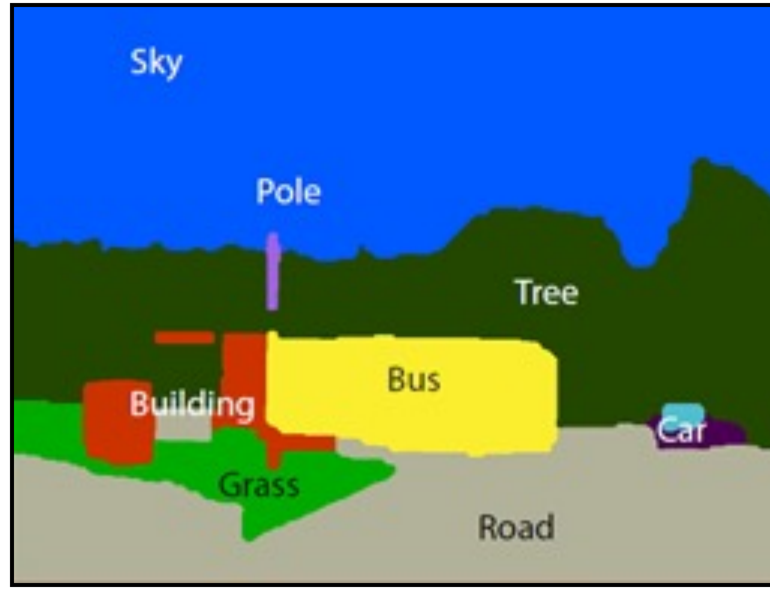
Finding Things: Image Parsing with Regions and Per-Exemplar Detectors



Joseph Tighe

University of North Carolina at Chapel Hill

Svetlana Lazebnik

University of Illinois at Urbana-Champaign

Image parsing

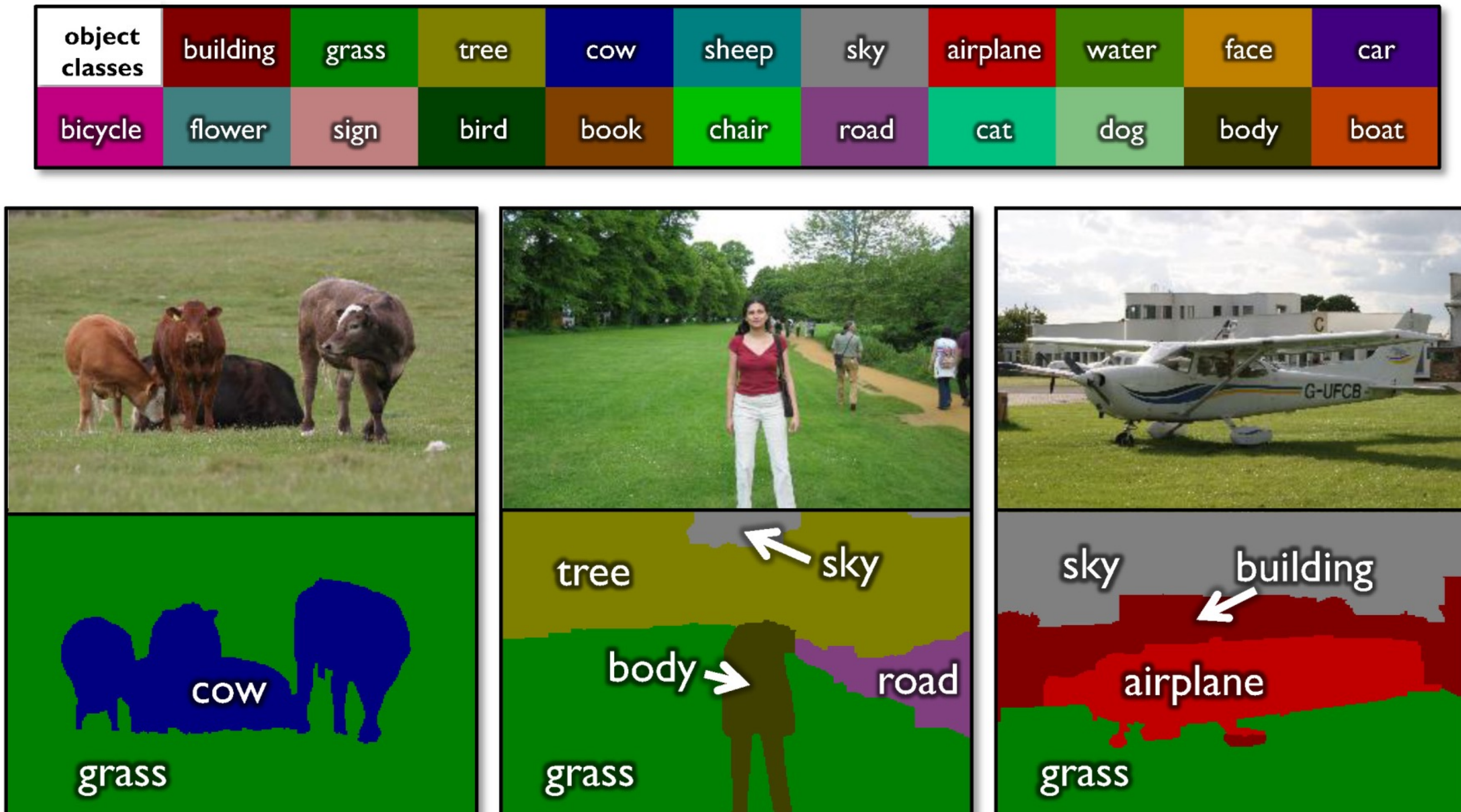
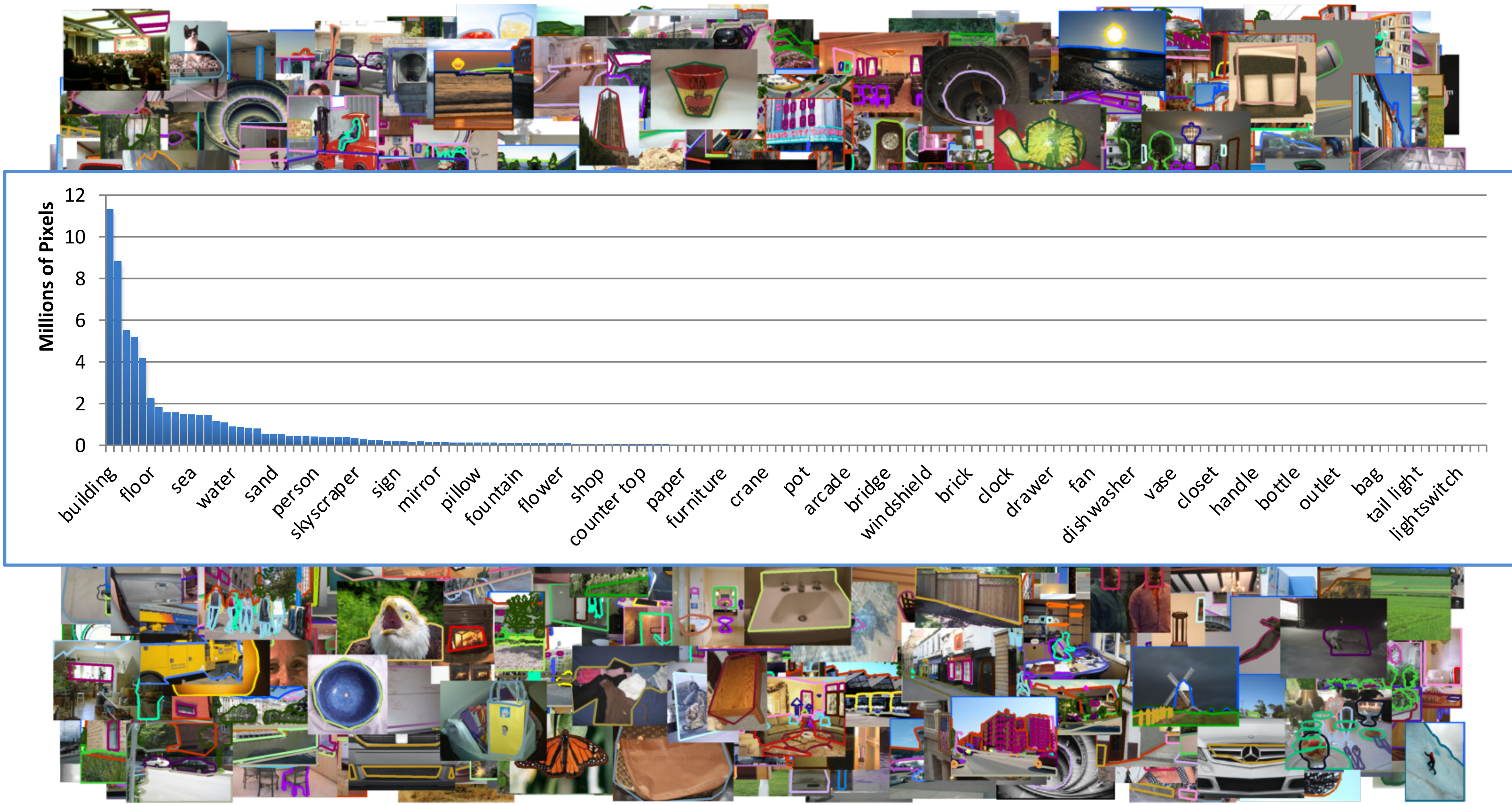


