GAPartNet: Cross-Category Domain-Generalizable Object Perception and Manipulation via Generalizable and Actionable Parts

CVPR2023 Highlight

Student: **Haoran Geng** *ghr@stu.pku.edu.cn* Advisor: **He Wang** *hewang@pku.edu.cn*

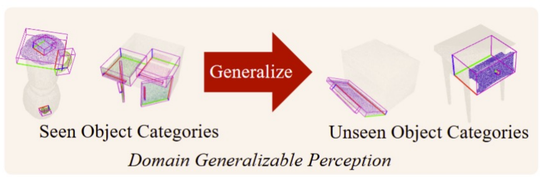
Motivation

Key insight: Parts are fundamental building blocks of our daily objects. We humans can identify a set of commonly used **parts**, which can **generalize to unseen object categories.** Some **part classes** are more **elementary** and **fundamental** than object categories and thus worthy of more research efforts.

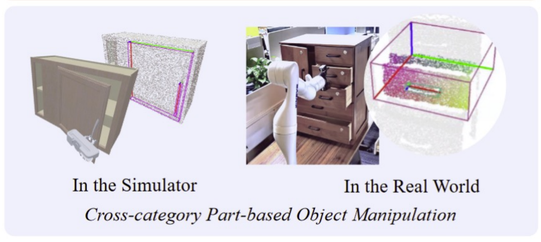
Goal: Learning cross-category skills via **G**eneralizable and **A**ctionable **Parts** (**GAParts**).

Tasks: Part Perception (Segmentation & Pose Estimation), Part-based Object Manipulation

Part Classes in the GAPartNet Dataset



Domain Generalizable Perception



Cross-category Part-based Object Manipulation

Contribution

Dataset: A novel concept **GAPart,** a large-scale interactive dataset, **GAPartNet**, with rich **part semantics** and **pose annotations**.

Perception: A first-ever pipeline for domain-generalizable 3D part segmentation and pose estimation

Manipulation: A new solution to generalizable object manipulation by leveraging the concept of **GAPart**

Dataset

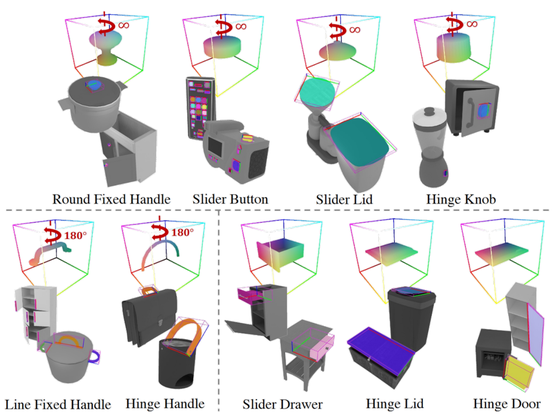
GAPart Concept

Rigorous Definition:

- Geometric similarity, actionability alignment

Semantics:

- 9 common GAPart classes

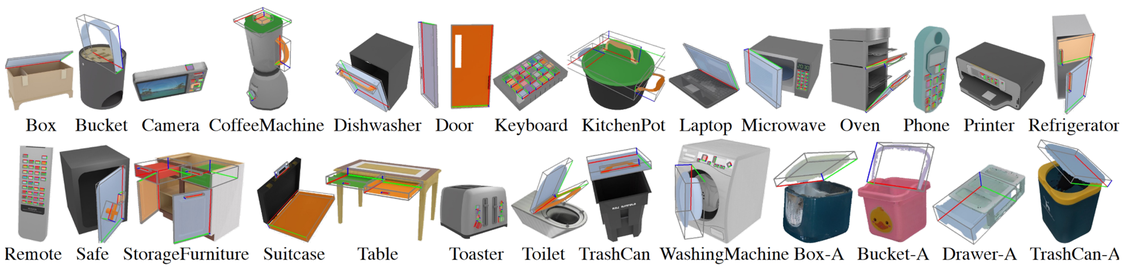


Poses:

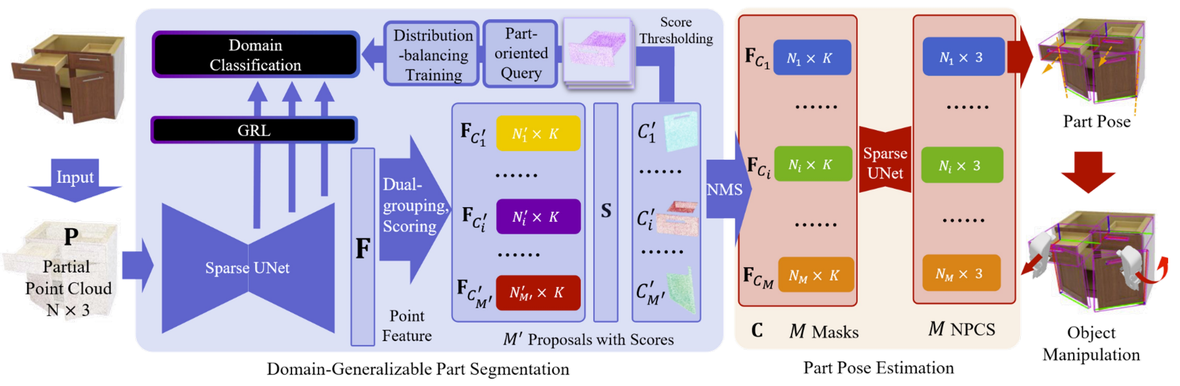
- Canonicalized part position and orientation in NPCS (Normalized Part Coordinate Space)

GAPartNet Dataset

- A large-scale **part-centric interactive** dataset
- **9 GAPart classes, 27 object categories, 8,489 part instances, 1,166 objects**
- Rich, part-level annotations (**semantics, poses**)



