

	avg	ObjNet	LVIS	COCO	BDD	
CLIP alone	.72	.64	.63	.90	.74	all queries
+ CLIP alignment objective	.76	.69	.68	.92	.75	
+ DB alignment objective	.77	.70	.69	.92	.76	
+ multiscale representation	.80	.70	.76	.96	.79	
CLIP alone	.19	.28	.19	.27	.02	hard subset
+ CLIP alignment objective	.31	.39	.28	.51	.05	
+ DB alignment objective	.33	.40	.30	.55	.07	
+ multiscale representation	.46	.40	.44	.75	.24	