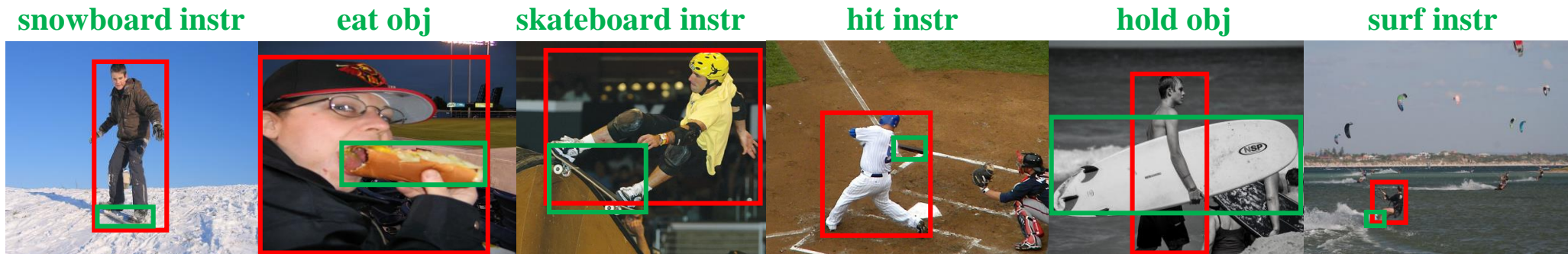


Human-Object Interaction Detection with Missing Objects

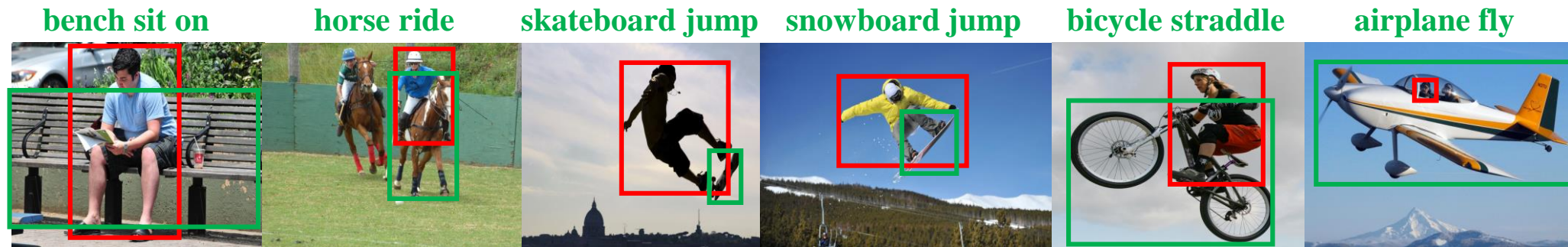
Kaen Kogashi, Yang Wu, Shohei Nobuhara, Ko Nishino
Kyoto University



What is Human-Object Interaction (HOI) Detection?



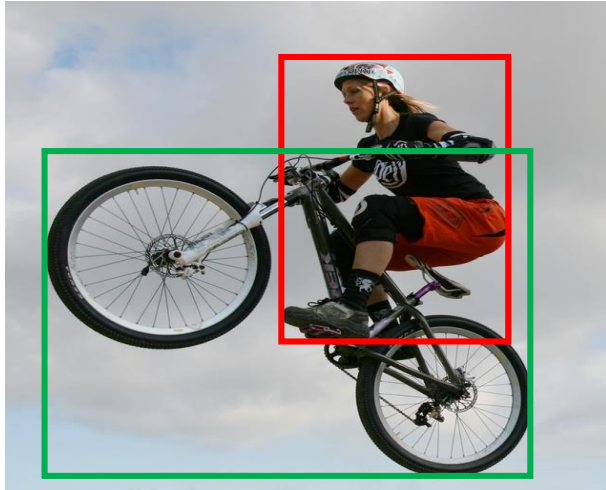
V-COCO test set examples



HICO-DET test set examples

Conventional HOI detection V/S. HOI detection with Missing objects

Conventional HOI



➡ Human and object pairs can be detected perfectly.