Figure from Shotton et al. (2009)

He et al. (2004), Hoiem et al. (2005), Shotton et al. (2006, 2008, 2009), Verbeek and Triggs (2007), Rabinovich et al. (2007), Galleguillos et al. (2008), Brostow et al. (2008), Gould et al. (2009), Sturgess et al. (2009), Zhang et al. (2010), Ladicky et al. (2010), Liu et al. (2011), Floros et al. (2011), Farabet et al. (2012), Eigen and Fergus (2012), Myeong et al. (2012)

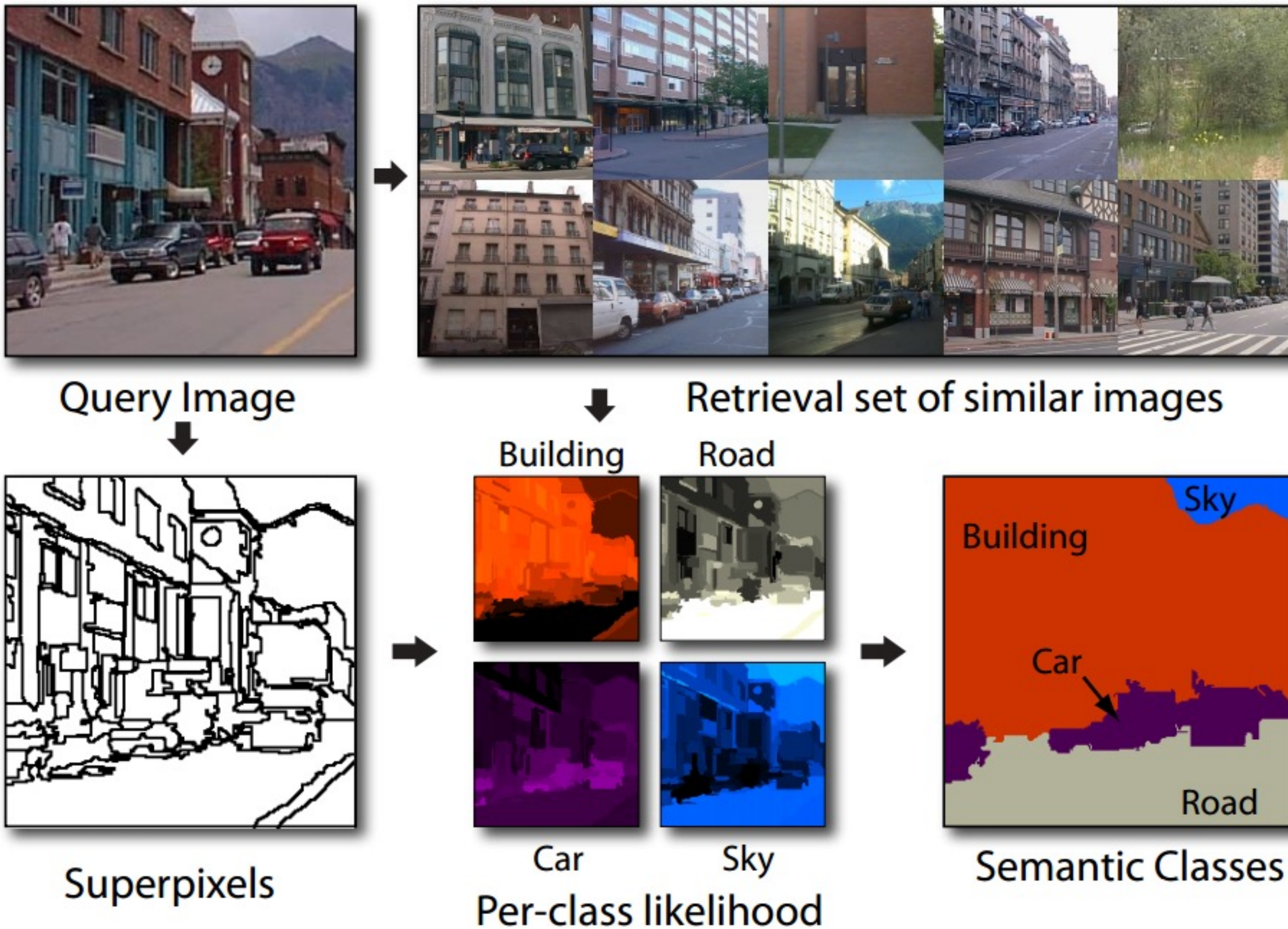
Towards broader coverage

Hundreds of classes and tens of thousands of images



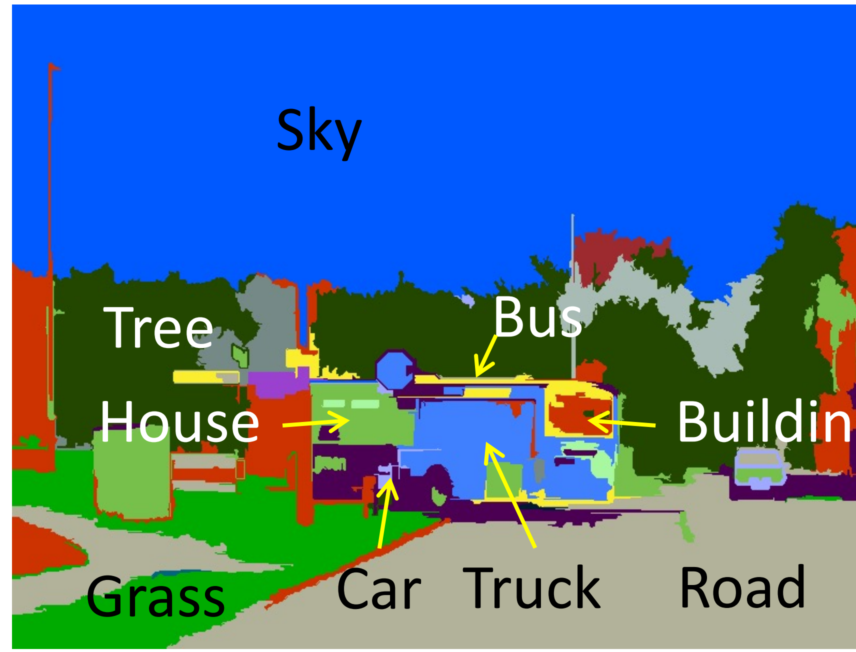
http://labelme.csail.mit.edu/

Our previous work region-based parsing

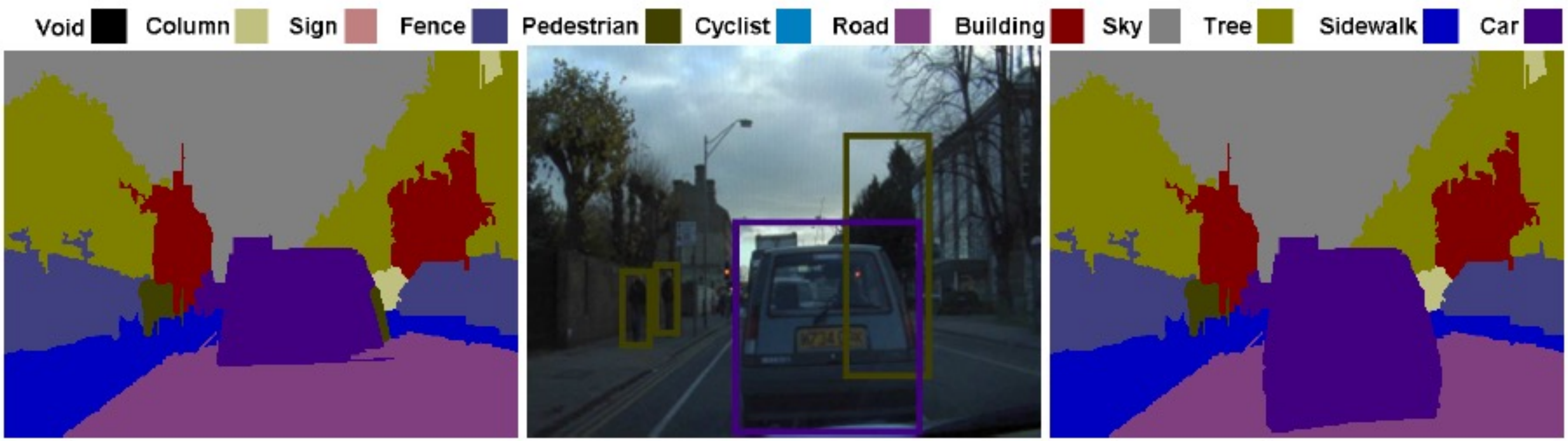


SuperParsing: Scalable Nonparametric Image Parsing with Superpixels
J. Tighe and S. Lazebnik, ECCV 2010, IJCV 2013

Finding Things



To get the things, use detectors



Result without detections

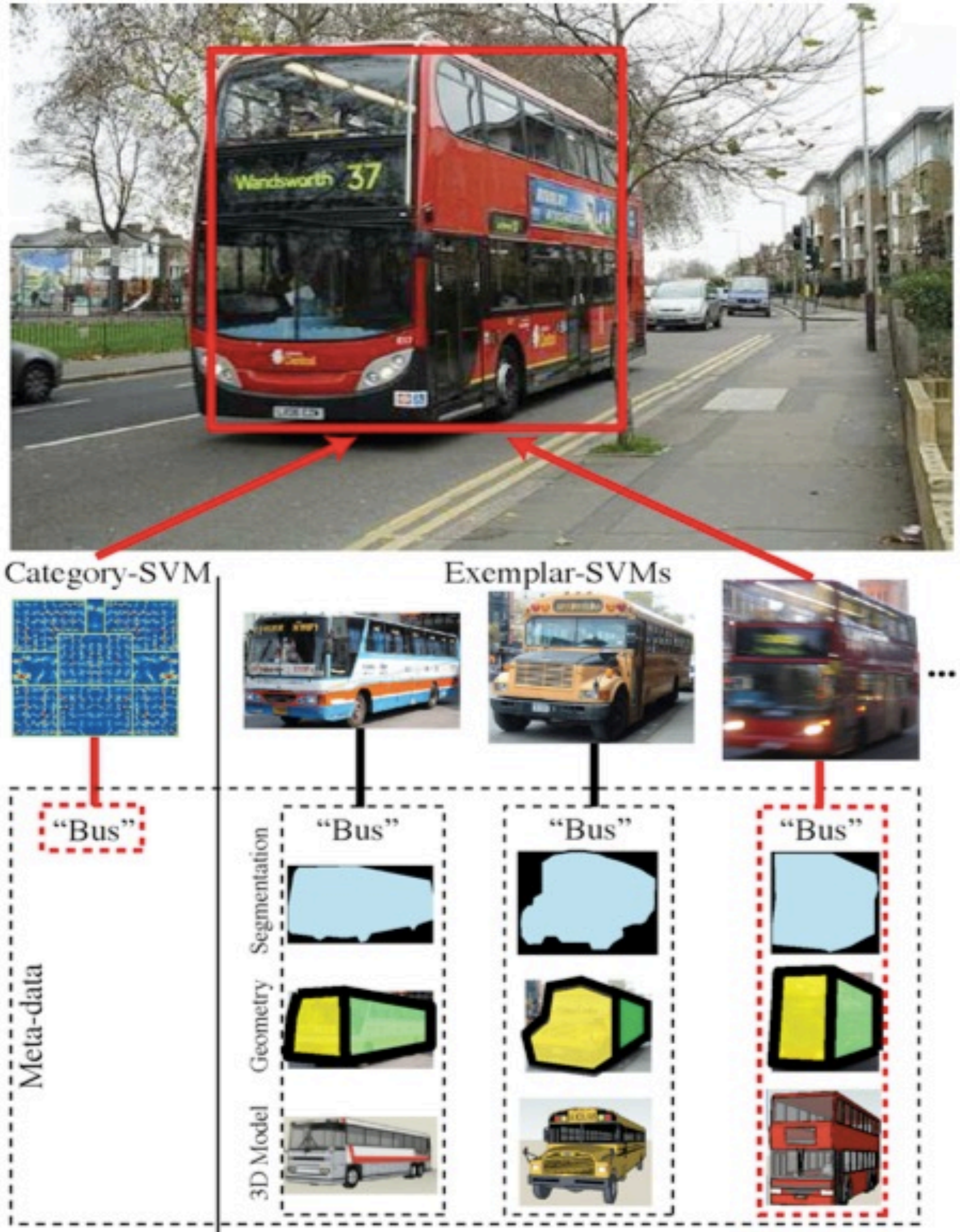
Set of detections

Final Result

Ľubor Ladický, Paul Sturgess, Karteek Alahari, Chris Russell, Philip H.S. Torr
What, Where & How Many? Combining Object Detectors and CRFs. In ECCV 2010

Problems with standard sliding window detectors

- They return only bounding box hypotheses, and obtaining segmentation hypotheses from them is challenging
- They do not work well for classes with few training examples and large intra-class variation

