Supplementary for Learning Meta-class Memory for Few-Shot Semantic Segmentation

Zhonghua Wu[1,2] Xiangxi Shi[3] Guosheng Lin[*1,2] Jianfei Cai[4]
[1]S-lab, Nanyang Technological University
[2]School of Computer Science and Engineering, Nanyang Technological University
[3]Electrical Engineering and Computer Science, Oregon State University
[4]Dept of Data Science and AI, Monash University
zhonghua001@e.ntu.edu.sg shixia@oregonstate.edu gslin@ntu.edu.sg jianfei.cai@monash.edu

1. Effect of Foreground Confidence Module.

Table 1 shows the effectiveness of the Foreground Confidence Module (FCM). The baseline method is that we directly pass the fused meta-class activate maps to FEM for the query mask prediction (denoted as 'Ours w/o FCM'). As shown, the FCM improves mIoU by 2.9%, which suggests that high level features are helpful for the query segmentation mask prediction.

Table 1. Ablation studies on the Foreground Confidence Module (FCM) under 1-shot setting on PASCAL 5^i dataset.

Methods	1 shot				
	Fold 0	Fold 1	Fold 2	Fold 3	Mean
Ours w/o FCM	57.5	68.7	55.8	53.7	58.9
Ours $w/$ FCM	**62.7**	**70.2**	**57.3**	**57.0**	**61.8**

2. Comparison between Quality Measurement Module and the attention mechanism.

We further implement the attention mechanism of CANet into our method. Specifically, same as CANet, we concatenate the support and query features and pass them to two convolutional layers and a global average pooling layer to obtain the importance weights. The weights are used to fuse class activation maps and generate the final weighted activation map for final mask prediction. As shown in Table 2, with the attention mechanism of CANet (denote as "Ours $w/$ Attn"), our model obtains mIoU of 61.7% in the PASCAL 5^i dataset, which is lower than ours with QMM (63.4%). This indicates that QMM is more suitable for our method.

Table 2. Comparsion between our QMM and attention mechanism under 5-shot setting on PASCAL 5^i dataset.

Methods	5 shot				
	Fold 0	Fold 1	Fold 2	Fold 3	Mean
Ours $w/$ Attn	60.1	69.8	55.9	61.1	61.7
Ours $w/$ QMM	**62.2**	**71.5**	**57.5**	**62.4**	**63.4**

3. Visualization results for ablation studies "Features for memory learning"

We visualize the meta-class activate maps obtained from different level features in Fig 1. As shown in the figure, the memory learned from 2nd level features has high activation on local information (e.g. edge) and the one from 3rd level features has high activation on the whole object or parts. A combination of them leads to learning better meta-class memory embeddings, which are able to have high activation on meta-class regions.

*Corresponding author: G. Lin (e-mail: gslin@ntu.edu.sg)

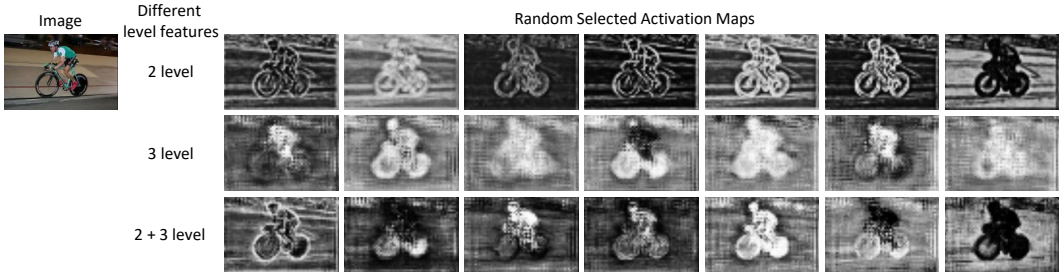

Figure 1. Visual results of the meta-class activation maps with different levels of image features. Here, $2 + 3$ refers to the fused features by channel-wise concatenating the 2nd and 3rd level features followed by a convolution. A combination of them leads the memory to have high activation on the meta-class regions (e.g. edge, part, object, etc.)