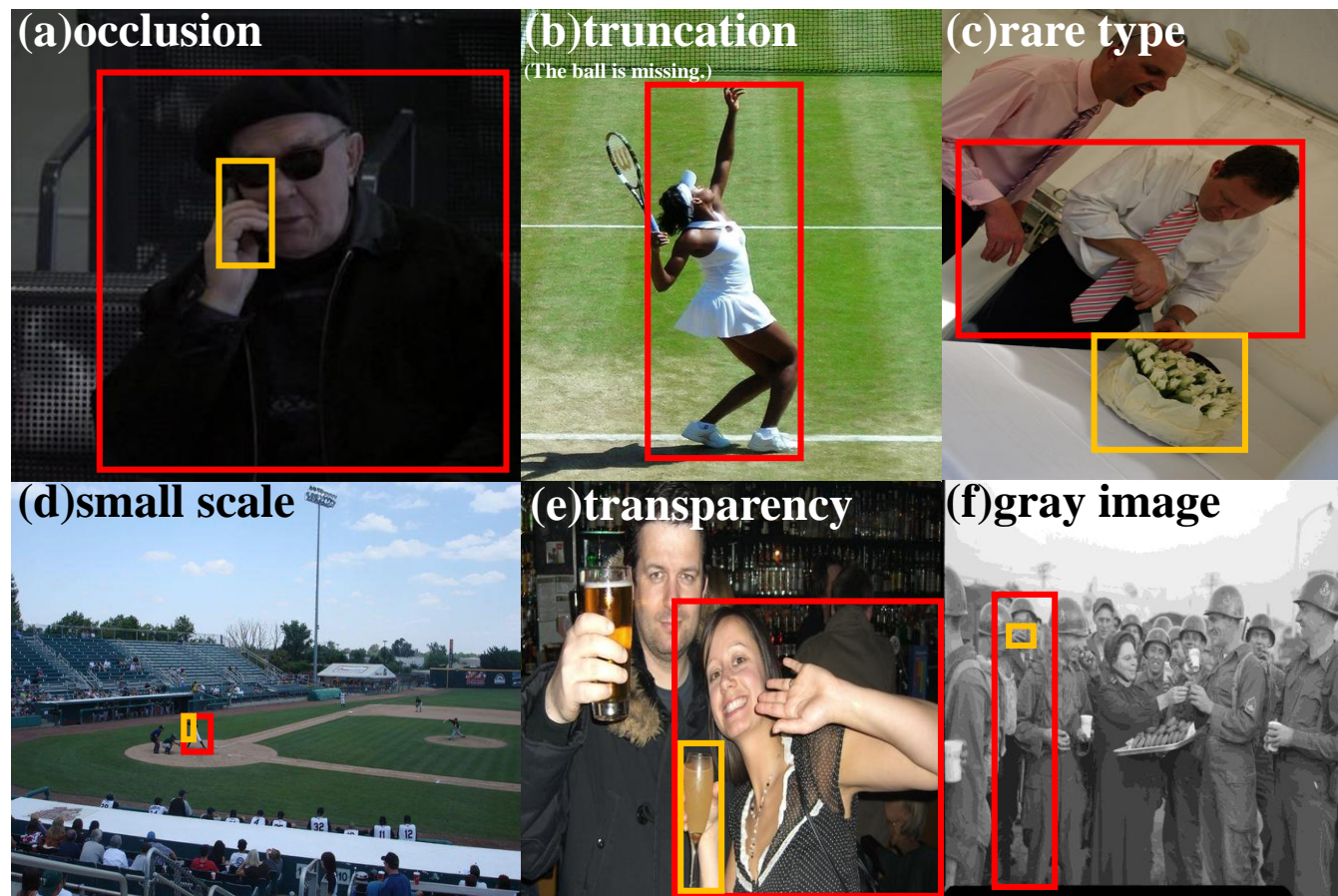
HOI with Missing Objects



➡ Since the real world is noisy, there are many cases where object detection does not work well.

