ZeroPS: High-quality Cross-modal Knowledge Transfer for Zero-Shot 3D Part Segmentation – Supplementary Material

S. Supplementary Material

S.1. Merging 3D Groups

The core steps of the merging algorithm are presented in the Algorithm 1. Starting from K_1 different viewpoints, we obtain a set of m_1 3D groups from all self-extensions, denoted as $A = \{G_1^{3D}, \ldots, G_{m_1}^{3D}\}$. Since G^{3D} with the same semantics in set A have similar areas, we sort set A in descending order. All sets in this algorithm are regarded as ordered sets. Then we iterate over A, and for each G^{3D}, it is either merged with an existing M^{3D} in B or added to B as a new M^{3D} (steps 3-12). Note that we use the Intersection over Union (IoU) as the criterion to determine whether G^{3D} is merged or not, while controlled by the merge threshold T. Second, we need to ensure that each point of Q^{3D} is associated with unique semantics. If a point simultaneously exists in different M^{3D}, we choose to assign it to the M^{3D} with higher granularity. Since M^{3D} in B is granularity from lower to higher, we iterate over B and add M^{3D} to C each time. After adding M^{3D} each time, it is needed to remove the points that each P^{3D} in C (except the currently added M^{3D}) shares with the current M^{3D} (step 14-16). Finally, we return set $C = \{P_1^{3D}, \ldots, P_{m_2}^{3D}\}$ which includes m_2 3D unlabeled parts P^{3D}.

S.2. Implementation Details of Viewpoints

In the experiment, we follow PartSLIP [4] to fix the viewpoint position and set the camera distance to 2.2 in Pytorch3D. We place 20 viewpoints around the 3D object, with 8 of these viewpoints each serving as a starting viewpoint. The viewpoint positions are detailed in Tab. S1 and Fig. S1.

S.3. AKBSeg Benchmark

Tab. S2 shows the statistics of the proposed AKBSeg benchmark. All 508 3D objects collect from the AKB-48 [3] dataset. To be consistent with PartNetE's requirements for the part instance segmentation, we provide additional instance labels for the original AKB-48 data. While the PartNetE and AKBSeg benchmarks have overlapping object categories, the 3D parts exhibit significant differences. For example, in the object category 'Trashcan', PartNetE includes 'footpedal', 'lid', and 'door', whereas AKBSeg includes 'lid' and 'wheel'. Such differences and diversity are advantageous for evaluating zero-shot baselines, as they are entirely grounded in real-world scenarios. We hope that the proposed AKBSeg benchmark could help the future evaluation of zero-shot 3D part segmentation.

Algorithm 1 Merge 3D Groups. T is the merge threshold.

Input: 3D groups $A = \{G_1^{3D}, \ldots, G_{m_1}^{3D}\}$
Output: 3D parts $C = \{P_1^{3D}, \ldots, P_{m_2}^{3D}\}$

1: sort the elements in A by area (the number of points) in descending order
2: initialize an empty set B
3: **for** each G^{3D} **in** A **do**
4: $\quad flag = False$
5: $\quad$ **for** each M^{3D} **in** B **do**
6: $\quad\quad$ calculate iou of G^{3D} and M^{3D}
7: $\quad\quad$ **if** $iou > T$ **then**
8: $\quad\quad\quad M^{3D} \leftarrow M^{3D} \cup G^{3D}$ ▷ update M^{3D} in B
9: $\quad\quad\quad flag = True$
10: $\quad\quad\quad$ **break**
11: $\quad$ **if not** $flag$ **then**
12: $\quad\quad$ add G^{3D} to B
13: initialize an empty set C
14: **for** each M^{3D} **in** B **do**
15: $\quad$ add M^{3D} to C
16: $\quad C[0 : \text{len}(C) - 1] \leftarrow C[0 : \text{len}(C) - 1] \setminus M^{3D}$
17: **return** C

Table S1. List of all viewpoints with elevation and azimuth angles.

