Self-Guided and Cross-Guided Learning for Few-Shot Segmentation: Supplement Material

Bingfeng Zhang[1,2], Jimin Xiao[1*], Terry Qin[3]
[1]XJTLU, [2]University of Liverpool, [3] Dinnar Automation Technology
{Bingfeng.Zhang,Jimin.Xiao}@xjtlu.edu.cn, terry.qin@outlook.com

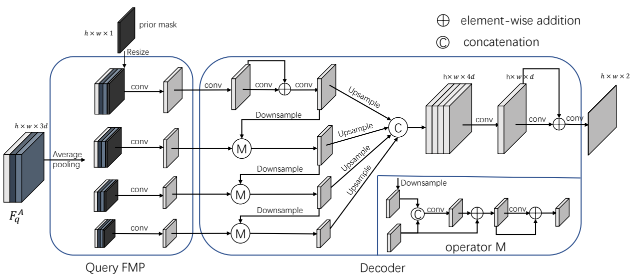

Figure 1. Architecture of the multi-scale query Feature Processing Module (FPM) and decoder in PFENet [7]. PFENet [7] used a prior mask which is generated from the pre-trained model on ImageNet [5] in its query FPM. The height (the width shares the same size) of the feature map after the average pooling is set as {60, 30, 15, 8}.

1. Overview

In this supplement material, we will firstly show the architecture of the multi-scale query FPM and its decoder, then we will provide more quantitative and qualitative results.

2. Multi-Scale Query FPM and Decoder

Fig. 1 shows the details of the multi-scale query FPM and the decoder in PFENet [7]. We follow the same multi-scale setting in our support FPM in SGM, and the only different is that the prior mask is not used in our support FPM. For more details such as how to compute the prior mask and the setting for COCO-20^i, please refer to PFENet [7].

3. More Quantitative Results

Table 1 shows the comparison between our approach and other state-of-the-art methods using FB-IoU as the evaluation metric on PASCAL-5^i. Our approach using PFENet [7] as the baseline achieves 71.9% and 72.8% FB-IoU on PASCAL-5^i for 1-shot and 5-shot tasks respectively, both of which are new state-of-the-art performances. Besides, adopting our approach on CANet [12] improves the performance of the baseline by a large margin, being 4.1% and 1.1% increase for 1-shot and 5-shot tasks, respectively.

Table 1. Comparison with other state-of-the-art methods using FB-IoU (%) on Pascal-5^i for 1-shot and 5-shot segmentation.

Method	Backbone	FB-IoU (%)	
		1-shot	5-shot
OSLSM (BMVC'17) [6]	vgg16	61.3	61.5
co-FCN [4]	vgg16	60.1	60.2
PL (BMVC'18) [1]	vgg16	61.2	62.3
PANet (ICCV'19) [9]	vgg16	66.5	70.7
PGNet (ICCV'19) [11]	resnet50	69.9	70.5
A-MCG (AAAI'19) [2]	resnet101	61.2	62.2
DAN (ECCV'20) [8]	resnet101	71.9	72.3
CANet (CVPR'19) [12]	resnet50	66.2	69.6
PFENet (TPAMI'20) [7] *	resnet50	71.4	-
ours-SCL (CANet)	resnet50	70.3	70.7
ours-SCL (PFENet)	resnet50	**71.9**	**72.8**

* The result is generated using the models provided by the author.

Table 2. Comparison with other state-of-the-art methods using FB-IoU (%) on COCO-20^i for 1-shot and 5-shot segmentation.

Method	Backbone	FB-IoU (%)	
		1-shot	5-shot
PANet (ICCV'19) [9]	vgg16	59.2	63.5
A-MCG (AAAI'19) [2]	resnet101	52.0	54.7
PFENet (TPAMI'20) [7]	resnet101	58.6	61.9
ours-SCL (PFENet)	resnet101	**61.9**	**63.7**

Table 2 shows FB-IoU results between ours and other state-of-the-art methods on COCO-20^i. It can be seen that using our approach achieves new state-of-the-art performance. Compared to the baseline, our approach obtain a large gain of 3.3% and 1.8% FB-IoU for 1-shot and 5-shot segmentation, respectively.

In Table 3, we make a comparison between ours and other approaches on COCO (2017)-20^i. The difference between COCO (2017)-20^i and COCO-20^i is that COCO-20^i

