End-to-End Affordance Learning for Robotic Manipulation

Student: Yiran Geng*, Boshi An*

Advisor: Hao Dong

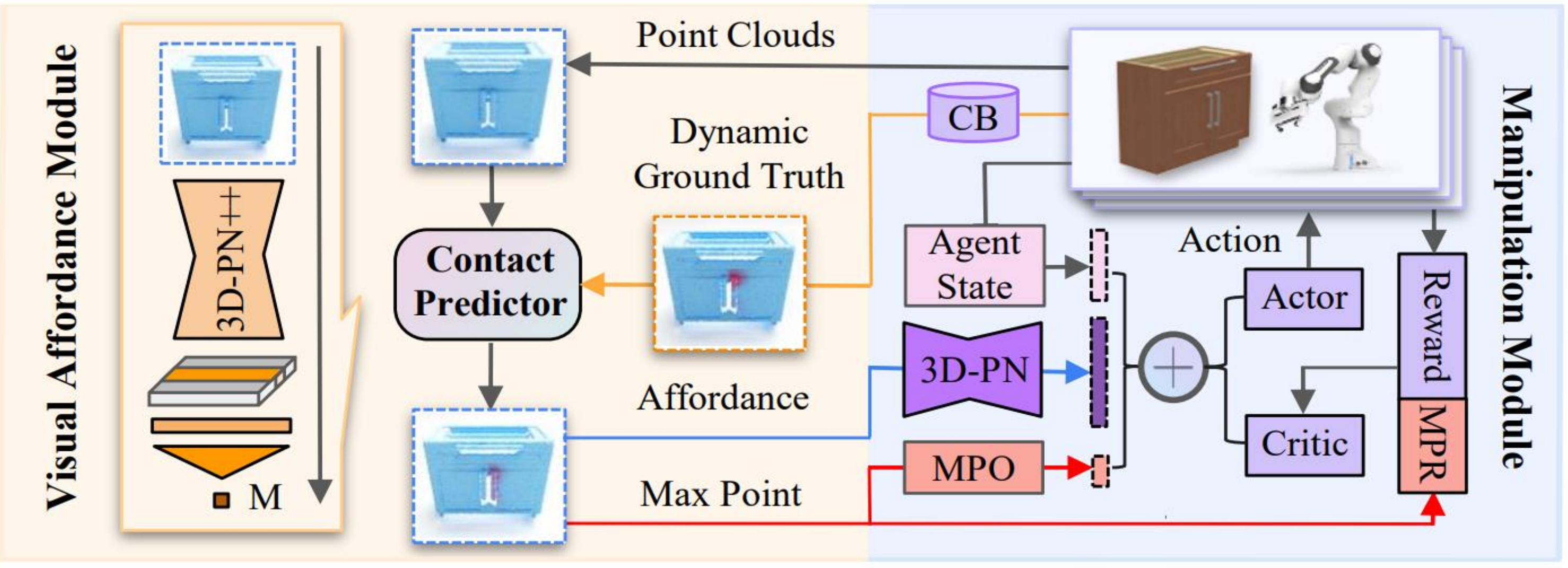
Hyperplane Lab, Center on Frontiers of Computing Studies, Peking University