id	elevation (°)	azimuth (°)
1*	35	-35
2	35	10
3*	35	55
4	35	100
5*	35	145
6	35	190
7*	35	235
8	35	280
9	-10	-35
10	-10	55
11	-10	145
12	-10	235
13*	-55	-35
14	-55	10
15*	-55	55
16	-55	100
17*	-55	145
18	-55	190
19*	-55	235
20	-55	280

* starting viewpoint

S.4. More Quantitative Comparison

Zero-shot Semantic Segmentation. We degrade the instance segmentation result into semantic segmentation and

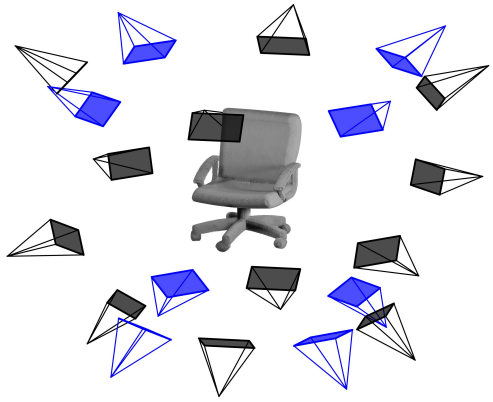

Figure S1. Visualization of viewpoint positions. In the experiment, we place 20 viewpoints around the 3D object, with 8 of these viewpoints (highlighted in blue) each as a starting viewpoint.

Table S2. The table shows the statistics of the AKBSeg benchmark.

category	parts	test
Ballpoint	cap,button	9
Bottle	lid	35
Box	lid	40
Bucket	handle	37
Condiment	lid,handle	10
Cup	lid,handle	34
Drink	lid	51
Eyeglasses	body,leg	93
Faucet	spout,switch	45
Foldingrack	hook,body,leg	11
Knife	blade,handle	9
Lighter	lid,wheel,button	19
Sauce	lid	43
Scissor	blade,handle	19
Shampoo	head,lid	31
Trashcan	lid,wheel	22
16 in total	29 in total	508

Table S3. Zero-shot semantic segmentation results, measured as mIoU(%).

	PointCLIP V2 [8]	PartSLIP [4]	Ours
PartNetE	16.1	34.4	**39.3**
AKBSeg	17.8	25.9	**35.7**

compare it with existing methods. We follow [4] to utilize the category mIoU as the metric. For PointCLIP V2's [8] text prompt, we follow it to prompt GPT-3 [1] to generate 3D specific text for each part category of the input object by constructing a 3D language command. Tab. S3 shows that our method on the semantic segmentation task consistently exhibits performance advantages over other methods, similar to the gap observed on the unlabeled and instance segmentation tasks.

Comparison with PartSLIP++. We conduct the quantitative comparison with PartSLIP++ [7], *a concurrent work* that introduces two main improvements to PartSLIP: 1) using SAM to refine the GLIP's bounding boxes, to achieve more accurate 2D predictions; 2) proposing an improved Expectation-Maximization (EM) algorithm, to lift 2D segmentation to 3D. Since the improved EM algorithm is based on the few-shot setting, we replaced it with the PartSLIP's voting strategy to ensure the zero-shot setting. As shown in Tabs. S4 and S5, our method demonstrates better zero-shot performance compared to PartSLIP++ across the different tasks and benchmarks.

Table S4. Quantitative comparison with PartSLIP++ on PartNetE.

	unlabeled seg.	instance seg.	semantic seg.
PartSLIP++ [7]	38.9	25.5	37.4
Ours	**56.0**	**28.5**	**39.3**

Table S5. Quantitative comparison with PartSLIP++ on AKBSeg.

	unlabeled seg.	instance seg.	semantic seg.
PartSLIP++ [7]	35.7	17.5	26.8
Ours	**58.9**	**26.5**	**35.7**

Table S6. Ablation Study on TDCM, measured as mAP50(%).

