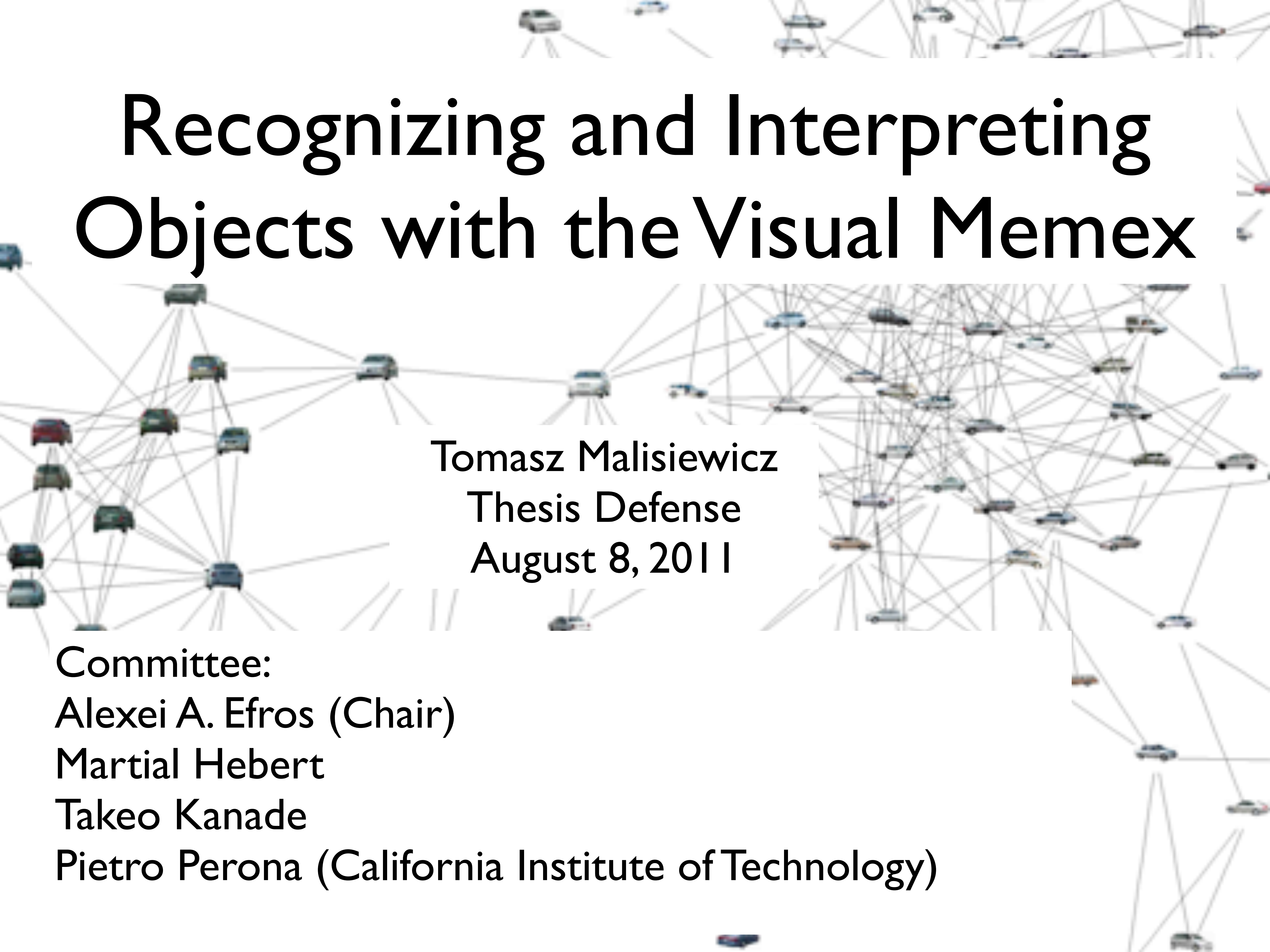


Recognizing and Interpreting Objects with the Visual Memex

Tomasz Malisiewicz
Thesis Defense
August 8, 2011

Committee:
Alexei A. Efros (Chair)
Martial Hebert
Takeo Kanade
Pietro Perona (California Institute of Technology)

Understanding an Image

Fei Fei, Fergus & Torralba 2005

Object naming

Object naming / Object categorization

Fei, Fergus & Torralba 2005

Object naming / Object categorization

sky

building

flag

face

banner

wall

street lamp

bus

bus

cars

Classical View of Categories

- Dates back to Plato & Aristotle
  - –Categories are defined by a list of properties shared by all members
  - –Category membership is binary
  - –Every member of a category is equal

Problems with Classical View

Problems with Classical View

- Humans don't do this! (Wittgenstein 1953)
  - People don't rely on abstract definitions
  - e.g. define the essential property shared by all "games"?

Problems with Classical View

- Humans don’t do this! (Wittgenstein 1953)
  - People don’t rely on abstract definitions
  - e.g. define the essential property shared by all “games”?

- Typicality and borderline-cases (Rosch 1973)
  - A robin is “more” of a bird than a penguin
  - Is an olive a fruit? Are curtains furniture?
  - Is Pluto a planet?

Problems with **Visual** Categories

Problems with **Visual** Categories

Chair

- A lot of categories are functional

Problems with **Visual** Categories

Chair

- A lot of categories are functional

Car

- Same object, different appearance!

The Dictatorship of Librarians

Weinberger, 2007

The Dictatorship of Librarians

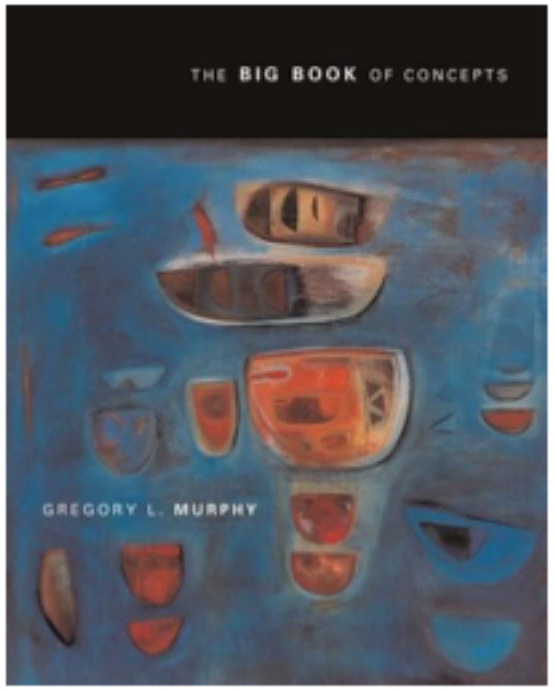
Weinberger, 2007

categories are losing...

YAHOO!® vs. Google

Who needs categories?

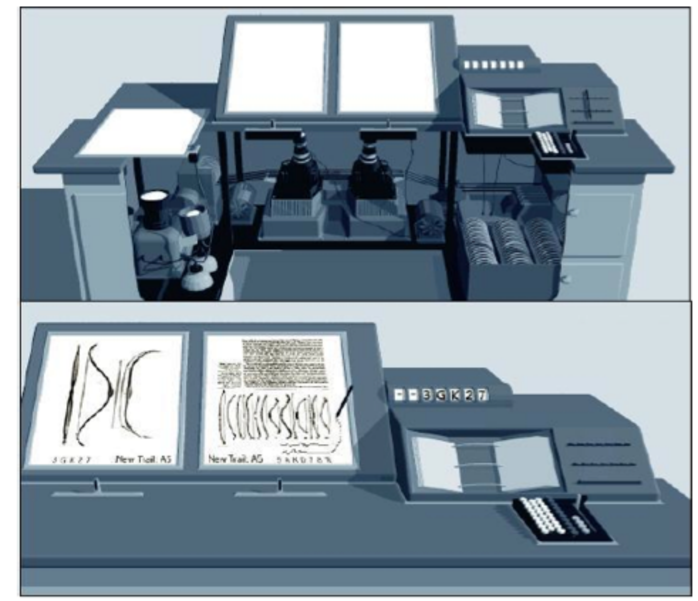
- Exemplar Theory (Medin & Schaffer 1978, Nosofsky 1986, Krushke 1992)
  - categories represented in terms of remembered objects (exemplars)
  - Similarity is measured between input and all exemplars



Murphy
Big Book of Concepts

- "What is this like?" vs. "What is this?" (Bar, 2007)
- Vannevar Bush's Memex (Bush 1945)

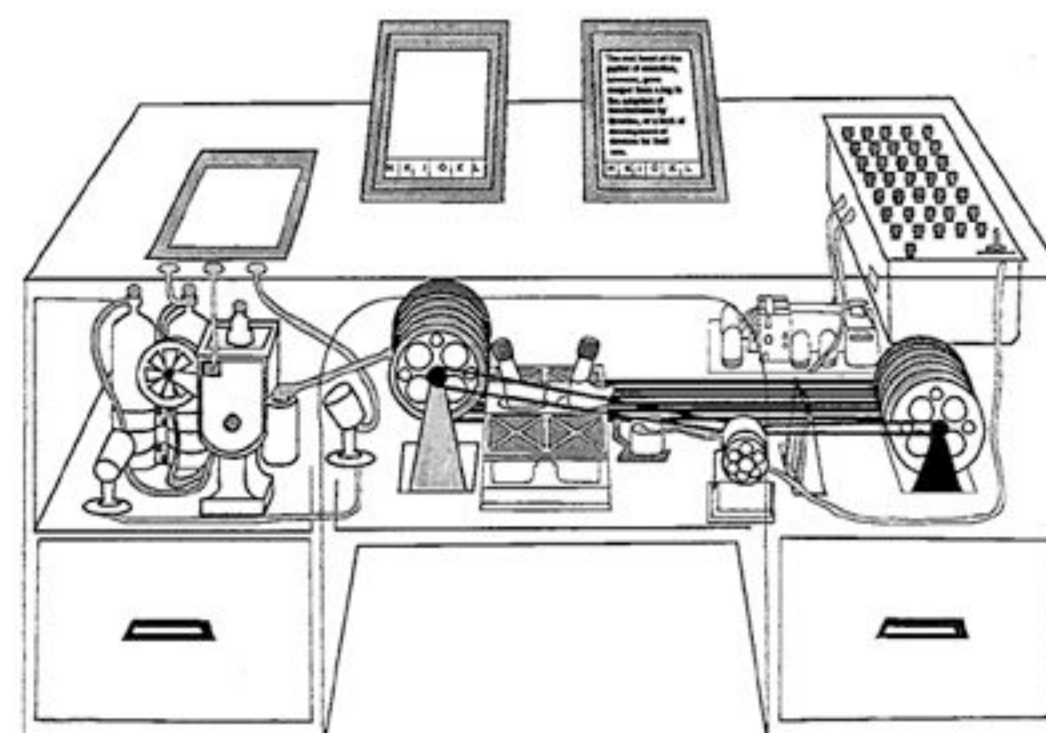
Bush’s Memex (1945)



A physical device which stores research papers, notes, books on microfilm

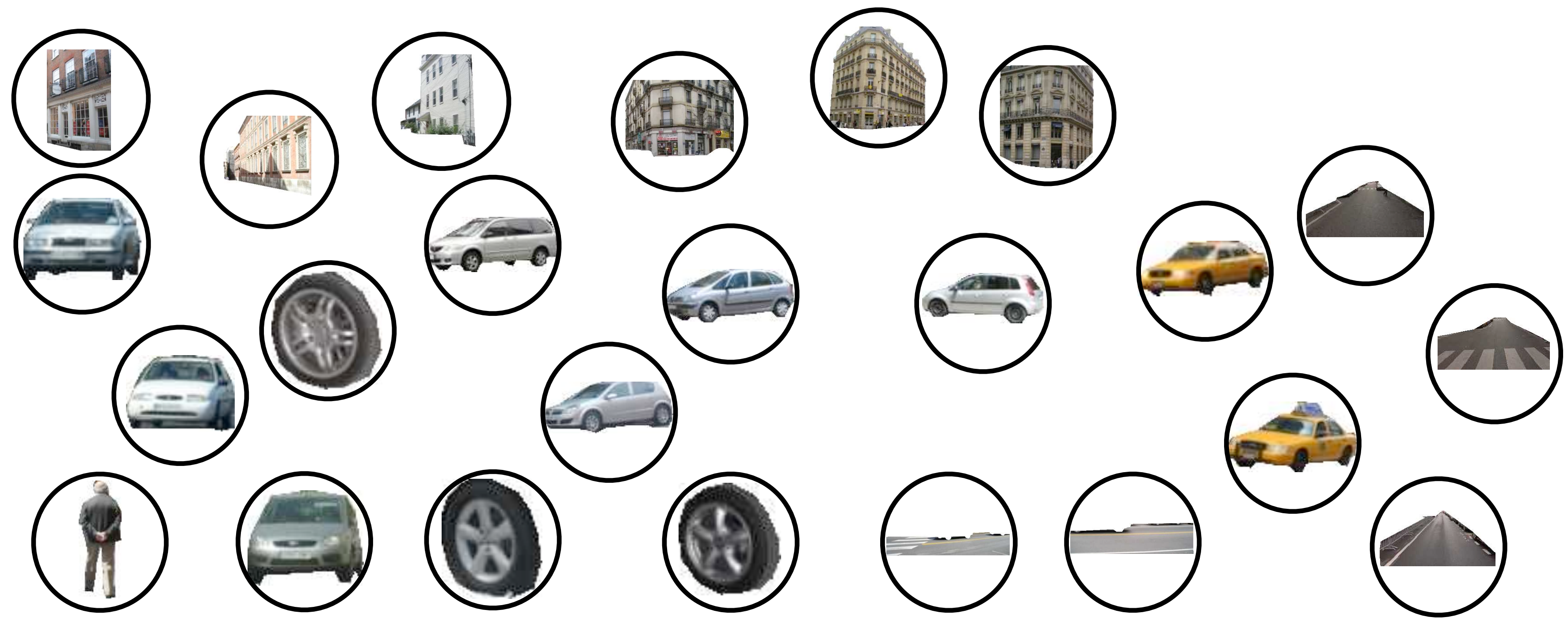
User creates “trails” between the materials in the memex

Acts as an external memory



The Visual Memex

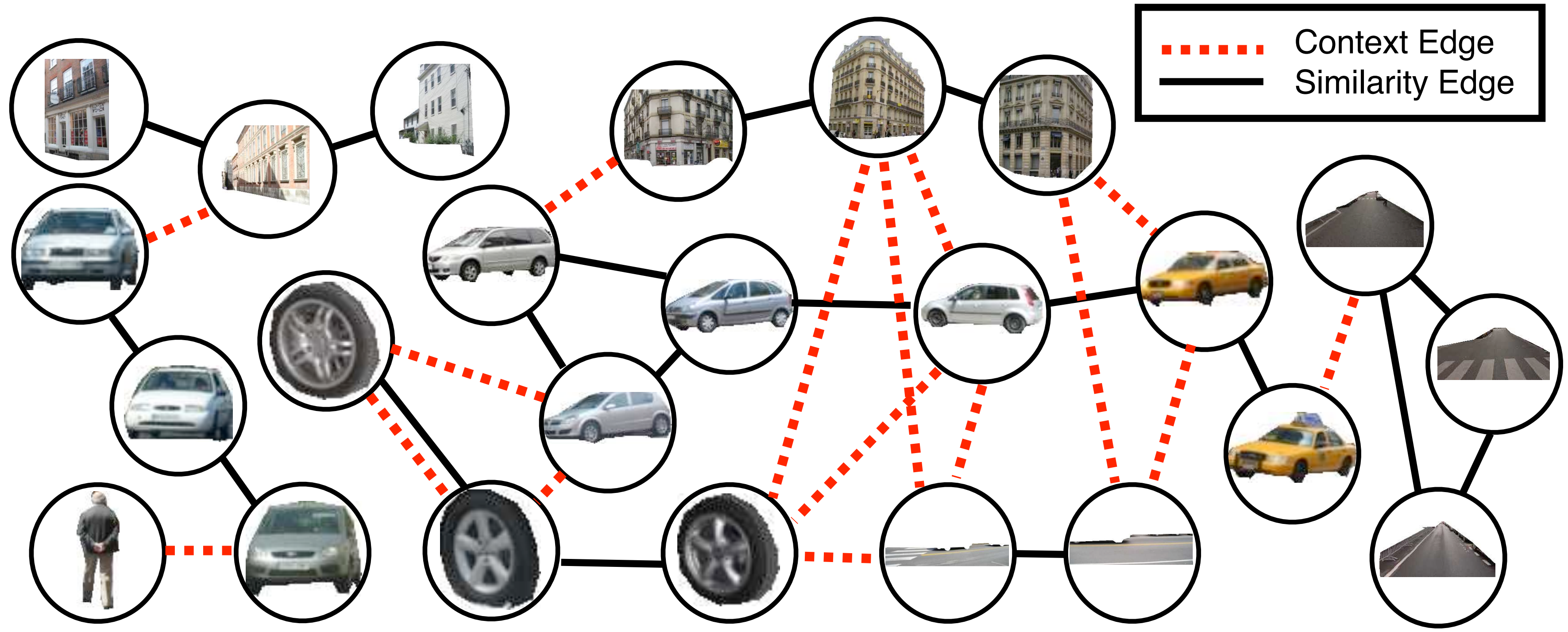
Input Image



Nodes = exemplars

The Visual Memex

Input Image



Nodes = exemplars

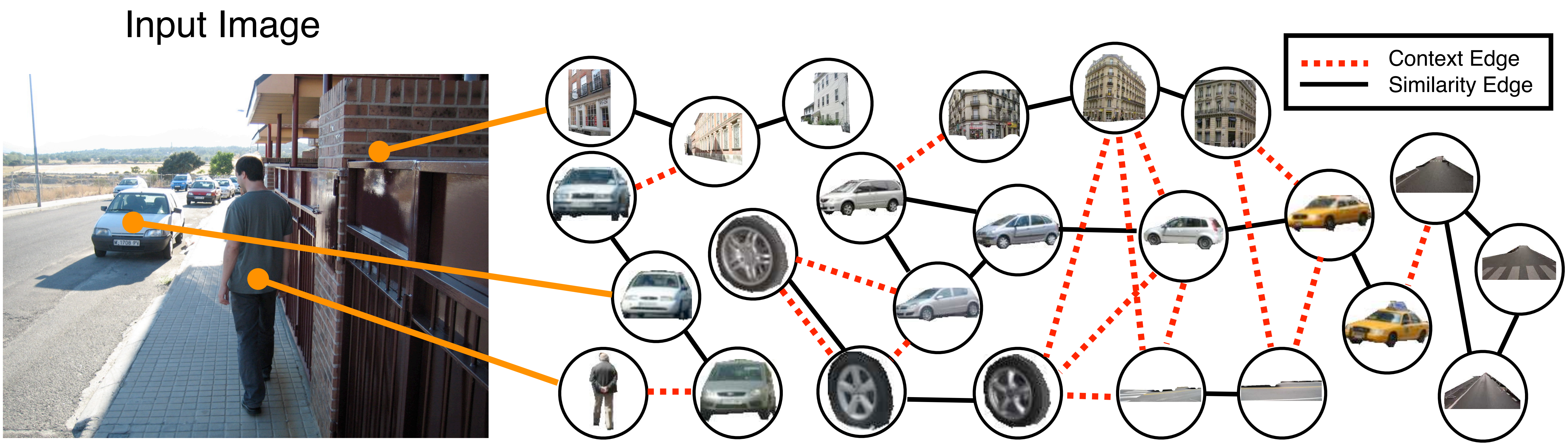
Edges = relationships

visual similarity

context

meta-data

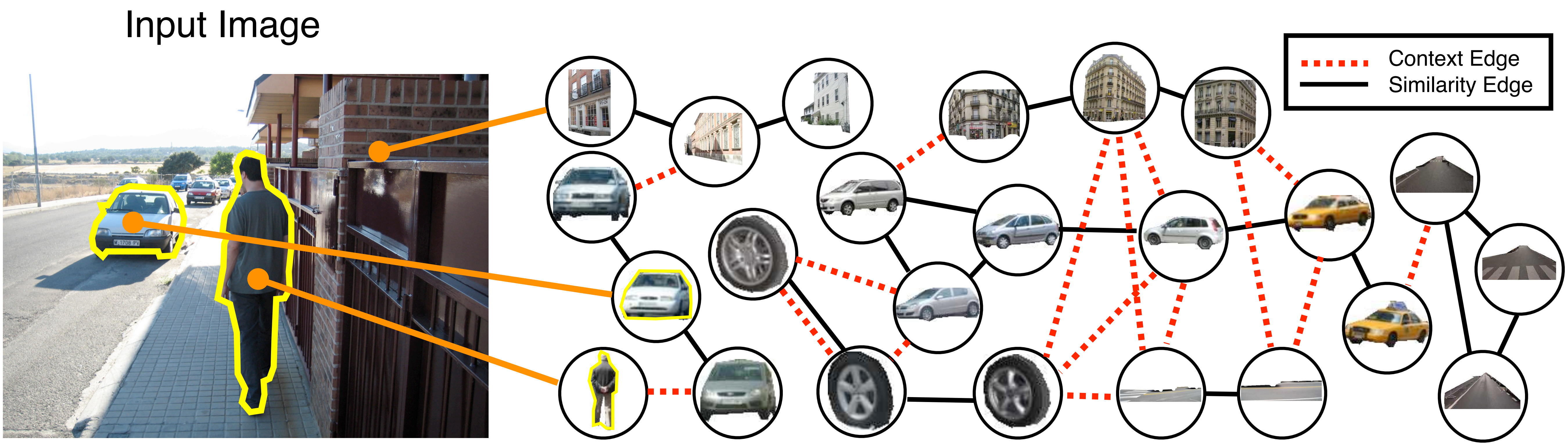
The Visual Memex



Nodes = exemplars

Edges = relationships
visual similarity
context
meta-data

The Visual Memex



Nodes = exemplars

Edges = relationships
visual similarity
context
meta-data

Overview

- Part I: Creating **Visual Associations**
  - Per-Exemplar Distance Functions & Multiple Segmentations [CVPR 2008]
  - Exemplar-SVMs [ICCV 2011]

- Part II: Utilizing **Visual Memex**
  - Object Interpretation [ICCV 2011]
  - Context Challenge [NIPS 2009]

Visual Associations

- How are objects similar?

