http://openaccess.thecvf.com/content_ECCV_2018/html/Siyuan_Qi_Learning_Human-Object_Interactions_ECCV_2018_paper.html

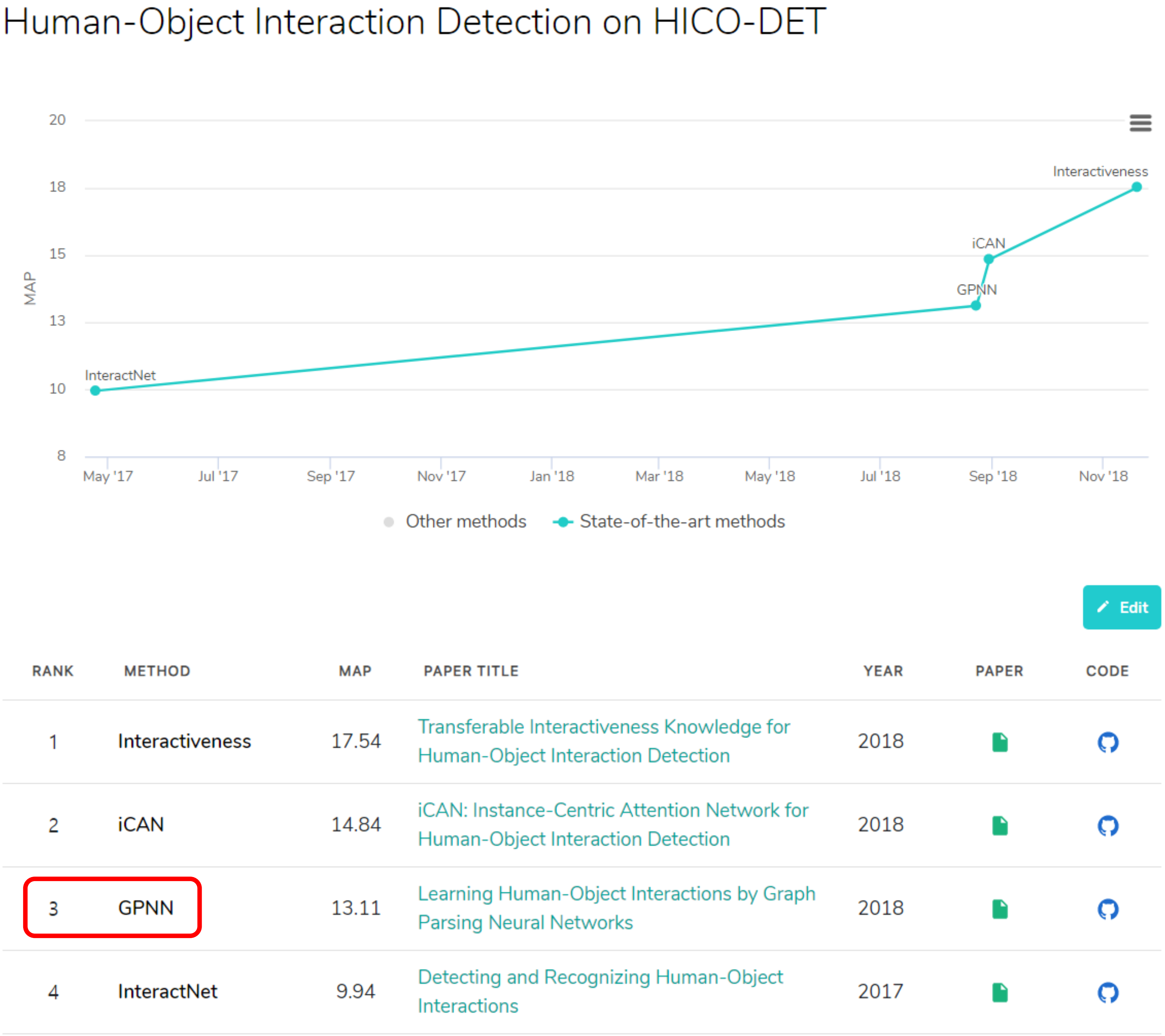
Learning Human-Object Interactions by Graph Parsing Neural Networks

Siyuan Qi, Wenguan Wang, Baoxiong Jia, Jianbing Shen, Song-Chun Zhu

1 University of California, Los Angeles 2 International Center for AI and Robot Autonomy (CARA) 3 Beijing Institute of Technology 4 Peking University 5 Inception Institute of Artificial Intelligence

ECCV 2018

인공지능 연구실
석사과정 구자봉



Human-Object Interaction Detection on HICO-DET

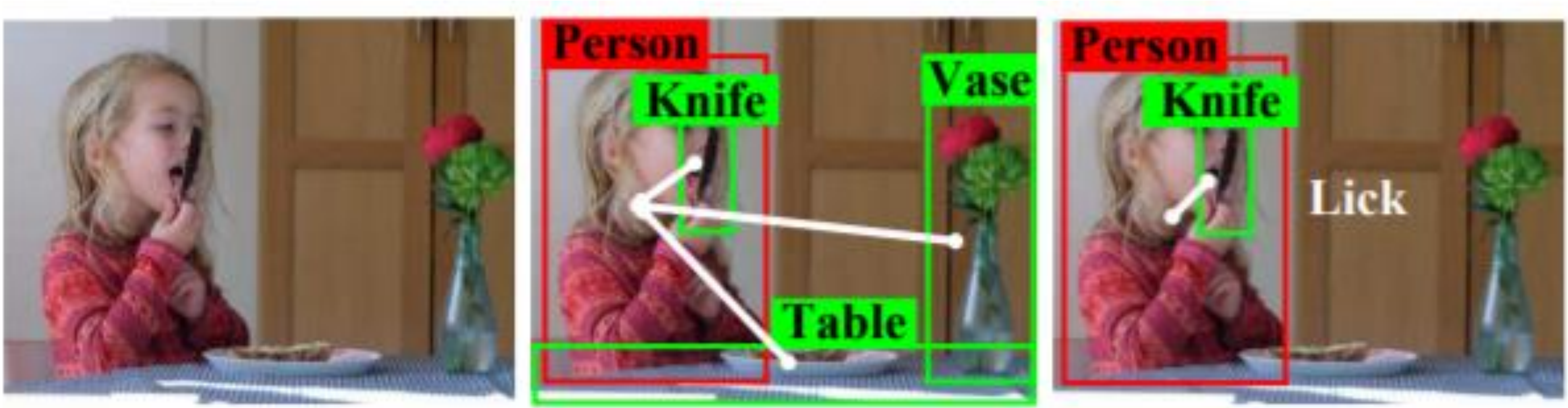
RANK	METHOD	MAP	PAPER TITLE	YEAR	PAPER	CODE
1	Interactiveness	17.54	Transferable Interactiveness Knowledge for Human-Object Interaction Detection	2018	📄	●
2	iCAN	14.84	iCAN: Instance-Centric Attention Network for Human-Object Interaction Detection	2018	📄	●
3	GPNN	13.11	Learning Human-Object Interactions by Graph Parsing Neural Networks	2018	📄	●
4	InteractNet	9.94	Detecting and Recognizing Human-Object Interactions	2017	📄	●