	2D	3D	2D & 3D
PartNetE	22.8	19.5	**24.1**
AKBSeg	22.6	20.1	**23.9**

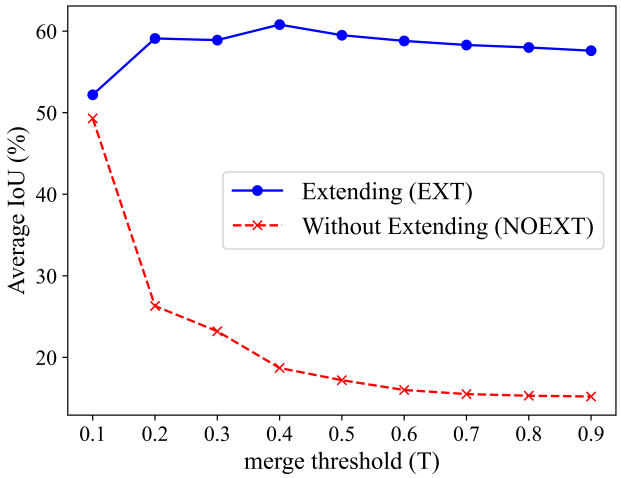

Figure S2. Ablation Study on self-extension by the 'Extending' and 'Without Extending' settings. The Average IoU is the overall result on AKBSeg.

S.5. More Ablation Studies

Self-extension. To further analyze the effectiveness of self-extension, we conduct the ablation study on the AKBSeg benchmark. As shown in Fig. S2, from PartNetE (See Fig.6 in main paper) to AKBSeg, EXT and NOEXT exhibit similar performance trends. Compared to NOEXT, EXT still demonstrates better robustness and stability.

Two-dimensional Checking Mechanism (TDCM). To evaluate the effectiveness of TDCM, we conduct the abla-

Table S7. Sensitivities to Object Rotation. Applying random rotation to the input object results in no significant fluctuation in performance, measured as Average IoU(%).

setting	Kettle	Pen	Stapler
original	**85.5**	71.7	80.7
rotated	85.2	**72.1**	**80.9**

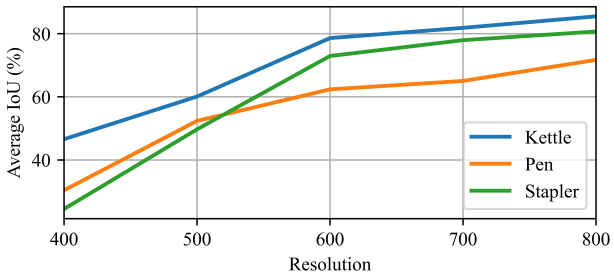

Figure S3. Ablation study on render resolution.

tion study. As shown in Tab. S6, when voting is performed only in either the 2D or 3D space, we observe the performance decrease. This indicates that aligning the voting results from both 2D and 3D spaces, to check and then discard unqualified bounding boxes is effective and reasonable.

Render Resolution. We conduct the ablation study on the rendering resolution. As shown in Fig. S3, increasing the resolution from 400 to 600 results in a significant performance gain. However, the gain is relatively small when increasing the resolution from 600 to 800.

Sensitivities to Object Rotation. We perform sensitivity analysis on the input object by random rotations. As shown in Tab. S7, we observe no significant fluctuation in performance. This indicates that our method is hardly sensitive to object rotation.

S.6. Metric Details

For the unlabeled segmentation, we follow [6] to utilize the Average IoU as its metric. For each 3D object, the Average IoU is determined by calculating the maximum IoU for each ground-truth instance part from the predicted parts and averaging them. Next, we calculate the average metrics for each object category. For the instance segmentation, We follow [4] to utilize mAP (50% IoU threshold) as its metrics. For each object category, the mAP50% is determined by calculating the AP50% of each part instance category and averaging them.

S.7. Text Prompts

The official Supplementary Material and code of PartSLIP demonstrate that the text prompt without the object name gives GLIP better performance (*e.g.*, removing 'of chair' in 'arm, back, leg, seat, wheel of chair'). Thus, our method and PartSLIPs' text prompts contain only the part names (*e.g.*, 'arm, back, leg, seat, wheel') in the experiment.

S.8. Density of Input Points

To obtain finer 3D parts, we follow [4, 6] to focus on the dense rather than sparse point clouds in this work.

