

Learning to Detect Human-Object Interactions

Yu-Wei Chao¹, Yunfan Liu¹, Xieyang Liu¹, Huayi Zeng², Jia Deng¹

¹University of Michigan, Ann Arbor

²Washington University in St. Louis



Detecting Human-Object Interactions (HOI)

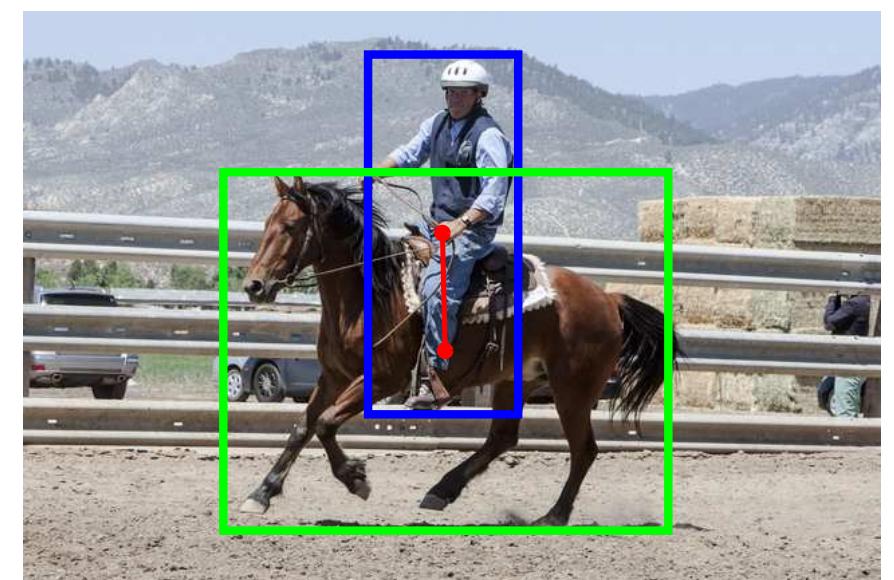
Motivation

1. Recognition of *human-object interactions (HOI)* (e.g. "riding a horse", "eating a sandwich") is an important image understanding problem.
2. Recent work by Chao et al. [1] introduces a new large-scale benchmark *HICO* and studies image-level HOI classification. We seek to extend the task to further detect each HOI instance.

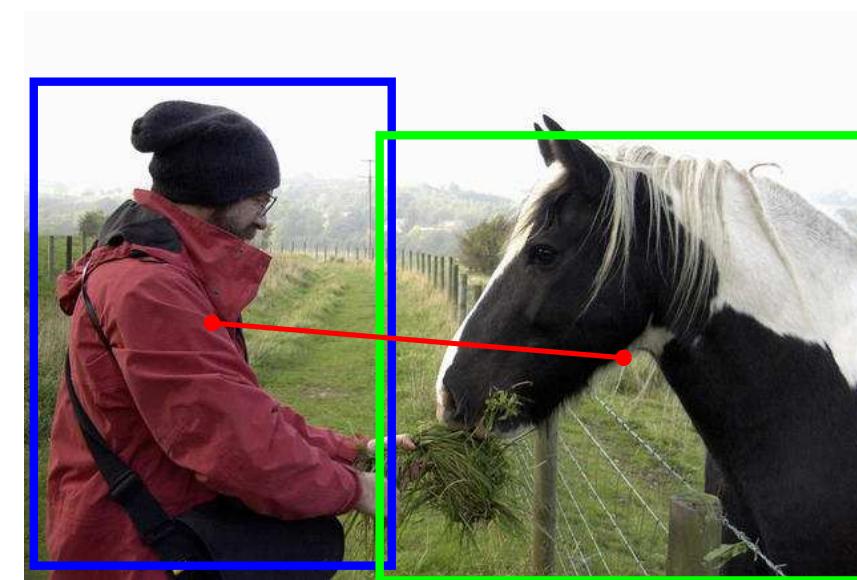
Problem Statement

Each detection instance consists of:

