Watch Less, Feel More: Sim-to-Real RL for Generalizable Articulated Object Manipulation via Motion Adaptation and Impedance Control

Tan-Dzung Do[1,2] **Nandiraju Gireesh**[1,2] **Jilong Wang**[2] **He Wang**[1,2]

[1]CFCS, School of CS, Peking University [2]Galbot

Figure 1: Our policy learns to infer object motion and intrinsic from history observation, following how humans open a door even with covered eyes. Our impedance-aware policy achieves 84% success rate in the real world, using only one first-frame RGBD image.

Abstract: Fine manipulation tasks like articulated object manipulation pose a unique challenge as the object itself represents a dynamic environment. In this work, we present a novel RL-based pipeline equipped with variable impedance control and motion adaptation for generalizable articulated object manipulation, focusing on smooth and dexterous motion during zero-shot sim-to-real transfer. To mitigate the sim-to-real gap, our pipeline diminishes reliance on vision by extracting useful low-dimensional data via off-the-shelf modules and inferring object motion and intrinsic properties via observation history. Furthermore, we develop a well-designed training setting with great randomization and a specialized reward system that enables multi-staged, end-to-end manipulation without heuristic motion planning. To the best of our knowledge, our policy is the first to report 84% success rate for extensive real-world experiments with various unseen objects. Project website: https://watch-less-feel-more.github.io/

Keywords: Sim-to-Real, Reinforcement Learning, Impedance Control

1 Introduction

A generalist robot represents a big milestone for the robot learning community, with the potential to revolutionize our daily life. Amid the great progress in the embodied AI field in these couple of years[1, 2, 3, 4], generalizable articulated object manipulation remains an open question due to various reasons. One major challenge is that the true articulation characteristics (e.g. pivot center, friction, stiffness) could only be identified after physical contact is made. As a result, it necessitates a closed-loop pipeline that can adaptively infer these characteristics during the manipulation stage. Additionally, fine manipulation tasks like articulated object manipulation are also hard due to the joint constraints of objects. These constraints require the applied actions to comply with the actual object joint motion to prevent potential damages.

Recent articulated object manipulation works often rely on visual information as the dominant input for their pipelines, either in the form of pointcloud [5, 6, 7] or RGB images [8, 9, 10, 11, 12, 6], to predict actionable parts and action sequence. This action sequence is then directly executed in an open-loop manner neglecting all possible physical interaction with objects as well as their intrinsic properties. Other works leverage RL backbones [13, 14, 15, 16] to output actions in a closed-loop

8th Conference on Robot Learning (CoRL 2024), Munich, Germany.

fashion based on vision feedback but suffer the substantial vision sim-to-real gap inherited from vision modules [9, 6]. Additionally, during the manipulation stage, this approach might output suboptimal action due to the occlusion of the actionable part.

In this project, we propose combining closed-loop RL with learnable impedance control for generalizable articulated object manipulation. First, we use observation history to manipulate objects in a closed-loop fashion as an alternative for vision input. We evidence our intuition by exemplifying how humans can open a door in the dark: we gradually adjust the opening actions while inferring the door handle position, even without direct vision input (Fig. 1). This paradigm helps mitigate the vision sim-to-real gap and implicitly learn the movement of objects, thus enabling a generalizable closed-loop pipeline. Second, we address the importance of compliant action for articulated object manipulation by introducing variable impedance control to our pipeline. While implementing a high-frequency variable impedance controller in simulation, we also learn its parameters jointly with our RL policy to generate smooth and continuous motions that comply with object joint movements. We find learning motion instead of a single action or discrete waypoints [17, 11, 12] can yield a higher success rate in the real world.

