PartManip: Learning Cross-Category Generalizable Part Manipulation Policy from Point Cloud Observations

CVPR2023

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Student: **Haoran Geng** ghr@stu.pku.edu.cn

Advisor: **He Wang** hewang@pku.edu.cn

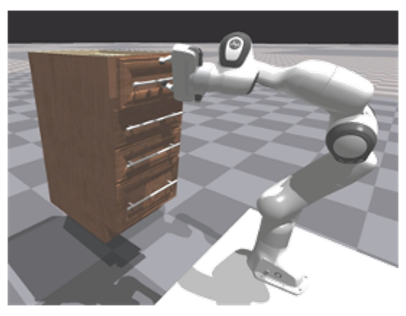
Motivation

Key insight: We humans can successfully manipulate certain types of parts across different objects in a similar way.

In this work, we tackle generalizable part-based manipulation policy learning.

Pressing Buttons on Microwave

Opening Drawer on Table

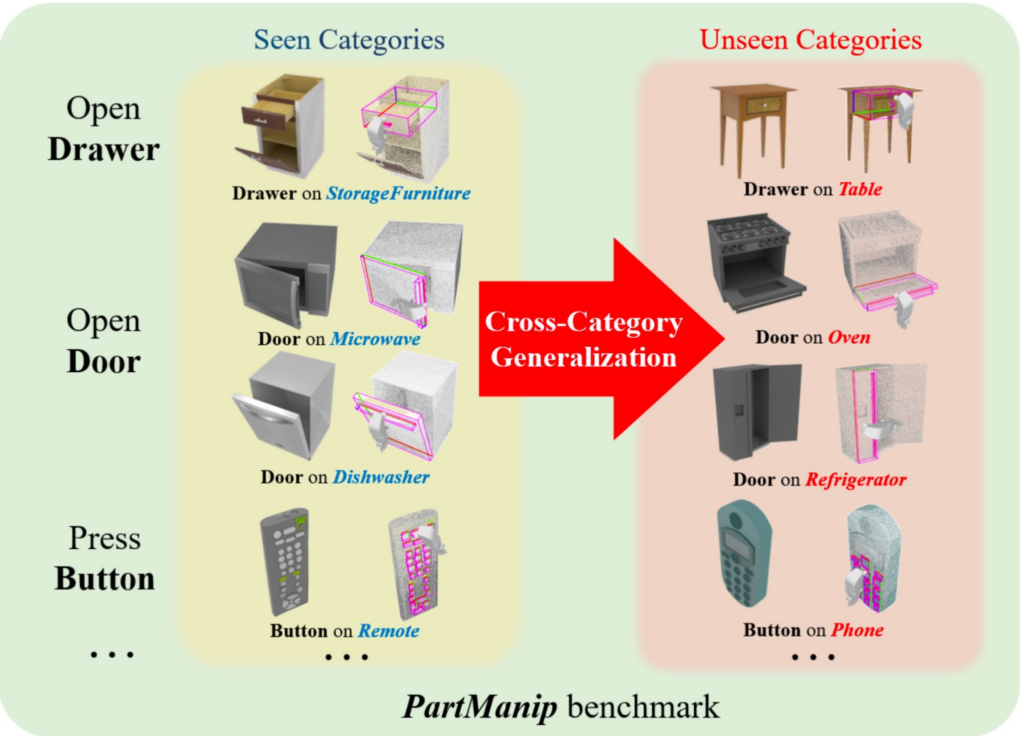


Open Drawer in Simulator

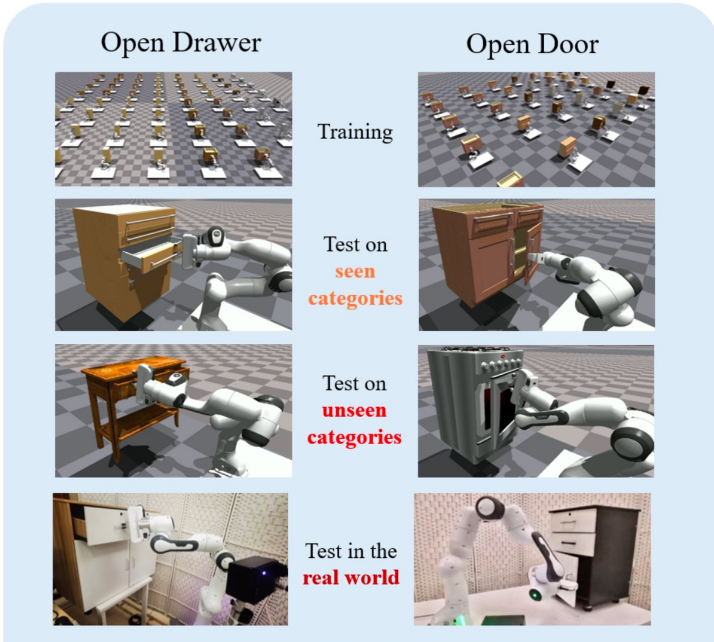
Open Drawer in Real World

Goal: Learning cross-category Manipulation skills via **G**eneralizable and **A**ctionable **Parts** (**GAParts**).

