Table 3. Performance comparison of our approach with other methods on eight semantic segmentation benchmarks. We report results on: **With background category**: VOC21, Context60, COCO-Obj; **Without background category**: VOC20, City, Context59, ADE, Stuff. The best and second-best results are marked in **bold** and underline, respectively. The shaded region ours(SD) represents results obtained by processing our final feature maps using the SD model following CLIPer's(Sun et al., 2024a) approach.

Method	Post-processing	With background			Without background					Avg.
		VOC21	Context60	COCO-Obj	VOC20	City	Context59	ADE	Stuff	
GroupViT(Xu et al., 2022)	No	50.4	18.7	27.5	79.7	11.1	23.4	9.2	15.3	27.7
TCL(Cha et al., 2023)	No	51.2	24.3	30.4	77.5	23.1	30.3	14.9	19.6	33.1
C-Dsier(Wysoczańska et al., 2023)	No	62.2	32.4	35.0	80.2	31.7	35.9	20.0	24.6	40.3
CoDe(Wu et al., 2024)	No	57.5	30.5	32.3	-	28.9	-	17.7	23.9	-
CLIP(Radford et al., 2021)	No	20.8	9.3	8.9	49.1	6.7	11.2	3.2	5.7	14.4
ReCo(Shin et al., 2022)	No	25.1	19.9	15.7	57.7	21.6	22.3	11.2	14.8	23.7
MaskCLIP(Ding et al., 2023)	No	38.8	23.6	20.6	74.9	12.6	26.4	9.8	16.4	27.9
CLIPSurgery(Li et al., 2023)	No	55.2	30.3	29.7	77.5	33.1	33.4	16.1	22.2	37.2
GEM(Bousselham et al., 2024)	No	46.2	32.6	33.9	79.9	21.2	35.9	15.7	23.7	36.1
SCLIP(Wang et al., 2025)	No	59.1	30.4	30.5	80.4	32.2	34.2	16.1	22.4	38.2
ClearCLIP(Lan et al., 2025b)	No	57.0	32.2	32.5	82.3	32.8	35.8	17.3	24.0	39.2
CAR(Sun et al., 2024b)	No	48.6	13.6	15.4	73.7	-	18.4	5.4	-	-
OVDiff(Karazija et al., 2023)	No	66.3	29.7	34.6	80.9	23.4	32.9	14.1	20.3	37.8
LAVG(Kang & Cho, 2025)	No	62.1	31.6	34.2	82.5	26.2	34.7	15.8	23.2	38.8
ProxyCLIP(Lan et al., 2025a)	No	61.3	35.3	37.5	80.3	38.1	39.1	20.2	26.5	42.3
CLIPer(Sun et al., 2024a)	No	60.1	34.8	36.0	84	-	38.5	19.8	25.3	-
SCCLIP(Bai et al., 2024)	No	64.6	36.8	37.7	84.3	41.0	40.1	20.1	26.6	43.9
Trident(Shi et al., 2024)	No	64.5	37.2	**39.5**	83.7	40.4	40.9	20.9	27.6	44.5
Ours	No	**66.1**	**37.7**	39.4	**85.4**	**42.8**	**41.2**	**21.0**	**27.3**	**44.8**
Post-processing										
NACLIP(Hajimiri et al., 2024)	Yes	64.1	35.0	36.2	83.0	38.3	38.4	19.1	25.7	42.5
CLIPer(Sun et al., 2024a)	Yes	65.9	37.6	39.0	85.2	-	41.7	21.2	27.5	-
Trident(Shi et al., 2024)	Yes	67.1	38.6	**41.1**	84.5	42.9	42.2	21.9	28.3	45.8
Ours	Yes	**70.4**	**39.8**	40.8	**86.2**	**46.4**	**43.6**	**22.2**	**28.7**	**47.3**
Ours(SD)	Yes	71.0	40.2	41.4	86.4	46.9	43.9	22.3	29.1	47.7

Table 4. Efficiency Evaluation and Ablation Study of Post-processing Methods. Performance and computational efficiency across seven datasets without post-processing, conducted on a single RTX4090 24G GPU. The table compares various post-processing approaches: Trident SAM (from Trident(Shi et al., 2024)), SD (from CLIPer(Sun et al., 2024a)), their combination (Trident SAM+SD), and our speed-optimized version (Optimization SAM). The bottom rows present ablation experiments on Optimization SAM. Note: Coco-obj dataset experiments were omitted due to GPU memory constraints (24GB). Optimization SAM maintains the same core principles as Trident SAM.

Method	With background		Without background					Avg.	Total Time
	VOC21	Context60	VOC20	City	Context59	ADE	Stuff		
Main Methods									
Trident SAM+SD	**71.0**	**40.2**	**86.4**	**46.9**	**43.9**	**22.3**	**29.1**	**48.5**	–
Time	22min 2s	87min 44s	22min	41min 16s	88min 27s	29min 31s	93min 51s	–	384min 51s
Trident SAM	70.4	39.8	86.2	46.4	43.6	22.2	28.7	48.2	–
Time	2min 50s	20min 03s	2min 39s	6min 33s	20min 25s	5min 26s	24min 11s	–	82min 7s
Optimization SAM	70.4	39.8	86.2	46.4	43.6	22.2	28.7	48.2	–
Time	**2min 30s**	**19min 28s**	**2min 26s**	**5min 52s**	**19min 46s**	**4min 41s**	**22min 03s**	–	**76min 46s**
NO SAM	66.1	37.7	85.4	42.8	41.2	21.0	27.3	45.9	-
Time	1min 36s	15min 33s	1min 33s	4min 8s	15min 23s	2min 23s	16min 16s	-	56min 52s
Ablation Studies (based on Optimization SAM)									
w/o CCA	69.1	39.1	86.0	42.4	42.9	21.5	28.16	47.0	–
w/o box	65.7	38.0	84.7	42.9	41.4	21.3	27.2	46.0	–
w/o mask	**71.0**	39.7	86.1	46.2	**43.6**	**22.2**	**28.7**	**48.2**	–
full model	70.4	**39.8**	**86.2**	**46.4**	**43.6**	**22.2**	**28.7**	**48.2**	–