

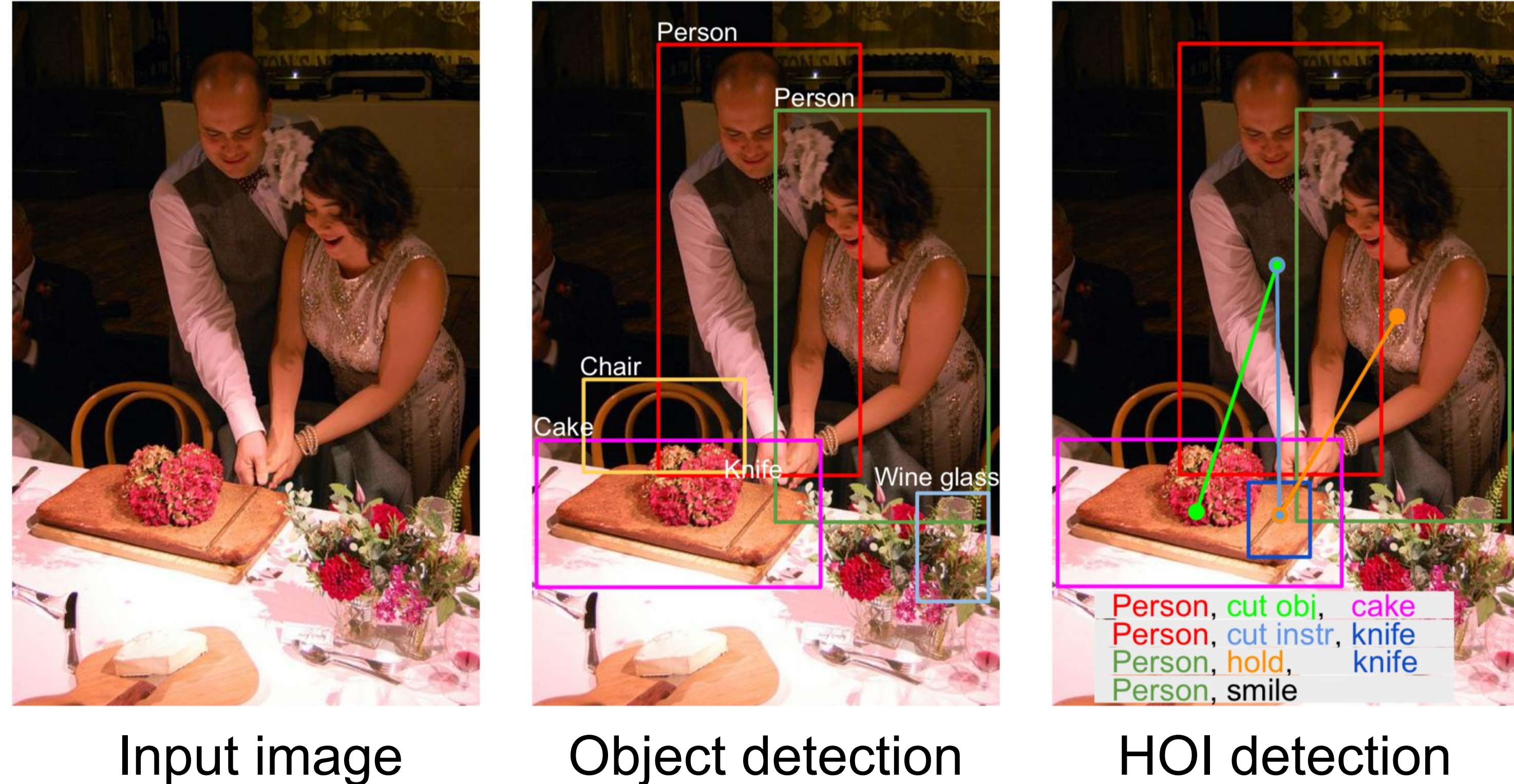
iCAN: Instance-Centric Attention Network for Human-Object Interaction Detection