2 Proposed method

2.1 Online policy distillation with Observation History

Articulated object manipulation poses a unique challenge compared to rigid object manipulation because the object itself is a dynamic environment. The fact that object motion can only be observed via physical interactions or that joint ground-truth position is hidden inside the object resembles locomotion tasks where environment parameters (e.g. terrain friction, slope) are difficult to predict. To this end, we adopt the online policy distillation pipeline, which is widely applied for locomotion tasks [18, 3, 4, 19], and learn two separate modules: Adaptation Module σ and Privileged Observation Encoder ϕ (Fig. 2).

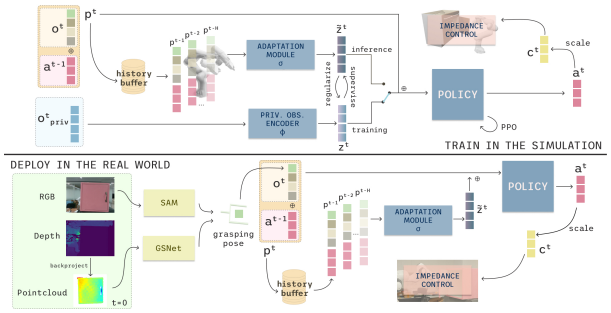

Figure 2: In the simulation, we train a Privileged Observation Encoder ϕ to extract the latent representation of privileged information z^t and simultaneously train an Adaptation Module σ to infer this representation $\tilde{z}^t$ from $H = 10$ previous (o^t, a^{t-1}) pairs. In the real world, we rollout trained policy with Adaptation Module σ in an end-to-end manner, executing smooth reaching, grasping, and manipulating with only first-frame RGBD image.

Privileged Observation Encoder ϕ is a shallow MLP, which is utilized during training to learn the latent representation z^t of privileged observations. This 20-dimensional vector is then concatenated with an (observation, action) pair $p^t = (o^t \oplus a^{t-1})$ at the current timestep to form actor inputs. We design the Adaptation Module σ to be a temporal architecture to extract latent information about the environment from $H = 10$ p^t pairs. We keep only parts of action history as inputs for σ: position command Δ^t_{xyz}, gripper command G^t, and controller gain k^t_p.

As the conventional two-staged teacher-student pipeline might result in realizability gap and sim-to-real gap [18], we simultaneously train Adaptation Module and Privileged Observation Encoder in a single training by formulating a supervision-regularization loss $\lambda\|z - sg[\tilde{z}]\|_2 + \|sg[z] - \tilde{z}\|_2$ on top of PPO objectives.

2.2 Reward Design and Domain Randomization

While the proposed framework is adopted widely for locomotion tasks, it remains non-trivial how to transfer this pipeline for fine-manipulation tasks like articulated object manipulation. To facilitate a single end-to-end policy that can efficiently perform multi-staged motions, we introduce stage-conditioned rewards, including task-aware rewards and motion-aware rewards. Task-aware rewards focus on executing a proper motion sequence, complying grasp-then-open order, rather than cheating to gain success rewards immediately. Motion-aware rewards encourage our policy to generate smooth motions while maintaining a high success rate. We argue that incorporating these regularization terms is crucial and helps bridge the sim-to-real gap by preventing unnecessary motion or non-achievable target poses. More details in Appendix B.

Recent manipulation works [13, 6, 8] demonstrate that training a policy with domain randomization may benefit sim-to-real transfer. To cover a reasonable workspace for real-world settings, we randomize object positions and object yaw rotations during training. In terms of physical intrinsic, we vary the joint friction, stiffness, and mass for more robust sim-to-real transfer. For desired grasping poses, after we infer a pose from part bounding boxes, we introduce random noise along y and z axes, together with a random rotation target from a pre-defined spherical cone.