Measuring Visual Similarity is not trivial

Measuring Visual Similarity is not trivial

Shape

Color

Measuring Visual Similarity is not trivial

Shape

~~Color~~

Measuring Visual Similarity is not trivial

Shape

~~Color~~

Measuring Visual Similarity is not trivial

Shape

~~Color~~

Shape

Color

Measuring Visual Similarity is not trivial

Shape

~~Color~~

~~Shape~~

Color

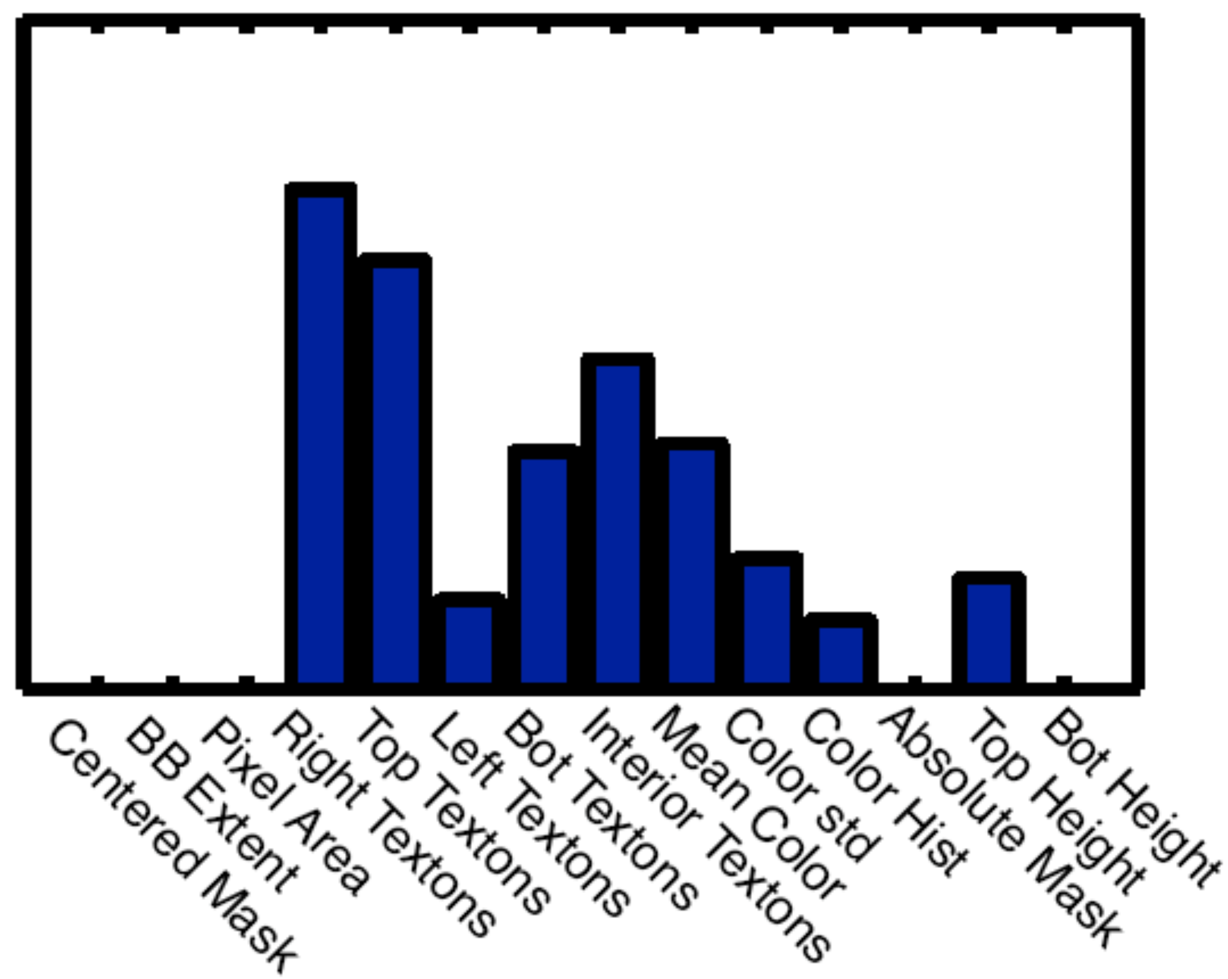
Per-Exemplar
Distance "Similarity" Functions

- Positive linear combination of elementary distances

$$D_e(z) = \mathbf{w}_e \cdot \mathbf{d}_{ez}$$

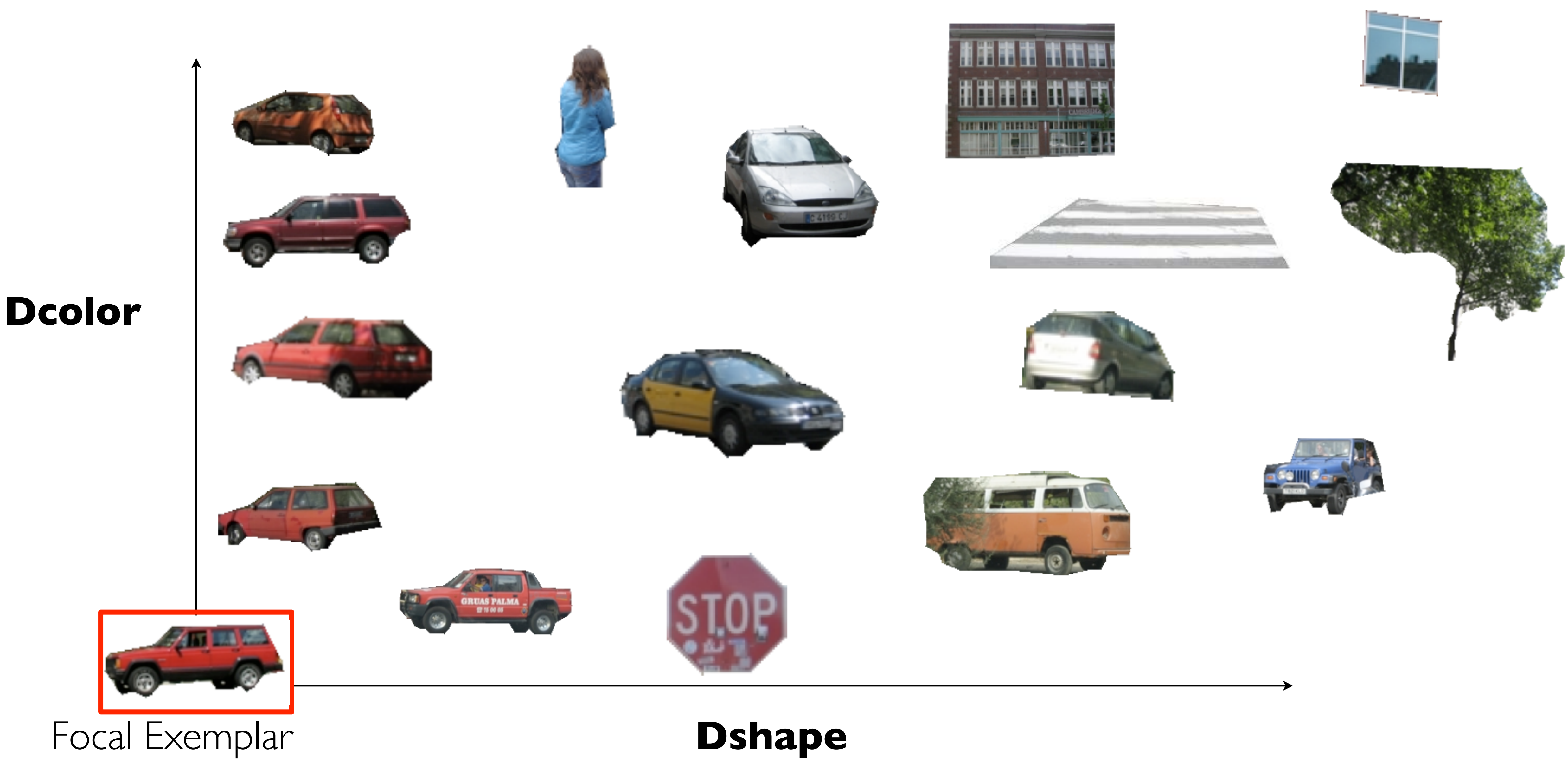
Exemplar **e**

Exemplar **e** Distance Function

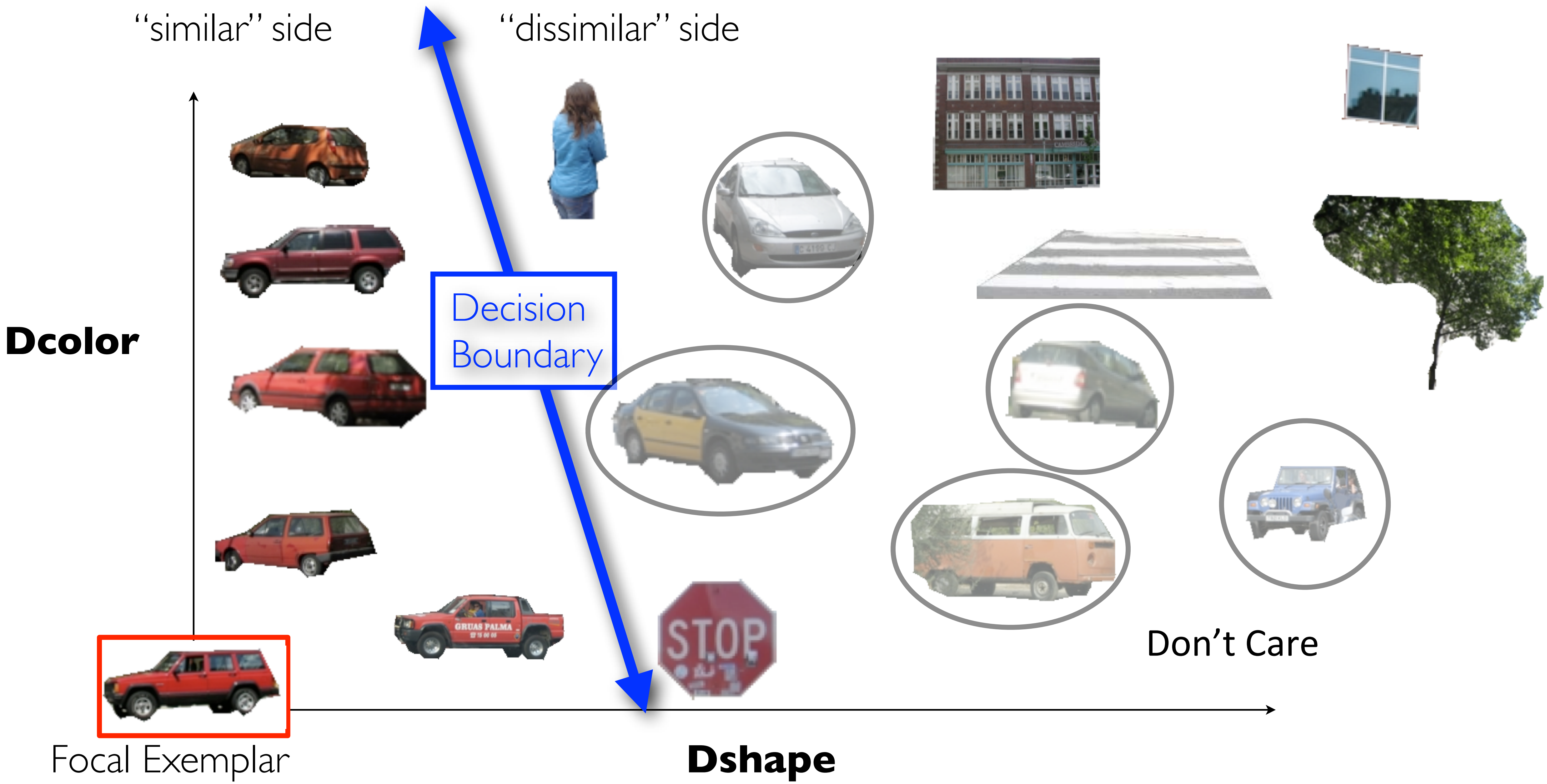


Malisiewicz et al. 2008

Learning Distance Function



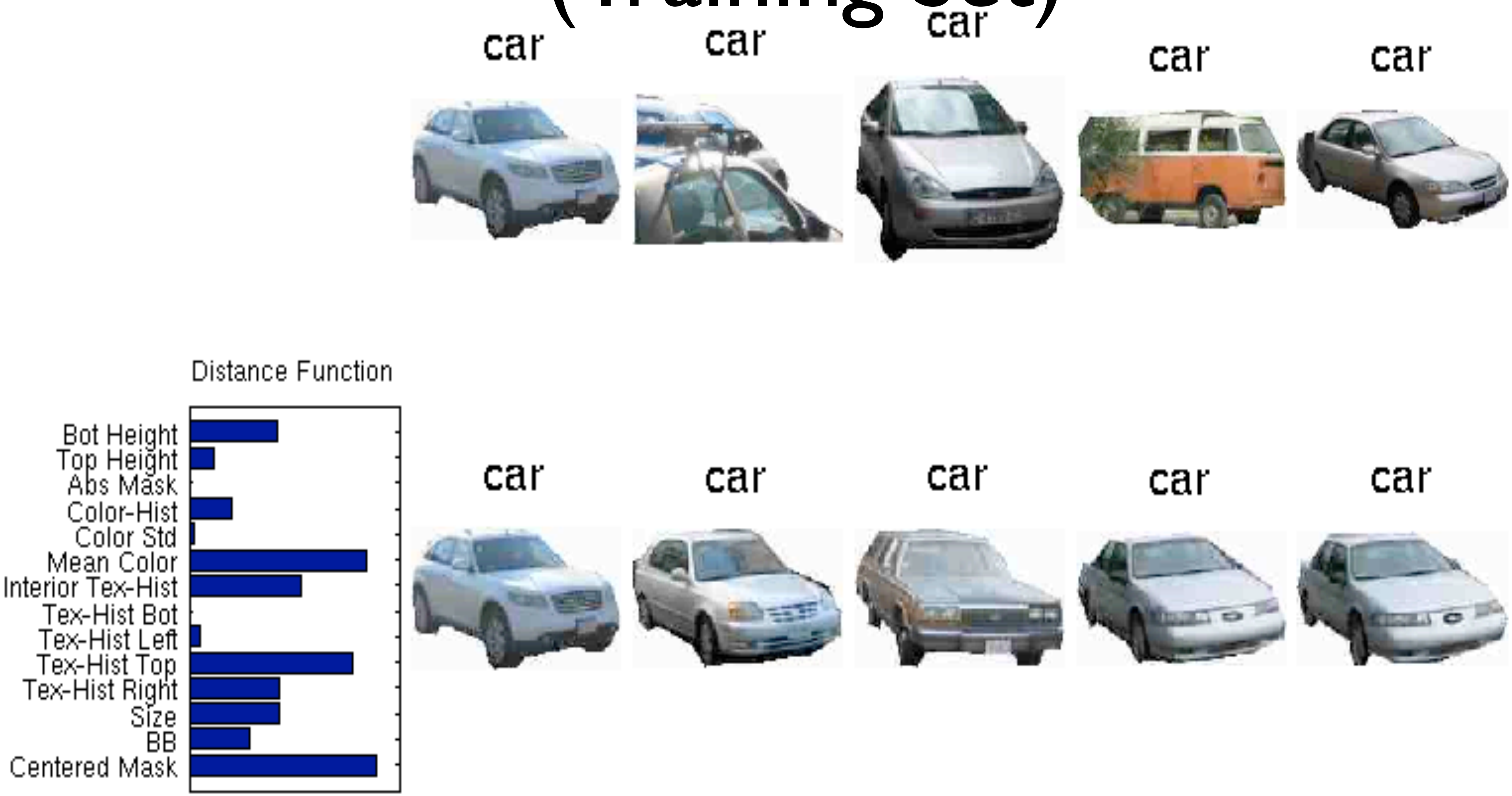
Learning Distance Function



LabelMe = Source of Exemplars

Russell et al. 2008

Visualizing Distance Functions (Training Set)

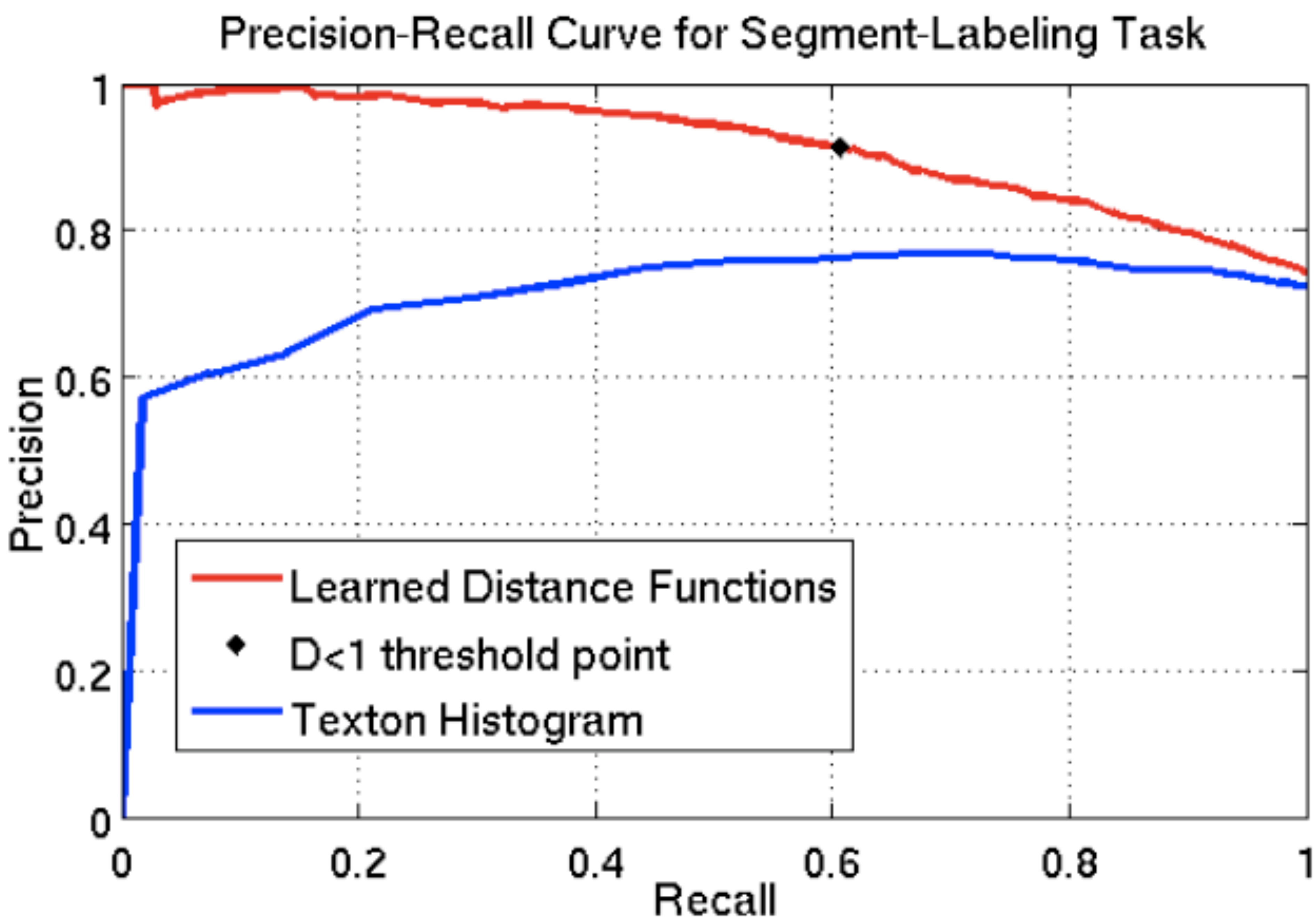


Top label confusions

stop sign	sign	7.8%
pole	streetlight	6.7%
motorcycle	motorbike	6.2%
mountains	mountain	6.2%
ground grass	sidewalk	3.7%
grass	lawn	3.6%
road highway	road	3.4%
painting	picture	3.4%
sidewalk	road	3.2%
cloud	sky	3.1%
grass	ground grass	3.1%
mountain	mountains	2.7%

Table 2: Top dozen label confusions discovered after distance function learning.

