

VOFA: Visual Object Goal Pushing with Force-Adaptive Control for Humanoids

Zichao Hu¹, Zifan Xu¹, Dongsik Chang², He Yin², Linh Tran¹, Roberto Martín-Martín^{1,2}
 Peter Stone^{1,3}, Jingyu Qiao², Joydeep Biswas¹

¹Department of Computer Science, The University of Texas at Austin

² Amazon Inc. ³Sony AI

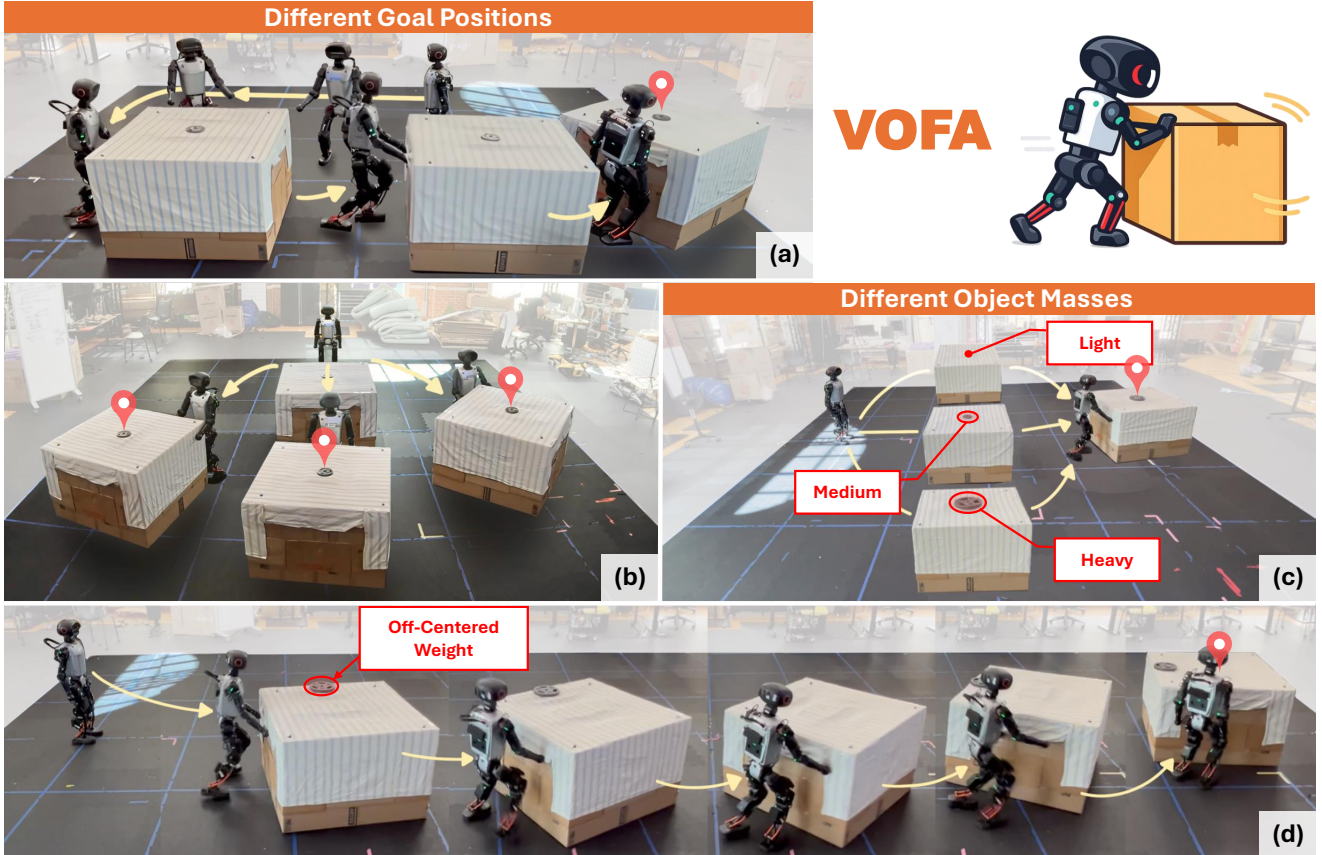


Fig. 1: We present VOFA, a visual goal-conditioned humanoid loco-manipulation system capable of pushing objects with unknown physical properties. The system adapts to different goal positions (a,b), object masses (c), and center-of-mass configurations (d) while maintaining stable, closed-loop control.

Abstract—The ability to push large objects in a goal-directed manner using onboard egocentric perception is an essential skill for humanoid robots to perform complex tasks such as material handling in warehouses. To robustly manipulate heavy objects to arbitrary goal configurations, the robot must cope with unknown object mass and ground friction, noisy onboard perception, and actuation errors; all in a real-time feedback loop. Existing solutions either rely on privileged object-state information without onboard perception or lack robustness to variations in goal configurations and object physical properties. In this work, we present VOFA, a visual goal-conditioned humanoid loco-manipulation system capable of pushing objects with unknown physical properties to arbitrary goal positions.

VOFA consists of a two-level hierarchical architecture with a high-level visuomotor policy and a low-level force-adaptive whole-body controller. The high-level policy processes noisy onboard observations and generates goal-conditioned commands to operate in closed loop across diverse object-goal configurations, while the low-level whole-body controller provides robustness to variations in object physical properties. VOFA is extensively evaluated in both simulation and real-world experiments on the Booster T1 humanoid robot. Our results demonstrate strong performance, achieving over 90% success in simulation and over 80% success in real-world trials. Moreover, VOFA successfully pushes objects weighing up to 17kg, exceeding half of the Booster T1’s body weight.¹

I. INTRODUCTION

Developing loco-manipulation capabilities for humanoid robots to perform useful tasks has gained increasing attention in robotics research [1]. One such capability is pushing large objects to designated locations to support real-world applications such as warehouse inventory reorganization. Successfully achieving this capability requires addressing several key challenges: (1) handling objects with unknown and diverse physical properties, such as mass and friction; (2) operating with noisy onboard visual observations without access to privileged object-state information; and (3) adapting in closed loop to different object-goal configurations. Recent progress in humanoid loco-manipulation driven by reinforcement learning has demonstrated promising capabilities; however, existing humanoid systems do not fully address the challenges presented by goal-directed object pushing. In particular, existing systems rely on privileged object-state sensing [2]–[6], lack adaptability to different object and goal configurations [7]–[9], or do not explicitly account for variations in object mass [10], [11].