https://paperswithcode.com/sota/human-object-interaction-detection-on-hico

INDEX

1. INTRODUCTION
2. RELATED WORKS
3. PROPOSED MODEL
4. EXPERIMENTS
5. CONCLUSION

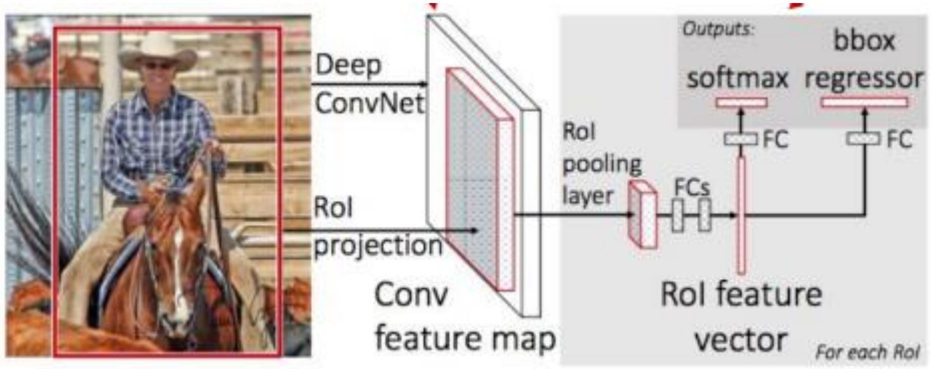
1. INTRODUCTION

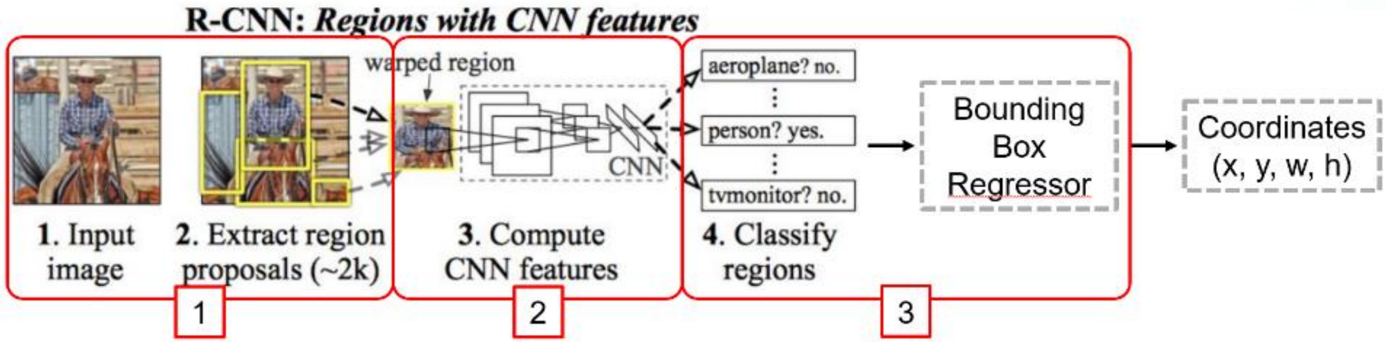


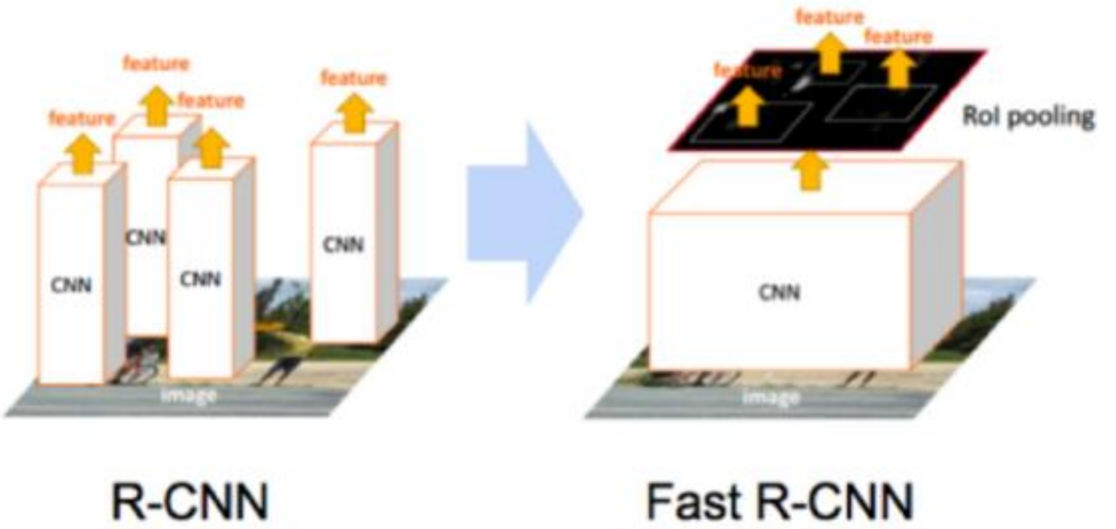
https://seongkyun.github.io/papers/2019/01/06/Object_detection/

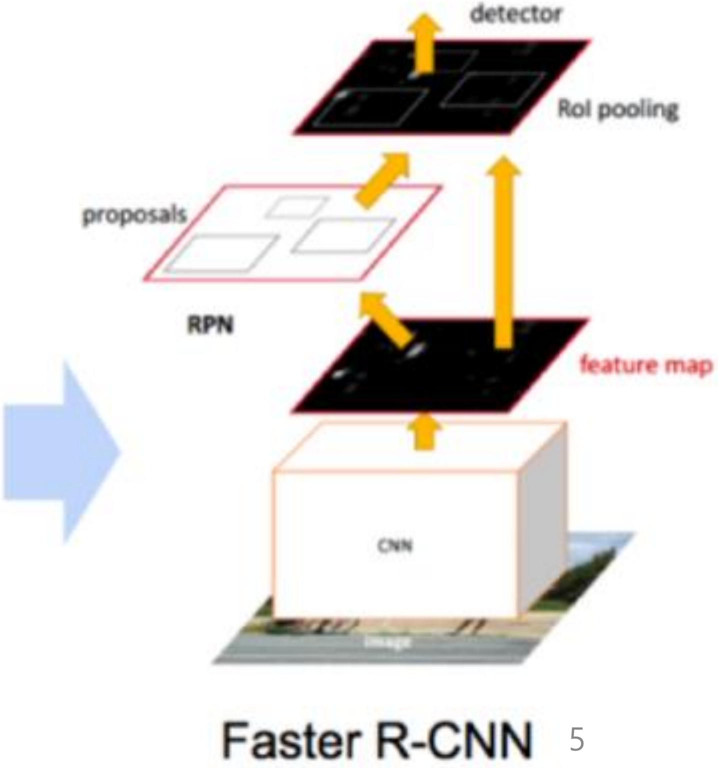
2. RELATED WORKS

- Object Detection





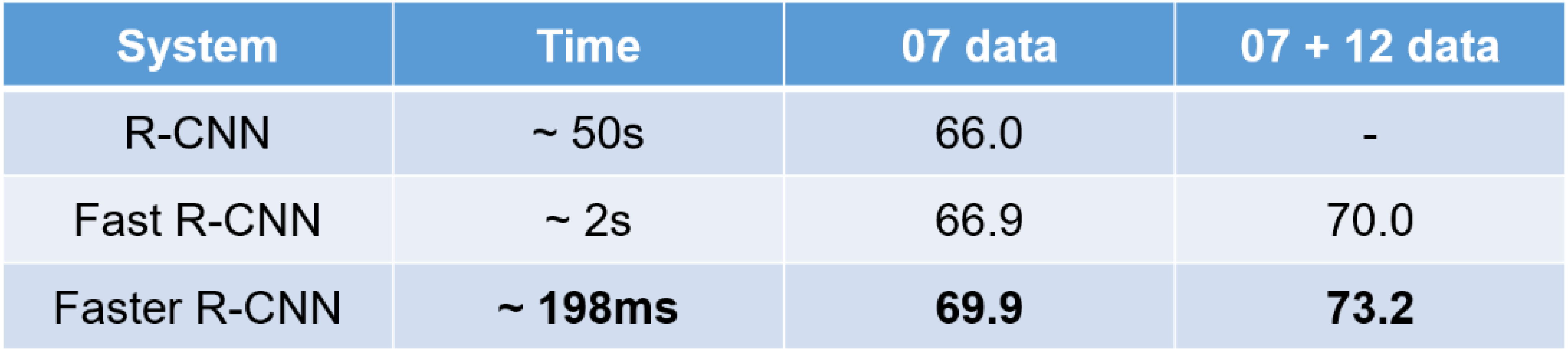




https://seongkyun.github.io/papers/2019/01/06/Object_detection/

2. RELATED WORKS

- Object Detection

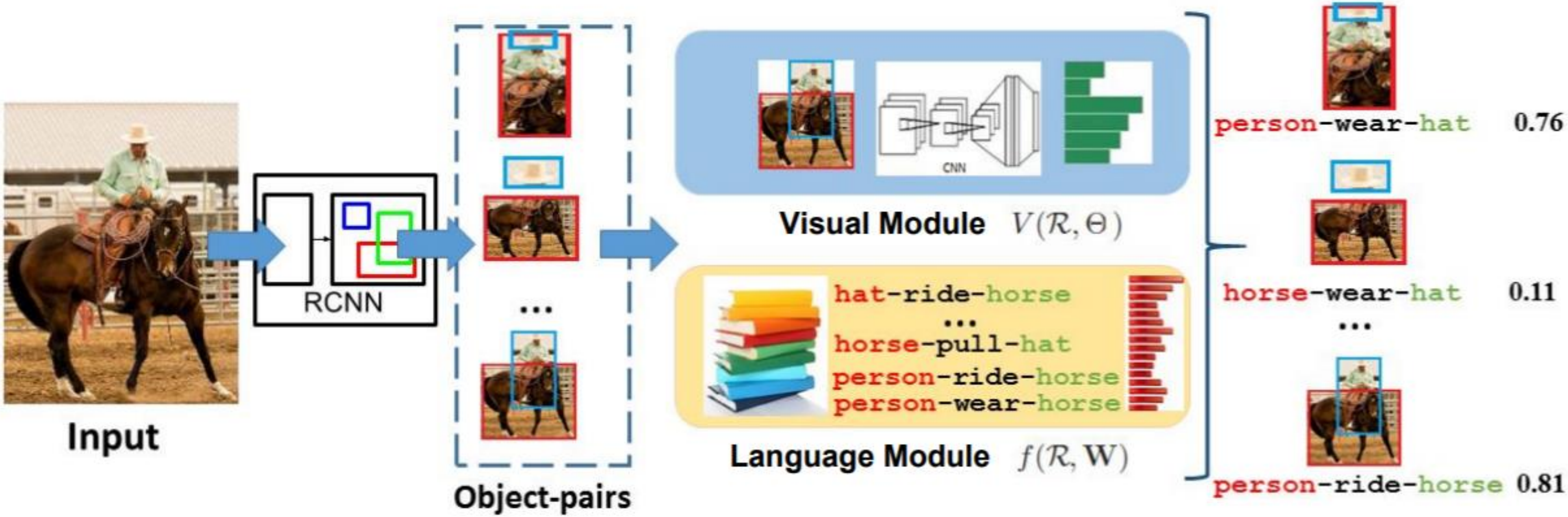


System	Time	07 data	07 + 12 data
R-CNN	~ 50s	66.0	-
Fast R-CNN	~ 2s	66.9	70.0
Faster R-CNN	**~ 198ms**	**69.9**	**73.2**