1. A pair of bounding boxes: one for a person (blue) and one for an object (green).
2. An interaction class label.



(a) riding a horse



(b) feeding horses

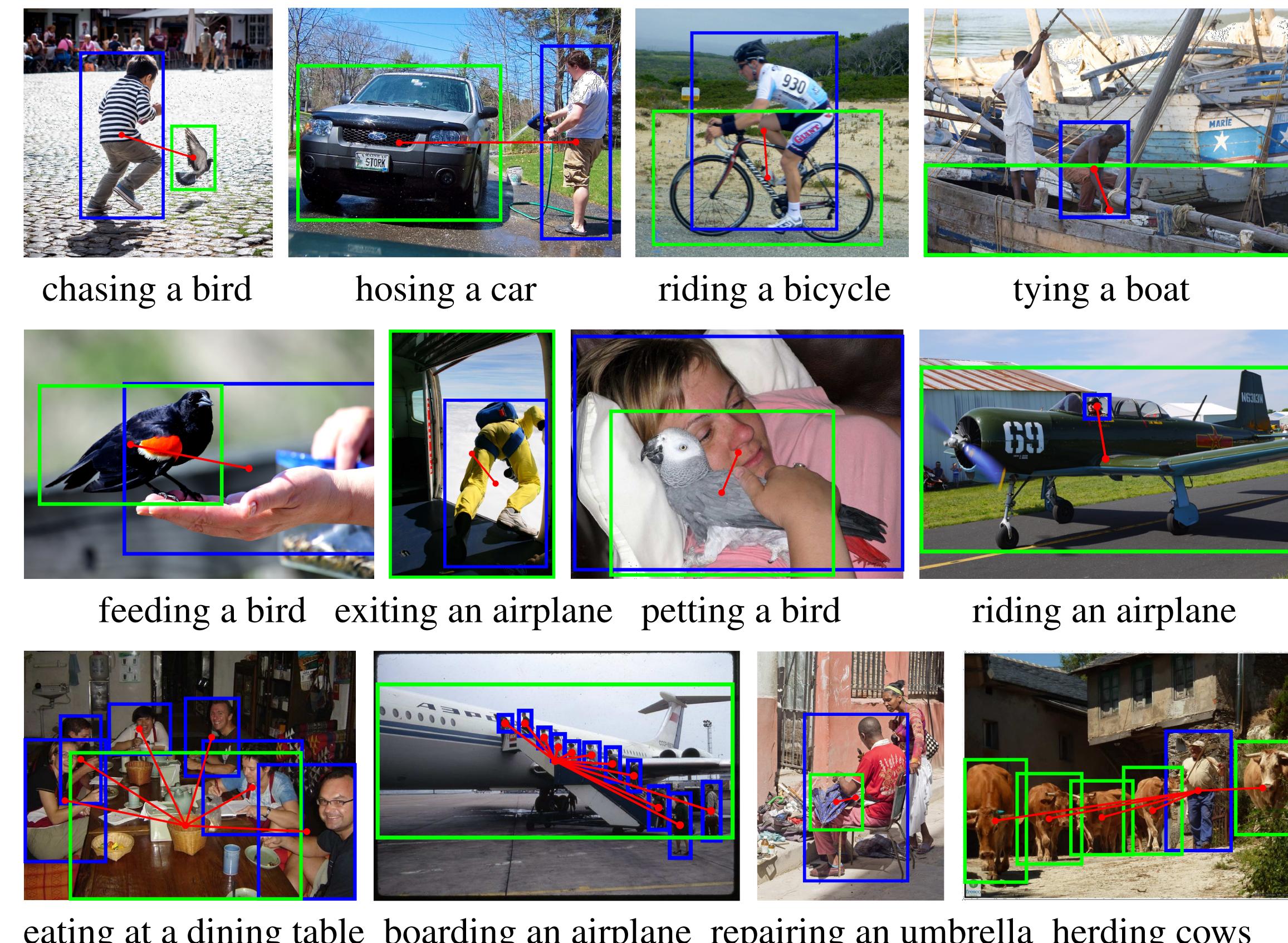
Contributions

1. *HICO-DET*: a new large benchmark for HOI detection.
2. *HO-RCNN*: a new multi-stream DNN-based framework that exploits features from a person, an object, and their spatial relations.

HICO-DET: A New Benchmark for HOI Detection

We augment *HICO* [1] with instance annotations

Sample Images and Annotations



Dataset Statistics

	HICO-DET			
	#image	#positive	#instance	#bounding box
Train	38118	70373	117871 (1.67/pos)	199733 (2.84/pos)
Test	9658	20268	33405 (1.65/pos)	56939 (2.81/pos)
Total	47776	90641	151276 (1.67/pos)	256672 (2.83/pos)