LabelMe Segment Labeling Task



Segment-then-recognize

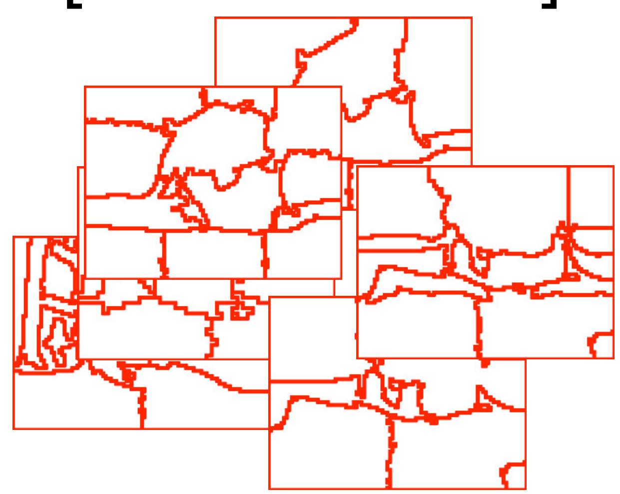
Input Image

Malisiewicz et al. 2008

Segment-then-recognize

Input Image

Multiple Segmentations
[Hoiem et al. 2005]



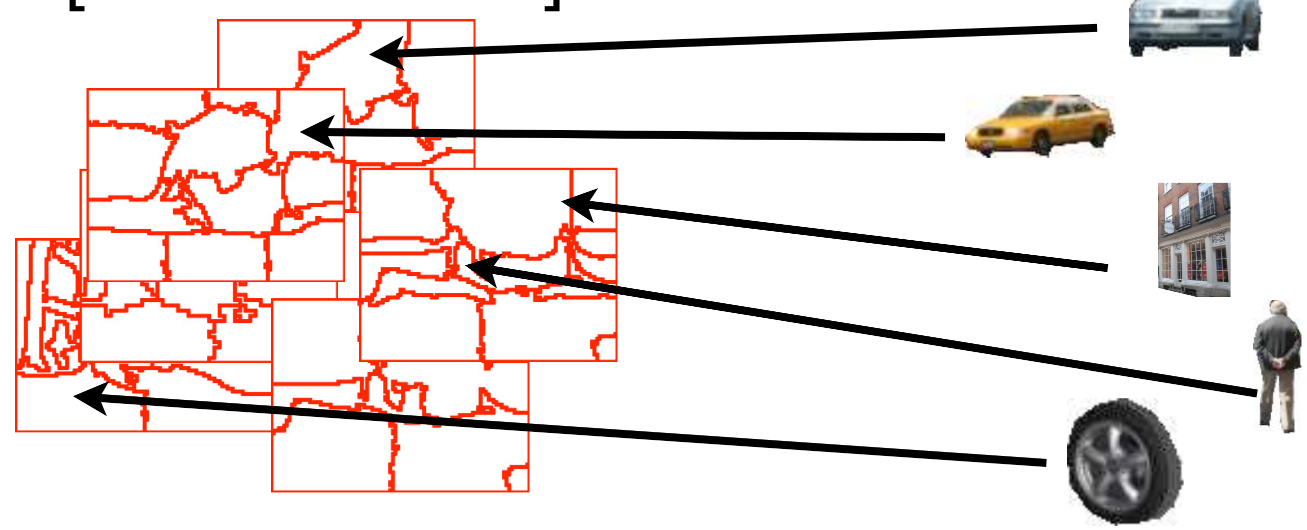
Malisiewicz et al. 2008

Segment-then-recognize

Input Image

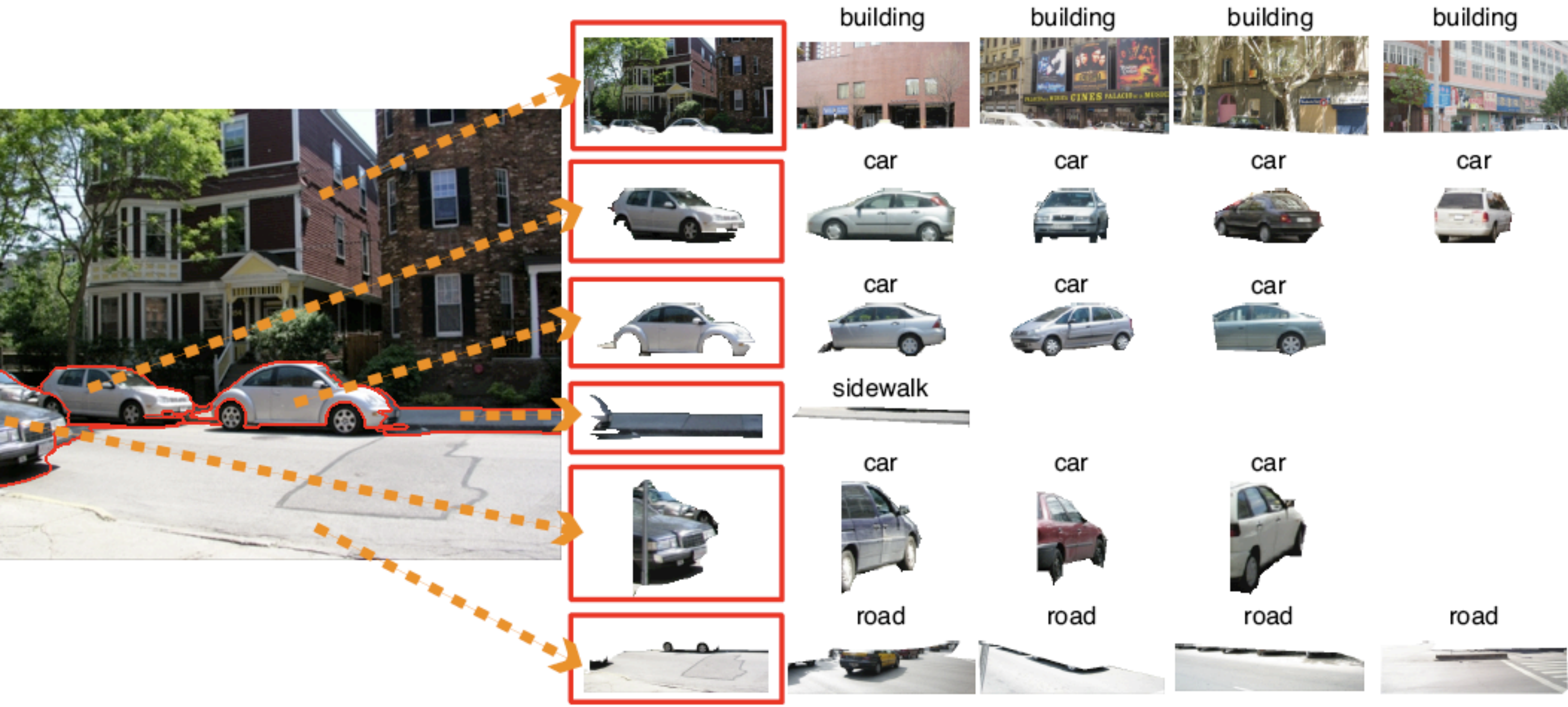
Multiple Segmentations
[Hoiem et al. 2005]

Exemplars



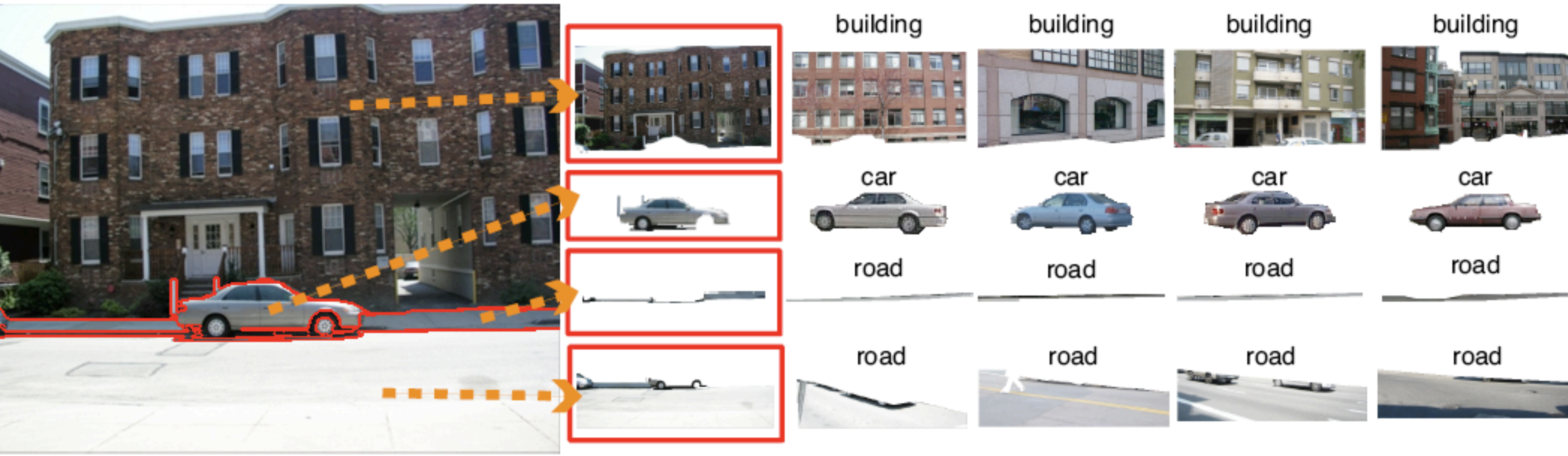
Malisiewicz et al. 2008

Segment-then-recognize Results



Malisiewicz et al. 2008

Segment-then-recognize Results



Malisiewicz et al. 2008

Limitations of CVPR 2008 approach

- Relying too much on bottom-up segmentation
- Not enough negative data

Limitations of CVPR 2008 approach

- Relying too much on bottom-up segmentation
- Not enough negative data
- State-of-the-art object detectors based on **multiscale sliding windows** and **hard negative mining** [Dalal-Triggs 2005, Felzenszwalb et al. 2008]

Limitations of CVPR 2008 approach

- Relying too much on bottom-up segmentation
- Not enough negative data
- State-of-the-art object detectors based on **multiscale sliding windows** and **hard negative mining** [Dalal-Triggs 2005, Felzenszwalb et al. 2008]

But these detectors are generally trained in a category-wise fashion

Best of both worlds?

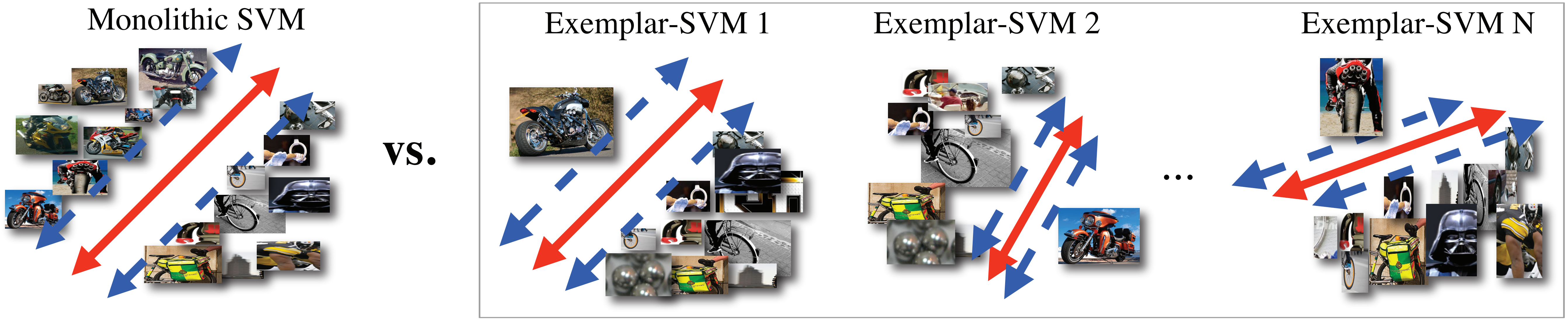
- Is it possible to combine:
  - State-of-the-art object detectors [Dalal-Triggs 2005, Felzenszwalb et al. 2008]
  - Per-exemplar models [Frome et al. 2007, Malisiewicz et al. 2008]

Best of both worlds?

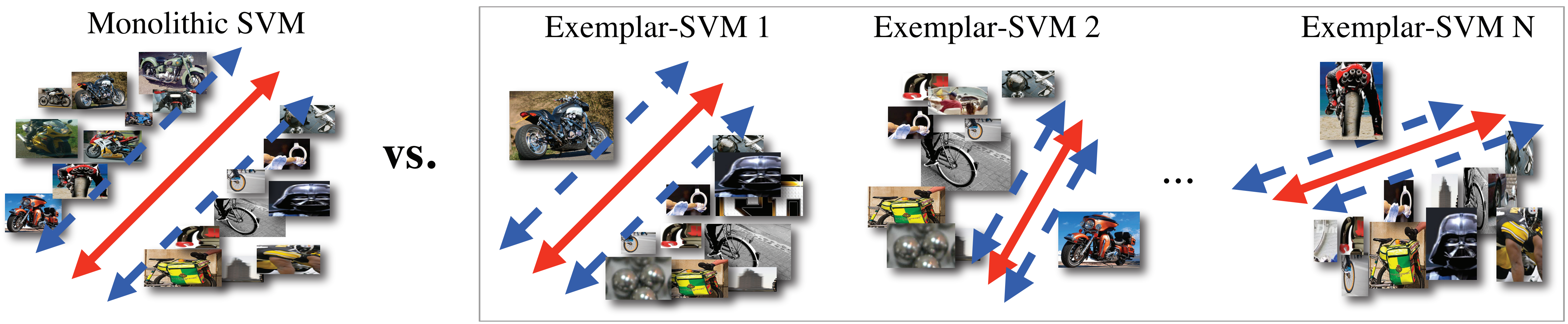
- Is it possible to combine:
  - State-of-the-art object detectors [Dalal-Triggs 2005, Felzenszwalb et al. 2008]
  - Per-exemplar models [Frome et al. 2007, Malisiewicz et al. 2008]

Yes :-)