Tasks: Cross-category Open/Close Door, Open/Close Drawer, Press Button, Grasp handle



PartManip benchmark



Cross-Category Part Manipulation Policy Learning

Contribution

Benchmark: The first large-scale, part-based, cross-category object manipulation benchmark, PartManip, is composed of 11 object categories, 494 objects, and 1432 tasks in 6 task classes.

Manipulation: first train a state-based expert with our proposed part-based canonicalization and part-aware rewards, and then distill the knowledge to a vision-based student

Benchmark

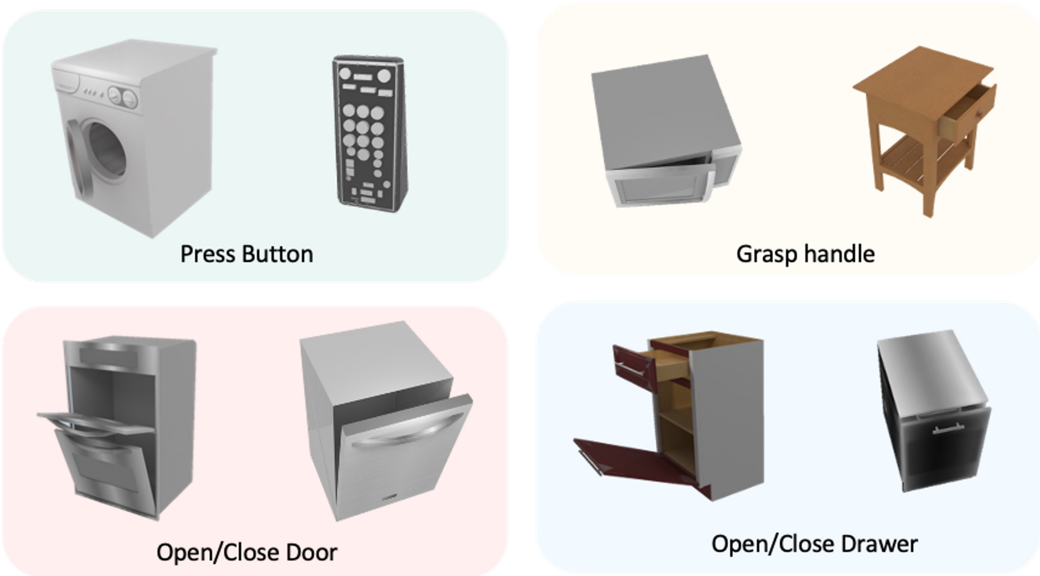
Benchmark Statistics

494 objects

11 object categories

Six tasks

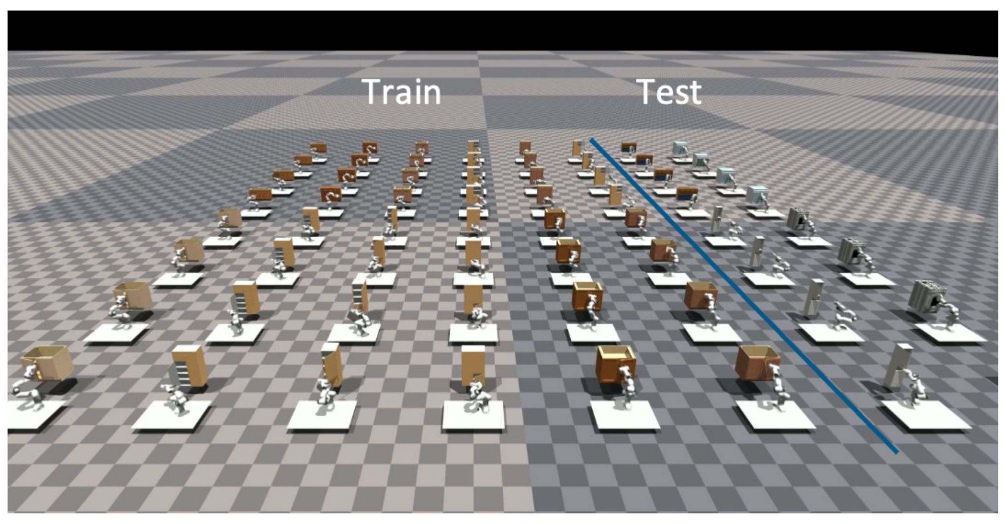
1432 different parts



Cross-category Generalization

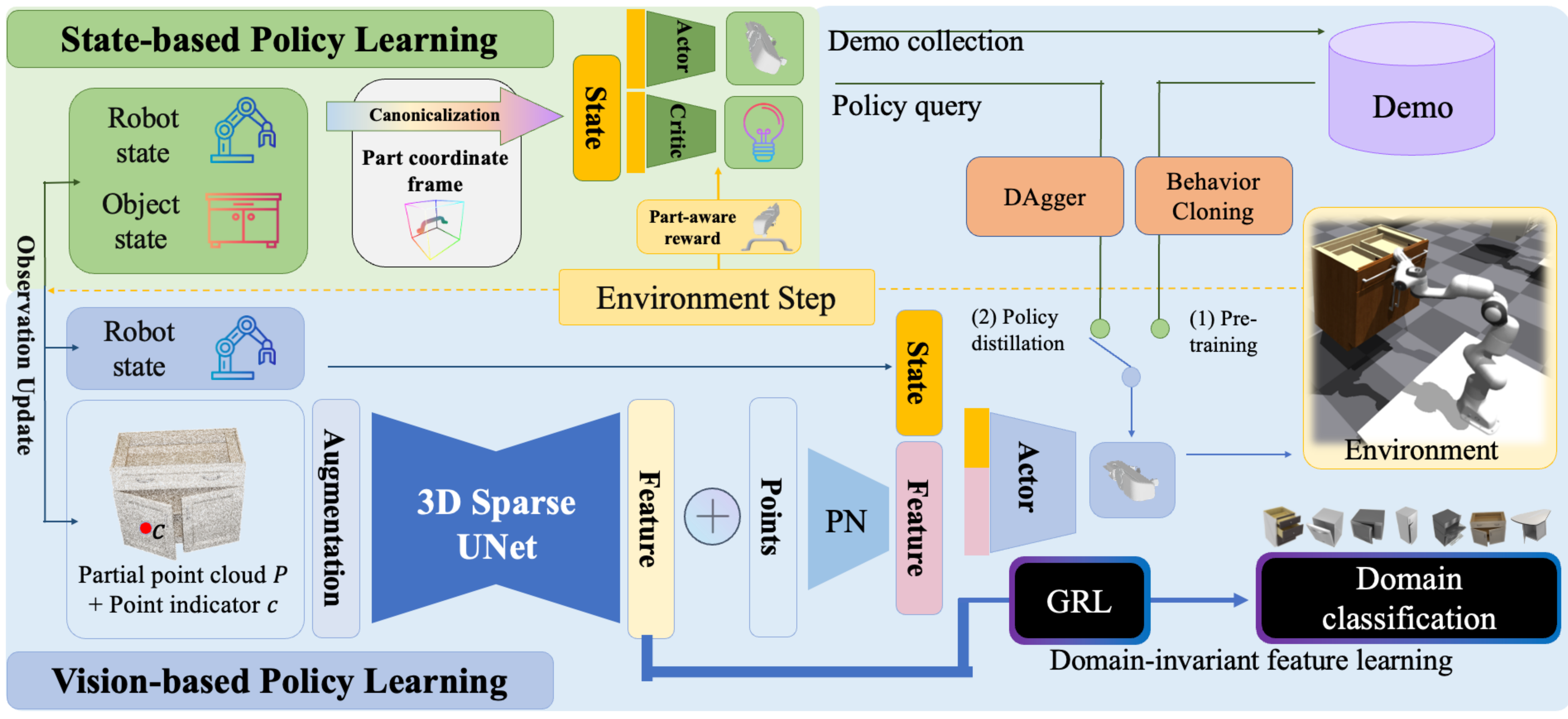
Training: Seen Categories

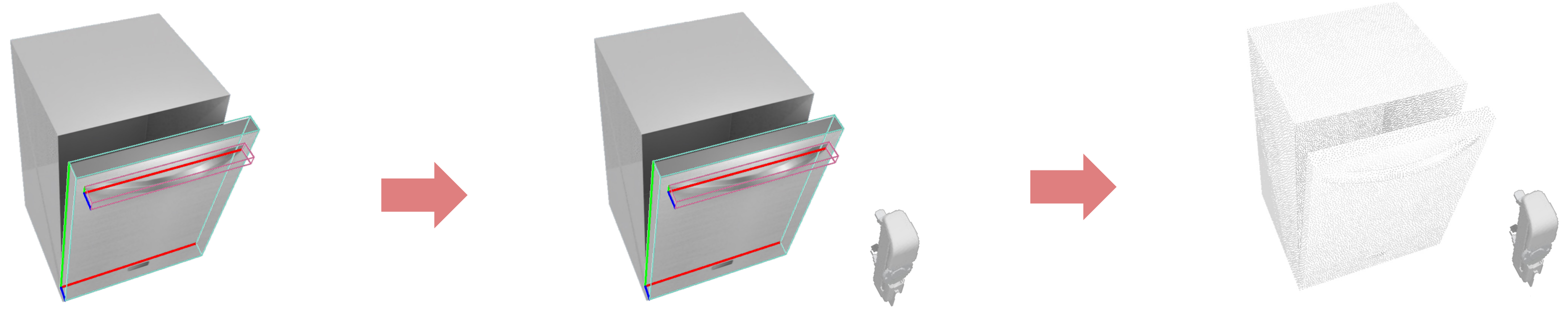
Testing: Unseen Categories



Object Categories in PartManip

Methods





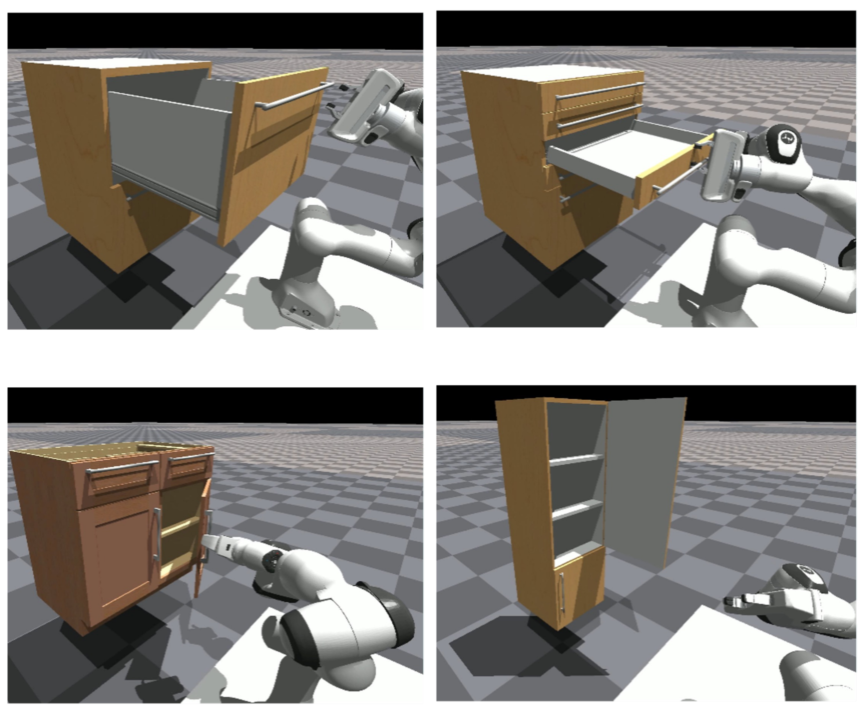
Results

Success rate (%)	OpenDoor			OpenDrawer		
	Training	Val-Intra	Val-Inter	Training	Val-Intra	Val-Inter
PPO [35]	4.5±3.8	4.9±3.5	0.2±0.2	8.9±2.8	11.3±2.8	3.3±1.6
ILAD [48]	13.3±4.9	6.3±2.5	5.0±4.1	18.7±3.6	18.3±2.9	3.3±2.9
Where2act [24]	25.4±0.1	23.4±0.0	15.2±0.1	39.6±0.2	37.2±0.2	20.5±0.1
SilverBullet3D [26]	54.6±2.5	49.9±1.0	26.9±2.2	77.7±3.3	60.0±2.0	31.2±5.1
Shen et. al [38]	1.5±0.6	0.3±0.6	2.3±4.0	9.7±0.5	18.0±2.2	2.7±1.9
Wu et. al [46]	45.9±2.3	34.1±3.8	17.8±1.4	70.5±2.2	53.3±3.3	28.5±2.4
Dubois et. al [5]	35.4±4.4	25.1±2.1	1.3±1.0	61.4±4.3	38.3±2.0	2.7±1.6
Ours	**68.4±1.1**	**57.2±0.4**	**49.1±1.5**	**82.3±2.1**	**78.7±2.0**	**54.7±4.2**

