

Interaction Compass: Multi-Label Zero-Shot Learning of Human-Object Interactions via Spatial Relations



Dat Huynh

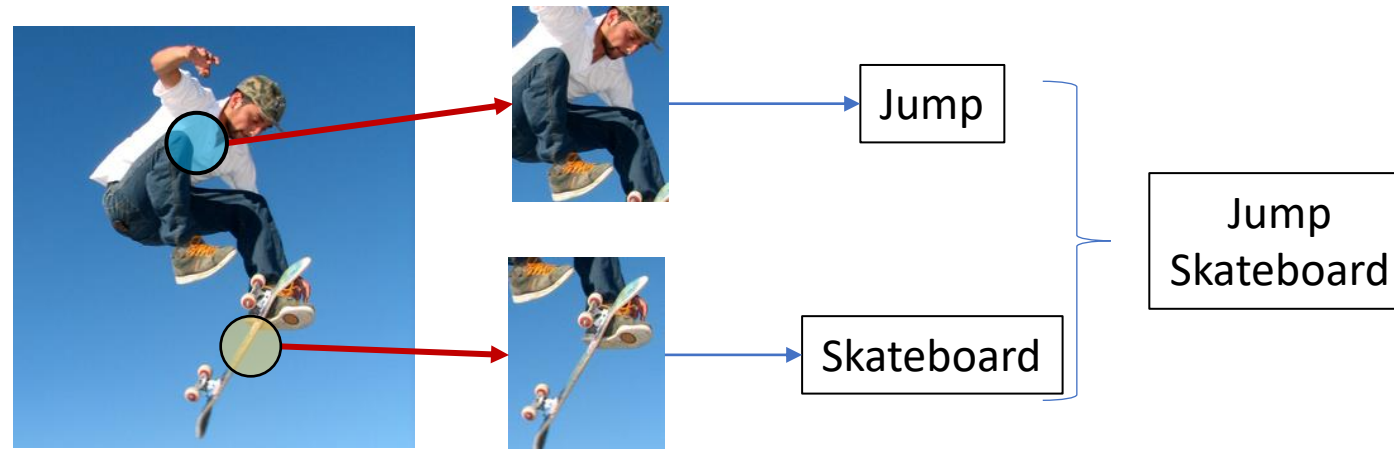


Ehsan Elhamifar

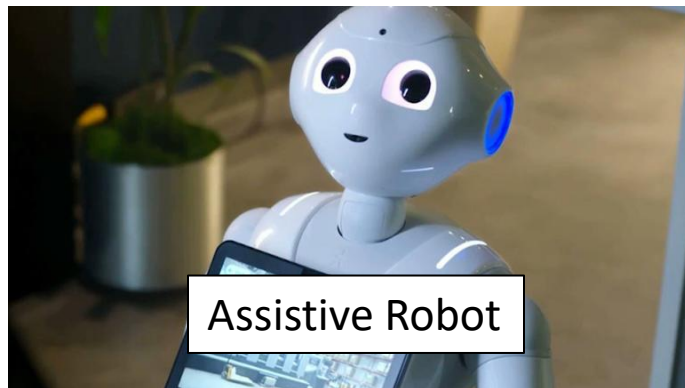
Khoury College of Computer Sciences
Northeastern University

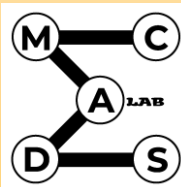


- Human-Object Interaction:
 - Classify/Localize *what actions* are performed on *what objects*



- Applications:



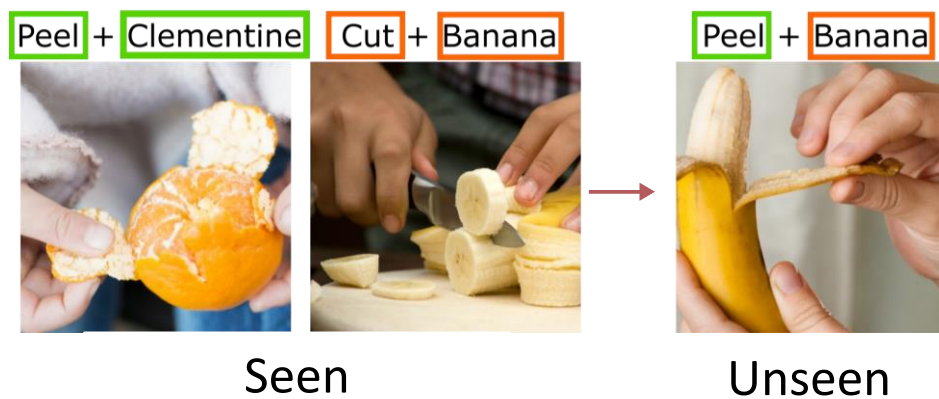


Contributions

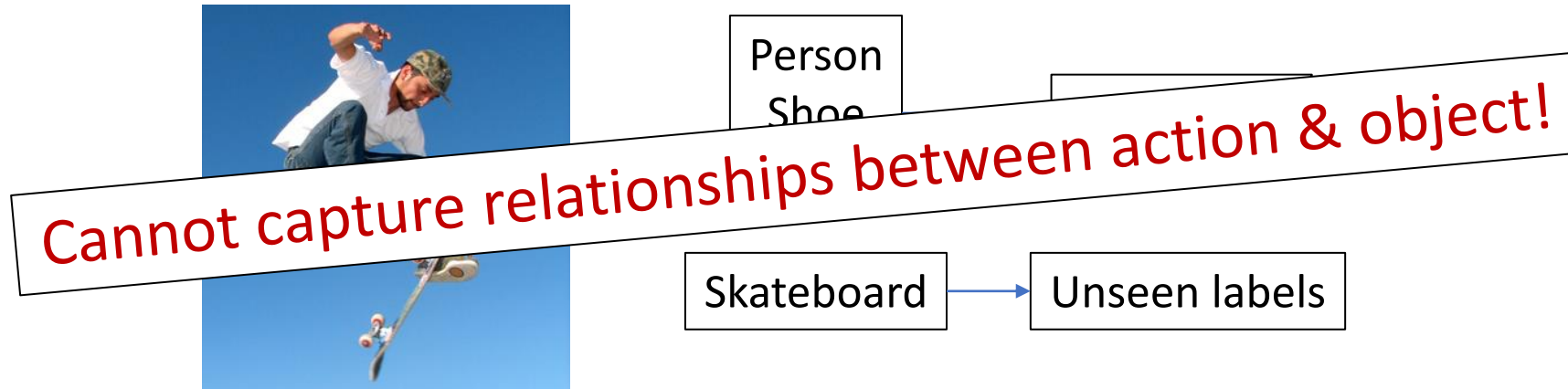
- Combinatorically larger number of possible human-object interactions



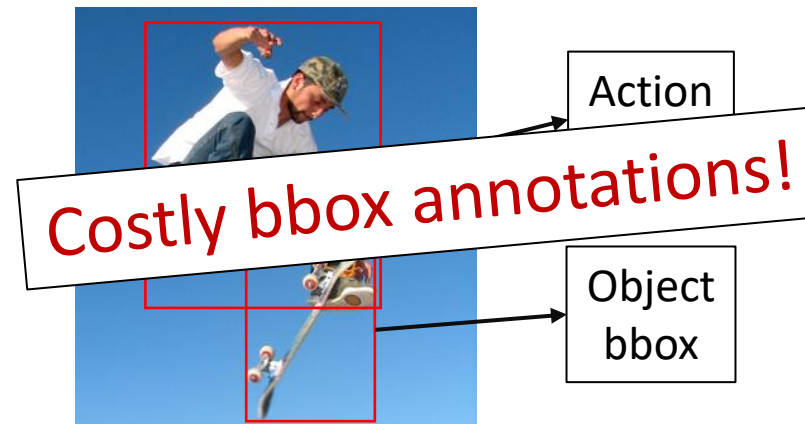
- Zero-Shot Human-Object Interaction
 - Recognize unseen interactions as novel combination of seen actions and objects



