Learning Whole-Body Quadrupedal Pushing Across Geometry and Physics Variation

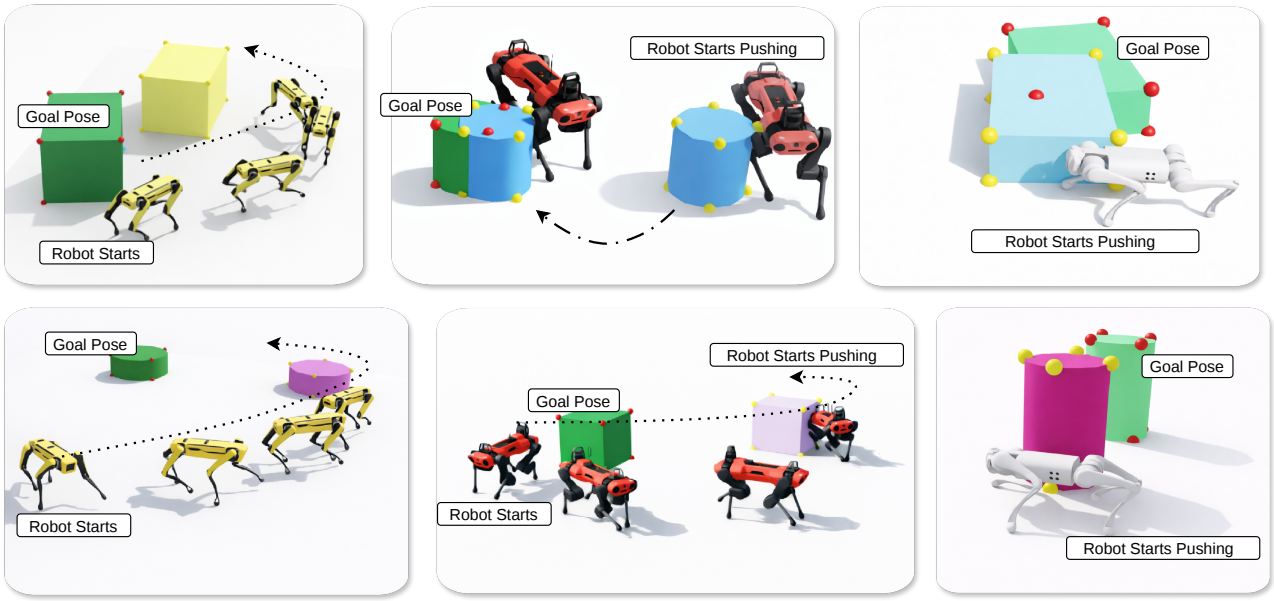


Fig. 1. **Whole-body pushing across robot embodiments and object geometries.** The learned hierarchical RL policy is shown on multiple robot platforms pushing diverse cuboid and cylindrical objects toward target poses. These examples illustrate that the same policy framework generalizes across different robot types and object shapes.

***Abstract*— Whole-body non-prehensile manipulation allows quadrupedal robots to reposition large objects through contact, but reliable performance remains difficult because pushing depends on object geometry, friction, mass, and intermittent contact dynamics. This paper presents a hierarchical reinforcement learning pipeline in IsaacLab for goal-conditioned whole-body quadrupedal pushing diverse cuboid and cylindrical objects. A high-level policy outputs planar velocity commands, while a frozen low-level locomotion controller provides stable execution. During training, the policy is conditioned on privileged object-context features available in simulation, allowing us to study what a strong simulation pipeline can achieve under substantial geometry and physics variation. We evaluate the learned policy under in-distribution conditions, out-of-distribution geometry, out-of-distribution physics, and on geometries not encountered during training. The results show that the privileged pipeline achieves robust whole-body pushing across objects. Project website: `https://quadrupedpushing-cr2.github.io/`.**

I. Introduction

Whole-body non-prehensile manipulation allows legged robots to reposition objects through contact, without requiring a grasp. Legged robots are well suited to object interaction in cluttered environments because mobility and manipulation can be combined within the same platform [1]. For quadrupedal robots, this capability is useful for tasks such as clearing a path, aligning a movable object, or pushing an obstacle to a target pose [2], [3]. At the same time, the problem is contact-rich: the outcome of a push depends on object geometry, friction, inertial properties, and intermittent contact. As a result, even tasks that appear similar can produce substantially different interaction dynamics, making reliable pushing difficult to model and challenging to generalize [4].