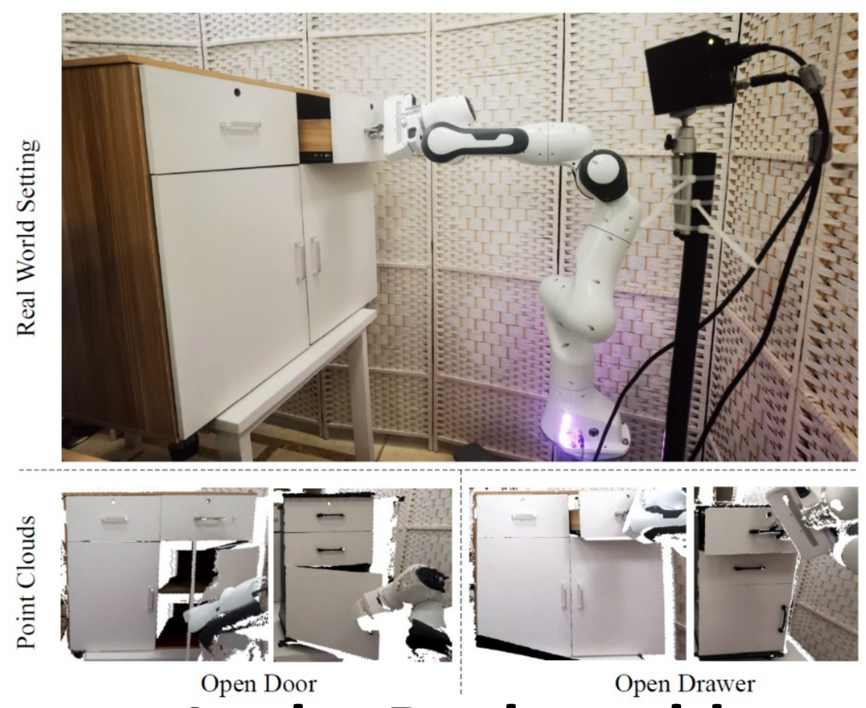
Comparison with Baselines.

Success rate (%)	Canon	DAgger	Augm	S-Unet	Pretrain	DomAdv	Opening Door			Opening Drawer		
							Training	Val-Intra	Val-Inter	Training	Val-Intra	Val-Inter
State-based Expert							67.8±3.4	50.2±1.9	23.4±3.9	71.5±2.1	62.5±2.3	37.5±5.2
	✓						**82.2±0.2**	**62.5±2.6**	**50.7±4.1**	**92.7±0.9**	**88.1±1.0**	**63.4±2.4**
Vision-based Student							4.5±3.8	4.9±3.5	0.2±0.2	8.9±2.8	11.3±2.8	3.3±1.6
				✓			0.8±0.5	0.4±0.2	0.0±0.0	5.9±2.3	3.9±0.6	1.0±0.2
		✓					60.3±0.7	49.2±1.1	31.5±2.9	70.9±0.6	62.0±1.1	42.7±1.8
		✓		✓			66.8±2.7	50.2±1.7	28.8±2.1	77.4±2.7	61.9±3.0	36.4±3.3
		✓	✓				60.0±1.7	54.4±2.3	40.2±3.9	69.7±2.4	69.8±2.5	49.0±2.1
		✓	✓	✓			65.5±1.5	55.9±2.7	41.7±2.5	74.6±3.4	63.8±4.7	49.1±3.4
		✓	✓		✓		61.1±3.3	55.0±1.2	37.8±2.9	71.9±3.3	72.2±3.5	50.3±2.6
		✓	✓	✓	✓		**71.2±1.8**	57.0±0.7	37.2±1.2	82.0±3.3	73.8±2.9	48.8±4.5
		✓	✓	✓	✓	✓	68.4±1.1	**57.2±0.4**	**49.1±1.5**	**82.3±2.1**	**78.7±2.0**	**54.7±4.2**

Ablation Study.



In Simulator



In the Real-world

Conclusion

We introduce a **large-scale part-based cross-category** object manipulation benchmark PartManip, with six tasks in realistic settings. To tackle the challenging problem of the generalizable vision-based policy learning, we first introduce a carefully designed **state-based part-aware expert learning method**, and then a well-motivated **state-to-vision distillation process**, as well as a **domain generalization technique** to improve the cross-category generalization ability.

Scan the QR code for more information and to contact us!