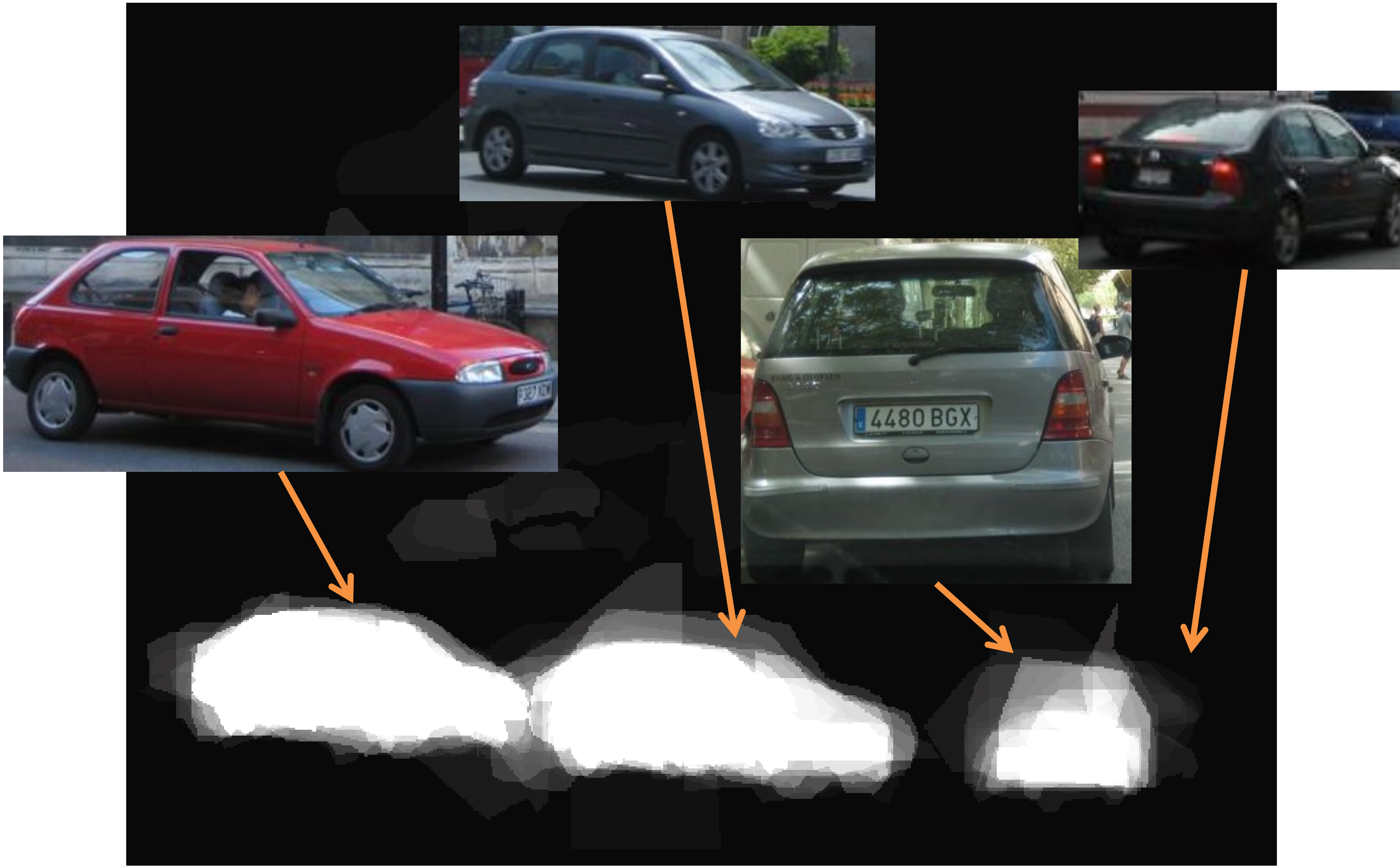


Tomasz Malisiewicz, Abhinav Gupta, Alexei A. Efros
Ensemble of Exemplar-SVMs for Object Detection and Beyond. In ICCV 2011