*Corresponding author

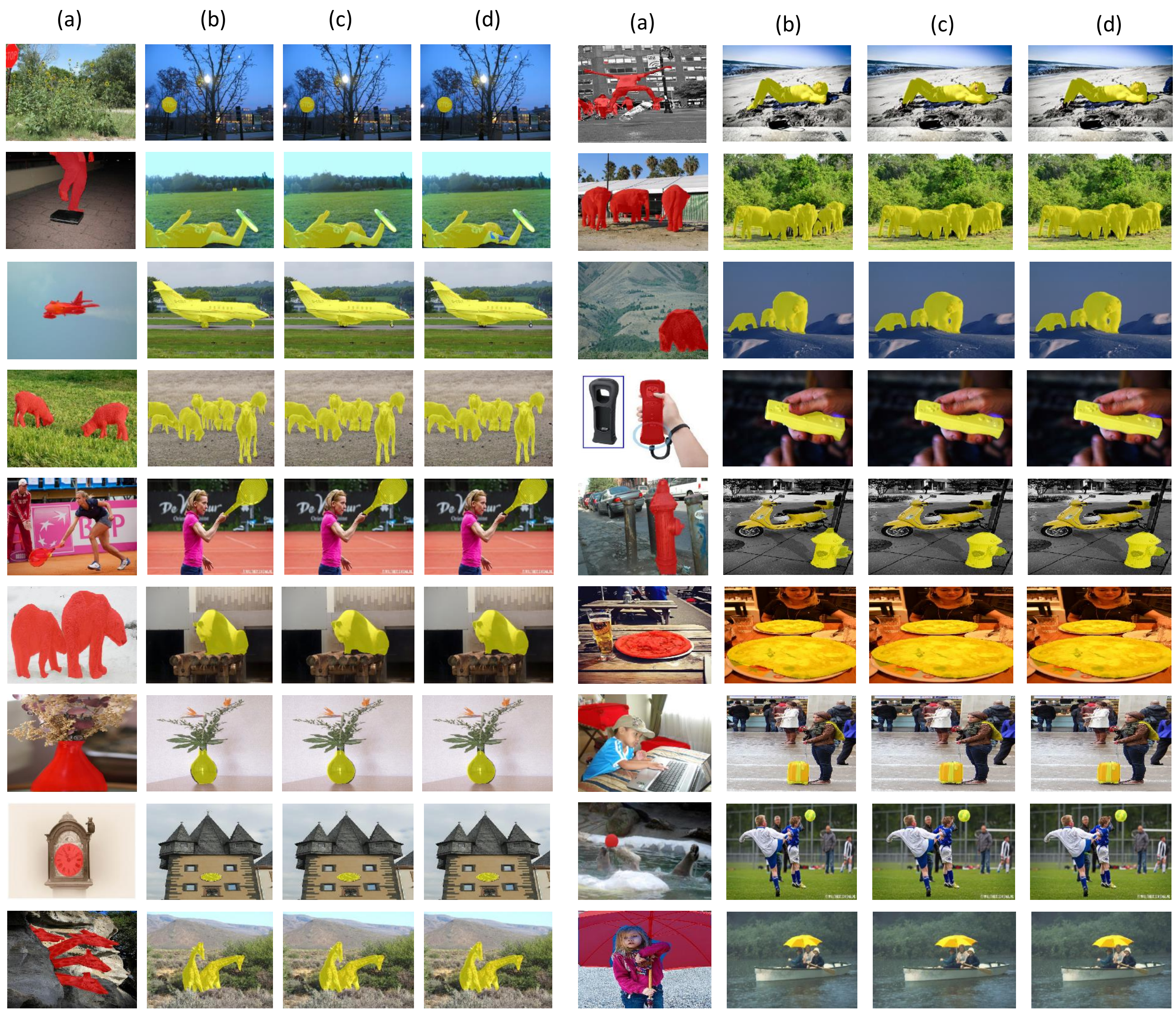

Figure 2. Qualitative results of our proposed approach using PFENet [7] as the baseline on COCO-20^i. (a) Support images for the 1-shot task and their masks. (b) Query images and their ground-truth. (c) *ours-SCL* (PFENet) 1-shot results. (d) *ours-SCL* (PFENet) 5-shot results.

Table 3. Comparison with other state-of-the-art methods using mIoU (%) on COCO (2017)-20^i for 1-shot and 5-shot segmentation.

Method	Backbone	mIoU (%)	
		1-shot	5-shot
CANet (CVPR'19) [12]	resnet50	-	-
RPMMs (ECCV'20) [10]	resnet50	30.6	35.5
ours-SCL (CANet)	resnet50	**32.5**	**35.8**

Table 4. Comparison with the baseline (CANet [12]) about multi-scale inference on Pascal-5^i. MS: multi-scale inference.

Method	MS	mIoU (%)	
		1-shot	5-shot
CANet (CVPR'19) [12]	-	54.0	55.8
CANet (CVPR'19) [12]	✓	55.4	57.1
ours-SCL (CANet)	-	56.3	58.2
ours-SCL (CANet)	✓	**57.5**	**59.2**

is built based on MS-COCO 2014 [3] while COCO (2017)-20^i is built based on MS-COCO 2017. It can be seen that compared to the state-of-the-art method RPMMs [10], our approach obtains a mIoU gain of 1.9% and 0.3% for 1-shot and 5-shot segmentation, respectively. Note that RPMMs [10] also adopted CANet [12] as its baseline.

In Table 4, we show the influence of using the multi-scale method during inference, it can be seen that using our

Figure 3. Qualitative results of our proposed approach using CANet [12] as the baseline on Pascal-5^i. (a) Support images for the 1-shot task and their masks. (b) Query images and their ground-truth. (c) *ours-SCL* (CANet) 1-shot results. (d) *ours-SCL* (CANet) 5-shot results.

