

## A Technical Appendices and Supplementary Material

This section provides additional visualization and ablation studies.

### A.1 Visualization

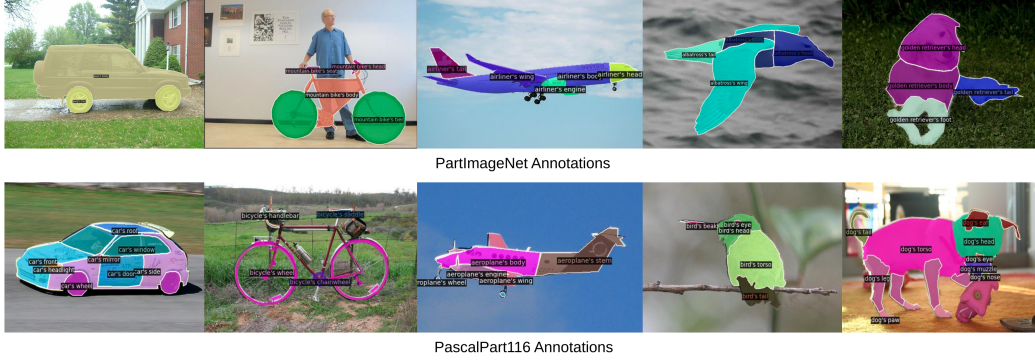


Figure 4: Visualization on Annotations of PartImageNet and PascalPart116 datasets.

**Granularity Difference Across Dataset.** In Fig. 4, we provide additional visualizations of the annotations of PartImageNet and PascalPart116 datasets. The figure shows that the two datasets provide annotations of parts in different granularity. Generally, PascalPart116 has finer part definition and thus it is more challenging to implement part segmentation on PascalPart116 than on PartImageNet, which explains that in both cross-dataset and in-domain settings, LangHOPS and baselines achieve less  $mAP$  on PascalPart116 than on PartImageNet.

**Failure Cases.** We further provide failure cases of LangHOPS in the cross-dataset setting. As shown in Fig. 5, LangHOPS can fail in several cases:

- when the object is distant to the camera and has small area in the image, LangHOPS may not be able to detect all the parts (one motorbike’s wheel missing);
- in the cross-dataset setting, LangHOPS have difficulties in generalizing to some novel parts which it has not see during training (bird’s eye, cat’s eye). As shown in Fig. 4, the training dataset (PartImagenet) only contain annotations of animal’s head and no annotation of eyes.
- when the training and evaluation dataset have different annotation styles, the trained model tends to predict the part segmentation in the style of training dataset (bicycle’s wheel, all the pixels within the wheel circle).

### A.2 Ablation study on Hyper-parameters

**Ablation on  $N_p$ .** We further provide ablation study on the number of repeated part queries for each object  $N_p$  in the cross-dataset setting of **PPS-116+INS+PART** (training)  $\rightarrow$  PartImageNet (evaluation). As shown in Tab. 6, the object-part segmentation performance drops when the  $N_p$  is too small (1, 2) or too large (4, 5, 6), .

$N_p$	1	2	3	4	5	6
Obj AP	61.7	62.1	62.8	62.4	61.4	61.8
Part AP	15.4	15.8	16.4	16.0	16.1	15.9

Table 6: Ablation Study on  $N_p$ .

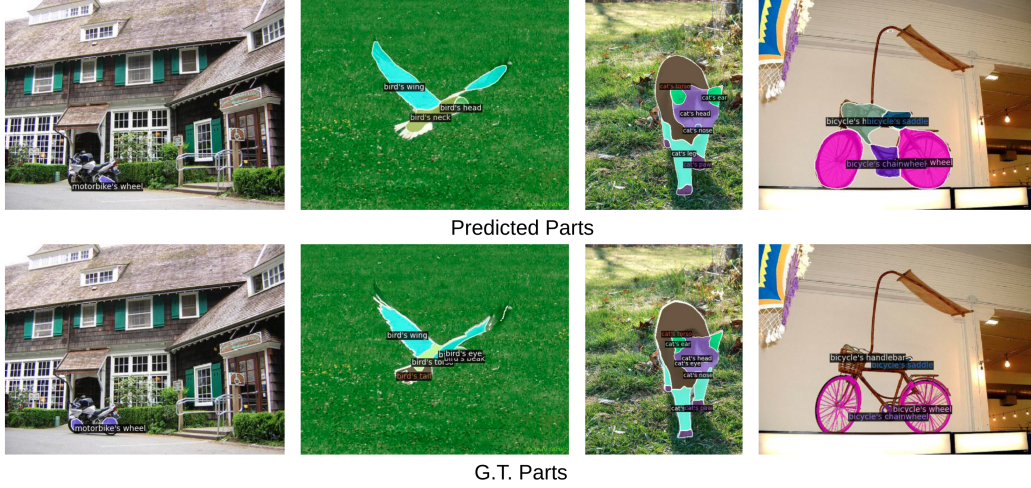


Figure 5: Failure cases of LangHOPS in the cross-dataset setting of **PartImageNet+INS+PART** (training) → PPS-116(evaluation).

Method	<b>PPS-116</b>			<b>+INS</b>			<b>+INS+PART</b>			<b>PartImageNet</b>		
	obj	part	AP	obj	part	AP	obj	part	AP	obj	part	AP
One-Stage	40.6	8.50	15.6	57.8	10.6	21.1	60.2	15.5	25.4	84.6	51.2	58.6
Two-Stage	44.5	8.86	16.7	60.5	11.4	22.3	62.8	16.4	26.7	83.9	49.2	56.9

Table 7: Ablations on training strategy in the cross-dataset setting of **PascalPart-116** (training) → PartImageNet (evaluation).

486 **Ablation on two-stage.** We also provide ablation study on the training strategy of the model.  
487 Two-stage means we firstly train the the model only on object segmentation and secondly train  
488 it on object-part segmentation. One-stage means we directly train the the model on object-part  
489 segmentation from scratch. As shown in Tab. 7, the model trained with two-stage strategy achieves  
490 better cross-dataset performance, though its in-domain performance is inferior compared to one-stage.  
491 The result shows that with the two-stage training, the model can avoid overfit to the training dataset  
492 and achieve better generalization ability.