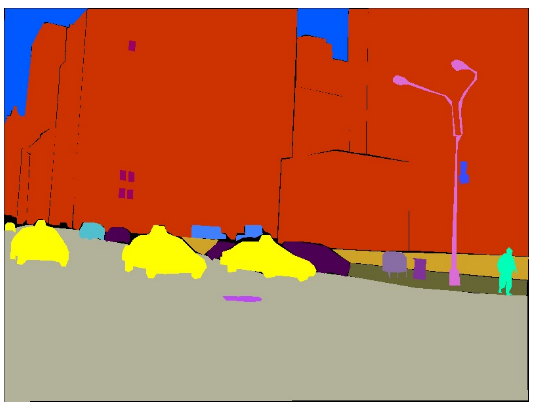
Our approach

Test image

Detector-based data term

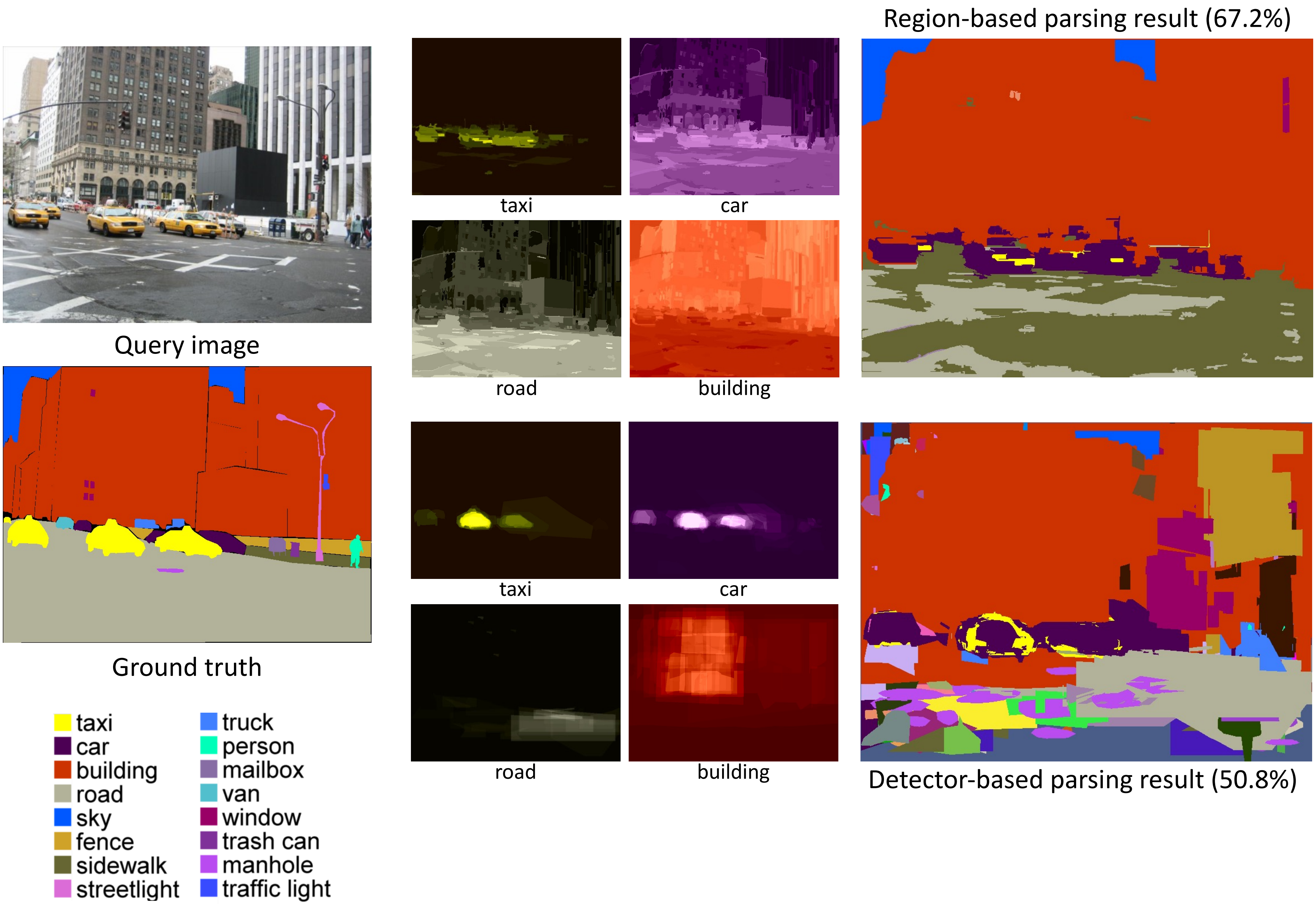


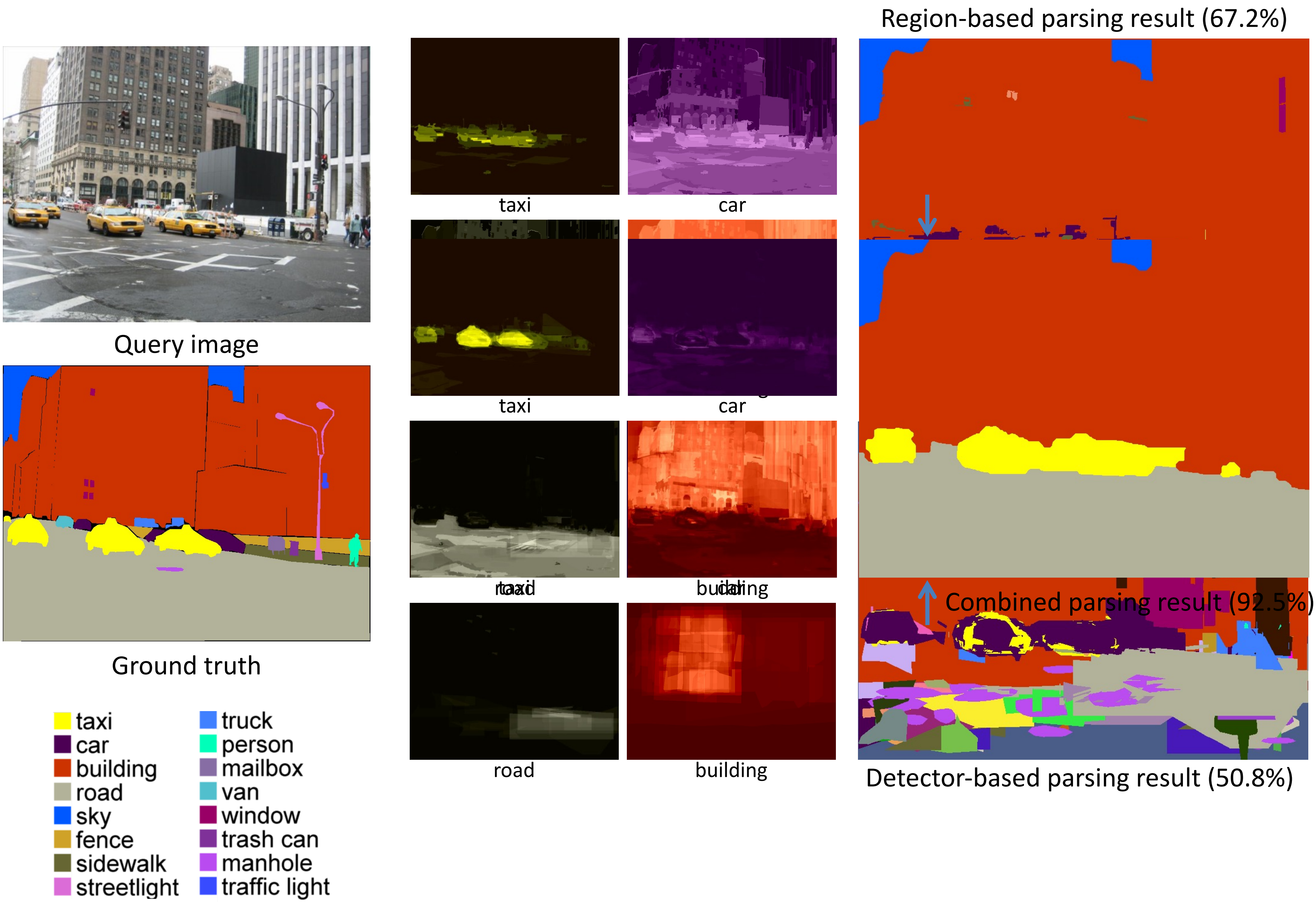
Query image



Ground truth

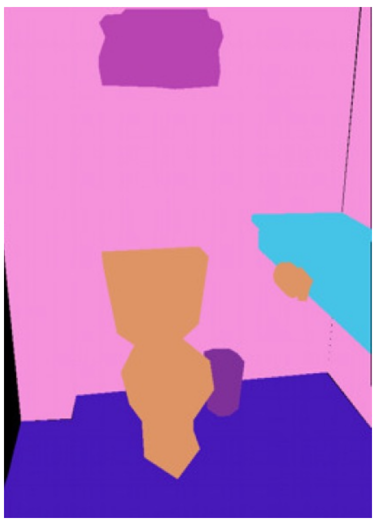
taxi
car
building
road
sky
fence
sidewalk
streetlight
truck
person
mailbox
van
window
trash can
manhole
traffic light


Region-based parsing result (67.2%)
Query image
taxi
car
road
building
Ground truth
taxi
car
road
building