https://arxiv.org/pdf/1608.00187.pdf

2. RELATED WORKS

- Visual Relationship Detection

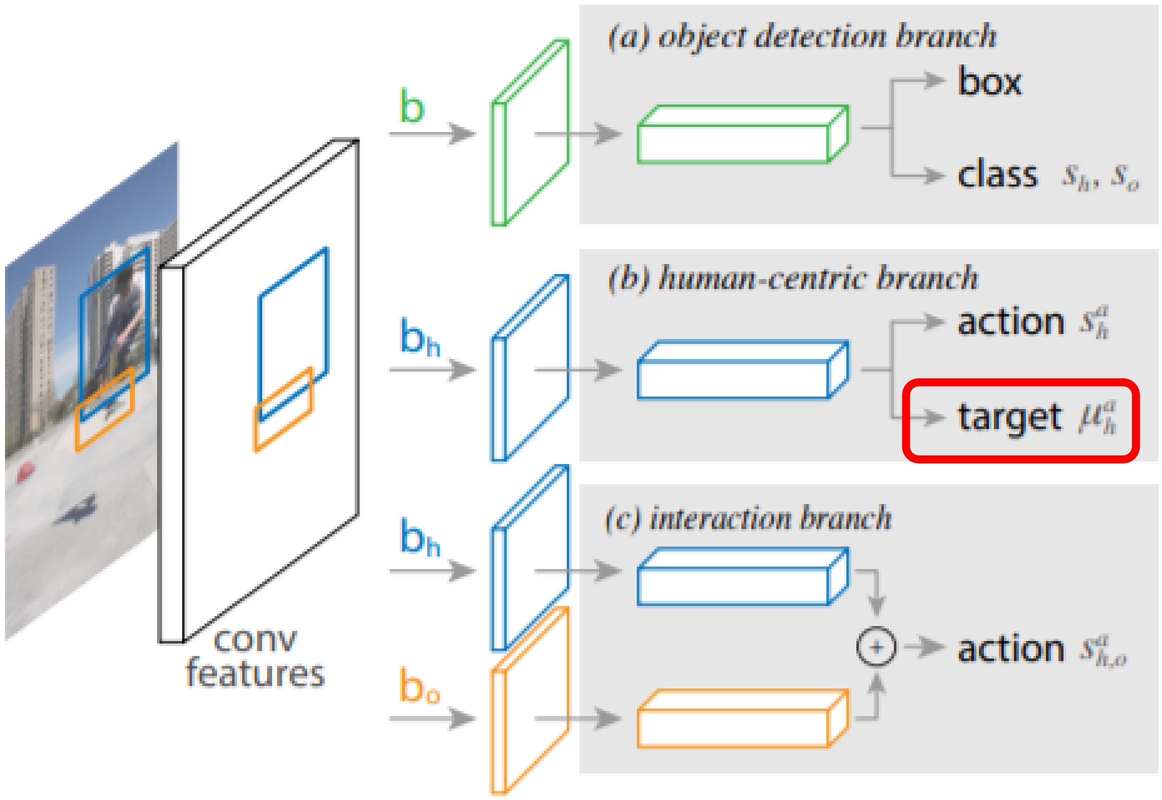


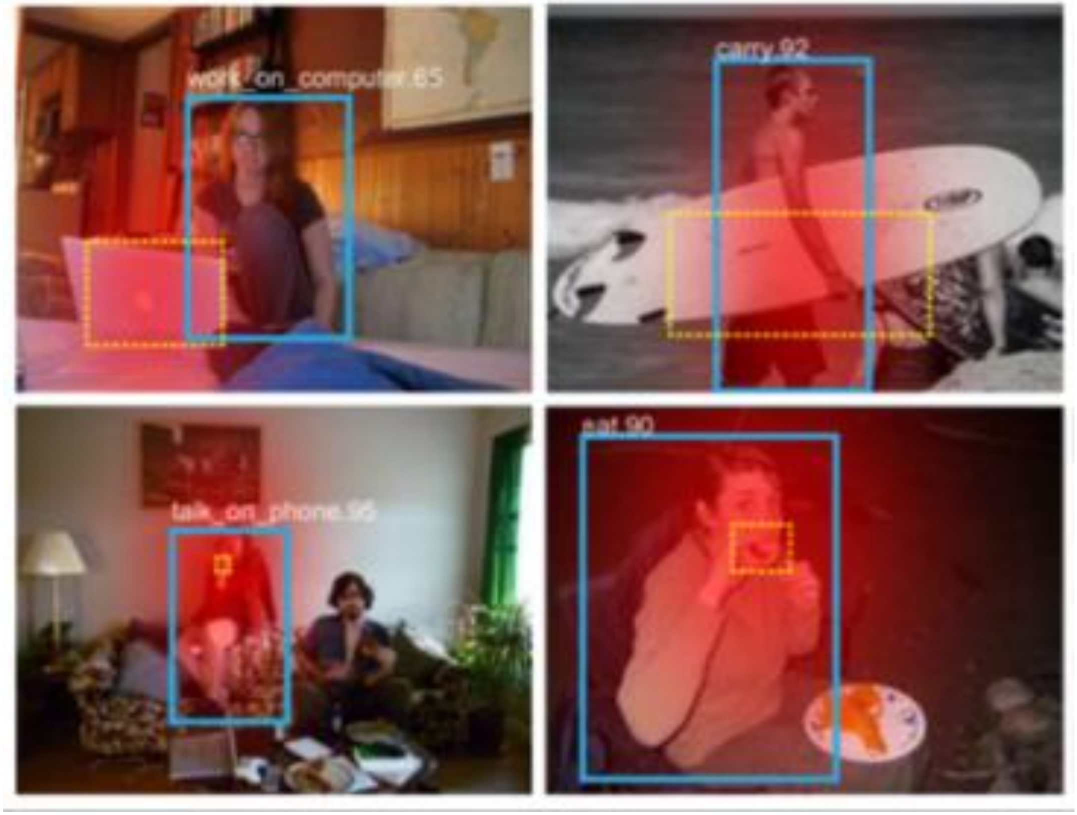
- Human-Object Interaction Detection

https://arxiv.org/pdf/1704.07333.pdf

Recently methods

InteractNet(action specific density map estimation method)





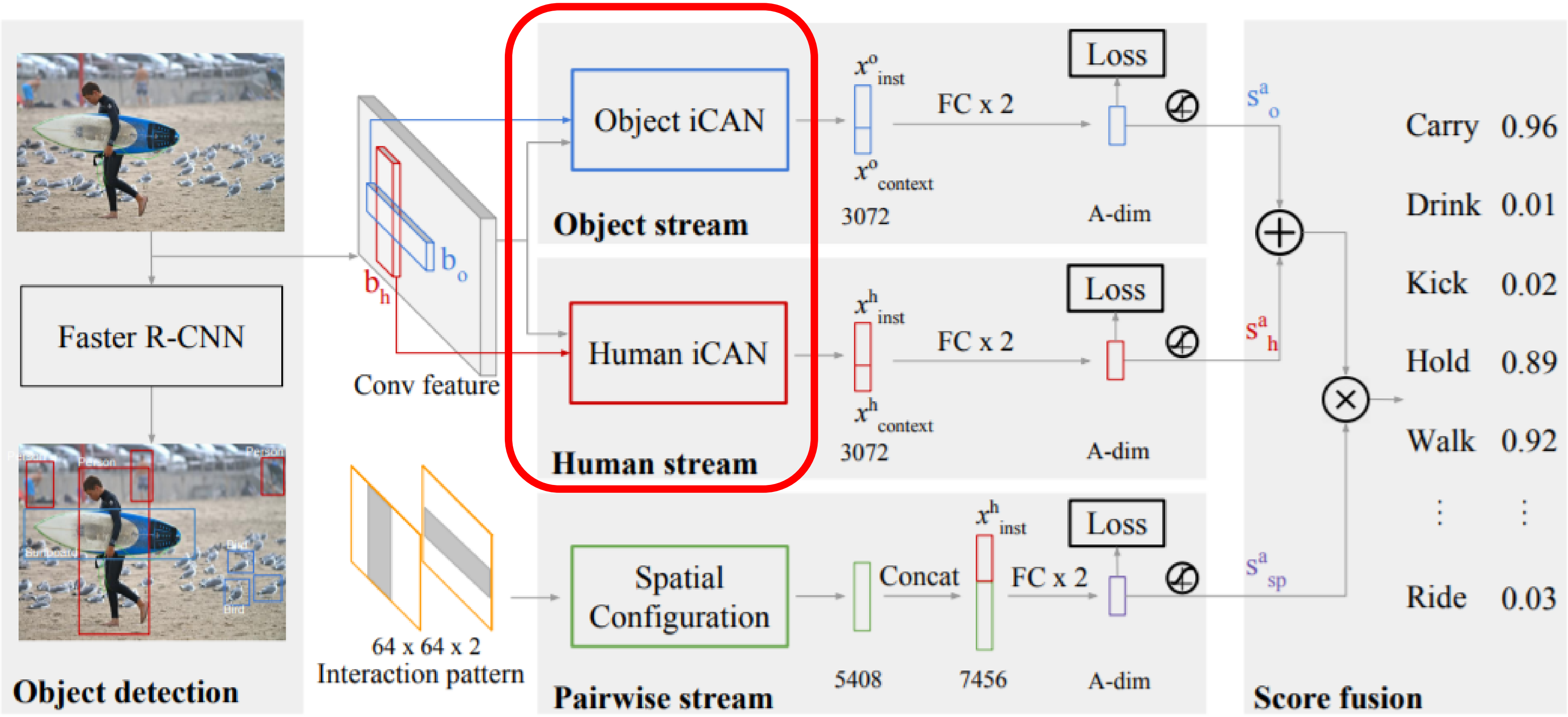
AI Lab.
ARTIFICIAL INTELLIGENCE LAB.

