# Supplementary Materials: Towards Open-World Segmentation of Parts

Tai-Yu Pan[1,*] Qing Liu[2] Wei-Lun Chao[1] Brian Price[2]
[1]The Ohio State University [2]Adobe Research
{pan.667, chao.209}@osu.edu {qingl, bprice}@adobe.com

In this supplementary material, we provide details and results omitted in the main text.

- Appendix A: **Pascal-Part-58 to PartImageNet.** In contrast with the main paper Section 4.4, we further perform the reverse direction of training and evaluation to validate our proposed method.
- Appendix B: **Multiple-Round Self-Training.** In the main paper Section 3.4 we report singe round self-training. In this supplementary, we further explore multiple rounds.
- Appendix C: **Robustness to Imperfect Object Masks** In this section, we visualize more predictions to show the effect of using imperfect object masks in pre-aware setting to support discussions in Section 3.2 of the main paper.
- Appendix D: **Robustness to Unseen Parts.** In this section, we show more cases of applying OPS to unseen objects and unseen parts. Some of our predictions are more reasonable and fine-grained than GT.
- Appendix E: **Comparison to More Baselines.** In this section, we compare to an additional baseline, SCOPS [3], besides Section 4.5 in the main paper.
- Appendix F: **Robustness to Multiple Objects in a Scene.** In this section, we show more cases of applying OPS to an image containing multiple objects.
- Appendix G: **More Qualitative Results.**

## A. Pascal-Part-58 to PartImageNet

As claimed in the main paper, our goal is to improve the robustness of the part segmentation model to the unseen parts. Our proposed method, OPS, utilizes the novel self-supervised (SS) and self-training (ST) fine-tuning approach to learn with unlabeled data. In Section 4.3 and Section 4.4, we investigate two settings: (1) train the base model on PartImageNet train, fine-tune the base model on PartImageNet val without ground-truth labels, evaluate on both PartImageNet val and PartImageNet test; (2) train the base model on PartImageNet train, fine-tune the base model on Pascal-Part train without ground-truth labels, evaluate on Pascal-Part val. In this supplementary, we further provide the result of using Pascal-Part-58 to train the base model, PartImageNet train set to fine-tune, and PartImageNet val/test set to evaluate.

In Tab. A, we see consistent improvements over the base models with our proposed SS and ST methods. Both SS and ST can improve part segmentation when working alone. By combining SS and ST, our proposed full OPS model achieves further gain, which improves the val set from AP 22.59 to 25.14 and the test set from 18.58 to 20.73 with imperfect object masks, and val set from AP 36.39 to 37.93 and test set from 32.08 to 33.71 with perfect object masks. This demonstrates our proposed OPS model indeed achieves improved robustness for part segmentation on unseen objects no matter how it is tested in the cross-dataset setting.

## B. Multiple-Round Self-Training

In the main paper, we report the result of single-round self-training: we use the base model to generate pseudo labels and perform fine-tuning for a single round. In many self-training works, multiple rounds of pseudo-label generation and fine-tuning are usually performed. In this setting, pseudo labels are updated by the fine-tuned model and additional fine-tuning can be applied on top of it. Here, we explore two rounds of self-training for our proposed OPS model, and results are shown in Tab. B.

On PartImageNet, we see the result of OPS gets slightly improved (AP 43.16 to 43.29 on the val set and 40.43 to 40.78 on test set) with imperfect object masks and the performance is nearly the same with perfect object masks. On Pascal-Part-58, we improve from AP 27.69 to 27.83 with perfect object masks, but the performance stays almost the

[*] This work was done during an internship at Adobe Research.

