组会分享 UniPolicy

2025.11.01 李佩泽

Learning a Unified Policy for Position and Force Control in Legged Loco-Manipulation

Peiyuan Zhi[1,2,*], **Peiyang Li**[1,3,*], **Jianqin Yin**[3], **Baoxiong Jia**[1,2,†], **Siyuan Huang**[1,2,†]
[1] State Key Laboratory of General Artificial Intelligence, BIGAI
[2] Joint Laboratory of Embodied AI and Humanoid Robots, BIGAI & UniTree Robotics
[3] Beijing University of Posts and Telecommunications
https://unified-force.github.io/
[*] Equal contribution. [†] Corresponding authors.

Abstract: Robotic loco-manipulation often involves contact-rich interactions with the environment, requiring the joint modeling of contact force and robot position. However, recent visuomotor policies often focus solely on learning position or force control, overlooking their co-learning. We propose the first unified policy for legged robots that jointly models force and position control learned without relying on force sensors. By simulating diverse combinations of position and force commands alongside external disturbance forces, we use reinforcement learning to learn a policy that estimates forces from historical robot states and compensates for them through position and velocity adjustments. This policy enables a wide range of manipulation behaviors under varying force and position inputs, including position tracking, force application, force tracking, and compliant interactions. Moreover, we demonstrate that the learned policy enhances trajectory-based imitation learning pipelines by incorporating essential contact information through its force estimation module, achieving approximately ∼39.5% higher success rates in four challenging contact-rich manipulation tasks over position-control policies. Experiments on both a quadrupedal manipulator and a humanoid robot validate the versatility and robustness of the proposed policy in diverse scenarios.

Keywords: Unified Force and Position Control, Force-aware Imitation Learning

Research Gap

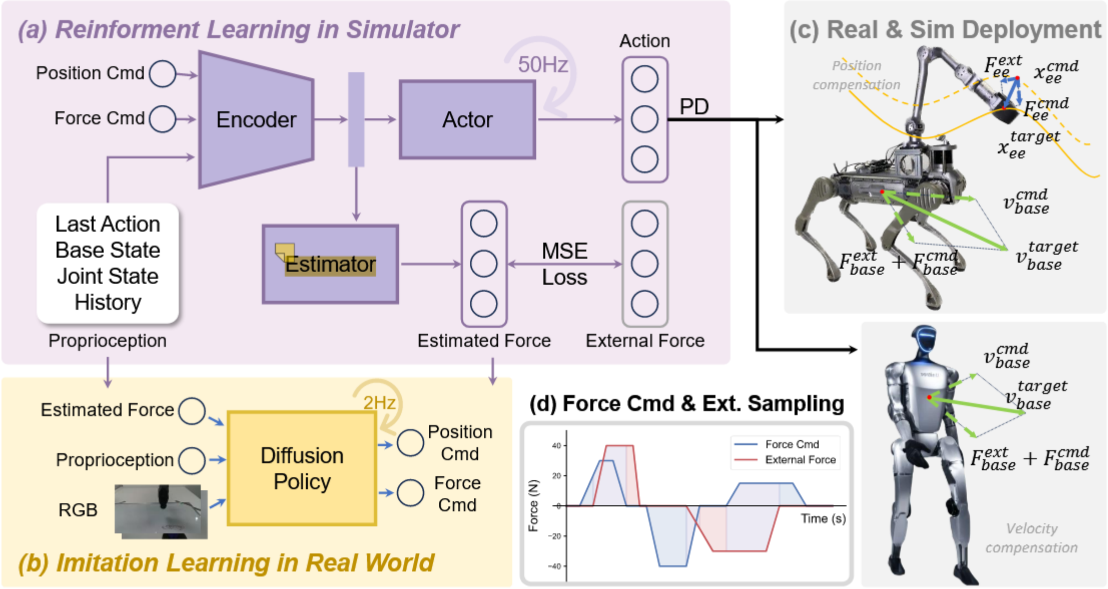
To tackle the control challenge of legged manipulators, reinforcement learning (RL) algorithms have emerged as effective alternatives to traditional control methods, offering robust and generalizable policies trained through domain randomization [2, 6, 7, 8, 9, 10]. These policies integrate locomotion and manipulation in complex tasks but primarily depend on precise position control, limiting their applicability in contact-rich scenarios. This reliance has also driven the rise of position-based robot imitation learning [11, 12, 13, 14, 15], with large datasets [13, 16, 17, 18] focused solely on robot trajectories, omitting crucial contact information due to the lack of force sensing. As shown in Section 4.2, such trajectory-only data is insufficient for training effective policies, even for basic contact-rich tasks (*e.g.*, wiping a blackboard). This underscores the limitation of position control and emphasizes the necessity of integrating force sensing and modeling into learning-based policies for more effective task execution.