Recent reinforcement learning approaches have shown that quadrupedal robots can learn whole-body pushing behaviors in simulation [5], [6], [7], [8]. These results are encouraging, but they leave an important practical question open. Much of the current evidence comes from settings [9] where useful task or object information is available inside simulation, while less is known about how such policies scale across broader geometry and physics variation in modern high-throughput training environments. This matters especially in contact-rich manipulation, where performance depends not only on the robot reaching the object, but also on how it approaches, redirects, and settles the object near the goal [10].

In this work, we study whole-body quadrupedal pushing in IsaacLab [11] and focus on building a strong privileged training pipeline for this setting. Our controller follows a hierarchical structure: a high-level reinforcement learning policy outputs planar velocity commands, and a frozen low-level locomotion controller executes them. Within this framework,

we study goal-conditioned pushing across diverse cuboid and cylindrical objects under randomized geometry and physics conditions. In addition to the control architecture, we introduce a revised reward design that separates coarse transport progress from near-goal pose precision and stabilization. This yields a clearer training signal for long-horizon pushing and improves the alignment between training objectives and the final task.

Prior work has shown that learned whole-body pushing can be effective, particularly when simulation provides informative task or dynamics context [5], [6]. Building on this, we establish a systematic IsaacLab benchmark for privileged whole-body pushing across object families. We evaluate the learned policies under in-distribution conditions, out-of-distribution geometry, out-of-distribution physics, and unseen geometries. This benchmark provides a clear view of the robustness that privileged training can achieve in a scalable contact-rich manipulation setting.

An important next step is to remove privileged object information and infer it online from interaction history. This direction is closely related to prior work on adaptation in legged control [12] and online system identification for dynamics-varying policies [13]. Establishing a strong benchmark is a necessary first step: it defines the performance target, clarifies which aspects of the task remain difficult under broad variation, and provides the foundation on which future deployable adaptation methods can be evaluated.

Contributions.

- We present a hierarchical reinforcement learning pipeline in IsaacLab for whole-body quadrupedal pushing across diverse object geometries and randomized physics conditions.
- We introduce a reward design that combines transport progress, directional shaping, near-goal pose precision, and stabilization, improving training behavior and final placement quality.
- We provide a systematic evaluation of whole-body non-prehensile pushing under in-distribution, out-of-distribution geometry, out-of-distribution physics, and unseen-geometry settings.

II. Method

A. Task setting

We study goal-directed whole-body non-prehensile pushing with a quadrupedal robot in simulation. At each episode, the robot starts near a pushable object and must move it toward a target goal pose on the ground plane. The task is defined over multiple objects, with variation in both geometry and physical properties. We focus on whole-body interaction, in which the robot primarily uses body and base contact to transport the object, rather than grasping or specialized manipulation primitives.

B. Hierarchical whole-body controller

The controller is organized hierarchically. A high-level policy receives the current task observation and a latent context variable, and outputs a planar velocity command for pushing. A frozen low-level locomotion controller tracks this command using robot proprioception and produces joint targets for execution. This separation allows the high-level policy to focus on task geometry and interaction strategy, while the low-level controller provides stable locomotion.