- Human-Object Interaction Detection

https://arxiv.org/pdf/1808.10437.pdf

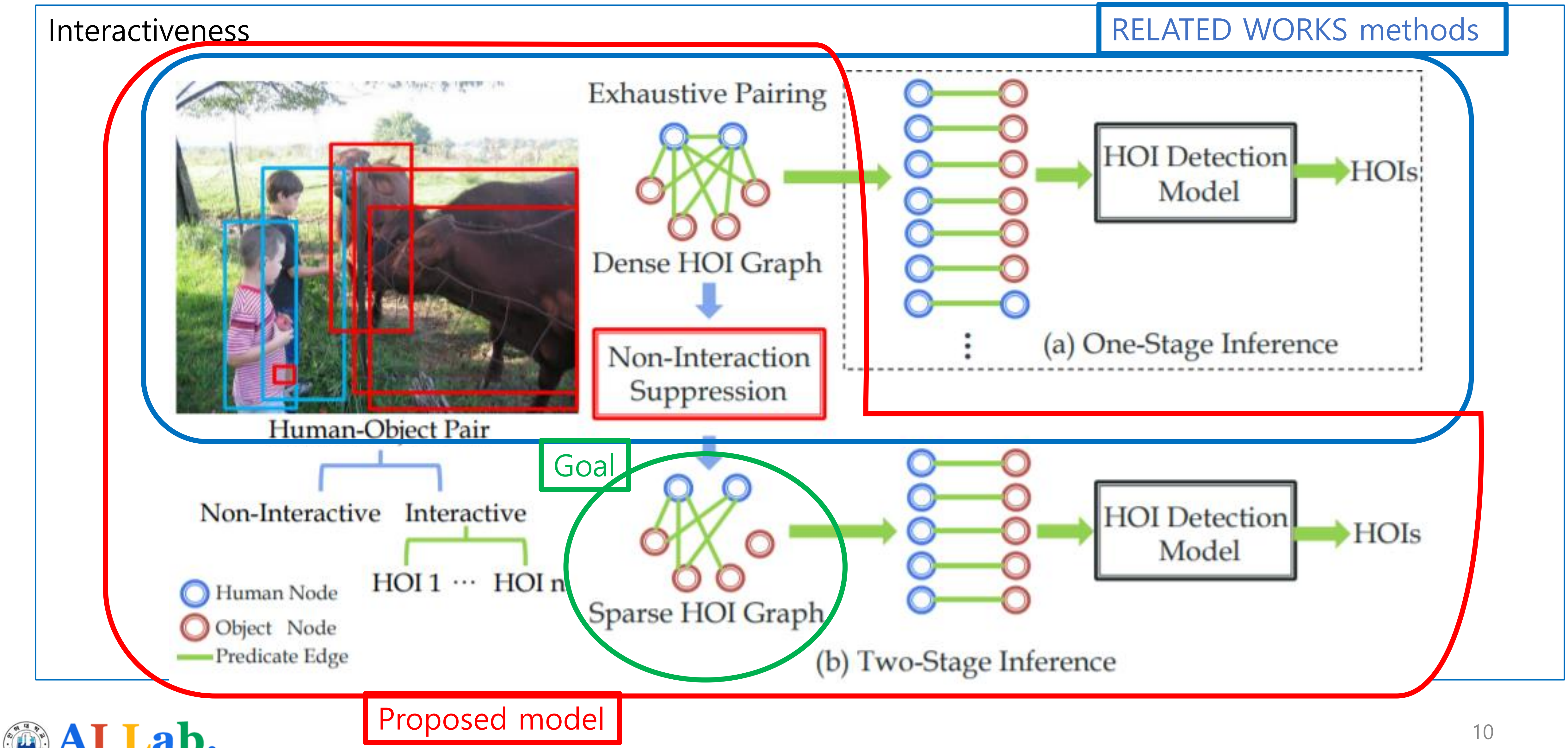
Recently methods

ICAN(instance centric attention module)