Exemplar-SVMs

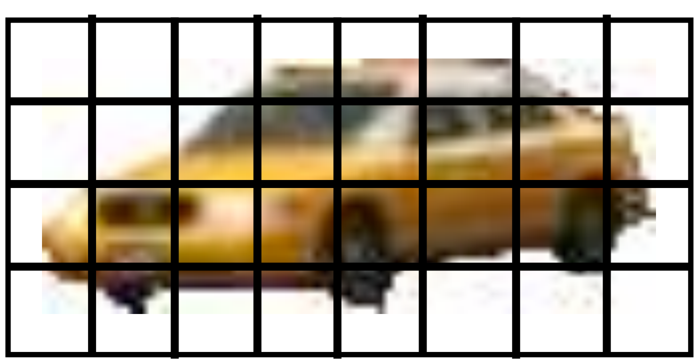


Exemplar-SVMs

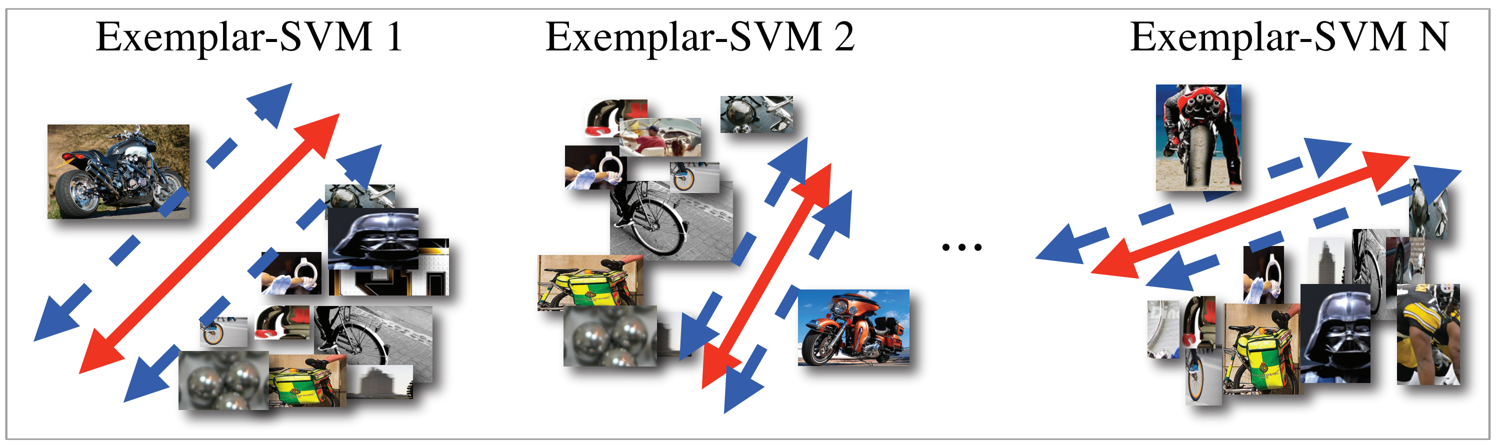


Solve many easy (convex) learning problems

Learn with a **single positive instance**

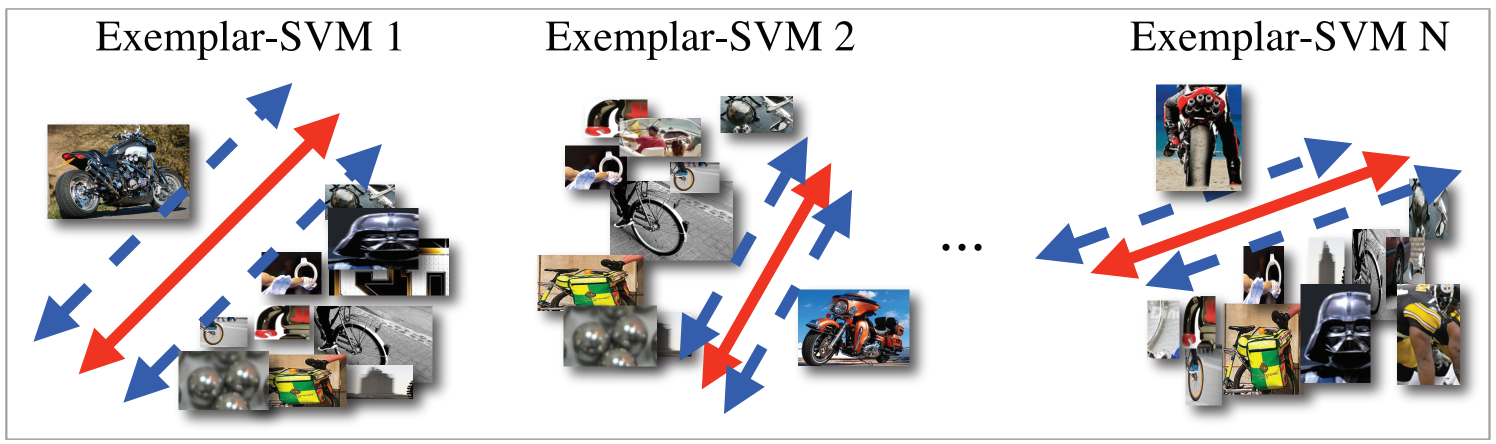


Exemplar-SVMs



CPU_1 CPU_2 CPU_N

Exemplar-SVMs



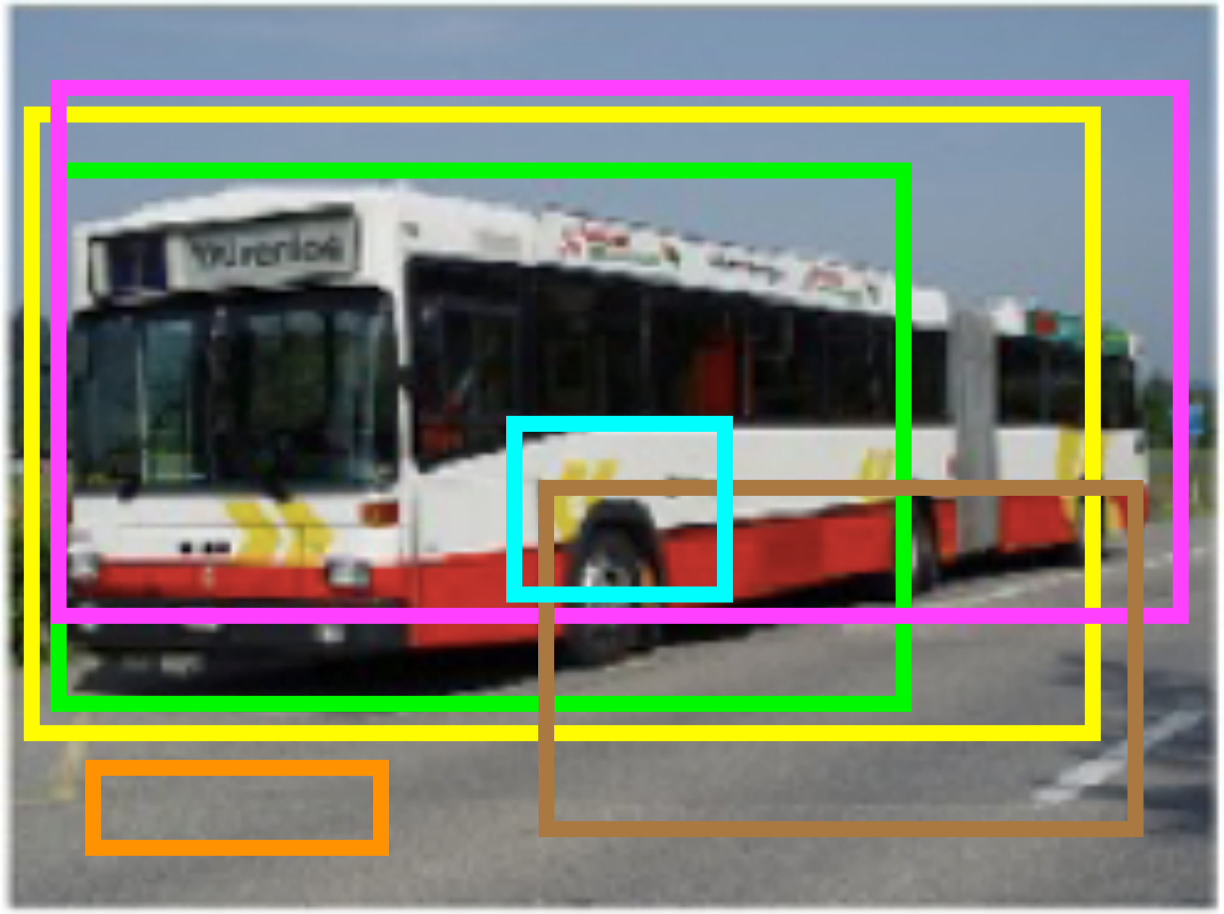
CPU$_1$ CPU$_2$ CPU$_N$

SVM after training

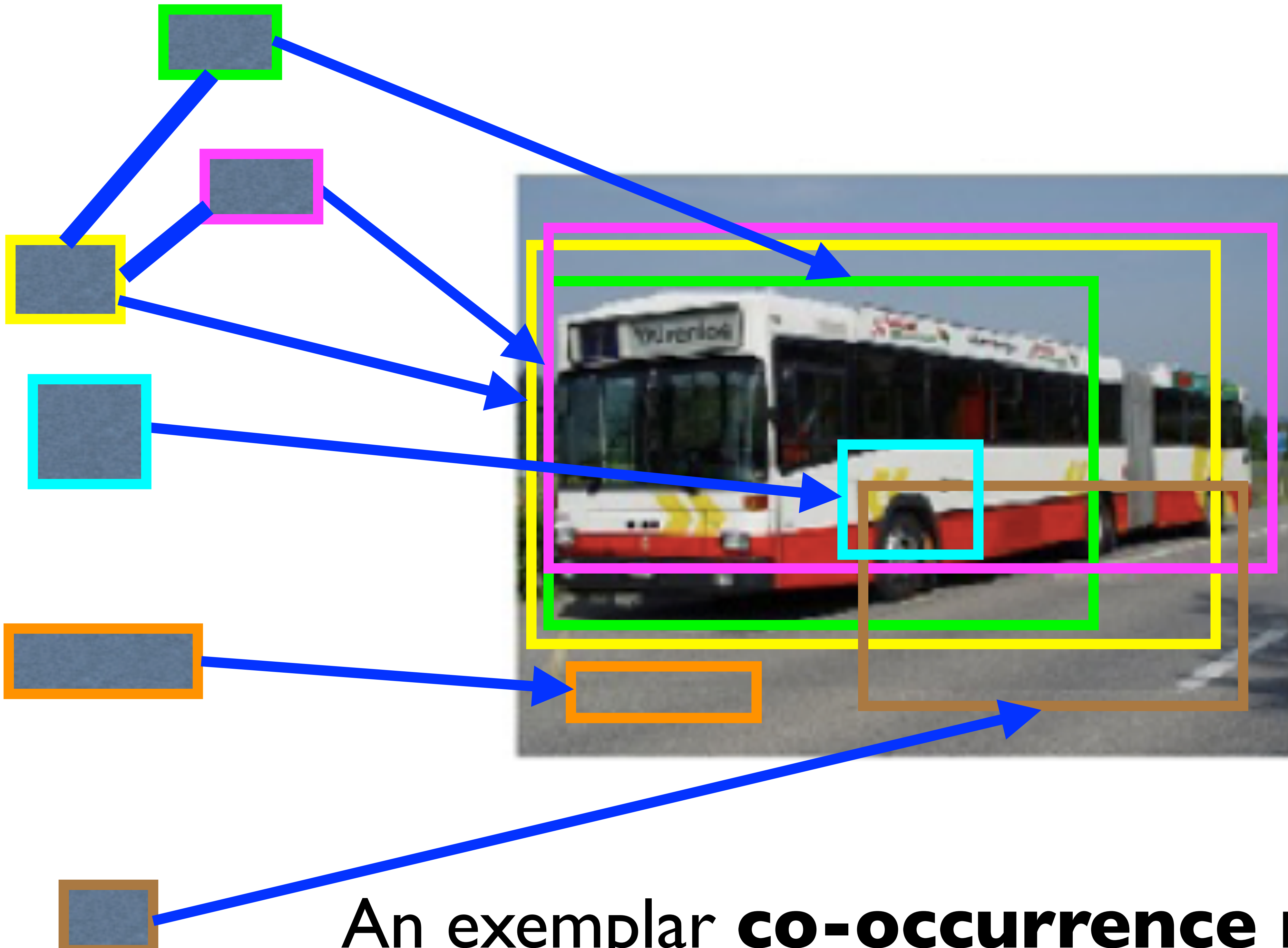
SVM after calibration

Platt 1999

Exemplar-SVMs



Exemplar-SVMs

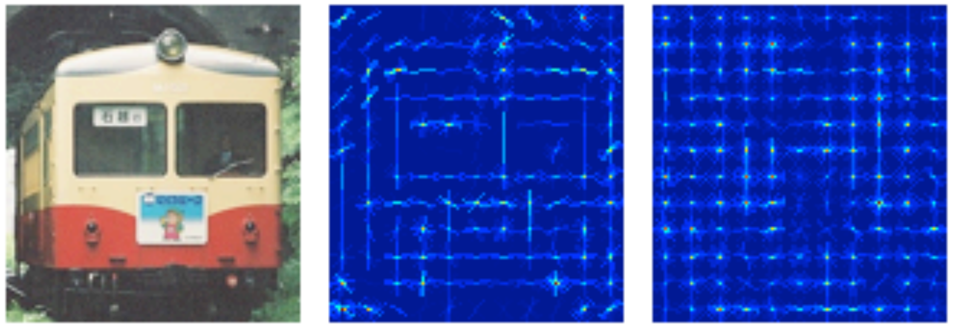


An exemplar **co-occurrence matrix**

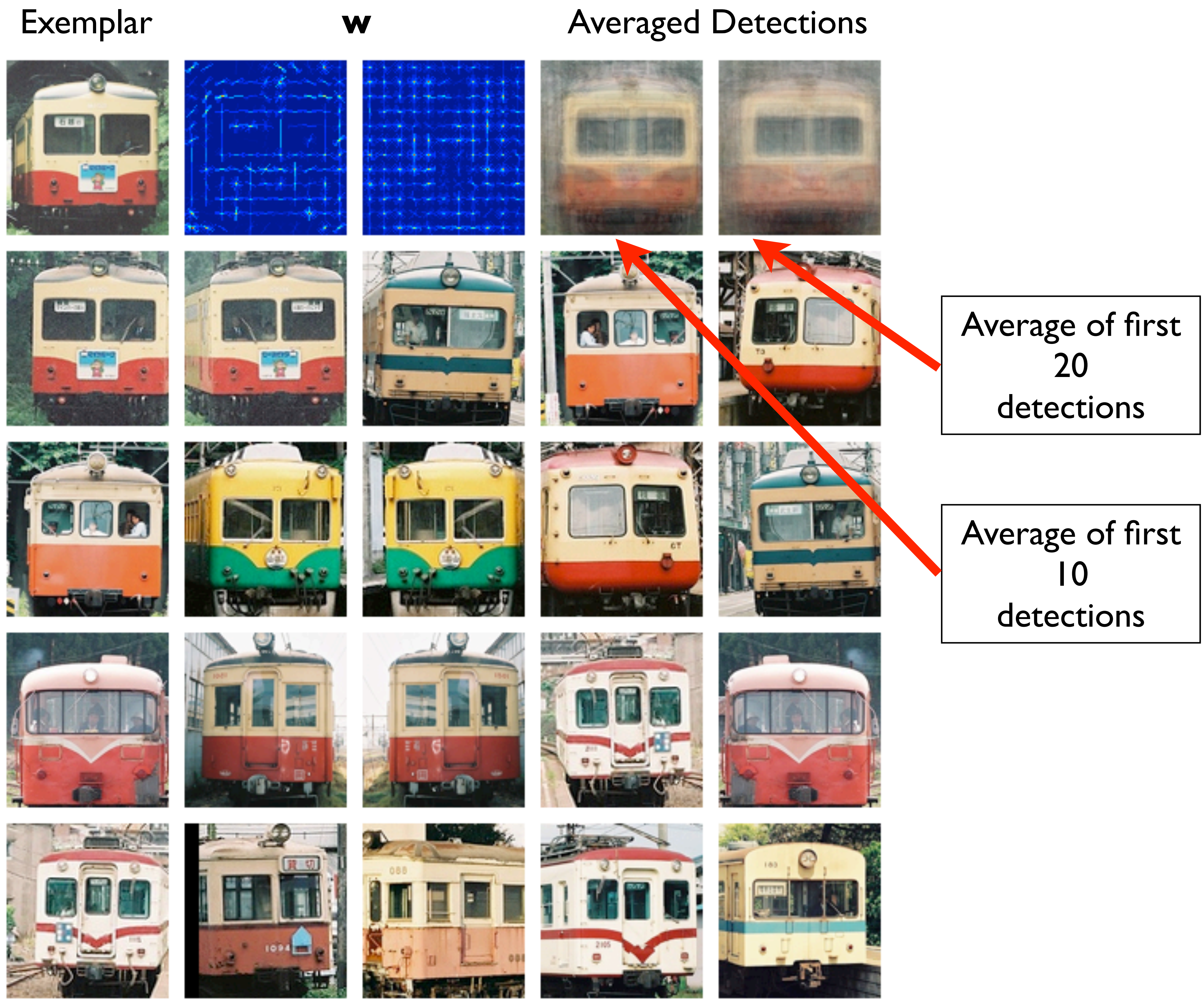
Qualitative Results

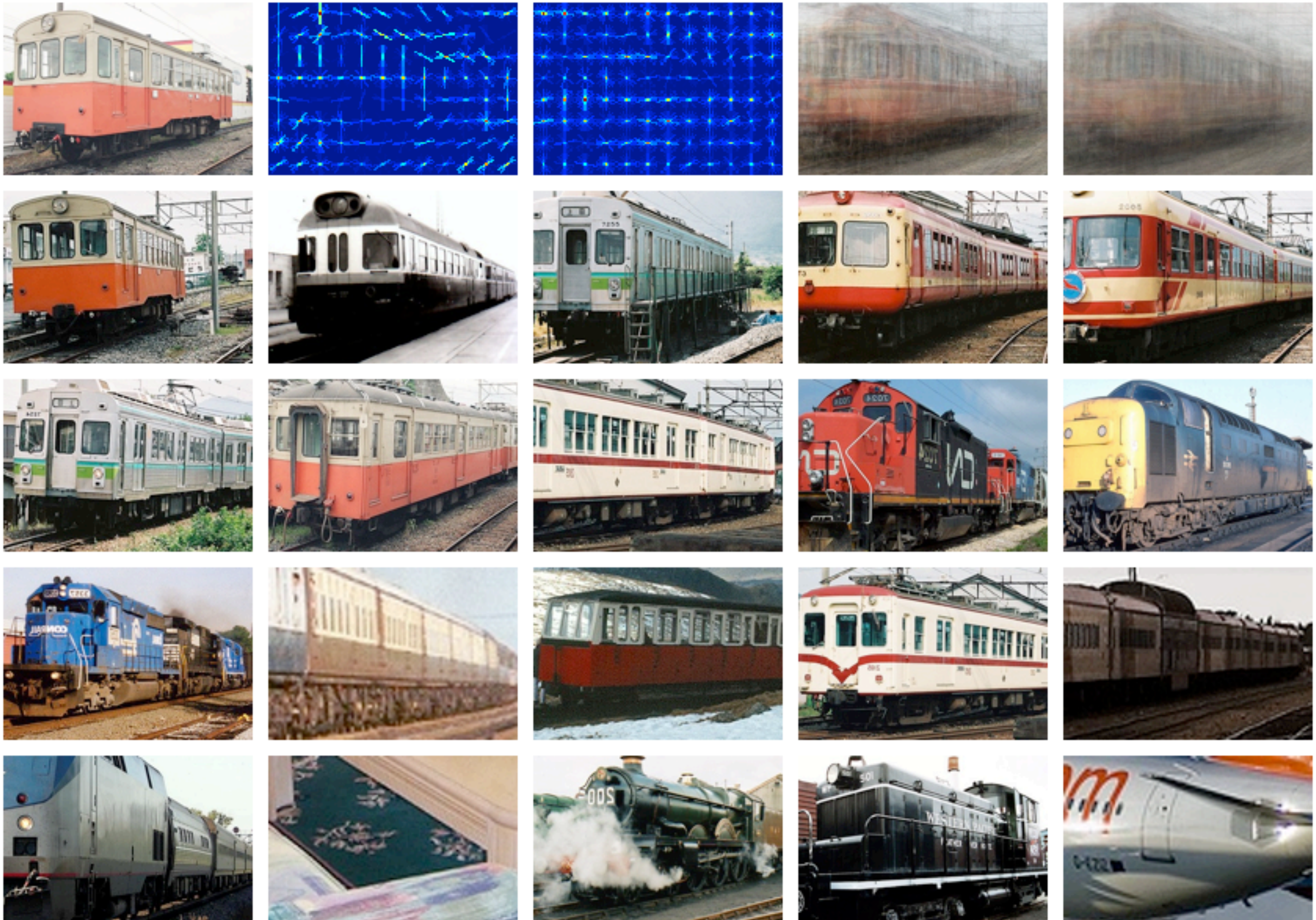
- Let's take a look at some Exemplar-SVM results in PASCAL VOC dataset