Contribution

In light of the aforementioned challenges and observations, we propose **the first unified policy for legged robots that seamlessly integrates force and position control without the need for force sensors.** Unlike previous methods [9] that handle force and position control independently, we train

1. We propose the first model for learning unified force and position control in legged loco-manipulation, enabling diverse control behaviors such as position tracking, force control, and compliance with a single policy.
2. Through 7 experiments on a quadrupedal manipulator and a humanoid robot, we demonstrate the effectiveness and robustness of our learned policy across diverse and challenging task scenarios.
3. We develop a force-aware robot imitation learning data collection pipeline using our learned force estimator, improving position-based imitation learning baselines by $\sim$39.5% on three challenging contact-rich manipulation tasks, highlighting our policy's promise as a general and efficient framework for contact-rich task demonstration curation.

Method



Impedance Formulation

$$\boldsymbol{F} = K(\mathbf{x} - \mathbf{x}^{\mathrm{des}}) + D(\dot{\mathbf{x}} - \dot{\mathbf{x}}^{\mathrm{des}}) + M(\ddot{\mathbf{x}} - \ddot{\mathbf{x}}^{\mathrm{des}}), \tag{1}$$

$$\mathbf{x}^{\mathrm{target}} = \mathbf{x}^{\mathrm{cmd}} + \frac{\boldsymbol{F}^{\mathrm{ext}} + (\boldsymbol{F}^{\mathrm{cmd}} - \boldsymbol{F}^{\mathrm{react}})}{K}, \tag{2}$$

$$\mathbf{v}^{\mathrm{target}}_{\mathrm{base}} = \mathbf{v}^{\mathrm{cmd}}_{\mathrm{base}} + \frac{\boldsymbol{F}^{\mathrm{ext}}_{\mathrm{base}} + (\boldsymbol{F}^{\mathrm{cmd}}_{\mathrm{base}} - \boldsymbol{F}^{\mathrm{react}}_{\mathrm{base}})}{D}. \tag{4}$$

Training Policy

$$\mathbf{o}_t = [\mathbf{g}^{\mathrm{base}}_t, \omega^{\mathrm{base}}_t, \mathbf{q}_t, \dot{\mathbf{q}}_t, \mathbf{a}_{t-1}, \mathbf{c}^{\mathrm{cmd}}_t, \theta^{\mathrm{feet}}_t] \tag{5}$$

Policy Design We provide an overview of our policy model in Fig. 2(a). Our policy model comprises three modules: the observation encoder, the state estimator, and the actor. The encoder processes the observation history $\mathbf{o}_{[t-H,\dots,t-1,t]}$ $(H = 32)$ and outputs a latent feature, which is then sent to the state estimator and the actor. The state estimator then predicts the robot's state, including the external force $\boldsymbol{F} = \boldsymbol{F}^{\mathrm{ext}} + \boldsymbol{F}^{\mathrm{react}}$, the end-effector position, and the base velocity. This estimated force could then be translated to command signals in certain desired control behaviors.

