Waste-Container Lifting Using Residual Reinforcement Learning On Large-Scale Crane with Underactuated Tools

Qi Li[†,*], Karsten Berns[†]
[†] Robotics Research Lab, University of Kaiserslautern-Landau

Abstract

This paper studies the container lifting phase of urban waste-container recycling task with a hydraulic loader crane and an underactuated discharge unit. The task requires accurate hook–ring alignment under tight geometric tolerances while suppressing oscillations of the suspended unit. To address this, we propose a residual reinforcement learning framework that combines a nominal Cartesian controller for trajectory tracking and anti-sway control with a learned residual policy for compensating unmodeled dynamics. The residual policy is trained with PPO. Simulation results show improved tracking accuracy, reduced oscillations, and higher lifting success than the nominal controller alone.

Keywords: Robotics, Residual Reinforcement Learning, Underactuated Systems

1. Introduction

Container recycling, such as waste-glass and garbage collection, is an important task in urban infrastructure. It is often performed using a truck-mounted hydraulic loader crane equipped with an underactuated discharge unit. To lift and empty a container into the truck, the crane must accurately engage small hooking rings attached to containers located above or below ground level. This task is challenging because cranes are typically commanded in joint space, successful hooking requires high TCP accuracy under tight tolerances, and the underactuated discharge unit can oscillate during motion. Together, these factors make operation physically and cognitively demanding, while the shortage of skilled operators further motivates automation.

Automation of hydraulic machinery has attracted growing attention in robotics, with applications in construction, forestry, recycling, warehousing, and port operations [1–3]. Prior work has studied hydraulic manipulation and lifting using model-based methods [4–6], as well as integrated autonomous systems that combine perception, planning, and control [7]. More recently, reinforcement learning (RL) has been explored for hydraulic equipment, including excavation and other dynamic manipulation tasks [8–10], as well as crane-like and forestry systems with underactuated or suspended loads [11–14]. These studies demonstrate adaptability and scalability, but most do not address accurate hooking under tight tolerances, where large-scale crane dynamics, structural compliance, and load oscillations play a central role. Although high-precision manipulation has been extensively studied for industrial robot arms [15], their methods do not transfer directly to hydraulic cranes because of substantial differences in scale, actuation, and dynamics.

To address this gap, we propose a residual reinforcement learning (RRL) framework for accurate container lifting. Residual RL combines a nominal controller with a learned residual policy that compensates for modeling errors and disturbances [16–18]. In our setting, the nominal Cartesian controller provides reliable trajectory tracking and swing suppression, while the residual policy improves robustness and final positioning accuracy. This hybrid design retains the stability and structure of model-based control while avoiding the difficulty of learning the full task end to end from scratch.

This is a short paper. The full version of this paper can be found online [19].

[*] qili@rptu.de

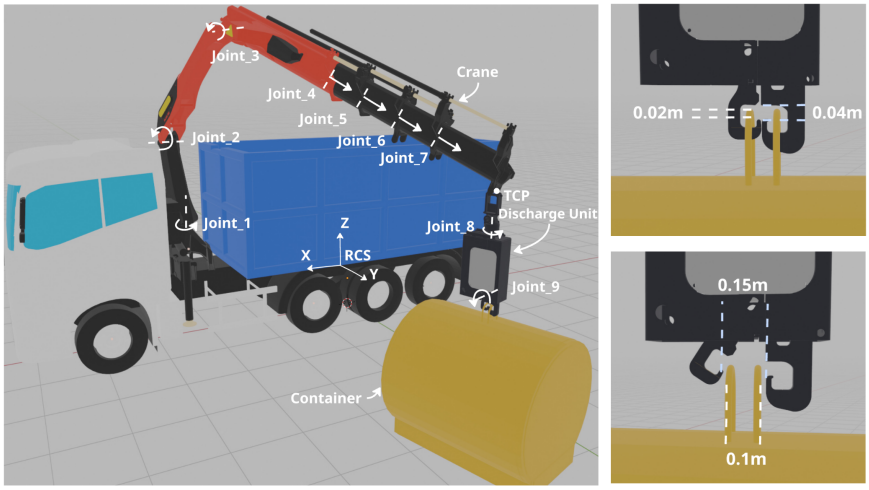

Figure 1. System overview. **Left:** crane kinematic model and task setup. **Right:** close-ups illustrating tight hook–ring tolerances.