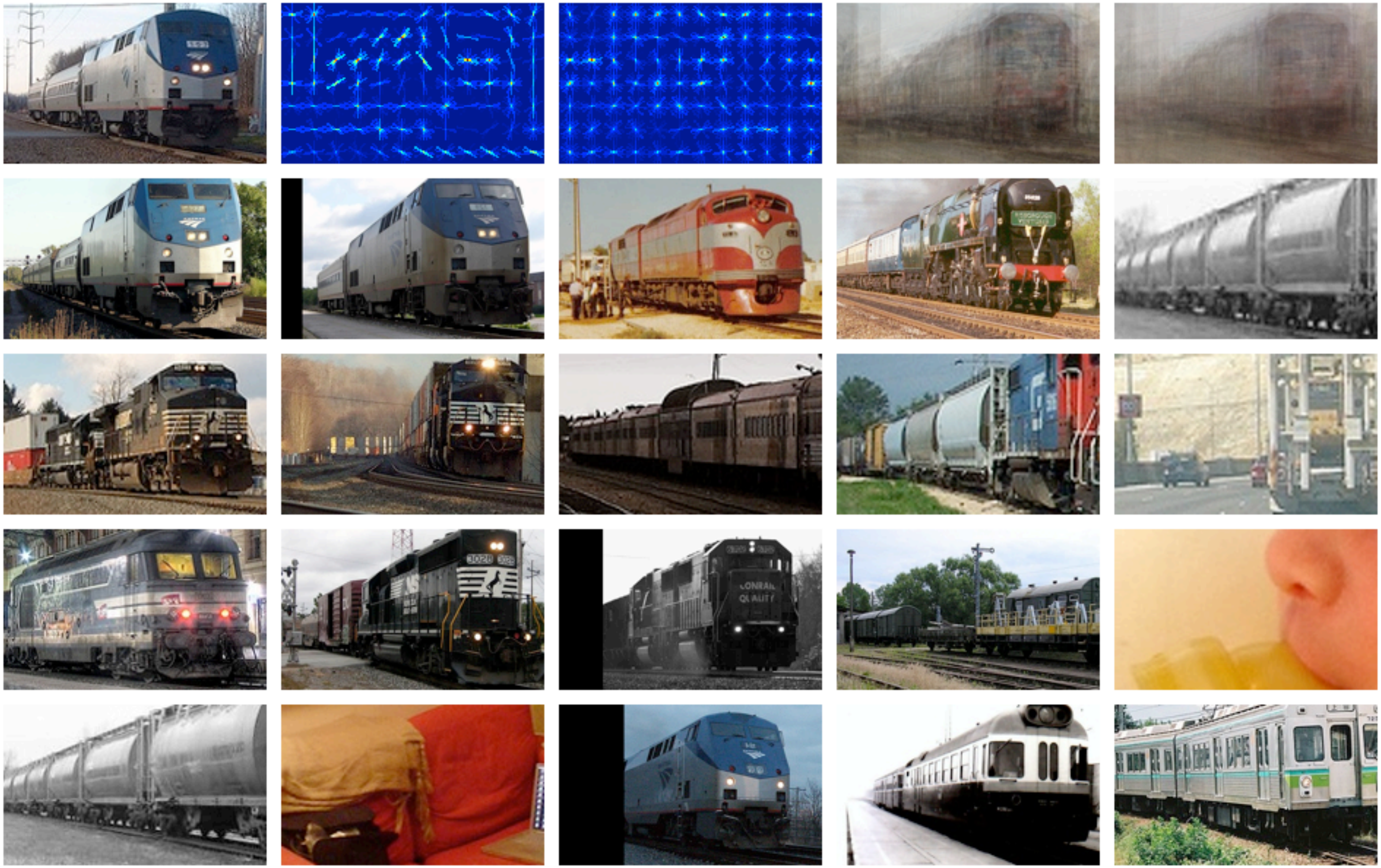
Exemplar **w**




Exemplar
w


Exemplar
w
Averaged Detections
Average of first 20 detections
Average of first 10 detections





NREX
Pan Am Railways
8006
BNSF
DANGER

Evaluating Exemplar-SVMs

- **Nearest Neighbor**
  - No Learning
- **Per-Exemplar Distance Functions**
  - Learning in distance-to-exemplar space [Malisiewicz et al. 2008]

Comparison of 3 methods

*Learned Distance Function

Quantitative: PASCAL VOC 2007 dataset

- A standard computer vision object detection benchmark
- 20 object categories
- Machine performance is far below human

Object Category Detection

mAP on PASCAL VOC 2007 detection task

NN	0.110
DFUN	0.157
Exemplar-SVMs	**0.150**
Exemplar-SVMs Cal	**0.198**
Exemplar-SVMs Co-occur	**0.227**
DT*	0.097
LDPM**	0.266

*Dalal et al. 2005 **Felzenszwalb et al. 2010

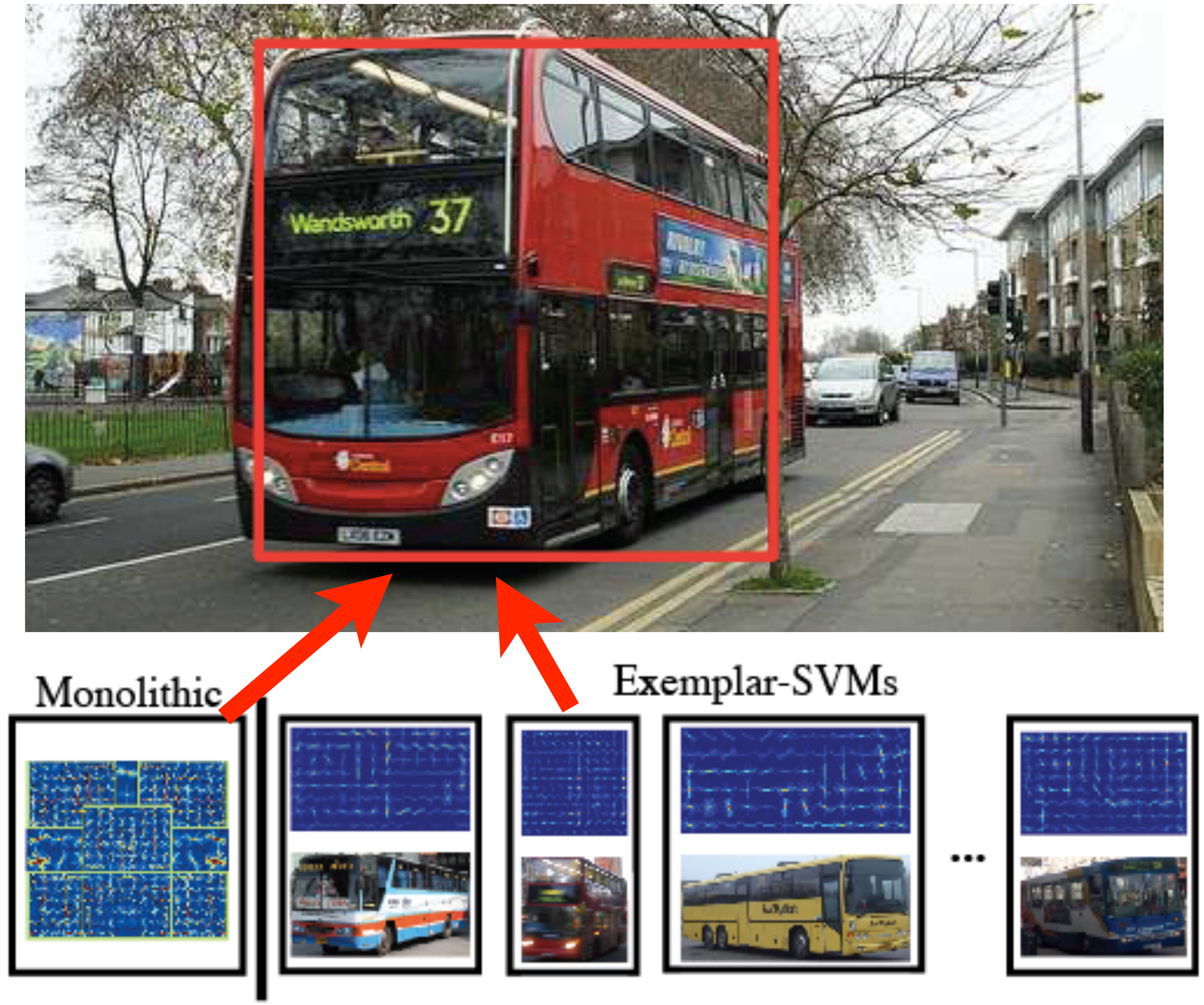
Overview

- Part I: Creating **Visual Associations**
  - Per-Exemplar Distance Functions & Multiple Segmentations [CVPR 2008]
  - Exemplar-SVMs [ICCV 2011]
- Part II: Utilizing **Visual Memex**
  - Object Interpretation [ICCV 2011]
  - Context Challenge [NIPS 2009]

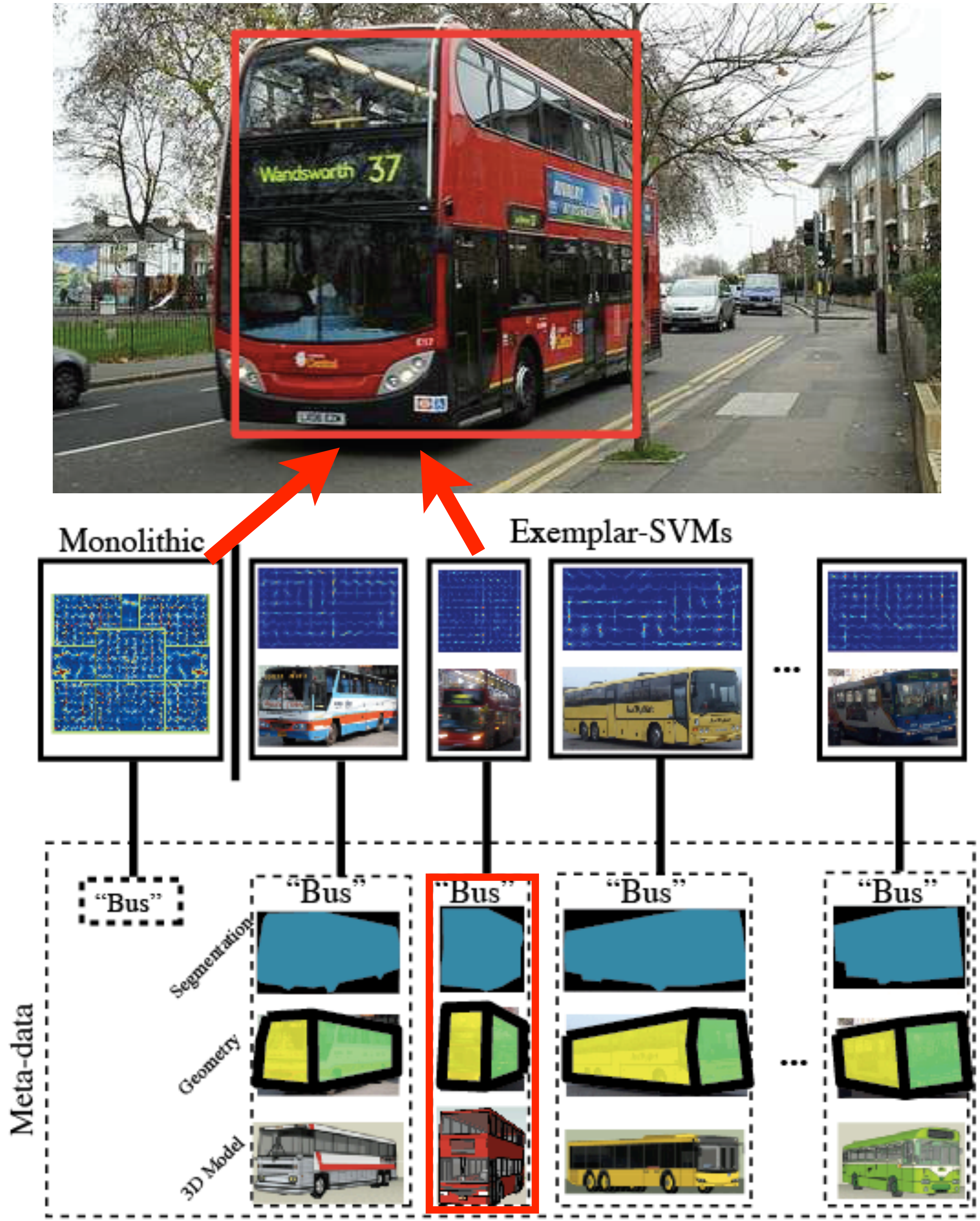
Object Interpretation: Beyond Bounding Boxes

- Let’s first take a look at the output of typical object category bounding box detector

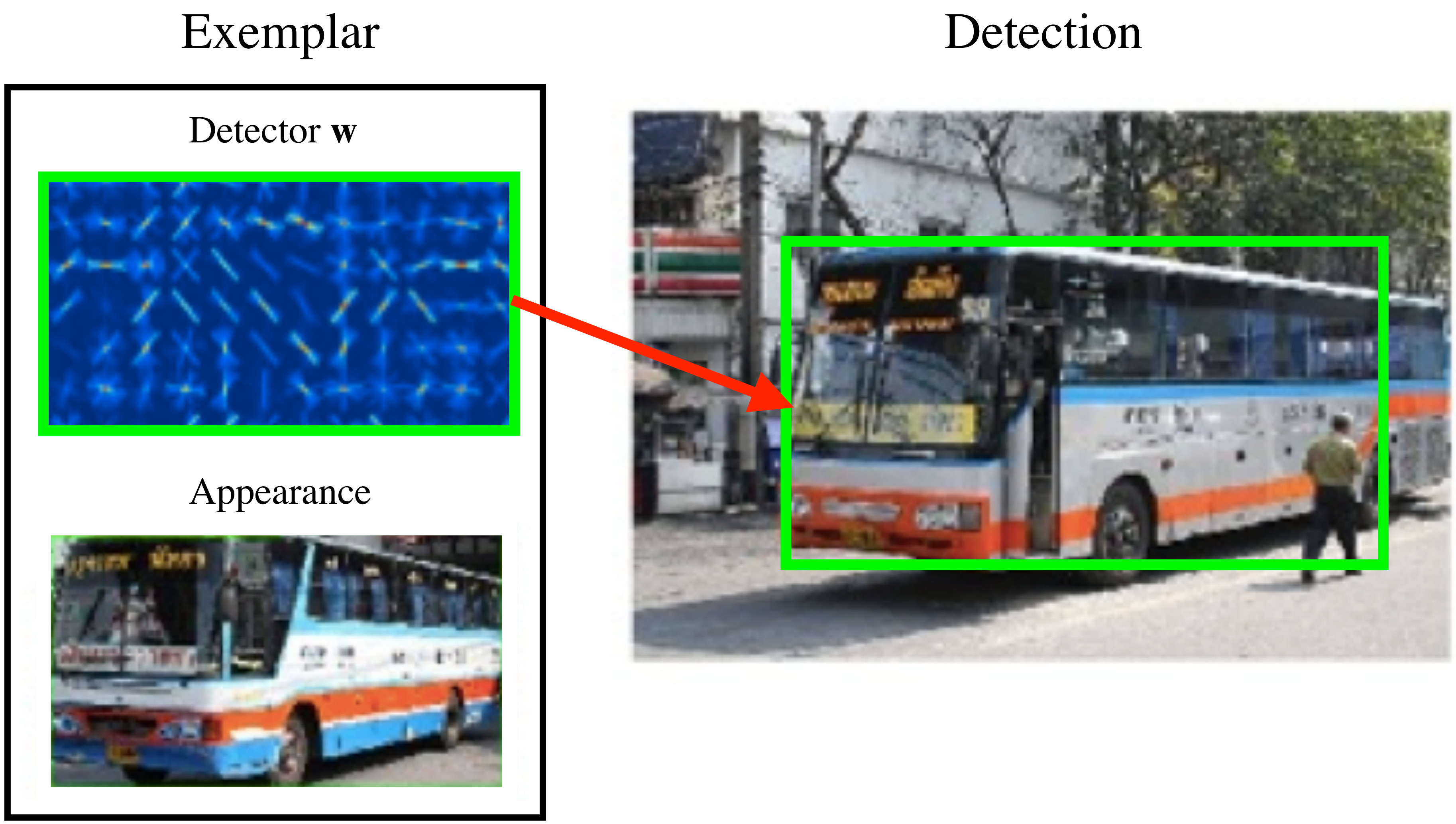

Wendsworth 37



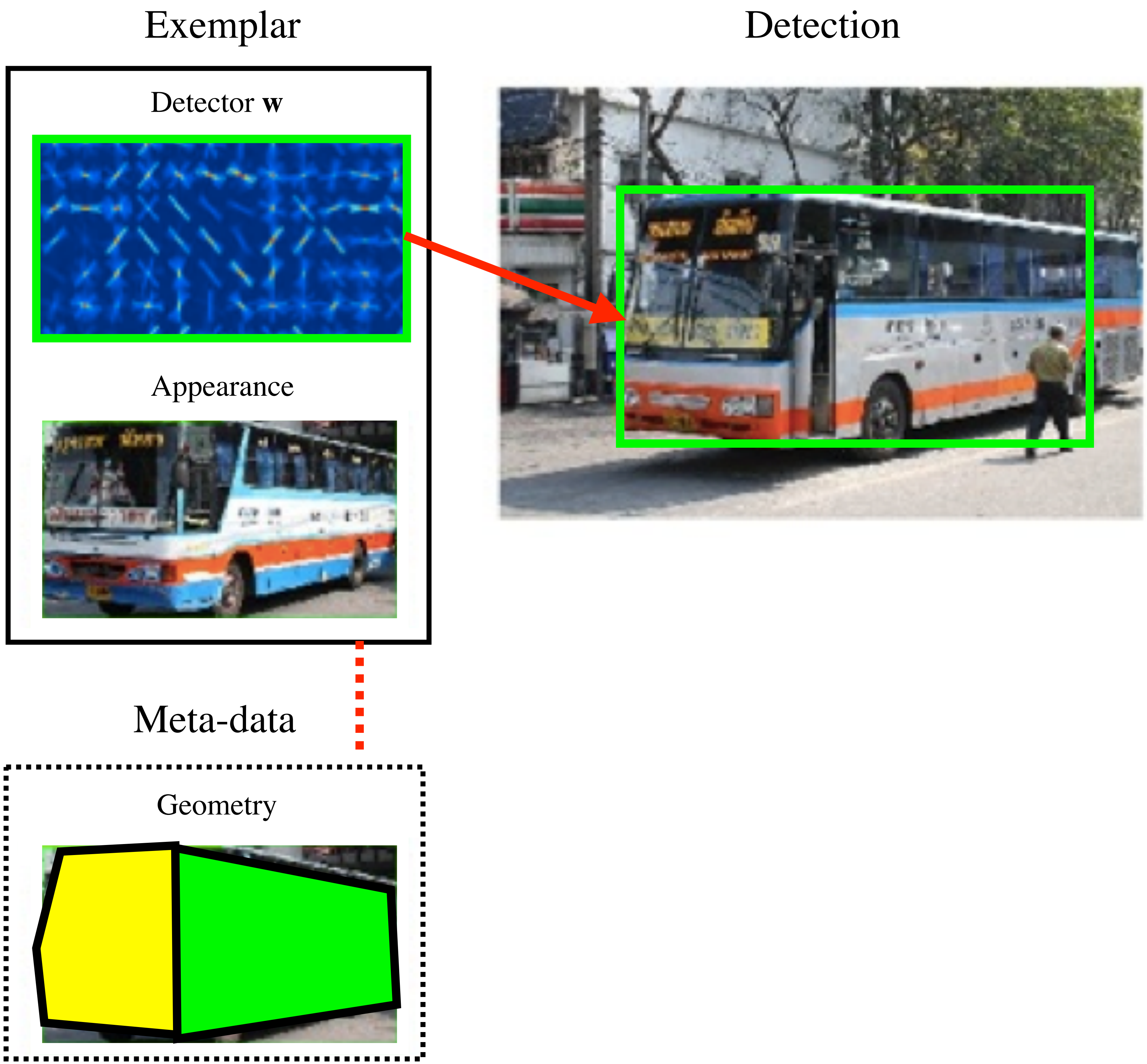
Monolithic
Exemplar-SVMs
...