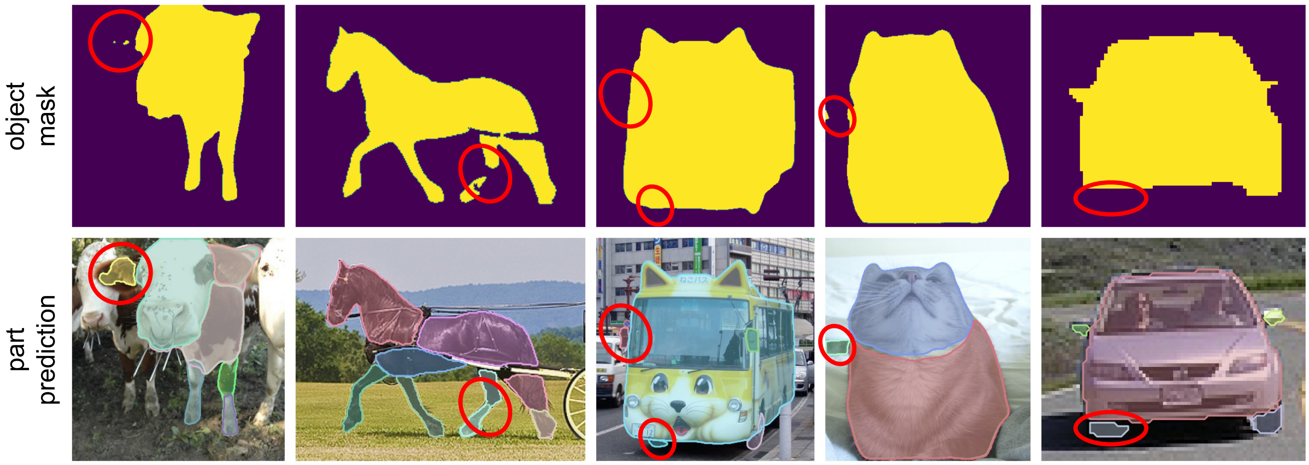

Figure A. **Imperfect object masks.** The red circles indicate the parts missed in the imperfect object masks but recovered by OPS part predictions. The model can recognize the shape of the objects and their belonging parts even though it is trained in a class-agnostic way.

Table A. **Results on PartImageNet [2]** We train the base model on Pascal-Part-58 [1,4] and fine-tune it on PartImageNet train set with proposed self-supervised (SS) and self-training (ST), with imperfect and perfect object masks. Both outperform the base model.

| | | | Val | | Test | |
|---|---|---|---|---|---|---|
| method | SS | ST | AP | $AP_{50}$ | AP | $AP_{50}$ |
| imperf. | | | | | | |
| base | | | 22.59 | 48.06 | 18.58 | 39.47 |
| | ✓ | | 23.60 | 49.90 | 19.27 | 40.64 |
| | | ✓ | 24.60 | 53.17 | 20.23 | 43.70 |
| OPS | ✓ | ✓ | 25.14 | 54.18 | 20.73 | 44.46 |
| perf. | | | | | | |
| base | | | 36.39 | 63.01 | 32.08 | 55.24 |
| | ✓ | | 38.40 | 65.29 | 33.40 | 56.90 |
| | | ✓ | 36.86 | 66.83 | 32.89 | 59.12 |
| OPS | ✓ | ✓ | 37.93 | 67.87 | 33.71 | 59.93 |

Table B. **Results of multiple rounds for pseudo labels.** In the main paper, we report the result of a single round, which generates the pseudo labels only once. In the supplementary, we further explore multiple rounds.

| | single round | | multi rounds | |
|---|---|---|---|---|
| datasets | AP | $AP_{50}$ | AP | $AP_{50}$ |
| PartImageNet val | | | | |
| w/ imperf. | 43.16 | 74.96 | 43.29 | 74.80 |
| w/ perf. | 86.19 | 96.43 | 86.18 | 96.41 |
| PartImageNet test | | | | |
| w/ imperf. | 40.43 | 71.18 | 40.78 | 71.20 |
| w/ perf. | 83.86 | 95.05 | 83.86 | 95.05 |
| Pascal-Part-58 | | | | |
| w/ imperf. | 24.02 | 50.10 | 23.96 | 49.80 |
| w/ perf. | 27.69 | 49.75 | 27.83 | 50.13 |

same (AP 24.02 vs 23.96) with imperfect object masks. Note that the performance on multi-rounds may not be optimal yet because it requires further mining on pseudo labels, which will be investigated more in our future work.

## C. Robustness to Imperfect Object Masks

In the main paper, we propose to apply object-aware learning which aims to capture the fact that parts are "compositions" of their objects. We extract imperfect object masks by an off-the-shelf segmentation model (see main paper for more information) and input with images as an additional channel to RGB. Fig. A shows that OPS part predictions are able to complete the missing parts even though the imperfect object masks are used for pre-awareness. For example, in the first column, the predictions of the cow recover the right ear region. Similarly, the predictions of the horse in the second column repair the leg. In addition, the model is able to exclude the rein since it is less likely to be a part of the horse.