2. Methodology

2.1. System and Task

As shown in Fig. 1, the simulated platform in Isaac Lab [20] is a truck-mounted loader crane with a fixed truck base and a 7-DoF crane (3 revolute, 4 prismatic). A discharge unit is attached at the tool center point (TCP) and includes two actuated joints for container rotation and hook opening/closing, as well as two unactuated revolute joints that capture the underactuated swinging behavior of the tool. The container is equipped with hooking rings and weighs between 100 and 700 kg. The task requires accurate TCP positioning to engage the rings under tight geometric tolerances while suppressing oscillations of the underactuated discharge unit. We assume that an external perception system provides accurate object poses. A reference Cartesian TCP trajectory is generated from the initial TCP pose and the sampled container pose, and divided into three phases: approach, horizontal alignment, and lift.

2.2. Residual Reinforcement Learning Framework

As shown in Fig. 2, we adopt a residual reinforcement learning (RRL) framework. A nominal controller provides stable baseline behavior, while a learned residual policy compensates for unmodeled dynamics and improves final positioning accuracy.

Nominal controller. The nominal controller operates in Cartesian space and combines three components: (1) an admittance controller for TCP trajectory tracking, (2) a pendulum-inspired anti-sway term based on the discharge-unit swing state, and (3) damped least-squares inverse kinematics to map desired TCP motion to crane joint velocities. The equations of the controller are provided in Appendix A.

Residual policy. The residual policy outputs a joint-velocity correction that is added to the nominal action. Since precise alignment is most critical near the hooking stage, the residual is applied only during the horizontal alignment phase, while the nominal controller alone is used during approach and lifting. This design preserves reliable baseline behavior and focuses learning on the most accuracy-critical part of the task.

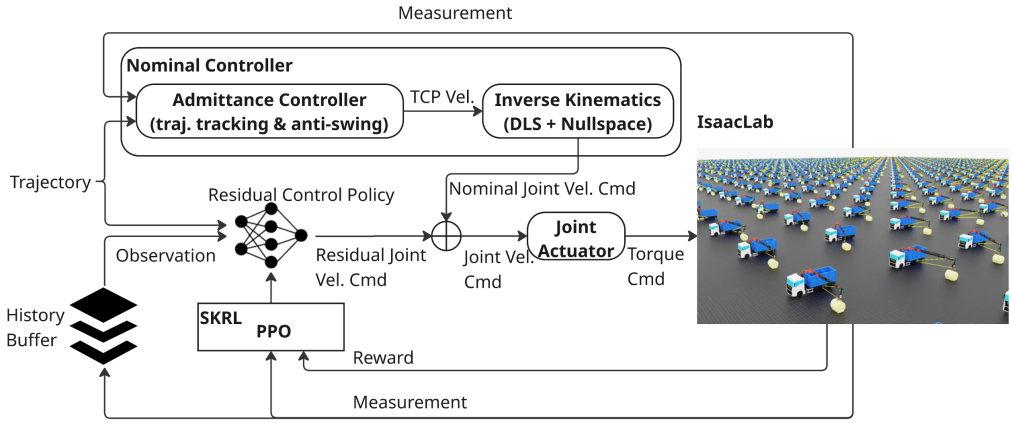

Figure 2. Overall control architecture. A nominal controller provides TCP tracking and swing suppression, while a learned residual policy improves robustness and precision.

2.3. **Policy Training**

The residual policy is trained with PPO using `skrl` [21]. The observation includes crane and discharge-unit states, reference trajectory information, and previous control actions over a short time history, allowing the policy to capture system dynamics and controller behavior. The action is a residual joint-velocity command for the crane.

The reward encourages accurate target tracking, adherence to the reference trajectory [13], forward progress, low oscillation of the discharge unit, successful lifting, and smooth residual actions. Episodes are initialized by randomizing the container pose and the initial TCP pose, from which a reference trajectory is generated. To improve robustness, we apply domain randomization to payload properties, passive-joint friction, actuator parameters, and nominal-controller gains. Full definitions of rewards is given in Appendix B.