- Human-Object Interaction Detection

Recently methods



Interactiveness

RELATED WORKS methods

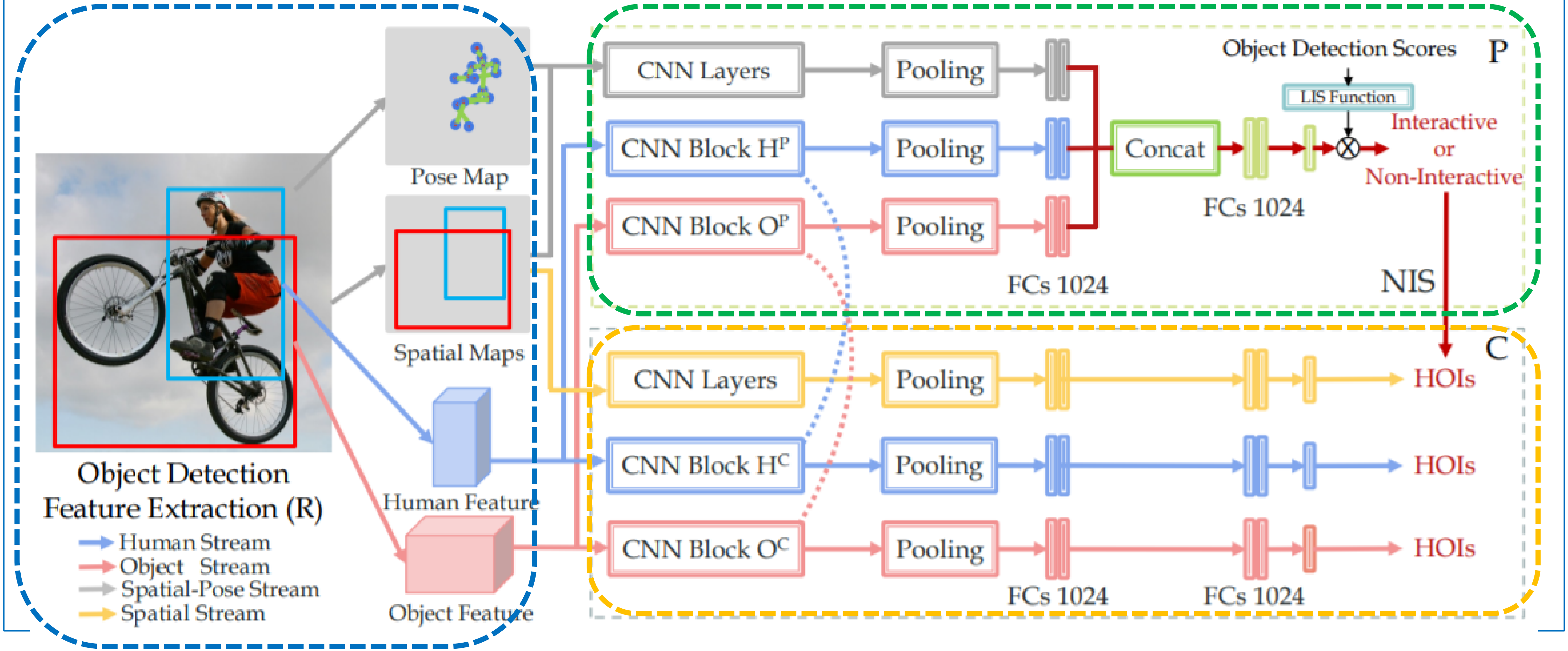
Goal

Proposed model

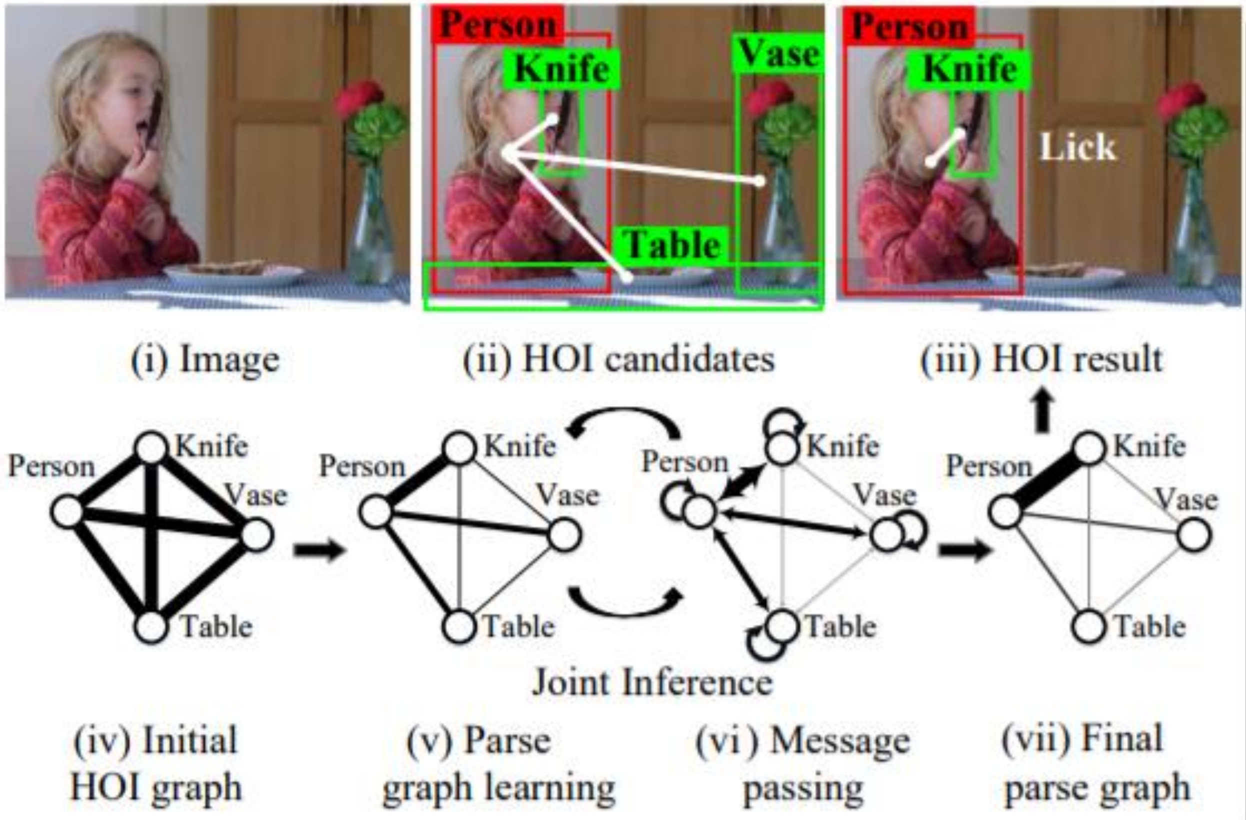
• Human-Object Interaction Detection

Recently methods

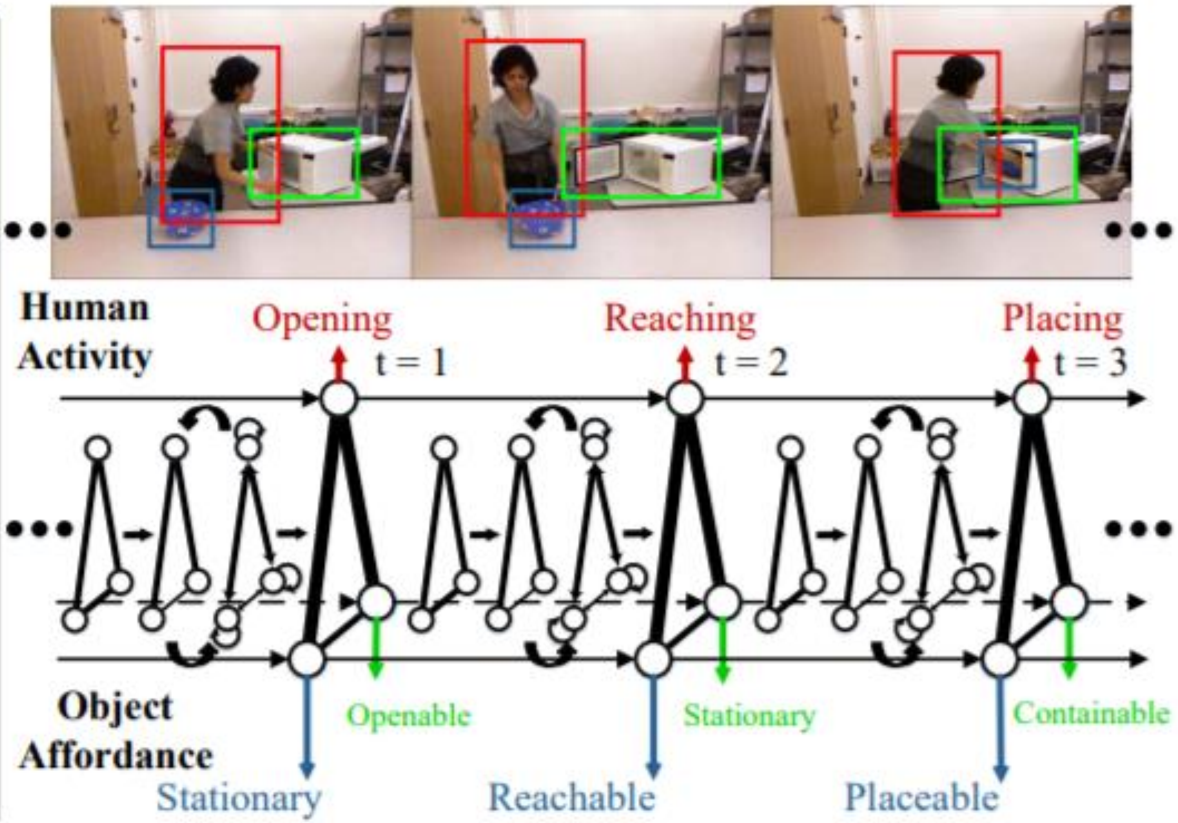
Interactiveness



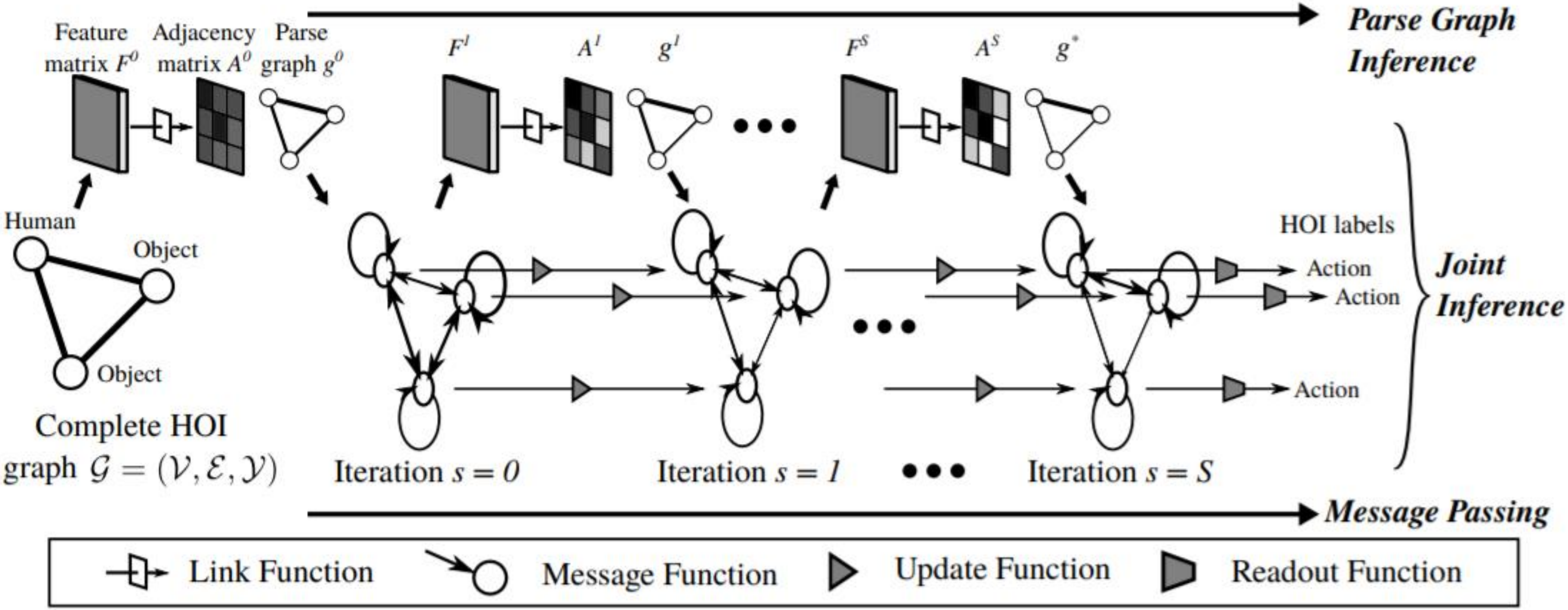
3. PROPOSED MODEL



(a) Human-Object Interaction Detection in Still Images



(b) Human-Object Interaction Recognition in Videos



$$\Gamma = \{\Gamma^{\mathcal{V}}, \Gamma^{\mathcal{E}}\} \qquad p(\mathcal{V}_g, \mathcal{E}_g | \Gamma, \mathcal{G})$$