Chen Gao Yuliang Zou Jia-Bin Huang
Virginia Tech



Code available at https://bit.ly/iCAN_BMVC18

Human-Object Interaction Detection



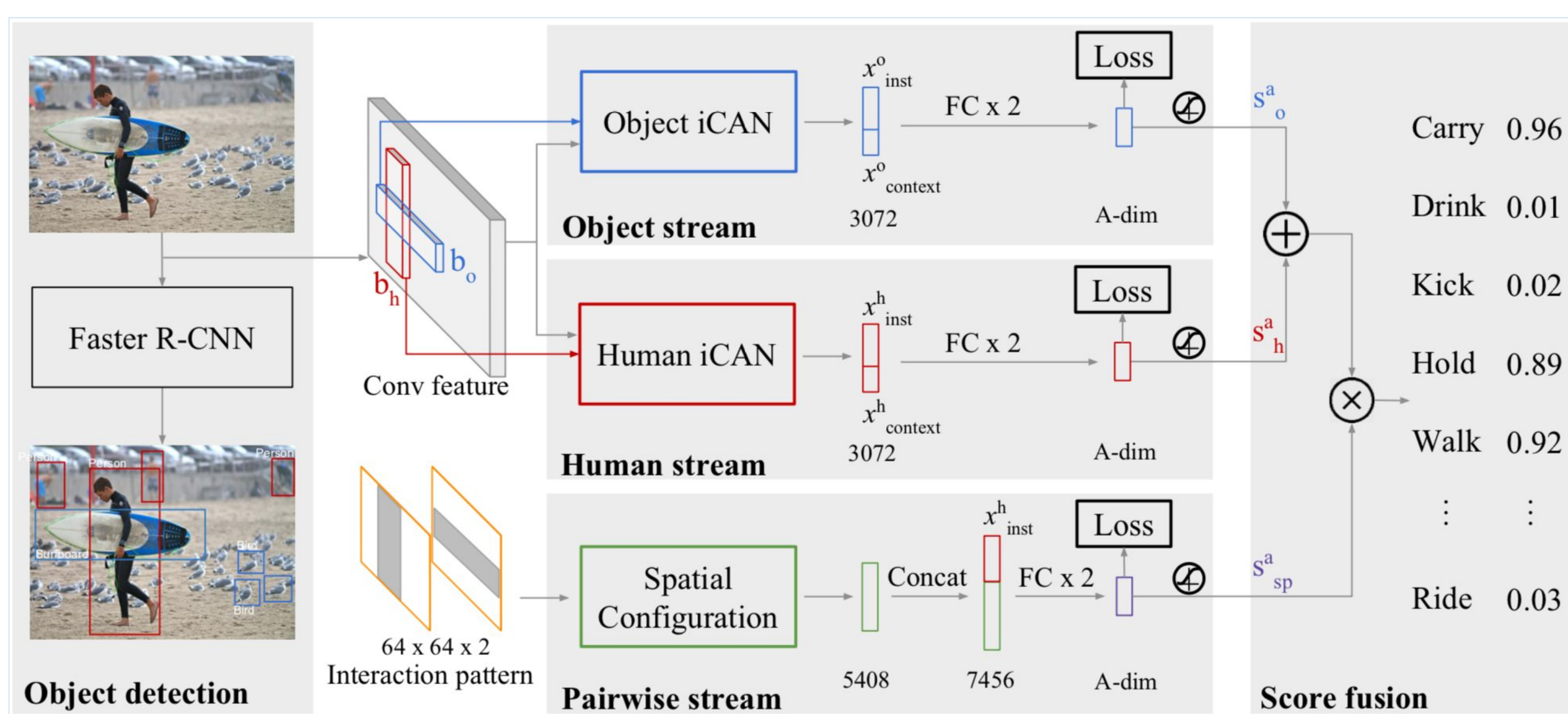
Input image

Object detection

HOI detection

What is where? $\Rightarrow$ What is happening?