2.3 Variable Impedance Control

The design of impedance control follows a mass-spring-damper system that can dynamically adjust target setpoints based on feedback force as well as the stiffness of the environment: $M(\ddot{x_c} - \ddot{x_d}) + D(\dot{x_c} - \dot{x_d}) + K(x_c - x_d) = F_{ext}$, where M is the mass-inertia matrix of the robot, D is the damping matrix, K is the stiffness matrix, and $[\ddot{x_c}, \dot{x_c}, x_c]$ is impedance trajectory outputs. In our pipeline, we learn to predict the stiffness factor k_p of our Cartesian impedance controller and expand it into a six-dimensional diagonal matrix K. Following [20, 21], we scale the gain by $c^t_{k_p} = clip(a^t_{k_p}, -1, 1) * 40 + 100$ to ensure system stability and then infer the damping matrix with the critical damping condition $D = 2\sqrt{MK}$.

3 Experiments

To verify the effectiveness of the proposed method, we conduct extensive evaluations in both simulation and real-world settings.

Figure 3: We extensively evaluate our policy in the real world with a wide range of unseen objects with diverse characteristics in a large workspace.

3.1 Data and Task Settings

In the simulation, following the settings of PartManip [13], we conduct our experiments with the large-scale PartNet-Mobility dataset [22] including 346 objects in IsaacGym. In the real-world setting, we perform experiments with a variety of household objects using a Franka Emika equipped with a RealSenseD415. We leverage Segment Anything (SAM) [23] for actionable part pointcloud extraction using a first-framed RGBD image and GSNet [24] for grasp prediction. We evaluate our proposed pipeline with two following tasks: 1) OpenDoor/OpenDoor+: A door is initially closed, the agent needs to open the door larger than 15%/80% of the maximum door swing; 2) OpenDrawer/OpenDrawer+: A drawer is initially closed, the agent needs to open the drawer larger than 20%/80% of the maximum opening length. The key requirement for our task setting is that the gripper should firmly grasp the handle while opening the door without cheating by opening from the side or with the robot body. We adopt Success Rate (SR) as the major evaluation metric.

3.2 Baselines and Ablation Study Design

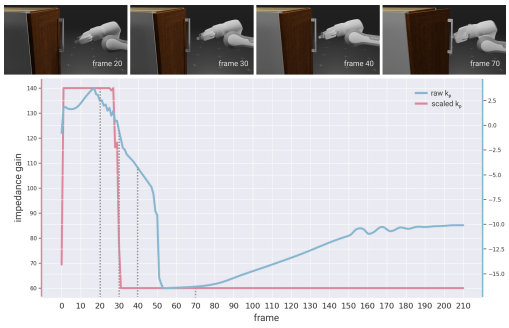

Figure 4: Our learned controller gain actively adapts to the manipulation stages: stiffer while reaching, softer while opening.

We compare our proposed method with articulated-object manipulation pipelines that follow sim-to-real RL paradigm including PPO, Where2Act [11], PartManip [13], RGBManip [8], GAPartNet [6].

To highlight the contribution and effectiveness of each module within our approach, we conducted four comprehensive ablation studies: Ours w/o Policy Distillation, Ours w/o Variable Impedance Control, Ours w/o Regularization, Ours w/o Randomization.

3.3 Results and Findings

Results of simulation experiments are shown in Table 1, from which we can see that our method maintains consistently strong performance on the evaluation set without a sharp drop, highlighting the excellent generalization ability of our approach. We also find our controller learns to adapt to different manipulation stages, even without any direct gain rewards (Fig. 4).

Our policy rollout performance in real world can be found in Table 2. We conduct 50 experiments for our pipeline and each ablated model (500 runs in total) on diverse objects (Fig. 3). For OpenDoor+, we find $6/50$ inferences fail during Grasping Stage while only $4/50$ fail during Opening Stage, suggesting that if a stable grasping pose is initiated, our policy might yield $40/44 = 0.90\%$ SR. For OpenDrawer+, $7/8$ failure cases are due to unsuccessful grasping.

