

# FALCON: Learning Force-Adaptive Humanoid Loco-Manipulation

Anonymous Author(s)

Affiliation

Address

email

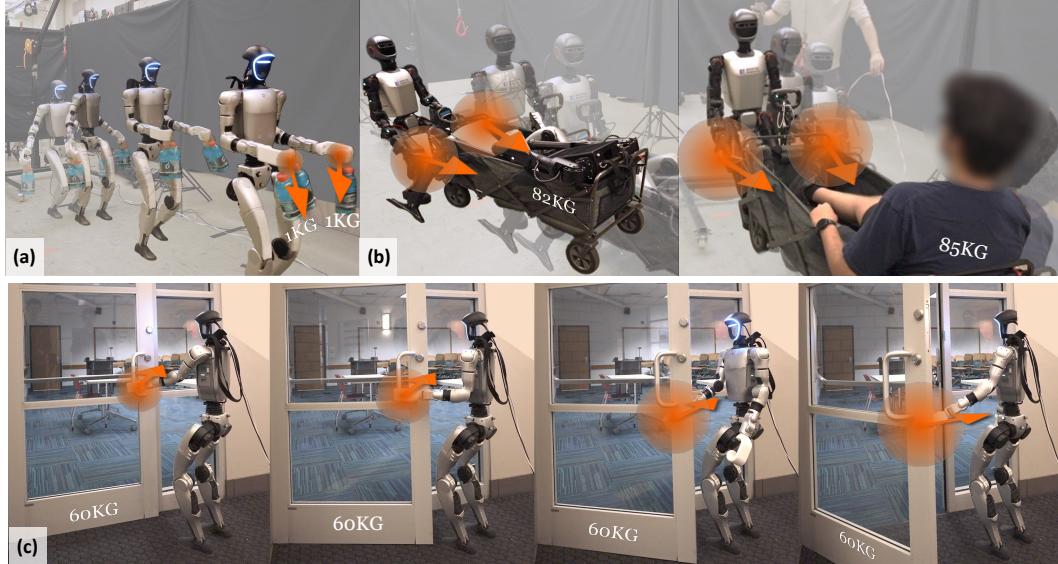


Figure 1: FALCON enables humanoid robots to perform versatile forceful loco-manipulation tasks: (a) **Transporting Payloads**: walk, squat, twist torso with payloads; (b) **Cart-Pulling** with significant lateral forces; (d) **Door-Opening** with multi-directional forces. **Videos:** see our website: <https://anonymous-falcon-corl25.github.io/>

1       **Abstract:** Humanoid loco-manipulation holds transformative potential for daily  
2 service and industrial tasks, yet achieving precise, robust whole-body control with  
3 3D end-effector force interaction remains a major challenge. Prior approaches  
4 are often limited to lightweight tasks or quadrupedal/wheeled platforms. To  
5 overcome these limitations, we propose FALCON, a dual-agent reinforcement-  
6 learning-based framework for robust force-adaptive humanoid loco-manipulation.  
7 FALCON decomposes whole-body control into two specialized agents: (1) a  
8 lower-body agent ensuring stable locomotion under external force disturbances,  
9 and (2) an upper-body agent tracking precise end-effector positions with implicit  
10 force compensation. These two agents are jointly trained in simulation with a  
11 force curriculum that progressively escalates the magnitude of external force ex-  
12 ereted on the end effector while respecting joint torque limits. Experiments demon-  
13 strate that, compared to the baselines, FALCON achieves 2× more accurate upper-  
14 body joint tracking, while maintaining robust locomotion under force disturbances  
15 and achieving faster training convergence. Importantly, FALCON enables policy  
16 training without embodiment-specific reward or curriculum tuning. Using the  
17 same training setup, we obtain policies that deploy across multiple humanoid  
18 robots, enabling forceful loco-manipulation tasks such as transporting payloads  
19 (0-20N), cart-pulling (0-100N), and door-opening (0-40N) in the real world.

20       **Keywords:** Humanoid Loco-Manipulation, Force Adaptation, RL

21 **1 Introduction**

22 Humanoid robots have demonstrated remarkable progress in locomotion and manipulation [1–5].  
23 However, extending these capabilities to forceful loco-manipulation remains fundamentally chal-  
24 lenging. Tasks such as door opening, highlighted in the 2015 DARPA Challenge [6], require not  
25 only precise manipulation under dynamic, multi-directional forces but also maintaining lower-body  
26 stability throughout the interaction. Meeting these demands calls for humanoid systems that can  
27 flexibly adapt to varying payloads and contact forces without compromising overall precision and  
28 robustness in loco-manipulation.

29 Reinforcement Learning (RL) has achieved impressive results for humanoid whole-body control [7–  
30 25], yet existing RL approaches succeed mostly on lightweight tasks but do not consider significant  
31 force adaptation during humanoid loco-manipulation. Currently, there are two main paradigms: (1)  
32 *Lower-RL-Upper-IK*, which applies RL to lower-body locomotion while using kinematic solvers for  
33 upper-body control [7, 8], lacks whole-body dynamics modeling for forceful interaction and has  
34 limited whole-body coordination; (2) *Monolithic-Whole-body-RL*, which directly learns to control  
35 all degrees of freedom [9, 10], suffers from inefficient exploration as a single policy must simulta-  
36 neously learn weakly correlated locomotion and manipulation skills. Though some advances have  
37 been made in force adaptation for quadrupeds [11–13], humanoids pose extra challenges like higher  
38 DOF and stricter torque limits, especially in certain joint configurations.

39 In this work, we aim to develop an RL framework that enables humanoid robots to perform a diverse  
40 set of force-adaptive loco-manipulation tasks. To this end, we introduce FALCON, a dual-agent RL  
41 architecture trained with a 3D force curriculum respecting joint torque limits. Our key innovations  
42 include: (1) A *dual-agent learning decomposition* that separates lower-body and upper-body policy  
43 training with tailored rewards while sharing the same whole-body proprioception and commands;  
44 (2) An *3D force curriculum* with joint torque feasibility that progressively scales applied 3D forces  
45 on both end-effectors while enforcing joint torque constraints through inverse dynamics. This ar-  
46 chitecture enables efficiently joint training of both stable locomotion and accurate EE tracking in  
47 forceful loco-manipulation tasks. We validate FALCON on Unitree G1 and Booster T1 humanoids,  
48 demonstrating its generalization across different platforms through: (1) Transporting Payloads (2)  
49 Cart-Pulling, and (3) Door-Opening (Figure. 1), which require significant and 3D force adaptability.  
50 In summary, our main contributions are:

- 51 • We introduce FALCON, a dual-agent reinforcement learning framework that enables humanoids  
52 to perform forceful loco-manipulation while adapting to substantial, unknown end-effector forces  
53 (0–100N, up to 30% of body weight). FALCON achieves twice the upper-body joint tracking  
54 accuracy of prior methods while maintaining robust locomotion performance.
- 55 • To facilitate RL training of FALCON, we design a 3D force curriculum with progressive force  
56 application while ensuring joint torque feasibility and maximizing its force-adaptive capability.
- 57 • We validate FALCON on two different humanoid platforms (Unitree G1, Booster T1), achieving  
58 strong cross-platform generalization with minimal tuning overhead.

59 **2 Related Works**

60 **2.1 Humanoid Loco-Manipulation**

61 Humanoid loco-manipulation remains a challenging control problem in robotics. While traditional  
62 model-based methods (e.g., simplified dynamics models and MPC) [19, 12, 20–24] offer real-  
63 time planning, their reliance on manual design limits flexibility and generalizability. In contrast,  
64 learning-based methods—particularly sim-to-real RL—have demonstrated promising results in ver-  
65 satile loco-manipulation tasks [26, 16, 10, 7, 8, 27, 28]. For humanoids, two primary paradigms  
66 have emerged: *Lower-RL-Upper-IK* and *Monolithic-Whole-body-RL*. For *Lower-RL-Upper-IK*, Lu  
67 *et al* [7] introduce PMP which uses inverse kinematics (IK) and PD control for upper body con-  
68 trol while locomotion is trained and conditioned on a Conditional Variational Autoencoder (CVAE)

representing upper-body motions. Then, Ben *et al* [8] propose HOMIE that follows the same decoupling framework but introduces a exoskeleton-based cockpit for more intuitive human teleoperation. For *Monolithic-Whole-body-RL*, Dao et al. [27] adopted a unified RL approach for box pick-and-place tasks, training distinct skills (e.g., lifting, walking, stance) and orchestrating them via a finite state machine. He *et al* [15, 16] and Ji *et al* [9] employ a teacher-student training framework to mimic human motions for loco-manipulation tasks.