Experiment

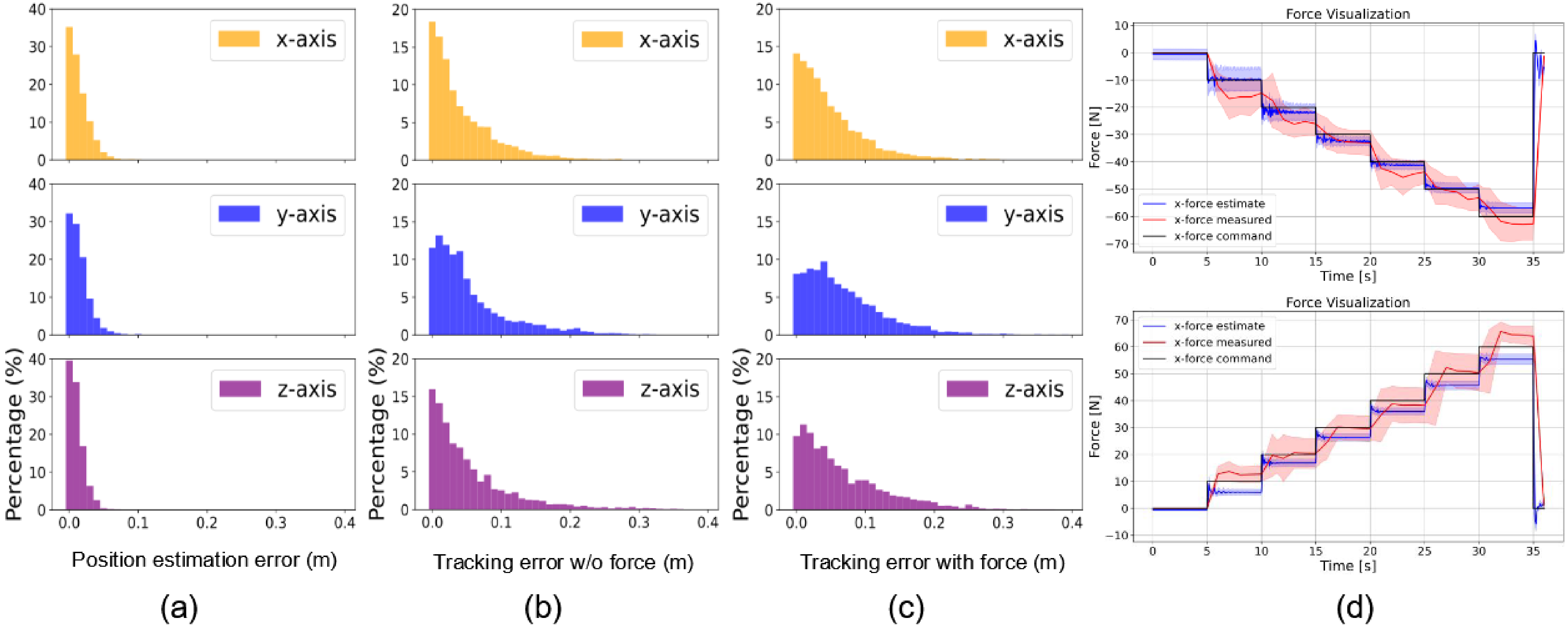


Figure 3: **Force and position control evaluation.** (a)–(c) Evaluation of force and position control tracking errors in simulation environments. (d) Real-world evaluation of force control, shaded areas indicate variance measured across 5 different end-effector positions.