$$p(\mathcal{Y}_g | \mathcal{V}_g, \mathcal{E}_g, \Gamma)$$

$$\begin{aligned} g^* &= \operatorname*{argmax}_g \; p(g|\Gamma, \mathcal{G}) = \operatorname*{argmax}_g \; p(\mathcal{V}_g, \mathcal{E}_g, \mathcal{Y}_g | \Gamma, \mathcal{G}) \\ &= \operatorname*{argmax}_g \; p(\mathcal{Y}_g | \mathcal{V}_g, \mathcal{E}_g, \Gamma) p(\mathcal{V}_g, \mathcal{E}_g | \Gamma, \mathcal{G}) \end{aligned}$$

Link Function

$$\Gamma^{\mathcal{V}} \quad \Gamma^{\mathcal{E}}$$

$$A \in [0,1]^{|\mathcal{V}|\times|\mathcal{V}|}$$

$$A_{vw} = L(\Gamma_v, \Gamma_w, \Gamma_{vw})$$

$$A^s = \sigma(\mathbf{W}^L * F^s)$$

Update Function

$$h_v^s = U(h_v^{s-1}, m_v^s)$$

$$h_v^s = U(h_v^{s-1}, m_v^s) = GRU(h_v^{s-1}, m_v^s)$$

$$A_{vw}^s = L(h_v^{s-1}, h_w^{s-1}, m_{vw}^{s-1})$$

Message Function

$$m_v^s = \sum_w A_{vw} M(h_v^{s-1}, h_w^{s-1}, \Gamma_{vw})$$

$$M(h_v, h_w, \Gamma_{vw}) = [\mathbf{W}_V^M h_v, \mathbf{W}_V^M h_w, \mathbf{W}_E^M \Gamma_{vw}]$$

Readout Function

$$y_v = R(h_v^S)$$

$$y_v = R(h_v^S) = \varphi(\mathbf{W}^R h_v^S)$$

$$m_v^s = \sum_w A_{vw}^s M(h_v^{s-1}, h_w^{s-1}, \Gamma_{vw})$$

4. EXPERIMENTS

- Datasets

V-COCO
Images 10,346(2,533, 2,867, 4,946)

People 16,199
HOI 29 (24(object), 5(no object))

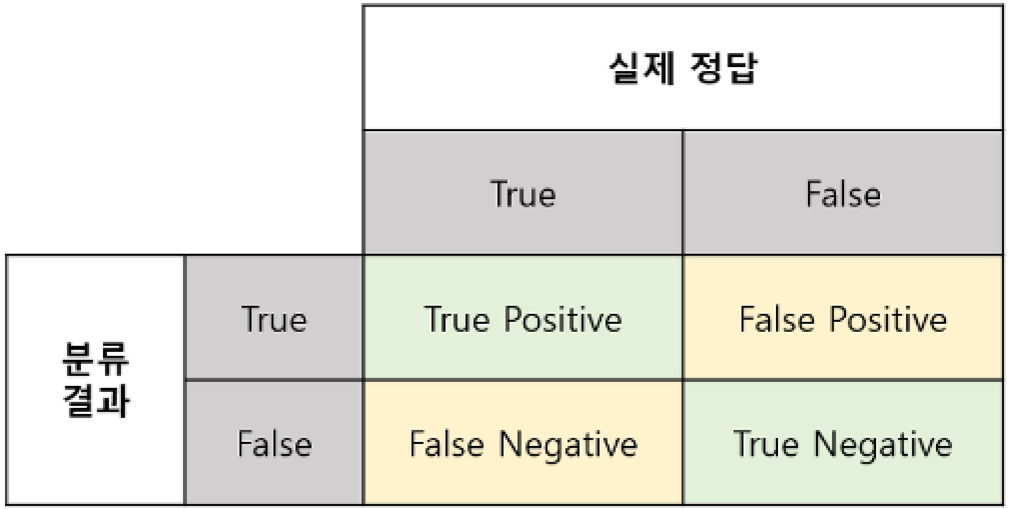
HICO-DET
Images 47,776 (38,118, 9,658)
Objects 80 (airplane, apple…)
Verbs 117 (carry, catch…)
HOI 600 (airplane – board, direct, exit, fly…)

HOI Remark >= 150k

CAD-120

- Environment

라이브러리 : PyTorch
GPU : Nvidia Titan Xp GPU
평가지표 : mAP



		실제 정답	
		True	False
분류 결과	True	True Positive	False Positive
	False	False Negative	True Negative

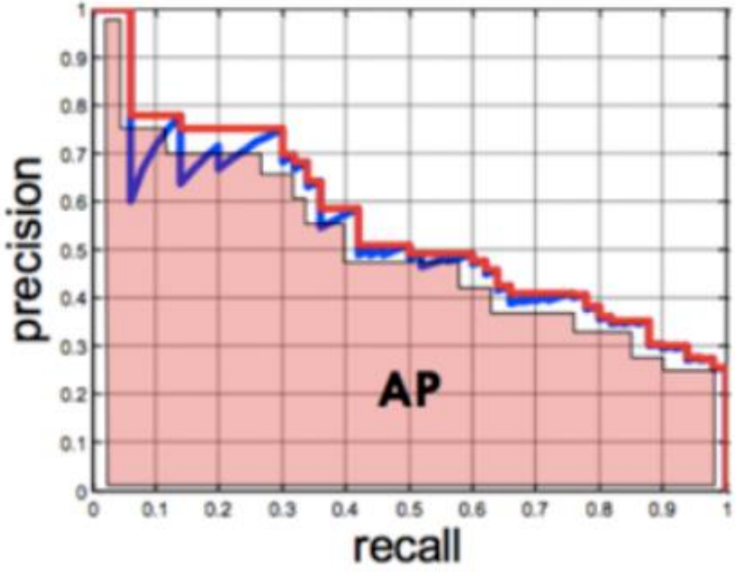


그림 2. average precision

$$(Precision) = \frac{TP}{TP + FP}$$

$$(Recall) = \frac{TP}{TP + FN}$$

$$(Accuracy) = \frac{TP + TN}{TP + FN + FP + TN}$$

$$(F1\text{-}score) = 2 \times \frac{1}{\frac{1}{Precision} + \frac{1}{Recall}} = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