In this work, we present VOFA (Visual Object-Goal Pushing with Force-Adaptive Control), a visual goal-conditioned humanoid loco-manipulation system capable of pushing objects with unknown physical properties (e.g., mass and friction) to arbitrary goal positions. VOFA employs a two-level hierarchical architecture (Fig. 2) composed of a high-level depth-image-based visuomotor policy and a low-level force-adaptive whole-body controller (WBC). The high-level policy processes onboard observations and outputs goal-conditioned commands, which are translated by the low-level controller into joint-level control actions. To address unknown object physical properties, the low-level controller is explicitly designed as a force-adaptive controller following the FALCON [12] framework. This controller enables stable object interaction across diverse object masses and ground frictions. Together, the hierarchical design decouples low-level force-adaptive control from high-level visuomotor planning, simplifying policy training [7] while preserving robustness to variations in object physical properties.

To efficiently learn a high-level policy driven by noisy onboard visual observations, VOFA follows a teacher-student design. The teacher policy is first trained as a goal-conditioned policy using privileged proprioceptive and exteroceptive information (e.g., the robot’s base linear velocity and the object’s orientation). DAGGER [13] is then employed to distill a vision-based student policy that replaces privileged inputs with proprioceptive signals and onboard depth observations. During distillation, visual augmentations including camera extrinsic randomization and depth-noise injection (Fig. 3) are incorporated to improve robustness to perception noise and background distractions at deployment.

To enable VOFA to adapt in closed loop to arbitrary goal positions, training is conducted in large-scale Isaac Gym simulation with extensive randomization over initial object

and goal placements. An object-goal alignment reward is further introduced to encourage the robot to first reposition itself to the opposite side of the object relative to the goal before initiating contact (Fig. 5). This simple yet effective reward design removes the need to manually define separate training stages [8], [14] to complete the full task. It promotes consistent object-robot alignment and enables corrective behavior when deviations from the desired object trajectory occur.

Our system, VOFA, achieves over 90% success rates in simulation and over 80% success rates in real-world experiments across a range of initial object and goal positions and diverse object physical properties. In real-world deployment (Fig. 1), VOFA successfully pushes objects weighing 17 kg (more than half of the Booster T1 robot’s body weight) despite not being trained on such object masses in simulation. Leveraging closed-loop visual feedback, VOFA remains robust even when the object’s center of mass is intentionally shifted to previously unseen locations, and is able to detect and recover from deviations in the object’s motion online. In addition, VOFA supports long-horizon box-pushing tasks, sequentially pushing the object to multiple goal locations. Together, these results demonstrate robust closed-loop performance and highlight the potential of VOFA to enable more complex downstream applications, such as object rearrangement. Overall, our key contributions are:

- 1) We present a visual humanoid object-goal pushing system that enables closed-loop, goal-directed pushing across diverse object-goal configurations.
- 2) Our system handles objects with diverse and even previously unseen masses, successfully pushing objects up to 17kg—more than half of the robot’s body weight.
- 3) We validate our approach in both large-scale simulation and real-world experiments on the Booster T1 humanoid robot, achieving high success rates across diverse scene configurations and object physical properties.

II. RELATED WORK

In this section, we review related work on humanoid whole-body control and autonomous legged loco-manipulation.

A. Humanoid Whole Body Control

A substantial body of recent work has explored learning-based whole-body control (WBC) for humanoid robots, demonstrating stable and expressive full-body behaviors such as locomotion, balance, and dynamic motion generation [16]–[20]. Many of these approaches rely on large-scale motion imitation and primarily focus on reproducing coordinated human motions. Beyond expressive whole-body motions, other work investigates WBC for teleoperated loco-manipulation, enabling humanoids to interact with objects while maintaining whole-body stability [12],

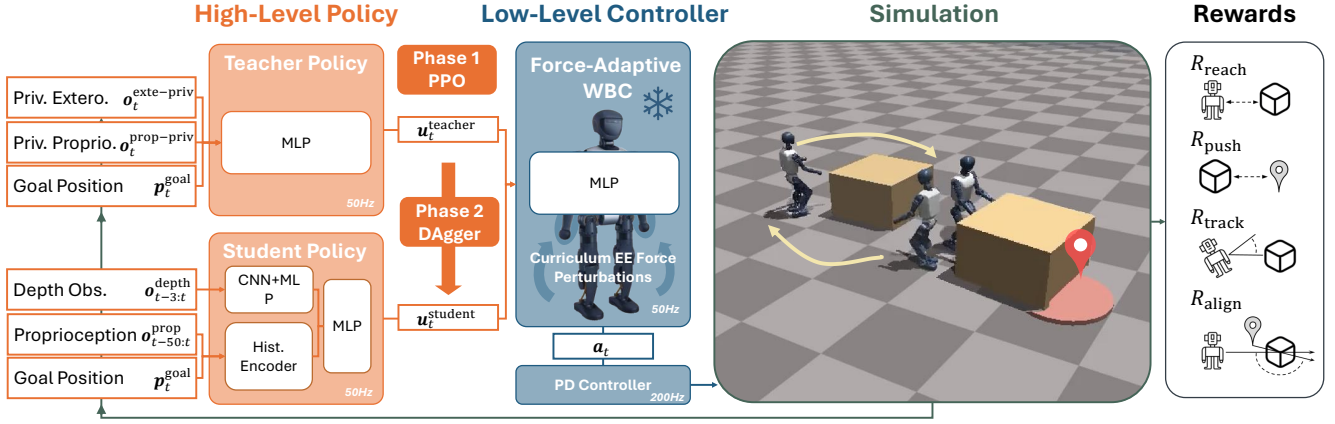


Fig. 2: **VOFA Design.** VOFA adopts a hierarchical architecture that combines a high-level visuomotor policy with a force-adaptive whole-body controller [12] for humanoid visual object–goal pushing. The high-level policy is trained in a teacher–student framework: a teacher policy is trained with privileged observations and goal positions using PPO [15], and a vision-based student policy is distilled via DAGger [13] using onboard sensory inputs and goal information. Training uses four reward terms—reach, push, track, and alignment—to promote robust closed-loop object–goal interaction.