Despite these advances, few RL methods address significant unknown force disturbances on the EEs for humanoid loco-manipulation, and both paradigms exhibit critical shortcomings accordingly. *Lower-RL-Upper-IK* approaches suffer from delayed force compensation for upper-body control. *Monolithic-Whole-body-RL* methods face sample inefficiency from unrelated task objectives between upper-body manipulation and lower-body locomotion, exacerbated by single agent that often leads to behavioral dominance. In this study, inspired by [29–31], we propose FALCON, a dual-agent RL framework employing task-specific reward formulations for upper-lower body decomposition. Unlike separately trained architectures, the two agents in FALCON are jointly trained with shared proprioception and commands, allowing mutual awareness of each other’s behaviors. This joint training prevents the agents from adapting in isolation and enables coordinated responses to external forces that affect the full-body dynamics.

## 2.2 Forceful Interaction in Legged Robots

Forceful interaction has been extensively studied in quadrupedal robots with mounted arms, where model-based approaches, particularly MPC combined with force planning and control for robust and adaptive locomotion and manipulation [11, 12, 32]. Recent advances in RL have further enhanced adaptability, enabling quadrupeds to learn adaptive and agile force interactions including impedance control [33] and aggressive force adaptation [13]. For humanoids, forceful interaction presents significantly greater challenges due to their more complex dynamics and stringent joint limits. Unlike quadrupeds with centralized mass distributions, humanoids exhibit coupled dynamics between their upper and lower bodies, making force adaptation particularly difficult. Recent model-based approaches have demonstrated force control for heavy-duty tasks [34, 21], but these require prior knowledge of manipulated objects’ mass, CoM, or pre-defined force trajectories, limiting their applicability to unknown disturbances. While some works have attempted explicit force estimation for humanoids [35], these are restricted to quasi-static scenarios and cannot handle force adaptation in dynamic loco-manipulation scenarios.

In this paper, FALCON implicitly learn to adapt to unknown external forces on the different EEs with a novel omni-directional EE force curriculum that considers humanoid joint torque limits. In this way, we can maximize the force adaptability of the learned loco-manipulation policy while ensuring the joint torque limits for robust and safe real-world deployment.

## 3 FALCON: Force-Adaptive Humanoid Loco-Manipulation

Humanoid loco-manipulation under external EE forces requires coordinated control of both the lower and upper body. We first formulate the problem as a unified *dual goal-conditioned* policy learning problem. Let the degrees of freedom (DoFs) of the humanoid be partitioned into lower-body joints and upper-body joints, with  $n^l$  denoting the number of lower-body DoFs,  $n^u$  the number of upper-body DoFs, and  $n = n^l + n^u$  the total number of actuated joints.

The robot proprioception  $s_t^p \in \mathcal{S}_t$  is defined as  $s_t^p \triangleq [q_{t-4:t}, \dot{q}_{t-4:t}, \omega_{t-4:t}^{root}, g_{t-4:t}, a_{t-5:t-1}]$ , which contains five-step histories of joint positions  $q_t \in \mathbb{R}^n$ , joint velocities  $\dot{q}_t \in \mathbb{R}^n$ , root angular velocity  $\omega_t \in \mathbb{R}^3$ , projected gravity  $g_t \in \mathbb{R}^3$ , and previous actions  $a_{t-1} \in \mathbb{R}^n$ . The goal space  $\mathcal{G}_t$  consists of locomotion goals  $\mathcal{G}_t^l \triangleq [\mathbf{v}_t^{lin,ang}, \phi_t^{stance}, h_t^{root}, w_t^{yaw}]$ , specifying desired root linear and angular velocities, stance indicators, root heights, and waist yaw angles, and manipulation goals  $\mathcal{G}_t^u \triangleq [\mathbf{q}_t^{upper}]$ , specifying target joint configurations for the upper body where  $\mathbf{q}_t^{upper} \in \mathbb{R}^{n^u}$ .

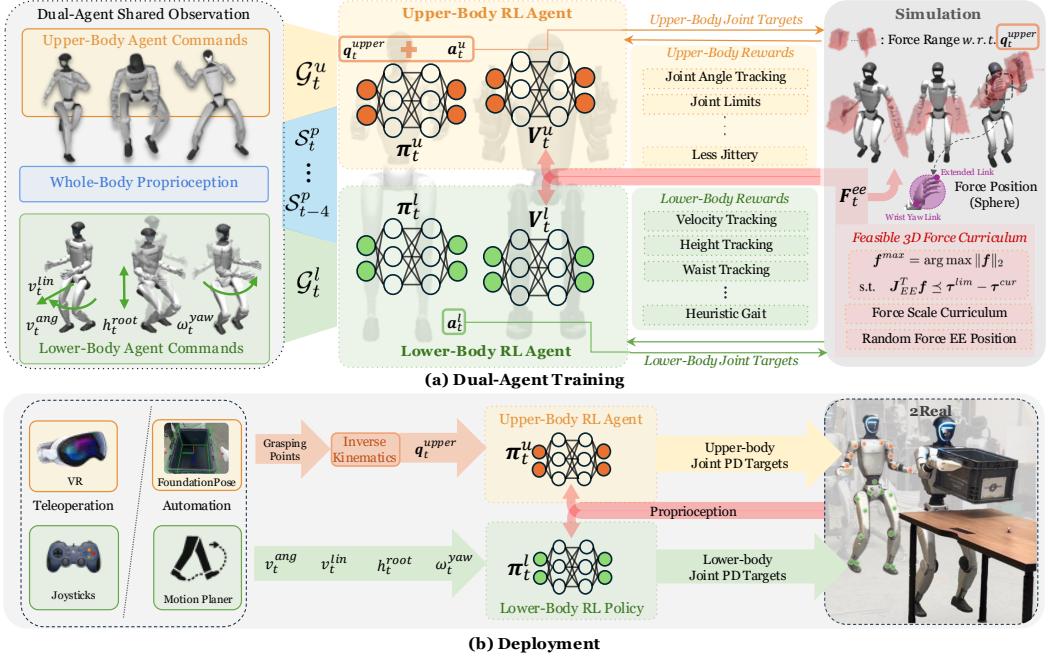


Figure 2: Overview of FALCON. (a) Two agents with different sub-tasks are jointly trained with shared whole-body proprioception. During training, we apply 3D external forces bounded by upper-body joint torque limits on the end-effectors; (b) FALCON is deployed with either teleoperation or an autonomy pipeline including FoundationPose for pose estimation and motion planning.

116 Under this unified formalism, conventional methods differ mainly in how they generate the action  
 117  $a_t \in \mathbb{R}^n$  that commands the robot joints:

- 118 • **Lower-RL-Upper-IK**: lower-body actions  $\mathcal{A}_t^l \in \mathbb{R}^{n^l}$  are generated by a policy  $\pi^l : s_t^p \times \langle \mathcal{G}_t^l, \mathcal{G}_t^u \rangle \mapsto$   
 119  $\mathcal{A}_t^l$  conditioned on whole-body proprioception and goals, while upper-body actions  $\mathcal{A}_t^u \in \mathbb{R}^{n^u}$  are  
 120 computed through inverse kinematics (IK) solvers based on  $\mathcal{G}_t^u$ .  
 121 • **Monolithic-Whole-body-RL**: a single policy  $\pi : s_t^p \times \mathcal{G}_t \mapsto a_t$  directly predicts the full-body action  
 122  $a_t \in \mathbb{R}^n$ , attempting to satisfy both locomotion and manipulation objectives simultaneously.

123 While *Lower-RL-Upper-IK* methods are sample-efficient, they neglect upper-body force compensation  
 124 and whole-body coupling under EE force disturbances. In contrast, *Monolithic-Whole-body-RL*  
 125 methods improve expressiveness but suffer from exploration inefficiency due to the large action  
 126 space spanning unrelated locomotion and manipulation objectives. To overcome these challenges,  
 127 we introduce FALCON, a dual-agent RL framework that achieves training efficiency and coordina-  
 128 tion through decomposition learning with shared whole-body observation.

