Part-aware Prototype Network for Few-shot Semantic Segmentation Supplementary Material

Yongfei Liu[1⋆], Xiangyi Zhang[1⋆], Songyang Zhang[1], and Xuming He[1,2]

[1]School of Information Science and Technology, ShanghaiTech University
[2]Shanghai Engineering Research Center of Intelligent Vision and Imaging
{liuyf3, zhangxy9, zhangsy1, hexm}@shanghaitech.edu.cn

In this material, we firstly report the binary-IoU for a clear comparison with previous works in PASCAL-5^i 1-way setting, and then detail the model complexity of our methods. We futher demonstrate the versatility of our model on multi-way setting, and effectiveness of graph attention network for utilizing unlabeled data, as a supplement to Sec.5.3. Finally, we provide more qualitative visualization results for PASCAL-5^i and COCO-20^i.

1 Binary-IoU for PASCAL-5^i

As in Tab.1, our model achieves **70.90**%(1-shot) and **77.45**%(5-shot) with gain of **1.0**% and **6.95**% respectively in terms of binary-IoU, compared with PGNet, which validates the superiority of our method when distinguishing complex background.

Table 1. Averaged binary-IoU over 4 folds of 1-way setting on PASCAL-5^i.

Methods	MetaSeg[30]	OSLSM[3]	co-FCN[23]	A-MCG[13]	PL[7]	AMP[28]	SG-one[38]	PANet[34]	CANet[37]	PGNet[36]	**PPNet(RN-50)**	**PPNet(RN-101)**
1-shot	-	61.30	60.10	61.20	61.20	62.20	63.90	66.50	66.20	69.90	69.19	**70.90**
5-shot	59.50	61.50	60.20	62.20	62.30	63.80	65.90	70.70	69.60	70.50	75.76	**77.45**

2 Model Complexity of 1-way 1-shot Setting

We compare the model complexity with PANet* in 1-way 1-shot setting. We note that the PANet* does not utilize unlabeled data and it is difficult to make direct comparison. Instead, we decompose the computation cost into two parts: prototype generation cost, C_g and the inference cost on a query image, C_q. Below we report model cost (GFLOPs) of the experiment in Tab. 2.

While our method uses more FLOPs in prototype generation due to extra unlabeled data, our inference cost is similar to the PANet*. In each task, as the prototype generation needs to be computed only once, the inference cost will dominate the average computation cost for sufficient number of queries.

⋆ Both authors contributed equally to the work. This work was supported by Shanghai NSF Grant (No. 18ZR1425100)

Table 2. GFLOPs results

Method	C_g	C_q	Average of N queries
PANet*	68.47	68.49	68.47/N+68.49
Ours (w/o unlabeled)	69.39	68.57	69.39/N+68.57
Ours	479.39	68.57	479.39/N+68.57

3 More Quantitative Results on COCO-20^i

3.1 Evaluation on Multi-Way Setting

To demonstrate the model's versatility, we perform more experiments of 2-way and 5-way setting on COCO-20^i. As shown in Tab.3, the results show that our model outperforms the baseline model PANet* by a sizeable margin both with/without unlabeled data. Even in the more challenging 5-way 1-shot setting, our model still improves mean-IoU consistently in each fold.

Table 3. Mean-IoU results of **2-way 1-shot** and **5-way 1-shot** on COCO-20^i *split*-A. Red numbers denote the averaged mean-IoU over 4 folds.

Methods	Backbone	2-way, 1-shot					5-way, 1-shot				
		fold-1	fold-2	fold-3	fold-4	mean	fold-1	fold-2	fold-3	fold-4	mean
PANet [34]	VGG16	29.88	21.13	20.46	15.37	21.71	24.94	19.85	19.28	14.11	19.55
PANet* [34]	RN50	31.86	21.47	21.31	16.43	22.76	27.20	21.50	19.66	15.35	20.93
PPNet(**w/o** $\mathcal{S}^u$)	RN50	33.87	23.98	22.75	17.59	24.55	29.12	22.29	21.10	16.37	22.22
PPNet	RN50	**34.20**	**24.21**	**23.39**	**19.06**	**25.22**	**30.84**	**23.03**	**21.32**	**17.93**	**23.28**

3.2 More Investigation of Graph Attention Network

We compare the effectiveness of the graph attention network in Sec.4.2 with a non-parametric graph attention network for utilizing unlabeled data. Here the non-parametric graph attention network means the network takes cosine distance as similarity function d in Equ.4, 5, and removes the linear mapping weight $\mathbf{W}$ in Equ.4. As in Tab.4, the performance will drop from 27.16% to 26.32%, which suggests that our graph attention network encode meta-knowledge of message propagation, and are effective for capturing informative features from unlabeled data.