S.9. Discussion and Limitation

While the current pipeline demonstrates strong zero-shot generalization and segmentation performance, it has certain limitations. The primary limitation is that the performance of the pretrained foundation models directly impacts our pipeline. However, since our pipeline relies only on the foundation models' prompt mechanism, the proposed three training-free manners (self-extension, TDCM, and CNVP) are independent of the internal structure of the foundation models. Therefore, replacing the pretrained foundation model with another model that uses the same prompt mechanism is a reasonable solution. Another generally applicable approach is to fine-tune the foundation model.

S.10. Full Table of Quantitative Comparison

Tabs. S8 to S11 show the full tables of quantitative comparison results of the main paper.

References

[1] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901, 2020. 2

[2] Li Jiang, Hengshuang Zhao, Shaoshuai Shi, Shu Liu, Chi-Wing Fu, and Jiaya Jia. Pointgroup: Dual-set point grouping for 3d instance segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. 5

[3] Liu Liu, Wenqiang Xu, Haoyuan Fu, Sucheng Qian, Qiaojun Yu, Yang Han, and Cewu Lu. Akb-48: A real-world articulated object knowledge base. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14809–14818, 2022. 1

[4] Minghua Liu, Yinhao Zhu, Hong Cai, Shizhong Han, Zhan Ling, Fatih Porikli, and Hao Su. Partslip: Low-shot part segmentation for 3d point clouds via pretrained image-language models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 21736–21746, 2023. 1, 2, 3, 4, 5

[5] Thang Vu, Kookhoi Kim, Tung M Luu, Thanh Nguyen, and Chang D Yoo. Softgroup for 3d instance segmentation on point clouds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2708–2717, 2022. 5

[6] Xiaogang Wang, Xun Sun, Xinyu Cao, Kai Xu, and Bin Zhou. Learning fine-grained segmentation of 3d shapes without part labels. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10276–10285, 2021. 3

Table S8. Full table of zero-shot unlabeled segmentation results on the PartNetE benchmark, corresponding to Table 1 of the main paper. Object category Average IoUs(%) are shown.

category	PartSLIP [4]	Ours (w/o Extending)	Ours
Bottle	78.0	55.7	**80.4**
Chair	**76.8**	60.1	71.8
Clock	17.8	**38.1**	33.8
Dishwasher	37.3	**63.3**	59.5
Display	**73.5**	56.9	68.7
Door	27.9	23.9	**37.8**
Faucet	16.5	45.0	**75.5**
Keyboard	0.5	**29.6**	27.7
Knife	18.3	44.5	**68.7**
Lamp	47.6	42.8	**72.9**
Laptop	41.2	**47.3**	39.8
Microwave	14.1	**14.6**	13.0
Refrigerator	32.2	**61.3**	57.9
Scissors	47.2	39.9	**51.1**
StorageFurniture	49.8	**55.8**	50.0
Table	46.9	47.9	**53.3**
TrashCan	51.1	58.1	**67.2**
Box	38.8	51.6	**63.1**
Bucket	35.6	57.3	**83.7**
Camera	36.7	**48.1**	40.9
Cart	**78.7**	58.2	78.5
CoffeeMachine	28.9	29.2	**31.1**
Dispenser	40.6	52.4	**74.6**
Eyeglasses	5.5	36.7	**70.4**
FoldingChair	**85.5**	44.9	76.2
Globe	**89.5**	33.4	53.0
Kettle	66.2	57.8	**85.5**
KitchenPot	60.5	70.7	**80.3**
Lighter	53.2	47.3	**64.4**
Mouse	21.7	**38.9**	34.9
Oven	22.8	**45.5**	37.0
Pen	44.6	56.1	**71.7**
Phone	11.8	**56.7**	38.4
Pliers	3.5	40.1	**61.3**
Printer	1.1	**6.2**	4.1
Remote	2.8	**38.0**	34.3
Safe	17.4	**26.5**	26.2
Stapler	27.3	39.7	**80.7**
Suitcase	**65.8**	51.8	62.2
Switch	6.5	68.0	**72.4**
Toaster	19.7	**65.0**	62.7
Toilet	45.4	56.0	**58.2**
USB	35.3	32.4	**85.0**
WashingMachine	14.8	**16.9**	15.6
Window	2.5	39.8	**46.7**
Overall (45)	36.4	45.6	**56.0**