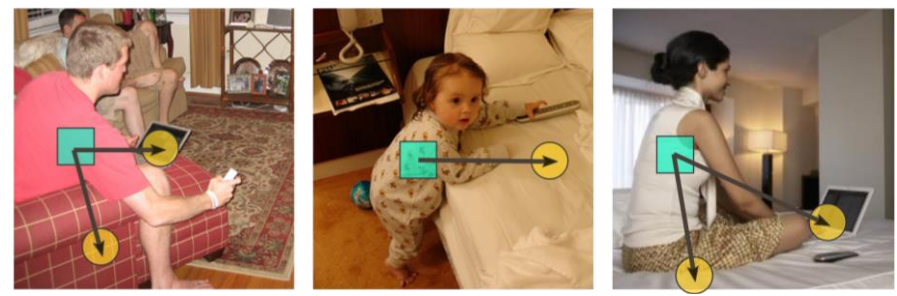
- (Multi-Label) Zero-Shot Learning [Y. Zhang et. al. '16, C. W. Lee et. al. '18, D. Huynh et. al. '20]



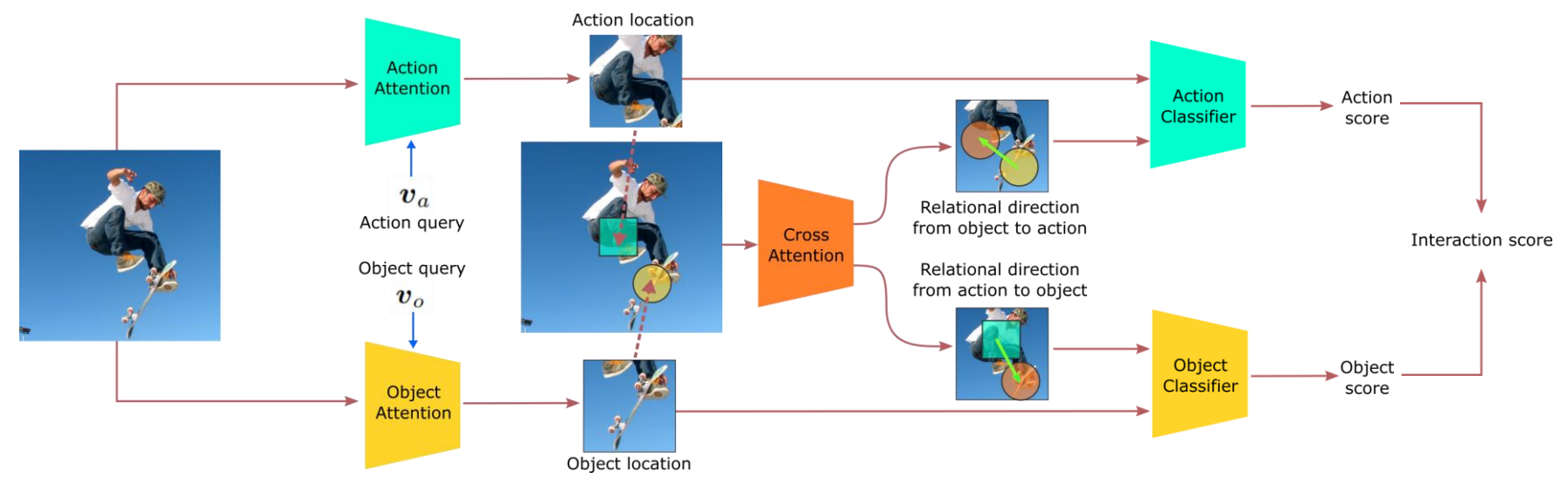
- Human-Object Interaction Detection [Y. W. Chao et. al. '18, Y. Li et. al. 19', C. Gao et. al. 20']