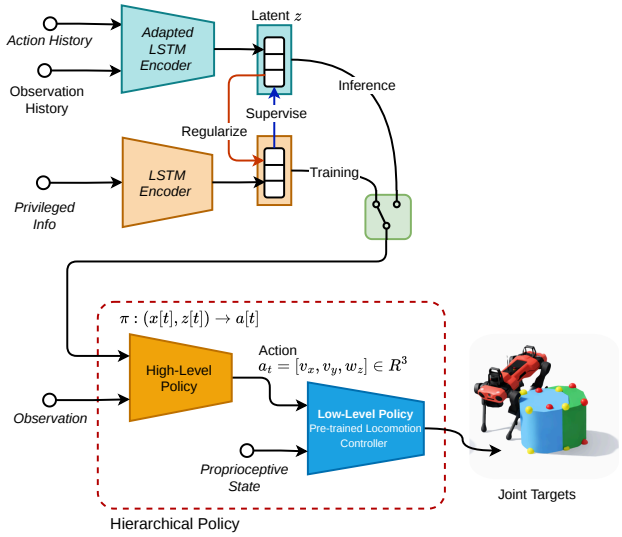


Fig. 2. **Privileged whole-body pushing pipeline.** A high-level policy receives robot and task observations together with privileged object-context input and produces planar velocity commands. A frozen low-level locomotion controller executes these commands.

The high-level action is a planar body-frame velocity command

$$\mathbf{a}_t = [v_x, v_y, \omega_z], \tag{1}$$

where v_x and v_y are linear velocities and ω_z is the yaw rate. This command is executed through a frozen low-level locomotion controller, which takes the robot's proprioceptive state as input and produces target commands for the 12 leg joints. The exact low-level controller differs slightly across the robot platforms used in this work, namely Spot, Anymal-C, and Unitree Go1, to account for embodiment-specific locomotion and actuation characteristics.

C. Privileged object-context input

During training, the high-level policy is conditioned on object-context features available in simulation, including latent physical properties that are difficult to observe directly from onboard sensing alone. These privileged features provide information about the object and its interaction dynamics, allowing the policy to adapt its behavior to variations in geometry and physics.

We use a teacher–student encoder module Regularized Online Adaptation (ROA) [14], with both encoders implemented as LSTMs [15]. During training, the teacher maps a history of privileged information to a latent context vector, while the student predicts the same latent from observation–action history alone. This provides a compact representation of object and environment variation.

TABLE I

POLICY OBSERVATIONS (HIGH-LEVEL).

Observation term	Dim.
Base linear velocity (v_{base})	3
Base angular velocity (ω_{base})	1
Projected gravity (g_{base})	3
Previous action (a_{t-1})	3
Robot–object relative vector (xy)	2
Object–goal relative vector in base frame (xy)	2
Object–goal distance (d_{bg})	1
Pushable heading in base frame	2
Goal heading in base frame	2
Object planar velocity in base frame (xy)	2
Speed toward goal ($v_{\parallel g}$)	1
Tangential speed ($v_{\perp g}$)	1
Low-level prior (frozen): locomotion-policy obs	48

TABLE II

REWARD SUMMARY.

Term	Definition	w
Core objective (task completion)		
Intrinsic	$r_{int}(t)$	0.15
Extrinsic	$\log\big(\|G_g - G_b\|_2 + 0.05\big)$	6
Approach and progress shaping		
Behind-box	$r_{\text{b}} = \big(\max(0, -\hat{u}_{rb}^{\top}\hat{u}_{bg})\big)^2\,\mathbf{1}[d_{rb} < d_e]$	1.0
Toward-goal vel.	$r_v = \text{clip}(v_b^{\top}\hat{u}_{bg}, 0, 1)\,\mathbf{1}[d_{rb} < d_e]$	2.0
Progress	$r_{\Delta d} = \text{clip}(d_{bg}^{t-1} - d_{bg}^{t}, -\delta, \delta)\,\mathbf{1}[d_{rb} < d_e]$	8.0
Near-goal precision		
Pose precision	$r_{xy} = \exp\big(-d_{bg}^2/\sigma_{xy}^2\big)\,\mathbf{1}[d_{bg} < d_{xy}]$	3.0
Yaw precision	$r_{\psi} = \exp\big(-e_{\psi}^2/\sigma_{\psi}^2\big)\,\mathbf{1}[d_{bg} < d_{\psi}]$	2.5
Settle	$r_{\text{settle}} = \exp(-\|v_b\|/\sigma_v)\,\mathbf{1}[d_{bg} < d_s]$	2.0
Tangent penalty	$r_{\tan} = -\|v_b - (v_b^{\top}\hat{r})\hat{r}\|^2\,\mathbf{1}[d_{bg} < d_t]$	1.0
Success dwell	$r_{\text{dwell}} = \mathbf{1}[c_t \geq K]$	20.0
Micro precision	$r_{\text{micro}} = \exp\big(-d_{bg}^2/\sigma_m^2\big)\,\mathbf{1}[d_{bg} < d_m]$	12.0
Stability and motion regularization		
Upright dense	$r_{\text{up}} = \text{clip}(\cos\theta_{\text{up}}, 0, 1)$	0.05
Upright threshold	$r_{\text{up-pen}} = -\,\text{ReLU}(\tau - \cos\theta_{\text{up}})$	2.0