3. **Experiments**

All experiments are conducted in Isaac Lab. We evaluate the proposed residual reinforcement learning (RRL) framework on the container lifting task with respect to four criteria: TCP tracking accuracy, trajectory-tube adherence, swing suppression of the discharge unit, and robustness to parameter variations. Test episodes are sampled symmetrically from container poses on both sides of the truck workspace.

3.1. **Trajectory Tracking and Swing Suppression**

Figure 3 shows a representative lifting episode, including the reference and executed TCP trajectories and the corresponding motion sequence. Figure 4 reports TCP tracking error and trajectory-tube deviation for representative episodes. Higher tube adherence is consistently associated with lower tracking error. At the lifting instant, the TCP error remains below 0.04 m, which is sufficient for reliable hook engagement under tight geometric tolerances.

Figure 5 shows the discharge-unit swing angle relative to gravity. Oscillations induced during the motion are effectively damped before lifting. In four of the six representative episodes, the swing angle at lift is below 2.5°, and all episodes remain within acceptable limits. These results indicate that the proposed controller achieves accurate tracking while maintaining low sway.

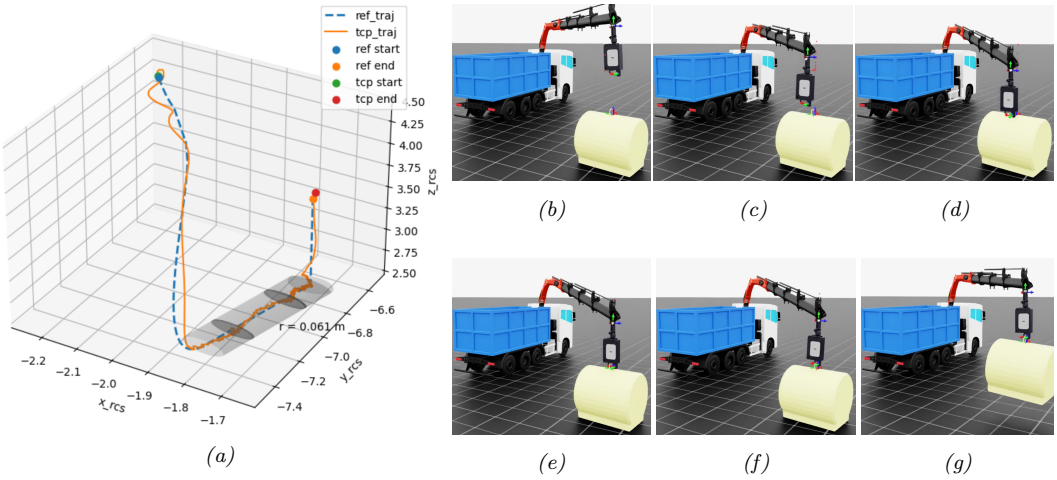

Figure 3. Representative lifting episode: (a) reference vs. executed TCP trajectory; (b–g) motion sequence.

3.2. **Robustness to Parameter Variations**

We further evaluate robustness under parameter settings outside the training randomization range. During training, actuator gains, passive-joint damping, and nominal-controller gains are randomized within a scale range of $[0.5, 1.5]$. At test time, we evaluate three disjoint ranges: a softer regime (0.1–0.49), the training regime (0.5–1.5), and a stiffer regime (1.51–2.0). Each setting is evaluated over 300 episodes. Detailed quantitative results are provided in Appendix C.

The method performs best in the stiff regime, where reduced structural compliance improves tracking and success rate. The soft regime is most challenging and leads to larger tracking errors and lower tube adherence. Nevertheless, the policy remains functional even under this mismatch, achieving a 47.3% lifting success rate. Overall, the results demonstrate that the proposed RRL framework generalizes beyond the training conditions and remains robust under substantial parameter variation.

4. **Ablation Studies**

We evaluate four controller variants: trajectory tracking only, trajectory tracking with RRL, trajectory tracking with anti-swing control, and the full method combining anti-swing control with RRL. For each RRL-based variant, the residual policy is retrained accordingly. All variants are evaluated over 300 episodes. Detailed quantitative results are provided in Appendix D.