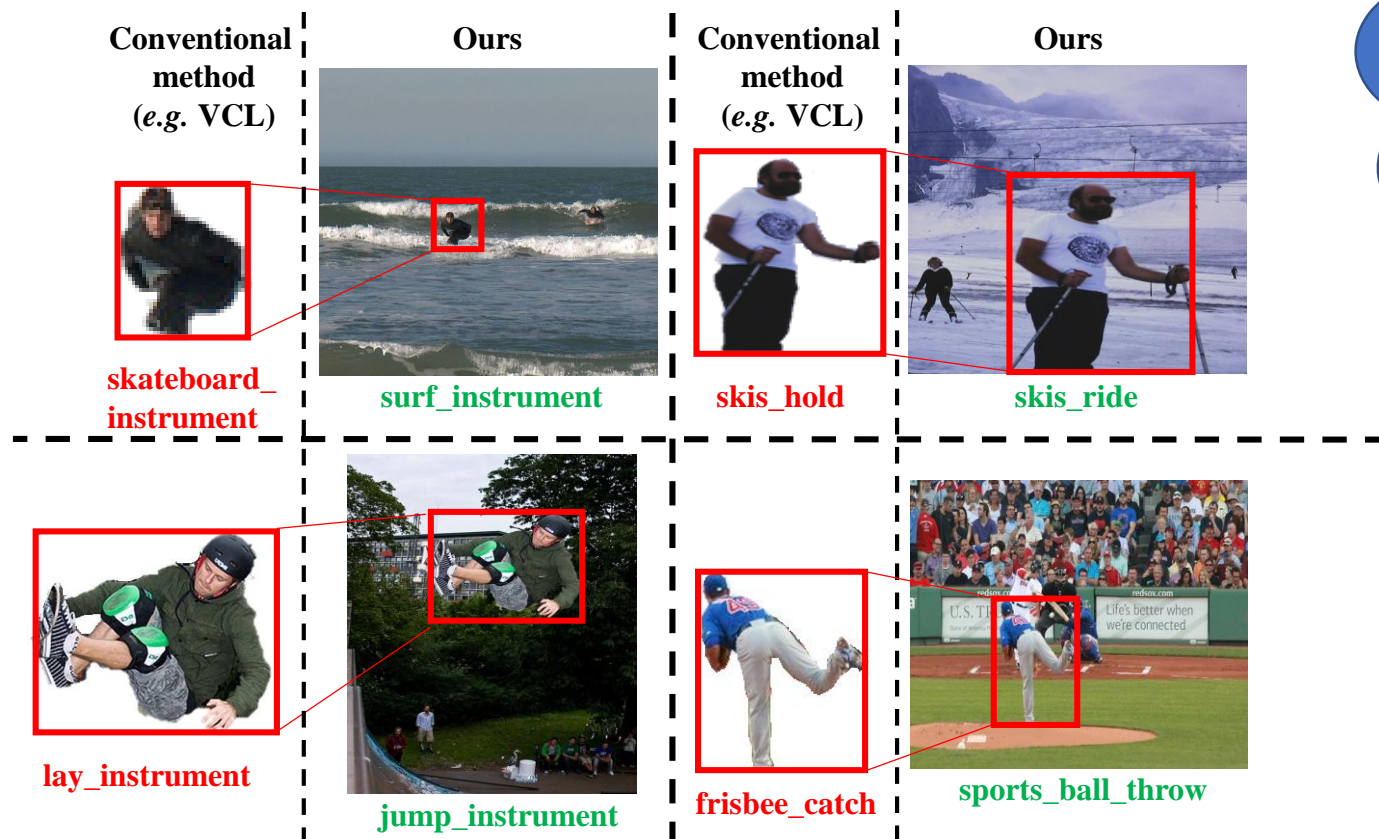
HOI with Missing objects

Six HOI-MO categories chosen in a data-driven manner are shown.



We found that even in widely adopted public HOI detection datasets (V-COCO and HICO-DET) which are carefully constructed to avoid HOI-MO cases, there still is a large amount of HOI-MO instances (more than 10%).