## D. Robustness to Unseen Parts

In Fig. B, we show the robustness of OPS to the unseen objects and parts. Here the test images are from Pascal-Part-58 val, while our OPS model is trained on PartImageNet train and fine-tuned on Pascal-Part-58 train. Some of the predicted parts are not annotated in Pascal-Part [1]. In the first column, the prediction excludes the baby on the chair while the GT labels the whole as a single part. In the third column, the prediction not only finds out the headrest, body, and bottom part of the chair but also discovers all wheels and the armrest. In these cases, our predictions won't get

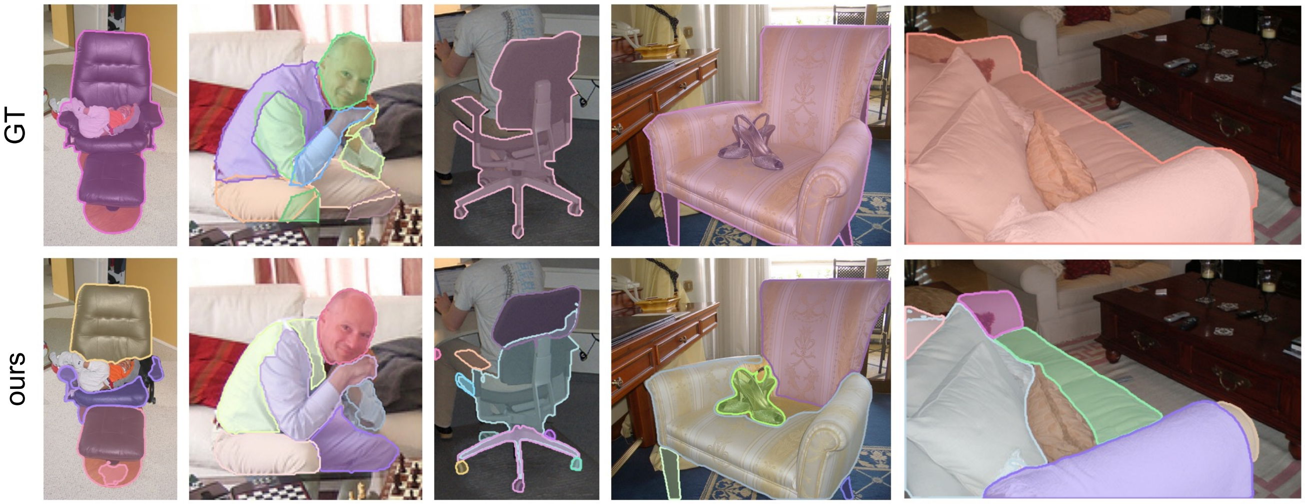

Figure B. **Robustness to unseen parts.** OPS shows strong generalizability to objects and parts that are unseen in PartImageNet [2]. Furthermore, our method is able to segment parts that are not annotated in Pascal-Part [1].

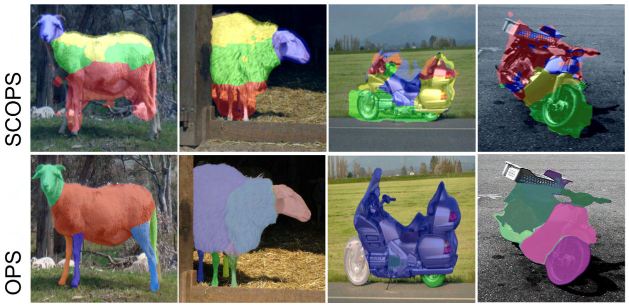

Figure C. **Comparison to SCOPS [3]** (The results in the first row are directly copied from its paper). OPS has a higher quality of parts.

any reward in terms of evaluation metric, e.g., AP or $AP_{50}$, and they will even get lower scores since the predicted parts are not annotated in the ground-truth. We will explore better evaluation metrics for unlabeled part discovery in our future work.

## E. Comparison to More Baselines

In this section, we try to compare OPS to SCOPS [3] as an additional baseline. We note that SCOPS [3] experimented with PASCAL-Part but did not release the checkpoint; it reported object-level IoU (by aggregating parts) but not part-level IoU or AP. Therefore, we perform a qualitative comparison by applying OPS on the PASCAL-Part images that SCOPS [3] showed in their paper. As shown in Fig. C, OPS generally leads to a higher quality of parts.

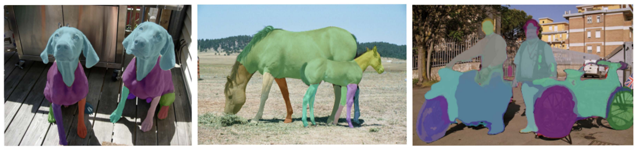

Figure D. **Qualitative results on multiple objects in an image.**

## F. Robustness to Multiple Objects in a Scene

We use single-object images for simplicity by following SCOPS [3]. Meanwhile, our OPS can be applied to a multi-object image *in one pass* by simply including a multi-object mask. As shown in Fig. D, although in the rightmost case, all object masks are connected without differentiation in the mask channel, OPS still correctly recognizes the parts of each object. In addition, simple post-processing with object masks can further refine the part predictions.

