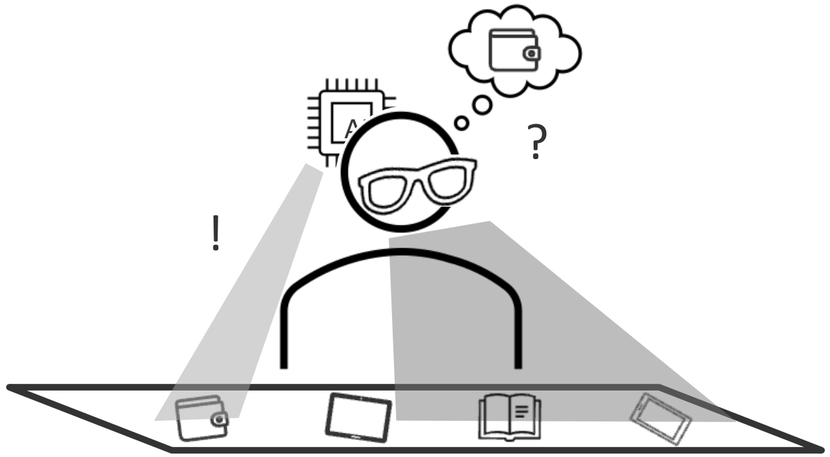
JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Where is my Wallet?
Modeling Object Proposal Sets for Egocentric Visual Query Localization

TAG: TUE-AM-247



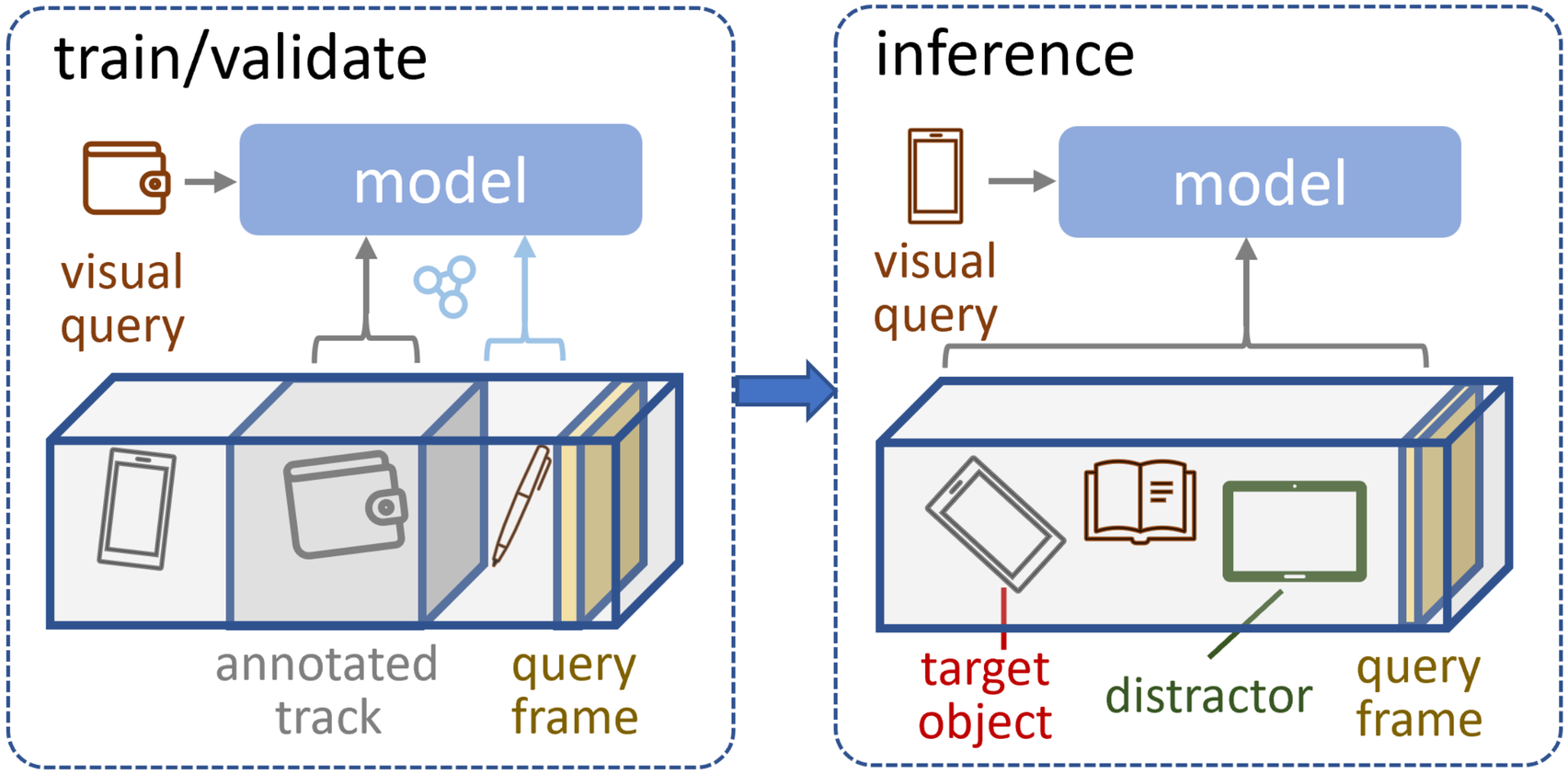
Mengmeng Xu · Yanghao Li · Cheng-Yang Fu · Bernard Ghanem · Tao Xiang · Juan-Manuel Pérez-Rúa