HO-RCNN

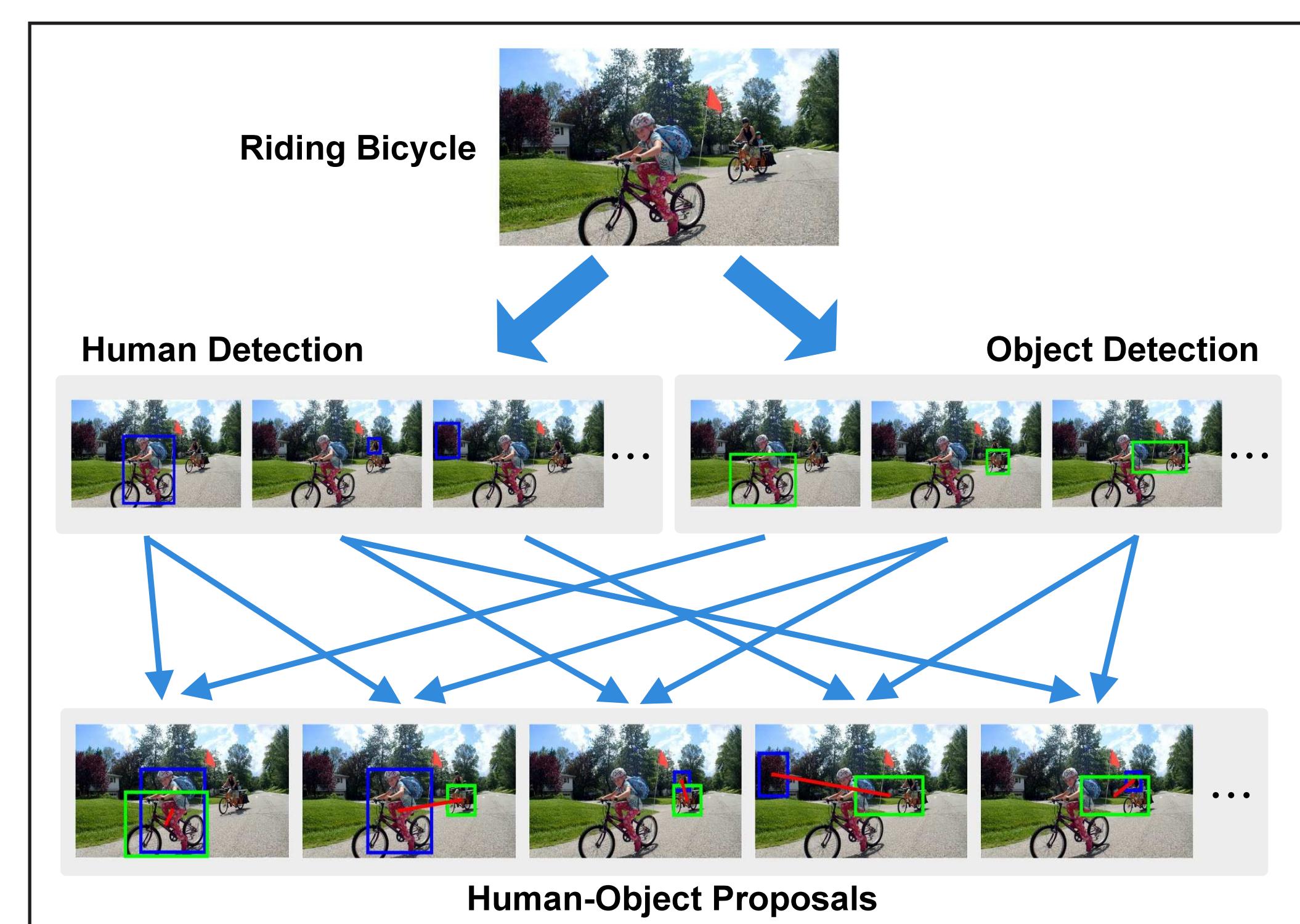
A two-stage framework inspired by the region-based object detectors:

1. Generating human-object proposals
2. Classifying HOI category for each proposal

Note the differences to object detection:

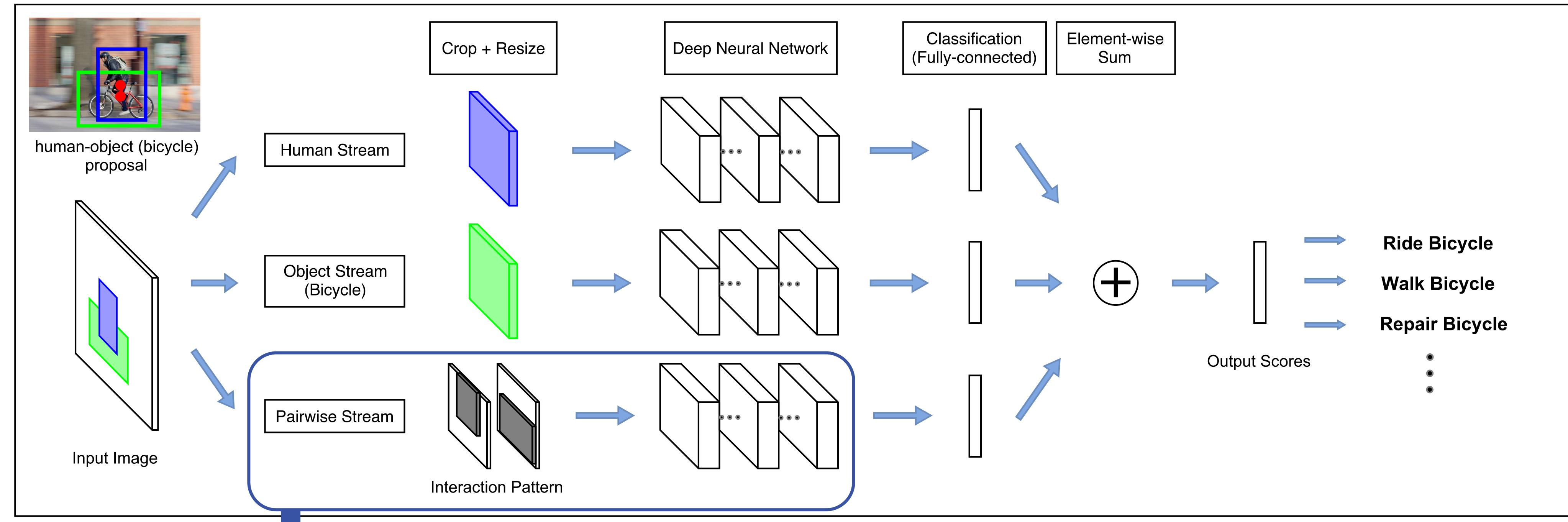
1. Each proposal is a pair of bounding boxes instead of a single one.
2. We classify the HOI category instead of the object category

1. Generating Human-Object Proposals



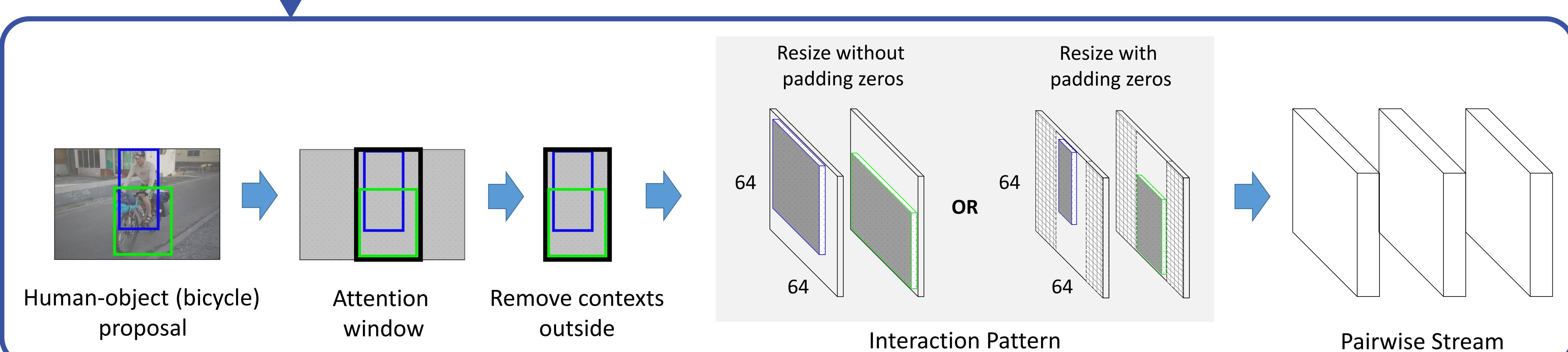
HO-RCNN: A New Multi-Stream DNN-based Framework

2. Classifying HOI Category for Each Proposal



Pairwise Stream: extracting features for human-object spatial relations

Interaction Pattern: a novel DNN input characterizing the spatial relations between two bounding boxes.



Experimental Results

Evaluation Metric: mean Average Precision (mAP)

- Define the overlap between a prediction and ground truth as the minimum of the overlap on human and the overlap on object.
- Declare a true positive if the overlap > 0.5

Evaluation Settings

1. **Known Object (KO)**: for each HOI category, evaluate only on the images containing the associated object category.
2. **Default**: for each HOI category, evaluate on the full test set.