Monolithic
Exemplar-SVMs
"Bus"
"Bus"
"Bus"
"Bus"
"Bus"
Segmentation
Geometry
3D Model
Meta-data

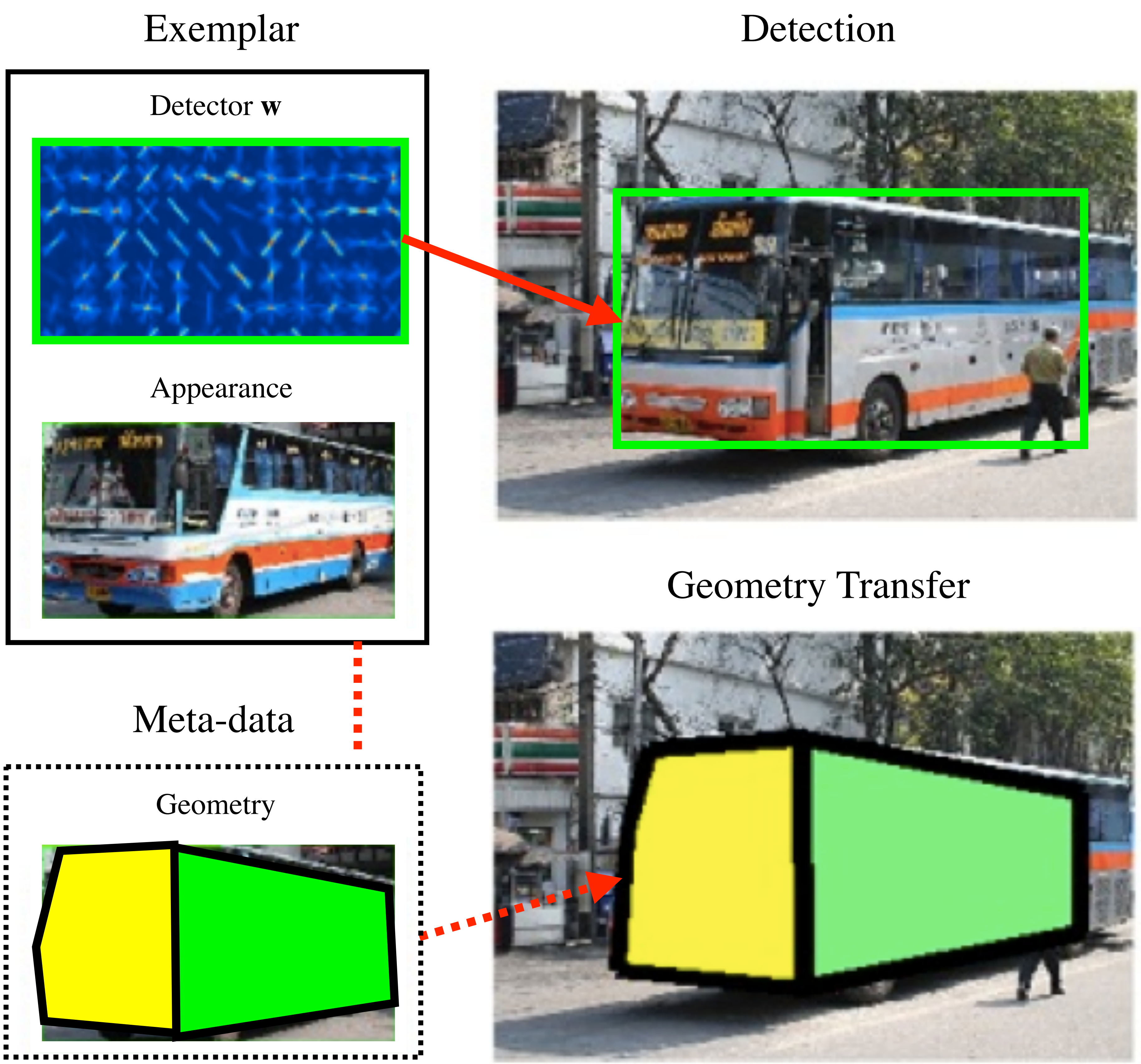
Task 1: Geometry

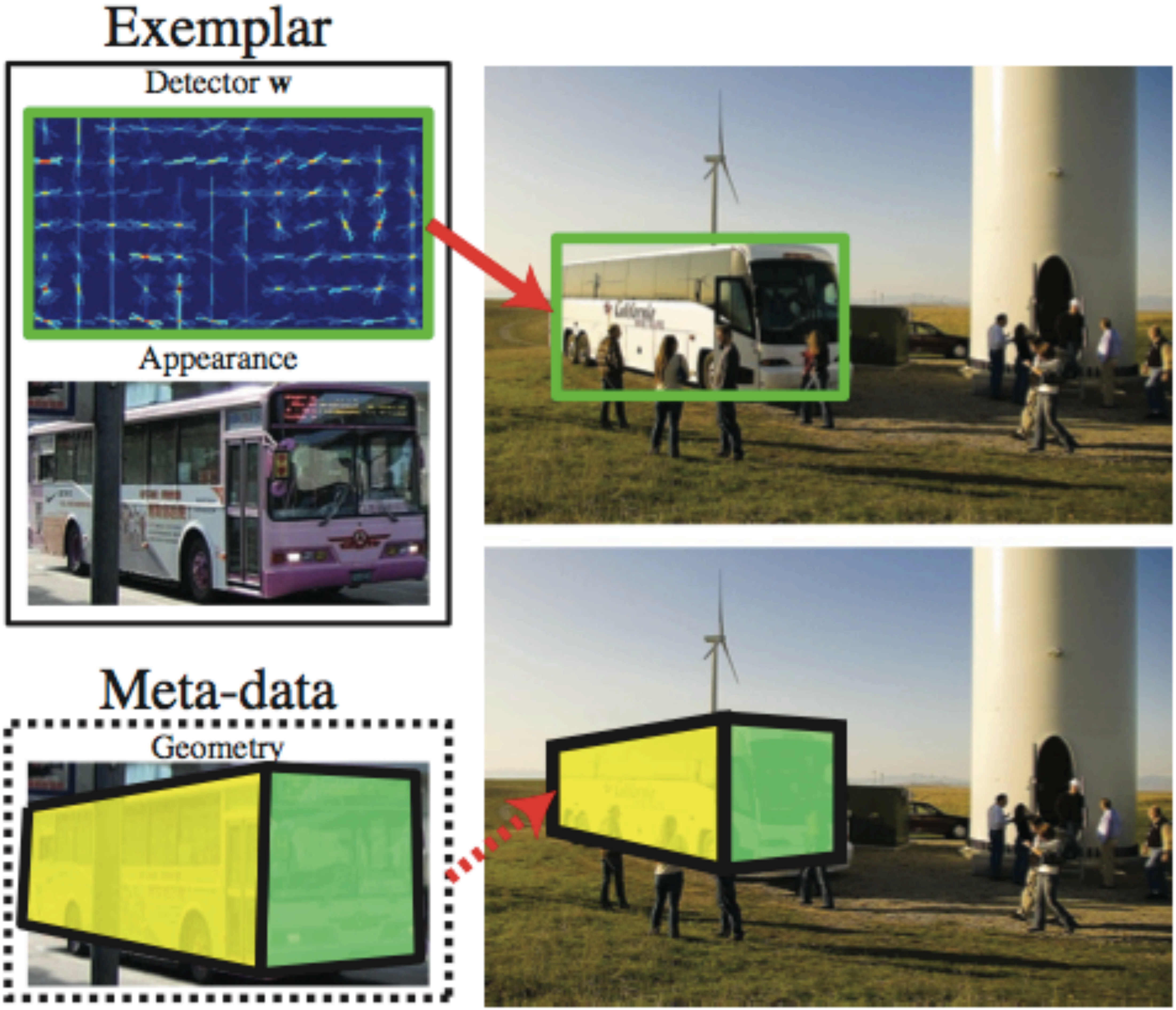


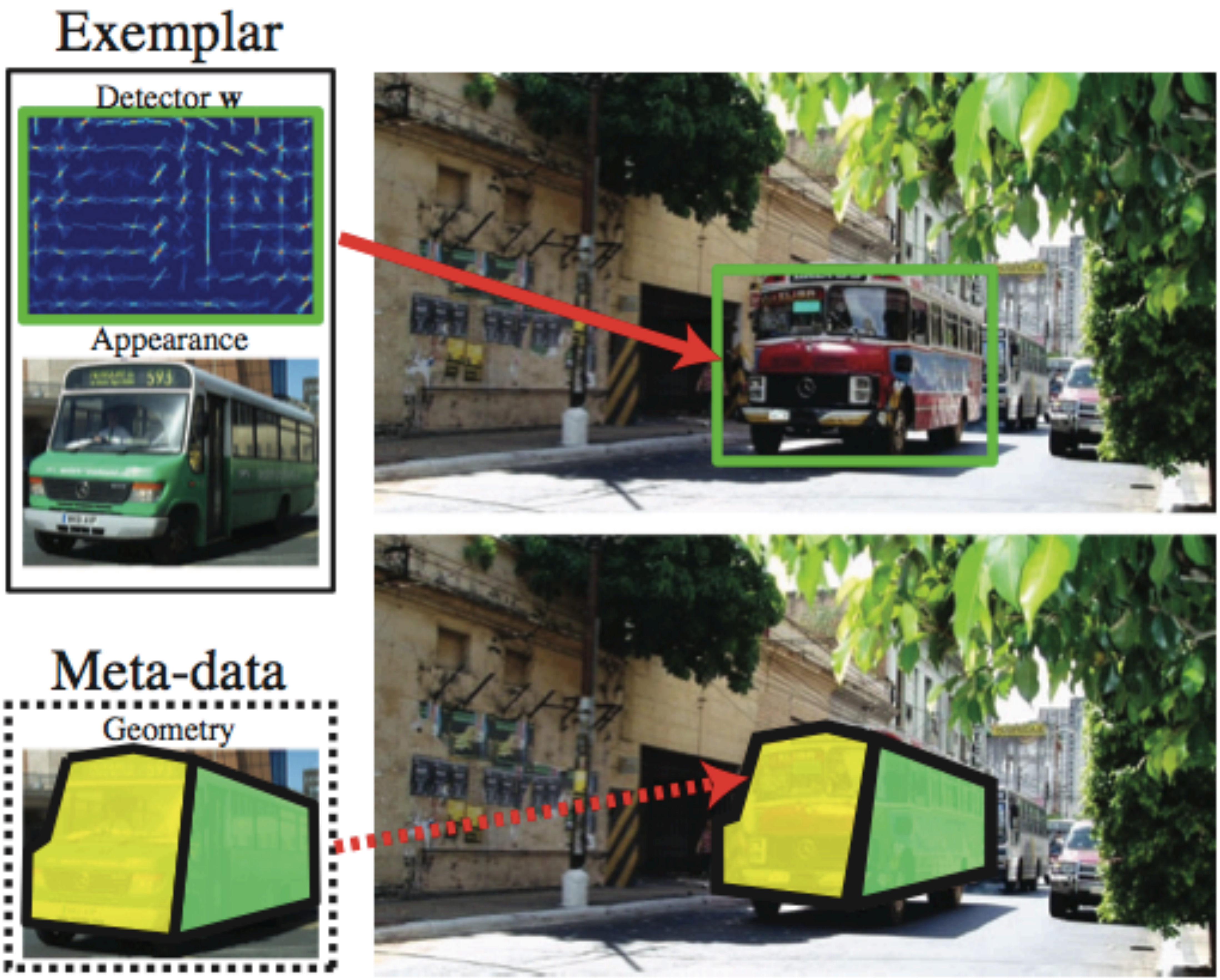
Task 1: Geometry



Task 1: Geometry




Exemplar
Detector w
Appearance
Meta-data
Geometry


Exemplar
Detector w
Appearance
Meta-data
Geometry


Exemplar
Detector w
Appearance
Meta-data
Geometry

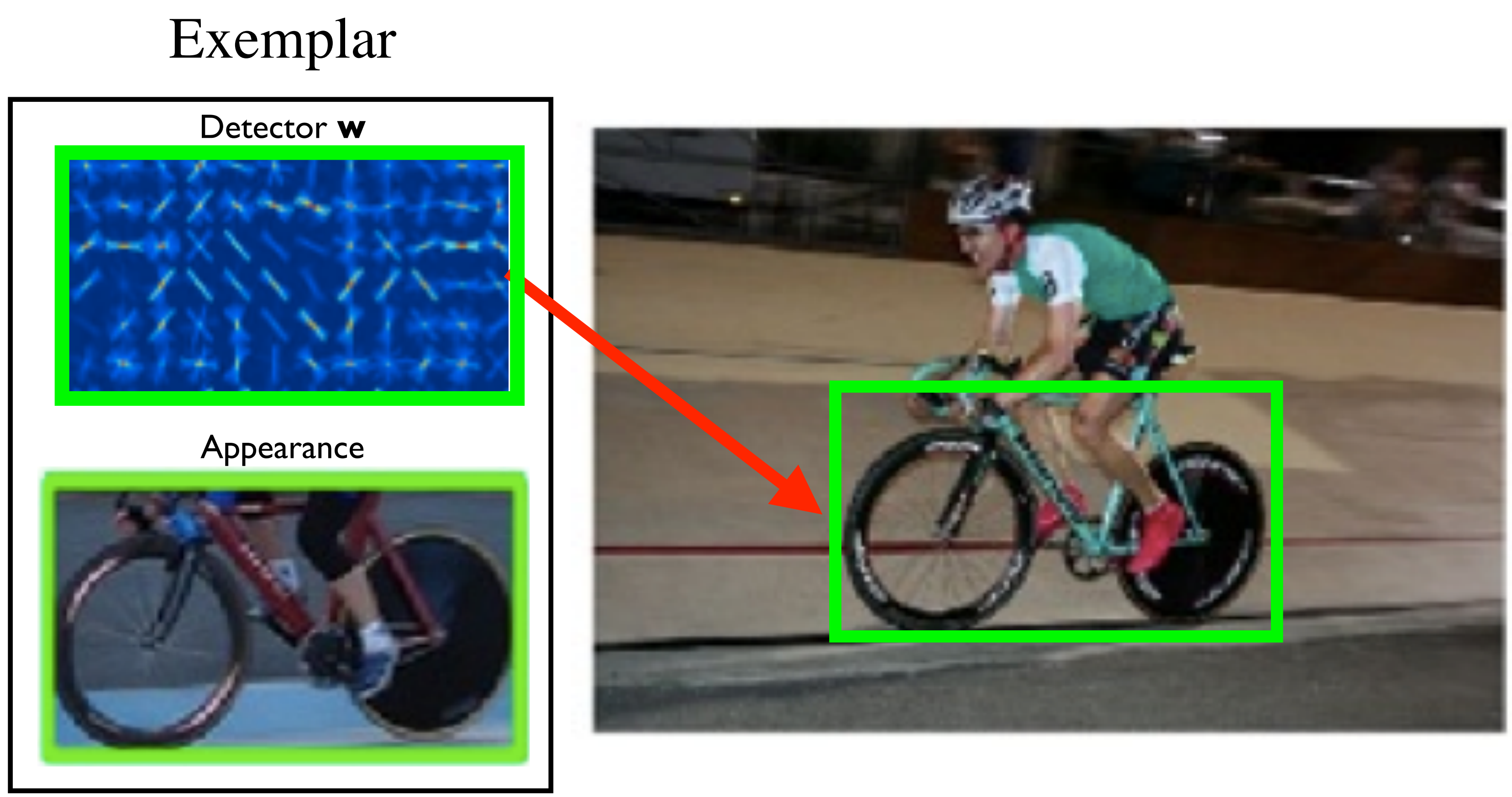
Task 1: Evaluation on Buses

- measure pixelwise accuracy on the 3-class geometric-labeling problem: "left," "front," "right"-facing

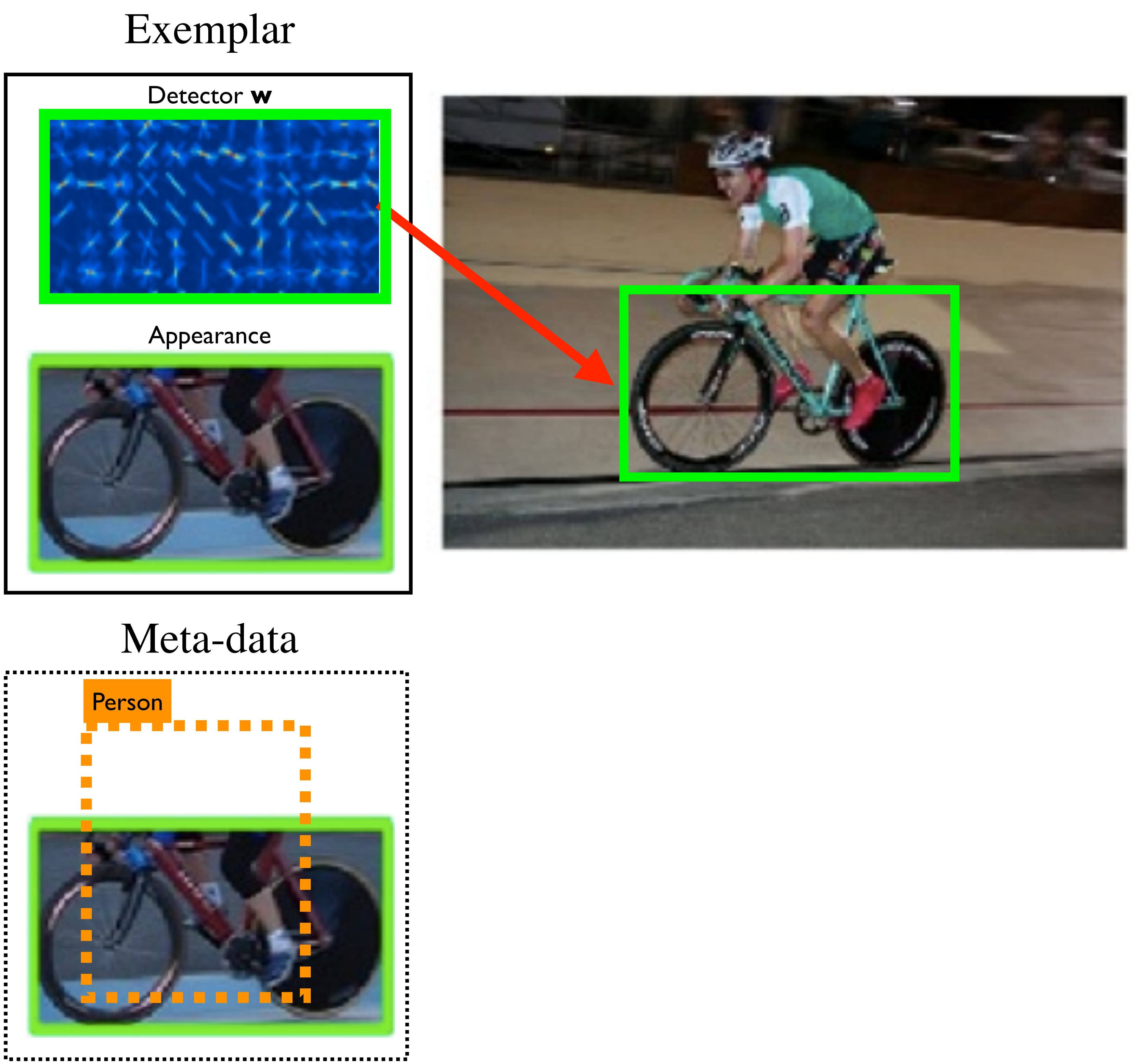
- 43.0% Hoiem et al. 2005
- 51.0% Monolithic Detector* + NN
- **62.3**% Exemplar-SVMs