[7] Yuchen Zhou, Jiayuan Gu, Xuanlin Li, Minghua Liu, Yunhao Fang, and Hao Su. Partslip++: Enhancing low-shot 3d part segmentation via multi-view instance segmentation and maximum likelihood estimation. *arXiv preprint arXiv:2312.03015*, 2023. 2

[8] Xiangyang Zhu, Renrui Zhang, Bowei He, Ziyu Guo, Ziyao Zeng, Zipeng Qin, Shanghang Zhang, and Peng Gao. Point-clip v2: Prompting clip and gpt for powerful 3d open-world learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2639–2650, 2023. 2

Table S9. Full table of zero-shot unlabeled segmentation results on the AKBSeg benchmark, corresponding to Table 3 of the main paper. Object category Average IoUs(%) are shown.

category	PartSLIP [4]	Ours (w/o Extending)	Ours
Ballpoint	3.0	33.7	**48.9**
Bottle	8.7	55.7	**65.7**
Box	35.0	**54.1**	52.5
Bucket	49.9	74.1	**75.5**
Condiment	44.4	50.7	**65.2**
Cup	**42.2**	28.2	39.5
Drink	10.5	49.9	**67.8**
Eyeglasses	1.0	25.0	**38.2**
Faucet	11.0	44.6	**52.8**
Foldingrack	32.4	50.6	**64.1**
Knife	73.0	74.8	**84.1**
Lighter	**39.1**	34.9	33.0
Sauce	22.5	35.7	**45.8**
Scissor	**73.1**	47.8	66.9
Shampoo	37.6	56.5	**63.2**
Trashcan	66.0	72.0	**79.8**
Overall (16)	34.3	49.3	**58.9**

Table S10. Full table of zero-shot instance segmentation results on the AKBSeg benchmark, corresponding to Table 4 of the main paper. Object category mAP50s(%) are shown.

category	parts	PartSLIP [4]	Ours (w/o CNVP)	Ours
Ballpoint	cap	**1.0**	**1.0**	**1.0**
	button	1.0	8.9	**12.1**
Bottle	lid	1.1	**26.0**	20.2
Box	lid	12.9	13.6	**16.3**
Bucket	handle	34.8	59.6	**77.8**
Condiment	lid	7.9	13.9	**21.6**
	handle	14.9	43.1	**60.4**
Cup	lid	1.1	4.7	**5.2**
	handle	**42.3**	16.4	24.4
Drink	lid	1.0	35.0	**36.8**
Eyeglasses	body	1.0	**5.4**	4.0
	leg	1.0	**12.1**	10.3
Faucet	spout	1.0	1.7	**3.3**
	switch	5.8	**6.3**	4.6
Foldingrack	hook	48.0	61.5	**76.6**
	body	1.0	**22.7**	20.0
	leg	1.0	**22.0**	12.0
Knife	blade	65.9	72.5	**73.6**
	handle	36.1	**88.1**	**88.1**
Lighter	lid	**1.0**	**1.0**	**1.0**
	wheel	**1.0**	**1.0**	**1.0**
	button	**6.1**	1.0	1.0
Sauce	lid	5.1	9.2	**10.1**
Scissor	blade	**41.9**	38.9	38.1
	handle	**87.3**	68.5	71.1
Shampoo	head	1.5	7.1	**10.9**
	lid	**4.2**	2.7	2.3
Trashcan	lid	1.8	4.6	**4.6**
	wheel	**16.8**	9.6	15.4
Overall (16)		15.0	23.9	**26.5**

Table S11. Full table of instance segmentation results on the PartNetE benchmark, corresponding to Table 2 of the main paper. Object category mAP50s(%) are shown.