HOI with Missing objects

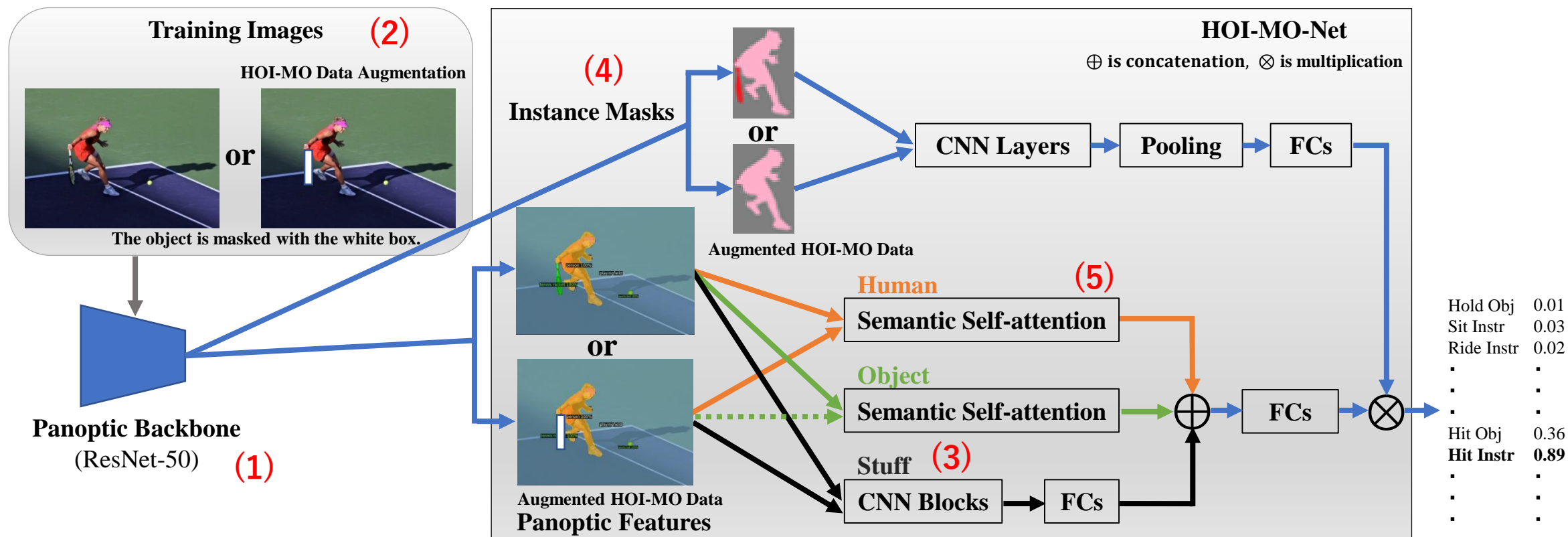


Without vs. with stuff (global scene) context

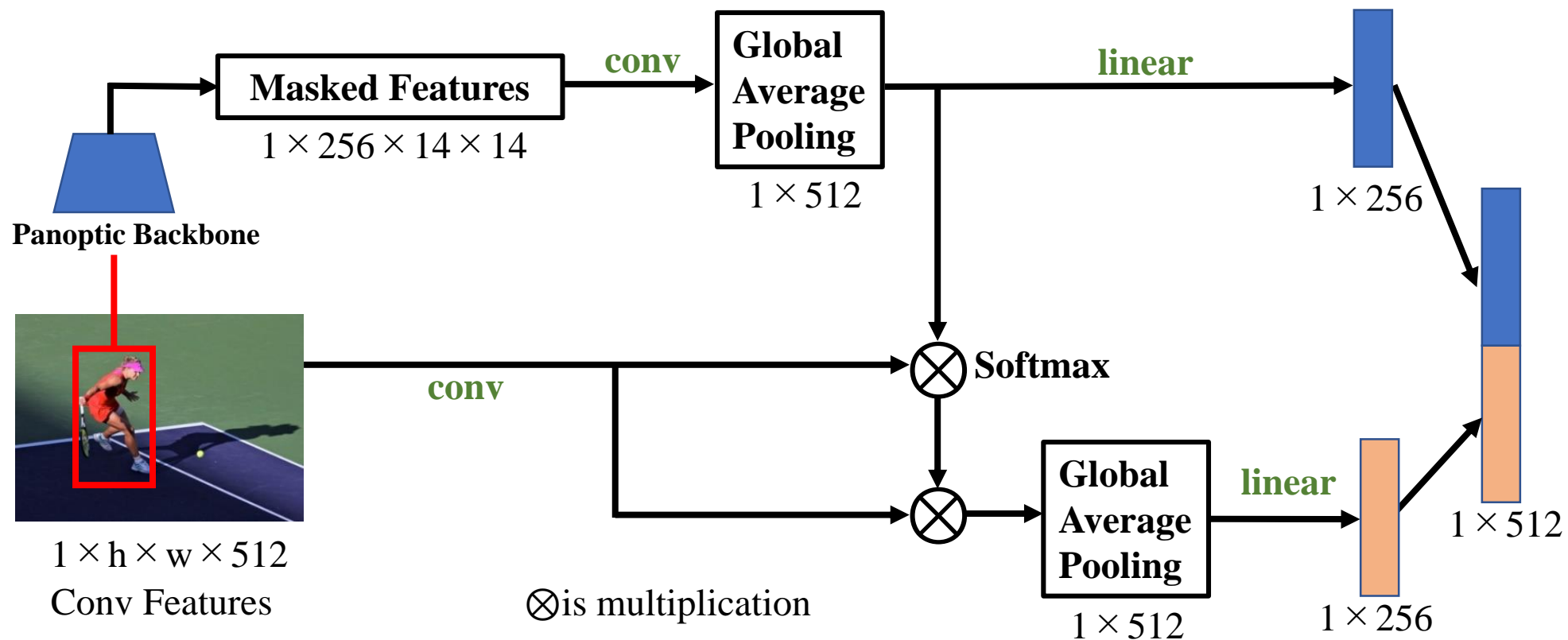
If the global scene context is taken into account, it seems that HOI can be detected even if the object cannot be detected.