[21]–[25]. Among these works, FALCON [12] introduces a force-adaptive whole-body controller for humanoid loco-manipulation, providing robustness to substantial and unknown external forces at the end-effector during object interaction. In this work, we build on FALCON and extend it with a hierarchical, vision-based framework to enable goal-conditioned visual object pushing for humanoid robots.

B. Sim-to-Real Reinforcement Learning for Autonomous Humanoid Loco-Manipulation

Enabling legged robots to autonomously perform loco-manipulation is important for deploying robots to perform useful tasks in human-inhabited environments. Prior work on autonomous loco-manipulation has studied both quadruped [2], [3] and humanoid [4]–[6] robots; however, these approaches typically rely on privileged object state information rather than onboard visual perception. More recently, vision-based humanoid loco-manipulation has been explored by incorporating onboard visual perception into whole-body policies [7], [8], [10]; however, these approaches do not explicitly study goal-conditioned object interaction or robustness to variations in object physical properties. Visual, goal-directed humanoid loco-manipulation under variations in object mass and physical properties is still challenging. In this work, we aim to address this gap through sim-to-real reinforcement learning.

III. METHOD

Problem Setup: We consider the problem of visual goal-directed object pushing with a humanoid robot. At each timestep t , the robot receives observations consisting of proprioceptive measurements $\mathbf{o}_t^{\text{prop}}$, depth observations $\mathbf{o}_t^{\text{depth}}$, and a desired object goal position $\mathbf{p}_t^{\text{goal}}$. The objective is to learn a control policy that maps a history of observations to

humanoid joint position targets $\mathbf{a}_t$ such that it controls the robot to push the object to the specified goal location.

System Overview: We present VOFA, a visual goal-conditioned loco-manipulation system with a two-level hierarchical architecture. The high-level policy $\pi_H(\mathbf{u}_t \mid \mathbf{o}_{t-h_1:t}^{\text{prop}}, \mathbf{o}_{t-h_2:t}^{\text{depth}}, \mathbf{p}_t^{\text{goal}})$ first maps a history of proprioceptive and depth observations and the goal position to a high-level command $\mathbf{u}_t$ (Sec. III-A). Then, the high-level command is provided as input to a low-level force-adaptive controller $\pi_L(\mathbf{a}_t \mid \mathbf{u}_t)$, which outputs joint position targets $\mathbf{a}_t$ (Sec. III-B). Both policies are trained in IsaacGym simulation and operate at 50Hz. Beyond the architectural design, we detail the reward formulation (Sec. III-C) and the domain randomization strategy enabling zero-shot transfer to the real robot (Sec. III-D) in the following subsections as well.

A. High-Level Policy Design and Training

To efficiently train the high-level policy, we adopt a teacher–student paradigm.

a) Teacher Policy Design: We first train a goal-conditioned teacher policy to output actions within the low-level WBC’s valid command range [7]. The teacher policy is trained using privileged observations with PPO [15] under an asymmetric actor–critic framework [26], which has been shown to be effective in prior work [3].

The actor receives privileged observations $\mathbf{o}_t^{\text{priv}} = (\mathbf{o}_t^{\text{prop-priv}}, \mathbf{o}_t^{\text{exte-priv}}, \mathbf{p}_t^{\text{goal}})$, consisting of privileged proprioceptive information, privileged exteroceptive information, and the goal position. Proprioceptive inputs are given by $\mathbf{o}_t^{\text{prop-priv}} = (\mathbf{v}_t, \boldsymbol{\omega}_t, \mathbf{g}_t, \mathbf{u}_{t-1}^{\text{teacher}}, \mathbf{q}_t^{\text{upper}}, \dot{\mathbf{q}}_t^{\text{upper}})$, including linear and angular velocities, projected gravity, the previous high-level action, upper-body joint positions and velocities. Exteroceptive inputs are given by $\mathbf{o}_t^{\text{exte-priv}} = (\mathbf{p}_t^{\text{ee,r-obj}}, \mathbf{p}_t^{\text{ee,l-obj}}, \mathbf{R}_t)$, comprising the relative positions of the

right and left end effectors with respect to the object and the object orientation. All quantities are expressed in the robot base frame.

The critic observes all actor inputs and additionally receives privileged object state information $\mathbf{o}_t^{\text{priv-critic}} = (c_t, \mathbf{p}_t^{\text{obj}}, m_t^{\text{obj}}, \mathbf{d}_t^{\text{obj}}, \mathbf{v}_t^{\text{obj}}, \boldsymbol{\omega}_t^{\text{obj}})$, where c_t denotes the object contact flag, $\mathbf{p}_t^{\text{obj}}$ is the object center-of-mass position, m_t^{obj} and $\mathbf{d}_t^{\text{obj}}$ represent the object mass and dimensions, and $\mathbf{v}_t^{\text{obj}}$ and $\boldsymbol{\omega}_t^{\text{obj}}$ are the object linear and angular velocities.

b) Student Policy Design: We employ DAGger [13] to distill a vision-based student policy from the teacher by replacing privileged proprioceptive and exteroceptive inputs ($\mathbf{o}_t^{\text{prop-priv}}, \mathbf{o}_t^{\text{exte-priv}}$) with a history of onboard proprioceptive observations and depth images ($\mathbf{o}_{t-50:t}^{\text{prop}}, \mathbf{o}_{t-3:t}^{\text{depth}}$). The proprioceptive observation $\mathbf{o}_t^{\text{prop}} = (\boldsymbol{\omega}_t, \mathbf{g}_t, \mathbf{u}_{t-1}^{\text{student}}, \mathbf{q}_t, \dot{\mathbf{q}}_t)$ consists of the base angular velocity, projected gravity, the previous high-level action, and the joint positions and velocities. The student policy encodes the proprioceptive history using temporal 1D convolutions [14], while each depth observation is processed by a convolutional neural network to extract visual features. The proprioceptive and visual embeddings are then fused and passed through an MLP projection head that aggregates the combined representation for action prediction. To improve training efficiency, the policy maintains a rolling buffer of the three most recent depth frames, processes depth observations at 5Hz, and downsamples the resolution of images (from 640×360 to 32×32).