*Felzenszwalb et al. 2010

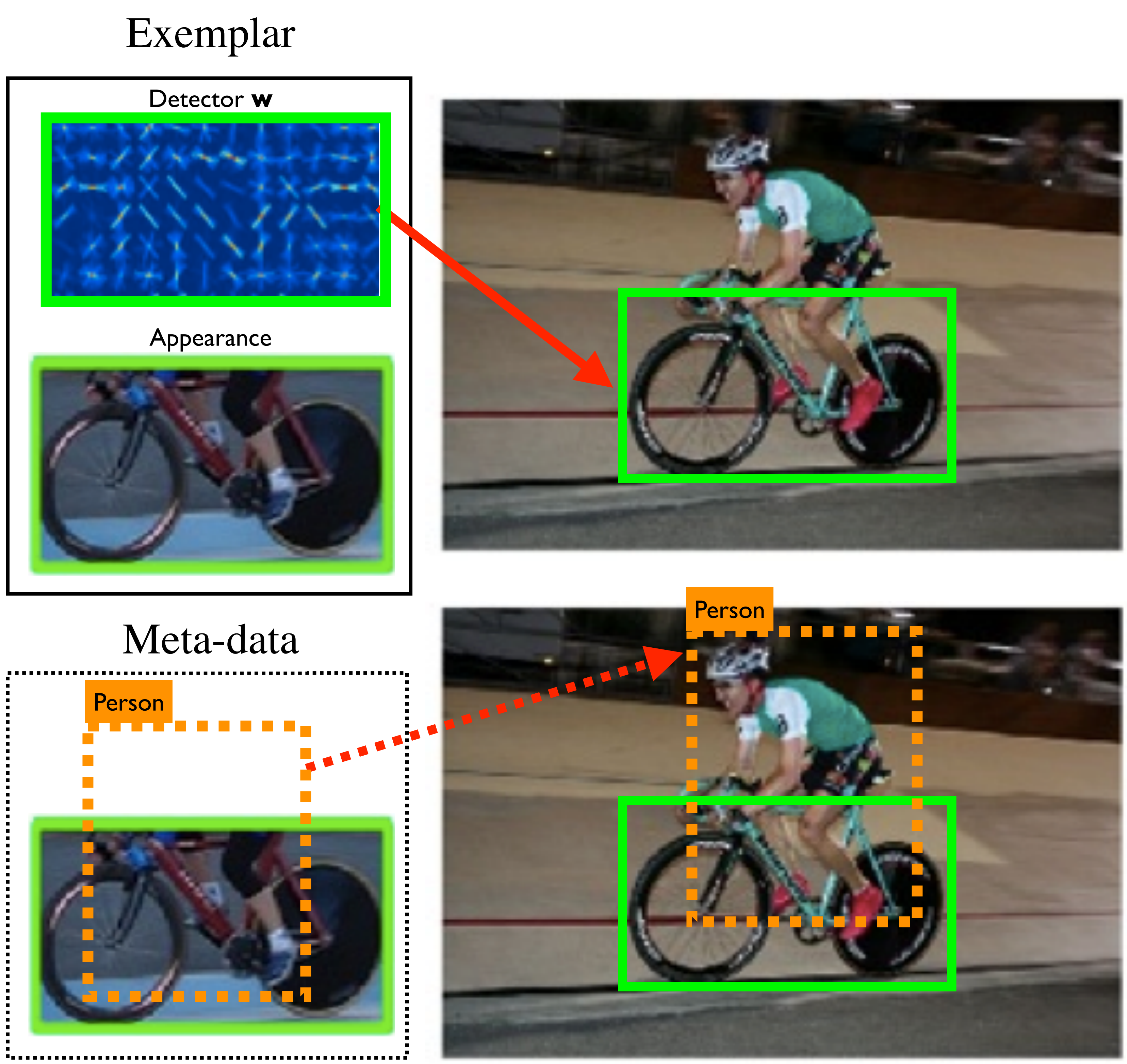
Task II: Person Prediction

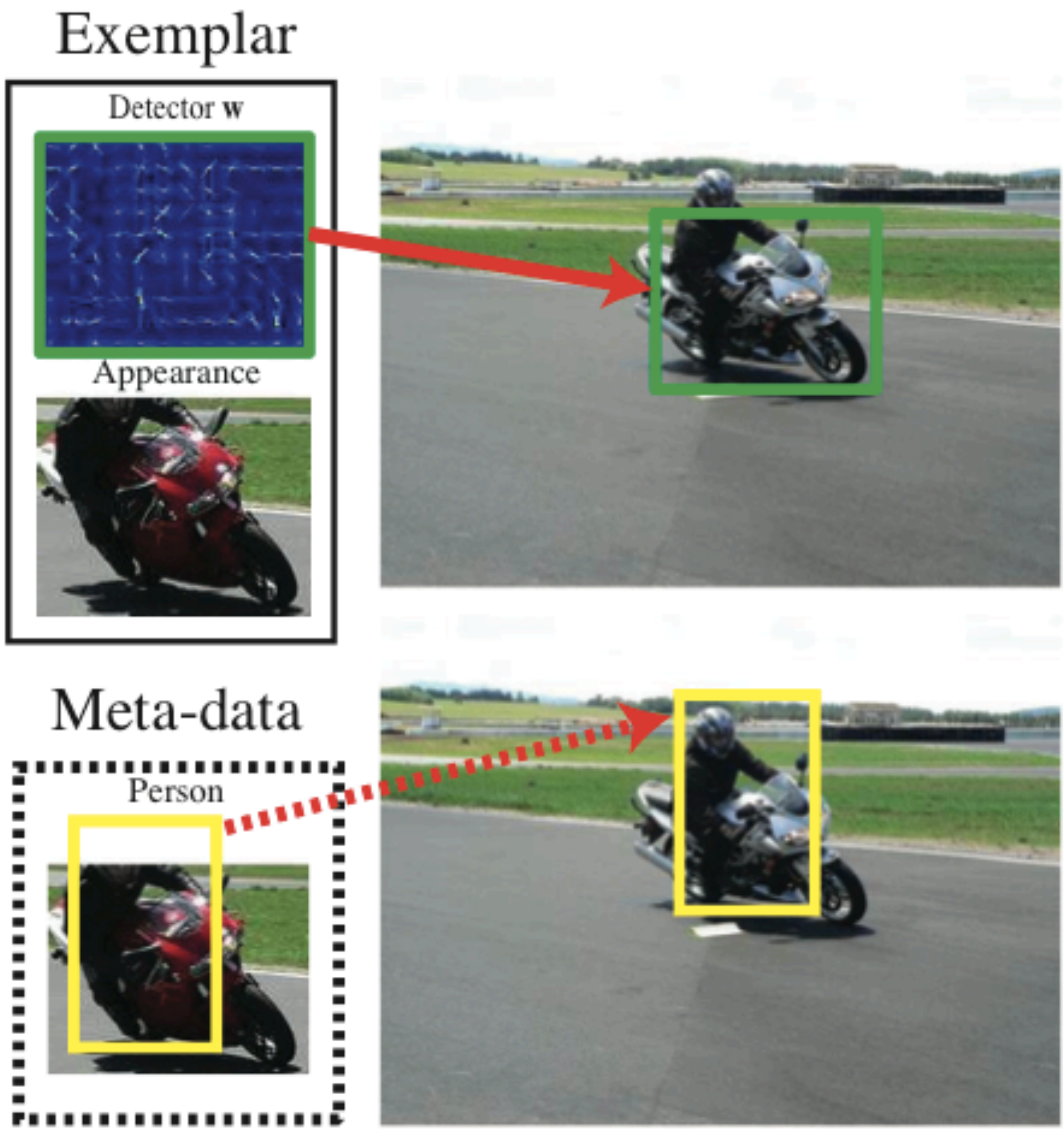


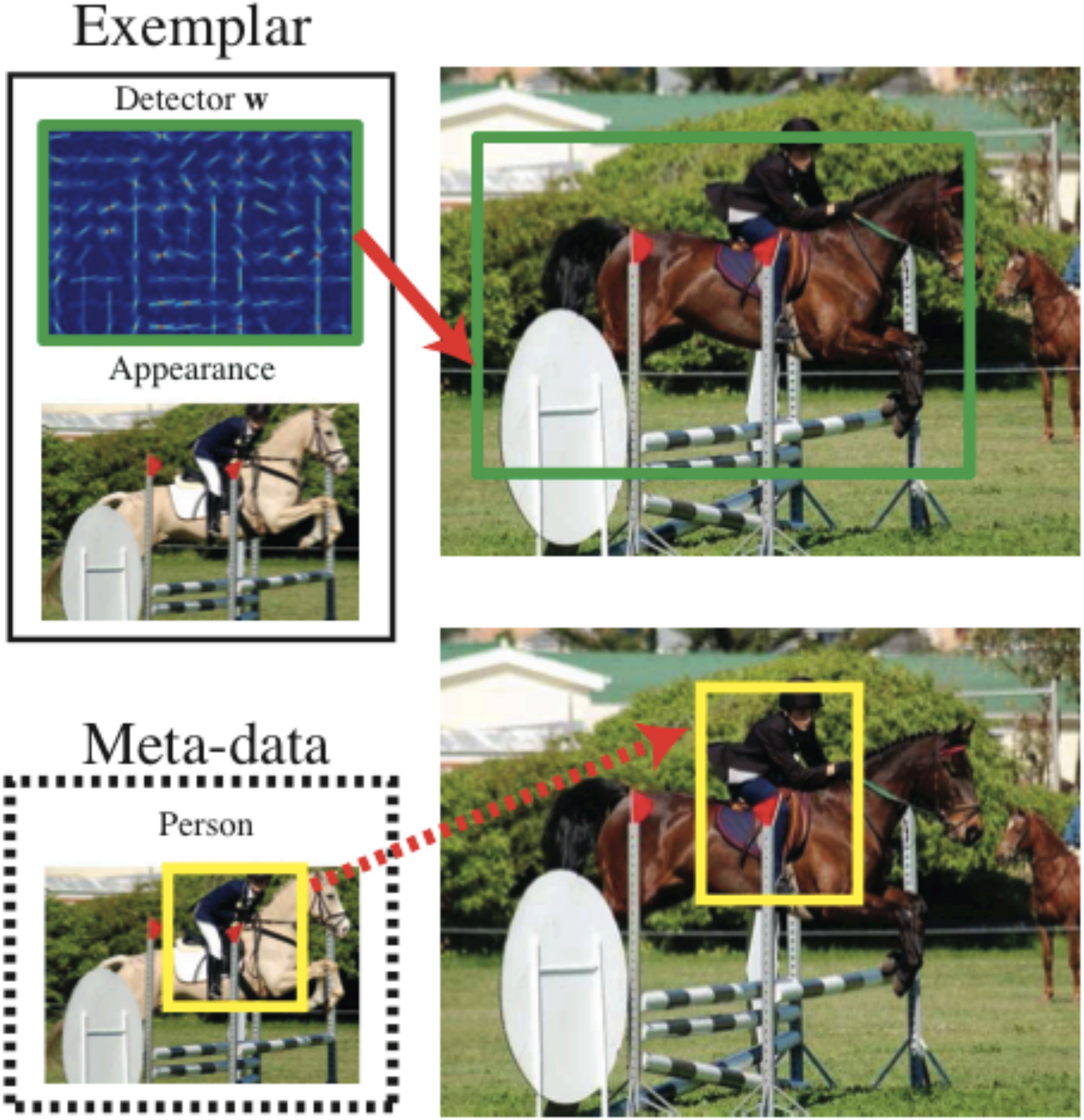
Task II: Person Prediction



Task II: Person Prediction




Exemplar
Detector w
Appearance
Meta-data
Person


Exemplar
Detector w
Appearance
Meta-data
Person

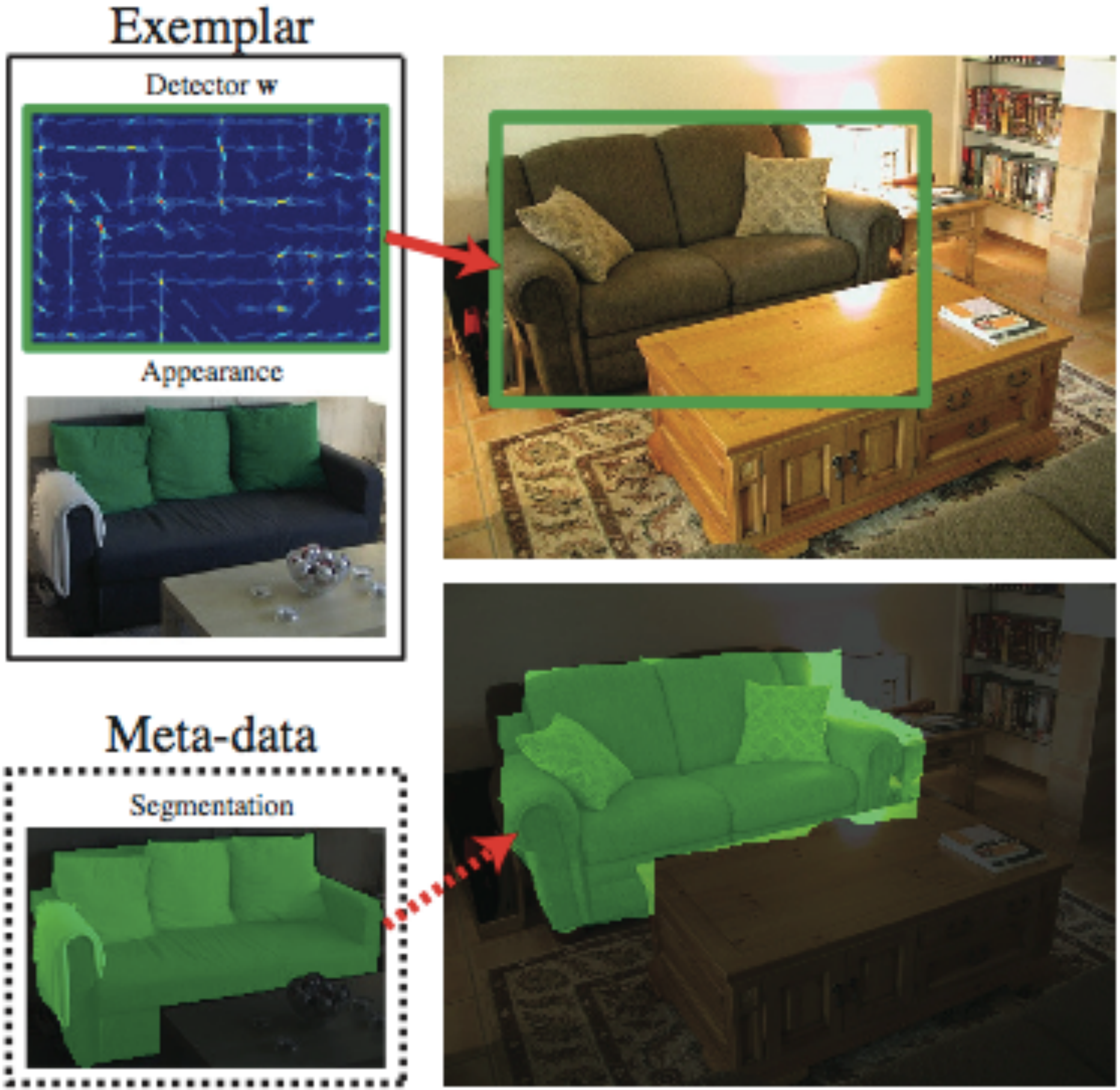
Task II: Evaluation

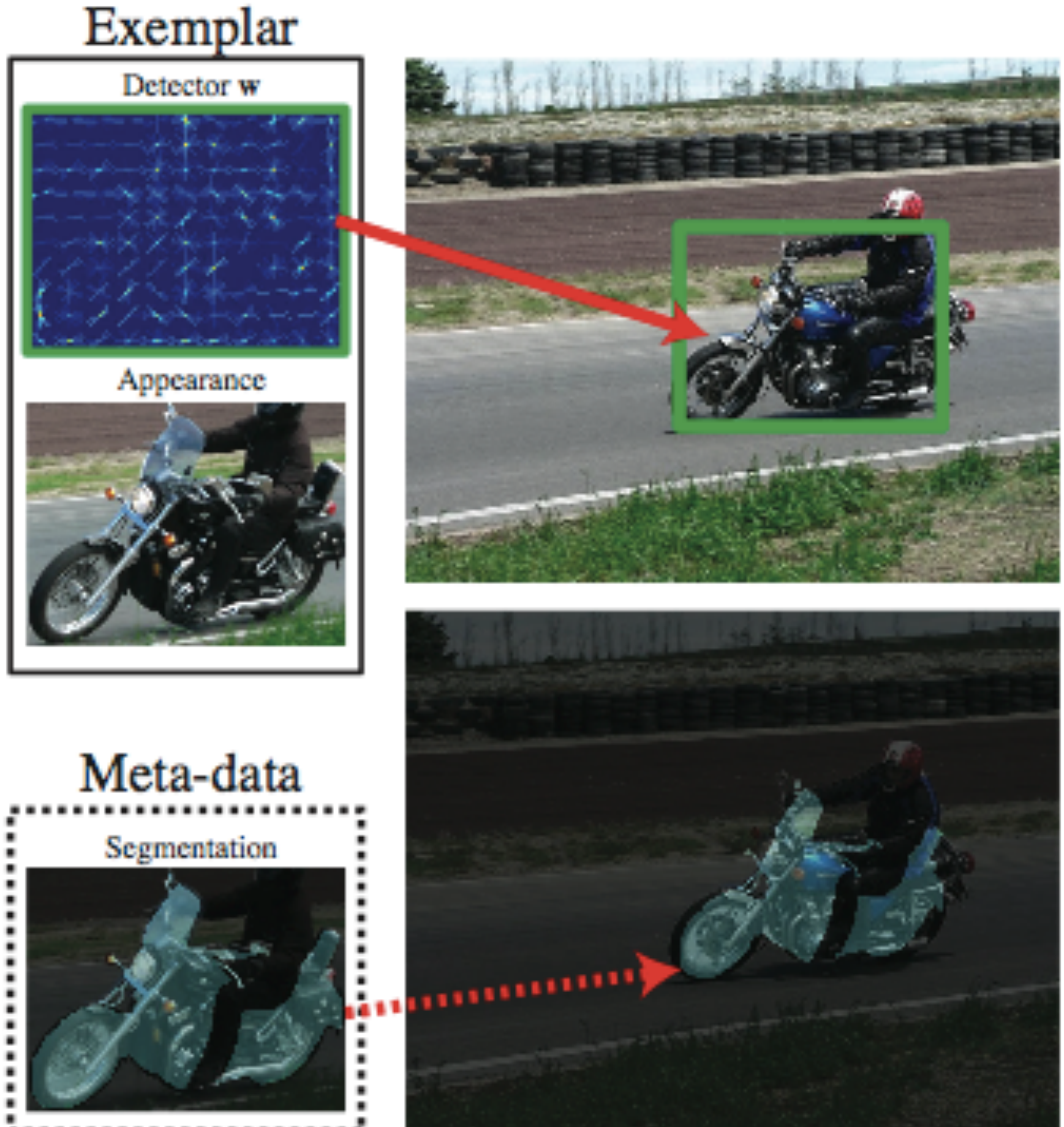
Category	Majority Voting	us
bicycle	63.4%	**72.8%**
motorbike	50.0%	**67.4%**
horse	62.6%	**77.2%**

Table 2. **Is there a person riding this horse?** We predict from our bicycle, motorbike, and horse detectors whether there is a person riding the object. Our approach is better than the majority vote baseline, suggesting that exemplars are useful at predicting nearby, related objects.