Detector-based parsing result (50.8%)
taxi
car
building
road
sky
fence
sidewalk
streetlight
truck
person
mailbox
van
window
trash can
manhole
traffic light

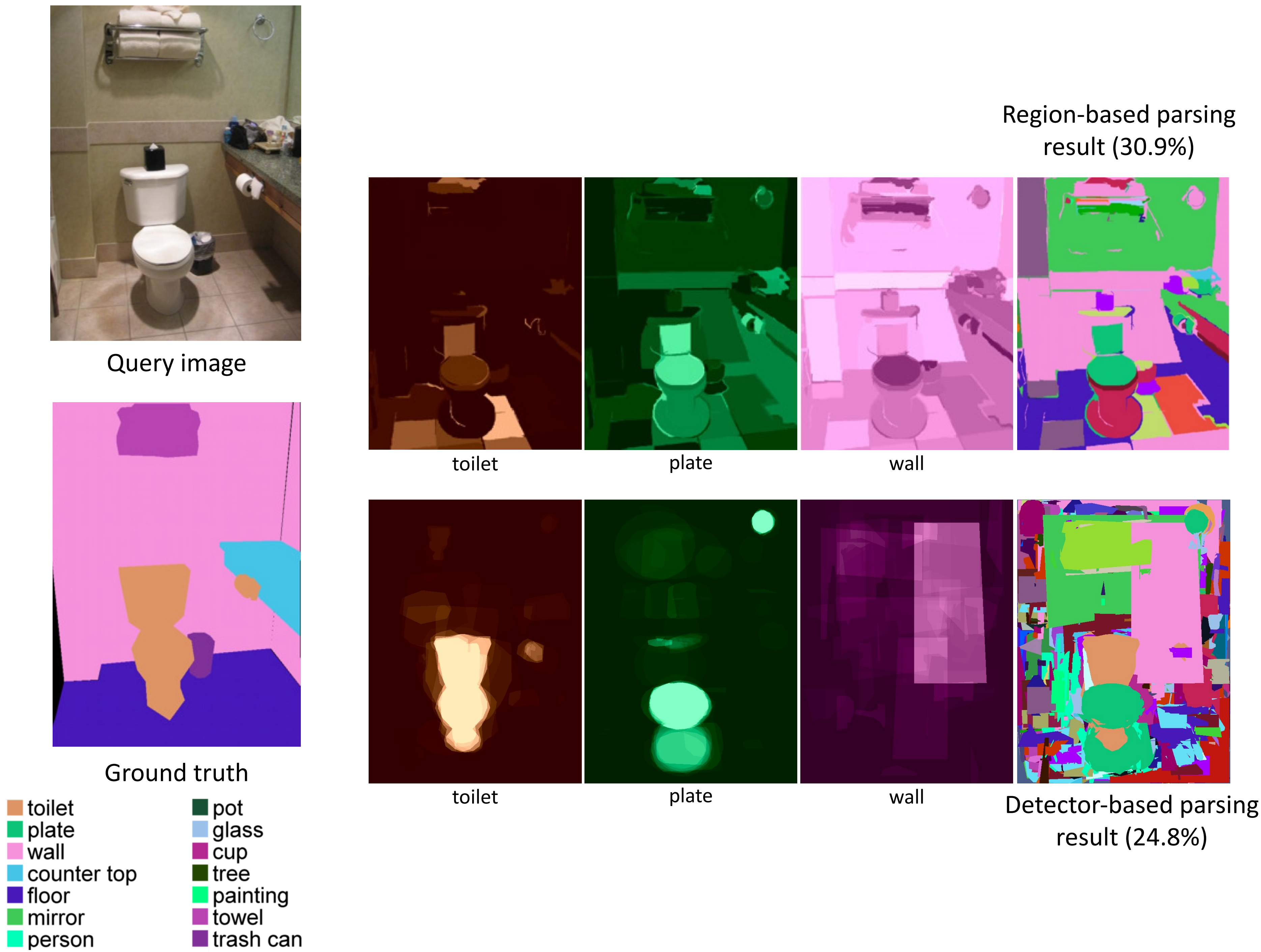


Region-based parsing result (67.2%)
Query image
Ground truth
taxi
car
taxi
car
road
taxi
building
road
building
Combined parsing result (92.5%)
Detector-based parsing result (50.8%)
taxi
car
building
road
sky
fence
sidewalk
streetlight
truck
person
mailbox
van
window
trash can
manhole
traffic light