B. Low-Level Policy Design and Training

We adopt a force-adaptive whole-body controller (WBC) as the low-level policy to handle substantial and unknown end-effector interaction forces, which commonly arise in object-pushing tasks due to variations in object physical properties (e.g., mass). In this work, we choose FALCON [12] as it is explicitly trained for force-adaptive control using a curriculum that progressively increases external forces at the end-effectors. In addition, FALCON is trained on the AMASS [27] motion dataset, which spans a broad range of natural human motions, including behaviors relevant to box pushing. FALCON receives a high-level command $\mathbf{u}_t = (\mathbf{u}_t^{\text{lower}}, \mathbf{q}_t^{\text{upper}})$, where the lower-body locomotion command $\mathbf{u}_t^{\text{lower}} = (\mathbf{v}_t^{\text{lin,ang}}, \phi_t^{\text{stance}}, h_t^{\text{root}}, \omega_t^{\text{yaw}})$ specifies the desired root linear and angular velocities, walking-mode indicator, root height, and waist yaw angle, and $\mathbf{q}_t^{\text{upper}}$ denotes upper-body target joint positions. In this work, the walking-mode indicator, root height, and waist yaw angle are set to their default values (1, 0.62, and 0, respectively), while learning a residual upper-body target joint position $\Delta \mathbf{q}_t^{\text{upper}}$. FALCON is trained for 10k epochs using PPO and is kept frozen during subsequent high-level policy training.

C. Reward Design

To enable the teacher policy to learn effective object goal pushing behavior, we design four main rewards:

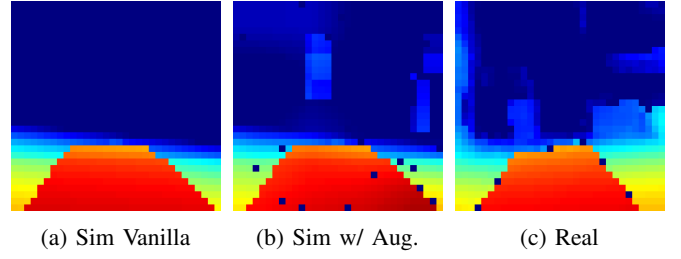


Fig. 3: **Egocentric vision of the humanoid robot.** During training, we apply three vision augmentations: *far-plane depth perturbation*, *correlated depth noise*, and *pixel dropout*.

- 1) **Reaching the object with both end-effectors.** To encourage both end-effectors to approach the object, we define the per-end-effector reaching terms as

$$r_t^{\text{ee}_i} = \exp\left(\frac{-\|\mathbf{p}_t^{\text{ee}, i-\text{obj}}\|^2}{\sigma_{\text{reach}}^2}\right), \quad i \in \{l, r\}. \quad (1)$$

The overall reaching reward is computed as the harmonic mean of the two end-effector terms:

$$R_{\text{reach}} = \text{HarmonicMean}(r_t^{\text{ee}_l}, r_t^{\text{ee}_r}). \quad (2)$$

- 2) **Pushing the object toward the goal.** The robot is rewarded for reducing the object–goal distance:

$$R_{\text{push}} = \exp\left(\frac{-\|\mathbf{p}_t^{\text{goal}} - \mathbf{p}_t^{\text{obj}}\|^2}{\sigma_{\text{push}}^2}\right). \quad (3)$$

- 3) **Head tracking reward.** The robot is encouraged to visually track the object:

$$R_{\text{track}} = \begin{cases} \exp\left(\frac{-\theta_t^2}{\sigma_{\text{track}}^2}\right), & \text{if } |\theta_t| \leq \theta_{\text{fov}}, \\ 0, & \text{otherwise,} \end{cases} \quad (4)$$

where θ_t is the angular deviation between the head (camera optical axis) and the object direction, and θ_{fov} denotes half of the camera field of view.

- 4) **Object–goal alignment reward.** To promote goal-directed pushing configurations, we define:

$$R_{\text{align}} = \exp\left(\frac{-(1 - \theta_t^{\text{align}})^2}{\sigma_{\text{align}}^2}\right), \quad (5)$$

where θ_t^{align} is the normalized angle between the object–robot and object–goal direction vectors.

Notably, the object–goal alignment reward is critical for learning effective goal-conditioned pushing behavior, as demonstrated by the ablation study in Sec. IV-C. In addition to these four main task rewards, we incorporate several penalty terms to regularize the robot’s behavior and improve stability.

TABLE I: Performance under different initial goal positions.

Method	Overall	Front			Lateral			Rear		
	SR(%) $\uparrow$	SR(%) $\uparrow$	Pre-C.(%) $\downarrow$	Post-C. $\downarrow$	SR (%) $\uparrow$	Pre-C.(%) $\downarrow$	Post-C.(%) $\downarrow$	SR(%) $\uparrow$	Pre-C.(%) $\downarrow$	Post-C.(%) $\downarrow$
Teacher w/ FA (ours)	97.84	98.12	0.72	0.96	97.58	1.00	1.16	97.34	1.12	1.08
Teacher w/o FA	76.94	83.78	2.92	5.52	79.00	4.88	6.58	71.86	6.46	7.38
Student w/ FA (ours)	91.38	91.97	1.84	2.42	91.40	1.30	3.22	90.22	1.92	3.98
Student w/o FA	59.32	71.16	2.68	7.00	64.90	4.52	7.78	56.06	5.64	8.84

TABLE II: Performance under different box mass.