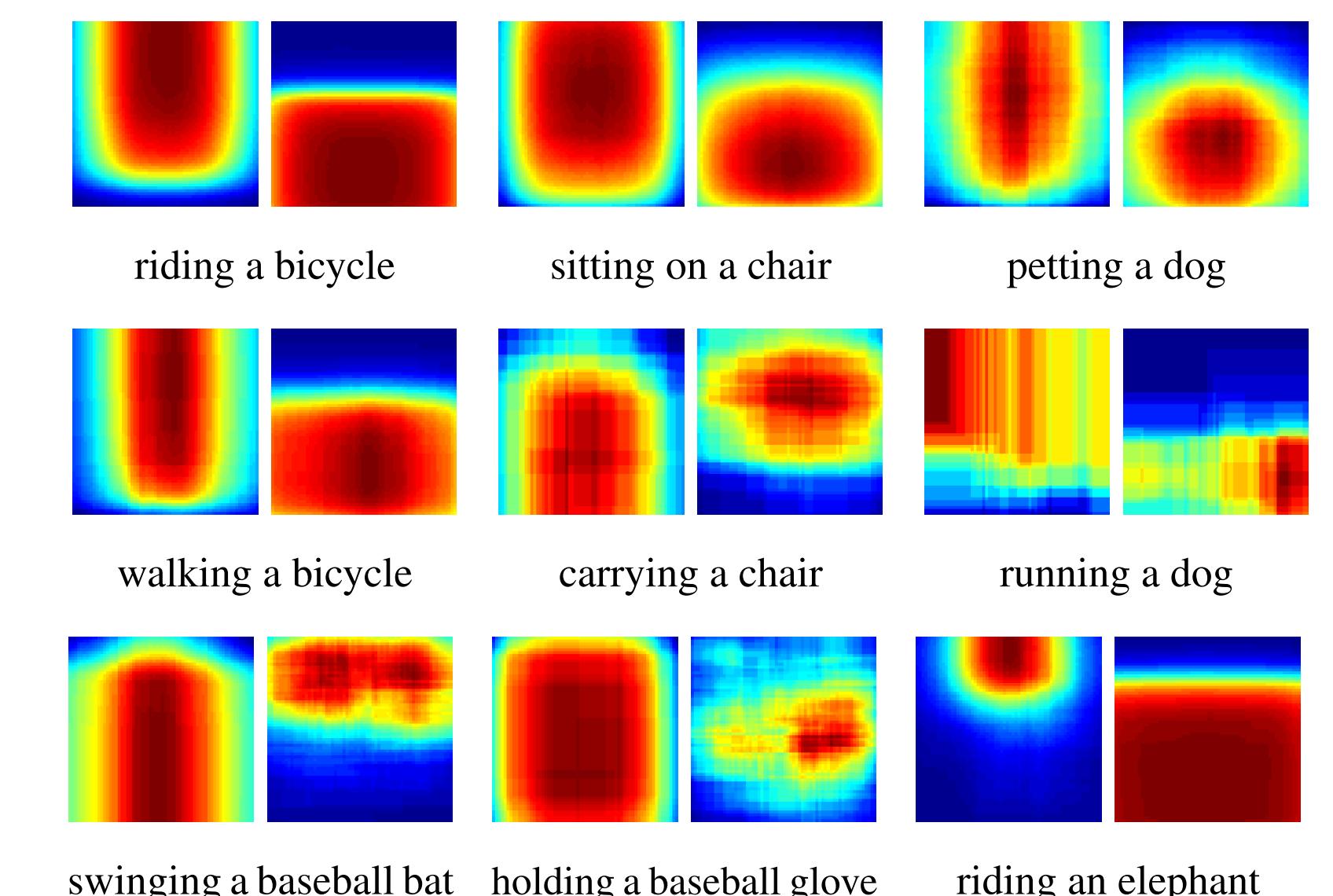
Ablation Study on the Pairwise Stream

Using Interaction Pattern achieves the highest mAP.