Methods



Part Segmentation and Pose Estimation

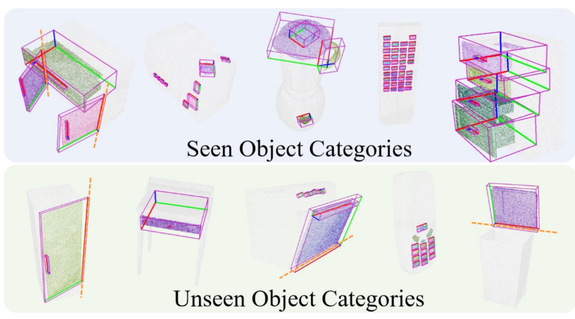
- A **part-oriented domain adversarial training strategy.**
- **Part-oriented feature query**, **multi-resolution,** and leveraging **focal loss**

Part-based Object Manipulation

- **Actionability** in GAPart pose definition → a simple yet efficient **heuristic algorithm**

Results

		Ln.F.Hl.	Rd.F.Hl.	Hg.Hl.	Hg.Ld.	Sd.Ld.	Sd.Bn	Sd.Dw.	Hg.Dr.	Hg.Kb.	Avg.AP	Avg.AP50
Seen (%)	PG [17]	86.1	23.0	84.6	80.01	88.3	49.3	62.6	92.8	34.6	57.3	66.8
	SG [48]	57.8	**93.6**	81.2	76.0	89.3	25.2	50.8	93.9	51.5	58.5	68.8
	AGP [31]	86.8	20.3	87.7	79.7	89.4	62.3	61.6	92.5	16.7	57.2	66.3
	Ours	**89.2**	54.9	**90.4**	**84.8**	**89.8**	**66.7**	**67.2**	**94.7**	**52.9**	**67.6**	**76.5**
Unseen (%)	PG [17]	32.44	9.8	2.1	26.8	0.0	42.6	57.0	63.9	1.7	21.9	26.3
	SG [48]	25.8	5.0	0.4	33.9	0.6	**51.5**	51.2	69.0	12.1	22.0	27.7
	AGP [48]	**45.6**	4.8	**3.1**	34.3	0.0	47.8	64.1	63.1	11.5	25.7	30.5
	Ours	**45.6**	**40.0**	**3.1**	**40.2**	**5.0**	49.1	**64.2**	**69.1**	**23.4**	**32.0**	**37.2**

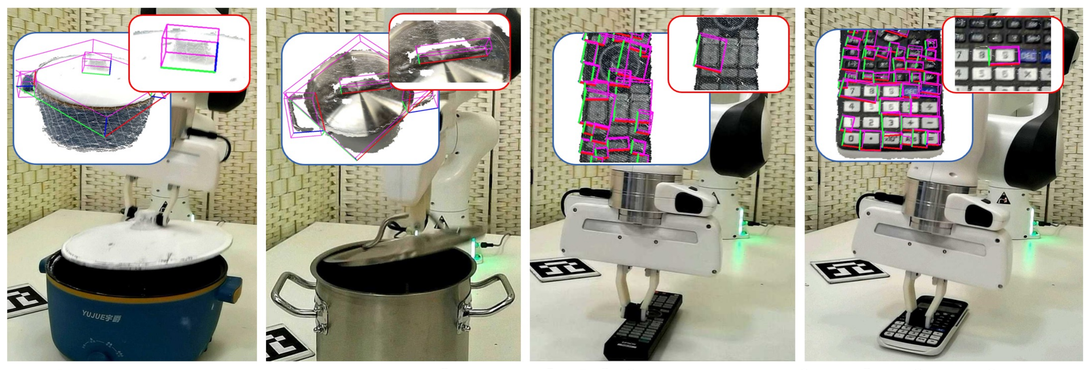


		$R_e\downarrow$	$T_e\downarrow$	$S_e\downarrow$	$\theta_e\downarrow$	$d_e\downarrow$	mIoU ↑	$A_5\uparrow$	$A_{10}\uparrow$
Seen	PG [17]	14.3	0.034	0.039	7.947	0.020	49.4	41.2	66.4
	AGP [31]	14.4	0.036	0.039	7.955	0.021	48.7	40.9	64.8
	Ours	**8.8**	**0.028**	**0.035**	**7.4**	**0.014**	**52.2**	**45.6**	**71.5**
Unseen	PG [17]	18.2	0.056	0.073	12.0	0.031	36.2	28.0	50.9
	AGP [31]	18.2	0.57	0.076	11.9	0.029	36.3	28.6	51.2
	Ours	**14.8**	**0.051**	**0.067**	**11.3**	**0.024**	**43.1**	**32.0**	**55.7**

Cross-category Part Segmentation and Pose Estimation

Success Rate(%)	Drawer		Door		Handle		Button	
	Seen	Unseen	Seen	Unseen	Seen	Unseen	Seen	Unseen
Where2act [6]	69.9	54.5	44.4	18.2	78.7	49.2	82.2	80.9
ManiSkill [7]	32.9	26.6	27.8	28.3	53.9	42.1	65.5	54.5
Ours	**95.0**	**90.0**	**70.0**	**55.0**	**90.0**	**85.0**	**100.0**	**95.0**

Cross-category Part-based Object Manipulation in the Simulator



Cross-category Part-based Object Manipulation in the Real World

Conclusion

We introduce the concept of **GAPart** and present the **GAPartNet dataset** and explore three cross-category tasks based on GAParts: part segmentation, part pose estimation, and part-based object manipulation. Our proposed approach, adopting a **domain generalization** perspective, outperforms previous works in **segmentation and pose estimation**. Furthermore, we design part-pose-based **interaction heuristics** that enable effective and generalizable object manipulation in both the simulator and the real world.

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