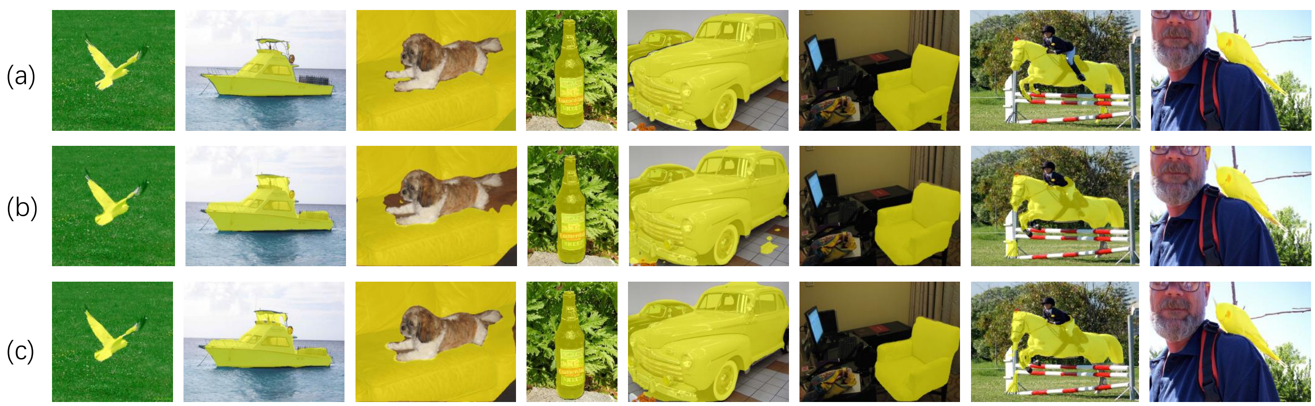

Figure 4. Qualitative results of our proposed approach using PFENet [7] as the baseline on Pascal-5^i. (a) Query images and their ground-truth. (b) PFENet [7] 5-shot segmentation. (c) *ours-SCL* (PFENet) 5-shot results.

method can improve the performance with or without using multi-scale. Besides, it can also be found that the improvement is more obvious for single scale inference.

4. More Qualitative Results

In Fig. 2, we report more qualitative results using PFENet [7] as the baseline on COCO-20^i. It can be seen that our approach retains integral object details for both large and small objects.

In Fig. 3, we report more qualitative results using CANet [12] as the baseline on Pascal-5^i. It can be seen that our approach produces integral segmentation masks covering object details.

In Fig. 4, we report the 5-shot results using PFENet [7] as the baseline on Pascal-5^i. It can be seen that our approach produces integral segmentation masks covering object details.

References

[1] Nanqing Dong and Eric P Xing. Few-shot semantic segmentation with prototype learning. In *Proceedings of the British Machine Vision Conference*, volume 3, 2018. 1

[2] Tao Hu, Pengwan Yang, Chiliang Zhang, Gang Yu, Yadong Mu, and Cees GM Snoek. Attention-based multi-context guiding for few-shot semantic segmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 8441–8448, 2019. 1

[3] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *Proceedings of the European Conference on Computer Vision*, pages 740–755, 2014. 2

[4] Kate Rakelly, Evan Shelhamer, Trevor Darrell, Alyosha Efros, and Sergey Levine. Conditional networks for few-shot semantic segmentation. 2018. 1

[5] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, 115(3):211–252, 2015. 1

[6] Amirreza Shaban, Shray Bansal, Zhen Liu, Irfan Essa, and Byron Boots. One-shot learning for semantic segmentation. *arXiv preprint arXiv:1709.03410*, 2017. 1

[7] Zhuotao Tian, Hengshuang Zhao, Michelle Shu, Zhicheng Yang, Ruiyu Li, and Jiaya Jia. Prior guided feature enrichment network for few-shot segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020. 1, 2, 3

[8] Haochen Wang, Xudong Zhang, Yutao Hu, Yandan Yang, Xianbin Cao, and Xiantong Zhen. Few-shot semantic segmentation with democratic attention networks. In *Proceedings of the European Conference on Computer Vision*, 2020. 1

[9] Kaixin Wang, Jun Hao Liew, Yingtian Zou, Daquan Zhou, and Jiashi Feng. Panet: Few-shot image semantic segmentation with prototype alignment. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 9197–9206, 2019. 1

[10] Boyu Yang, Chang Liu, Bohao Li, Jianbin Jiao, and Qixiang Ye. Prototype mixture models for few-shot semantic segmentation. *arXiv preprint arXiv:2008.03898*, 2020. 2

[11] Chi Zhang, Guosheng Lin, Fayao Liu, Jiushuang Guo, Qingyao Wu, and Rui Yao. Pyramid graph networks with connection attentions for region-based one-shot semantic segmentation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 9587–9595, 2019. 1

[12] Chi Zhang, Guosheng Lin, Fayao Liu, Rui Yao, and Chunhua Shen. Canet: Class-agnostic segmentation networks with iterative refinement and attentive few-shot learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5217–5226, 2019. 1, 2, 3