A Joint Framework Towards Class-aware and Class-agnostic Alignment for Few-shot Segmentation

Kai Huang[1⋆[0000−0003−3700−9052]], Mingfei Cheng[2⋆[0000−0002−8982−1483]], and Yang Wang[1] and Bochen Wang[1] and Ye Xi[1] and feigege wang[1] and Peng Chen[1]

[1] Alibaba Group, China
{zhouwan.hk, wanyuan.wy, bochen.wbc, yx150449, feigege.wfgg, yuanshang.cp}@alibaba-inc.com
[2] Singapore Management University, Singapore
snowbirds.mf@gmail.com

In this material, we firstly report the further ablation study on Hybrid Prototype Alignment Module (HPAM) to explore the application of multi-scale, and then further show the effect of block sizes (m) and top$k$ ratio (k) in the P2P module.

1 More Study on HPAM with Multi-Scale Scheme

Table 1 reports the ablation studies of multi-scale scheme. "BL" means the baseline of this ablation study which removes the multi-scale scheme as well as the Hybrid Prototype Alignment Module, "BL*" indicates the baseline with multi-scale scheme. We also conduct the single HPAM with only choosing the middle layer features, which are represented as "BL $w/$ single HPAM" and "BL* $w/$ single HPAM" respectively. And "BL* $w/$ HPAM" is the complete framework of our JC^{2}A, in which the HPAM acts in each scale feature maps. From Table 1, we can draw the following conclusion as: 1) It is clear that multi-scale scheme can greatly improve the segmentation performance which has been verified in many FSS methods. 2) Not only for normal baseline but also the baseline with multi-scale scheme, our single HPAM can still make positive effect. 3) the model with multi-scale HPAM achieves highest accuracy in all folds, proving the superiority of our HPAM with multi-scale scheme.

2 Block size and top k settings

We further analyzed the effect of different block sizes (m) and the choice of top$k$ ratio (k) in the P2B. As shown in Table 2, we find that the performance drops with the expansion of block size. These result with different m settings indicate that the large block is able to introduce more irrelevant semantic noises which weaken the representation of target parts within specific regions. Besides, setting k to 20% achieves the best performance while using all blocked features brings

⋆ Equal contribution.

Table 1. The performance comparison of 1-shot and 5-shot segmentation on PASCAL-5^i. Best-performing results are highlighted in bold.

Method	1-shot						5-shot					
	fold-0	fold-1	fold-2	fold-3	mIoU	FB-IoU	fold-0	fold-1	fold-2	fold-3	mIoU	FB-IoU
BL	59.3	67.0	52.9	53.5	58.2	69.5	62.8	67.3	51.6	55.7	59.4	70.2
BL*	62.8	69.6	54.8	56.5	60.9	72.2	65.8	68.7	54.9	58.9	62.1	72.9
BL *w/* single HPAM	63.7	69.5	54.0	57.9	61.3	72.7	65.1	69.9	55.4	59.0	62.4	73.0
BL* *w/* single HPAM	65.4	71.2	56.5	59.0	63.0	75.2	67.7	71.8	57.3	61.4	64.6	74.8
BL* *w/* HPAM (our)	**67.3**	**72.4**	**57.7**	**60.7**	**64.5**	**76.5**	**68.6**	**72.9**	**58.7**	**62.0**	**65.4**	**76.8**

negative effects. We can infer that our block selection of P2B can keep most relevant class-aware information without reducing computational efficiency.

Table 2. Ablation study on different block sizes (m) and top$k$ ratio (k) settings in P2B. "·/·" represents the results of "1-shot/5-shot".

k	10%	20%	30%	40%	50%	60%	70%	80%
m=3	62.1/63.7	**64.5/65.4**	64.3/65.0	64.2/65.1	64.4/65.3	64.5/65.2	64.1/64.9	63.0/64.3
m=5	61.4/62.8	63.2/64.1	63.0/64.0	63.5/64.7	62.9/63.4	63.3/63.8	62.2/63.2	62.0/63.7
m=7	60.7/61.9	63.0/64.4	62.9/63.7	62.5/63.7	63.1/64.2	62.7/64.4	62.0/63.6	61.6/62.9