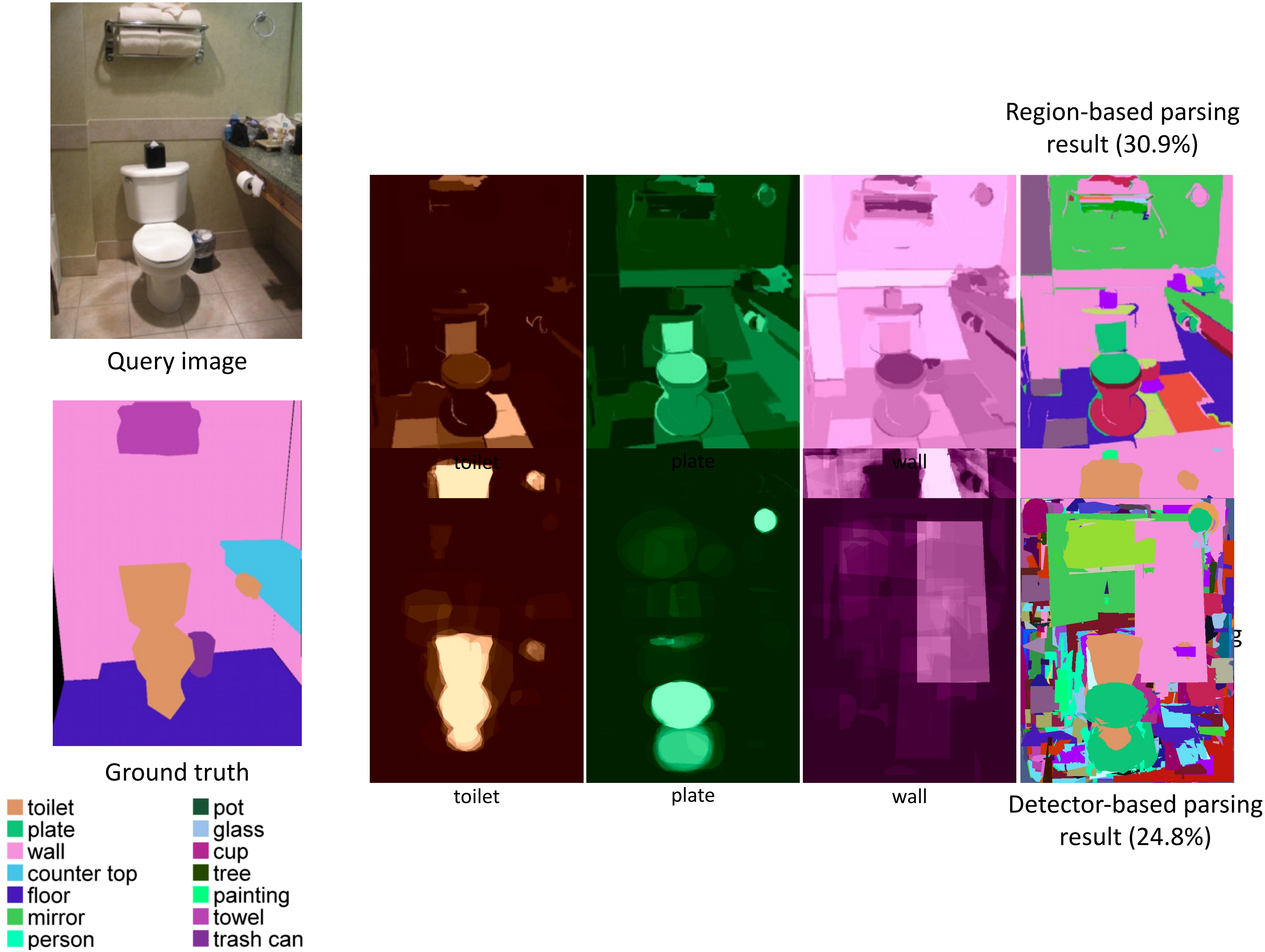
Query image



Ground truth

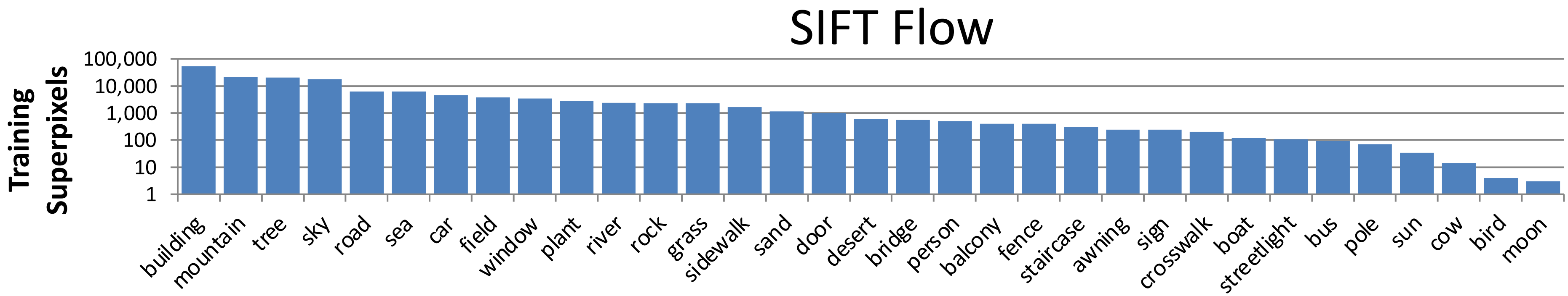


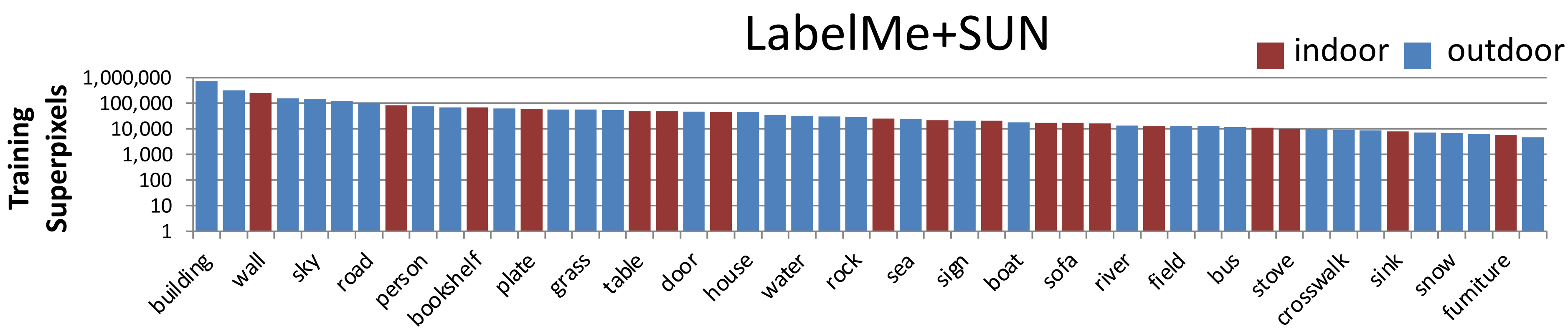
Query image
Region-based parsing result (30.9%)
toilet
plate
wall
Ground truth
toilet
plate
wall
Detector-based parsing result (24.8%)
toilet
plate
wall
counter top
floor
mirror
person
pot
glass
cup
tree
painting
towel
trash can


Region-based parsing result (30.9%)
Query image
toilet
plate
wall
Ground truth
toilet
plate
wall
Detector-based parsing result (24.8%)
toilet
plate
wall
counter top
floor
mirror
person
pot
glass
cup
tree
painting
towel
trash can

Evaluation: Datasets

	Training Images	Test Images	Labels
SIFT Flow (Liu et al., 2009)	2,488	200	33
LabelMe+SUN	45,176	500	232