The ablation study shows that both anti-swing control and RRL improve task performance. Adding anti-swing control to the nominal controller increases the success rate from 57.3% to 71.3%, confirming its benefit for reducing oscillations. Adding RRL yields a larger improvement: without anti-swing, success increases from 57.3% to 88.3%, and with anti-swing, from 71.3% to 91.7%. Tracking-error differences across variants are relatively small, indicating that coarse trajectory tracking is largely handled by the nominal controller, while anti-swing control and residual learning mainly improve robustness, oscillation reduction, and final task success.

Overall, the best performance is achieved by combining the model-based anti-swing controller with residual reinforcement learning, showing that the two components provide complementary benefits.

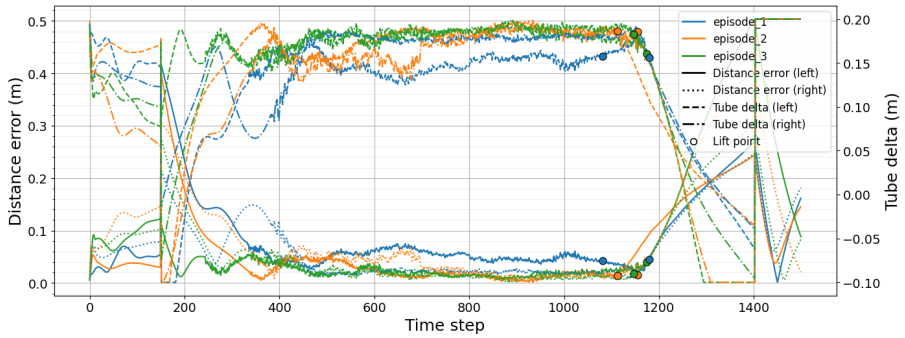

Figure 4. TCP tracking error and trajectory tube delta for representative episodes.

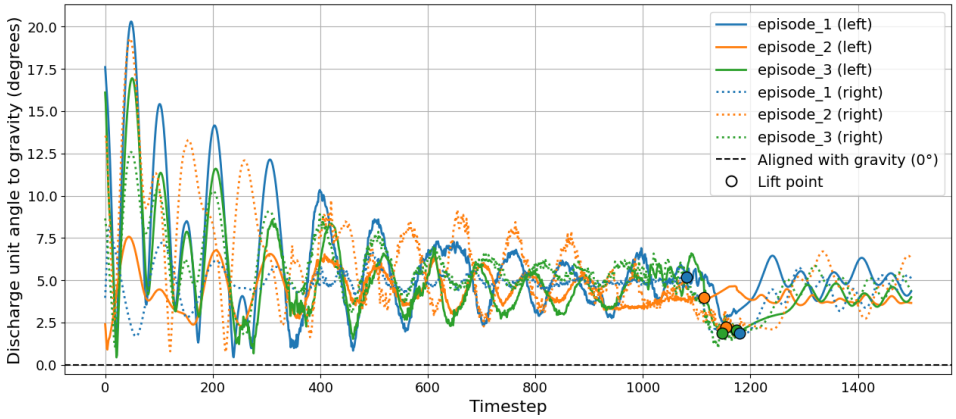

Figure 5. Swing angle of the discharge unit relative to gravity over time.

5. Conclusion and Future Work

We presented a residual reinforcement learning approach for accurate container lifting with a large-scale hydraulic loader crane and an underactuated discharge unit. By combining a nominal Cartesian controller with anti-swing control and a learned residual policy, the method improves tracking accuracy, reduces oscillations, and remains robust to parameter variations in simulation.

Future work will incorporate more realistic hydraulic dynamics, reduce the sim-to-real gap, and validate the approach on a real loader crane platform.

Acknowledgements

This work was carried out within a publicly funded project of the Commercial Vehicle Cluster (CVC), Rheinland-Pfalz, Germany, in collaboration with Palfinger AG. The CAD model of the loader crane and the discharge unit used in this research is provided by Palfinger AG.

