compliance information in the RL structure, and introduced its controller based on RL and compliance modulation for humanoid robot loco-manipulation. The main contribution of this work is the integration of model-based compliance control into the RL training framework, enabling the robot to benefit from the advantages of RL-based motion generation while ensuring quantitatively adjustable compliance at the low-level controller. By incorporating randomization of the required task-level compliance values, the controller no longer needs to be retrained each time the compliance setting is modified.

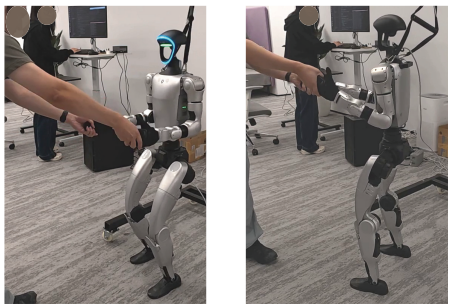


Fig. 8. Experimental snapshots of the G1 robot under different control policies. Left: whole-body PD-based baseline. Right: proposed CoTaP method (task stiffness $\boldsymbol{K}^{ee} = 200$ N/m, modulation ratio $\alpha = 0.7$). Under the same manually applied external force from the operator, the baseline remains almost stationary at the initial target position, whereas under the proposed method the robot's arm is correspondingly displaced.

After achieving this, task-space compliance becomes a parameterizable and adjustable quantity, both for future teleoperation and real-world control. This broadens the overall state space of humanoid robot operation, enhances the adaptability of robotic loco-manipulation, and enables the robot to go beyond simply imitating human motions by compliance modulation according to the actual situation.

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 Transactions On Graphics (TOG)*, vol. 37, no. 4, pp. 1–14, 2018.

[2] X. B. Peng, Z. Ma, P. Abbeel, S. Levine, and A. Kanazawa, "Amp: Adversarial motion priors for stylized physics-based character control," *ACM Transactions on Graphics (ToG)*, vol. 40, no. 4, pp. 1–20, 2021.

[3] A. Tang, T. Hiraoka, N. Hiraoka, F. Shi, K. Kawaharazuka, K. Kojima, K. Okada, and M. Inaba, "Humanmimic: Learning natural locomotion and transitions for humanoid robot via wasserstein adversarial imitation," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2024, pp. 13 107–13 114.

[4] Q. Zhang, P. Cui, D. Yan, J. Sun, Y. Duan, G. Han, W. Zhao, W. Zhang, Y. Guo, A. Zhang, *et al.*, "Whole-body humanoid robot locomotion with human reference," in *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2024, pp. 11 225–11 231.

[5] 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.

[6] D. Rakita, B. Mutlu, M. Gleicher, and L. M. Hiatt, "Shared control–based bimanual robot manipulation," *Science Robotics*, vol. 4, no. 30, p. eaaw0955, 2019.

[7] 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 *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2024, pp. 8944–8951.

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

[9] T.-H. Pham, N. Kyriazis, A. A. Argyros, and A. Kheddar, "Hand-object contact force estimation from markerless visual tracking," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 12, pp. 2883–2896, 2017.

[10] T.-H. Pham, S. Caron, and A. Kheddar, "Multicontact interaction force sensing from whole-body motion capture," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 6, pp. 2343–2352, 2017.

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

[12] E. Dean-Leon, J. R. Guadarrama-Olvera, F. Bergner, and G. Cheng, "Whole-body active compliance control for humanoid robots with robot skin," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 5404–5410.

[13] K. Yamamoto, "Robust Walking by Resolved Viscoelasticity Control Explicitly Considering Structure-Variability of a Humanoid," in *Proc. of IEEE ICRA*, 2017, pp. 3461–3468.

[14] K. Yamamoto, T. Ishigaki, and Y. Nakamura, "Humanoid motion control by compliance optimization explicitly considering its positive definiteness," *IEEE Transactions on Robotics*, 2021.

[15] S. G. Khan, G. Herrmann, M. Al Grafi, T. Pipe, and C. Melhuish, "Compliance control and human–robot interaction: Part 1—survey," *International journal of humanoid robotics*, vol. 11, no. 03, p. 1430001, 2014.

[16] C. Ott, A. Dietrich, and A. Albu-Schäffer, "Prioritized multi-task compliance control of redundant manipulators," *Automatica*, vol. 53, pp. 416–423, 2015.

[17] S. Lee, P. S. Chang, and J. Lee, "Deep compliant control," in *ACM SIGGRAPH 2022 conference proceedings*, 2022, pp. 1–9.

[18] A. Hartmann, D. Kang, F. Zargarbashi, M. Zamora, and S. Coros, "Deep compliant control for legged robots," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2024, pp. 11 421–11 427.

[19] D. Spoljaric, Y. Yan, and D. Lee, "Variable stiffness for robust locomotion through reinforcement learning," *arXiv preprint arXiv:2502.09436*, 2025.

[20] R. Watanabe, T. Miki, F. Shi, Y. Kadokawa, F. Bjelonic, K. Kawaharazuka, A. Cramariuc, and M. Hutter, "Learning quiet walking for a small home robot," *arXiv preprint arXiv:2502.10983*, 2025.

[21] 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.

[22] 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.

[23] L. Wei, X. Peng, R.-Z. Qiu, X. Cheng, and X. Wang, "Hmc: Learning heterogeneous meta-control for contact-rich loco-manipulation," in *RSS 2025 Workshop on Whole-body Control and Bimanual Manipulation: Applications in Humanoids and Beyond*.

[24] J. Ho and S. Ermon, "Generative adversarial imitation learning," *Advances in neural information processing systems*, vol. 29, 2016.

[25] 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.

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

[27] Z. Ding, H. Jiang, Y. Wang, Z. Sun, Y. Zhang, X. Niu, M. Yang, W. Zeng, X. Xu, and Z. Lu, "Jaeger: Dual-level humanoid whole-body controller," *arXiv preprint arXiv:2505.06584*, 2025.

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

[29] V. Arsigny, P. Fillard, X. Pennec, and N. Ayache, "Fast and simple calculus on tensors in the log-euclidean framework," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2005, pp. 115–122.

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

[31] N. Mahmood, N. Ghorbani, N. F. Troje, G. Pons-Moll, and M. J. Black, "Amass: Archive of motion capture as surface shapes," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 5442–5451.

[32] Z. Luo, J. Cao, K. Kitani, W. Xu, *et al.*, "Perpetual humanoid control for real-time simulated avatars," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*.

[33] E. Todorov, T. Erez, and Y. Tassa, "Mujoco: A physics engine for model-based control," in *2012 IEEE/RSJ international conference on intelligent robots and systems*. IEEE, 2012, pp. 5026–5033.