جامعة الملك عبدالله للعلوم والتقنية
King Abdullah University of Science and Technology

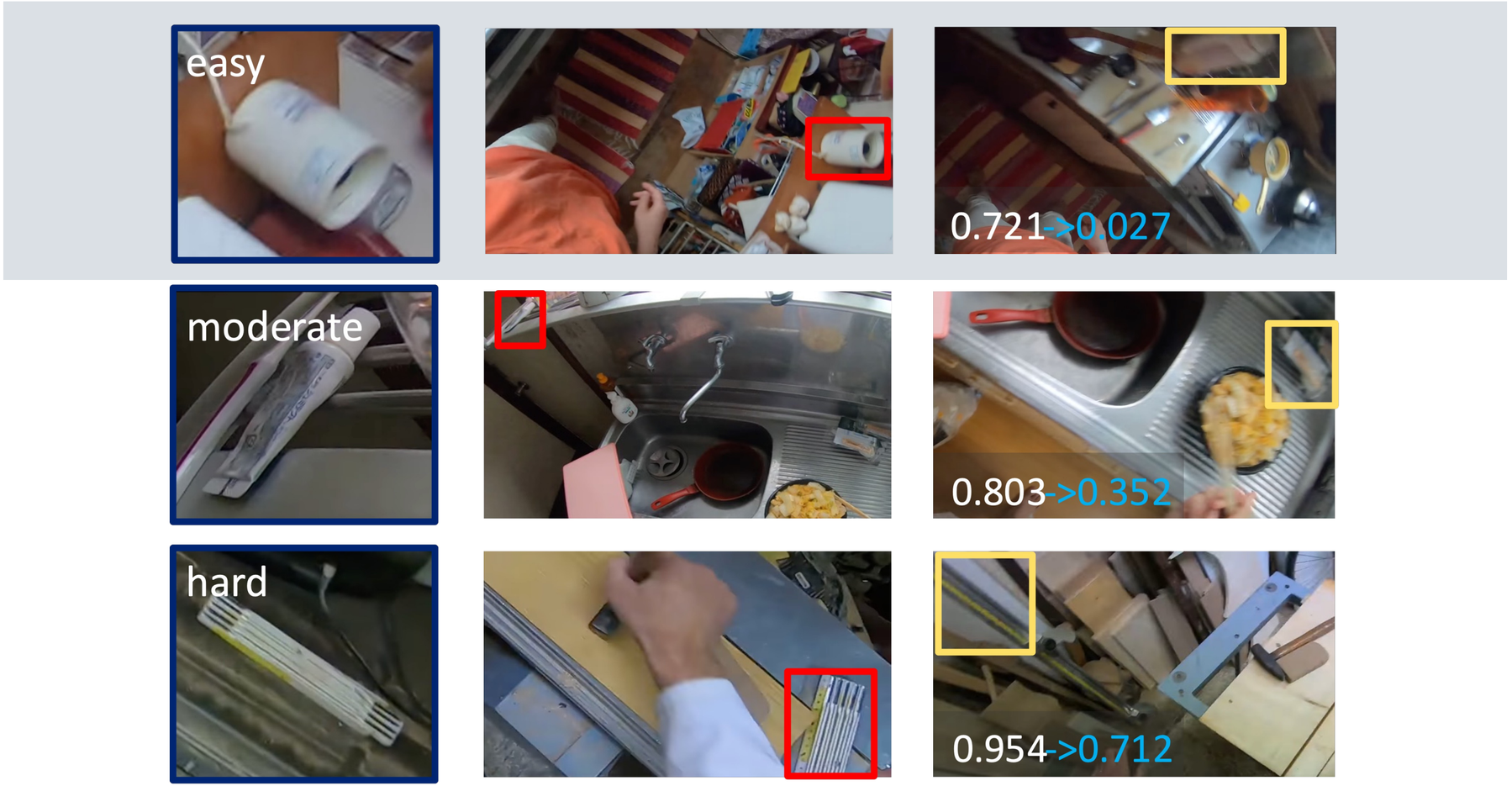
Meta AI

VQ bias = domain bias + task bias

Conditional detector with balanced samples



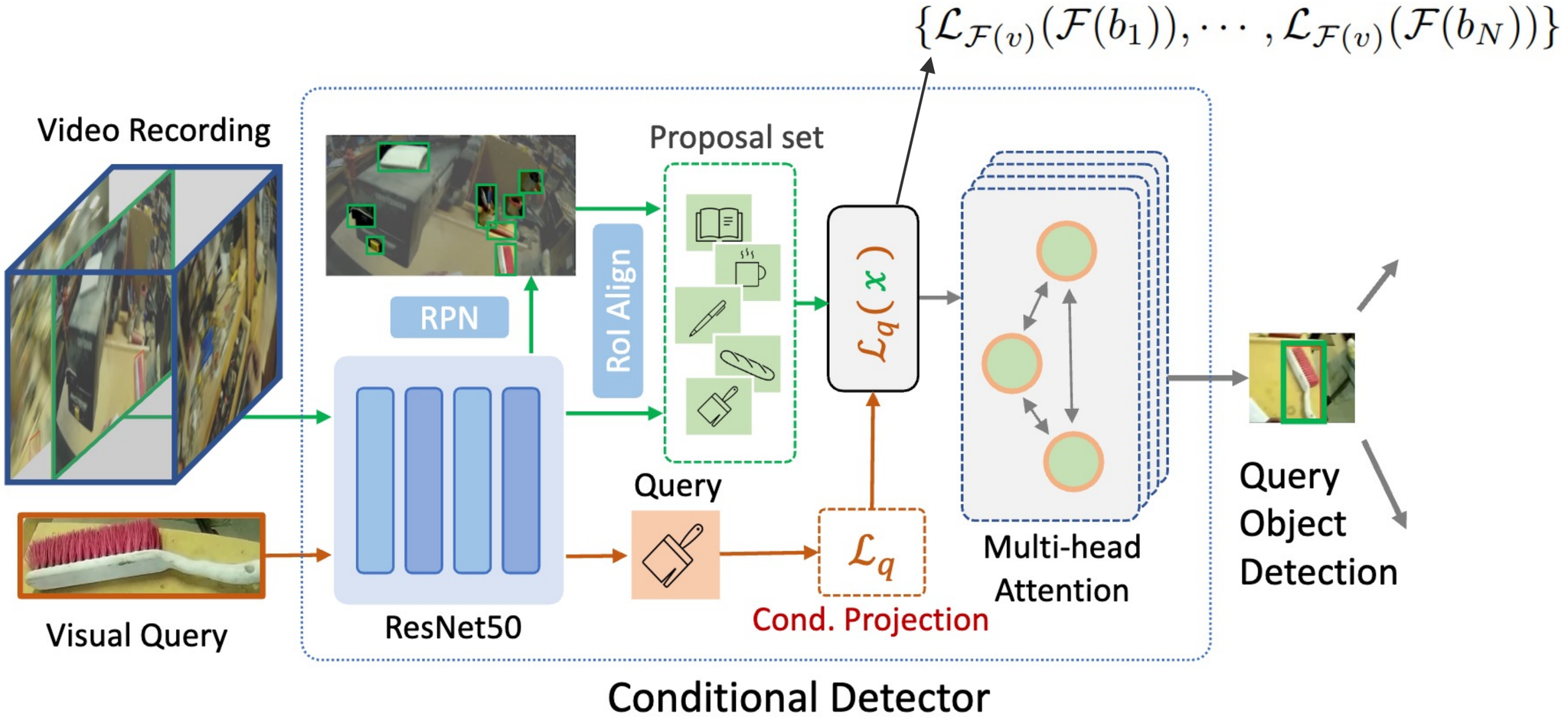
VQ bias = domain bias + task bias



VQ bias = domain bias + task bias

VQ bias = domain bias + task bias