### 129 3.1 Dual-Agent Learning Framework

130 As shown in Figure 2, FALCON jointly trains two agents, each specialized for a different subtask.  
 131 The lower-body locomotion agent learns a policy  $\pi^l : s_t^p \times \mathcal{G}_t^l \mapsto \mathcal{A}_t^l$  with value function  $V^l(\cdot)$ , while  
 132 the upper-body manipulation agent learns a policy  $\pi^u : s_t^p \times \mathcal{G}_t^u \mapsto \mathcal{A}_t^u$  with value function  $V^u(\cdot)$ .  
 133 Both agents observe the same proprioceptive input  $s_t^p$  but optimize independent goal-conditioned  
 134 objectives:

$$r_t^l = \mathcal{R}^l(s_t^p, \mathcal{G}_t^l) \text{ (locomotion)} \quad r_t^u = \mathcal{R}^u(s_t^p, \mathcal{G}_t^u) \text{ (manipulation)} \quad (1)$$

135 These two policy parameters  $\theta_l$  and  $\theta_u$  are updated via proximal policy optimization (PPO [36]):

$$\max_{\theta_l} \mathbb{E} \left[ \sum_{t=1}^T \gamma^{t-1} r_t^l \right] \text{ (Lower-body)} \quad \max_{\theta_u} \mathbb{E} \left[ \sum_{t=1}^T \gamma^{t-1} r_t^u \right] \text{ (Upper-body)} \quad (2)$$

136 The upper-body target joint angles  $q_t^{upper}$  (target joints of shoulders, elbows, wrists), are randomly  
 137 sampled from the AMASS dataset [37] during training, and calculated via IK during deployment.  
 138 The combined action from the two agents  $a_t = [a_t^l; a_t^u]$  is sent to a joint-level PD controller. As  
 139 the real-world humanoid control is inherently partially observable, we adopt asymmetric actor-critic  
 140 training, where critics additionally access privileged information including root linear velocities and  
 141 EE forces  $F_t^{ee}$  during training but not during deployment. Detailed reward designs and domain  
 142 randomization during training are provided in Appendix A.1 and Appendix A.2.

### 143 3.2 Torque-Limit-Aware 3D Force Curriculum

144 For humanoid robots—particularly those with relatively weak joint torque limits, such as the wrist  
 145 joints on the Unitree Humanoid G1—it is crucial to explicitly account for these torque constraints  
 146 when large external disturbances are applied to the end-effectors (EEs). Ignoring these limits during  
 147 upper body policy training can lead to unexpected or unsafe behaviors due to torque saturation or  
 148 joint limit violations in real-robot deployment. Additionally, it's important to gradually increase the  
 149 external force during training, allowing the policy to progressively learn effective force adaptation  
 150 strategies. To achieve these, our force application framework follows through three principles:

151 **Torque-Aware Force Computation:** Before applying forces, we first need to estimate the max-  
 152 imum forces that we can exert on the left or right end-effector. Given the left or right end-effector  
 153 Jacobian  $J_{EE} \in \mathbb{R}^{3 \times \frac{n_u}{2}}$  at its Center of Mass (CoM) and current joint torques  $\tau^{cur} \in \mathbb{R}^{\frac{n_u}{2}}$ , we es-  
 154 timate the maximum/minimum admissible force  $f_i^{max}, f_i^{min}$  along each Cartesian axis  $i \in \{x, y, z\}$   
 155 by analyzing the worst-case joint torque induced by a unit force in that direction. The element-wise  
 156 force bound can be computed in parallel as:

$$f_i^{max} = \min_j \left( \frac{\tau_j^{\lim} - \tau_j^{cur}}{|J_{EE}^{ji}| + \epsilon} \right), \quad f_i^{min} = \max_j \left( \frac{-\tau_j^{\lim} - \tau_j^{cur}}{|J_{EE}^{ji}| + \epsilon} \right), \quad (3)$$

157 where  $J_{EE}^{ji}$  denotes the  $(j, i)$ -th element of the end-effector Jacobian matrix, and  $\epsilon$  is a small positive  
 158 constant to prevent division by zero. After that, we sample the relatively ratio  $\gamma = [\gamma_x, \gamma_y, \gamma_z]$   
 159 among x, y and z axis through Dirichlet Distribution [38], which satisfy  $\sum_{i \in \{x, y, z\}} \gamma_i = 1$ . The  
 160 feasible applied force will be uniformly sampled within the estimated range and expressed as:

$$f_t^{ee} = \sum_{i \in \{x, y, z\}} F_i \cdot e_i, \quad \text{where } F_i \sim \mathcal{U}[\gamma_i \cdot f_i^{min}, \gamma_i \cdot f_i^{max}] \quad (4)$$

161 This approach maximizes force adaptivity while respecting torque limits, leading to more effective  
 162 training than random sampling (??). Note that applied forces may differ between left and right EEs  
 163 due to asymmetric upper-body configurations (Figure 2).

164 **Progressive Force Curriculum:** To facilitate progressive force adaptation, the estimated EE  
 165 forces are scaled by a global factor  $\alpha_g \in (0, 1)$ , increasing over training, so the applied force  
 166 becomes  $F_t^{ee} = \alpha_g \cdot f_t^{ee}$ . During walking, planar forces are projected opposite to the velocity. A  
 167 low-pass filter is applied to reduce force jitter.

168 **Position Randomization of the Applied Force:** Learning-based force adaptation leverages pro-  
 169 prioceptive history to implicitly compensate for external forces, removing the need for explicit force  
 170 estimation [35]. To improve robustness to variations in end-effector (EE) contact points—which  
 171 alter the torque mapping via the EE Jacobian—we randomize force application along the EE link,  
 172 from the wrist yaw to the distal segment, as illustrated in Figure 2.

## 173 4 Simulation and Real-World Experiments

174 In this section, we present extensively quantitative comparison between FALCON and the baselines  
 175 as well as qualitative results on real-world deployment. We choose Unitree Humanoid G1 and  
 176 Booster T1 as our humanoid platforms. Specifically, we address the following key questions:

Methods	$E_{\text{tracking}}^{\text{upper}} \downarrow$			$E_{\text{tracking}}^{\text{root}} \downarrow$			
	N-Force	M-Force	L-Force	N-Force	M-Force	L-Force	
Lower-RL	PD-w/o-Force-Curr.	$0.46 \pm 0.04$	$0.94 \pm 0.05$	$1.44 \pm 0.06$	$0.38 \pm 0.04$	$0.66 \pm 0.05$	$1.14 \pm 0.06$
	PD-Force-Curr.	$0.44 \pm 0.03$	$0.93 \pm 0.03$	$1.42 \pm 0.04$	$0.33 \pm 0.03$	$0.38 \pm 0.03$	$0.46 \pm 0.03$
	PID-Force-Curr.	$0.24 \pm 0.03$	$0.31 \pm 0.03$	$0.60 \pm 0.04$	$0.32 \pm 0.03$	$0.35 \pm 0.03$	$0.46 \pm 0.04$
	PD-ID-Force-Curr.	$0.29 \pm 0.03$	$0.38 \pm 0.03$	$0.53 \pm 0.04$	$0.40 \pm 0.03$	$0.42 \pm 0.03$	$0.47 \pm 0.04$
M-WB-RL	w/o-Force-Curr.	$0.46 \pm 0.05$	$1.03 \pm 0.07$	$1.65 \pm 0.08$	$0.34 \pm 0.04$	$0.67 \pm 0.05$	$1.28 \pm 0.07$
	with-Force-Curr.	$0.43 \pm 0.04$	$0.50 \pm 0.04$	$0.73 \pm 0.05$	$0.28 \pm 0.03$	$0.32 \pm 0.03$	$0.44 \pm 0.04$
<b>FALCON</b>	w/o-Force-Curr.	<b><math>0.14 \pm 0.03</math></b>	$0.55 \pm 0.04$	$1.06 \pm 0.06$	$0.32 \pm 0.03$	$0.66 \pm 0.05$	$1.24 \pm 0.06$
	with-Force-Curr.	$0.21 \pm 0.03$	<b><math>0.24 \pm 0.03</math></b>	<b><math>0.37 \pm 0.04</math></b>	<b><math>0.27 \pm 0.03</math></b>	<b><math>0.30 \pm 0.03</math></b>	<b><math>0.45 \pm 0.04</math></b>

Table 1: Loco-Manipulation Evaluation of FALCON and Baselines in IsaacGym.

