A Parameter Amount and FLOPs

Table 1: Impacts of AENet on parameters and FLOPs.

Method	#Parameters	#FLOPs
SCCAN	35.0M	480.9G
SCCAN + AENet	$40.7\text{M}_{14.0\%\uparrow}$	$528.4\text{G}_{9.0\%\uparrow}$

Our proposed AENet serves as a plugin for existing cross attention-based few-shot segmentation (FSS) methods. Take SCCAN as an example, it has 8 self-calibrated cross attention (SCCA) blocks, thus we insert 8 ambiguity eliminators (AE) in total, with each of them inserted before 1 SCCA block. Besides, we replace its pseudo mask aggregation (PMA) module with our prior generator (PG). We summarize the parameter amount, as well as the FLOPs, of SCCAN and SCCAN + AENet in Tab. 1, and could observe that our proposed AENet is lightweight in both parameters and computations, *e.g.*, the there is only a 9% increase in terms of FLOPs.

B Weak Support Labels

Table 2: Performance comparisons with weak support labels (bounding boxes). The backbone is ResNet50. * show the performance with accurate pixel-wise labels.

Method	1-shot					
	5^0	5^1	5^2	5^3	Mean	FB-IoU
PANet	-	-	-	-	45.1	-
CANet	-	-	-	-	52.0	-
DPCN	59.8	70.5	63.2	55.5	62.3	-
SCCAN	67.3	71.8	65.6	58.0	65.7	75.5
SCCAN*	68.3	72.5	66.8	59.8	66.8	77.7
SCCAN + AENet	71.8	74.6	67.1	61.7	68.8	80.5
SCCAN + AENet*	72.2	75.5	68.5	63.1	69.8	80.8

Given a specific class, semantic segmentation relies on large number of manually annotated samples to learn its representative features for segmentation. The most inspiring thing of few-shot segmentation (FSS) is it can greatly reduce the annotation cost from more than thousands of samples to only 1 or 5 samples. However, some existing studies think that even 1 or 5 pixel-wise labels still cost much, they further conduct experiments under the scenario where cheaper weak support labels are provided, *e.g.*, bounding boxes. It can be observed from Tab. 2

Table 3: Testing results of 20,000 episodes on COCO-20^i in terms of mIoU and FB-IoU. "20^i" shows the mIoU scores of 20 novel classes in fold i, "Mean" is the averaged mIoU score from all folds. * means testing with 4,000 episodes, and † means testing with 20,000 episodes.

Backbone	Method	1-shot						5-shot					
		5^0	5^1	5^2	5^3	Mean	FB-IoU	5^0	5^1	5^2	5^3	Mean	FB-IoU
VGG16	SCCAN + AENet*	40.3	50.4	47.9	44.9	45.9	71.2	45.8	56.3	55.8	53.4	52.8	74.3
	SCCAN + AENet†	39.4	49.9	46.2	44.9	45.1	71.1	45.2	56.0	55.0	52.6	52.2	74.3
ResNet50	SCCAN + AENet*	43.1	56.0	50.3	48.4	49.4	73.6	51.7	61.9	57.9	55.3	56.7	76.5
	SCCAN + AENet†	42.6	56.3	48.8	48.6	49.1	73.5	49.5	61.8	56.5	55.6	55.8	76.6

that the proposed AENet also works well with weak support labels, validating the effectiveness of our idea, *i.e.*, mining discriminative query regions with a "subtraction" operation, which can mitigate the side-effects of the BG features mingled in FG features.

C More testing episodes on COCO-20^i

Compared to PASCAL-5^i, COCO-20^i contains much more images. Therefore, we follow PFENet to test SCCAN + AENet with 20,000 testing episodes, so as to obtain more robust results on COCO-20^i. The results are shown in Tab. 3, and it could be observed that there is no prominent performance drop, which means the proposed method is stable.

Table 4: Error bars evaluation on COCO-20^i. The random seeds are taken from {0, 1, 2, 3, 321} to generate 4,000 testing episodes. **Bold** values denote the best cases.

Method	Seed	1-shot					
		20^0	20^1	20^2	20^3	Mean	FB-IoU
HDMNet	0	45.5	55.3	49.6	46.7	49.3	71.9
	1	45.3	54.9	50.8	48.3	49.8	71.8
	2	44.9	54.2	50.0	48.7	49.5	72.2
	3	44.1	54.9	51.9	48.6	49.9	72.1
	321	44.8	54.9	50.0	48.7	49.6	72.0
	Mean	44.9	54.9	50.5	48.2	49.6	72.0
	Std	**0.6**	0.4	0.9	0.9	0.3	0.2
HDMNet + AENet	0	46.7	57.7	52.0	49.5	51.5	74.3
	1	47.2	57.1	51.5	51.1	51.7	74.4
	2	45.7	57.6	52.0	49.1	51.1	74.4
	3	47.1	57.9	52.8	49.7	51.9	74.4
	321	45.4	57.1	52.6	50.0	51.3	74.4
	Mean	**46.4**	**57.5**	**52.2**	**49.9**	**51.5**	**74.4**
	Std	0.8	**0.4**	**0.5**	**0.7**	**0.3**	**0.1**

