

Edge-Aware Self-Attention in Vision Transformers for Food Image Segmentation

Supplementary Material

7. Results in Supplementary

This section provides additional ablations and raw values corresponding to Fig. 3.

Table 5. Ablation on EdgeAttn variants under 4-/8-connected neighborhoods.

Connectivity	Method	Lr	Iters	GPUs	mIoU (%) ↑	mAcc (%) ↑
4-conn	SETR+EdgeAttn(QK)	1e-3	320000	1	44.91	56.03
4-conn	SETR+EdgeAttn(V)	1e-3	320000	1	45.88	57.85
4-conn	SETR+EdgeAttn(QKV)	1e-3	320000	1	44.42	56.20
8-conn	SETR+EdgeAttn(QK)	1e-3	320000	1	44.44	56.19
8-conn	SETR+EdgeAttn(V)	1e-3	320000	1	45.69	57.46
8-conn	SETR+EdgeAttn(QKV)	1e-3	320000	1	44.43	56.15

Table 6. Raw values corresponding to Fig. 3 for the baseline. The reported quantities are the boundary-alignment proxy $\text{align_cos_gt_edge_mean_layers}$, BF1@4, and Boundary IoU@4.

img	cls	align_cos	BF1@4	BIOU@4
1056.jpg	tea	0.135	0.405	0.277
1063.jpg	melon	0.136	0.328	0.208
1090.jpg	milk	0.142	0.545	0.331
131.jpg	tea	0.130	0.229	0.150
1339.jpg	milk	0.132	0.505	0.302
1353.jpg	milk	0.100	0.739	0.504
1388.jpg	milk	0.139	0.204	0.116
1437.jpg	melon	0.126	0.456	0.307
1465.jpg	melon	0.151	0.611	0.390
1598.jpg	tea	0.110	0.056	0.038
1617.jpg	melon	0.107	0.181	0.114
1639.jpg	milk	0.097	0.708	0.447
173.jpg	tea	0.107	0.577	0.338
1738.jpg	melon	0.141	0.431	0.265
1819.jpg	tea	0.147	0.104	0.060
2223.jpg	cashew	0.108	0.432	0.267
2404.jpg	cashew	0.130	0.398	0.231
2469.jpg	date	0.153	0.034	0.017
2634.jpg	date	0.104	0.307	0.169
354.jpg	date	0.107	0.081	0.044
4474.jpg	cashew	0.106	0.537	0.298
4500.jpg	cashew	0.147	0.205	0.116
4545.jpg	date	0.147	0.730	0.460
4558.jpg	cashew	0.128	0.622	0.400
4610.jpg	date	0.114	0.741	0.469

Table 7. Raw values corresponding to Fig. 3 for EdgeAttn(QK).

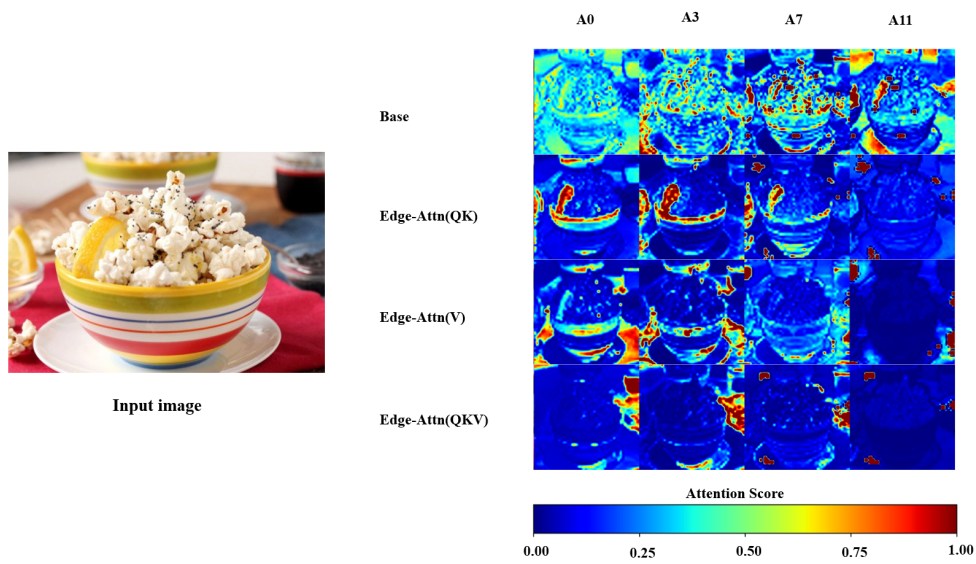
img	cls	align_cos	BF1@4	BloU@4
1056.jpg	tea	0.095	0.984	0.793
1063.jpg	melon	0.078	0.989	0.722
1090.jpg	milk	0.095	0.981	0.742
131.jpg	tea	0.069	0.990	0.845
1339.jpg	milk	0.080	0.991	0.785
1353.jpg	milk	0.073	0.947	0.656
1388.jpg	milk	0.104	0.617	0.387
1437.jpg	melon	0.080	0.984	0.815
1465.jpg	melon	0.082	0.975	0.770
1598.jpg	tea	0.039	0.990	0.820
1617.jpg	melon	0.036	0.768	0.549
1639.jpg	milk	0.101	0.987	0.788
173.jpg	tea	0.059	0.985	0.792
1738.jpg	melon	0.084	0.951	0.719
1819.jpg	tea	0.097	0.881	0.604
2223.jpg	cashew	0.068	0.936	0.701
2404.jpg	cashew	0.073	0.887	0.672
2469.jpg	date	0.107	0.383	0.224
2634.jpg	date	0.074	0.961	0.767
354.jpg	date	0.083	0.593	0.369
4474.jpg	cashew	0.094	0.921	0.635
4500.jpg	cashew	0.075	0.580	0.357
4545.jpg	date	0.051	0.989	0.830
4558.jpg	cashew	0.084	0.932	0.715
4610.jpg	date	0.060	0.948	0.738