177 **Q1:** Can FALCON outperform other baselines in terms of both upper-body manipulation and lower-  
 178 body locomotion performance?

179 **Q2:** Why does FALCON has better training-efficiency compared to Monolithic-Whole-body-RL (M-  
 180 WB-RL) for force-adaptive loco-manipulation?

181 **Q3:** Does FALCON work for different humanoids to show cross-platform generalizability?

#### 182 4.1 Evaluation Criterion

183 To evaluate the performance of the learned low-body locomotion and upper-body manipulation capa-  
 184 bilities, we consider the following metrics under dynamically unknown and 3D EE forces  $\mathbf{F}_t \in \mathbb{R}^3$ ,  
 185 given a sequence of target upper-body joints  $\mathbf{q}_t^{\text{upper}*}$ , target root velocities  $\mathbf{v}_t^{\text{lin,ang}*}$  and stance signal  
 186  $\phi_t^{\text{stance}}$ , where  $t = 1, 2, \dots, T$  and  $T$  is the sequence length:

187 (ii) *Upper-Body Joints Tracking Error:*  $E_{\text{tracking}}^{\text{upper}}(\mathbf{q}_t^{\text{upper}*}) = \frac{1}{T} \sum_{t=1}^T \left| \mathbf{q}_t^{\text{upper}} - \mathbf{q}_t^{\text{upper}*} \right|$

188 (iii) *Root Velocity Tracking Error:*  $E_{\text{tracking}}^{\text{root}}(\mathbf{v}_t^{\text{lin,ang}*}) = \frac{1}{T} \sum_{t=1}^T \left| \mathbf{v}_t^{\text{lin,ang}} - \mathbf{v}_t^{\text{lin,ang}*} \right|$

#### 189 4.2 Baselines

190 We consider two types of baseline methods for force adaptation, both trained under the same goal  
 191 space (e.g., commands) in Section 3.1 and force curriculum described in Section 3.2, with each type  
 192 further including relevant ablation variants.

193 **Decoupled Lower-body RL with Upper-body IK Controllers.** For all variants, RL is used for  
 194 lower-body locomotion, and IK provides target upper-body joint angles from end-effector poses.  
 195 The key differences lie in the use of force curriculum and the upper-body joint tracking strategy:

- 196 (a) *Upper-PD-w/o-Force-Curr.:* A baseline following [7, 8], using PD control for upper-body joint  
 197 tracking without force randomization.
- 198 (b) *Upper-PD:* Extends (a) by incorporating force curriculum, enabling lower-body adaptation to  
 199 external forces; upper-body remains PD-controlled.
- 200 (c) *Upper-PID:* Extends (b) by adding an integral term to the upper-body controller to reduce  
 201 steady-state tracking error.
- 202 (d) *Upper-PD-ID:* Extends (a) with a learned force estimator [33] and inverse dynamics-based  
 203 torque compensation under quasi-static assumptions (details in Appendix A.4).

#### 204 Monolithic Whole-body RL

- 205 (e) *Monolithic-WB-RL-w/o-Force-Curr.:* Built upon prior designs [10, 14], a single agent is trained  
 206 with the same goal commands as FALCON, but without applying any force during training.

207 (f) *Monolithic-WB-RL*: Based on (e), we adopt force randomization into the training curriculum  
 208 for force adaptation, while keeping the other training settings identical.

### 209 4.3 Simulation Results

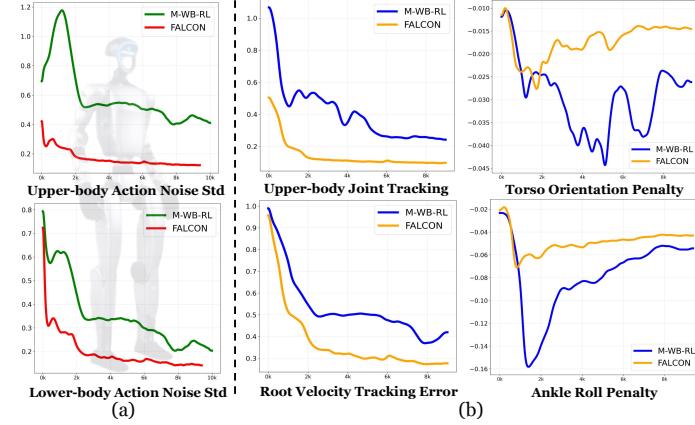
210 To answer **Q1** (*Can FALCON outperform other baselines in terms of both upper-body manipulation*  
 211 *and lower-body locomotion performance?*) and **Q2** (*Why does FALCON has better*  
 212 *training-efficiency compared to Monolithic-Whole-body-RL (M-WB-RL) for force-adaptive loco-*  
 213 *manipulation?*), we conduct quantitative comparisons of our method with other two baselines in  
 214 IsaacGym on Unitree Humanoid G1.

215 **Loco-Manipulation Performance:** We evaluate FALCON and baselines on 252 ACCAD [39] motion  
 216 targets under three force levels: (i) No-Force ( $\alpha_g = 0$ ), (ii) Middle-Force ( $\alpha_g = 0.5$ ), and  
 217 (iii) Large-Force ( $\alpha_g = 1.0$ ), applied to both end-effectors. As shown in Table 1, across all settings,  
 218 FALCON with force curriculum achieves the lowest tracking errors in both upper-body motion  
 219 ( $E_{\text{tracking}}^{\text{upper}}$ ) and root velocity ( $E_{\text{tracking}}^{\text{root}}$ ), demonstrating robust manipulation under disturbance. Under  
 220 L-Force, it reduces upper-body error to 0.37, outperforming PID-Force-Curr. (0.60) and M-WB-  
 221 RL (0.73). Root error remains low at 0.45, indicating stable locomotion. While force curriculum  
 222 benefits all methods, FALCON gains most due to its decomposed learning structure.

### 223 Exploration and Learning:

224 (i) *Action Noise Std*: As shown in Figure 3 (a), FALCON exhibits faster and smoother noise decay in both upper and lower body actions, indicating more efficient and stable exploration.  
 225 In contrast, M-WB-RL suffers from prolonged noise due to entangled control objectives, especially for upper body actions.

226 (ii) *Reward and Postural Stability*: As shown in Figure 3 (b), FALCON achieves less tracking errors in both upper-body joints and base angular velocity while in M-WB-RL these two reward terms tend to fluctuate. Additionally, M-WB-RL suffers from larger torso and ankle penalties due to excessive whole-body compensation, resulting in unnatural bending and CoM upright misalignment as shown in Figure 4.



227 Figure 3: Comparison of FALCON and M-WB-RL: (a) action  
 228 noise std; (b) tracking errors and penalties.

### 242 4.4 Real-World Quantitative Tracking Results

243 We evaluate FALCON on Unitree G1 with each hand loaded with 1.2kg payload in a real-world task,  
 244 which is walking at (0.5, 0.0)m/s with zero angular velocity, fixed height and waist, and keeping the  
 245 upper body in its default position. We compare against two baselines: (i) *Upper-PD with Force Cur-*  
 246 *riculum*, and (ii) *Monolithic-WB-RL with Force Curriculum*. As shown in Table 2, FALCON achieves  
 247 the lowest tracking errors, and perform stable and natural motion in heavy-duty loco-manipulation.

### 248 4.5 Real-world Deployment with Teleoperation

249 To answer **Q3** (*Does FALCON work for different humanoids to show cross-platform generalizability?*)  
 250 ), we deploy policies trained in simulation on the Unitree G1 and Booster T1 humanoids **without**  
 251 **any reward or force curriculum modifications**, thanks to FALCON efficient dual-agent training and  
 252 torque-limit-aware 3D force curriculum. As shown in Figure 1, we evaluate the policies on three  
 253 forceful loco-manipulation tasks: (1) **Transporting Payloads**, with 0-20N vertical forces while

Method	$E_{\text{tracking}}^{\text{upper}}$	$E_{\text{tracking}}^{\text{root}}$
Upper-PD-Force-Curr.	$1.81 \pm 0.13$	<b>0.40</b> $\pm 0.04$
Monolithic-WB-RL	$0.81 \pm 0.11$	$0.58 \pm 0.05$
<b>FALCON</b>	<b>0.39</b> $\pm 0.08$	<b>0.42</b> $\pm 0.03$

Table 2: Real-world Tracking Errors.

