Supplementary Materials for Learning Environment-Aware Affordance for 3D Articulated Object Manipulation under Occlusions

Anonymous Author(s)
Affiliation
Address
email

1 Introductory Video

We have attached a **video** introducing our work, with real-world demonstrations in the last.

2 More Details on Simulation and Settings

Following Where2Act [5], we design our interactive simulation environment based on SAPIEN, using the same set of simulation parameters for all interaction trials.

For **general simulation settings**, we use frame rate 500 fps, tolerance length 0.001, tolerance speed 0.005, solver iterations 20 (for constraint solvers related to joints and contacts), with Persistent Contact Manifold (PCM) disabled (for better simulation stability), with disabled sleeping mode (*i.e.* no locking for presumably still rigid bodies in simulation), and all the other settings as default in SAPIEN release.

For **physical simulation**, we use the standard gravity 9.81, static friction coefficient 4.0, dynamic friction coefficient 4.0, and restitution coefficient 0.01. For the object articulation dynamics simulation, we use stiffness 0 and damping 10.

For the **rendering**, we use OpenGL-based rasterization rendering for the fast speed of simulation. We set three point lights around the object (one at the front, one from back-left and one from back-right) for lighting the scene, with mild ambient lighting as well. The camera is set to have near plane 0.1, far plane 100, resolution 448, and field of view 35°.

For **3D partial point cloud scan inputs**, we back-project the depth image into a foreground point cloud, by rejecting the far-away background depth pixels, and then perform furthest point sampling to get a 10K-size point cloud scan.

For **robot arm movement**, we use RRT Planner [7, 2, 4] equipped with PID controller to generate and execute a certain path towards the target.

For an **interaction trial** to be considered successful, it not only needs to cause considerable part motion along intended direction. To avoid the extreme data unbalance in pulling data, we manually set handle mask on our simulator and assign half of the interactions on the handles. To simplify the consideration of different interaction directions' impact on affordance, we set every interaction to move along the normal direction of the target point.

Submitted to 37th Conference on Neural Information Processing Systems (NeurIPS 2023). Do not distribute.

3 More Data Details and Visualization

In Table 1, we summarize our data statistics. In Fig. 1, we visualize our simulation assets from ShapeNet [1] and PartNet [6] that we use in this work.