4 More Qualitative Visualization for PASCAL-5^i and COCO-20^i

4.1 Visualization for PASCAL-5^i

As in Fig.1, our model can cope with the large appearance and scale variations between support and query images by utilizing the unlabeled data, both in 1-way 1-shot and 2-way 1-shot setting.

Table 4. Ablation studies of 1-way 1-shot on COCO-20^i split-A in every fold. Red numbers denote the averaged mean-IoU over 4 folds.

Model	PAP	SEM	UD	1-shot fold-1	fold-2	fold-3	fold-4	mean
Baseline (PANet*)	-	-	-	31.50	22.58	21.50	16.20	22.95
	✓	✓	✓(w/o params)	36.34	23.59	26.39	18.97	26.32
PPNet	✓	✓	✓	**36.48**	**26.53**	**25.99**	**19.65**	**27.16**

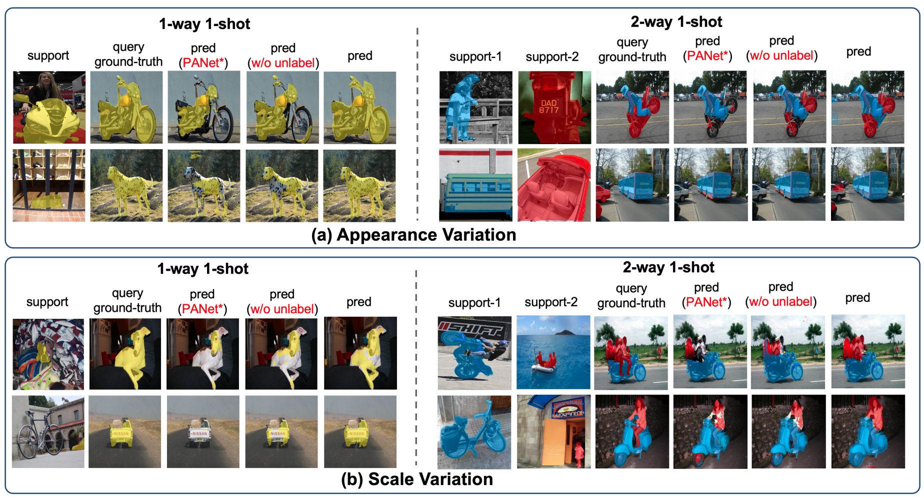

Fig. 1. Qualitative Visualization of 1-way 1-shot and 2-way 1-shot on PASCAL-5^i. (a) demonstrates the prediction results in the appearance variation scenario while (b) shows the prediction in scale variation

4.2 Visualization for COCO-20^i

We also provide more visualization results of 1-way 1-shot setting for COCO-20^i as in Fig.2. Our part-aware prototype network is still capable of modeling one semantic class at a fine-grained level and further coping with variations between support and query images in this more challenging benchmark.

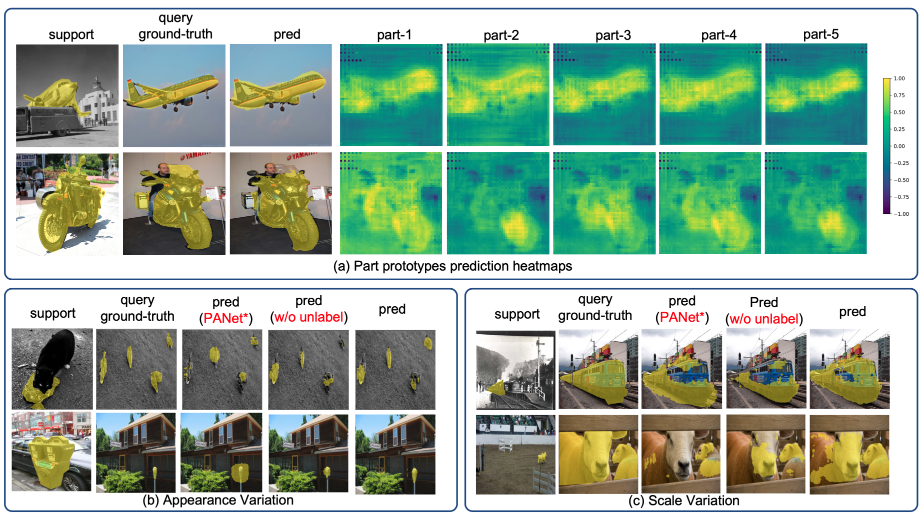

Fig. 2. Qualitative Visualization of 1-way 1-shot on COCO-20^i *split*-A. (a) shows the part prototypes prediction heatmaps. The prediction results in the appearance variation and scale variation scenario are demonstrated in (b),(c) respectively.