D. High-level observations

The high-level policy receives the robot base state, planar task geometry, and object-motion information. The observations include the robot's base velocities, projected gravity, previous action, relative object and goal geometry, and object motion features relevant to pushing. Table I summarizes the observation terms used in our implementation. Together, these signals provide the policy with enough information to approach the object, maintain directional progress, and adjust behavior near the goal.

E. Reward design

A key part of our formulation is the reward design for whole-body pushing. We structure the reward around four roles: (i) task completion, (ii) approach and directional progress, (iii) near-goal pose precision, and (iv) stability and motion regularization. Relative to earlier formulations, this design places greater emphasis on consistent progress toward the goal and accurate settling near the target pose.

The core objective combines an intrinsic term (adopting the intrinsic whole-body pushing term from [5]) with an extrinsic goal term. On top of this, we add shaping terms that encourage the robot to move into a useful pushing configuration, maintain velocity in the goal direction, and make measurable progress in object transport. Near the goal, additional terms encourage precise position alignment, correct final yaw, and low residual object motion. We also include regularization terms that help maintain stable behavior during contact-rich interaction.

This structure is motivated by the fact that pushing is not a single-mode behavior. Early in an episode, the main challenge is to establish good contact and drive the object in the correct direction. Near the goal, the challenge shifts to accurate pose correction and controlled settling. By separating these stages in the reward design, the policy receives a training signal that better reflects the structure of the task.

Table II summarizes the reward components used in our implementation.

F. Training objective

The high-level policy is trained with reinforcement learning to maximize discounted return:

$$J(\pi_\theta) = \mathbb{E}\left[\sum_{t=0}^{T-1} \gamma^t r_t\right], \tag{2}$$

where r_t is the whole-body pushing reward and γ is the discount factor. In practice, we optimize the policy using PPO [16]. The low-level locomotion controller remains frozen throughout task training, which stabilizes learning and reduces the complexity of the high-level policy's optimization problem.

III. EXPERIMENTAL SETUP

A. IsaacLab environment

All experiments are conducted in IsaacLab using a high-throughput simulated quadrupedal pushing environment. In each episode, the robot starts near a pushable object and must move it toward a target goal pose. The environment supports large-scale training under randomized initial conditions and object variation, making it well suited for studying contact-rich generalization.

B. Object families and distribution shifts

We evaluate across multiple object families to test whether the learned policy generalizes beyond a single canonical pushable. Our primary quantitative evaluation focuses on cuboid and cylindrical objects, with controlled variation in geometry and physics. We consider three evaluation suites: in-distribution (ID), out-of-distribution geometry (OOD-Geometry), and out-of-distribution physics (OOD-Physics). In addition, we include qualitative and targeted quantitative examples on unseen geometries not encountered during training.

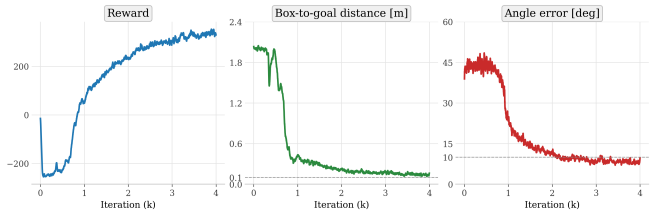


Fig. 3. Learning performance over training iterations. Each grouped plot shows total reward (left), object-to-goal position error (middle), and yaw error (right). Dashed lines indicate the reference thresholds used for the error metrics.

TABLE III

SUCCESS RATES (500 VARIANTS) FOR DIFFERENT DISTRIBUTIONS AND TOLERANCE SETTINGS FOR DIFFERENT OBJECTS.