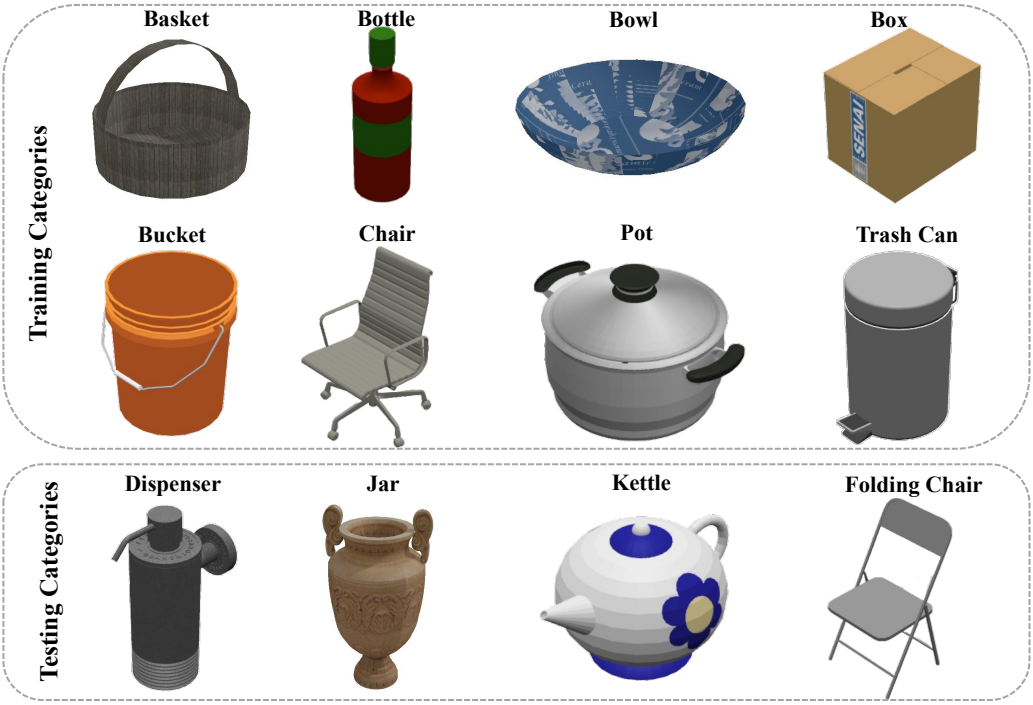

Figure 1: **Our simulation assets from ShapeNet [1] and PartNet [6].**

Train-Cats	All	Basket	Bottle	Bowl	Box	Bucket	Chair	Pot	TrashCan
Train-Data	367	77	16	128	17	27	61	16	25
Test-Data	128	31	4	44	5	9	20	5	10
Test-Cats	**All**	**Dispenser**	**Jar**	**Kettle**	**FoldingChair**				
Test-Data	589	9	528	26	26				

Table 1: **Occluder Dataset Statistics.** We use 1,084 different shapes in ShapeNet [1] and PartNet-Mobility [6], covering 12 commonly seen indoor occluder categories. We use 8 training categories (split into 367 training shapes and 128 test shapes), and 4 test categories with 589 shapes networks have never seen in training.

4 More Training Details

4.1 Hyper-parameters

We set the batch size to 30, and use Adam Optimizer [3] with 0.001 as the initial learning rate.

We use const 2.00 as the boundary constant in α contrastive learning, and 1.00 as the balancing coefficient λ_{CL} in the total loss.

4.2 Computing Resources

We use PyTorch as our Deep Learning framework, and RTX GeForce 3090 (20GB GPU) for training and inference.

4.3 Error Bar

We run an experiment three times and report the average result.

5 More Results and Analysis

Fig. 2 3 4 5 demonstrate comparions with baselines and ablations. Fig. 6 shows the whole occlusion fields. Fig. 7 shows real-world demonstrations with analysis in the caption.

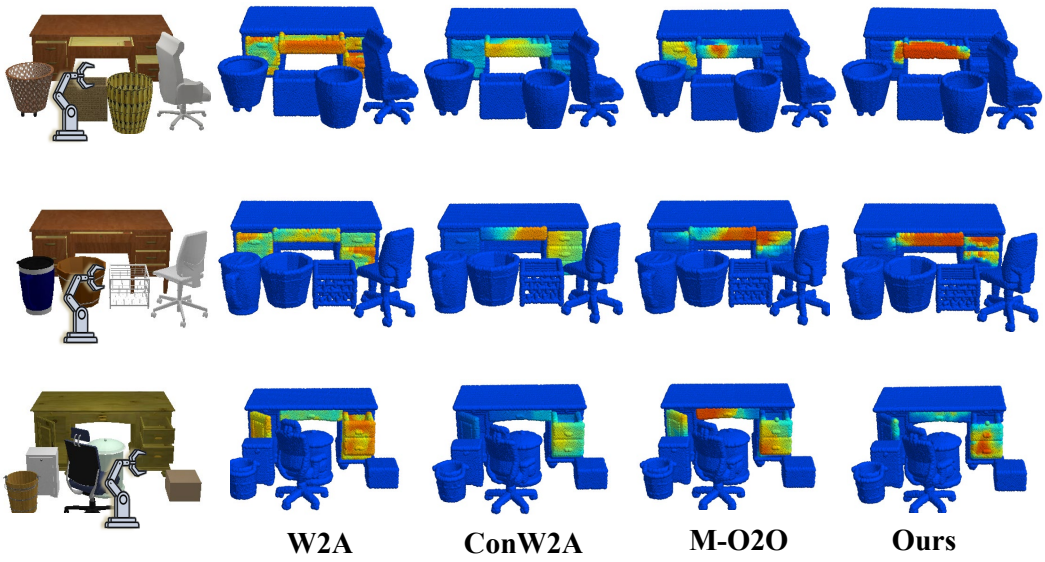

Figure 2: **More Qualitative Comparisons between Our Method and Baselines in Pushing.**

