Revisit The Open Nature of Open-Vocabulary Semantic Segmentation

Qiming Huang, Han Hu, Jianbo Jiao
The MIx Group @ University of Birmingham

ICLR

Project Page

What is the Open Nature of OVS?

Are these correct or incorrect?

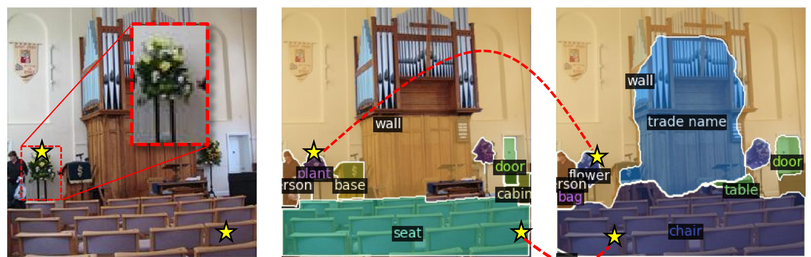


Image | Human Labels | Prediction

OVS Inference $\hat{Y} = \arg\max_{y_i \in \mathcal{V}} P(y_i|X, \Theta, \mathcal{V})$

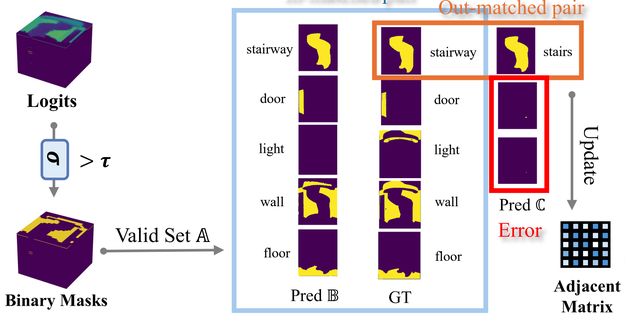
Open Nature of $\mathcal{V}$

$$\hat{\Theta}_{MAP} = \arg\max_{\Theta} P(\Theta|X, \mathcal{V}) \propto \arg\max_{\Theta} \underbrace{\log P(X|\Theta, \mathcal{V})}_{likelihood} + \underbrace{\log P(\mathcal{V}|\Theta)}_{language\ likelihood} + \underbrace{\log P(\Theta)}_{prior}$$

OVS Training

Motivation and Evaluation

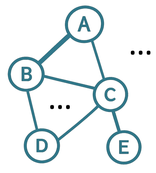
- From pixel-wise to mask-wise
- Retain all predictions, not just the highest-probability class

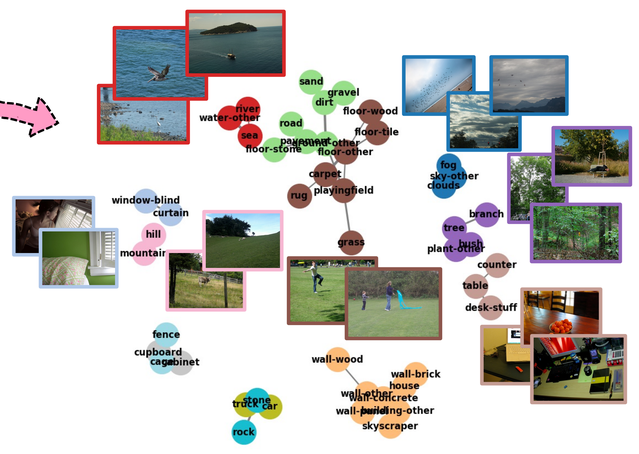


Mask-wise Evaluation Protocol

[chair, other] | [chair, ceiling, other] | [chair, ceiling, wall, curtain,other] | [chair, ceiling, wall, curtain,table,other] | [chair, ceiling, wall, curtain,table,floor,other]

Vocabulary co-occurring relationship





Ambiguous Vocabulary Graph

Results and Analysis

❑ Re-benchmarking

Method	Venue	PC59			ADE150			PC459			ADE847		
SAN	CVPR'23	57.70			32.10			15.70			12.40		
CAT-Seg	CVPR'24	63.30			37.90			23.80			16.00		
SED	CVPR'24	60.90			35.30			22.10			13.70		
MAFT+	ECCV'24	59.40			36.10			21.60			15.10		
		front↑	back↑	err↓	front↑	back↑	err↓	front↑	back↑	err↓	front↑	back↑	err↓
SAN	CVPR'23	65.91	93.75	9.99	42.89	93.12	8.56	27.65	70.87	6.67	22.84	92.46	8.41
CAT-Seg	CVPR'24	68.46	94.24	Null	45.74	94.61	5.53	30.95	68.96	3.86	26.39	93.66	5.20
SED	CVPR'24	66.29	94.21	6.43	44.90	93.50	5.20	31.41	70.72	4.93	26.99	92.61	5.07
MAFT+	ECCV'24	64.95	93.57	9.10	46.51	93.10	7.31	31.89	70.82	7.12	28.72	92.15	7.84
SAN⋆	CVPR'23	64.32	91.83	10.99	42.18	91.50	8.32	27.85	69.06	6.20	21.01	91.04	5.10
CAT-Seg⋆	CVPR'24	66.35	92.24	2.19	50.04	92.68	2.30	11.56	67.32	2.00	12.83	91.20	2.20
SED⋆	CVPR'24	63.35	91.32	5.31	42.65	91.28	4.52	30.04	68.40	3.23	27.45	90.05	4.10
MAFT+⋆	ECCV'24	62.05	91.55	8.56	44.30	91.32	6.70	29.04	69.01	4.40	26.01	90.50	6.40

❑ Vocabulary Removal

Method	PC59			ADE150			PC459			ADE847		
	front↑	back↑	error↓	front↑	back↑	error↓	front↑	back↑	error↓	front↑	back↑	error↓
MAFT+	-0.87	-0.23	-1.98	-1.56	-0.63	-2.30	-1.12	-0.08	-1.24	-1.81	-1.33	-1.96
SED	+2.03	+0.61	-2.15	+5.28	+0.46	-0.21	+0.50	-0.04	-1.57	+1.51	-0.62	-2.00
SED w/ 0.7	+1.10	+0.40	-1.80	+3.40	+0.30	-0.70	+0.25	-0.02	-1.50	+1.00	-0.30	-1.80
SED w/ 0.5	+1.40	+0.50	-1.70	+3.90	+0.40	-0.60	+0.35	-0.01	-1.40	+1.30	-0.40	-1.70
SED w/ 0.3	+1.70	+0.55	-1.60	+4.20	+0.42	-0.50	+0.40	+0.01	-1.30	+1.40	-0.50	-1.60
SED w/ 0.1	+2.00	+0.60	-1.50	+4.50	+0.45	-0.40	+0.45	+0.02	-1.20	+1.50	-0.55	-1.50