	Default			Known Object		
	Full	Rare	Non-Rare	Full	Rare	Non-Rare
HO	5.73	3.21	6.48	8.46	7.53	8.74
HO+vec0 (fc)	6.47	3.57	7.34	9.32	8.19	9.65
HO+vec1 (fc)	6.24	3.59	7.03	9.13	8.09	9.45
HO+IP0 (fc)	7.07	4.06	7.97	10.10	8.38	10.61
HO+IP1 (fc)	6.93	3.91	7.84	10.07	8.43	10.56
HO+IP0 (conv)	7.15	4.47	7.95	10.23	8.85	10.64
HO+IP1 (conv)	7.30	4.68	8.08	10.37	9.06	10.76

Average Interaction Patterns

Left: human channel. Right: object channel.



Leveraging Object Detection Scores

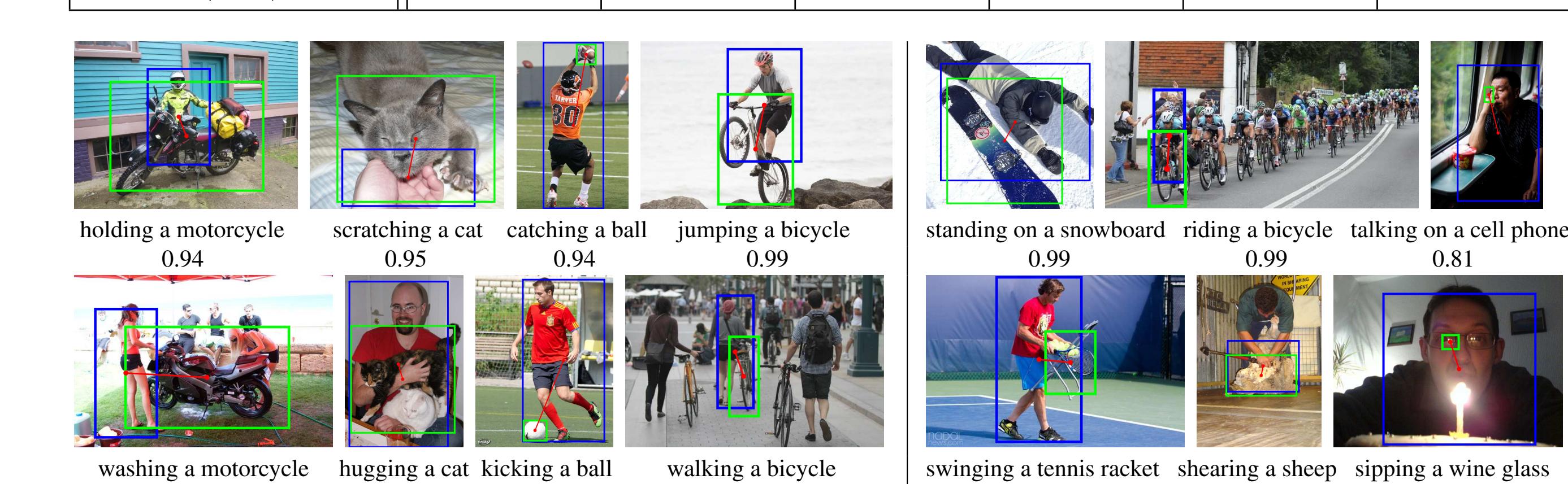
Improves mAP in the Default setting.

	Default			Known Object		
	Full	Rare	Non-Rare	Full	Rare	Non-Rare
HO	5.73	3.21	6.48	8.46	7.53	8.74
HO+S	6.07	3.79	6.76	8.09	6.79	8.47
HO+IP1 (conv)	7.30	4.68	8.08	10.37	9.06	10.76
HO+IP1 (conv)+S	7.81	5.37	8.54	10.41	8.94	10.85

Comparison with Prior Approaches

	Default			Known Object		
	Full	Rare	Non-Rare	Full	Rare	Non-Rare
Random	1.35×10^{-3}	5.72×10^{-4}	1.62×10^{-3}	0.19	0.17	0.19
Fast-RCNN [8] (union)	1.75	0.58	2.10	2.51	1.75	2.73
Fast-RCNN [8] (score)	2.85	1.55	3.23	4.08	2.37	4.59
HO+IP1 (conv)	7.30	4.68	8.08	10.37	9.06	10.76
HO+IP1 (conv)+S	7.81	5.37	8.54	10.41	8.94	10.85

Left: human channel. Right: object channel.



References

- [1] Y.-W. Chao, Z. Wang, Y. He, J. Wang, and J. Deng. *HICO*: A benchmark for recognizing human-object interactions in images. In ICCV, 2015.
[8] R. Girshick. Faster R-CNN. In ICCV, 2015.

Dataset and Code

<http://www.umich.edu/~ywchao/hico/>