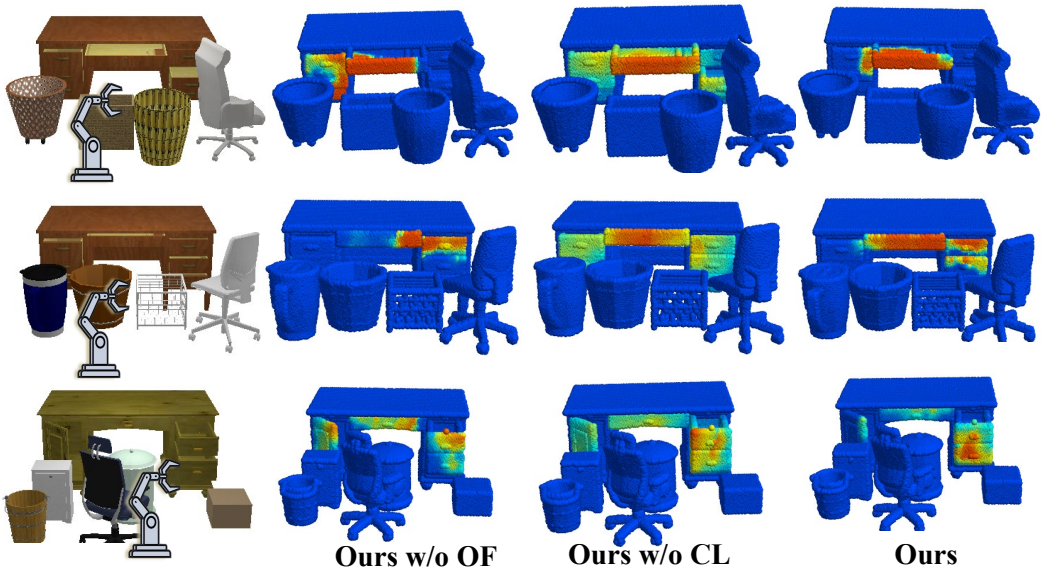

Figure 3: **More Qualitative Comparisons between Our Method and Ablations in Pushing.**

Figure 4: **More Qualitative Comparisons between Our Method and Baselines in Pulling.**

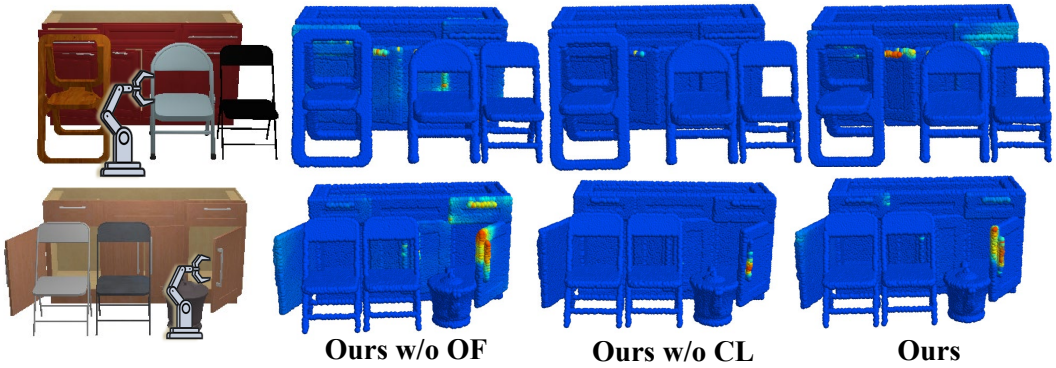

Figure 5: **More Qualitative Comparisons between Our Method and Ablations in Pulling.**