gyr@stu.pku.edu.cn & hao.dong@pku.edu.cn

Introduction

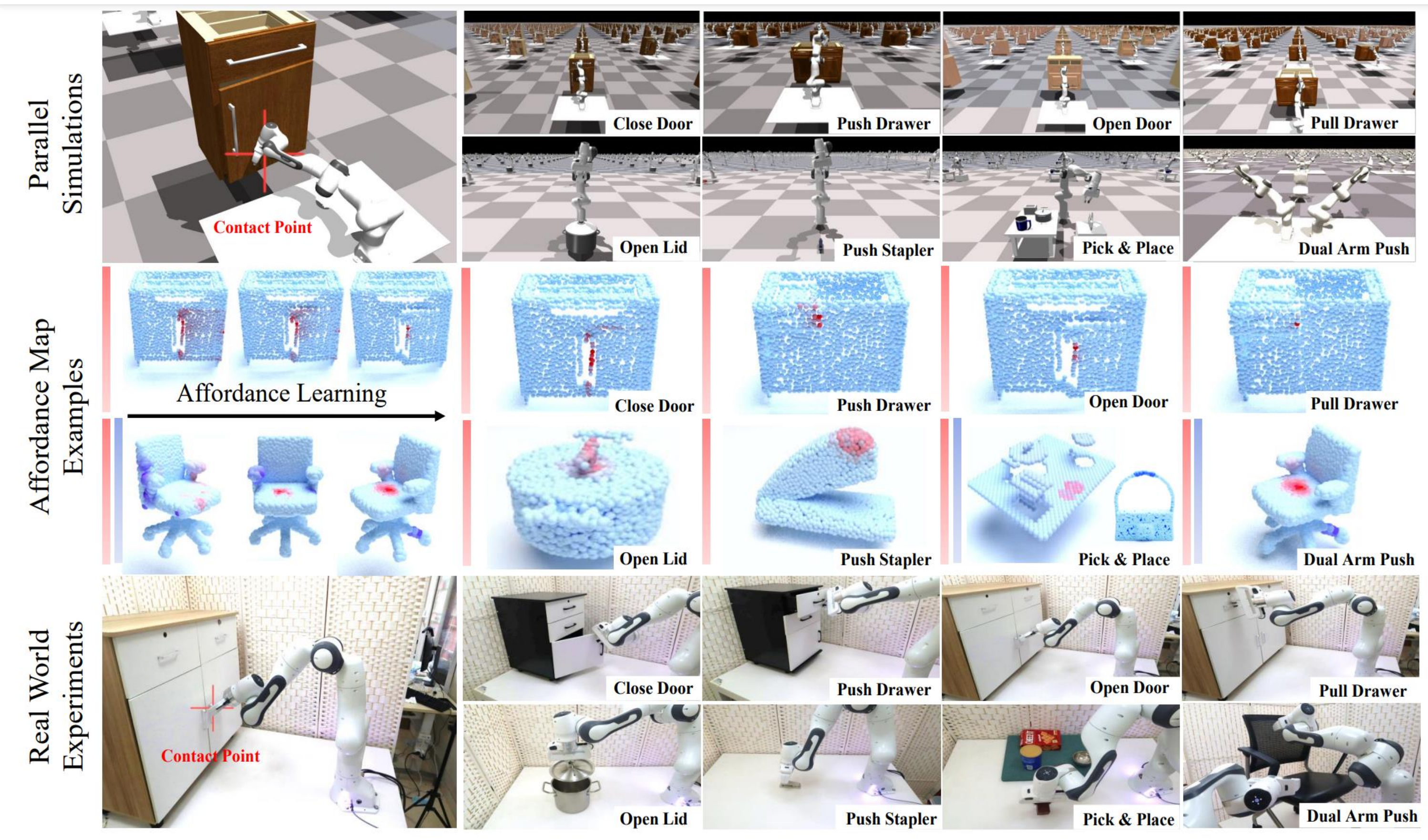
Learning to manipulate 3D objects in an interactive environment has been a challenging problem in Reinforcement Learning (RL). In particular, it is hard to train a policy that can generalize over objects with different semantic categories, diverse shape geometry and versatile functionality. Recently, the technique of visual affordance has shown great prospects in providing object-centric information priors with effective actionable semantics. As such, an effective policy can be trained to open a door by knowing how to exert force on the handle. However, to learn the affordance, it often requires human-defined action primitives, which limits the range of applicable tasks. In this study, we take advantage of visual affordance by using the contact information generated during the RL training process to predict contact maps of interest. Such contact prediction process then leads to an end-to-end affordance learning framework that can generalize over different types of manipulation tasks. Surprisingly, the effectiveness of such framework holds even under the multistage and the multi-agent scenarios. We tested our method on eight types of manipulation tasks. Results showed that our methods outperform baseline algorithms, including visual-based affordance methods and RL methods, by a large margin on the success rate.

Materials and methods



Our pipeline contains two main modules: Manipulation Module (MA Module) generating interaction trajectories and Visual Affordance Module (VA Module) learning to generate per-point affordance map M based on the real-time point cloud. The Contact Predictor (CP), shared across two modules, serves as a bridge between them: 1) MA Module uses the affordance map (indicated by the blue arrow) and Max-affordance Point Observation (MPO) (indicated by the upper red arrow) predicted by the CP as a part of the input observation. A Max-affordance Point Reward (MPR) feedback (indicated by the lower red arrow) is also incorporated in training MA Module; 2) MA Module maintains a Contact Buffer (CB) by collecting collision information and generating Dynamic Ground Truth (DGT) (indicated by the orange arrow), where VA Module uses the DGT as the target for training CP.

Results



QUANTITATIVE RESULTS OF SINGLE-STAGE TASKS. (MORE RESULTS ON OUR WEBSITE.)