Method	Light			Medium			Heavy			Extra Heavy (Unseen)		
	SR(%) $\uparrow$	Pre-C.(%) $\downarrow$	Post-C.(%) $\downarrow$	SR(%) $\uparrow$	Pre-C.(%) $\downarrow$	Post-C.(%) $\downarrow$	SR (%) $\uparrow$	Pre-C.(%) $\downarrow$	Post-C.(%) $\downarrow$	SR(%) $\uparrow$	Pre-C.(%) $\downarrow$	Post-C.(%) $\downarrow$
Teacher w/ FA (ours)	97.40	1.12	1.34	98.22	0.92	0.74	97.46	1.06	1.08	87.92	0.88	4.16
Teacher w/o FA	82.18	5.40	9.00	80.98	5.08	6.06	69.76	4.64	6.06	53.68	4.78	5.66
Student w/ FA (ours)	93.50	1.46	3.06	93.30	1.40	2.74	89.16	1.22	3.22	69.00	1.30	6.30
Student w/o FA	70.98	6.60	9.46	62.68	5.98	7.64	48.66	6.56	9.58	32.00	5.36	9.94

D. Environment Setup and Domain Randomization

During training, we randomize object physical properties and initial object–goal configurations. We consider a box-shaped object and randomize its mass (1–8 kg) and friction coefficients (0.1–1.0) to promote robustness across diverse physical conditions. Goals are placed uniformly over a 360° arc around the robot at a radius of 2.5–3.5 m. Objects are randomly initialized within the robot’s field of view ($\pm 60^\circ$) and initial distances of 1.8–2.2 m.

To reduce the sim-to-real gap in visual perception, we match the camera intrinsics in simulation to those of the real robot’s cameras (ZED 2i). However, small differences in camera mounting can result in slight viewpoint variations, even across robots of the same model. To account for this variability, we apply random extrinsic perturbations during training, including translations of ± 3 cm in x,y,z and rotations of $\pm 5^\circ$ (yaw), $\pm 5^\circ$ (pitch), and $\pm 2^\circ$ (roll). In addition, real-world depth images exhibit sensor artifacts such as interference from background objects and isolated pixels with spuriously large depth values, as shown in Fig. 3b. To emulate these effects, we apply three visual randomizations as shown in Fig. 3c. First, we introduce *far-plane depth perturbation*, where random rectangular patches are placed in the upper image region (above an estimated ground boundary) and overwritten with blocky far-depth noise to simulate transient background clutter. Second, we apply *correlated depth noise* by adding spatially correlated low-frequency Gaussian noise to the depth map, modeling structured sensor distortions. Third, we apply *pixel dropout*, randomly zeroing depth pixels to simulate missing or unreliable returns. An ablation study on the effectiveness of these randomizations is presented in Sec. IV-E.

IV. EXPERIMENTAL RESULTS

In this section, we conduct a series of simulation and real-world experiments, with policies trained in IsaacGym and

deployed on the Booster T1 humanoid robot, to answer the following research questions:

- 1) How robust is VOFA across varying initial goal configurations and object masses (Sec. IV-A)?
- 2) Is using a force-adaptive whole-body control beneficial for VOFA (Sec. IV-B)?
- 3) Does the object–goal alignment reward design mitigate failure modes in challenging initial configurations (Sec. IV-C)?
- 4) How well does the policy transfer from simulation to the real world (Sec. IV-D)?
- 5) How effective is the proposed visual randomization strategy for sim-to-real transfer (Sec. IV-E)?
- 6) Can the policy perform with closed-loop control when pushing the object (Sec. IV-F)?

A. Performance Across Initial Configurations and Masses

We conduct two sets of quantitative simulation experiments using three metrics: *Success Rate (SR)*, *Terminated Pre-Contact (Pre-C.)*, and *Terminated Post-Contact (Post-C.)*. An episode is successful if the object is pushed within 0.3 m of the goal and remains there for 2 s. Pre-C. and Post-C. denote that the robot falls before and after object contact, respectively.

As shown in Tab. I, we evaluate three goal-position categories based on the relative goal angle θ : *front* ($|\theta| \in [0^\circ, 60^\circ]$), *lateral* ($|\theta| \in [60^\circ, 120^\circ]$), and *rear* ($|\theta| \in [120^\circ, 180^\circ]$). Performance remains consistently high across all configurations, with success rates above 97% for the teacher and 90% for the student. Even in rear configurations, which require longer-horizon repositioning, performance degrades only marginally.

As shown in Tab. II, we evaluate four mass ranges: *light* (1–3 kg), *medium* (3–5 kg), *heavy* (5–8 kg), and *extra-heavy* (8–12 kg), with the last range outside the training distribution. The policy remains robust across all in-distribution

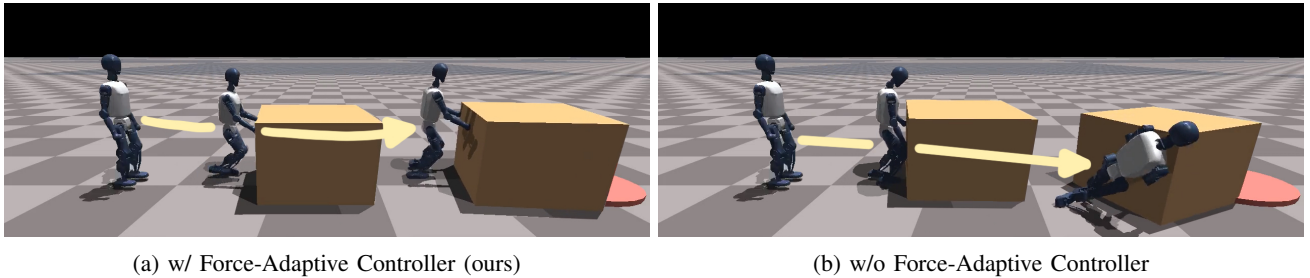


Fig. 4: **Force-Adaptive Controller Ablation.** With the force-adaptive controller, the robot stably pushes the object using its end-effector (a). Without it, the robot struggles to apply consistent forces, often resorting to kicking, which results in unstable behavior and a higher risk of falling.