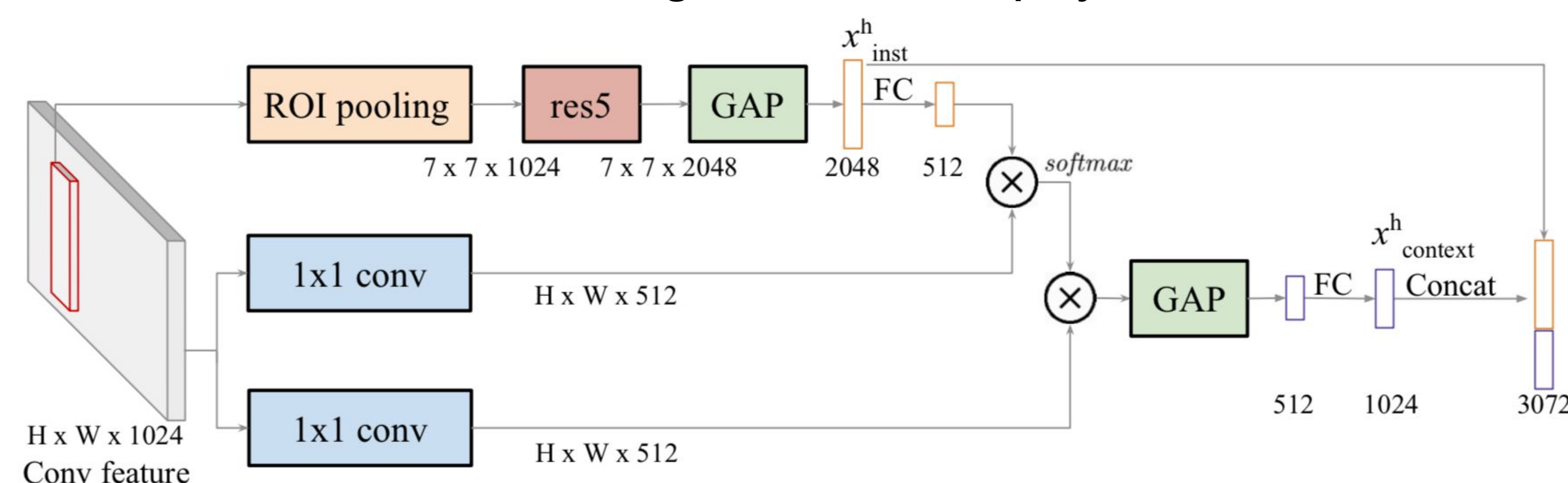
Instance-centric Attention Network



- Detecting objects with Faster R-CNN
- Predicting action scores with object/human/pairwise streams
- Fusing score to produce final predictions