Evaluation Setting	Cuboid		Cylinder	
	C1 (%)	C2 (%)	C1 (%)	C2 (%)
ID	86.69	79.38	82.33	79.56
OOD-Geometry	59.20	35.40	73.50	64.50
OOD-Physics	67.50	38.12	51.00	47.50

C. Policy Evaluation during Training

To better understand training behavior, we also report learning curves over training iterations. These include total reward, object-to-goal distance error, and yaw error. Together, they show whether the policy is learning both transport progress and final pose control over time.

IV. RESULTS

A. Overall privileged performance

The learning curves in Fig. 3 show that the privileged whole-body pushing pipeline learns stable goal-directed behavior in IsaacLab. During training, total reward improves while both position and yaw errors decrease, indicating that the policy learns not only to transport the object toward the goal but also to refine the final pose.

B. Generalization across object families

Table III reports per-family performance for cuboid and cylinder objects across the ID, OOD-Geometry, and OOD-Physics tests. We evaluate success under two tolerance settings: *Criterion 1 (C1)* uses $(\varepsilon_{xy}, \varepsilon_{\psi}) = (15\,\text{cm}, 15^{\circ})$ and *Criterion 2 (C2)* uses $(10\,\text{cm}, 10^{\circ})$. Performance is strong in-distribution for both object families. Under C1, cuboids achieve 86.69% success and cylinders 82.33%, while under the stricter C2 criterion both remain near 80%. These results show that the privileged pipeline supports both reliable transport and reasonably accurate final placement on trained variants. Out-of-distribution geometry and physics shifts remain challenging, but the policy still retains meaningful performance in both settings. The degradation pattern differs by family, which suggests that the difficulty of whole-body pushing depends not only on the size of the shift but also on how object geometry shapes the contact dynamics.

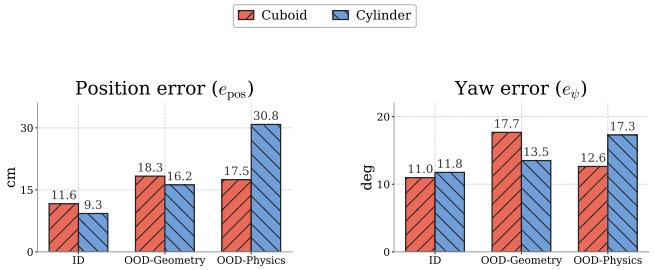


Fig. 4. Final position error e_{pos} and final yaw error e_{ψ} for cuboid and cylinder objects under ID, OOD-Geometry, and OOD-Physics.

C. Qualitative transfer to an unseen chair geometry

Beyond the main cuboid and cylinder evaluation suites, we tested the learned policy on an unseen chair geometry not used during training. Figure 5 shows six snapshots from a representative rollout. The policy approaches the chair, establishes contact, and pushes it toward the target pose. This suggests that the learned behavior is not limited to the exact training geometries. At the same time, this result is qualitative evidence rather than a broad generalization, since performance on unseen objects still depends strongly on whether the new geometry induces contact patterns similar to those encountered during training.

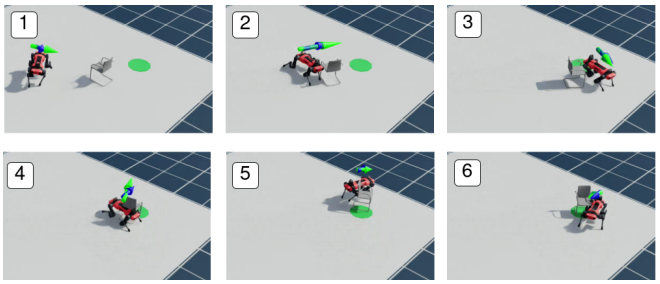


Fig. 5. **Qualitative transfer to an unseen chair geometry.** Six snapshots from a representative rollout on a chair instance.

D. Across-embodiment validation

In addition to the main object-family evaluation, we tested the proposed formulation on Spot, Anymal-C, and Unitree Go1. Across the three platforms, the hierarchical controller produced stable goal-directed whole-body pushing behavior.