Experiment

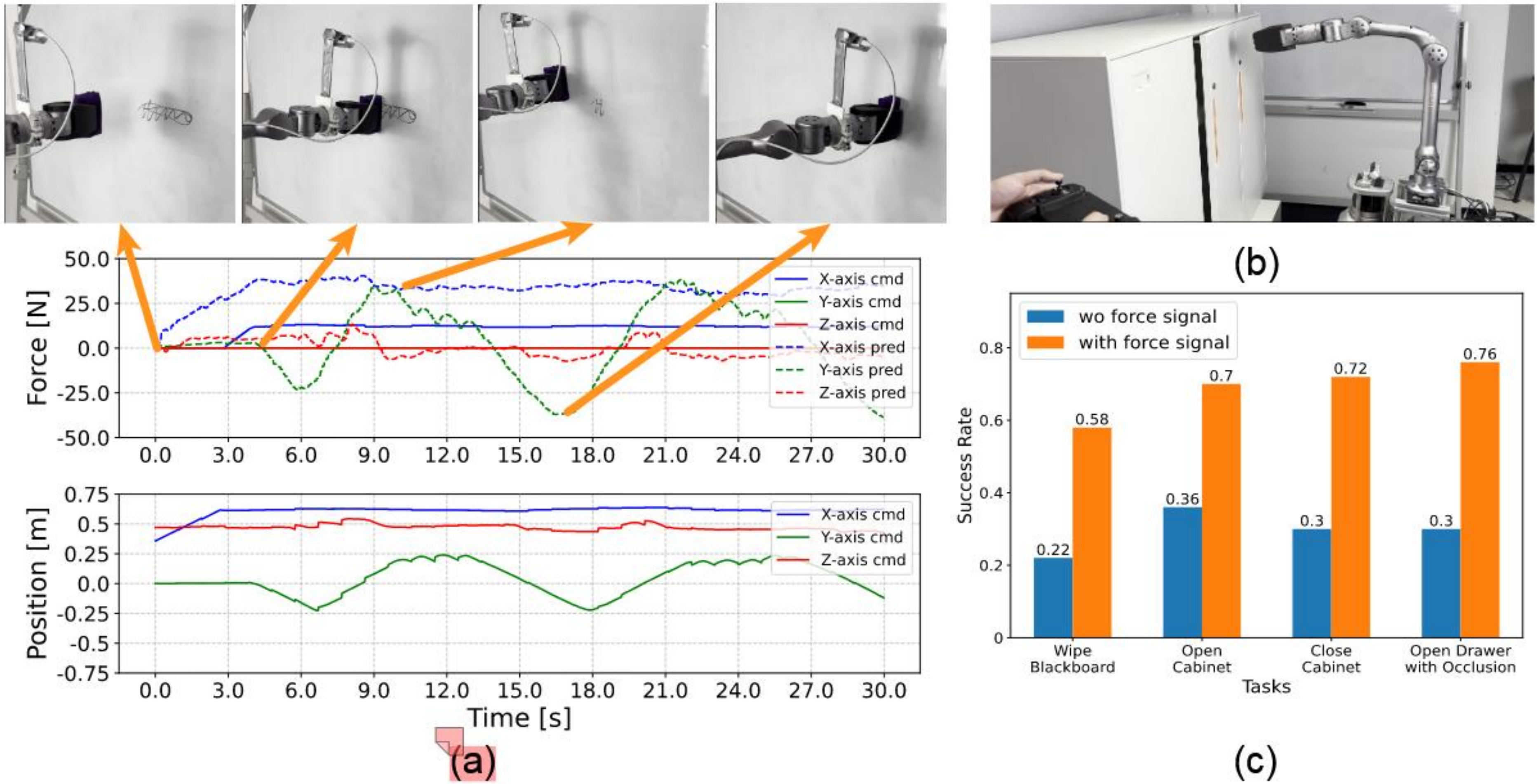


Figure 4: **Force-aware imitation learning.** (a) Time-series outputs of position and force commands to the trained force-aware imitation policy in the `wipe-blackboard` task. *cmd* denotes the output of the imitation learning policy, while *pred* indicates the external force estimated by the low-level policy. (b) A visualization of the data collection process. (c) The performance comparison between our policy and a baseline vision-only policy over 50 trials across four tasks.