category	part	PointGroup* [2]	SoftGroup* [5]	PartSLIP† [4]	Ours (w/o CNVP)†	Ours†
Bottle	lid	38.2	**43.9**	67.0	62.0	**74.5**
Chair	arm	94.6	**95.1**	44.4	47.0	**67.3**
	back	**82.0**	73.2	**86.4**	67.1	72.0
	leg	88.6	**93.6**	**52.3**	46.4	52.1
	seat	75.0	**85.9**	**87.2**	50.7	69.4
	wheel	**98.0**	97.7	**92.6**	58.4	67.2
Clock	hand	**1.0**	**1.0**	3.1	**10.9**	9.5
Dishwasher	door	**76.7**	75.0	13.4	13.1	**21.1**
	handle	55.6	**56.4**	16.1	22.9	**32.3**
Display	base	95.2	**97.4**	71.1	51.8	**72.8**
	screen	46.0	**55.4**	25.5	30.9	**32.5**
	support	**54.0**	53.2	38.0	30.4	**45.1**
Door	frame	**36.8**	28.3	1.0	**22.0**	19.2
	door	32.4	**34.3**	15.4	21.1	**26.4**
	handle	**1.0**	**1.0**	**15.3**	1.3	1.4
Faucet	spout	85.4	**86.3**	8.0	21.2	**23.3**
	switch	**74.5**	72.5	2.3	**15.3**	12.7
Keyboard	cord	**42.6**	39.7	**75.4**	38.3	60.1
	key	37.2	**37.7**	1.0	**4.3**	2.6
Knife	blade	19.3	**27.2**	11.6	**28.0**	21.5
Lamp	base	64.3	**71.1**	**75.6**	59.7	72.2
	body	**48.6**	36.5	1.4	18.1	**20.3**
	bulb	54.5	**59.2**	1.4	**3.8**	2.2
	shade	83.5	**86.4**	32.7	41.8	**48.7**
Laptop	keyboard	**0.0**	**0.0**	**34.5**	5.9	7.1
	screen	**1.0**	**1.0**	34.5	41.6	**50.6**
	shaft	1.2	**3.5**	**1.0**	**1.0**	**1.0**
	touchpad	**0.0**	**0.0**	14.5	30.1	**35.7**
	camera	**0.0**	**0.0**	**1.0**	**1.0**	**1.0**
Microwave	display	**4.2**	1.0	**22.8**	**22.8**	**22.8**
	door	**62.6**	57.1	12.6	18.0	**35.1**
	handle	**1.0**	**1.0**	**100.0**	14.6	16.4
	button	**100.0**	**100.0**	**4.1**	2.2	1.9
Refrigerator	door	**57.1**	54.2	16.9	15.8	**17.7**
	handle	**19.3**	17.2	23.0	26.2	**33.3**
Scissors	blade	6.2	**6.5**	5.4	**7.0**	**7.0**
	handle	82.0	**82.9**	40.1	67.0	**67.7**
	screw	27.2	**28.4**	**9.0**	4.3	4.6
StorageFurniture	door	**86.9**	85.6	**10.2**	4.7	7.5
	drawer	3.9	**4.2**	**7.9**	5.2	4.9
	handle	56.4	**57.5**	33.0	25.3	**33.6**
Table	door	44.4	**49.3**	5.7	**5.9**	5.1
	drawer	35.7	**36.5**	7.4	6.9	**7.9**
	leg	**33.8**	27.4	24.4	35.4	**43.9**
	tabletop	81.2	**82.0**	47.4	63.8	**73.2**
	wheel	1.0	**1.3**	**75.4**	45.0	43.3
	handle	**81.9**	80.8	**11.1**	2.4	3.1
TrashCan	footpedal	34.8	**35.3**	0.0	**1.0**	**1.0**
	lid	**0.0**	**0.0**	40.3	17.1	**45.0**
	door	**0.0**	**0.0**	**7.7**	2.5	2.4
Box	lid	7.2	**8.6**	18.9	26.0	**32.2**
Bucket	handle	1.5	**1.6**	8.8	60.0	**75.6**

