

VII. SUPPLEMENTARY MATERIAL

In addition to the UMD dataset we evaluated our method on the the AGD20K dataset.

1) *AGD20K*: The AGD20K dataset is intended for affordance grounding tasks and includes over 20,000 images across 36 affordance categories. It utilizes three evaluation metrics: Kullback-Leibler Divergence (KLD), similarity metric (SIM), and Normalized Scanpath Saliency (NSS). Detailed information on their design and implementation can be found in the supplementary material of Luo et al.'s paper [3]. For the performance of other models, we referenced the data from the paper by Li et al. [23].

Our method has higher KLD than the best seen model by 9.429 points and the unseen by 7.427. Our SIM and NSS scores were competitive with other models beating several models and only being surpassed by LOCATE [23] for both seen and unseen categories, suggesting our segmentation closely aligns with the ground truth. The high KLD stems from our method producing a binary mask for the affordance area, in contrast to the ground truth comparison which uses a value distribution, the peak of which indicates the affordance region. Consequently, our non-distributional output incurs a significant penalty in KLD calculation.

TABLE IV: AGD20K Dataset Evaluation

Method	Seen			Unseen		
	KLD↓	SIM↑	NSS↑	KLD↓	SIM↑	NSS↑
EIL [37]	1.931	0.285	0.522	2.167	0.227	0.330
SPA [38]	5.528	0.221	0.357	7.425	0.169	0.262
TS-CAM [39]	1.842	0.260	0.336	2.104	0.201	0.151
Hotspots [19]	1.773	0.278	0.615	1.994	0.237	0.577
Cross-view-AG [3]	1.538	0.334	0.927	1.787	0.285	0.829
Cross-view-AG+ [40]	1.489	0.342	0.981	1.765	0.279	0.882
AffCorrs† [41]	1.407	0.359	1.026	1.618	0.348	1.021
LOCATE [23]	1.226	0.401	1.177	1.405	0.372	1.157
OVAL-Prompt(Ours)	10.649	0.339	1.044	8.832	0.365	0.925