OV-COAST: Cost Aggregation with Optimal Transport for Open-Vocabulary Semantic Segmentation

Aditya Gandhamal*, Aniruddh Sikdar*, Suresh Sundaram
Indian Institute of Science

Department of Aerospace Engineering
Indian Institute of Science, Bangalore

CVPR Nashville JUNE 11-15, 2025

Objective

- To improve out-of-domain generalization for Open-Vocabulary Semantic Segmentation .
- To use optimal transport theory to align visual and textual features for improving segmentation performance on unseen classes.

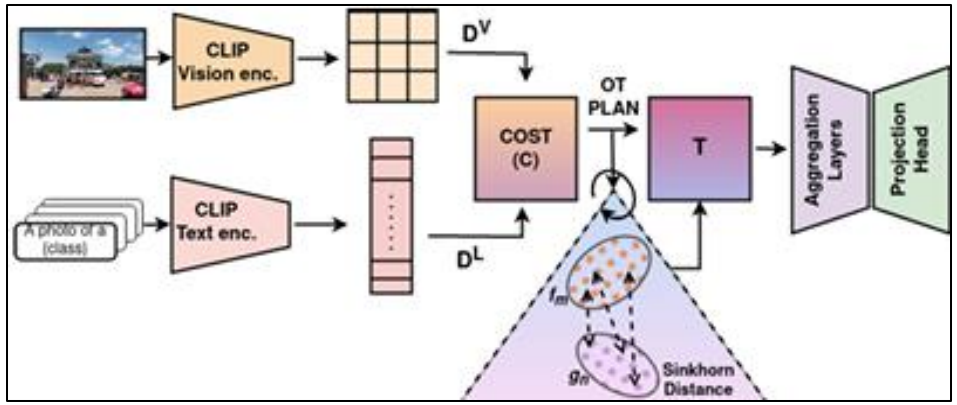
Comparison between existing techniques

Method	MPG	MTP	ITM
ZegFormer [6]	✓	✗	-
ZSseg [28]	✓	✗	-
ZegCLIP [36]	✗	✓	MHA
OVSeg [14]	✓	✓	Contrastive Loss
SAN [29]	✓	✗	-
CAT-Seg [4]	✗	✓	Cost Aggregation
OV-COAST (ours)	✗	✓	Cost Aggregation + Optimal Transport

Contributions

- A two-stage optimization strategy using optimal transport and cost aggregation to achieve improved multi-modal alignment in the CLIP embedding space.
- State-of-the-art performance on the MESS Benchmark

Architecture Overview



$$U = \sum_{m=1}^{M} u_m \delta_{D^V} \quad and \quad V = \sum_{n=1}^{N} v_n \delta_{D^L}$$

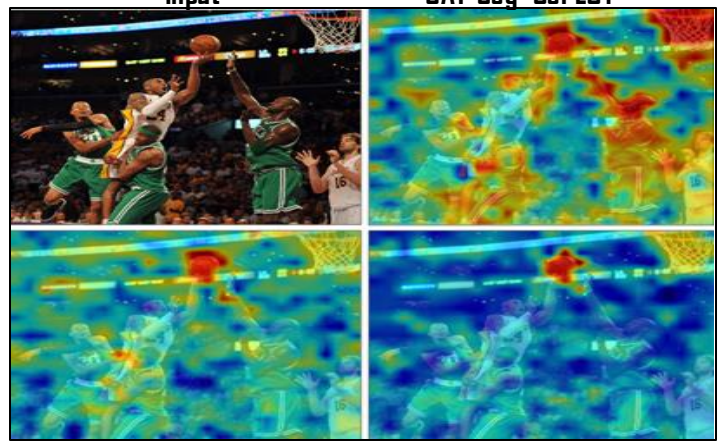
$$< T, C >= \sum_{m=1}^{M} \sum_{n=1}^{N} T_{m,n} C_{m,n}$$

$$d_{OT}(u, v|C) = \arg\min_{T} < T, C >$$

$$T^* = diag(u^{(t)}) exp(-C/\lambda) diag(v^{(t)})$$

Ablation of optimal transport with prompt learning techniques

Method	mIoU (on MESS)
CAT-Seg + VPT + Optimal Transport	20.68
CAT-Seg + CoPLOT	23.98
CAT-Seg + PLOT	26.61
OV-COAST	**31.15**



Performance Comparison on the MESS Benchmark

	Dark Zurich	MHP v1	FoodSeg103	ATLANTIS	iSAID	ISPRS	WorldFloods	FloodNet	UAVid	Kvasir-inst.	CHASE DB1	CryoNuSeg	Corrosion CS	DeepCrack	PST900	ZeroWaste-f	SUIM	CUB-200	CWFID	Mean
Random (L.B)	1.31	1.27	0.23	0.56	0.56	8.02	18.43	3.39	5.18	27.99	27.25	31.25	9.3	26.52	4.52	6.49	5.3	0.06	13.08	10.27
Best Sup. (U.B)	63.9	50.0	45.1	42.22	65.3	87.56	92.71	82.22	67.8	93.7	97.05	73.45	49.92	85.9	82.3	52.5	74.0	84.6	87.23	70.99
ZSSeg-B	16.86	7.08	8.17	22.19	3.8	11.57	23.25	20.98	30.27	46.93	37.0	**38.7**	3.06	25.39	18.76	8.78	30.16	4.35	32.46	22.73
ZegFormer-B	4.52	4.33	10.01	18.98	2.68	14.04	25.93	22.74	20.84	27.39	12.47	11.94	4.78	29.77	19.63	17.52	28.28	16.8	32.24	17.57
X-Decoder-T	24.16	3.54	2.61	27.51	2.43	31.47	26.23	8.83	25.65	55.77	10.16	11.94	1.72	24.65	19.44	15.44	24.75	0.51	29.25	19.8
SAN-B	24.35	8.87	19.27	36.51	4.77	37.56	31.75	37.44	41.65	**69.88**	17.85	11.95	19.73	50.27	19.67	**21.27**	22.64	**16.91**	5.67	26.21
OpenSeeD-T	28.13	2.06	9.0	18.55	1.45	31.07	30.11	23.14	39.78	59.69	**46.88**	33.76	13.38	47.84	2.5	2.28	19.45	0.13	11.478	21.82
Gr.-SAM-B	20.91	**29.38**	10.48	17.33	12.22	26.68	33.41	19.19	38.35	46.82	23.56	38.06	**20.88**	**59.02**	**21.39**	16.74	14.13	0.43	38.41	25.65
ODISE	28.88	7.31	16.62	36.73	11.71	**54.83**	26.30	35.06	**43.46**	28.02	20.60	12.27	10.74	24.34	20.18	5.74	26.24	4.62	30.43	23.37
CAT-Seg	28.86	23.74	**26.69**	**40.31**	19.34	45.36	35.72	37.57	41.55	48.2	16.99	15.7	12.29	31.67	19.88	17.52	**44.71**	10.23	42.77	29.42
OV-COAST	**28.94**	25.15	25.92	39.59	**19.87**	46.02	**37.22**	**37.59**	43.16	46.56	30.30	15.76	14.47	44.22	19.66	18.0	44.61	10.52	**44.34**	**31.15**