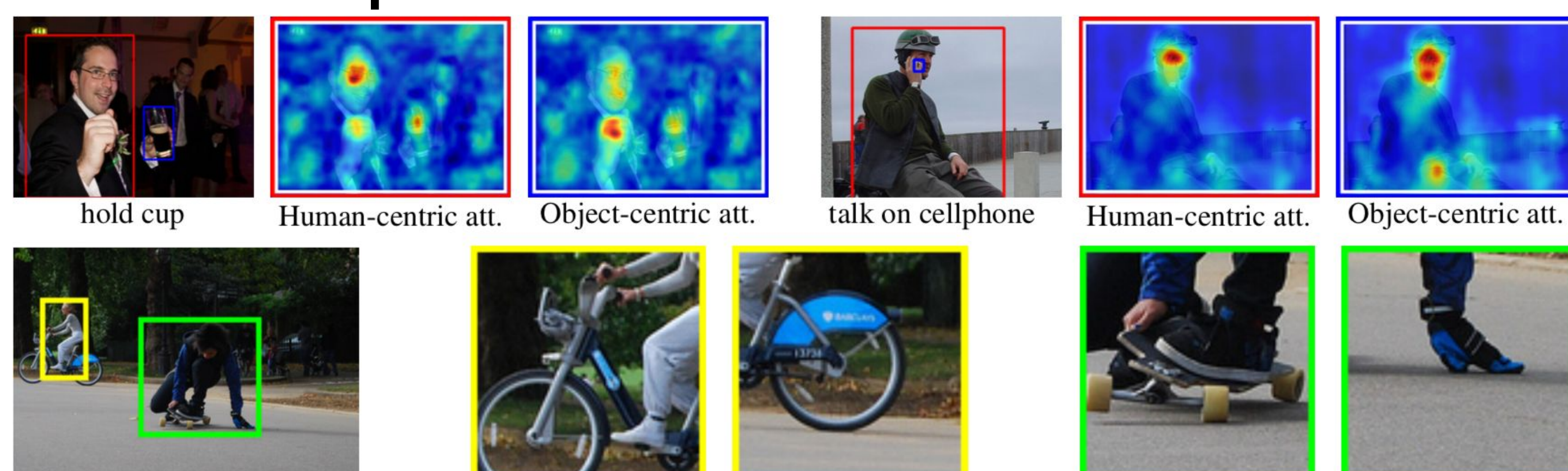
Instance-centric Attention Module

Core idea: Appearance of an instance provides informative cues on where in the image we should pay attention to



- Generating attention map conditioned on instance appearance
- Measuring the similarity in embedding space
- Using instance-centric attentional feature to complement the instance appearance feature

Attention map visualization



Experimental Results

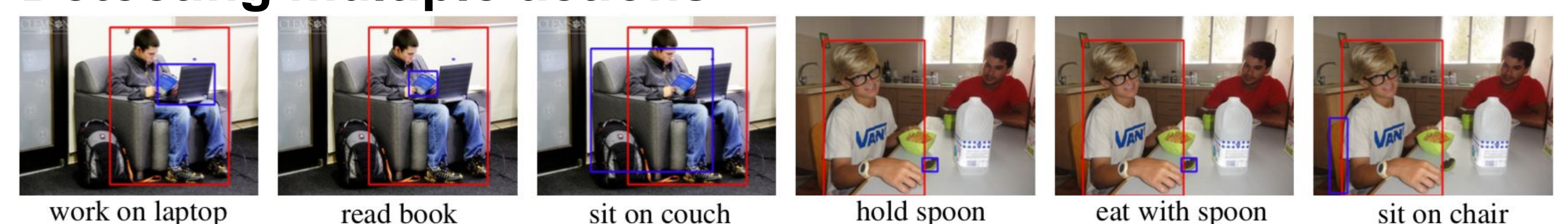
Qualitative results on V-COCO



Qualitative results on HICO-DET



Detecting multiple actions



Quantitative results on V-COCO

Method	Feature backbone	AP_{role}
Model C of [Gupta et al. 2015]	ResNet-50-FPN	31.8
InteractNet [Gkioxari et al. 2018]	ResNet-50-FPN	40.0
BAR-CNN [Kolesnikov et al. 2018]	Inception-ResNet	41.1
iCAN (ours) w/ late fusion	ResNet-50	<u>44.7</u>
iCAN (ours) w/ early fusion	ResNet-50	45.3

Quantitative results on HICO-DET

Method	Feature backbone	Default			Known Object		
		Full	Rare	Non Rare	Full	Rare	Non Rare
Zero-shot [Shen et al. 2018]	VGG-19	6.46	4.24	7.12	-	-	-
HO-RCNN [Chao et al. 2017]	CaffeNet	7.81	5.37	8.54	<u>10.41</u>	<u>8.94</u>	<u>10.85</u>
InteractNet [Gkioxari et al. 2018]	ResNet-50-FPN	9.94	7.16	10.77	-	-	-
iCAN (ours)	ResNet-50	14.84	10.45	16.15	16.26	11.33	17.73

Ablation study: mAP vs. time/model size