## G. More Qualitative Results

Fig. E shows the result on PartImageNet test set by our OPS model trained on PartImageNet train and fine-tuned on PartImageNet val. The part predictions perform well on OOD objects and parts. Some of them even discover more reasonable parts than annotations in GT (*e.g.* chimpanzee on 2nd row and 4th column).

Fig. F shows the result on Pascal-Part-58 val set by our OPS model trained on PartImageNet train and fine-tuned on Pascal-Part-58 train. As mentioned in Appendix D, this demonstrates the robustness of OPS to unseen objects and parts in an even more challenging cross-dataset setting.

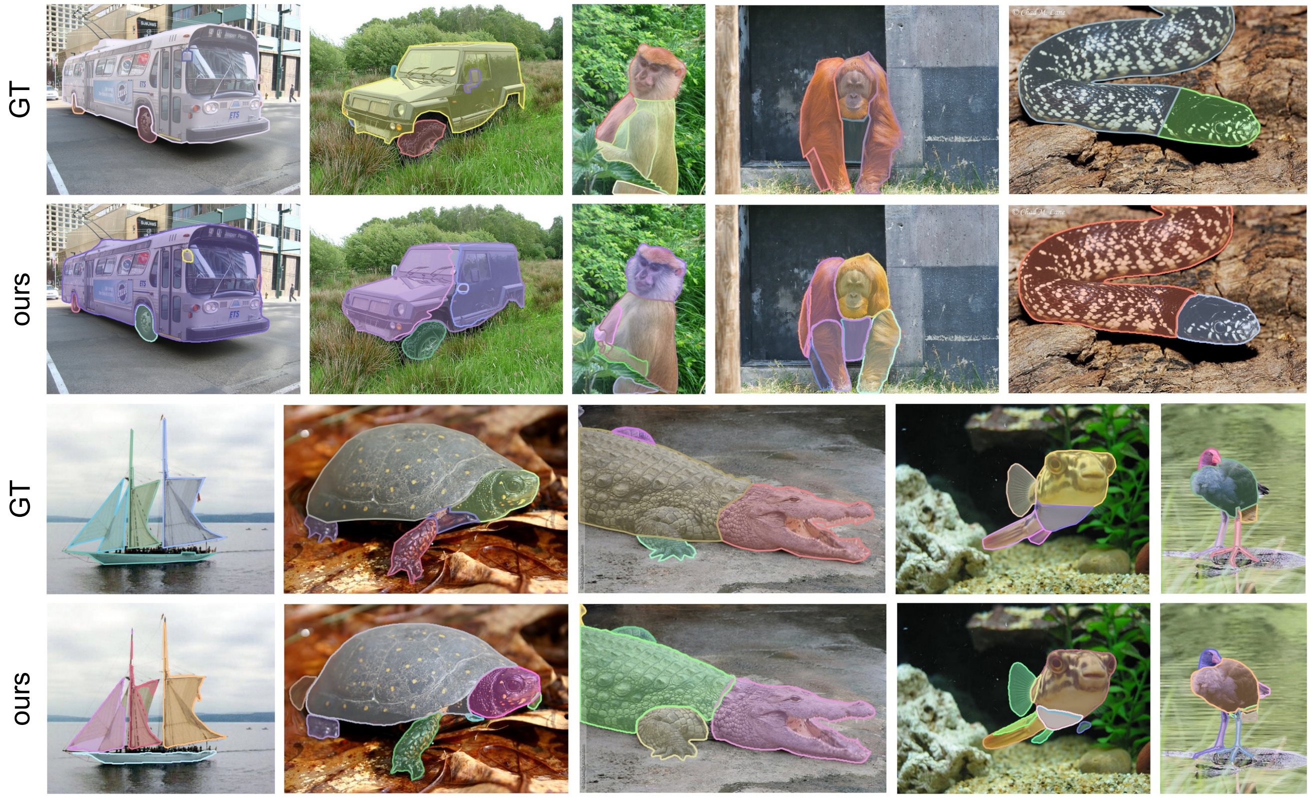

Figure E. **Qualitative results on PartImageNet test set.**

## References

[1] Xianjie Chen, Roozbeh Mottaghi, Xiaobai Liu, Sanja Fidler, Raquel Urtasun, and Alan Yuille. Detect what you can: Detecting and representing objects using holistic models and body parts. In *CVPR*, 2014. 2, 3

