



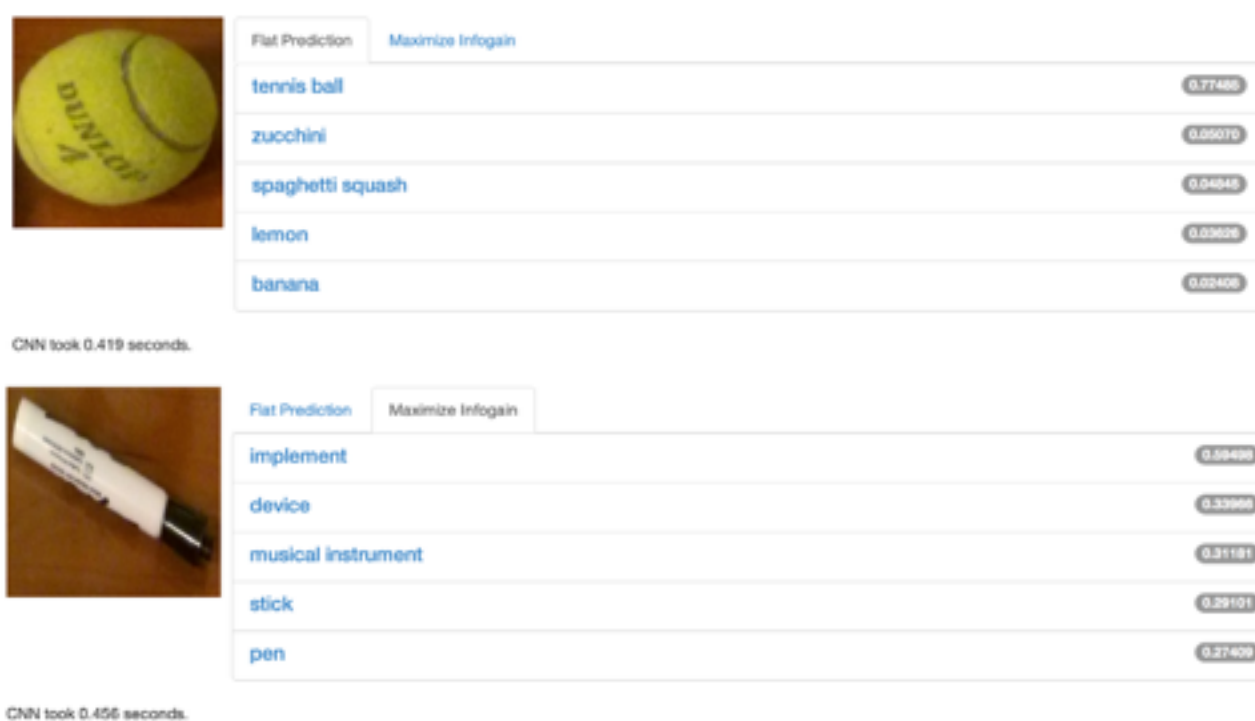
# Open-Vocabulary Object retrieval



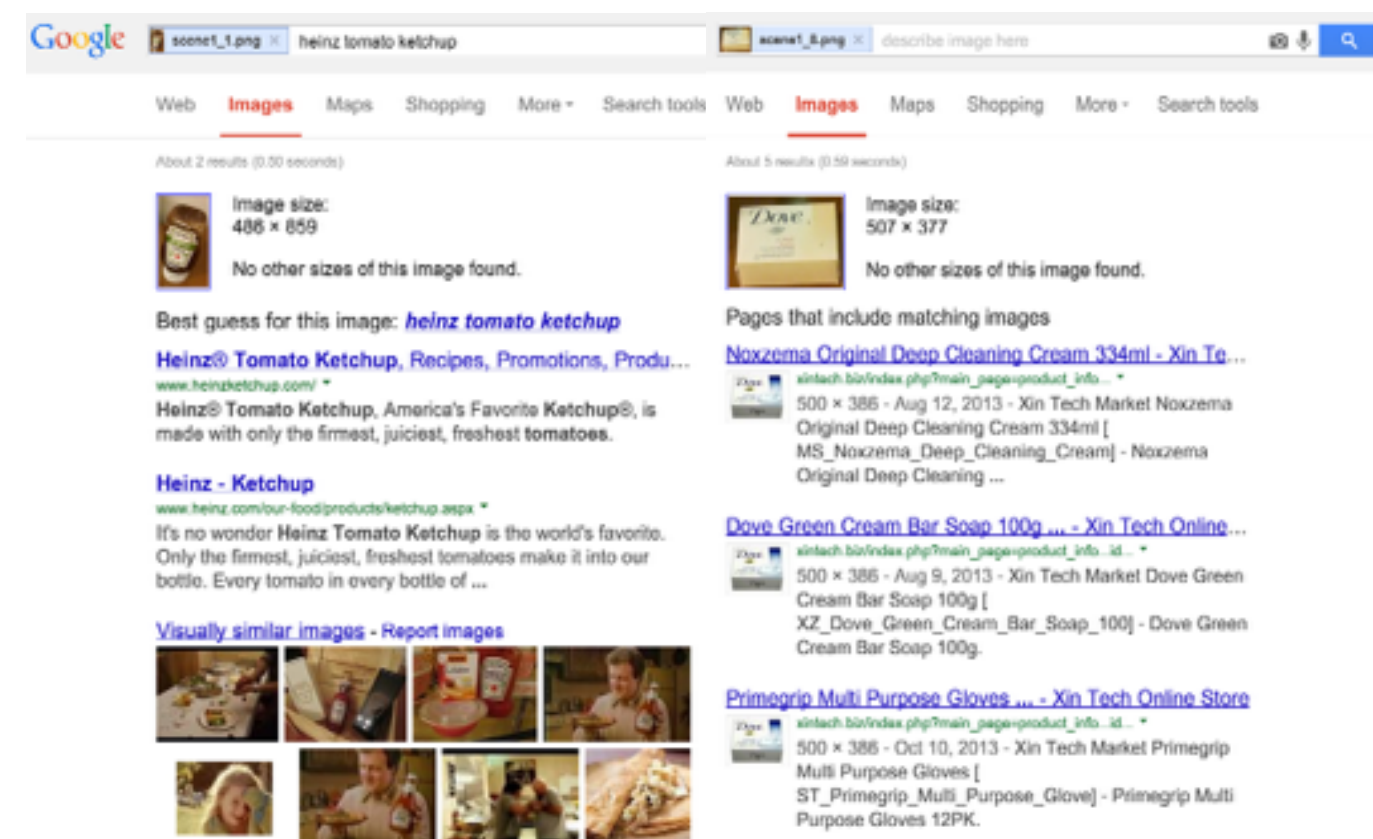
S. Guadarrama<sup>1</sup>, E. Rodner<sup>2</sup>, K. Saenko<sup>3</sup>, N. Zhang<sup>1</sup>, R. Farrell<sup>4</sup>, J. Donahue<sup>1</sup>, T. Darrell<sup>1,5</sup>