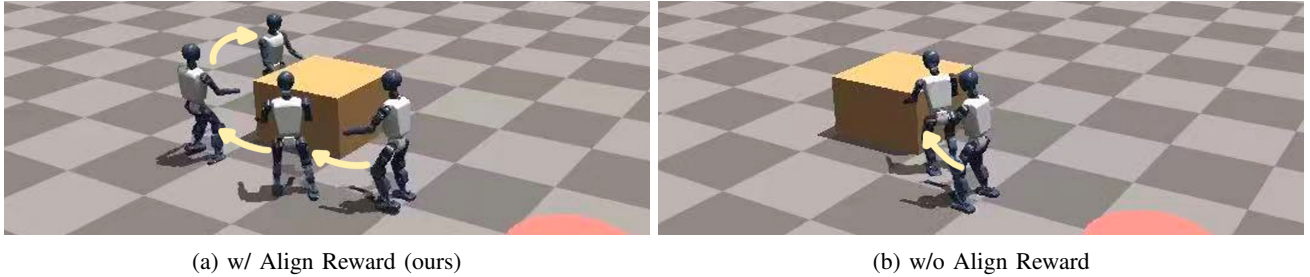


Fig. 5: **Object-Goal Alignment Reward Ablation.** The alignment reward enables the policy to reposition around the object before pushing it toward the goal (a). Without it, the robot tends to make premature contact and fails to push the object to the goal (b).

masses and achieves a 69% success rate in the extra-heavy setting.

Across these experiments, we observe three failure modes: failing to push the box to the target, falling before contact, and falling after contact. Compared with the teacher, the student exhibits more target-reaching failures and post-contact falls, highlighting the added challenge of replacing privileged object-state information with noisy onboard visual observations while maintaining closed-loop performance.

B. Effect of Force-Adaptive Whole-Body Control on VOFA

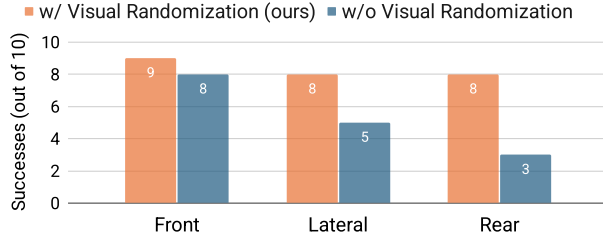
To evaluate the impact of force-adaptive whole-body control, we train ablated variants of VOFA by disabling random end-effector force perturbations during FALCON pretraining. We evaluate both teacher and student policies under varying initial goal configurations and object masses using the same setup as Sec. IV-A (Tab. I and Tab. II). Across all settings, force-adaptive control yields higher success rates and fewer terminations. Without force adaptation, performance degrades notably in lateral and rear configurations and for heavier objects, as the robot struggles to push the object to the target goal location. Fig. 4 presents a qualitative comparison of the robot’s behavior with and without the force-adaptive controller. Without the force-adaptive controller, the robot struggles to apply consistent contact forces and often resorts to impulsive kicking motions, leading to unstable behavior and an increased risk of falling. Overall, these results show that force-adaptive whole-body control is beneficial for robust loco-manipulation in VOFA.

C. Impact of Object–Goal Alignment Reward

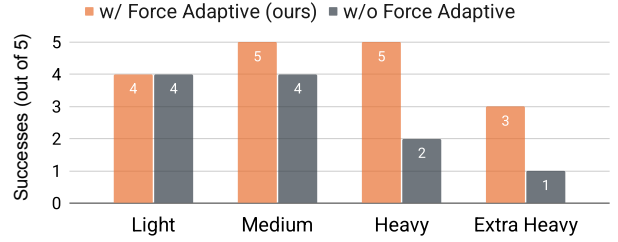
Without the object–goal alignment reward, we observe a consistent failure mode in which the robot prematurely approaches the object when the object lies on the opposite side of the robot from the goal as shown in Fig. 5. In such cases, the policy is driven by the R_{reach} reward to make early contact, which prevents effective pushing toward the goal. As a result, the policy struggles to explore behaviors that require first repositioning the robot to a more favorable location before pushing. In contrast, introducing the object–goal alignment reward encourages deliberate repositioning prior to contact, significantly improving task success in these challenging configurations by promoting long-horizon execution rather than myopic contact strategies.

D. Sim-to-Real Transfer Performance

To evaluate sim-to-real transfer, we deploy the policy on the Booster T1 humanoid to push a 5kg cardboard box measuring approximately $1.0 \times 1.0 \times 0.7\text{m}$. During deployment, we use motion capture to provide the relative goal position. While this quantity could also be obtained from a separate localization system, using motion capture allows us to focus the evaluation on the proposed visual object-pushing system. We first evaluate performance across three goal configurations: *front* (goal at 0°), *lateral* (goal at $\pm 90^\circ$), and *rear* (goal at 180°). For each configuration, we perform 10 real-world trials, with the object randomly placed within the robot’s initial field of view following the ranges

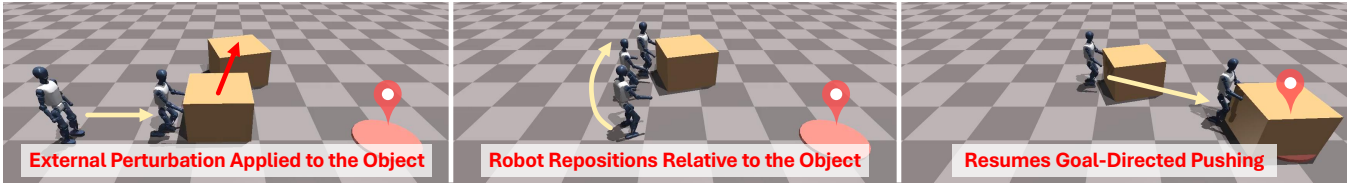


(a) Different Goal Positions

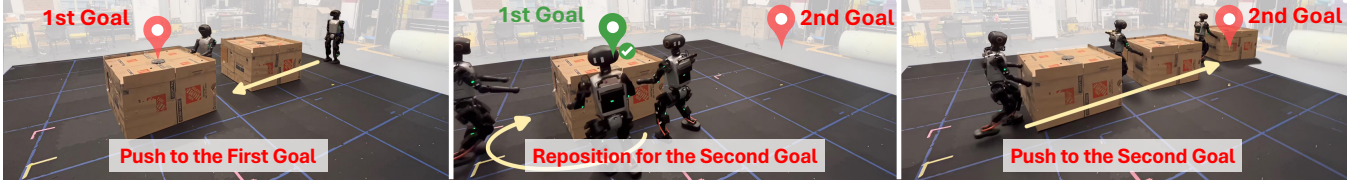


(b) Different Object Masses

Fig. 6: **Real World Deployment Results.** VOFA demonstrates strong sim-to-real transfer performance. Real-world experiments highlight the benefits of visual randomization across different goal positions (a) and the force-adaptive low-level controller across varying object masses (b).