Datasets / Methods	Open Door				Pull Drawer				Push Stapler				Open Pot Lid			
	ASR		MP		ASR		MP		ASR		MP		ASR		MP	
	train	test	train	test	train	test	train	test	train	test	train	test	train	test	train	test
Where2act	22.8	14.1	6.8	8.3	19.0	12.9	2.3	0.0	16.4	14.4	13.0	13.0	10.5	5.4	8.7	4.3
VAT-Mart	23.2	21.9	31.8	33.3	5.5	5.1	0.0	0.0	21.9	20.9	17.4	13.0	27.4	21.5	17.4	17.4
Multi-task RL	18.8	9.2	11.4	5.0	0.1	2.4	0.0	2.8	34.9	30.2	30.4	26.1	35.2	32.6	21.7	17.4
RL	21.5	5.5	22.7	0.0	23.1	22.4	19.6	19.5	45.5	40.6	34.8	30.4	32.5	28.6	21.7	21.7
RL+Where2act	20.5	8.0	19.3	9.4	25.2	22.2	24.4	21.9	48.9	45.2	39.1	34.8	38.2	30.6	26.1	21.7
Ours	**52.9**	**32.6**	**61.4**	**41.7**	59.7	**58.6**	62.8	**63.3**	**69.5**	**53.2**	**47.8**	**39.1**	**49.5**	**44.6**	**34.8**	**30.4**
Ours w/o MPO	48.0	23.8	50.0	16.7	41.9	42.5	38.6	43.8	60.6	52.5	43.5	**39.1**	44.2	40.7	**34.8**	**30.4**
Ours w/o MPR	28.2	8.4	29.5	8.3	**62.3**	44.0	**65.9**	43.8	50.8	39.9	39.1	30.4	44.8	40.1	30.4	26.1
Ours w/o E2E	21.2	12.4	20.5	8.3	57.7	57.3	61.1	61.7	40.2	36.6	39.1	34.8	32.1	30.6	30.4	26.1

QUANTITATIVE RESULTS OF PICK-AND-PLACE.

Metrics / Methods	ASR		MP	
	train	test	train	test
RL	25.2	22.1	19.2	11.5
RL+O2OAfford	26.1	22.2	19.2	11.5
RL+Where2act	28.6	23.5	23.1	15.4
RL+O2OAfford+Where2act	30.5	26.2	23.1	15.4
Ours	**46.5**	**39.2**	**30.7**	**26.9**
Ours w/o A2O Map	26.7	22.3	23.1	19.2
Ours w/o O2O Map	31.9	26.2	23.1	15.4
Ours w/o MPO	40.1	30.2	19.2	15.4
Ours w/o MPR	36.2	33.5	**30.7**	23.1
Ours w/o E2E	30.2	21.4	26.9	19.2

QUANTITATIVE RESULTS OF DUAL-ARM-PUSH.