We hereby acknowledge that some parts of the manuscript were edited with the assistance of ChatGPT for improving readability and clarity. All technical content and data analysis were developed and verified by authors.

References

[1] R. L. Johns, M. Wermelinger, R. Mascaro, D. Jud, I. Hurkxkens, L. Vasey, M. Chli, F. Gramazio, M. Kohler, and M. Hutter. "A framework for robotic excavation and dry stone construction using on-site materials". In: *Science Robotics* (2023).

[2] J. Andersson, K. Bodin, D. Lindmark, M. Servin, and E. Wallin. *Reinforcement Learning Control of a Forestry Crane Manipulator*. 2021. URL: https://arxiv.org/abs/2103.02315.

[3] Y.-G. Sun, H.-Y. Qiang, J. Xu, and D.-S. Dong. "The Nonlinear Dynamics and Anti-Sway Tracking Control for Offshore Container Crane on a Mobile Harbor". In: *Journal of Marine Science and Technology* (2017).

[4] Q. Ha, M. Santos, Q. Nguyen, D. Rye, and H. Durrant-Whyte. "Robotic excavation in construction automation". In: *IEEE Robotics & Automation Magazine* (2002).

[5] P. H. Chang and S.-J. Lee. "A straight-line motion tracking control of hydraulic excavator system". In: *Mechatronics* ().

[6] J. Mattila, J. Koivumäki, D. G. Caldwell, and C. Semini. "A Survey on Control of Hydraulic Robotic Manipulators With Projection to Future Trends". In: *IEEE/ASME Transactions on Mechatronics* (2017).

[7] D. Jud, S. Kerscher, M. Wermelinger, E. Jelavic, P. Egli, P. Leemann, G. Hottiger, and M. Hutter. "HEAP - The autonomous walking excavator". In: *Automation in Construction* (2021).

[8] P. Egli, L. Terenzi, and M. Hutter. "Reinforcement Learning-Based Bucket Filling for Autonomous Excavation". In: *IEEE Transactions on Field Robotics* (2024).

[9] Y. Zhai, L. Terenzi, P. Frey, D. G. Soto, P. Egli, and M. Hutter. *ExT: Towards Scalable Autonomous Excavation via Large-Scale Multi-Task Pretraining and Fine-Tuning*. 2025. URL: https://arxiv.org/abs/2509.14992.

[10] J. Gruetter, L. Terenzi, P. Egli, and M. Hutter. *Towards Learning Boulder Excavation with Hydraulic Excavators*. 2025. URL: https://arxiv.org/abs/2509.17683.

[11] Q. Wu, N. Sun, T. Yang, and Y. Fang. "Deep Reinforcement Learning-Based Control for Asynchronous Motor-Actuated Triple Pendulum Crane Systems With Distributed Mass Payloads". In: *IEEE Transactions on Industrial Electronics* (2024).

[12] L. Werner, F. Nan, P. Eyschen, F. A. Spinelli, H. Yang, and M. Hutter. *Dynamic Throwing with Robotic Material Handling Machines*. 2024. URL: https://arxiv.org/abs/2405.19001.

[13] F. A. Spinelli, Y. Zhai, F. Nan, P. Egli, J. Nubert, T. Bleumer, L. Miller, F. Hofmann, and M. Hutter. *Large Scale Robotic Material Handling: Learning, Planning, and Control*. 2025. URL: https://arxiv.org/abs/2508.09003.

[14] E. Wallin, V. Wiberg, and M. Servin. *Multi-log grasping using reinforcement learning and virtual visual servoing*. 2024. URL: https://arxiv.org/abs/2309.02997.

[15] L. Han, J. Mao, C. Zhang, R. W. Kay, R. C. Richardson, and C. Zhou. "A systematic trajectory tracking framework for robot manipulators: An observer-based nonsmooth control approach". In: *IEEE Transactions on Industrial Electronics* (2023).

[16] P. Kulkarni, J. Kober, R. Babuška, and C. Della Santina. "Learning Assembly Tasks in a Few Minutes by Combining Impedance Control and Residual Recurrent Reinforcement Learning". In: *Advanced Intelligent Systems* (2022).