Experiment

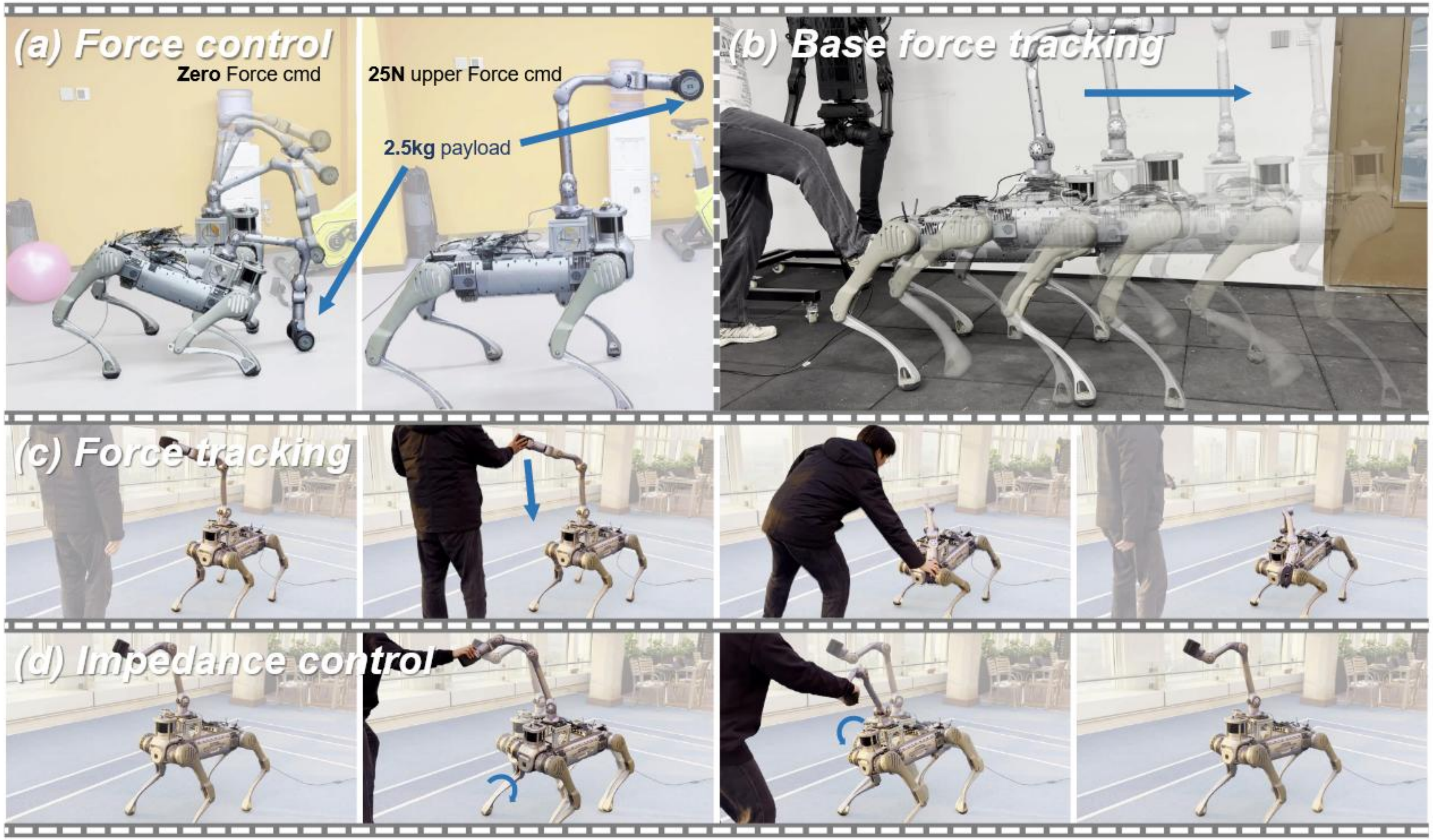


Figure 5: **Diverse skills facilitated by our policy.** (a) Force control: The robot counteracts gravity to support a payload when given a $25N$ force command. (d) Base force tracking: The robot responds compliantly to pushes on its base, enabling intuitive human guidance. (c) Force tracking: The robot tracks a zero-force command by minimizing external force interactions. (d) Impedance control: The robot adjusts its whole-body posture to counteract and comply with external disturbances.

Conclusion

We propose a unified force-position control policy for legged robots, enabling contact-rich loco-manipulation tasks without explicit force sensors. Using reinforcement learning, our policy estimates external forces from historical states and compensates for them through position and velocity adjustments. This approach supports diverse behaviors like position tracking, force application, and compliance. Additionally, integrating force estimation into imitation learning improves task success in contact-rich environments. Experiments on quadrupedal and humanoid robots validate the policy's adaptability and robustness in real-world scenarios.

Limitations and Future Work

First, while the policy successfully estimates external forces without direct force sensing, its accuracy tends to degrade in high-frequency interactions and at the edges of the robot's workspace. Future work could focus on improving force estimation in these corner cases. One possible direction is to incorporate velocity and acceleration terms from Eq. (2) to enhance force estimation, allowing the model to better capture dynamic interactions.

Second, while our policy generalizes well from simulation to real-world deployment, discrepancies remain due to the sim-to-real gap, particularly in force accuracy along different coordinate axes. These differences likely stem from mismatches in actuator dynamics and contact modeling between simulation and real hardware. Future work could explore techniques such as domain randomization and real-to-sim corrections to improve robustness across varying real-world conditions.

Additionally, our current framework primarily focuses on estimating force at a single interaction point. Future work could explore multi-point force estimation and whole-body force interaction tasks. For example, in scenarios such as a quadrupedal robot opening a heavy door, the robot could use its body to brace against the door while simultaneously using its manipulator to press down on the handle. Developing policies that coordinate multiple contact forces across different body parts could enable more complex and effective real-world interactions.