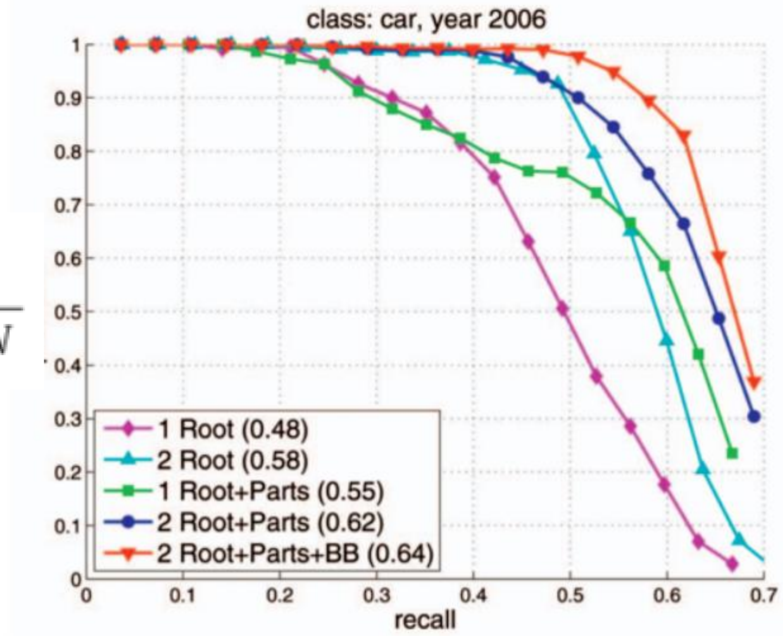


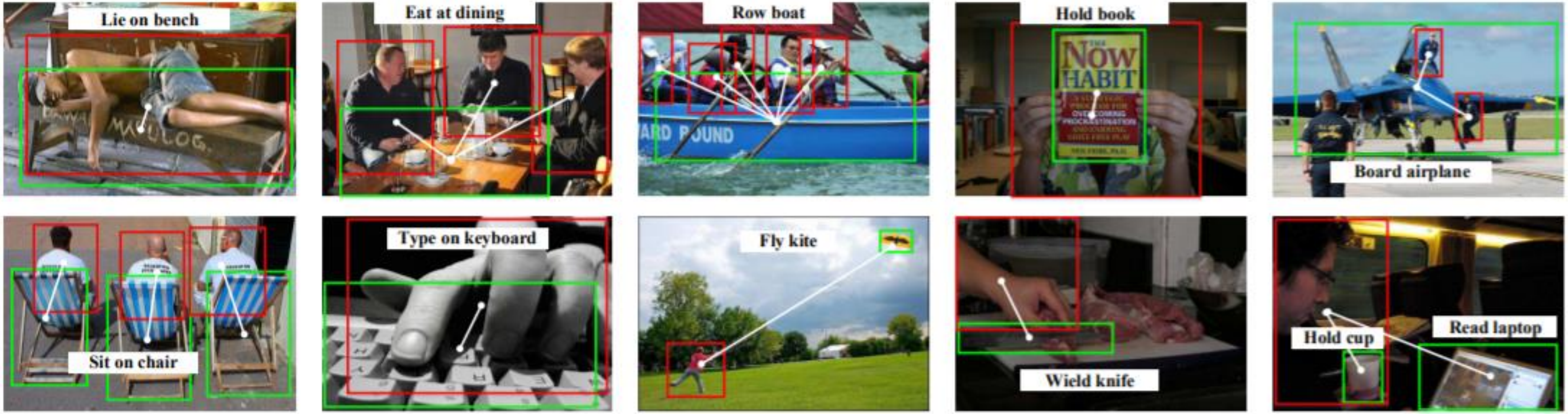
그림 1. precision-recall 그래프의 예

mAP(**mean Average Precision**) :
멀티 오브젝트 디텍션 문제에 있어 AP 들의 mean 값

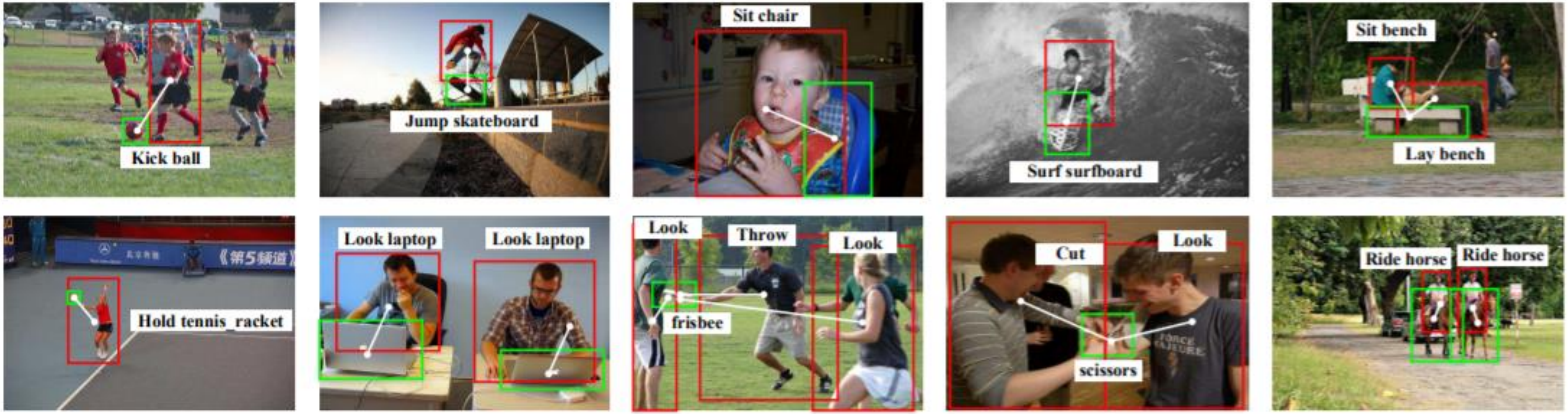
https://sumniya.tistory.com/26

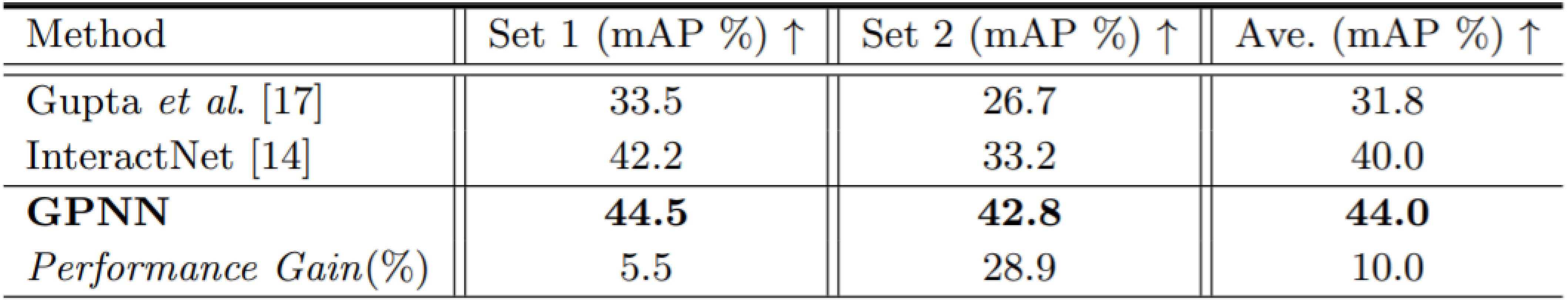
https://darkpgmr.tistory.com/162

https://eehoeskrap.tistory.com/237

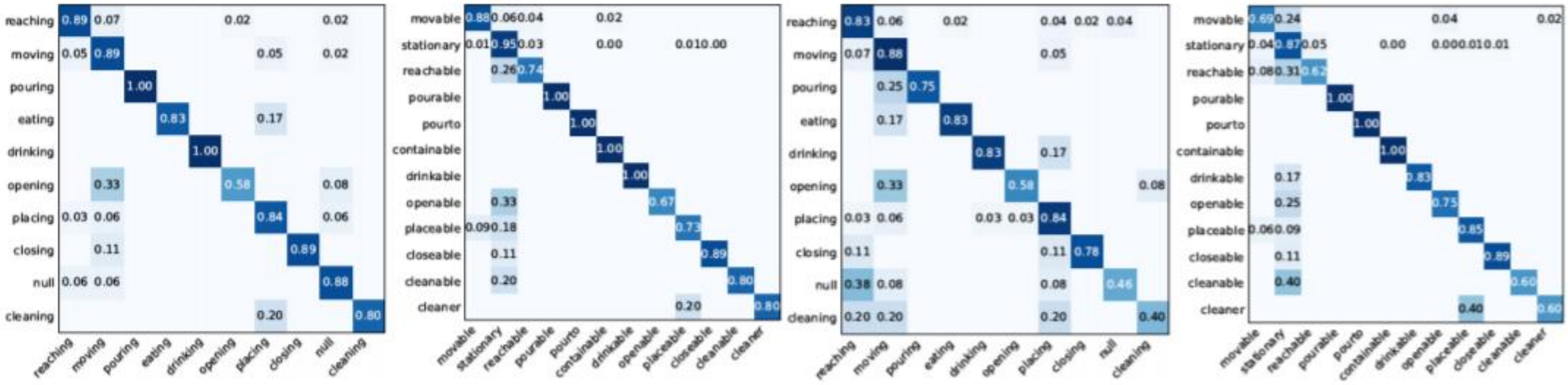


Methods	Full (mAP %) ↑	Rare (mAP %) ↑	Non-rare (mAP %) ↑
Random	1.35×10^{-3}	5.72×10^{-4}	1.62×10^{-3}
Fast-RCNN(union) [13]	1.75	0.58	2.10
Fast-RCNN(score) [13]	2.85	1.55	3.23
HO-RCNN [1]	5.73	3.21	6.48
HO-RCNN+IP [1]	7.30	4.68	8.08
HO-RCNN+IP+S [1]	7.81	5.37	8.54
Gupta *et al.* [17]	9.09	7.02	9.71
Shen *et al.* [38]	6.46	4.24	7.12
InteractNet [14]	9.94	7.16	10.77
GPNN	**13.11**	**9.34**	**14.23**
Performance Gain(%)	31.89	30.45	32.13