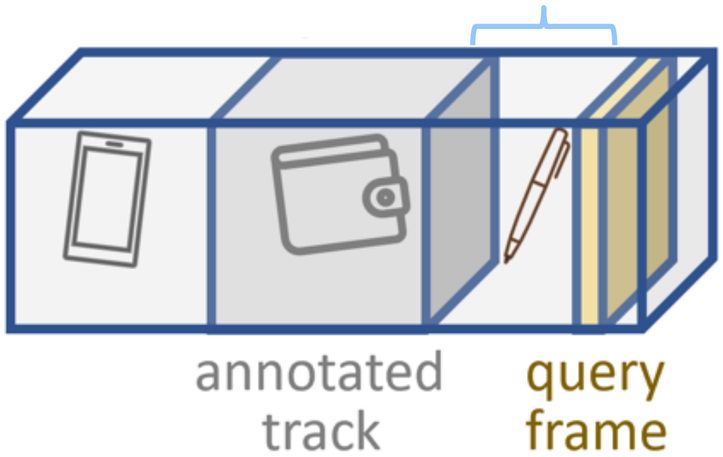
Main architecture



Training with balanced samples

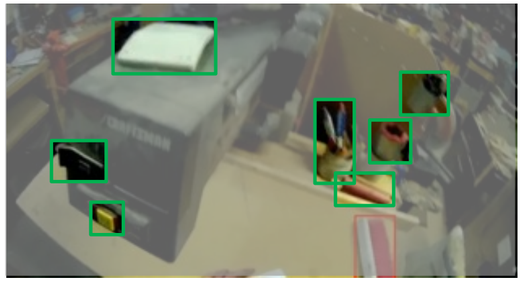
P-UFS

N-UFS



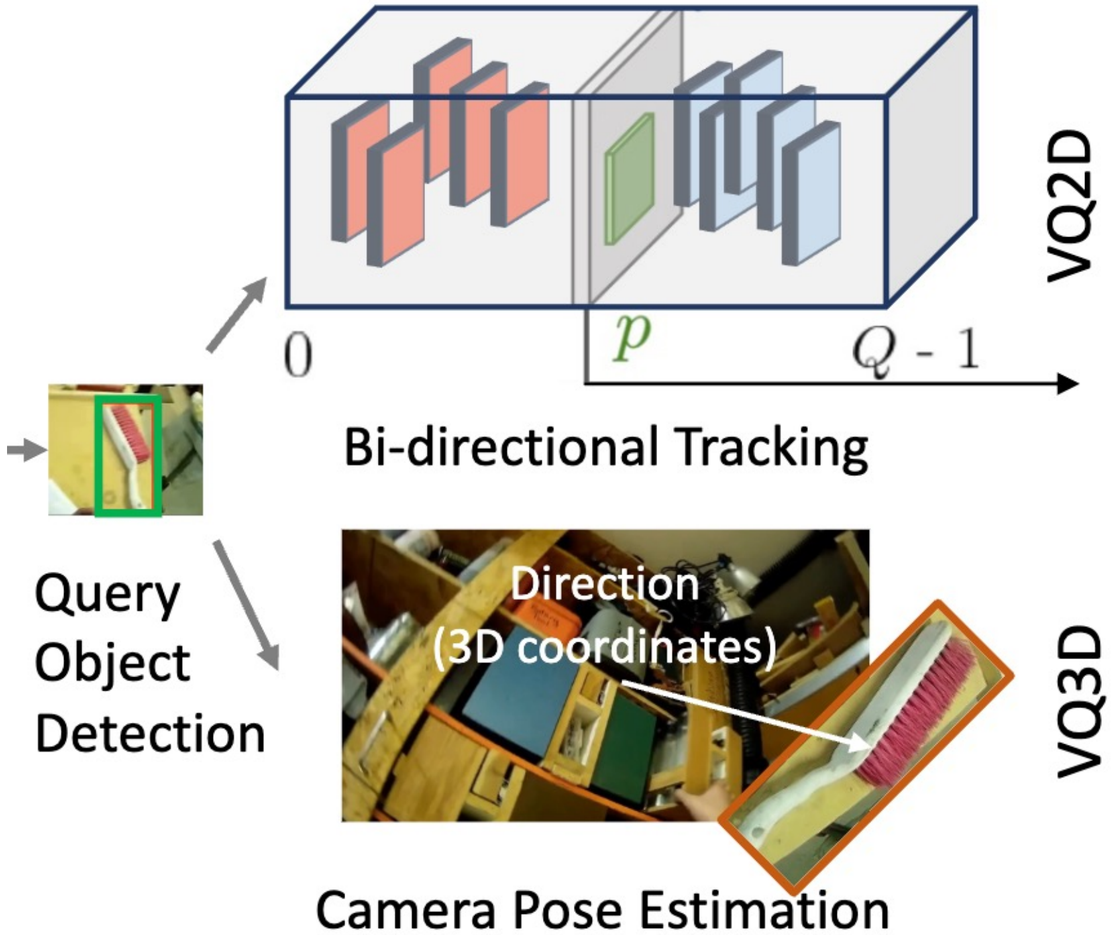
Balanced Proposal Set

positive proposal set



negative proposal set

Evaluated on VQ2D and VQ3D



Experimental Results

Compare with SOTA

method	AP	AP50	AP75
Baseline [5]	22.51	44.68	18.24
Xu. *etal* [9]	26.28	49.63	23.91
Our Solution	31.26	57.96	28.88
gain	(+18%)	(+17%)	(+21%)

Query object detection

method	tAP25	stAP25	rec%	Succ
Baseline [5]	0.20	0.12	32.2	39.8
Xu. *etal* [9]	0.22	0.15	35.29	43.07
Our Solution	**0.27**	**0.20**	**42.34**	**48.37**
gain	(+22%)	(+33%)	(+20%)	(+12%)