Table 8. Raw values corresponding to Fig. 3 for EdgeAttn(QKV).

img	cls	align_cos	BF1@4	BIoU@4
1056.jpg	tea	0.028	0.971	0.793
1063.jpg	melon	0.023	0.988	0.761
1090.jpg	milk	0.072	0.976	0.775
131.jpg	tea	0.105	0.976	0.770
1339.jpg	milk	0.091	0.937	0.719
1353.jpg	milk	0.047	0.939	0.645
1388.jpg	milk	0.098	0.520	0.317
1437.jpg	melon	0.080	0.972	0.761
1465.jpg	melon	0.127	0.933	0.701
1598.jpg	tea	0.124	0.984	0.781
1617.jpg	melon	0.094	0.764	0.526
1639.jpg	milk	0.072	0.891	0.651
173.jpg	tea	0.066	0.931	0.702
1738.jpg	melon	0.097	0.938	0.691
1819.jpg	tea	0.123	0.817	0.538
2223.jpg	cashew	0.072	0.894	0.650
2404.jpg	cashew	0.043	0.851	0.622
2469.jpg	date	0.060	0.375	0.217
2634.jpg	date	0.089	0.907	0.682
354.jpg	date	0.067	0.402	0.227
4474.jpg	cashew	0.065	0.871	0.582
4500.jpg	cashew	0.058	0.518	0.315
4545.jpg	date	0.071	0.969	0.794
4558.jpg	cashew	0.073	0.885	0.656
4610.jpg	date	0.075	0.748	0.502

Table 9. Raw values corresponding to Fig. 3 for EdgeAttn(V).

img	cls	align_cos	BF1@4	BIoU@4
1056.jpg	tea	0.033	0.995	0.817
1063.jpg	melon	0.097	0.979	0.756
1090.jpg	milk	0.097	0.985	0.744
131.jpg	tea	0.089	0.980	0.811
1339.jpg	milk	0.106	0.995	0.802
1353.jpg	milk	0.099	0.970	0.703
1388.jpg	milk	0.104	0.537	0.340
1437.jpg	melon	0.093	0.989	0.804
1465.jpg	melon	0.082	0.943	0.763
1598.jpg	tea	0.105	0.982	0.775
1617.jpg	melon	0.082	0.713	0.458
1639.jpg	milk	0.115	0.993	0.786
173.jpg	tea	0.102	0.981	0.784
1738.jpg	melon	0.159	0.926	0.679
1819.jpg	tea	0.135	0.766	0.516
2223.jpg	cashew	0.093	0.911	0.657
2404.jpg	cashew	0.117	0.896	0.660
2469.jpg	date	0.054	0.356	0.207
2634.jpg	date	0.080	0.911	0.709
354.jpg	date	0.079	0.592	0.361
4474.jpg	cashew	0.081	0.844	0.550
4500.jpg	cashew	0.065	0.479	0.285
4545.jpg	date	0.107	0.988	0.817
4558.jpg	cashew	0.081	0.857	0.602
4610.jpg	date	0.121	0.933	0.735



(a) Attention maps (A0, A3, A7, A11).

Figure 4. Attention heatmaps for the “**popcorn**” category from FoodSeg103. The input image is shown on the left, and the attention responses for the baseline, EdgeAttn(QK), EdgeAttn(V), and EdgeAttn(QKV) are shown on the right. Attention maps from four transformer blocks (A0, A3, A7, and A11) are visualized to demonstrate how the boundary-aware attention modules affect feature aggregation and segmentation accuracy.

Table 10. Efficiency of EdgeAttn. Latency is measured on a single GPU with batch size 1 and averaged over 300 runs (warm-up excluded).

Model	Params (M)	Δ	Latency (ms)	Δ	FPS	Δ
SETR (Base)	93.196	0.000	86.016	0.000	11.626	0.000
SETR + EdgeAttn(QK)	93.197	+0.001	89.091	+3.075	11.224	-0.402
SETR + EdgeAttn(V)	93.197	+0.001	87.824	+1.808	11.386	-0.240
SETR + EdgeAttn(QKV)	93.197	+0.001	88.778	+2.762	11.264	-0.362