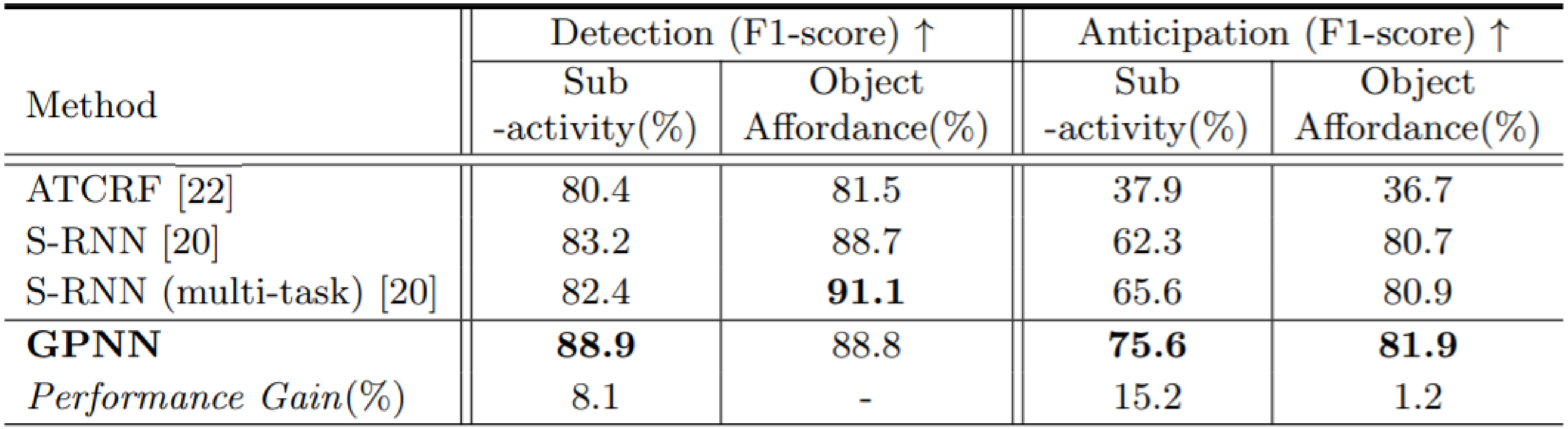




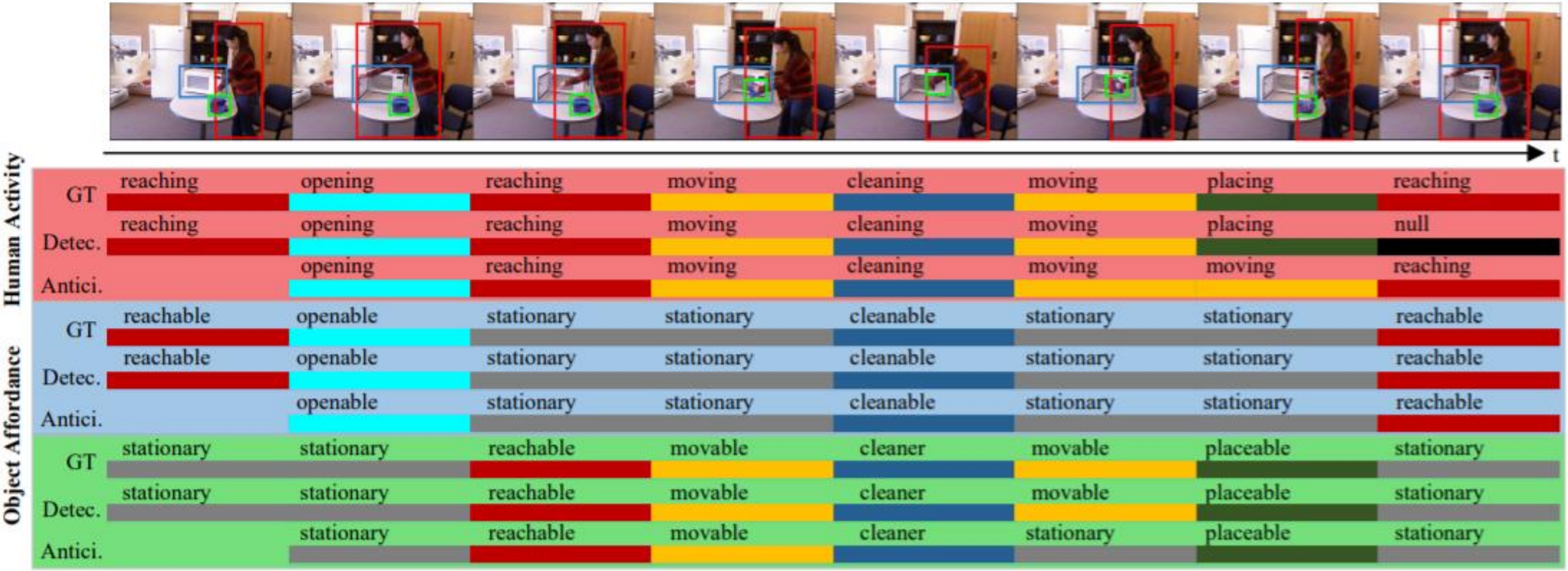
Method	Set 1 (mAP %) ↑	Set 2 (mAP %) ↑	Ave. (mAP %) ↑
Gupta *et al.* [17]	33.5	26.7	31.8
InteractNet [14]	42.2	33.2	40.0
GPNN	**44.5**	**42.8**	**44.0**
Performance Gain(%)	5.5	28.9	10.0



(a) Action (b) Affordance (c) Action (d) Affordance



Method	Detection (F1-score) ↑		Anticipation (F1-score) ↑	
	Sub -activity(%)	Object Affordance(%)	Sub -activity(%)	Object Affordance(%)
ATCRF [22]	80.4	81.5	37.9	36.7
S-RNN [20]	83.2	88.7	62.3	80.7
S-RNN (multi-task) [20]	82.4	**91.1**	65.6	80.9
GPNN	**88.9**	88.8	**75.6**	**81.9**
Performance Gain(%)	8.1	-	15.2	1.2

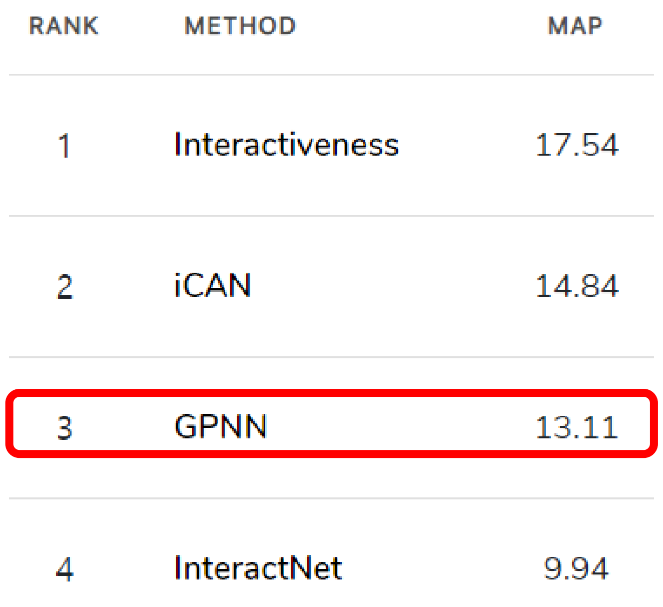


Human Activity	GT	reaching	opening	reaching	moving	cleaning	moving	placing	reaching
	Detec.	reaching	opening	reaching	moving	cleaning	moving	placing	null
	Antici.		opening	reaching	moving	cleaning	moving	moving	reaching
Object Affordance	GT	reachable	openable	stationary	stationary	cleanable	stationary	stationary	reachable
	Detec.	reachable	openable	stationary	stationary	cleanable	stationary	stationary	reachable
	Antici.		openable	stationary	stationary	cleanable	stationary	stationary	reachable
	GT	stationary	stationary	reachable	movable	cleaner	movable	placeable	stationary
	Detec.	stationary	stationary	reachable	movable	cleaner	movable	placeable	stationary
	Antici.		stationary	reachable	movable	cleaner	stationary	placeable	stationary

5. CONCLUSION

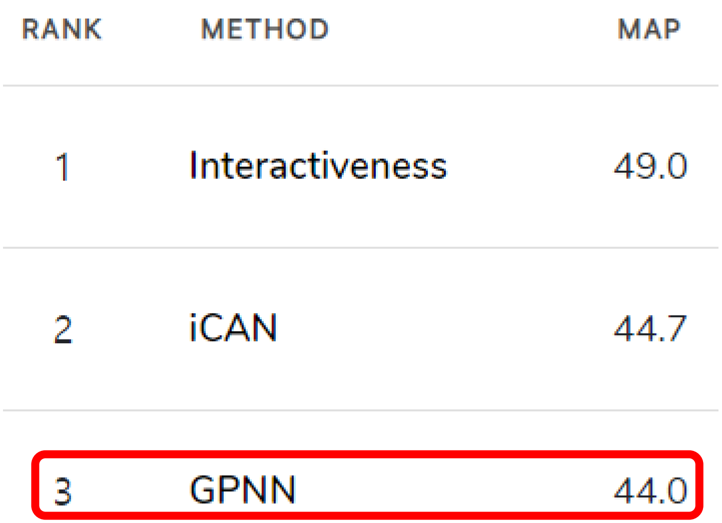
- GPNN (link functions, message functions, update functions and readout functions)

HICO-DET



RANK	METHOD	MAP
1	Interactiveness	17.54
2	iCAN	14.84
3	GPNN	13.11
4	InteractNet	9.94

V-COCO



RANK	METHOD	MAP
1	Interactiveness	49.0
2	iCAN	44.7
3	GPNN	44.0

Q & A