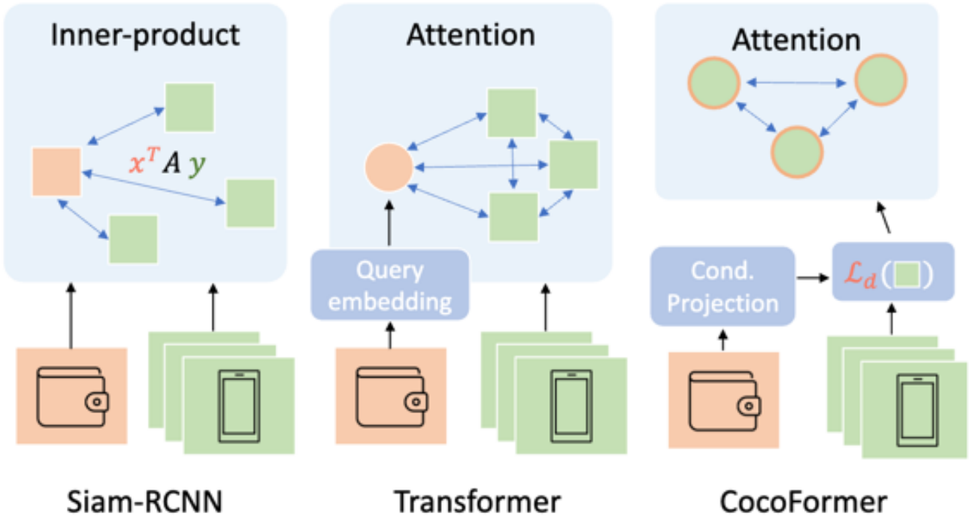
VQ2D

method	$L2 \downarrow$	$angle \downarrow$	$O.Succ \uparrow$	$Succ* \uparrow$
Siam-RCNN [24]	4.64	1.31	0.08	0.49
Our Method	**4.46**	**1.23**	**0.09**	**0.51**
gain	(+3.8%)	(+6.5%)	(+13%)	(+4.1%)

VQ3D

Experimental Results

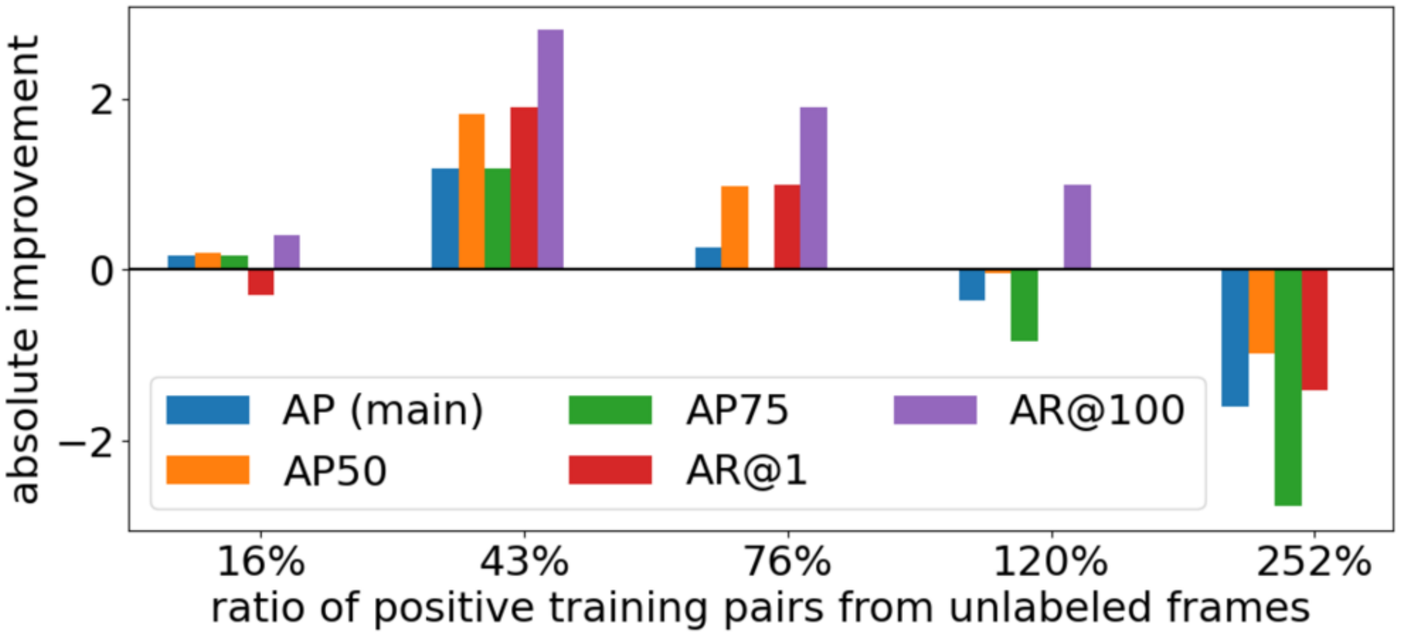
Compare with designs



model	query operation	visual query detection AP	AP50	AP75	AR@10	visual query 2D localisation tAP25	stAP25	rec%	Succ
SiamRCNN	Inner Prod.	28.74	52.25	27.35	50.1	0.253	0.190	40.29	47.70
Transformer	Self-Att.	17.46	37.61	13.01	39.7	0.148	0.007	26.86	30.18
Transformer	Cross-Att.	29.17	54.87	26.59	46.1	0.212	0.140	37.93	41.51
CocoFormer	Concat	**31.96**	**60.07**	28.77	**48.0**	0.257	0.191	42.15	47.81
CocoFormer	Cond. Proj.	31.26	57.96	**28.88**	47.1	**0.267**	**0.195**	**42.34**	**48.37**