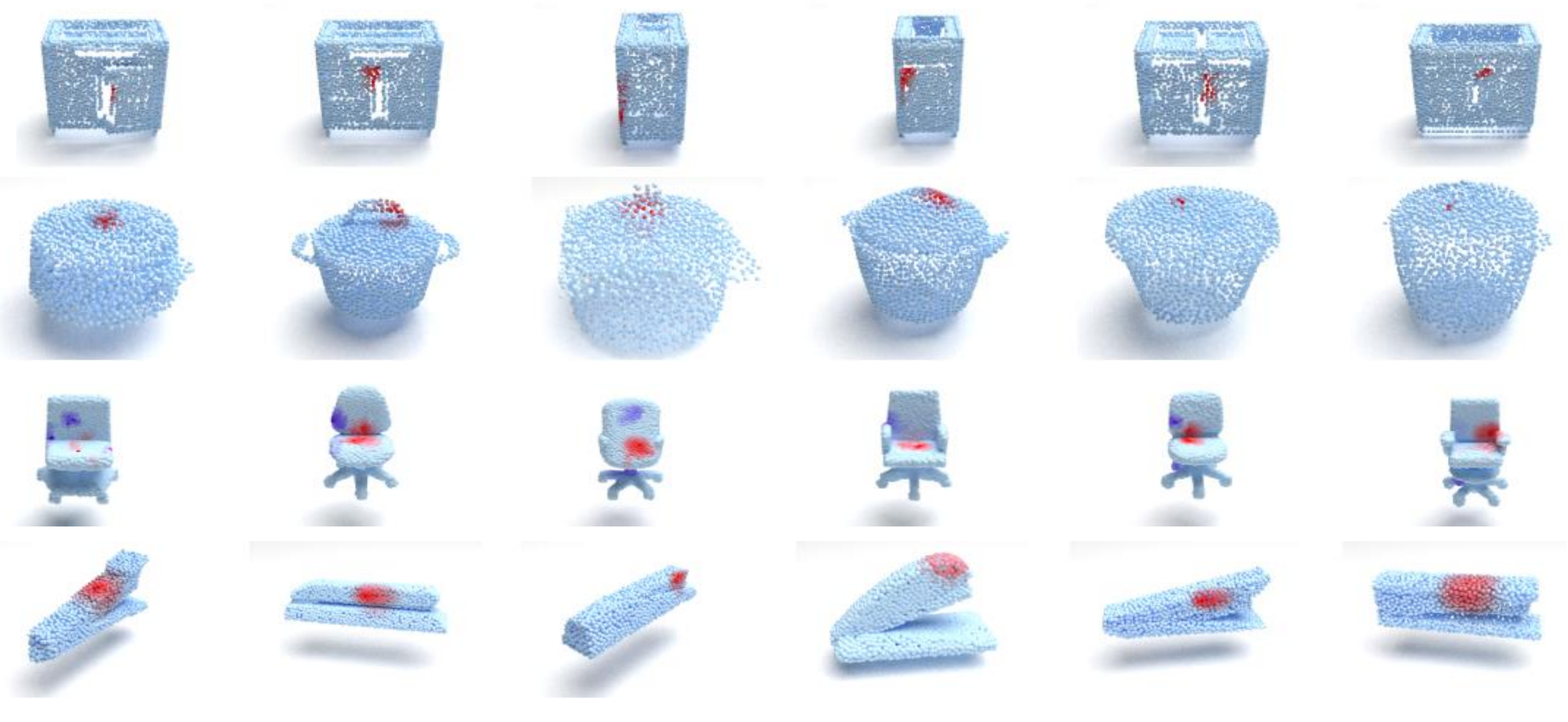
Metrics / Methods	ASR		MP	
	train	test	train	test
MAPPO	7.8	9.0	0.0	0.0
RL	37.2	36.1	36.4	31.3
Multi-task RL	51.6	52.9	54.5	56.3
Ours	83.9	78.5	90.9	93.8
Ours w/o MPO	**95.9**	**96.3**	**100.0**	**100.0**
Ours w/o MPR	63.9	55.3	63.6	56.3
Ours w/o E2E	53.5	55.9	56.8	50.0

- Average Success Rate (ASR): The ASR is the average of the algorithm's success rate on all objects in the training / testing dataset.
- Master Percentage (MP): The master percentage is the percentage of objects which the algorithm can success with a probability greater than or equal to 50%.

Discussion

From tables in Results section, the results of Where2act and RL show the visual affordance can improve the RL performance. However, our method achieves a more significant improvement over baselines in both training and testing sets. In dual-arm-push, our method outperforms both RL and MARL methods. From all tables, we see the MPO, MPR and E2E components play important roles in our method except that E2E on dual-arm-push. The potential reason is that the predicted max affordance point on the object is changing during object movement, which may influence the RL training. This may be something worth looking into in the future.

Figure in Results section shows the change in affordance maps during endto-end training and examples of final affordance maps. We can see that as the training proceeds, the affordance map gradually concentrates.



Conclusion

To the best of our knowledge, this the first work that proposes an end-to-end affordance RL framework for robotic manipulation tasks. In RL training, affordance can improve the policy learning by providing additional observation and reward signals. Our framework automatically learns affordance semantics through RL training without human demonstration or other artificial designs dedicated to data collection. The simplicity of our method, together with the superior performance over strong baselines and the wide range of applicable tasks, has demonstrated the effectiveness of learning from contact information. We believe our work could potentially open a new way for future RL-based manipulation developments

Reference

[1] 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 (ICCV), October 2021, pp. 6813–6823.

[2] Y. Wang, R. Wu, K. Mo, J. Ke, Q. Fan, L. Guibas, and H. Dong, "Adaafford: Learning to adapt manipulation affordance for 3d articulated objects via few-shot interactions," arXiv preprint arXiv:2112.00246, 2021.

[3] R. Wu, Y. Zhao, K. Mo, Z. Guo, Y. Wang, T. Wu, Q. Fan, X. Chen, L. J. Guibas, and H. Dong, "Vat-mart: Learning visual action trajectory proposals for manipulating 3d articulated objects," in ICLR. OpenReview.net, 2022.

[4] Y. Zhao, R. Wu, Z. Chen, Y. Zhang, Q. Fan, K. Mo, and H. Dong, "Dualafford: Learning collaborative visual affordance for dual-gripper object manipulation," arXiv preprint arXiv:2207.01971, 2022.

[5] K. Mo, Y. Qin, F. Xiang, H. Su, and L. Guibas, "O2O-Afford: Annotation-free large-scale object-object affordance learning," in Conference on Robot Learning (CoRL), 2021.