



DOCK: Detecting Objects by transferring Common-sense Knowledge



ALLEN INSTITUTE
for ARTIFICIAL INTELLIGENCE

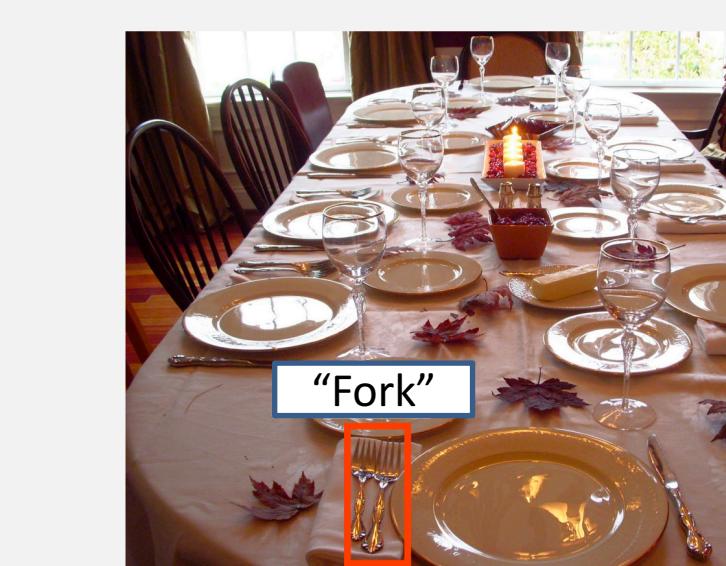
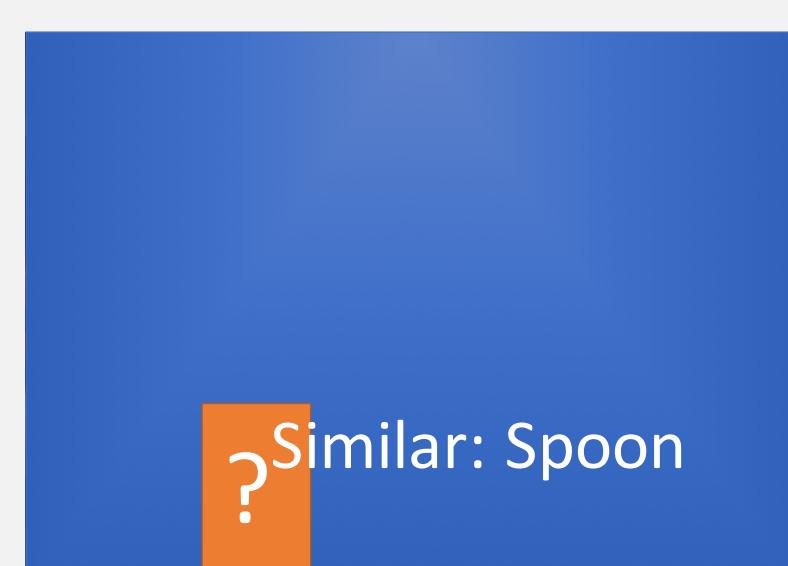
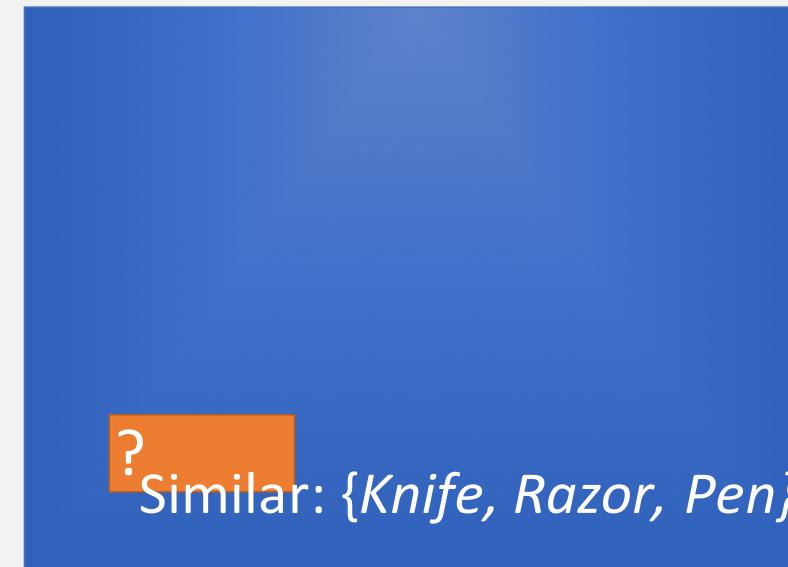
Krishna Kumar Singh