Experimental Results

P-UFS: More noisy positives from novel views



Experimental Results

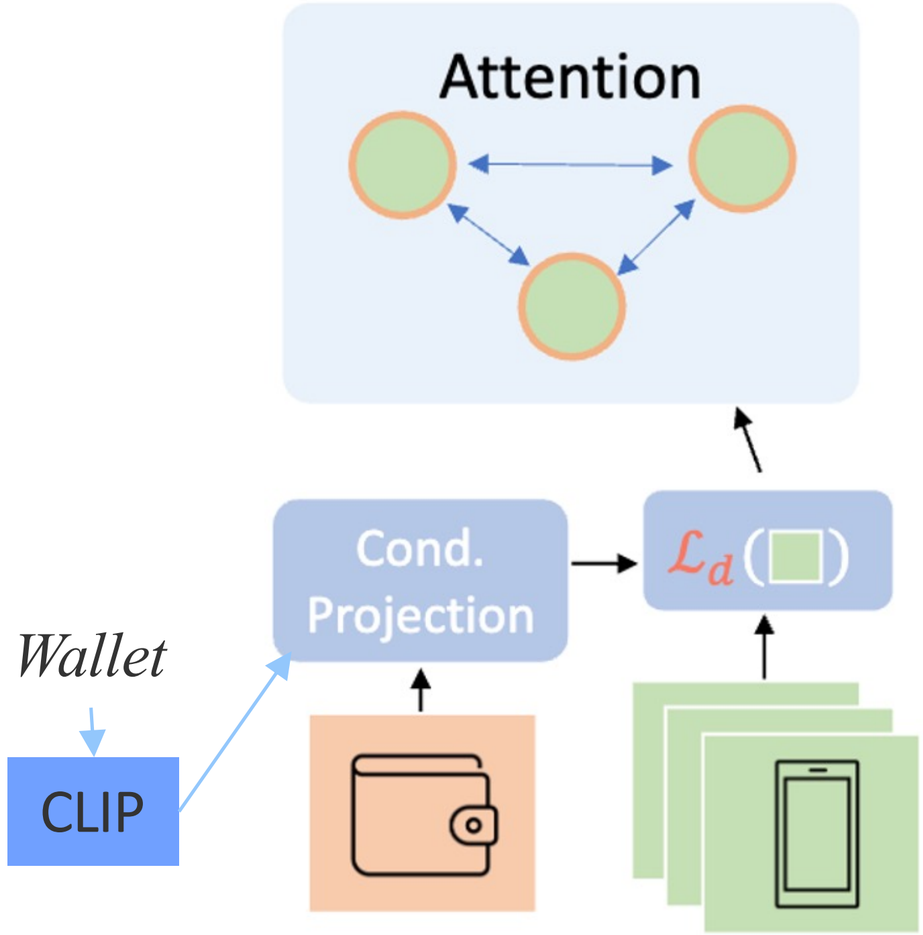
N-UFS&BPS: Negatives reduce task bias

N-UFS	BPS	VQ detection		VQ 2D localization		
		AP	AP_{50}	tAP_{25}	$stAP_{25}$	$Succ$
✗	✗	26.99	51.12	0.21	0.14	42.45
✓	✗	26.28	49.63	0.22	0.15	43.01
✓	✗	33.28	62.51	0.22	0.15	42.98
✓	✓	31.26	57.96	0.27	0.20	48.37

Experimental Results

Cond. Project: Multi-modal query

Use Text?	Query detection AP	Query detection AP50	VQ2D localization tAP25	VQ2D localization stAP25	VQ2D localization Succ.
False	31.26	57.96	0.257	0.195	48.37
True	**32.65**	**62.64**	**0.269**	**0.201**	**49.03**