With the ablation study results demonstrated in Table 2, apart from SR drop in both simulation and the real world, we aim to highlight the non-smooth motions of real-world executions. For *W/o Impedance Control*, we find the main reason for failure cases (40% drop) is the low flexibility of position control, which requires each predicted action to be executed precisely. This would generate large joint torque to overcome the feedback force of objects, resulting in the robot arm being triggered to stop. In simulation, this behavior does not seem to severely hurt the performance, as evidenced by > 0.8 success rate. However, in the real world, large torque is substantially dangerous and would trigger an emergency stop, emphasizing the necessity for impedance control.

Baselines	Type	OpenDoor		OpenDrawer		OpenDoor+		OpenDrawer+	
		Train	Test	Train	Test	Train	Test	Train	Test
PPO	Closed-loop	0.04	0.05	0.09	0.11	0.02	0.02	0.03	0.02
Where2act [11]	Open-loop	0.22	0.14	0.31	0.27	0.02	0.02	0.01	0.01
RGBManip [8]	Closed-loop	0.62	0.59	0.63	0.67	0.38	0.41	0.49	0.42
GAPartNet [6]	Open-loop	0.70	0.75	0.51	0.59	0.40	0.44	0.45	0.49
PartManip [13]	Closed-loop	0.75	0.70	0.83	0.77	0.68	0.57	0.62	0.59
Ours	Closed-loop	**0.96**	**0.95**	**0.97**	**0.96**	**0.96**	**0.93**	**0.97**	**0.96**

Table 1: Comparison with Baselines in Simulation

For *W/o Distillation* and *W/o Randomization*, the policy often finishes the task halfway, even when we manually tune a stiffer base value for the impedance controller. We claim that this behavior is due to the physics sim-to-real gap which resulted from non-diverse training settings and short-term observation. For *W/o Regularization*, the reaching and opening motions are jerky, which are highly undesirable and result in grasp failure and contact lost during execution.

Methods	OpenDoor+			OpenDrawer+		
	Train	Test	Real	Train	Test	Real
W/o Distillation	0.80	0.77	0.62	0.78	0.74	0.60
W/o Imp. Ctr.	0.84	0.82	0.40	0.90	0.90	0.44
W/o Reg.	0.88	0.86	0.64	0.92	0.87	0.70
W/o Rand.	0.91	0.89	0.66	0.93	0.91	0.64
Ours	**0.96**	**0.93**	**0.80**	**0.97**	**0.96**	**0.84**

Table 2: Ablation Study and Real-world Performance

4 Conclusion

In this work, we introduce a reliable RL policy for articulated object manipulation that can be seamlessly deployed in diverse real-world settings. Our experiments, conducted in both simulation and real-world scenarios, achieve over 80% SR to unseen objects and demonstrate the great generalizability of our policy.

Acknowledgments

We thank all reviewers for their helpful feedback and fruitful discussions.

References

[1] J. Zhang, N. Gireesh, J. Wang, X. Fang, C. Xu, W. Chen, L. Dai, and H. Wang. Gamma: Graspability-aware mobile manipulation policy learning based on online grasping pose fusion. *arXiv preprint arXiv:2309.15459*, 2023.

[2] H. Ha, Y. Gao, Z. Fu, J. Tan, and S. Song. UMI on legs: Making manipulation policies mobile with manipulation-centric whole-body controllers. 2024.

[3] T. Portela, G. B. Margolis, Y. Ji, and P. Agrawal. Learning force control for legged manipulation. *arXiv preprint arXiv:2405.01402*, 2024.

[4] T. He, C. Zhang, W. Xiao, G. He, C. Liu, and G. Shi. Agile but safe: Learning collision-free high-speed legged locomotion. *arXiv preprint arXiv:2401.17583*, 2024.

[5] S. Ling, Y. Wang, R. Wu, S. Wu, Y. Zhuang, T. Xu, Y. Li, C. Liu, and H. Dong. Articulated object manipulation with coarse-to-fine affordance for mitigating the effect of point cloud noise. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 10895–10901. IEEE, 2024.