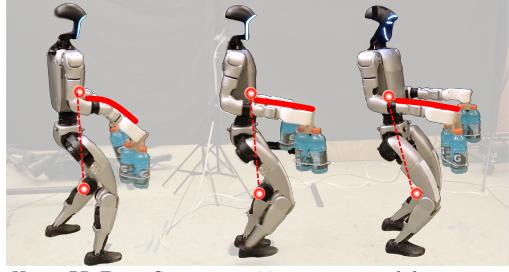
Upper-PD-Force-Curr.    **FALCON**    Monolithic-WB-RL

Figure 4: Real-World Payload Transportation.

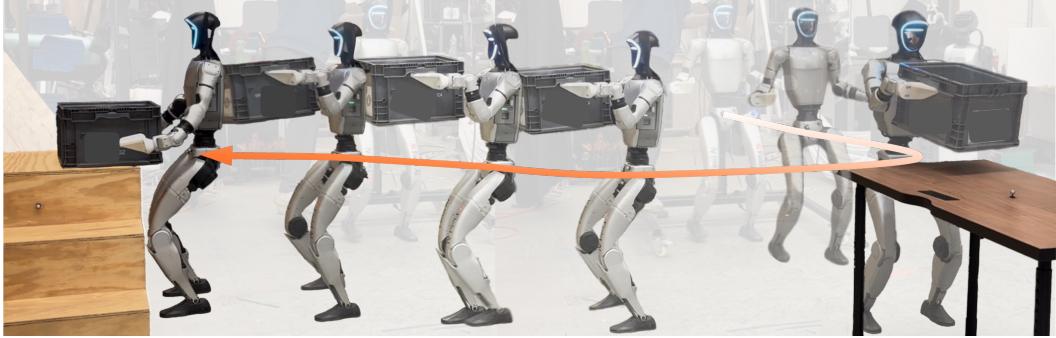


Figure 5: Autonomous Tote Logistics

254 maintaining stable locomotion and precise upper-body joint tracking; (2) **Cart-Pulling**, with up to  
 255 100N lateral (X-Y) forces during walking; and (3) **Door-Opening**, with up to 40N 3D forces during  
 256 stance. These force ranges are measured through a force gauge shown in Appendix A.5.

257 These results demonstrate that FALCON enables robust policy transfer across platforms with differ-  
 258 ent morphologies and actuation. The learned policies exhibit effective whole-body compensation:  
 259 the upper body responds adaptively to 3D forces, the lower body leans against significant lateral  
 260 forces, and the base height remains stable under vertical loads.

#### 261 4.6 Real-World Deployment with Autonomy

262 We also deploy FALCON on the Unitree G1 for autonomous tote logistics, a representative ware-  
 263 house task. As illustrated in Figure 5, the robot is required to walk from an initial location to a  
 264 pickup station, lift a tote of unknown weight, and transport it to a designated area for precise place-  
 265 ment. The detailed implementation of the autonomous pipeline can be found in Appendix A.6.

## 266 5 Conclusion

267 In this paper, we introduce FALCON, a dual-agent reinforcement learning framework designed for  
 268 force-adaptive humanoid loco-manipulation. By decoupling the learning of the upper and lower  
 269 body, while maintaining coordination through shared proprioceptive feedback, FALCON achieves  
 270 superior adaptability in handling 3D end-effector forces during complex tasks. Our extensive eval-  
 271 uation demonstrates that FALCON outperforms both Lower-RL-Upper-IK and Monolithic-Whole-  
 272 body-RL baselines, achieving faster training convergence, reduced tracking errors, and more stable  
 273 performance across a variety of force regimes. Moreover, FALCON exhibits strong cross-platform  
 274 generalizability, successfully transferring policies from simulation to physical humanoids, includ-  
 275 ing tasks like transporting payloads, cart-pulling, and door-opening. These results underscore  
 276 FALCON’s potential for real-world deployment in dynamic, forceful environments.

277 **6 Limitations**

278 Despite its strong performance, FALCON has two key limitations. First, it focuses solely on force  
279 disturbances applied to the end-effectors, without accounting for contact forces on other body parts  
280 or supporting multi-contact interactions. This restricts its applicability in scenarios involving whole-  
281 body support, such as leaning, bracing, or collaborative lifting. Second, the current force curriculum  
282 only considers linear external forces and ignores external torques. As a result, FALCON may strug-  
283 gle in tasks that involve rotational disturbances, such as operating handles or tools with eccentric  
284 loading. Addressing these limitations by incorporating multi-contact reasoning and torque-adaptive  
285 policies remains an important avenue for future research.

286 **References**

- 287 [1] Z. Gu, J. Li, W. Shen, W. Yu, Z. Xie, S. McCrory, X. Cheng, A. Shamsah, R. Griffin, C. K.  
288 Liu, et al. Humanoid locomotion and manipulation: Current progress and challenges in control,  
289 planning, and learning. [arXiv preprint arXiv:2501.02116](#), 2025.
- 290 [2] K. Darvish, L. Penco, J. Ramos, R. Cisneros, J. Pratt, E. Yoshida, S. Ivaldi, and D. Pucci.  
291 Teleoperation of humanoid robots: A survey. [IEEE Transactions on Robotics](#), 39(3):1706–  
292 1727, 2023.
- 293 [3] B. Dynamics. Atlas® and beyond: the world’s most dynamic robots, 2024. URL <https://bostondynamics.com/atlas/>.
- 294 [4] Unitree. Humanoid agent ai avatar, 2024. URL <https://www.unitree.com/g1>.
- 295 [5] F. AI. Figure 02 humanoid robot, 2024. URL <https://www.figure.ai/>.
- 296 [6] E. Krotkov, D. Hackett, L. D. Jackel, M. Perschbacher, J. Pippine, J. Strauss, G. A. Pratt,  
297 and C. Orlowski. The DARPA robotics challenge finals: Results and perspectives. [J. Field](#)  
298 [Robotics](#), 34(2):229–240, 2017. doi:[10.1002/ROB.21683](https://doi.org/10.1002/ROB.21683). URL <https://doi.org/10.1002/rob.21683>.
- 300 [7] C. Lu, X. Cheng, J. Li, S. Yang, M. Ji, C. Yuan, G. Yang, S. Yi, and X. Wang. Mobile-  
301 television: Predictive motion priors for humanoid whole-body control. [arXiv preprint](#)  
302 [arXiv:2412.07773](https://arxiv.org/abs/2412.07773), 2024.
- 304 [8] Q. Ben, F. Jia, J. Zeng, J. Dong, D. Lin, and J. Pang. HOMIE: humanoid loco-manipulation  
305 with isomorphic exoskeleton cockpit. [CorR](#), abs/2502.13013, 2025. doi:[10.48550/ARXIV.2502.13013](https://doi.org/10.48550/ARXIV.2502.13013). URL <https://doi.org/10.48550/arXiv.2502.13013>.
- 306 [9] M. Ji, X. Peng, F. Liu, J. Li, G. Yang, X. Cheng, and X. Wang. Exbody2: Advanced expressive  
307 humanoid whole-body control. [arXiv preprint arXiv:2412.13196](#), 2024.
- 309 [10] X. Cheng, Y. Ji, J. Chen, R. Yang, G. Yang, and X. Wang. Expressive whole-body control for  
310 humanoid robots. [arXiv preprint arXiv:2402.16796](#), 2024.
- 312 [11] M. Sombolestian and Q. Nguyen. Hierarchical adaptive loco-manipulation control for  
313 quadruped robots. In [2023 IEEE International Conference on Robotics and Automation](#)  
314 ([ICRA](#)), pages 12156–12162. IEEE, 2023.
- 316 [12] M. Sombolestian and Q. Nguyen. Adaptive force-based control of dynamic legged locomotion  
317 over uneven terrain. [IEEE Transactions on Robotics](#), 2024.
- 319 [13] N. Fey, G. B. Margolis, M. Peticco, and P. Agrawal. Bridging the sim-to-real gap for athletic  
320 loco-manipulation. [arXiv preprint arXiv:2502.10894](#), 2025.
- 322 [14] T. He, W. Xiao, T. Lin, Z. Luo, Z. Xu, Z. Jiang, C. Liu, G. Shi, X. Wang, L. Fan, and  
323 Y. Zhu. Hover: Versatile neural whole-body controller for humanoid robots. [arXiv preprint](#)  
324 [arXiv:2410.21229](https://arxiv.org/abs/2410.21229), 2024.
- 326 [15] T. He, Z. Luo, W. Xiao, C. Zhang, K. Kitani, C. Liu, and G. Shi. Learning human-to-humanoid  
327 real-time whole-body teleoperation. [arXiv preprint arXiv:2403.04436](#), 2024.
- 329 [16] T. He, Z. Luo, X. He, W. Xiao, C. Zhang, W. Zhang, K. Kitani, C. Liu, and G. Shi. Omnih2o:  
330 Universal and dexterous human-to-humanoid whole-body teleoperation and learning. [arXiv](#)  
331 preprint [arXiv:2406.08858](https://arxiv.org/abs/2406.08858), 2024.
- 333 [17] T. He, J. Gao, W. Xiao, Y. Zhang, Z. Wang, J. Wang, Z. Luo, G. He, N. Sobanbab, C. Pan,  
334 et al. Asap: Aligning simulation and real-world physics for learning agile humanoid whole-  
335 body skills. [arXiv preprint arXiv:2502.01143](#), 2025.