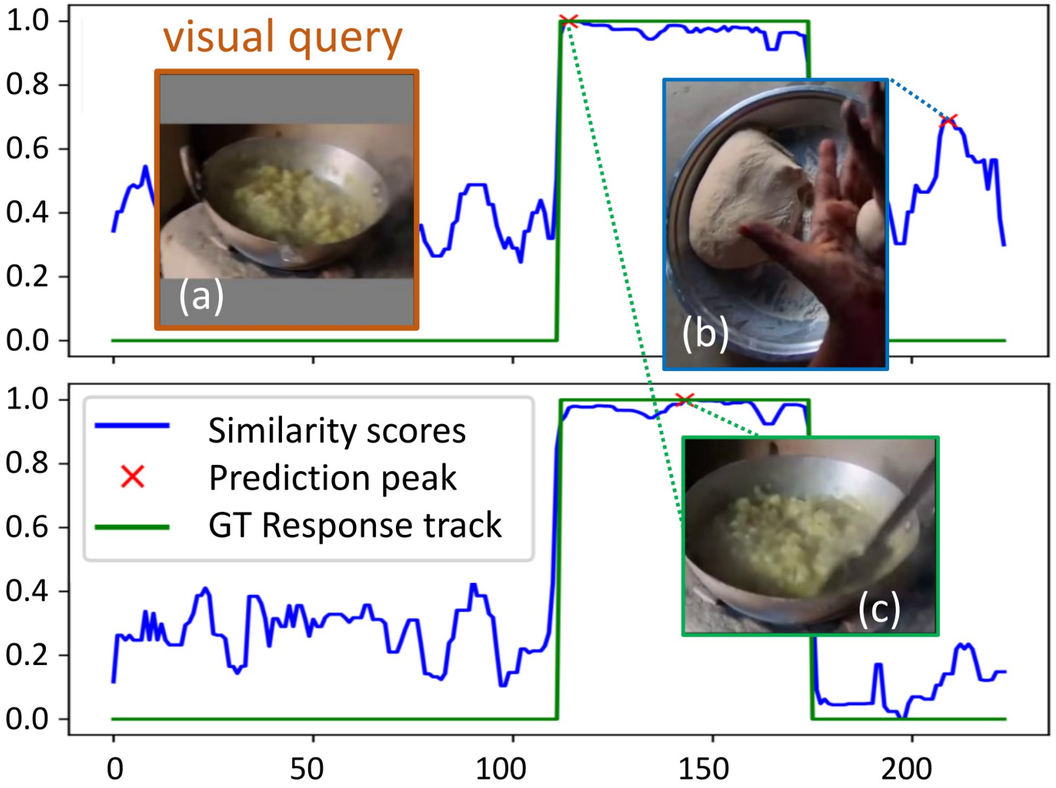


Experimental Results

FSD

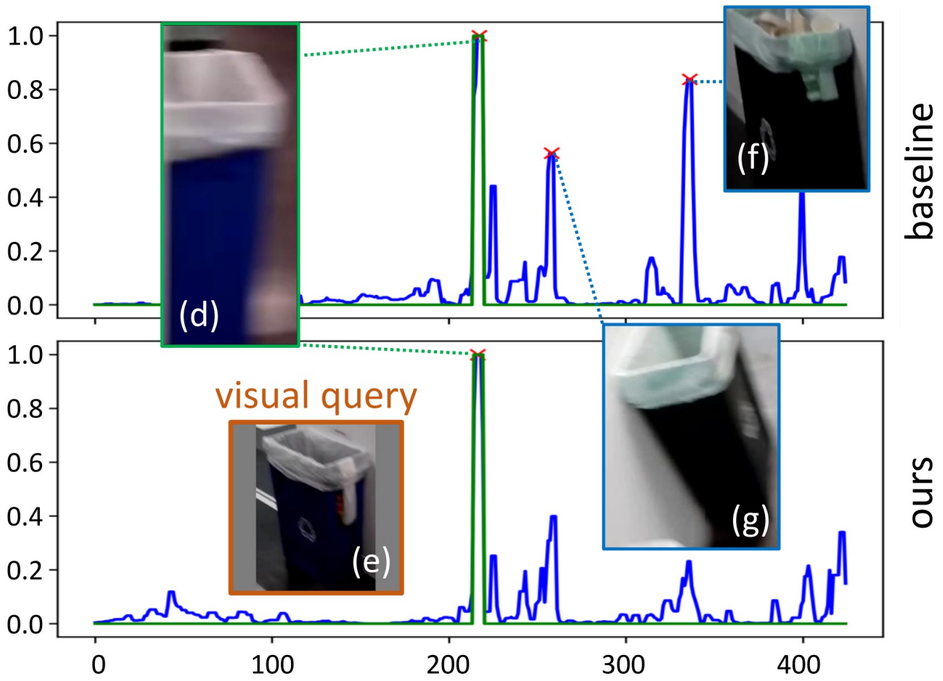
Method	novel ft.	1-shot nAP	1-shot AP50	1-shot AP75	3-shot nAP	3-shot AP50	3-shot AP75	5-shot nAP	5-shot AP50	5-shot AP75
TFA [15]	True	3.4	5.8	3.8	6.6	12.1	6.5	8.3	15.3	8.0
CoRPN [22]	True	4.1	7.2	4.4	-	-	-	-	-	-
Meta-DETR [20]	True	7.5	12.5	7.7	-	-	-	-	-	-
FADI [4]	True	5.7	10.4	6.0	-	-	-	-	-	-
Xiao *et al.* [18]	True	3.2	8.9	1.4	6.7	18.6	2.9	8.1	20.1	4.4
MPSR [16] †	True	2.3	4.1	2.3	5.2	9.5	5.1	6.7	12.6	6.4
Fan *et al.* [6] †	True	4.2	9.1	3.0	6.6	15.9	4.9	8.0	18.5	6.3
Zhang *et al.* [21]	True	4.4	7.5	4.9	7.2	13.3	7.4	-	-	-
QA-FewDet [7]	True	4.9	10.3	4.4	8.4	18.0	7.3	9.7	20.3	8.6
DeFRCN [13]	True	9.3	-	-	14.8	-	-	16.1	-	-
Fan *et al.* [6] †	False	4.0	8.5	3.5	5.9	12.5	5.0	6.9	14.3	6.0
Meta Faster-RCNN [8]	False	5.0	10.5	4.5	-	-	-	-	-	-
QA-FewDet [7]	False	5.1	10.5	4.5	8.6	17.7	7.5	9.5	19.3	8.5
FS-DETR [2]	False	7.0	13.6	7.5	9.8	18.5	9.8	10.7	20.5	10.8
DAnA [5]	False	11.9	**25.6**	10.4	14.0	**28.9**	12.3	14.4	**30.4**	13.0
hANMCL [12]	False	12.9	25.0	12.1	14.4	28.0	13.3	14.5	27.9	13.3
ours	False	**13.3**	**25.6**	**12.6**	**14.7**	28.8	**13.4**	**14.8**	28.9	**13.6**