Santosh Divvala

Ali Farhadi

Yong Jae Lee

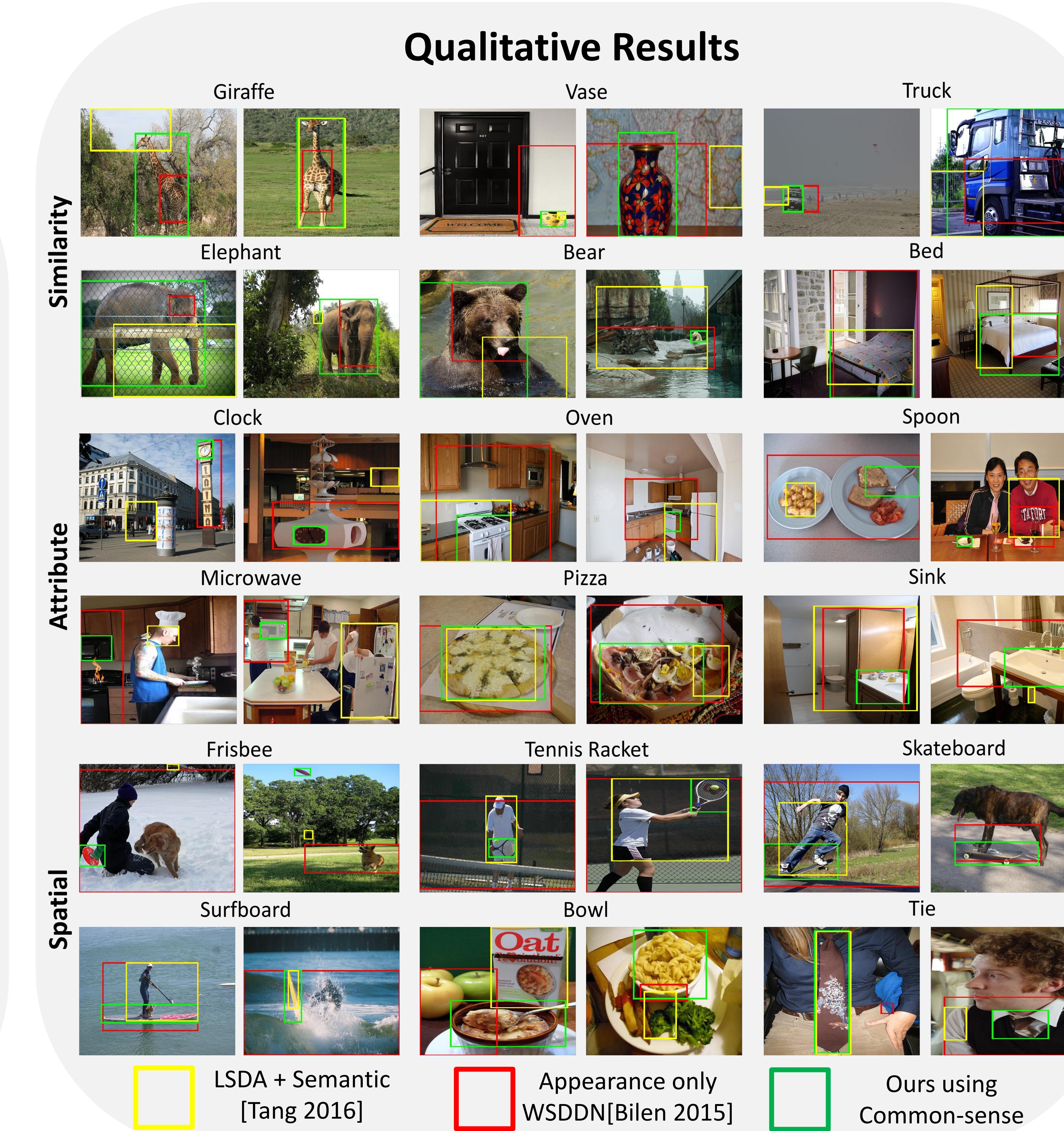
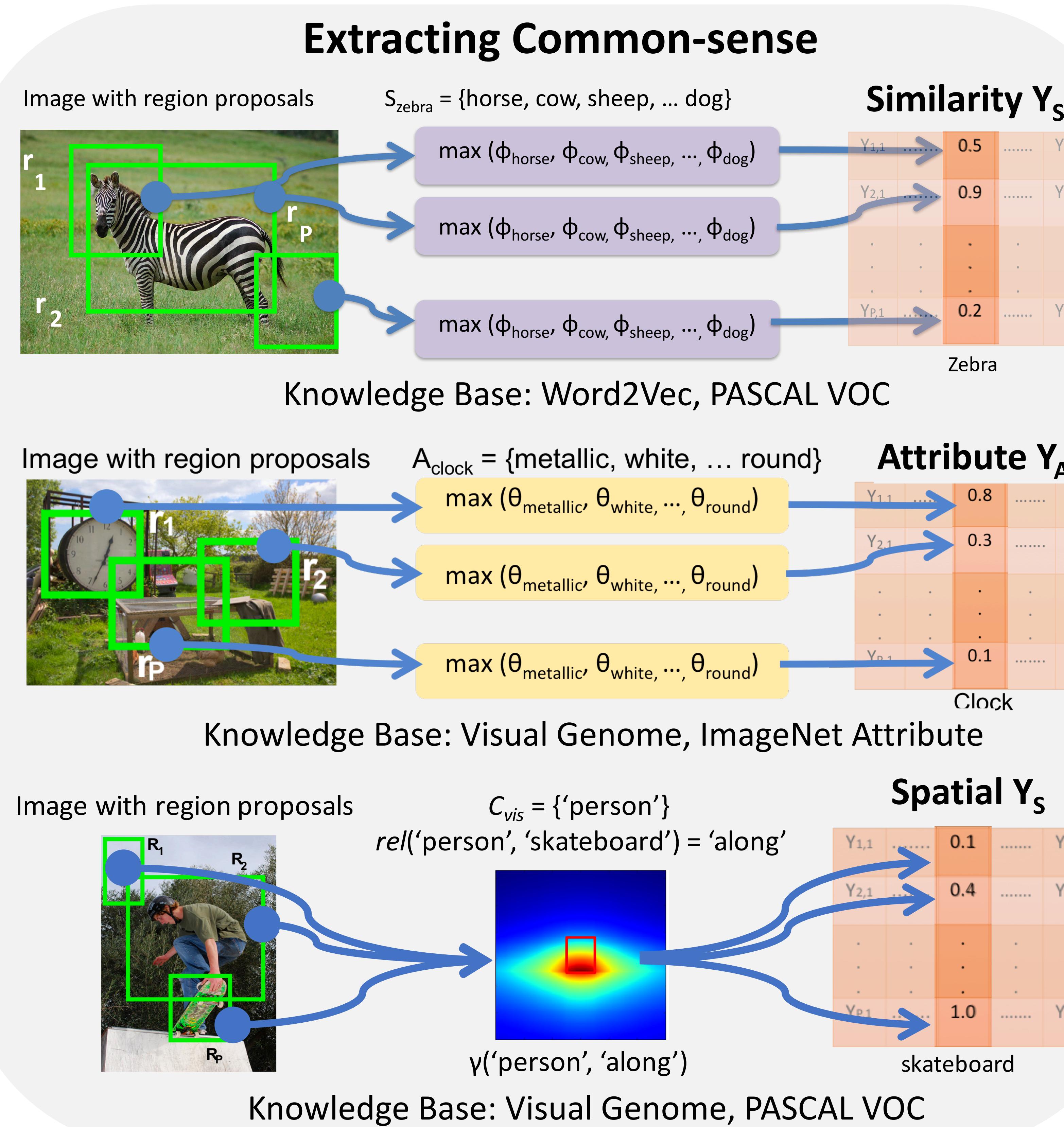
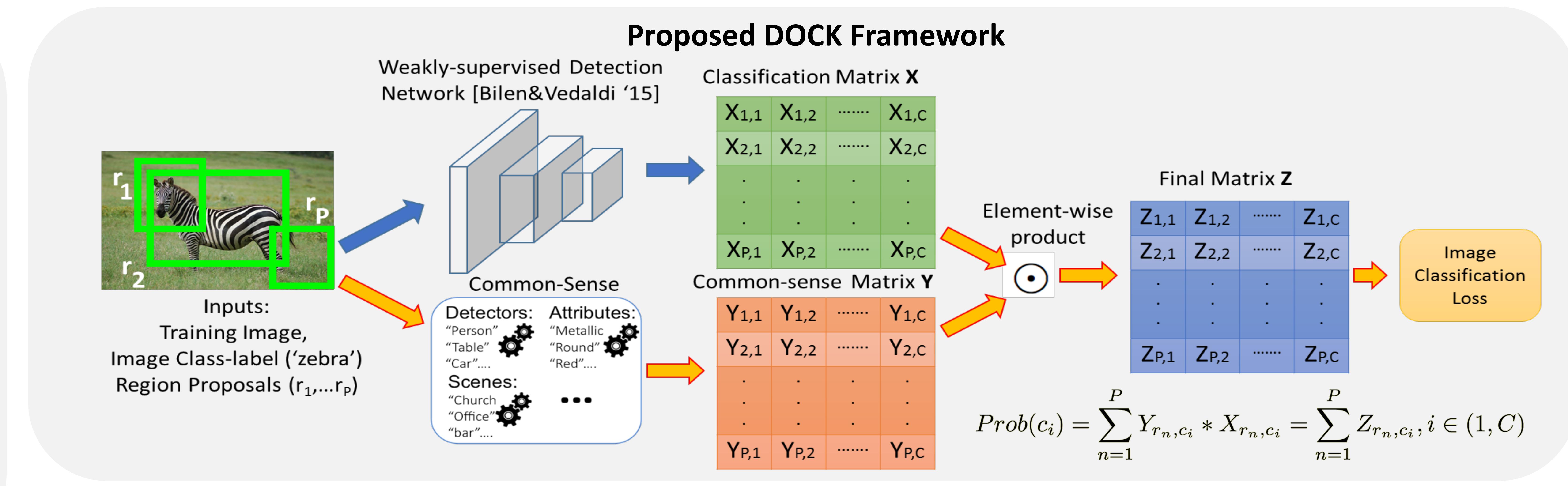
Goal: Scalable transfer learning using readily available Common-Sense knowledge for object detection on real world datasets (e.g. MS COCO)