D Error Bars Evaluation

We further report the 1-shot error bars evaluation of HDMNet and HDMNet + AENet in Tab. 4. It can be observed that with our plug-in AENet, (1) HDMNet can consistently be improved by considerable margins, and (2) the results are stable, as the standard deviation of mean mIoU score is only 0.3.

E More Qualitative Results

More qualitative results of SCCAN and SCCAN + AENet are displayed in Fig. 1, where we can observe a stable improvement of AENet over SCCAN.

Fig. 1: More qualitative results of SCCAN and SCCAN + AENet.

F More Visualizations of Prior Masks

The visualizations of discriminative prior masks serve as the direct evidences to the effectiveness of our main idea. In this section, we depict more examples in Fig. 2. We could observe that existing prior mask generation methods would mistakenly activate many wrong areas, and become ineffective, while our discriminative prior mask (M_{Prior}^{Disc}) can consistently suppress them well, which demonstrates the effectiveness of our "subtraction" operation in Eq. (4).

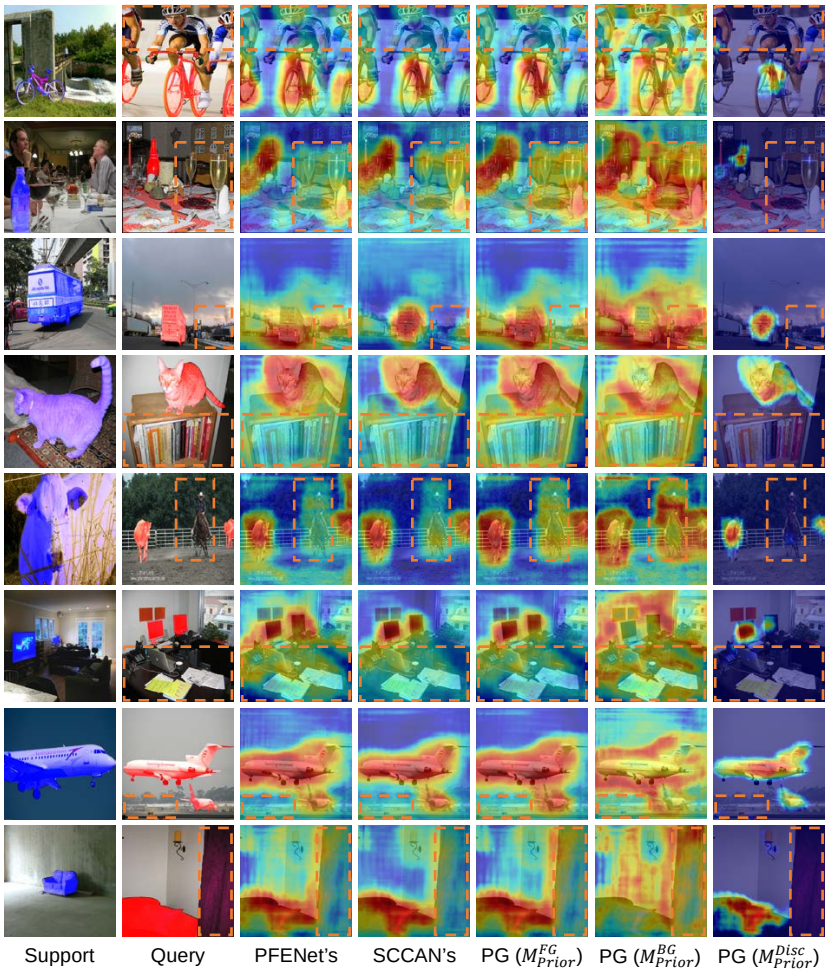

Fig. 2: More visualizations results of different prior masks, including the prior masks from PFENet, SCCAN and our prior generator (PG). We use some orange rectangles to highlight some challenging areas.

G Limitation and Future Direction

The main motivation of the designed plugin ambiguity elimination network (AENet) is to improve the query-support FG-FG matching for existing cross attention-based FSS methods. More concretely, this is achieved by mining discriminative query FG regions and then using them for query and support features refinement. Although the query and support FG pixels can consequently contain more FG information (so as to enhance FG-FG matching naturally), the query BG pixels would also be fused with the discriminative query FG features, making the refined query FG and BG features hard for separation. Therefore, a possible future direction is to design a module to prevent query BG pixels from fusing discriminative FG features.