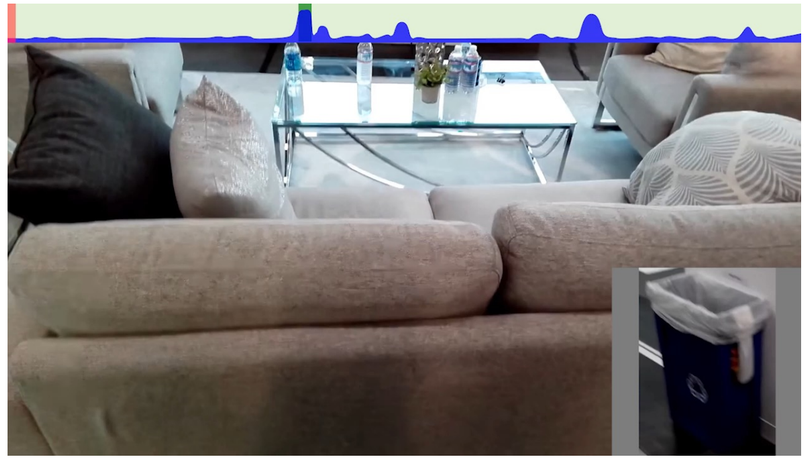
Visualization: Frying Pan (easy)

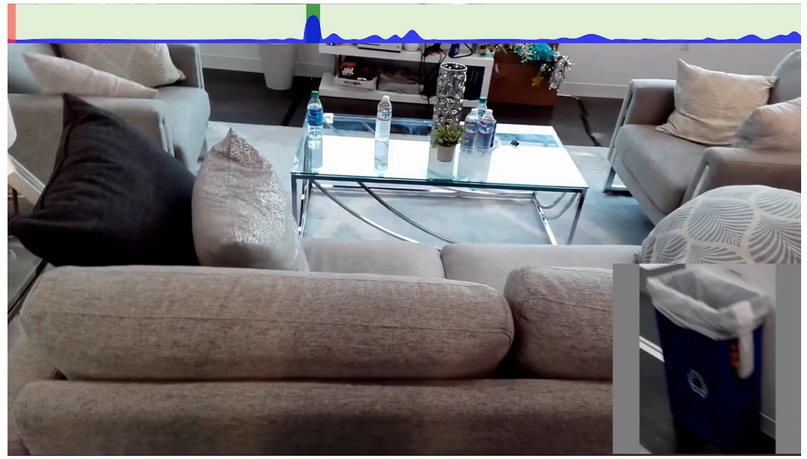


Visualization: Blue Bin (hard)

Hard negatives

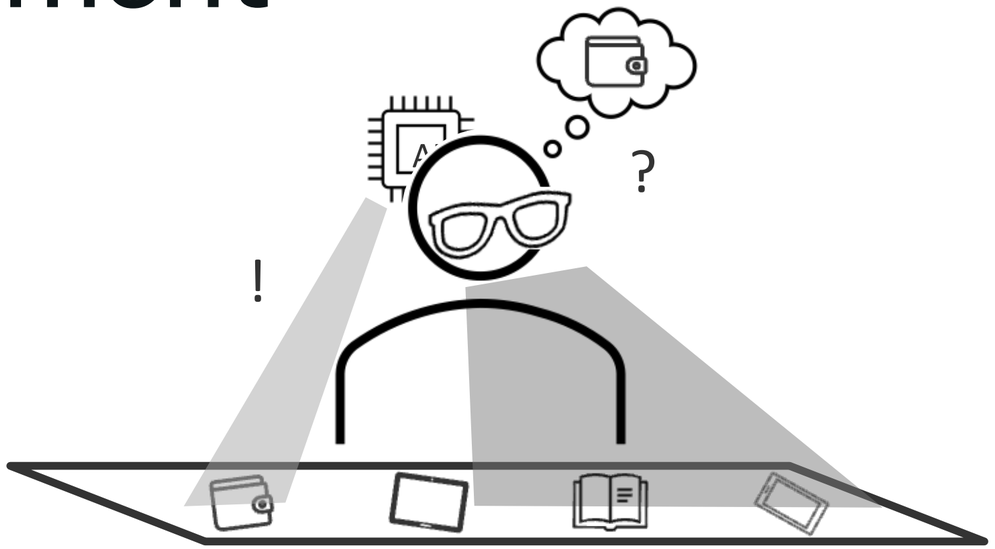






Acknowledgement

TAG: TUE-AM-247

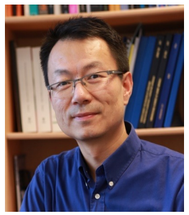



Mengmeng Xu


Yanghao Li


Cheng-Yang Fu


Bernard Ghanem


Tao Xiang


Juan-Manuel Pérez-Rúa


Meta AI