Model	VLM	Additional Backbone	Training Dataset	A-847	PC-459	A-150	PC-59	PAS-20	PAS-20^b
OpenSeg [Ghiasi *et al.*, 2022]	ALIGN	ResNet-101	COCO Panoptic	4.4	7.9	17.5	40.1	-	63.8
OpenSeg [Ghiasi *et al.*, 2022]	ALIGN	Eff-B7	COCO Panoptic	8.1	11.5	26.4	44.8	-	70.2
ZegFormer [Ding *et al.*, 2022]	CLIP ViT-B/16	ResNet-101	COCO-Stuff	5.6	10.4	18.0	45.5	89.5	65.5
ZSSeg [Xu *et al.*, 2022]	CLIP ViT-B/16	ResNet-101	COCO-Stuff	7.0	-	20.5	47.7	88.4	-
OVSeg [Liang *et al.*, 2023]	CLIP ViT-B/16	ResNet-101c	COCO-Stuff	7.1	11.0	24.8	53.3	92.6	-
SAN [Xu *et al.*, 2023b]	CLIP ViT-B/16	-	COCO-Stuff	10.1	12.6	27.5	53.8	94.0	-
CAT-Seg [Cho *et al.*, 2023]	CLIP ViT-B/16	ResNet-101	COCO-Stuff	8.9	16.6	27.2	57.5	93.7	77.3
SCAN [Liu *et al.*, 2024]	CLIP ViT-B/16	ResNet-101	COCO-Stuff	10.8	13.2	30.8	**58.4**	**97.0**	-
EBSeg [Shan *et al.*, 2024]	CLIP ViT-B/16	SAM	COCO-Stuff	11.1	17.3	30.0	56.7	94.6	-
SED [Xie *et al.*, 2024b]	ConvNeXt-B	-	COCO-Stuff	11.4	18.6	31.6	57.3	94.4	-
PAGSeg(Ours)	CLIP ViT-B/16	-	COCO-Stuff	**12.4**	**18.9**	**32.1**	57.8	95.0	**78.2**
OVSeg [Liang *et al.*, 2023]	CLIP ViT-L/14	Swin-B	COCO-Stuff	9.0	12.4	29.6	55.7	94.5	-
SAN [Xu *et al.*, 2023b]	CLIP ViT-L/14	-	COCO-Stuff	12.4	15.7	32.1	57.7	94.6	-
ODISE [Xu *et al.*, 2023a]	CLIP ViT-L/14	Stable Diffusion	COCO-Stuff	11.1	14.5	29.9	57.3	-	-
CAT-Seg [Cho *et al.*, 2023]	CLIP ViT-L/14	Swin-B	COCO-Stuff	11.4	20.4	31.5	62.0	96.6	81.8
FC-CLIP [Yu *et al.*, 2023]	ConvNeXt-L	-	COCO Panoptic	14.8	18.2	34.1	58.4	95.4	81.8
SCAN [Liu *et al.*, 2024]	CLIP ViT-L/14	Swin-B	COCO-Stuff	14.0	16.7	33.5	59.3	**97.2**	-
EBSeg [Shan *et al.*, 2024]	CLIP ViT-L/14	SAM	COCO-Stuff	13.7	21.0	32.8	60.2	96.4	-
SED [Xie *et al.*, 2024b]	ConvNeXt-L	-	COCO-Stuff	13.9	22.6	35.2	60.6	96.1	-
PAGSeg(Ours)	CLIP ViT-L/14	-	COCO-Stuff	**16.2**	**24.0**	**38.1**	**62.9**	97.1	**82.3**

Table 1: **Performance comparison with state-of-the-art methods.** We perform experiments on six well-established open-vocabulary semantic segmentation benchmarks and use mIoU as the evaluation metric. The best-performing results are presented in bold, while the second-best results are underlined.