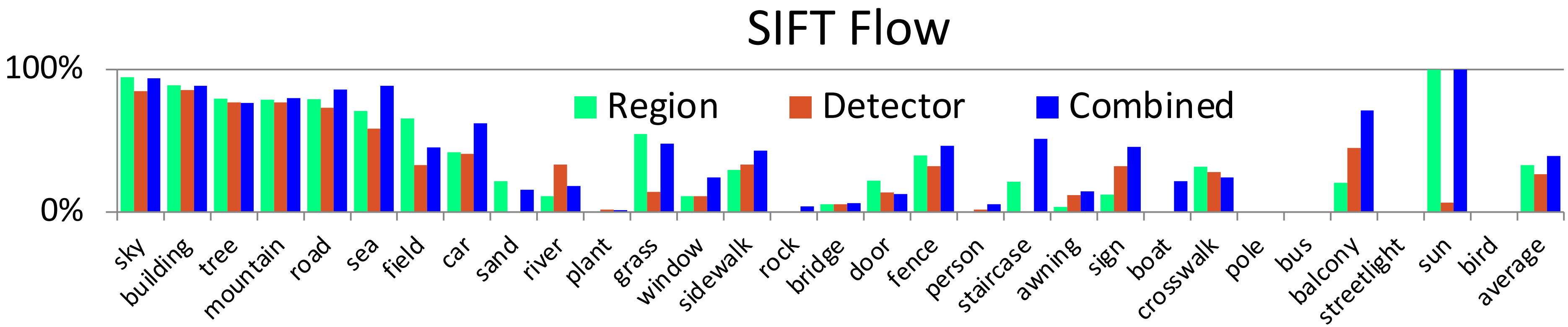


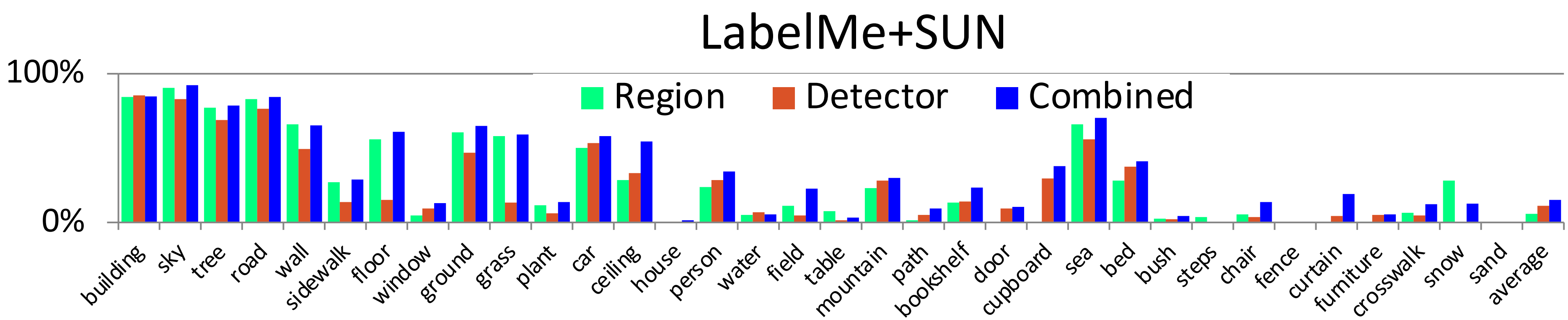


Quantitative evaluation

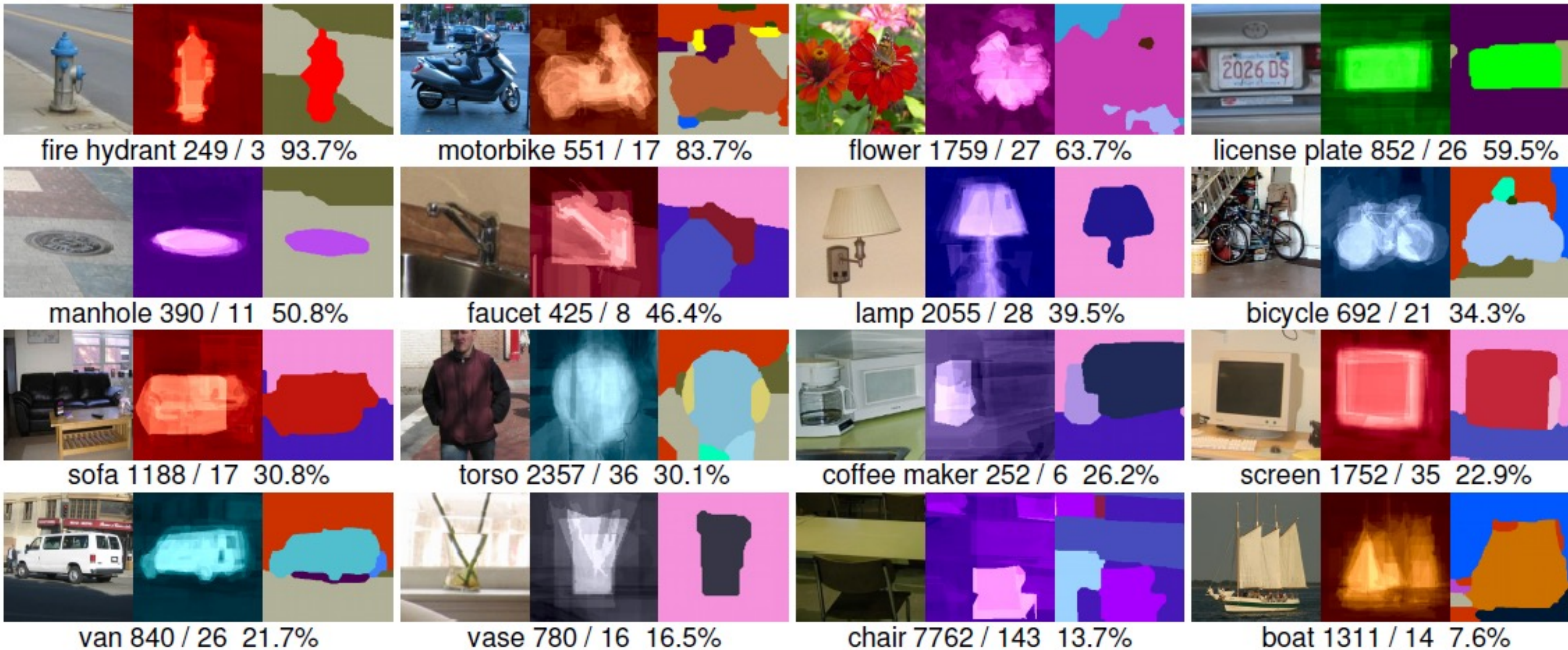
	Region-based	Detector-based	Region + Detector Combined
SIFT Flow (Liu et al., 2009)	77.7 (32.8)	71.1 (26.7)	**78.6 (39.2)**
LabelMe+SUN	58.3 (5.9)	52.5 (11.3)	**61.4 (15.2)**

Per-pixel rate (average per-class rate)





Toward Broad Coverage



Comparison to state of the art

SIFT Flow	Per-Pixel	Per-Class
Our approach	**78.6**	**39.2**
Tighe and Lazebnik (2013)	77.0	30.1
Liu et al. (2011)	76.7	N/A
Farabet et al. (2012)	78.5	29.6
Farabet et al. balanced (2012)	74.2	46.0
Eigen and Fergus (2012)	77.1	32.5
Myeong et al. (2012)	77.1	32.3