(a) External perturbation recovery.



(b) Sequential multi-goal pushing.

Fig. 7: **Closed-Loop Demonstration.** The policy reacts to external disturbances (a) and supports sequential goal switching without reset (b), illustrating closed-loop control.

used during training. As shown in Fig. 6a, the policy achieves consistently strong real-world performance across all goal configurations. In addition, we evaluate robustness to object mass by attaching additional weights to the box. We perform 5 real-world trials for each of the four mass conditions: *light* (no additional weight), *medium* (2.5kg), *heavy* (4.5kg), and *extra-heavy* (12kg). To isolate the effect of object mass, we fix the goal location 4m directly in front of the robot. As shown in Fig. 6b, the policy equipped with force-adaptive whole-body control maintains stable performance across all tested mass conditions, whereas the non-force-adaptive baseline degrades substantially as object mass increases. Notably, VOFA successfully pushes a 17kg object (12kg payload + 5kg box), exceeding half of the Booster T1 robot’s body weight, despite not being trained on such heavy objects in simulation. This real-world trend mirrors the simulation results, further validating the role of force-adaptive control in enabling robust sim-to-real transfer.

E. Effectiveness of Visual Randomization

To assess the effectiveness of the proposed visual randomization strategy, we conduct an ablation study by training a policy without visual randomization and evaluating it under

the same real-world conditions described in the previous subsection (Fig. 6a). Without visual randomization, policy performance degrades substantially for tasks that require longer-horizon execution, particularly in lateral and rear goal configurations. As the robot moves through the environment, it encounters a wider range of background structures and viewpoints, increasing sensitivity to visual noise and domain mismatch. In the absence of visual randomization during training, these effects lead to unstable behaviors, including confusion between background objects and the target object or failure to correctly localize the goal. In contrast, visual randomization improves robustness to such visual variations, enabling more reliable sim-to-real transfer.

F. Closed-Loop Object Pushing

To evaluate whether the policy performs closed-loop control during object pushing, we conduct three experiments in both simulation and the real world.

a) *External Perturbation.*: As shown in Fig. 7a, we apply a lateral impulse perpendicular to the pushing direction, displacing the object from its nominal path. The robot reacts by halting forward motion, reorienting to locate the displaced object, navigating around it to re-establish

object-goal alignment, and subsequently resuming pushing toward the target. Successful task completion under this disturbance demonstrates online feedback correction rather than open-loop execution.

b) *Off-Centered Mass Distribution.*: As shown in Fig. 1d, we attach additional weights to one side of the object to create an off-centered center of mass (COM). This causes the box to rotate and deviate from the desired trajectory when being pushed. Similar to the external-perturbation example in Fig. 7a, the robot detects this deviation and repositions itself to push the box to the goal.

c) *Sequential Goal Switching.*: As shown in Fig. 7b, we evaluate reactive re-planning by assigning multiple goals sequentially. The robot first pushes the object to a front goal location. Upon reaching this target, a new goal is issued behind the robot. The robot repositions itself by navigating around the object and then pushes it toward the updated goal position without reset.

Collectively, these experiments qualitatively demonstrate that the policy operates in a closed-loop manner: actions are continuously conditioned on updated observations, enabling disturbance rejection, dynamic re-alignment, and goal adaptation during object pushing.

V. CONCLUSION, LIMITATIONS, AND FUTURE WORK

In this work, we present VOFA, a visual goal-conditioned humanoid object pushing system. By integrating a force-adaptive whole-body controller with a depth-based visuomotor policy trained via teacher-student distillation and extensive domain randomization, VOFA achieves robust closed-loop performance under diverse goal configurations and object physical properties. Extensive simulation and real-world experiments demonstrate strong generalization across goal positions, robustness to object masses, and effective sim-to-real transfer. Despite these results, our approach has several limitations. First, we focus on single-object scenarios and do not explicitly consider environments with multiple objects or highly diverse background variations, which may introduce additional perceptual ambiguity. Second, the system assumes access to the relative goal position during deployment. Although this information could be supplied by a localization system, integrating such a system would introduce additional complexity. In future work, we aim to extend VOFA to more cluttered multi-object environments and develop an end-to-end framework that jointly reasons about object and goal localization from onboard perception.