## Category Recognition

Brew with Caffe [caffe.berkeleyvision.org](http://caffe.berkeleyvision.org)

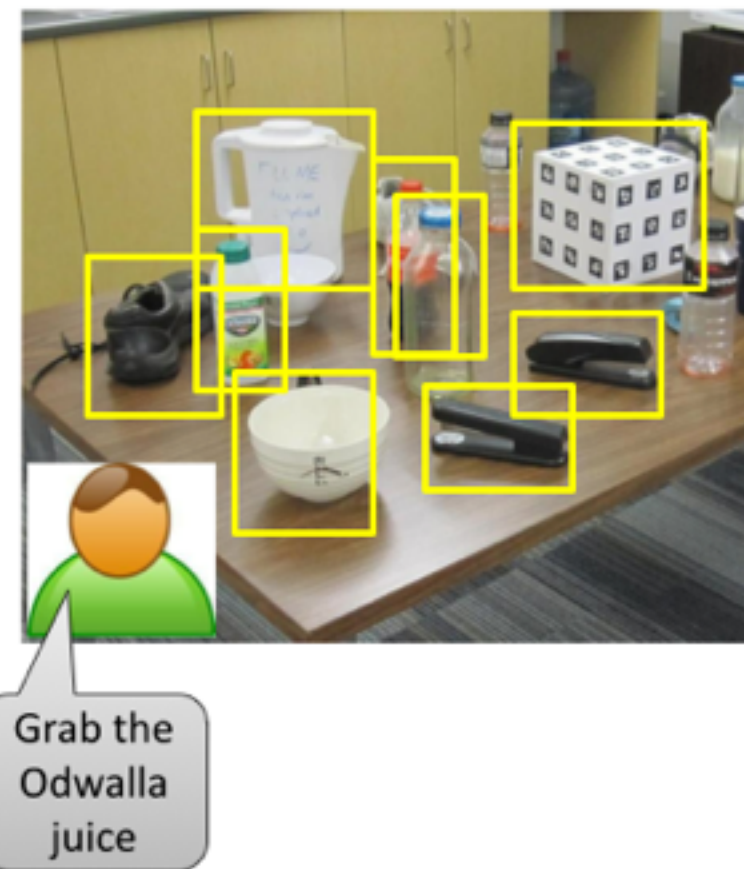


## Instance Matching



## Open-Vocabulary Interface

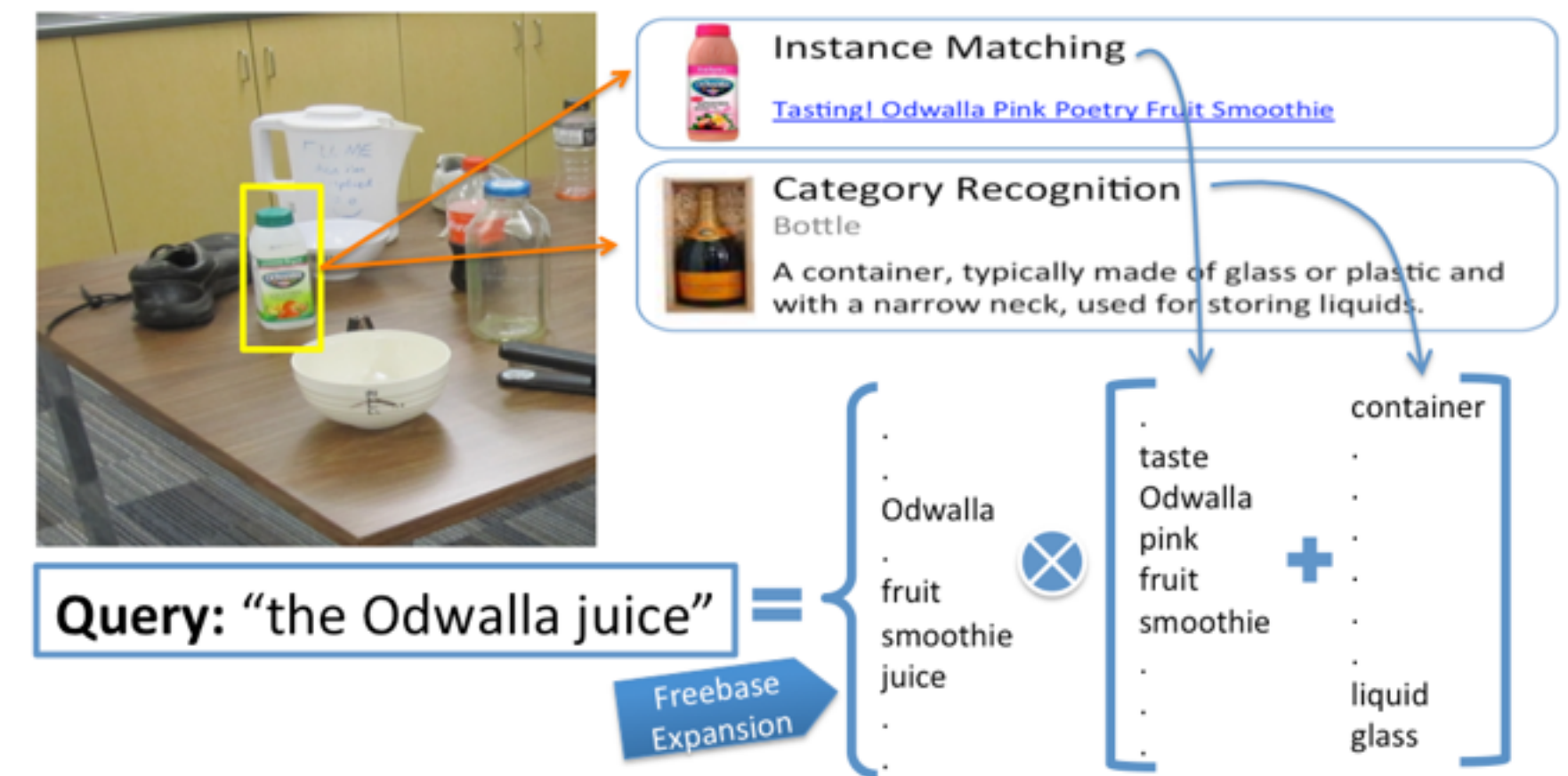
- the **Odwalla** juice
- the **kettle**
- black** shoe
- the **big** bowl
- the **black** stapler
- the **white** box
- the **coke** bottle
- the **empty** bottle
- ...



## Combining Category & Instance



## Projecting into a Text Space



## Results on Household Objects

	Method	Lab	Kitchen	ImageNet
Baselines	MQ-Max [5]	35.26%	40.34%	43.43%
	MQ-Avg [5]	33.40%	38.86%	41.42%
Instance Based	IQ-Engine (IQE)	48.59%	32.85%	24.46%
	Google-Image (GIS)	48.30%	44.45%	39.19%
	DECAF-1K (DEC)	43.36%	50.73%	53.13%
	DEC+FB (DEC+)	42.19%	52.13%	54.05%
Category Based	CAFFE-7K (CAF)	44.70%	51.34%	57.50%
	CAF+FB (CAF+)	42.19%	50.04%	56.82%
	LLC-10K (LLC)	40.05%	49.57%	57.24%
	LLC+FB (LLC+)	37.85%	41.25%	56.27%
Combined	Linear-SVM (LSVM)	45.75%	58.90%	63.65%
	Rank-SVM (RSVM)	56.40%	62.49%	72.62%
	Max-Kernel (MAX)	49.51%	61.11%	68.49%
Cascaded	IQE,GIS	56.37%	51.86%	58.86%
	IQE,GIS,DEC	64.09%	60.95%	75.04%
	IQE,GIS,DEC,CAF	66.45%	64.15%	80.13%
	IQE,GIS,DEC,CAF,LLC	66.76%	65.10%	81.50%
	Full Cascade (CAS)	67.07%	66.20%	81.93%

Code and paper available at <http://openvoc.berkeleyvision.org>