[2] Ju He, Shuo Yang, Shaokang Yang, Adam Kortylewski, Xiaoding Yuan, Jie-Neng Chen, Shuai Liu, Cheng Yang, and Alan Yuille. Partimagenet: A large, high-quality dataset of parts. *arXiv*, 2021. 2, 3

[3] Wei-Chih Hung, Varun Jampani, Sifei Liu, Pavlo Molchanov, Ming-Hsuan Yang, and Jan Kautz. Scops: Self-supervised co-part segmentation. In *CVPR*, 2019. 1, 3

[4] Rishubh Singh, Pranav Gupta, Pradeep Shenoy, and Ravi Sarvadevabhatla. Float: Factorized learning of object attributes for improved multi-object multi-part scene parsing. In *CVPR*, 2022. 2

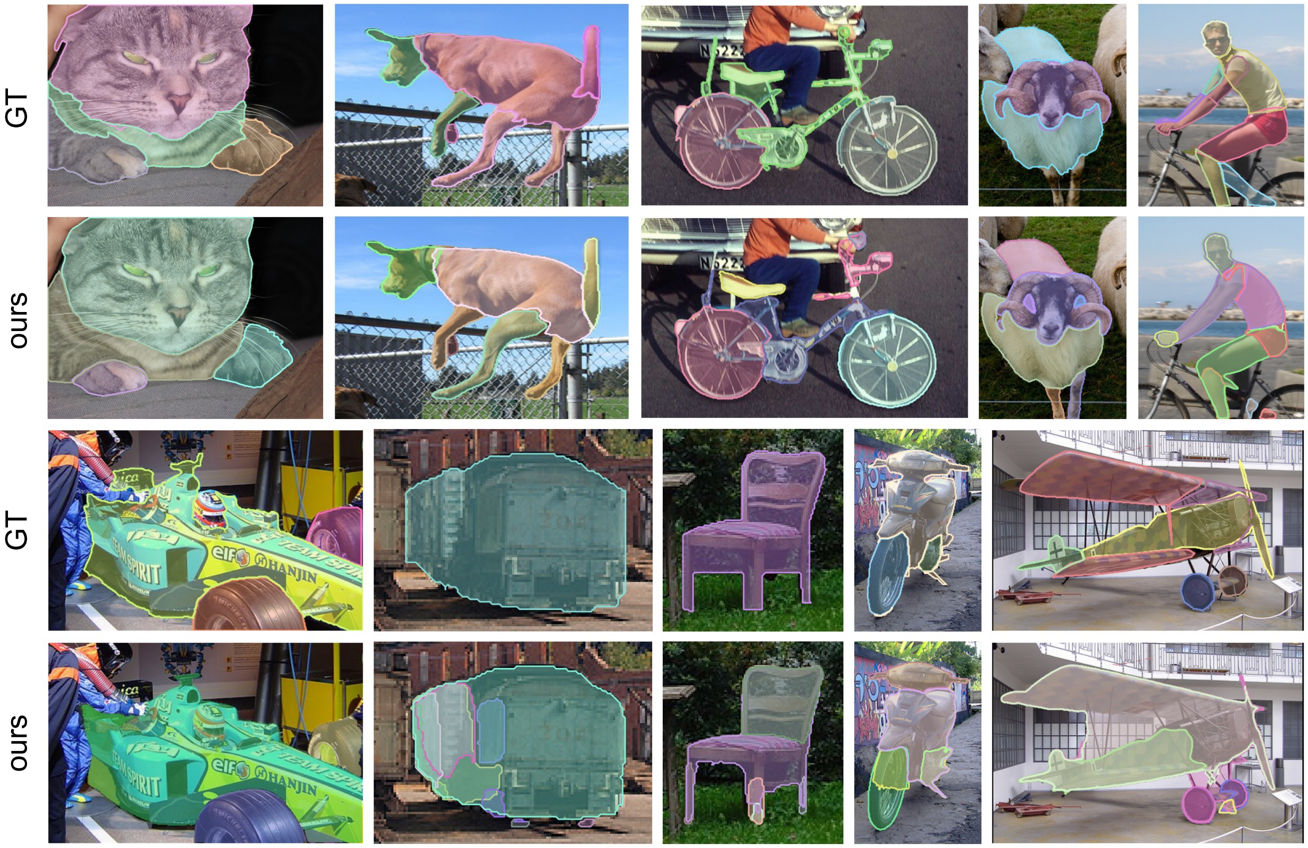

Figure F. **Qualitative results on Pascal-Part-58 val set.**