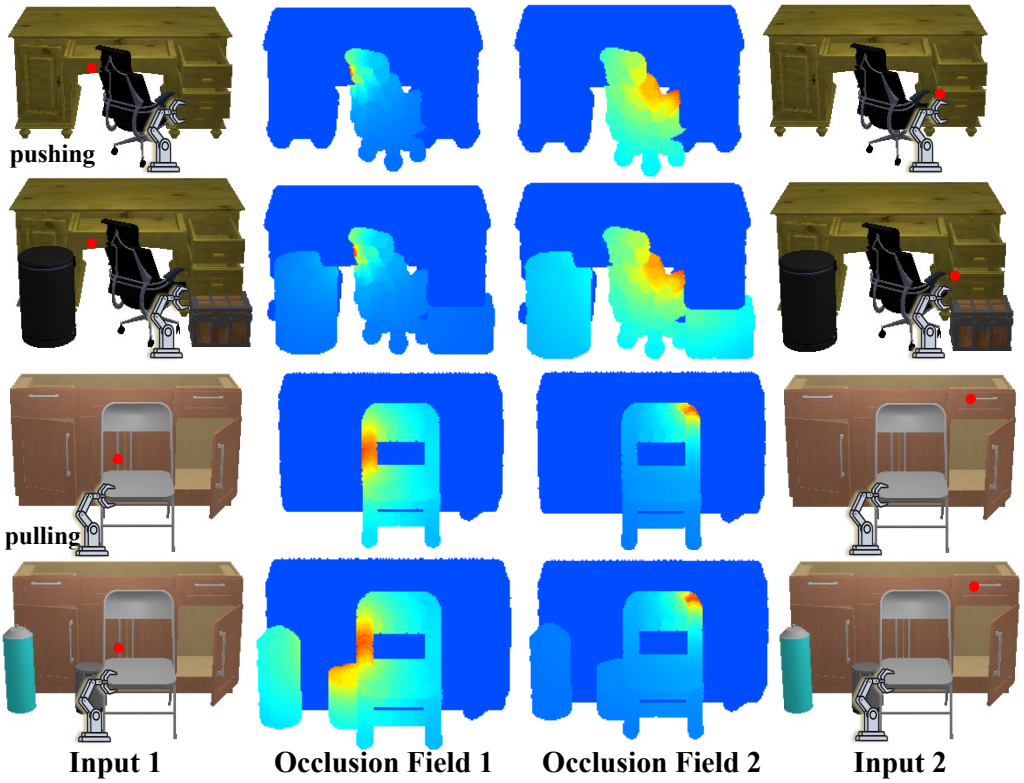

Figure 6: **Visualization of the whole Occlusion Fields.**

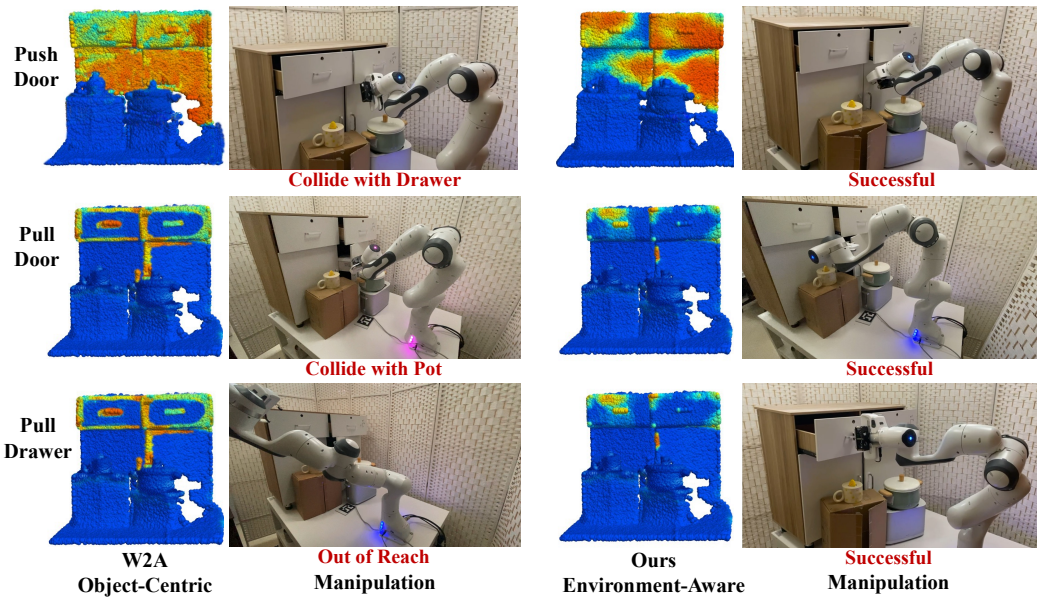

Figure 7: **Real-World Demonstrations of Manipulation Policy Guided by Object-Centric Affordance and Our Proposed Environment-Aware Affordance.** It is clear that environment-aware affordance can help avoid out-of-reach situations and collisions with other self-parts or objects.

6 Future Work on Robot-Target Conditioned Contrastive Learning

Limited to simulator configuration, our contrastive learning method only considers a limited augmentation distribution $A(\cdot \mid \bar{x})$ for each anchor scene $\bar{x} \in \mathcal{X}$ while the marginal distribution $A(\cdot) = \mathbb{E}_{\bar{x}} A(\cdot \mid \bar{x})$ is complete. The augmentation distribution $A(\cdot \mid \bar{x})$ only includes one more occluder at the edge of $\bar{x}$, and neglects the potential augmentation methods by choosing similar target points. Future methods can be applied with a better similarity metric of comparing different things and improve our self-supervised learning paradigm. Nevertheless, our current implementation version is already simple and efficient with good performance, and the above discussion is just the direction worth future study.

7 Code

We will release our code upon acceptance.

References