- 329 [18] Z. Zhuang and H. Zhao. Embrace collisions: Humanoid shadowing for deployable contact-  
 330 agnostics motions. [arXiv preprint arXiv:2502.01465](#), 2025.
- 331 [19] Whole body model predictive control with a memory of motion: Experiments on a torque-  
 332 controlled talos. In [2021 IEEE International Conference on Robotics and Automation \(ICRA\)](#),  
 333 pages 8202–8208. IEEE, 2021.
- 334 [20] J. Li, J. Ma, O. Kolt, M. Shah, and Q. Nguyen. Dynamic loco-manipulation on hector: Hu-  
 335 manoid for enhanced control and open-source research. [arXiv preprint arXiv:2312.11868](#),  
 336 2023.
- 337 [21] M. Murooka, K. Chappellet, A. Tanguy, M. Benallegue, I. Kumagai, M. Morisawa, F. Kane-  
 338 hiro, and A. Kheddar. Humanoid loco-manipulations pattern generation and stabilization con-  
 339 trol. [IEEE Robotics and Automation Letters](#), 6(3):5597–5604, 2021.
- 340 [22] K. Bouyarmane, K. Chappellet, J. Vaillant, and A. Kheddar. Quadratic programming for mul-  
 341 tirobot and task-space force control. [IEEE Transactions on Robotics](#), 35(1):64–77, 2018.
- 342 [23] J. Di Carlo, P. M. Wensing, B. Katz, G. Bledt, and S. Kim. Dynamic locomotion in the  
 343 mit cheetah 3 through convex model-predictive control. In [2018 IEEE/RSJ international](#)  
 344 [conference on intelligent robots and systems \(IROS\)](#), pages 1–9. IEEE, 2018.
- 345 [24] S. Kajita, F. Kaneko, K. Kaneko, K. Fujiwara, K. Harada, K. Yokoi, and H. Hirukawa. Biped  
 346 walking pattern generation by using preview control of zero-moment point. In [2003 IEEE](#)  
 347 [international conference on robotics and automation \(Cat. No. 03CH37422\)](#), volume 2, pages  
 348 1620–1626. IEEE, 2003.
- 349 [25] Z. Fu, Q. Zhao, Q. Wu, G. Wetzstein, and C. Finn. Humanplus: Humanoid shadowing and  
 350 imitation from humans. In [Conference on Robot Learning \(CoRL\)](#), 2024.
- 351 [26] Y. Zhang, T. Liang, Z. Chen, Y. Ze, and H. Xu. Catch it! learning to catch in flight with mobile  
 352 dexterous hands. [arXiv preprint arXiv:2409.10319](#), 2024.
- 353 [27] J. Dao, H. Duan, and A. Fern. Sim-to-real learning for humanoid box loco-manipulation.  
 354 In [2024 IEEE International Conference on Robotics and Automation \(ICRA\)](#), pages 16930–  
 355 16936. IEEE, 2024.
- 356 [28] F. Liu, Z. Gu, Y. Cai, Z. Zhou, S. Zhao, H. Jung, S. Ha, Y. Chen, D. Xu, and Y. Zhao.  
 357 Opt2skill: Imitating dynamically-feasible whole-body trajectories for versatile humanoid loco-  
 358 manipulation. [arXiv preprint arXiv:2409.20514](#), 2024.
- 359 [29] S. Gronauer and K. Diepold. Multi-agent deep reinforcement learning: a survey. [Artificial](#)  
 360 [Intelligence Review](#), 55(2):895–943, 2022.
- 361 [30] K. Zhang, Z. Yang, and T. Başar. Multi-agent reinforcement learning: A selective overview  
 362 of theories and algorithms. [Handbook of reinforcement learning and control](#), pages 321–384,  
 363 2021.
- 364 [31] Y. Wang, R. Sagawa, and Y. Yoshiyasu. Learning advanced locomotion for quadrupedal robots:  
 365 A distributed multi-agent reinforcement learning framework with riemannian motion policies.  
 366 [Robotics](#), 13(6):86, 2024.
- 367 [32] A. Rigo, M. Hu, S. K. Gupta, and Q. Nguyen. Hierarchical optimization-based control for  
 368 whole-body loco-manipulation of heavy objects. In [2024 IEEE International Conference on](#)  
 369 [Robotics and Automation \(ICRA\)](#), pages 15322–15328. IEEE, 2024.
- 370 [33] T. Portela, G. B. Margolis, Y. Ji, and P. Agrawal. Learning force control for legged manip-  
 371 ulation. In [2024 IEEE International Conference on Robotics and Automation \(ICRA\)](#), pages  
 372 15366–15372. IEEE, 2024.

- 373 [34] J. Li and Q. Nguyen. Kinodynamics-based pose optimization for humanoid loco-manipulation.  
374 [arXiv preprint arXiv:2303.04985](#), 2023.
- 375 [35] T. Mattioli and M. Vendittelli. Interaction force reconstruction for humanoid robots. [IEEE](#)  
376 [Robotics and Automation Letters](#), 2(1):282–289, 2016.
- 377 [36] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov. Proximal policy optimization  
378 algorithms. [CoRR](#), abs/1707.06347, 2017. URL <http://arxiv.org/abs/1707.06347>.
- 380 [37] N. Mahmood, N. Ghorbani, N. F. Troje, G. Pons-Moll, and M. J. Black. AMASS: Archive  
381 of motion capture as surface shapes. In [International Conference on Computer Vision](#), pages  
382 5442–5451, Oct. 2019.
- 383 [38] K. W. Ng, G.-L. Tian, and M.-L. Tang. Dirichlet and related distributions: Theory, methods  
384 and applications. 2011.
- 385 [39] Advanced Computing Center for the Arts and Design. ACCAD MoCap Dataset. URL <https://accad.osu.edu/research/motion-lab/mocap-system-and-data>.
- 387 [40] B. Wen, W. Yang, J. Kautz, and S. Birchfield. Foundationpose: Unified 6d pose estimation and  
388 tracking of novel objects. In [Proceedings of the IEEE/CVF Conference on Computer Vision](#)  
389 and Pattern Recognition
- 390 [41] E. Todorov, T. Erez, and Y. Tassa. Mujoco: A physics engine for model-based control. In  
391 [2012 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2012,](#)  
392 Vilamoura, Algarve, Portugal, October 7-12, 2012, pages 5026–5033. IEEE, 2012. doi:[10.1109/IROS.2012.6386109](https://doi.org/10.1109/IROS.2012.6386109). URL <https://doi.org/10.1109/IROS.2012.6386109>.

394 **A Appendix**

395 **A.1 Reward Terms**

396 We adopt the similar reward terms from [16, 10], but introduce some important penalties to ensure  
 397 the locomotion stability under significant external forces, and other tracking rewards for squat and  
 398 waist twist. The additional reward terms are summarized in Table 3:

Table 3: Additional Reward components and weights: penalty rewards for preventing undesired behaviors for sim-to-real transfer, and task rewards to achieve desired loco-manipulation capability.