category	part	PointGroup* [2]	SoftGroup* [5]	PartSLIP† [4]	Ours (w/o CNVP)†	Ours†
Camera	button	1.0	**1.5**	**19.9**	10.4	8.5
	lens	**16.1**	0.0	**23.0**	21.0	19.8
Cart	wheel	**29.2**	28.4	80.7	83.0	**86.7**
CoffeeMachine	button	**1.0**	**1.0**	**6.9**	3.1	1.6
	container	2.5	**4.0**	16.8	14.3	**17.4**
	knob	**5.6**	5.0	**7.5**	4.9	5.8
	lid	**3.3**	1.4	11.3	14.2	**22.2**
Dispenser	head	27.5	**29.2**	1.8	**2.8**	**2.8**
	lid	20.5	**23.6**	8.2	8.7	**12.1**
Eyeglasses	body	31.7	**39.5**	4.2	**9.9**	8.0
	leg	**68.0**	62.7	1.0	**60.3**	56.1
FoldingChair	seat	**16.8**	**16.8**	**83.3**	70.3	75.0
Globe	sphere	**63.1**	**63.1**	**90.9**	17.6	25.5
Kettle	lid	64.0	**64.4**	28.8	25.8	**40.2**
	handle	51.4	**54.3**	50.4	39.1	**55.9**
	spout	68.5	**72.6**	1.0	1.0	**4.3**
KitchenPot	lid	68.3	**68.5**	**85.2**	50.0	65.7
	handle	**50.6**	50.1	32.7	58.8	**63.1**
Lighter	lid	**30.7**	**30.7**	30.7	20.1	**32.5**
	wheel	**6.0**	5.3	9.0	9.4	**12.4**
	button	64.1	**67.8**	5.5	**19.3**	18.5
Mouse	button	**1.0**	**1.0**	6.3	**8.8**	7.7
	cord	**1.0**	**1.0**	**55.4**	27.1	27.1
	wheel	**83.2**	**83.2**	35.3	46.3	**50.5**
Oven	door	26.5	**31.9**	**42.8**	24.5	23.4
	knob	**1.0**	**1.0**	7.9	**21.7**	19.0
Pen	cap	**48.2**	44.4	2.3	5.5	**10.5**
	button	**16.9**	**16.9**	1.0	**1.1**	**1.1**
Phone	lid	1.0	**1.1**	13.4	51.3	**55.4**
	button	**1.0**	**1.0**	8.8	**19.7**	6.2
Pliers	leg	28.2	**40.4**	1.0	38.6	**40.7**
Printer	button	**1.0**	**1.0**	**1.3**	1.0	1.0
Remote	button	**23.4**	22.5	3.0	**5.2**	2.7
Safe	door	11.0	**12.3**	4.4	**10.9**	10.7
	switch	4.8	**5.4**	1.5	3.5	**4.6**
	button	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**
Stapler	body	86.6	**96.7**	**0.0**	**0.0**	**0.0**
	lid	90.0	**91.8**	32.4	44.7	**89.9**
Suitcase	handle	**25.5**	24.2	35.2	50.9	**71.4**
	wheel	**5.7**	2.9	17.7	18.7	**24.5**
Switch	switch	**7.5**	5.6	3.2	**4.6**	4.4
Toaster	button	9.0	**10.1**	13.8	12.9	**13.9**
	slider	**5.0**	**5.0**	**0.0**	**0.0**	**0.0**
Toilet	lid	5.5	**6.1**	23.7	31.0	**36.2**
	seat	**0.0**	**0.0**	2.4	**3.5**	2.6
	button	**1.0**	**1.0**	**12.5**	8.2	7.6
USB	cap	67.3	**75.7**	14.6	20.4	**25.5**
	rotation	**16.3**	15.0	1.0	1.0	**17.6**
WashingMachine	door	25.0	**34.3**	**48.8**	27.4	38.8
	button	**0.0**	**0.0**	**4.1**	2.4	3.2
Window	window	21.2	**26.4**	1.0	4.0	**4.5**
Overall (45)		31.0	**31.9**	23.3	24.1	**28.5**

* fully supervised; † zero-shot; PartSLIP's overall result reproduces by the official code, with the official paper being 18.0% mAP50.