REFERENCES

- [1] Z. Gu, J. Li, W. Shen, *et al.*, “Humanoid locomotion and manipulation: Current progress and challenges in control, planning, and learning,” 2025. [Online]. Available: <https://arxiv.org/abs/2501.02116>
- [2] S. Jeon, M. Jung, S. Choi, *et al.*, “Learning whole-body manipulation for quadrupedal robot,” *IEEE Robotics and Automation Letters*, vol. 9, no. 1, pp. 699–706, 2024.
- [3] I. Dadiotis, M. Mittal, N. Tsagarakis, *et al.*, “Dynamic object goal pushing with mobile manipulators through model-free constrained reinforcement learning,” in *2025 IEEE International Conference on Robotics and Automation (ICRA)*, 2025, pp. 13 363–13 369.
- [4] Z. Zhang, C. Chen, H. Xue, *et al.*, “Unleashing humanoid reaching potential via real-world-ready skill space,” *IEEE Robotics and Automation Letters*, vol. 11, no. 2, pp. 2082–2089, 2026.
- [5] H. Weng, Y. Li, N. Sobanbabu, *et al.*, “Hdmi: Learning interactive humanoid whole-body control from human videos,” 2025. [Online]. Available: <https://arxiv.org/abs/2509.16757>
- [6] Z. Su, B. Zhang, N. Rahmanian, *et al.*, “Hitter: A humanoid table tennis robot via hierarchical planning and learning,” 2025. [Online]. Available: <https://arxiv.org/abs/2508.21043>
- [7] S. Yin, Y. Ze, H.-X. Yu, *et al.*, “Visualmimic: Visual humanoid loco-manipulation via motion tracking and generation,” 2025. [Online]. Available: <https://arxiv.org/abs/2509.20322>
- [8] H. Xue, T. He, Z. Wang, *et al.*, “Opening the sim-to-real door for humanoid pixel-to-action policy transfer,” 2025. [Online]. Available: <https://arxiv.org/abs/2512.01061>
- [9] P. Zhi, P. Li, J. Yin, B. Jia, and S. Huang, “Learning unified force and position control for legged loco-manipulation,” 2025. [Online]. Available: <https://arxiv.org/abs/2505.20829>
- [10] T. He, Z. Wang, H. Xue, *et al.*, “Viral: Visual sim-to-real at scale for humanoid loco-manipulation,” 2025. [Online]. Available: <https://arxiv.org/abs/2511.15200>
- [11] J. Hu, P. Stone, and R. Martín-Martín, “Slac: Simulation-pretrained latent action space for whole-body real-world rl,” 2025. [Online]. Available: <https://arxiv.org/abs/2506.04147>
- [12] Y. Zhang, Y. Yuan, P. Gurunath, *et al.*, “Falcon: Learning force-adaptive humanoid loco-manipulation,” *arXiv preprint arXiv:2505.06776*, 2025.
- [13] S. Ross, G. Gordon, and D. Bagnell, “A reduction of imitation learning and structured prediction to no-regret online learning,” ser. *Proceedings of Machine Learning Research*. PMLR, 2011. [Online]. Available: <https://proceedings.mlr.press/v15/ross11a.html>
- [14] Z. Xu, M. Seo, D. Lee, *et al.*, “Learning agile striker skills for humanoid soccer robots from noisy sensory input,” 2025. [Online]. Available: <https://arxiv.org/abs/2512.06571>
- [15] J. Schulman, F. Wolski, P. Dhariwal, *et al.*, “Proximal policy optimization algorithms,” *ArXiv*, vol. abs/1707.06347, 2017. [Online]. Available: <https://api.semanticscholar.org/CorpusID:28695052>
- [16] T. He, J. Gao, W. Xiao, Y. Zhang, *et al.*, “Asap: Aligning simulation and real-world physics for learning agile humanoid whole-body skills,” *arXiv preprint arXiv:2502.01143*, 2025.
- [17] M. Ji, X. Peng, F. Liu, *et al.*, “Exbody2: Advanced expressive humanoid whole-body control,” *arXiv preprint arXiv:2412.13196*, 2024.
- [18] Q. Liao, T. E. Truong, X. Huang, *et al.*, “Beyondmimic: From motion tracking to versatile humanoid control via guided diffusion,” 2025. [Online]. Available: <https://arxiv.org/abs/2508.08241>
- [19] T. Zhang, B. Zheng, R. Nai, *et al.*, “Hub: Learning extreme humanoid balance,” *arXiv preprint arXiv:2505.07294*, 2025.
- [20] Z. Luo, Y. Yuan, T. Wang, *et al.*, “Sonic: Supersizing motion tracking for natural humanoid whole-body control,” *arXiv preprint arXiv:2511.07820*, 2025.
- [21] J. Li, X. Cheng, T. Huang, *et al.*, “Amo: Adaptive motion optimization for hyper-dexterous humanoid whole-body control,” *Robotics: Science and Systems* 2025, 2025.
- [22] Y. Li, Y. Lin, J. Cui, *et al.*, “Clone: Closed-loop whole-body humanoid teleoperation for long-horizon tasks,” 2025.
- [23] Q. Ben, F. Jia, J. Zeng, *et al.*, “Homie: Humanoid loco-manipulation with isomorphic exoskeleton cockpit,” *arXiv preprint arXiv:2502.13013*, 2025.
- [24] T. He, Z. Luo, X. He, *et al.*, “Omni2o: Universal and dexterous human-to-humanoid whole-body teleoperation and learning,” in *8th Annual Conference on Robot Learning*, 2024.
- [25] Y. Ze, Z. Chen, J. P. Araujo, *et al.*, “Twist: Teleoperated whole-body imitation system,” in *Proceedings of The 9th Conference on Robot Learning*, vol. 305. PMLR, 27–30 Sep 2025, pp. 2143–2154. [Online]. Available: <https://proceedings.mlr.press/v305/ze25a.html>
- [26] L. Pinto, M. Andrychowicz, P. Welinder, *et al.*, “Asymmetric actor critic for image-based robot learning,” 2017. [Online]. Available: <https://arxiv.org/abs/1710.06542>
- [27] N. Mahmood, N. Ghorbani, N. F. Troje, G. Pons-Moll, *et al.*, “Amass: Archive of motion capture as surface shapes,” in *The IEEE International Conference on Computer Vision (ICCV)*, Oct 2019. [Online]. Available: <https://amass.is.tue.mpg.de>

APPENDIX

Actor observations	Dimensions
EE-object relative position	9
Object rotation matrix	3
Arm dof position	10
Arm dof velocity	10
Base linear velocity	3
Base angular velocity	3
Projected gravity	3
Relative goal position	2
Last action	13
Total dim	56
Priv. critic observations	Dimensions
Object contact	1
Object CoM	3
Object mass	9
Object dimension	9
Object linear velocity	3
Object angular velocity	3
Total dim	84

TABLE III: Observation dimensions for an asymmetric actor-critic setup used to train the high-level teacher policy.

Student proprioception observations	Dimensions
Projected gravity	3
Base angular velocity	3
Dof position	23
Dof velocity	23
Last action	13
Relative goal position	2
Total dim	67
Student image observations	Dimensions
Depth image ($N_{\text{img}} \times H \times W$)	$3 \times 32 \times 32$

TABLE IV: Observation dimensions for the high-level student policy.

Main Reward	Param. σ	Weight
Reach object	$\sigma_{\text{reach}} = 2.0$	1.25
Push object to goal	$\sigma_{\text{push}} = 10.0$	2.5
Head track object	$\sigma_{\text{track}} = 0.5$	1.0
Object-goal alignment	$\sigma_{\text{align}} = 0.25$	2.0
Auxiliary Penalty	Expression	Weight
Command smoothness	$\ \mathbf{u}_t - \mathbf{u}_{t-1}\ _2^2$	-1.0
Object balance	$\ \theta_{\text{object tilt}}\ _2^2$	-0.5
Upper joint action	$\ \mathbf{q}_t^{\text{upper}}\ _2^2$	-0.5

TABLE V: Reward terms and weights