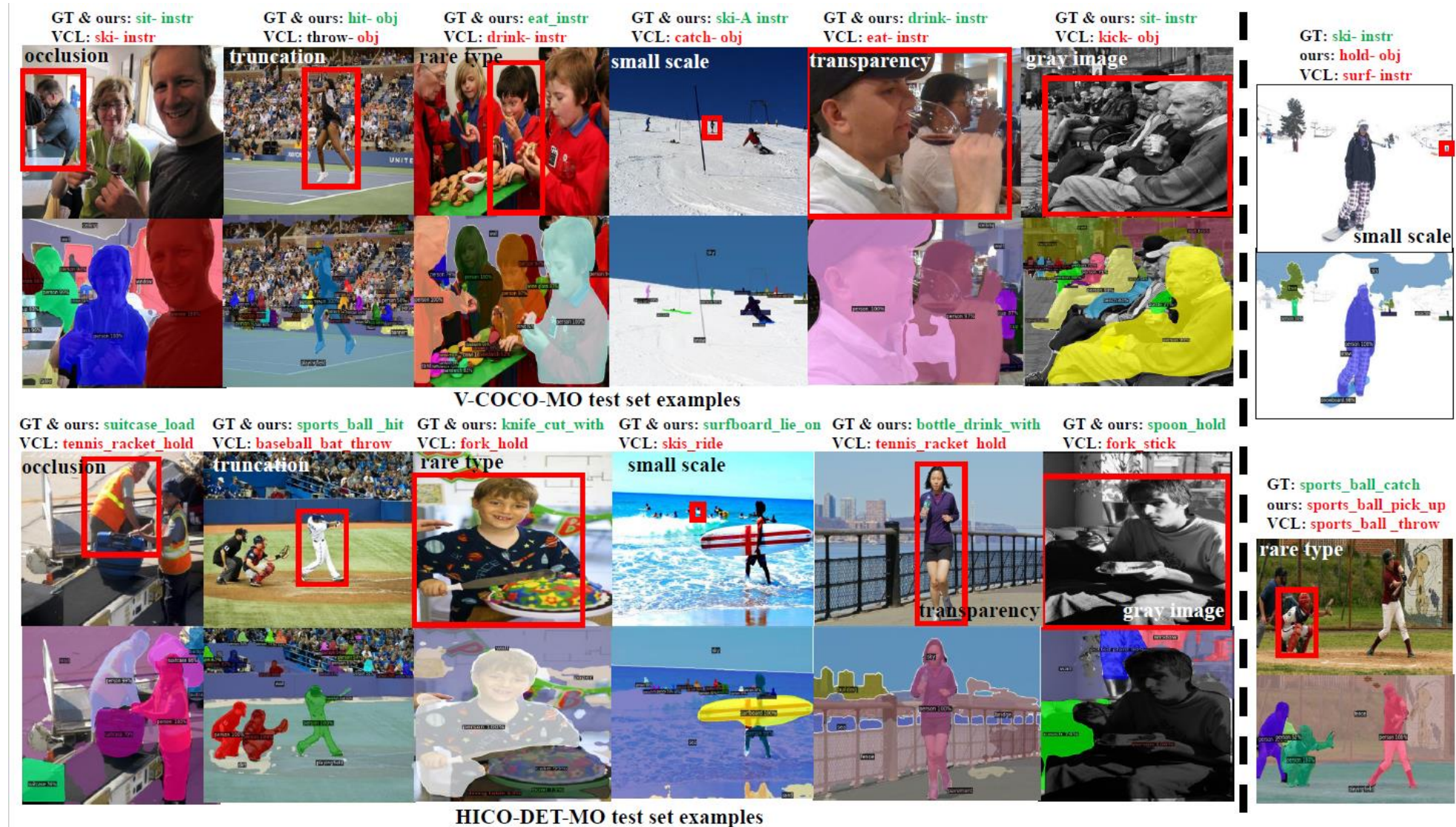
Overview



Method: Semantic Self-attention



More Visual Examples



Experimental Results

Table 1. Results on Mixed-V-COCO/HICO-DET and V-COCO/HICO-DET-MO test sets in mAP.

Method	Mixed-V-COCO	Mixed-HICO-DET			V-COCO-MO	HICO-DET-MO		
		Full	Rare	Non-Rare		Full	Rare	Non-Rare
iCAN [5]	38.06	13.93	10.35	15.11	5.37	5.42	4.30	5.67
TIN-net [11]	40.88	16.04	13.36	17.02	5.60	4.97	4.31	5.12
VCL [8]	41.07	18.00	16.36	18.75	8.45	4.70	4.55	4.74
HOI-MO-Net	48.76	19.83	16.39	20.93	31.23	13.21	12.45	13.38

Table 2. Performance on V-COCO-MO by individual HOI-MO categories(HC) in mAP, with (a)-(f) denoting the HOI-MO categories shown in previous page. `NS' denotes the number of samples. `Ours' denotes HOI-MO-Net.

HC	NS	iCAN [5]	TIN-net [11]	VCL [8]	Ours
(a)occlusion	1360	5.99	6.35	8.54	33.37
(b)truncation	198	0.72	0.85	1.06	60.37
(c)rare type	258	8.46	9.77	15.9	46.40
(d)small scale	514	1.87	1.66	3.67	31.67
(e)transparency	16	0.00	0.00	0.00	50.00
(f)gray image	30	0.00	0.00	0.00	36.38

Conclusion

- Introduced HOI-MO problem that has been overlooked in the past.
- Two new benchmark test sets for HOI-MO based on commonly used public datasets.
- A novel HOI-MO method that leverages global scene context.
- Demonstration of the effectiveness of the proposed model with extensive experiments.

Thank you for listening!