[6] H. Geng, H. Xu, C. Zhao, C. Xu, L. Yi, S. Huang, and H. Wang. Gapartnet: Cross-category domain-generalizable object perception and manipulation via generalizable and actionable parts. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7081–7091, 2023.

[7] B. Eisner, H. Zhang, and D. Held. Flowbot3d: Learning 3d articulation flow to manipulate articulated objects. *arXiv preprint arXiv:2205.04382*, 2022.

[8] B. An, Y. Geng, K. Chen, X. Li, Q. Dou, and H. Dong. Rgbmanip: Monocular image-based robotic manipulation through active object pose estimation. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 7748–7755. IEEE, 2024.

[9] H. Xiong, R. Mendonca, K. Shaw, and D. Pathak. Adaptive mobile manipulation for articulated objects in the open world. *arXiv preprint arXiv:2401.14403*, 2024.

[10] S. Bahl, R. Mendonca, L. Chen, U. Jain, and D. Pathak. Affordances from human videos as a versatile representation for robotics. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13778–13790, 2023.

[11] K. Mo, L. J. Guibas, M. Mukadam, A. Gupta, and S. Tulsiani. Where2act: From pixels to actions for articulated 3d objects. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6813–6823, 2021.

[12] C. Ning, R. Wu, H. Lu, K. Mo, and H. Dong. Where2explore: Few-shot affordance learning for unseen novel categories of articulated objects. *Advances in Neural Information Processing Systems*, 36, 2024.

[13] H. Geng, Z. Li, Y. Geng, J. Chen, H. Dong, and H. Wang. Partmanip: Learning cross-category generalizable part manipulation policy from point cloud observations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2978–2988, 2023.

[14] Z. Xu, Z. He, and S. Song. Universal manipulation policy network for articulated objects. *IEEE robotics and automation letters*, 7(2):2447–2454, 2022.

[15] Y. Geng, B. An, H. Geng, Y. Chen, Y. Yang, and H. Dong. Rlafford: End-to-end affordance learning for robotic manipulation. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 5880–5886. IEEE, 2023.

[16] Y. Li, X. Zhang, R. Wu, Z. Zhang, Y. Geng, H. Dong, and Z. He. Unidoormanip: Learning universal door manipulation policy over large-scale and diverse door manipulation environments. *arXiv preprint arXiv:2403.02604*, 2024.

[17] R. Wu, Y. Zhao, K. Mo, Z. Guo, Y. Wang, T. Wu, Q. Fan, X. Chen, L. Guibas, and H. Dong. Vat-mart: Learning visual action trajectory proposals for manipulating 3d articulated objects. *arXiv preprint arXiv:2106.14440*, 2021.

[18] Z. Fu, X. Cheng, and D. Pathak. Deep whole-body control: learning a unified policy for manipulation and locomotion. In *Conference on Robot Learning*, pages 138–149. PMLR, 2023.

[19] A. Kumar, Z. Fu, D. Pathak, and J. Malik. Rma: Rapid motor adaptation for legged robots. 2021.

[20] X. Zhang, C. Wang, L. Sun, Z. Wu, X. Zhu, and M. Tomizuka. Efficient sim-to-real transfer of contact-rich manipulation skills with online admittance residual learning. In *Conference on Robot Learning*, pages 1621–1639. PMLR, 2023.

[21] X. Zhang, M. Tomizuka, and H. Li. Bridging the sim-to-real gap with dynamic compliance tuning for industrial insertion. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 4356–4363. IEEE, 2024.

[22] K. Mo, S. Zhu, A. X. Chang, L. Yi, S. Tripathi, L. J. Guibas, and H. Su. Partnet: A large-scale benchmark for fine-grained and hierarchical part-level 3d object understanding. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 909–918, 2019.

[23] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, et al. Segment anything. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4015–4026, 2023.