Term	Expression	Weight
Penalty		
Hip pos	$\ q_{roll,pitch}^{hip}\ $	-2.5
Negative knee joint	$\sum_j \mathbb{1}[q_j < q_j^{\min}]$	-1.0
Stance tap feet	$ (\mathbf{p}_{left\_foot} - \mathbf{p}_{right\_foot})_x $ in base frame	-5.0
Stance root	$ (\mathbf{p}_{root} - \text{mid}(\mathbf{p}_{feet}))^y $	-5.0
Stand still	$\mathbb{1}[\text{no contact}]$	-0.15
Ankle roll	$\sum_j  q_j^{roll} $	-2.0
Task Reward		
Root linear velocity x	$\exp(-4.0\ \mathbf{v}_t^x - \mathbf{v}_t^{x*}\ _2)$	2
Root linear velocity y	$\exp(-4.0\ \mathbf{v}_t^x - \mathbf{v}_t^{x*}\ _2)$	1.5
Root angular velocity	$\exp(-4.0\ \mathbf{v}_t^{ang} - \mathbf{v}_t^{ang*}\ _2)$	4
Root walk height	$\exp\left(-\frac{ \text{command}_z - p_z^{\text{root}} }{0.05}\right)$	2
Waist dofs	$\exp\left(-\frac{\sum_{\theta \in \{\text{yaw, roll, pitch}\}} (\theta^{\text{sim}} - \theta^{\text{cmd}})^2}{0.05}\right)$	2
Upper body dofs	$\exp\left(-\frac{\ \mathbf{q}_{upper} - \mathbf{q}_{ref}\ _2^2}{0.01}\right)$	4

399 **A.2 Domain Randomization**

400 We apply the following domain randomization terms during training, which are important for suc-  
 401 cessful sim-to-real transfer.

Table 4: Domain randomization terms including dynamics randomization and external perturbation.

Term	Value
<b>Dynamics Randomization</b>	
Friction	$\mathcal{U}(0.5, 1.25)$
Link mass	$\mathcal{U}(0.9, 1.2) \times \text{default kg}$
Base mass	$\mathcal{U}(-1.0, 3.0) \text{ kg}$
P Gain	$\mathcal{U}(0.9, 1.1) \times \text{default}$
D Gain	$\mathcal{U}(0.9, 1.1) \times \text{default}$
Control delay	$\mathcal{U}(0, 20)\text{ms}$
<b>External Perturbation</b>	
Push robot	interval = 5s, $v_{xy} = 1\text{m/s}$

402 **A.3 Force Curriculum**

403 To evaluate the effectiveness of the proposed torque-limit-aware force curriculum introduced in  
 404 Section 3, we compare it against a baseline that applies randomly sampled forces within a large  
 405 clipping range ( $X : [-100 \text{ N}, 100 \text{ N}]$ ,  $Y : [-100 \text{ N}, 100 \text{ N}]$ ,  $Z : [-100 \text{ N}, 5 \text{ N}]$ ), without considering  
 406 torque feasibility. The training curves and quantitative comparisons are presented in ?? and Table 5.  
 407 During evaluation, forces are still constrained within the estimated maximum admissible limits.

408 ?? (a) shows that force curriculum saturates at a force scale  $\alpha_g = 0.6$  due to frequent violations of  
 409 torque limits, which hinder further progression. Additionally, as illustrated in ?? (b), the force cur-

Methods	$E_{\text{tracking}}^{\text{upper}} \downarrow$			$E_{\text{tracking}}^{\text{root}} \downarrow$		
	N-Force	M-Force	L-Force	N-Force	M-Force	L-Force
w/o-Torque-Limit-Aware.	$0.30 \pm 0.07$	$0.39 \pm 0.08$	$0.81 \pm 0.10$	$0.07 \pm 0.03$	$0.11 \pm 0.05$	$0.27 \pm 0.05$
<b>FALCON</b> with-Torque-Limit-Aware.	$0.17 \pm 0.03$	$0.21 \pm 0.03$	$0.38 \pm 0.04$	$0.05 \pm 0.03$	$0.08 \pm 0.03$	$0.19 \pm 0.02$

Table 5: Evaluation of FALCON using torque-limit-aware (Max-Force-Estimation) curriculum versus w/o torque-limit-aware force curriculum in IsaacGym. Our curriculum achieves significantly better tracking performance, especially for upper-body manipulation under large forces.

riculum w/o torque-limit-aware results in larger upper-body tracking errors, since excessive forces regularly exceed the feasible torque bounds, impairing the learning of effective upper body force compensation. Consequently, as shown in Table 5, policies trained w/o torque-limit-aware force curriculum tend to overfit to the locomotion objective, compromising upper-body accuracy. In contrast, our torque-limit-aware curriculum facilitates balanced learning of both upper-body joint tracking and root velocity tracking under significant external disturbances.

Note that in Table 1, we apply a smaller force clipping range ( $X : [-50, N, 50, N]$ ,  $Y : [-50, N, 50, N]$ ,  $Z : [-60, N, 5, N]$ ) compared to the wider range used in Table 5 ( $X : [-100, N, 100, N]$ ,  $Y : [-100, N, 100, N]$ ,  $Z : [-100, N, 5, N]$ ). The results suggest that expanding the force range can further improve locomotion performance while maintaining the upper body joint tracking accuracy, as shown in Table 6

Methods	$E_{\text{tracking}}^{\text{upper}} \downarrow$			$E_{\text{tracking}}^{\text{root}} \downarrow$		
	N-Force	M-Force	L-Force	N-Force	M-Force	L-Force
<b>FALCON</b> Smaller-Force-Clip.	$0.21 \pm 0.03$	$0.24 \pm 0.03$	$0.30 \pm 0.04$	$0.27 \pm 0.03$	$0.30 \pm 0.03$	$0.45 \pm 0.04$
Larger-Force-Clip.	$0.17 \pm 0.03$	$0.21 \pm 0.03$	$0.38 \pm 0.04$	$0.05 \pm 0.03$	$0.08 \pm 0.03$	$0.19 \pm 0.02$

Table 6: Evaluation of FALCON with a smaller force clip range in Table 1 versus a larger force clip range in Table 5.

#### A.4 Lower-RL-Upper-IK with Force Estimator

We jointly train a 3D force estimator, following a similar approach to [33], using the robot’s proprioception as input  $s_t^p \triangleq [q_{t-4:t}, \dot{q}_{t-4:t}, \omega_{t-4:t}^{\text{root}}, g_{t-4:t}, a_{t-5:t-1}]$ . As illustrated in Figure 6 (a), the estimator predicts the end-effector forces  $\tilde{F}_t^{\text{ee}}$ , which are then concatenated with full-body proprioception and fed into the lower-body RL policy. Meanwhile, the upper-body joint torques with force compensation are computed as  $\tau = K_p(q_t^{\text{upper}} - q_t^{\text{upper}*}) + K_d\dot{q}_t^{\text{upper}} + J_{EE}^T \tilde{F}_t^{\text{ee}}$ .

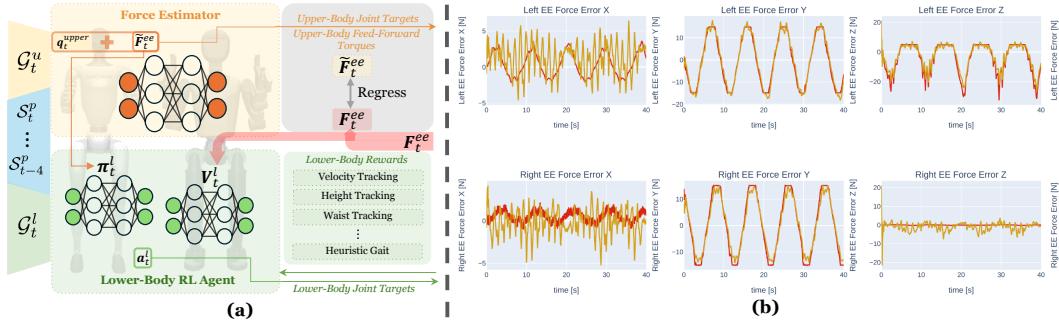


Figure 6: (a) Framework of Lower-RL-Upper-IK with Learning-based Force Estimator; (b) Force Estimator Results

We compare the estimated and applied forces in Figure 6 (b), showing close alignment between the two. However, even with accurate force estimates, changes in the contact point on the end-effector

429 alter the effective Jacobian  $\mathbf{J}'_{EE}^T$ , making the compensation term  $\mathbf{J}_{EE}^T \tilde{\mathbf{F}}_t^{ee}$  inaccurate. Therefore,  
 430 a force sensor is still necessary during deployment to localize the force application and compute the  
 431 correct  $\mathbf{J}'_{EE}^T$ . Moreover, the compensation assumes quasi-static conditions, introducing additional  
 432 error when upper-body joints are moving.