✓ Beyond similarity common-sense (spatial, attribute, etc.)

✓ Region based modelling

✓ Common-sense used only as prior during training



Quantitative Analysis

Detection results on MS COCO validation set
(Source: 20 classes [PASCAL VOC] Target: remaining 60 classes)

Methods	AP	AP ^S	AP ^M	AP ^L
Classification Network ('No Adapt')	3.4	0.9	2.9	6.1
LSDA [Hoffman 2014]	4.6	1.2	5.1	7.8
LSDA + Sematic [Tang 2016]	4.7	1.1	5.1	8.0
LSDA+MIL [Hoffman 2015]	5.9	1.5	8.3	10.7
Fine-tuned Detection Network	10.8	1.2	8.9	18.6
Ours (DOCK)	14.4	2.0	12.8	24.9

- Our approach of using common-sense leads to best detection results
- Other transfer learning approaches only rely on image-level similarity

Common-sense matrix (\mathbf{Y}) influence

Training	Testing	AP
No	No	11.7
No	Yes	11.8
Yes	No	14.4
Yes	Yes	14.1

Ablation Study

Methods	AP
Base Network	11.7
Base Network + Similarity	13.7
Base Network + Attribute	12.2
Base Network + Spatial	13.0
Base Network + Joint	14.4

Webly-supervised

Methods	AP [MS COCO]	AP [ImageNet Det]
Base Network	6.8	6.2
Ours (DOCK)	8.3	8.8