[17] M. Alakuijala, G. Dulac-Arnold, J. Mairal, J. Ponce, and C. Schmid. *Residual Reinforcement Learning from Demonstrations*. 2021. URL: https://arxiv.org/abs/2106.08050.

[18] L. Ankile, Z. Jiang, R. Duan, G. Shi, P. Abbeel, and A. Nagabandi. *Residual Off-Policy RL for Finetuning Behavior Cloning Policies*. 2025. URL: https://arxiv.org/abs/2509.19301.

[19] Q. Li and K. Berns. *Residual Reinforcement Learning for Waste-Container Lifting Using Large-Scale Cranes with Underactuated Tools*. 2026. URL: https://arxiv.org/abs/2602.05895.

[20] NVIDIA et al. *Isaac Lab: A GPU-Accelerated Simulation Framework for Multi-Modal Robot Learning*. 2025. URL: https://arxiv.org/abs/2511.04831.

[21] A. Serrano-Muñoz, D. Chrysostomou, S. Bøgh, and N. Arana-Arexolaleiba. "skrl: Modular and Flexible Library for Reinforcement Learning". In: *Journal of Machine Learning Research* (2023).

Appendix A. Nominal Controller

The nominal controller combines Cartesian tracking, anti-swing compensation, and inverse kinematics. The Cartesian control law is

$$F_{\mathrm{cmd}} = K_p(x_{\mathrm{ref}} - x) + K_v(v_{\mathrm{ref}} - v) + M_d a_{xy}, \tag{A.1}$$

with admittance dynamics

$$M_d \dot{v}_d + D_d v_d + K_d(x_d - x_{\mathrm{ref}}) = F_{\mathrm{cmd}}. \tag{A.2}$$

To suppress swing of the underactuated discharge unit, we apply

$$a_{xy} = w_s \begin{bmatrix} k_\theta \hat{\theta}_x + k_\omega \dot{\hat{\theta}}_x \\ k_\theta \hat{\theta}_y + k_\omega \dot{\hat{\theta}}_y \\ 0 \end{bmatrix}, \qquad k_\theta(L) = L\omega_n^2 - g, \quad k_\omega(L) = 2\zeta L \omega_n. \tag{A.3}$$

The desired TCP velocity is mapped to joint velocities using damped least-squares inverse kinematics:

$$^{\mathrm{nor}}u = J_\lambda^+ v_d + (I - J_\lambda^+ J)\, k_{\mathrm{ns}}(q_c - q). \tag{A.4}$$

Appendix B. Residual Policy

Residual reinforcement learning augments the nominal action with a learned correction,

$$u = {}^{\mathrm{nor}}u + {}^{\mathrm{res}}u, \tag{B.1}$$

and the residual is applied only during the horizontal alignment phase. The policy is trained with PPO. The reward at time step k is defined as a weighted sum of task-relevant components,

$$\begin{aligned} R_k = {} & c_1 r_k^{\mathrm{target\_coarse}} + c_2 r_k^{\mathrm{target\_fine}} + c_3 r_k^{\mathrm{tube}} + c_4 r_k^{\mathrm{progress}} \\ & + c_7 r_k^{\mathrm{oscillation}} + c_8 r_k^{\mathrm{lifting}} + c_9 r_k^{\mathrm{smooth}}, \end{aligned} \tag{B.2}$$

where c_i are scalar weights.

The reward terms are computed using the TCP position p_k^{tcp}, discharge unit position p_k^d, container position p_k^c, and the current reference control point p_m^{ref}. The individual reward components are defined as

$$r_k^{\mathrm{target\_coarse}} = -\frac{1}{\sigma}\max(0, d_{m,k} - \sigma),$$

$$r_k^{\mathrm{target\_fine}} = 1 - \tanh\left(\frac{d_{m,k}}{\sigma}\right),$$

$$\begin{aligned} r_k^{\mathrm{tube}} = \mathbb{I}\Big[& \delta_{\mathrm{tube},k} \geq 0 \wedge (p_k^{\mathrm{tcp}} - p_{m-1}^{\mathrm{ref}})^\top (p_m^{\mathrm{ref}} - p_{m-1}^{\mathrm{ref}}) \geq 0 \\ & \wedge (p_k^{\mathrm{tcp}} - p_m^{\mathrm{ref}})^\top (p_m^{\mathrm{ref}} - p_{m-1}^{\mathrm{ref}}) \leq 0 \Big], \end{aligned}$$