Comparison to state of the art

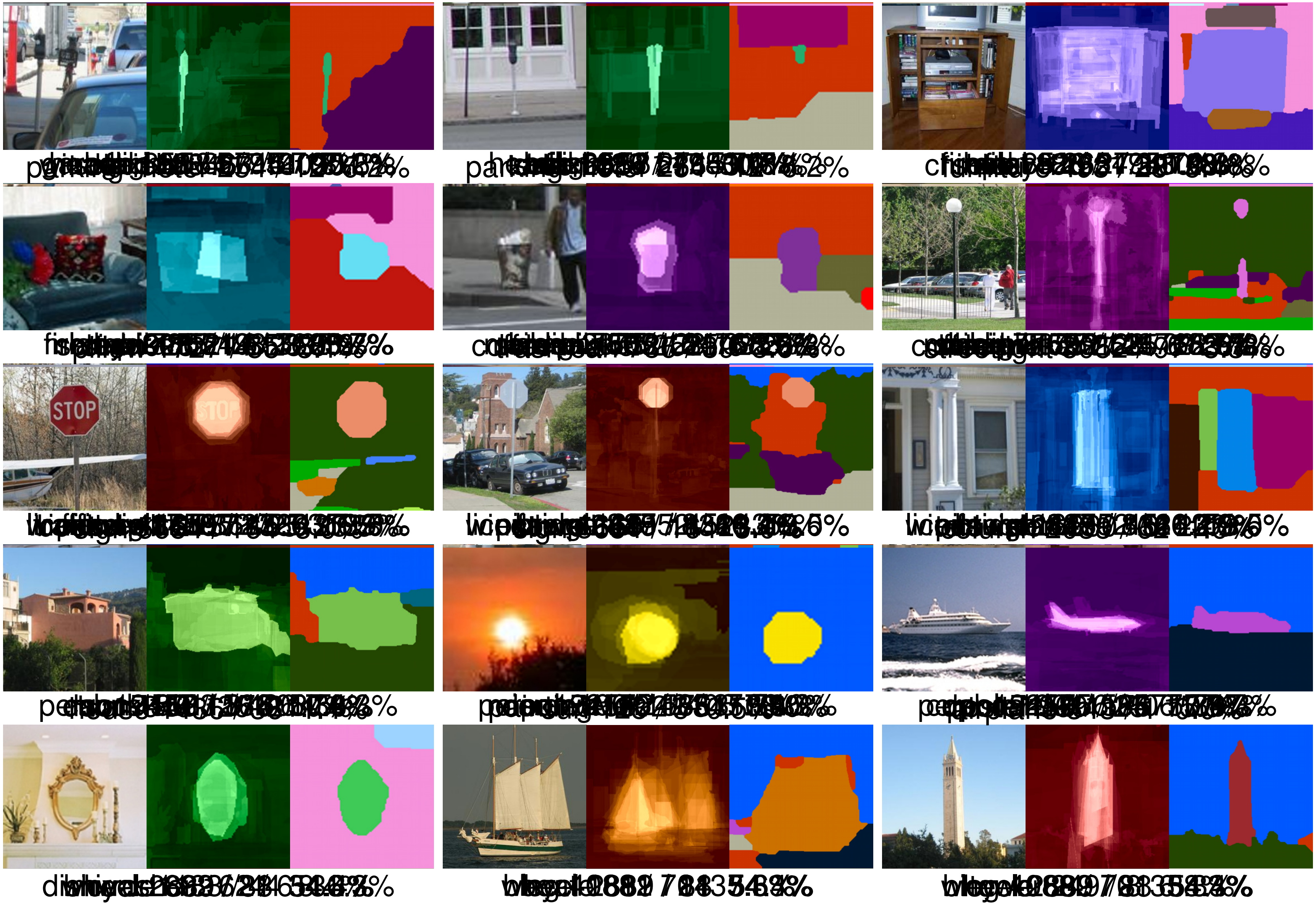
LabelMe+SUN	Per-Pixel	Per-Class
Our approach	**61.4**	**15.2**
Outdoor	65.5	15.3
Indoor	46.3	12.2
Tighe and Lazebnik (2013)	54.9	7.1
Outdoor	60.8	7.7
Indoor	32.1	4.8

Now what?

- Other researchers should push for bigger datasets, broader coverage
- For us – lots more work to do
  - Improve computational efficiency of exemplar SVM training: try whitened HOG approach of Hariharan et al. (ECCV 2012)
  - Leverage the object separation the per-exemplar detectors are already providing to separate the objects in our final parsing

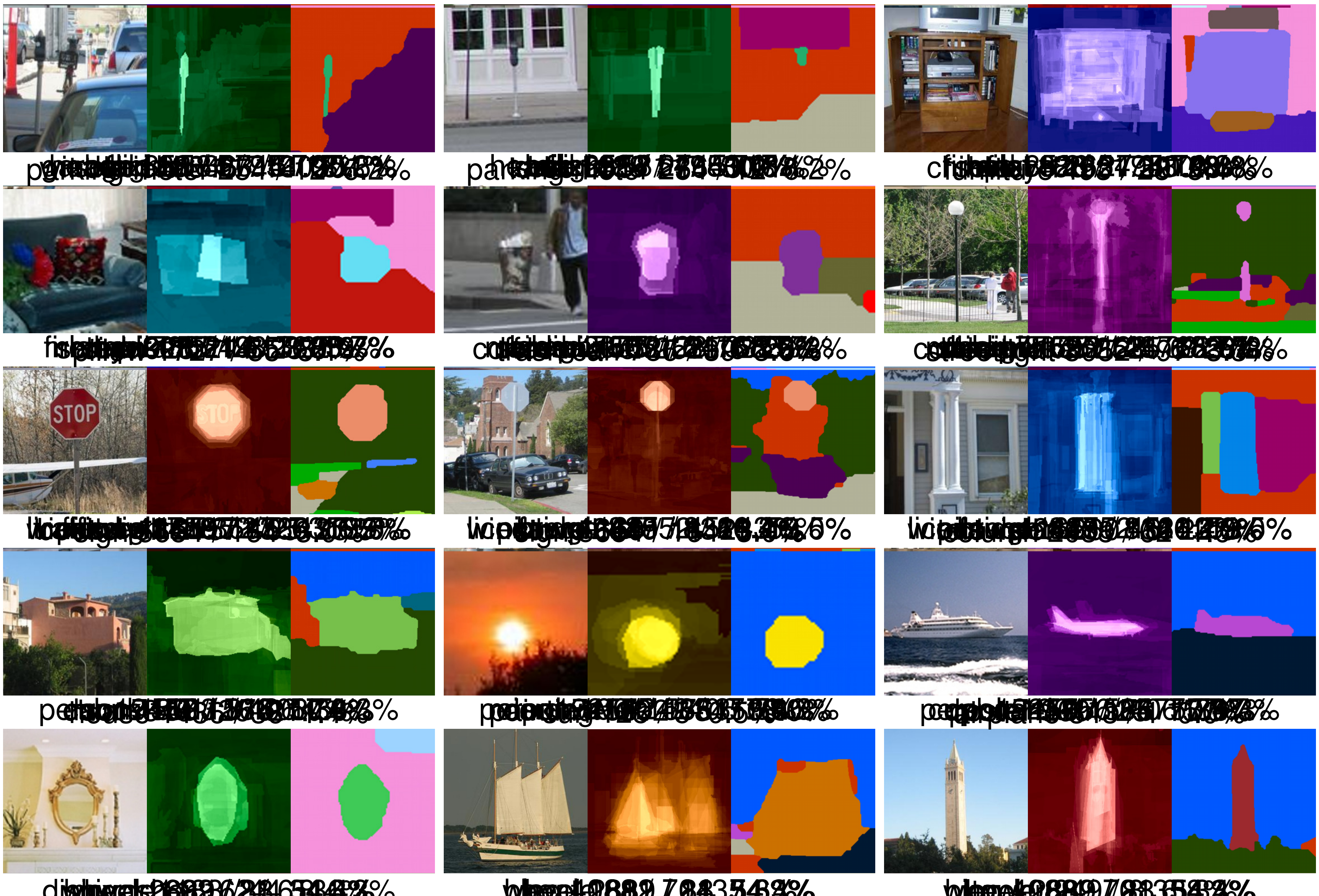
Code and data publicly available on our websites:

http://www.cs.unc.edu/~jtighe/Papers/CVPR13/



Code and data publicly available on our websites:

http://www.cs.unc.edu/~jtighe/Papers/CVPR13/



Code and data publicly available on our websites:
http://www.cs.unc.edu/~jtighe/Papers/CVPR13/