[24] C. Wang, H. Fang, M. Gou, H. Fang, J. Gao, C. Lu, and S. J. Tong. Graspness discovery in clutters for fast and accurate grasp detection. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 15944–15953, 2021.

A Action and Observation Space

We design our framework to facilitate one dexterous action prediction at a time instead of short-horizon primitive actions. Our action for each step $a^t \in R^{11}$ includes the target delta position $\Delta^t_{xyz} \in R^3$, target 6D orientation $R^t \in R^6$, gripper action $G^t \in R^1$, and impedance control parameter $k^t_p \in R^1$. Our raw robot action a^t is later converted into robot commands $c^t \in R^9$ using an action scaler.

Our observation o^t consists of desired grasping pose $g^t \in R^7$, robot joint configuration $q^t \in R^7$, robot-object relative distance $\delta^t \in R^1$, end-effector pose $ee^t \in R^9$ with three-dimensional position and 6D rotation, and graspability $1^t_{grasp} \in R^1$. Here, desired grasping poses are directly inferred from the handle bounding box in the simulation and from off-the-shelf grasp prediction modules in the real world. Our graspability signal is a distance-based and contact-aware condition, rather than a direct command for open/close gripper. In terms of task-aware observation, for instance, with DoorOpen task, we incorporate noisy pivot center $\tilde{r}^t_{pivot} \in R^3$, noisy pivot radius $\tilde{r}^t_{radius} \in R^1$, and right-hinged boolean $\tilde{r}^t_{rh} \in R^1$. These motion-related arguments serve as high-level guidance for smoother implementation.

$$o^t = [g^t, q^t, \delta^t, ee^t, 1^t_{grasp}, \tilde{r}^t_{pivot}, \tilde{r}^t_{radius}, \tilde{r}^t_{rh}] \in R^{30}$$

Our privileged observation o^t_{priv}, including values that are difficult to track in real-world settings, is used only in simulation for better environment understanding. These values are: pivot center $r^t_{pivot} \in R^3$, pivot radius $r^t_{radius} \in R^1$, object stiffness $r^t_{stiff} \in R^1$, object mass $r^t_m \in R^1$, object joint position $q^t_{obj} \in R^1$, handle grasped signal $1^t_{grasped} \in R^1$.

$$o^t_{priv} = [r^t_{pivot}, r^t_{radius}, r^t_m, r^t_{stiff}, q^t_{obj}, 1^t_{grasped}] \in R^8$$

B Reward functions

Term	Formula	Weight
	Nomenclature	
1_d	$\delta \leq 0.05$	-
1_{dy}	$0.02 \leq \delta \leq 0.08$	-
1_g	$\delta \leq 0.015 \wedge 1_{contact}$	-
τ	joint torque	-
$\dot{q}$	joint velocity	-
w_{len}	episode length weight	-
$a_t[y]$	action on y axis	-
$a_t[z]$	action on z axis	-
	Task-aware rewards	
success	$0.05^{1_d} * 0.5^{1_g} * 1_s$	40.0
distance	$\exp(-10 * (2\delta^{0.5}))/2 * 0.8^{1_g}$	0.6
object state	$q_{obj} * 0.5^{1_g} * 0.5^{1_d} * w_{len}$	1.0
grasp	$0.2 * 1_g$	0.05
	Motion-aware rewards	
energy	$\sum(\tau\dot{q})^{0.5} * 1_g$	-0.05
track pos.	$\exp(-4(c_{pos} - ee_{pos})) * 1_d$	0.025
track rot.	$\exp(-4\Delta(c_{ori} - ee_{ori})) * 1_d$	0.004
smoothness	$\sum 1_{[sgn(a_t) \neq sgn(a_{t-1})]} * (a_t - a_{t-1})$	-0.001
y reg.	$1_{dy} * (a_t[y] * 15)^2$	-0.005
z reg.	$1_g * (a_t[z] * 15)^2$	-0.07

Table 3: Reward functions