[1] Angel X Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, et al. Shapenet: An information-rich 3d model repository. *arXiv preprint arXiv:1512.03012*, 2015.

[2] Jiayuan Gu, Fanbo Xiang, Xuanlin Li, Zhan Ling, Xiqiang Liu, Tongzhou Mu, Yihe Tang, Stone Tao, Xinyue Wei, Yunchao Yao, et al. Maniskill2: A unified benchmark for generalizable manipulation skills. *arXiv preprint arXiv:2302.04659*, 2023.

[3] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *The 3rd International Conference for Learning Representations*, 2015.

[4] James J Kuffner and Steven M LaValle. Rrt-connect: An efficient approach to single-query path planning. In *Proceedings 2000 ICRA. Millennium Conference. IEEE International Conference on Robotics and Automation. Symposia Proceedings (Cat. No. 00CH37065)*, volume 2, pages 995–1001. IEEE, 2000.

[5] Kaichun Mo, Leonidas Guibas, Mustafa Mukadam, Abhinav Gupta, and Shubham Tulsiani. Where2act: From pixels to actions for articulated 3d objects. In *International Conference on Computer Vision (ICCV)*, 2021.

[6] Kaichun Mo, Shilin Zhu, Angel X. Chang, Li Yi, Subarna Tripathi, Leonidas J. Guibas, and Hao Su. PartNet: A large-scale benchmark for fine-grained and hierarchical part-level 3D object understanding. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

[7] Ioan A Sucan, Mark Moll, and Lydia E Kavraki. The open motion planning library. *IEEE Robotics & Automation Magazine*, 19(4):72–82, 2012.