#### 433 A.5 Force Measurement

434 Here, we use Mxmoonfree-Digital-500N-Force-Gauge to measure the peak forces needed for the  
 435 following four force-adaptive tasks: (1) **Cart-Pulling** for Booster T1 with a Unitree G1 and a Uni-  
 436 tree H1 in the cart; (2) **Door-Opening**; (3) **Stance-Pulling** for the Unitree G1; (4) **Stance-Pulling**  
 437 for the Booster T1.

438 Here, **Stance-Pulling** refers to applying lateral forces along the X-Y plane while the robot maintains  
 439 a static stance, and measuring the maximum force it can resist without losing balance. Notably, the  
 440 Booster T1 demonstrates a higher peak resistive force compared to the Booster T1, primarily due to  
 its lower center of mass (CoM), which contributes to a greater moment arm for lateral resistance.



(a) Cart-Pulling (peak: 107.9 N)



(b) Door-Opening (peak: 47.3 N)



(c) Stance-Pulling: Unitree G1 (peak: 45.3 N)



(d) Stance-Pulling: Booster T1 (peak: 66.3 N)

Figure 7: Maximum force readings captured during different force-adaptive tasks using a handheld force gauge. Subfigures (a)–(d) show peak force values during individual tasks.

442 **A.6 Autonomy Pipeline**

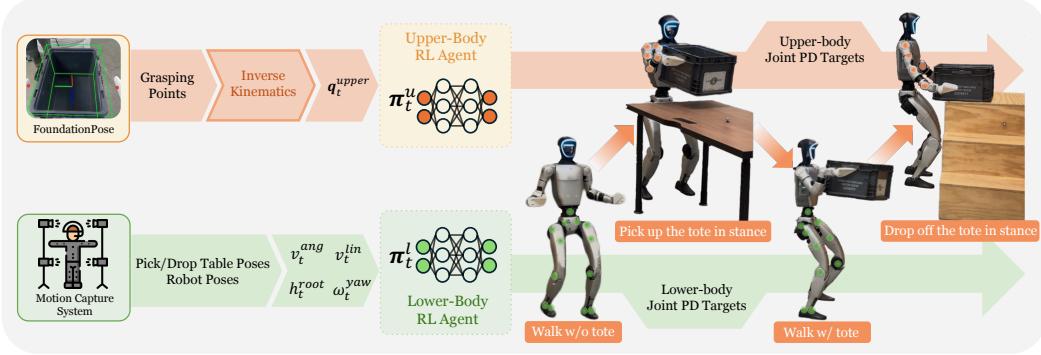


Figure 8: Overview of the autonomy pipeline for FALCON. The system integrates FALCON with 6-DoF object pose estimation via FoundationPose, MoCap-based global localization, and inverse kinematics for grasp planning to enable a humanoid robot to perform tote logistics tasks: walk without a tote, pick up the tote, walk with the tote, and drop off the tote.

443 We devise an end-to-end autonomous pipeline for the tote manipulation task which consists of the  
 444 Motion Capture System (Mocap) used to retrieve positions of objects of interest. These global pose  
 445 estimates serve as the foundation for spatial reasoning and task execution. The robot’s behavior is  
 446 governed by a state-machine-based control framework, which defines a sequence of discrete operational  
 447 states. As shown in Fig. 8 the state-machine includes 4 states: (1) walk without tote, (2) pick  
 448 up the tote in stance (3) walk with tote (4) drop off the tote in stance. This modular control strat-  
 449 egy enables robust and interpretable execution of complex manipulation behaviors in a structured  
 450 and repeatable manner. To estimate the pose of the industrial tote relative to the camera frame, we  
 451 employ a state-of-the-art pose estimation method, FoundationPose [40] that provides accurate and  
 452 reliable 6-DoF pose predictions.

453 **A.6.1 Perception - Pose Estimation**

454 To set up the FoundationPose pipeline, we require a few priors, including a high-fidelity 3D model  
 455 of the industrial tote. To obtain these, we begin by acquiring a raw 3D scan of the tote, which is  
 456 then imported into a 3D modeling software for manual post-processing and refinement. From this  
 457 process, we generate a high-quality 3D mesh in .obj format, along with an associated .png texture  
 458 file to preserve appearance information. These assets serve as critical inputs to the FoundationPose  
 459 [40] framework, which we use to perform pose inference on the image stream from the G1 robot,  
 460 ultimately estimating the six-degree-of-freedom (6-DoF) pose of the object relative to the camera.

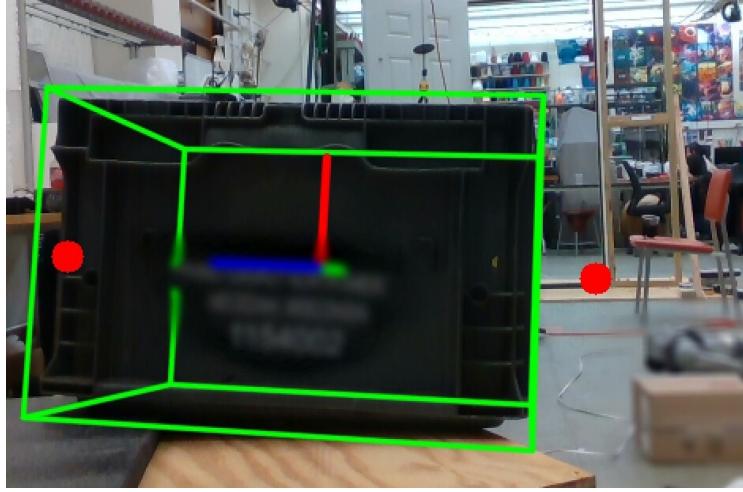


Figure 9: Industrial tote and the manipulation approach points shown in red.

461 In addition to pose estimation, we identify task-relevant points of interest on the tote, such as design-  
 462 nated grasp positions located laterally outward from the tote’s surfaces. These task-relevant points  
 463 (shown in Fig. 9) are subsequently transformed into the robot’s base frame using the extrinsic cali-  
 464 bration parameters between the camera and the robot base.

#### 465 **A.6.2 Motion Capture System**

466 The Motion Capture System is used to obtain precise six-degree-of-freedom (6-DoF) pose informa-  
 467 tion, comprising position ( $x$ ,  $y$ ,  $z$ ) and orientation (yaw, pitch, roll) for three key objects within the  
 468 global reference frame: the robot base, the pickup table, and the drop table. This global localiza-  
 469 tion capability ensures consistent spatial awareness across all components of the system.

470 Upon reaching the ”(2) Pick up the tote in stance” state—triggered by the state machine, which is  
 471 itself governed by the motion capture system, as illustrated in Fig. 8—FoundationPose is run in real-  
 472 time. The grasp positions are then supplied to an inverse kinematics (IK) solver. The IK problem  
 473 is formulated as a go-to-pose task. Once in position, the end effectors engage in a synchronized  
 474 squeezing motion to establish a secure, two-handed grip on the tote. The robot then proceeds to lift  
 475 the tote vertically while maintaining a solid grasp on the tote.

#### 476 **A.7 Hardware Limits in Real-World Deployment**

477 During sim-to-real deployment of the policy trained with FALCON, we observe that the humanoid  
 478 robot struggles to sustain high joint torques over extended periods, often leading to rapid motor over-  
 479 heating—particularly at the wrists. This significantly limits our ability to perform payload transport  
 480 exceeding 2kg per arm. In contrast, as shown in Figure 10, the same policy evaluated in simula-  
 481 tion using MuJoCo [41], with torque clipping to respect joint limits but without modeling thermal  
 482 constraints, successfully transports payloads over 4kg per end-effector. This highlights a key gap  
 483 between simulated and real-world actuator endurance.

484 However, for heavy-duty tasks such as cart-pulling—which require only brief bursts of high  
 485 torque—the motors are less prone to overheating, as sustained high torque output is not necessary.  
 486 This enables the robot to successfully perform such tasks in the real world.

figures/appendix/mujoco.pdf

Figure 10: Transporting 4kg Payloads in Mujoco