$$r_k^{\mathrm{oscillation}} = 1 - \tanh\left(\frac{\arccos\left(\dfrac{\vec{v}_k^\top \mathbf{g}}{\|\vec{v}_k\|}\right) - \theta_{\max}}{\theta_{\max}}\right),$$

$$r_k^{\mathrm{lifting}} = \mathbb{I}(z_k^c > z_{\min}), \qquad r_k^{\mathrm{progress}} = \frac{m}{M}, \qquad r_k^{\mathrm{smooth}} = -\sum_{i=1}^{N} \left(a_{k,i}^{\mathrm{res}}\right)^2,$$

where

$$d_{m,k} = \|p_m^{\mathrm{ref}} - p_k^{\mathrm{tcp}}\|_2, \qquad \vec{v}_k = p_k^d - p_k^{\mathrm{tcp}},$$

Appendix C. Robustness Results

Robustness is evaluated under parameter settings outside the training randomization range. During training, actuator gains, passive-joint damping, and nominal-controller gains are randomized within a scale range of $[0.5, 1.5]$. At test time, we consider three disjoint ranges: $[0.1, 0.49]$, $[0.5, 1.5]$, and $[1.51, 2.0]$. Each setting is evaluated over 300 episodes.

Randomization	**Tracking Error(m)**				**Swing Angle(deg)**			
	Mean	**Std**	**Mean** [700,1200)	**Std** [700,1200)	**Mean @lift**	**Std @lift**	**Mean** [700,1200)	**Std** [700,1200)
scale_0.1_0.49	0.147	0.087	0.103	0.127	3.096	3.555	6.271	5.166
scale_0.5_1.5	0.073	0.019	0.026	0.015	2.046	1.316	4.016	1.378
scale_1.51_2.0	0.049	0.008	0.014	0.004	2.584	1.641	4.544	1.414

Randomization	**Tube Delta(m)**				**Success Rate**
	Mean	**Std**	**Mean** [700,1200)	**Std** [700,1200)	$z^c > 0.5\,\mathrm{m}$
scale_0.1_0.49	0.089	0.050	0.114	0.077	47.3%
scale_0.5_1.5	0.140	0.015	0.174	0.015	90.0%
scale_1.51_2.0	0.158	0.008	0.187	0.004	92.3%

Table 1. Performance metrics aggregated over three randomization groups.

Appendix D. Ablation Results

We compare four controller variants: trajectory tracking only, trajectory tracking with RRL, trajectory tracking with anti-swing control, and the full method combining anti-swing control with RRL. Each variant is evaluated over 300 episodes.

Models				**Tracking Error(m)**			**Swing Angle(deg)**			
Nominal Controller			**RRL**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**
Traj. Tracking	**Anti-Swing**					[700,1200)	[700,1200)	**@lift**	**@lift**	[700,1200)	[700,1200)
•	–	–		0.100	0.025	0.089	0.030	3.318	3.100	7.792	2.819
•	–	•		0.071	0.018	0.021	0.014	2.361	1.520	4.355	1.533
•	•	–		0.105	0.031	0.098	0.039	2.719	2.353	6.545	2.880
•	•	•		0.073	0.021	0.027	0.014	2.090	1.440	4.033	1.303

Models				**Tube Delta(m)**				**Success Rate**
Nominal Controller			**RRL**	**Mean**	**Std**	**Mean**	**Std**	$z^c > 0.5\,\mathrm{m}$
Traj. Tracking	**Anti-Swing**					[700,1200)	[700,1200)	
•	–	–		0.110	0.023	0.111	0.030	57.3%
•	–	•		0.140	0.014	0.179	0.014	88.3%
•	•	–		0.105	0.027	0.103	0.039	71.3%
•	•	•		0.139	0.016	0.173	0.014	91.7%

Table 2. Performance metrics aggregated over four controller variants.