V. CONCLUSIONS

We presented a hierarchical RL pipeline in IsaacLab for whole-body quadrupedal pushing across diverse object families and dynamics variation. Using privileged object-context information during training, the proposed system achieves robust goal-conditioned pushing under in-distribution conditions and retains meaningful performance under out-of-distribution geometry, out-of-distribution physics, and unseen-geometry evaluation. The proposed reward design supports both coarse transport and accurate final placement, yielding a strong benchmark in simulation. The next step will be to remove the simulator-side object information and replace it with online inference during interaction.

REFERENCES

[1] Y. Gong, G. Sun, A. Nair, A. Bidwai, R. CS, J. Grezmak, G. Sartoretti, and K. A. Daltorio, "Legged robots for object manipulation: A review," *Frontiers in Mechanical Engineering*, vol. 9, p. 1142421, 2023.

[2] X. Cheng, A. Kumar, and D. Pathak, "Legs as manipulator: Pushing quadrupedal agility beyond locomotion," *arXiv preprint arXiv:2303.11330*, 2023.

[3] X. He, C. Yuan, W. Zhou, R. Yang, D. Held, and X. Wang, "Visual manipulation with legs," in *Conference on Robot Learning*. PMLR, 2025, pp. 4218–4234.

[4] Z. He, K. Lei, Y. Ze, K. Sreenath, Z. Li, and H. Xu, "Learning visual quadrupedal loco-manipulation from demonstrations," in *2024 IEEE/RSJ international conference on intelligent robots and systems (IROS)*. IEEE, 2024, pp. 9102–9109.

[5] S. Jeon, M. Jung, S. Choi, B. Kim, and J. Hwangbo, "Learning whole-body manipulation for quadrupedal robot," *IEEE Robotics and Automation Letters*, 2023.

[6] J. Cheng, D. Kang, G. Fadini, G. Shi, and S. Coros, "Rambo: Rl-augmented model-based whole-body control for loco-manipulation," *IEEE Robotics and Automation Letters*, 2025.

[7] M. Liu, Z. Chen, X. Cheng, Y. Ji, R.-Z. Qiu, R. Yang, and X. Wang, "Visual whole-body control for legged loco-manipulation," *arXiv preprint arXiv:2403.16967*, 2024.

[8] X. Zhu, Y. Chen, L. Sun, F. Niroui, S. L. Cleac'h, J. Wang, and K. Fang, "Versatile loco-manipulation through flexible interlimb coordination," *arXiv preprint arXiv:2506.07876*, 2025.

[9] J. A. Barreiros, A. Ö. Önol, M. Zhang, S. Creasey, A. Goncalves, A. Beaulieu, A. Bhat, K. M. Tsui, and A. Alspach, "Learning contact-rich whole-body manipulation with example-guided reinforcement learning," *Science Robotics*, vol. 10, no. 105, p. eads6790, 2025.

[10] I. Dadiotis, M. Mittal, N. Tsagarakis, and M. Hutter, "Dynamic object goal pushing with mobile manipulators through model-free constrained reinforcement learning," in *2025 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2025, pp. 13 363–13 369.

[11] M. Mittal, P. Roth, J. Tigue, A. Richard, O. Zhang, P. Du, A. Serrano-Munoz, X. Yao, R. Zurbrügg, N. Rudin *et al.*, "Isaac lab: A gpu-accelerated simulation framework for multi-modal robot learning," *arXiv preprint arXiv:2511.04831*, 2025.

[12] A. Kumar, Z. Fu, D. Pathak, and J. Malik, "Rapid motor adaptation for legged robots," in *Robotics: Science and Systems (RSS)*, 2021.

[13] W. Yu, J. Tan, C. K. Liu, and G. Turk, "Preparing for the unknown: Learning a universal policy with online system identification," in *Robotics: Science and Systems (RSS)*, 2017.

[14] Z. Fu, X. Cheng, and D. Pathak, "Deep whole-body control: Learning a unified policy for manipulation and locomotion," in *Conference on Robot Learning (CoRL)*, 2022.

[15] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

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