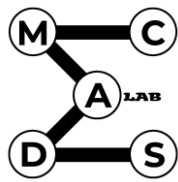


- Relational Direction:
 - *predictable displacement between action and object*

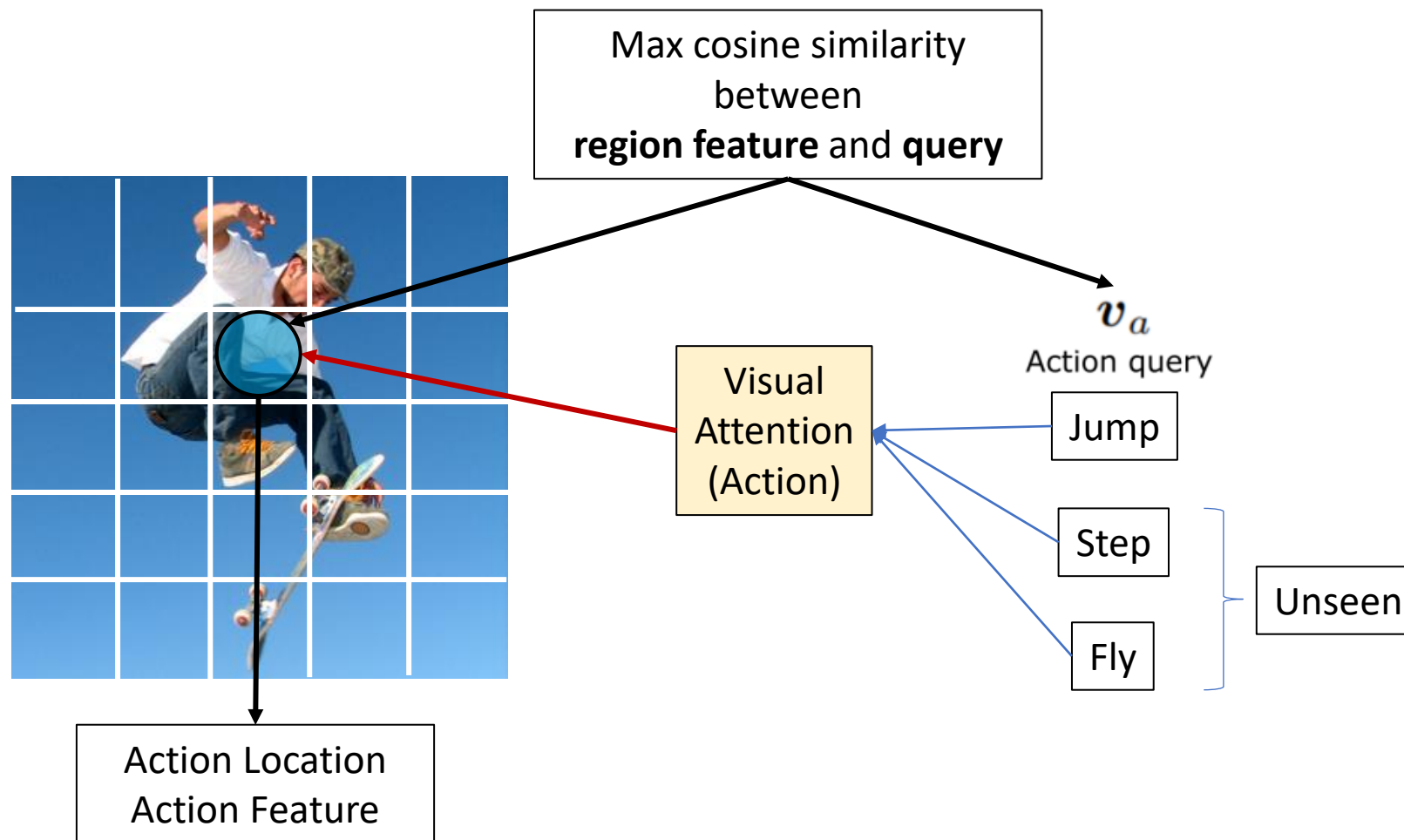


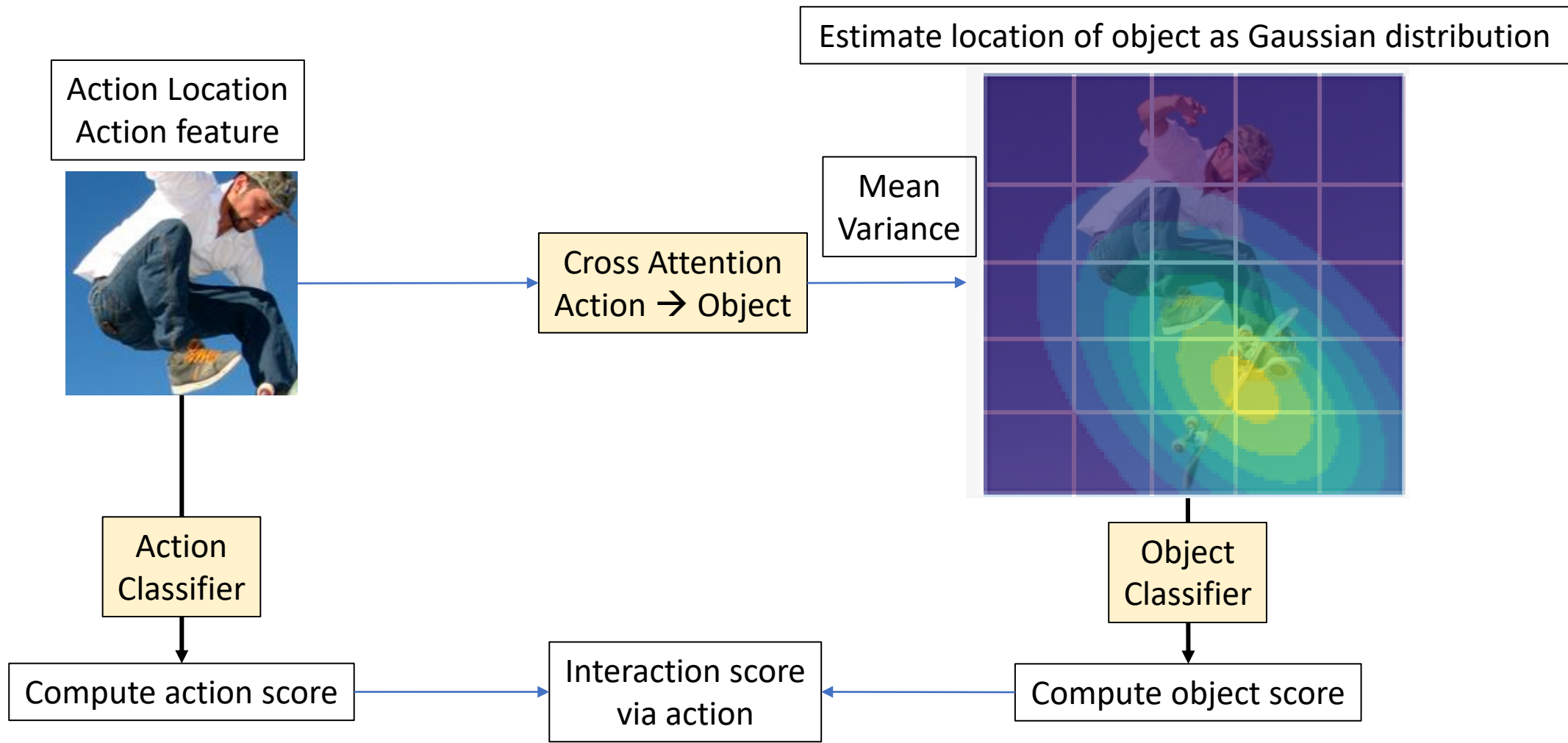
- Overview:





Visual Attention





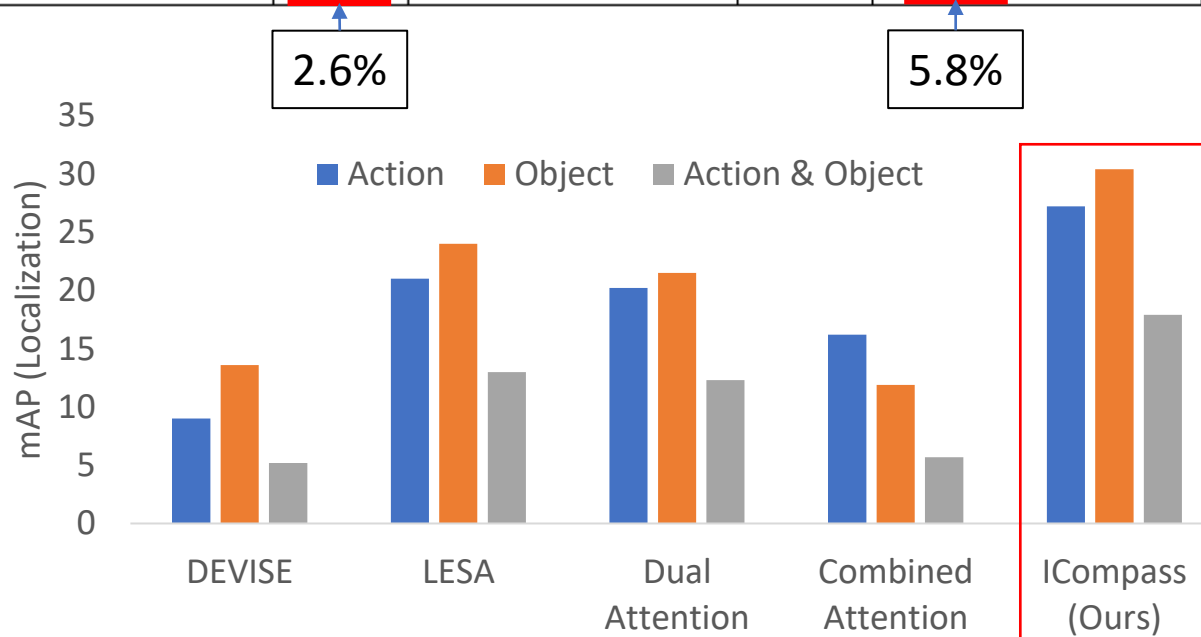


Experimental Results

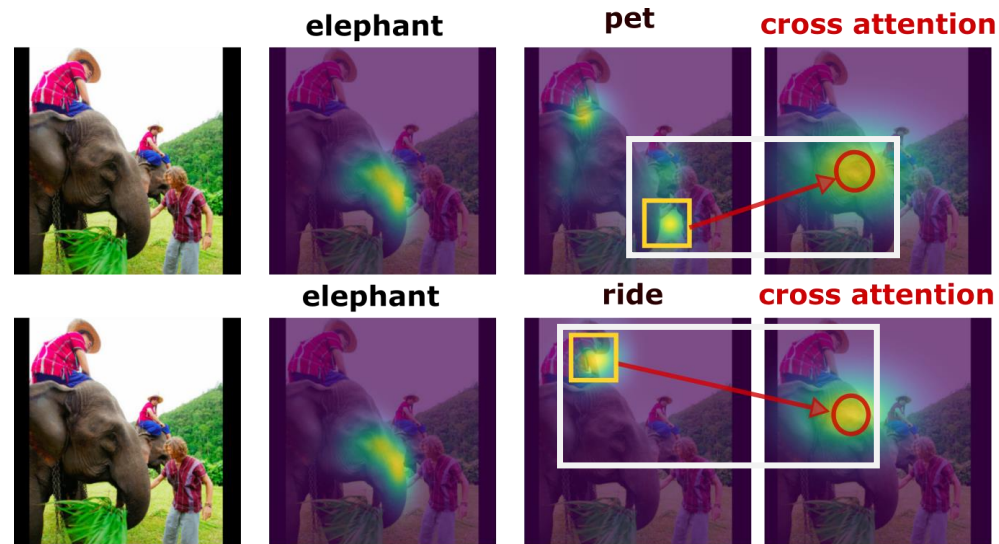
- Recognition

Method	HICO						Visual Genome					
	Unseen			All			Unseen			All		
	R@10	F1@10	mAP	R@10	F1@10	mAP	R@10	F1@10	mAP	R@10	F1@10	mAP
DEWISE	54.4	8.8	10.7	59.0	19.5	16.9	35.1	3.9	3.0	31.5	6.8	3.7
Fast0Tag	<u>76.8</u>	<u>12.5</u>	<u>19.9</u>	<u>75.8</u>	<u>25.1</u>	26.2	49.1	5.4	<u>7.1</u>	43.9	9.5	<u>8.0</u>
LESA	71.1	11.5	21.8	75.0	24.8	<u>28.3</u>	41.6	4.6	6.9	43.2	9.3	8.3
Dual Attention	71.2	11.6	19.1	73.9	24.5	25.8	51.2	<u>5.7</u>	6.1	<u>47.8</u>	<u>10.3</u>	7.5
Combined Attention	71.1	11.5	14.3	72.0	23.8	22.1	41.9	4.6	4.9	40.8	8.8	5.9
ICompass (Ours)	82.7	13.4	24.4	81.6	27.0	30.4	57.0	6.3	7.8	52.1	11.2	8.8

- Localization

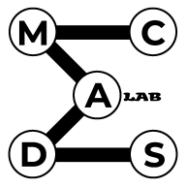


- Unseen object-action combination:



- Unseen action:





Code



Code is available at:

https://github.com/hbdat/iccv21_relational_direction