Qualitative Examples

- Segmentation Transfer

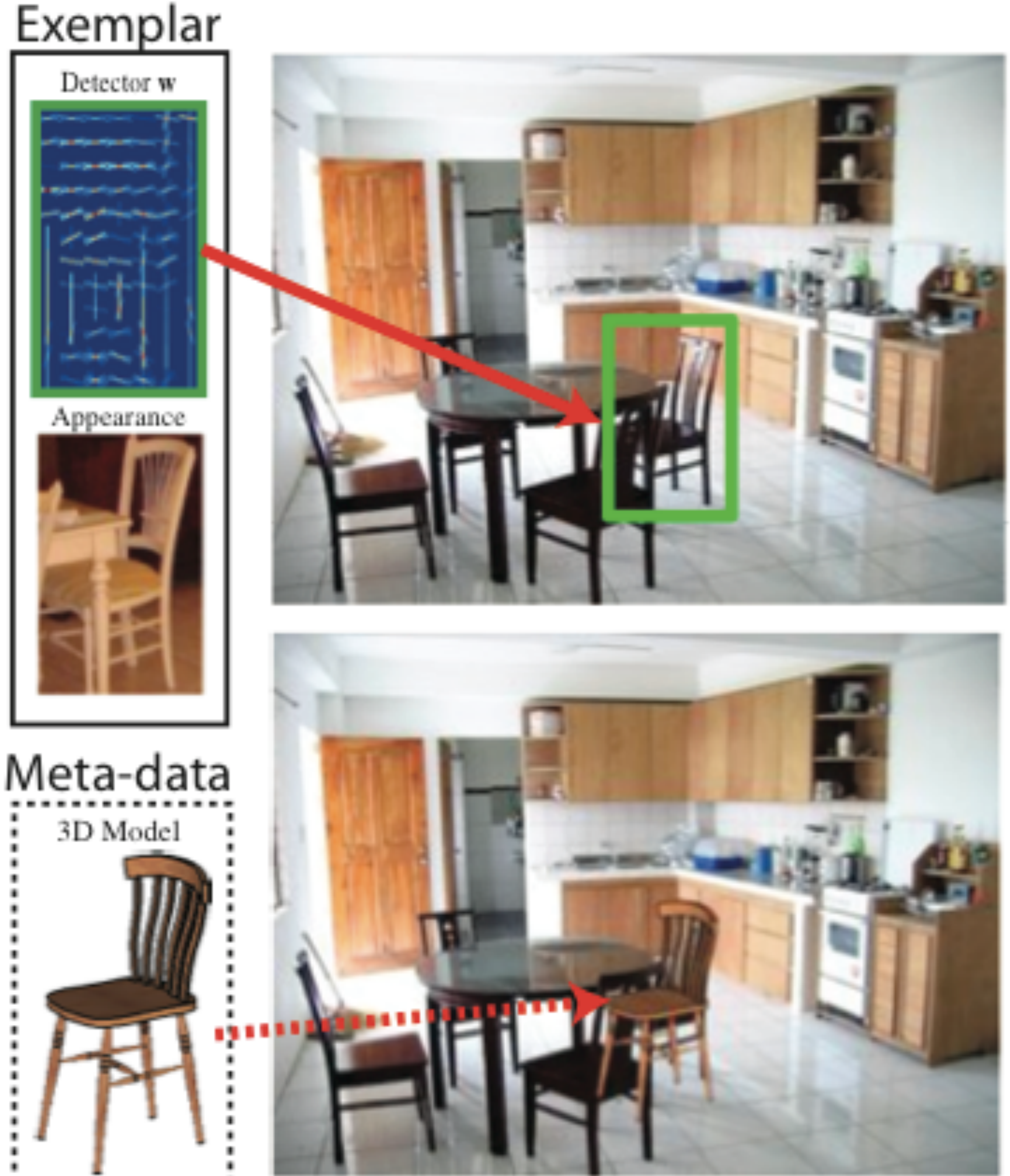

Exemplar
Detector w
Appearance
Meta-data
Segmentation


Exemplar
Detector w
Appearance
Meta-data
Segmentation

3D Model Transfer

Manually align 3D model from Google 3D Warehouse with a subset of PASCAL VOC "chair" exemplars

Exemplar
Detector w
Appearance
Meta-data
3D Model



Exemplar
Detector w
Appearance
Meta-data
3D Model

Overview

- Part I: Creating **Visual Associations**
  - Per-Exemplar Distance Functions & Multiple Segmentations [CVPR 2008]
  - Exemplar-SVMs [ICCV 2011]

- Part II: Utilizing **Visual Memex**
  - Object Interpretation [ICCV 2011]
  - Context Challenge [NIPS 2009]

“How far can you go without running an object detector?”

Antonio Torralba, 2003

Torralba's Context Challenge

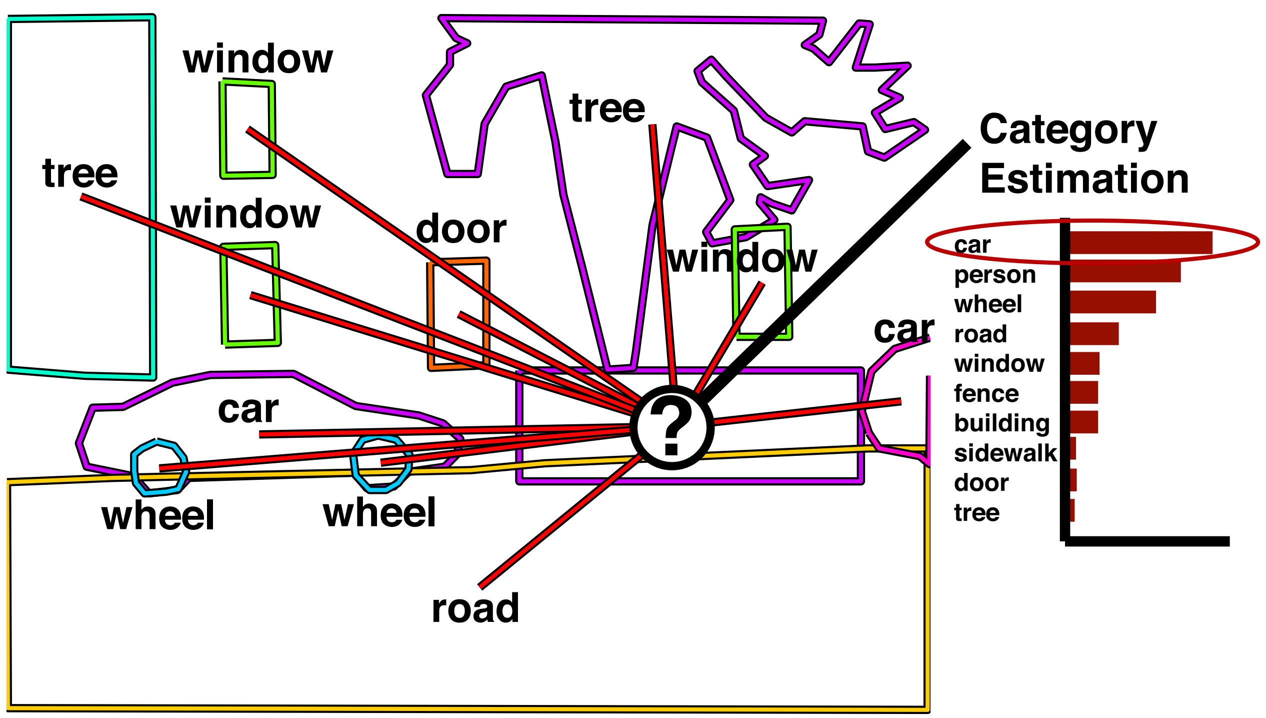
Slide by Antonio Torralba

Torralba's Context Challenge

Slide by Antonio Torralba

Our Context Challenge

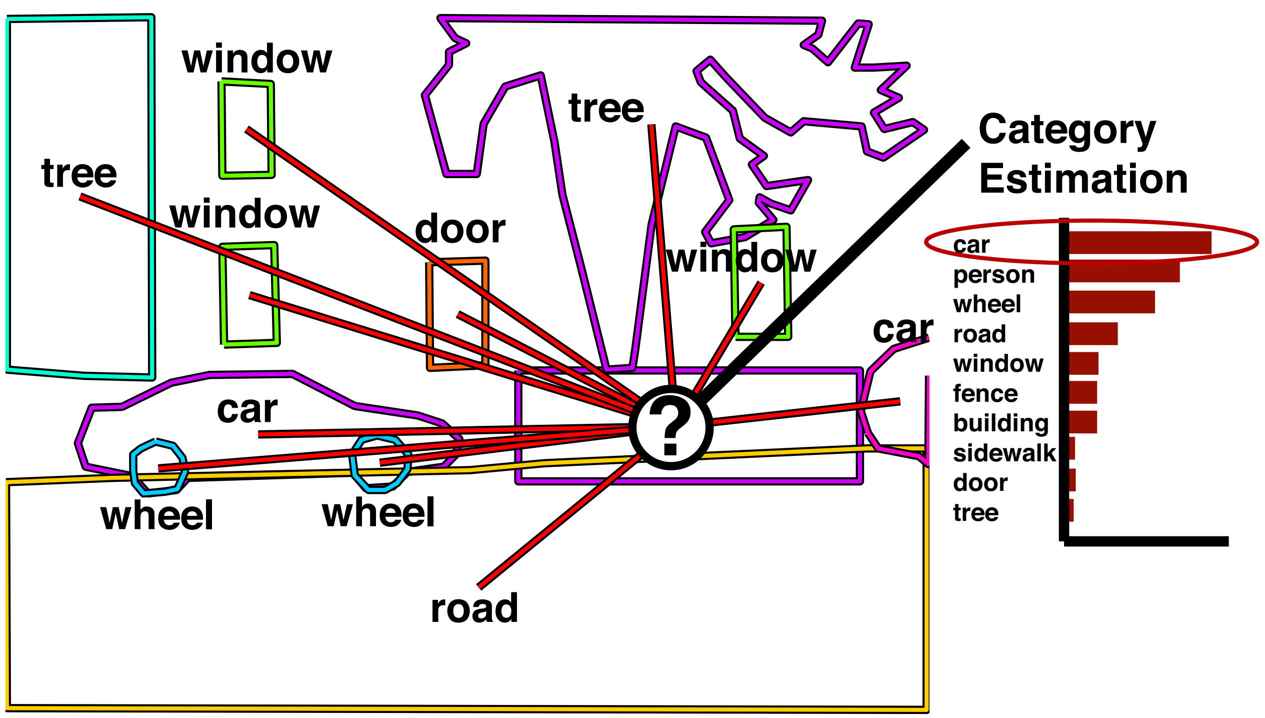
Given Categories



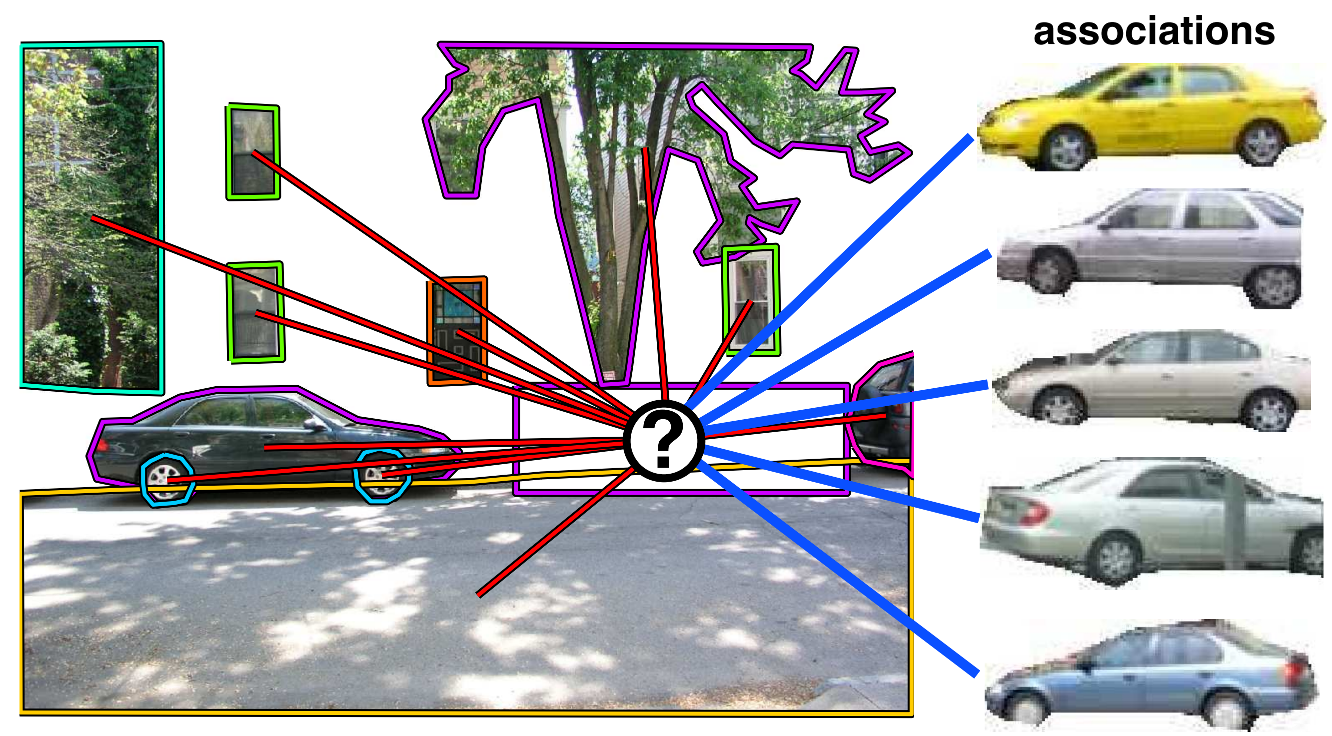
Malisiewicz et al. 2009

Our Context Challenge

Given Categories



Given Appearances

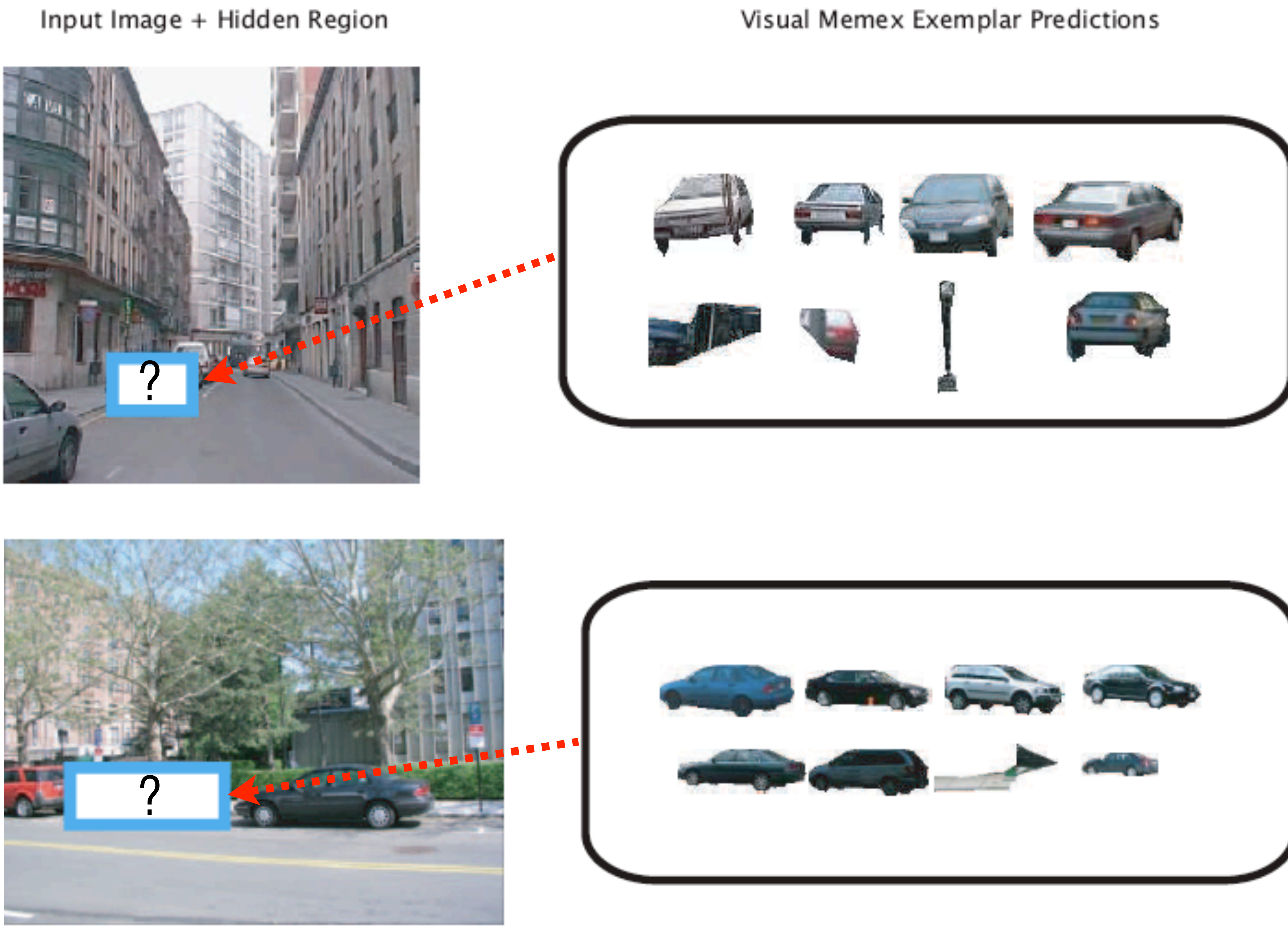


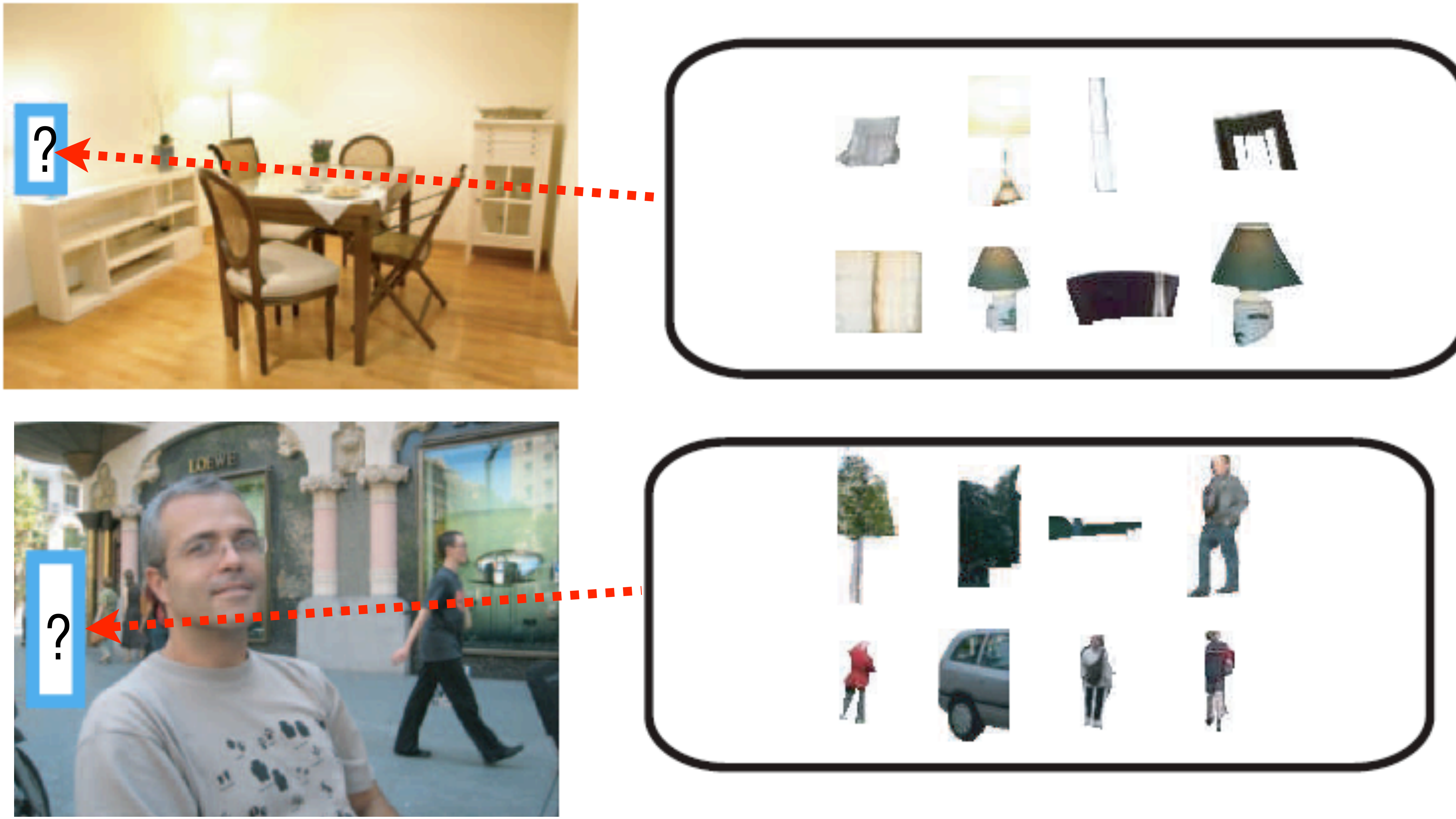
Malisiewicz et al. 2009

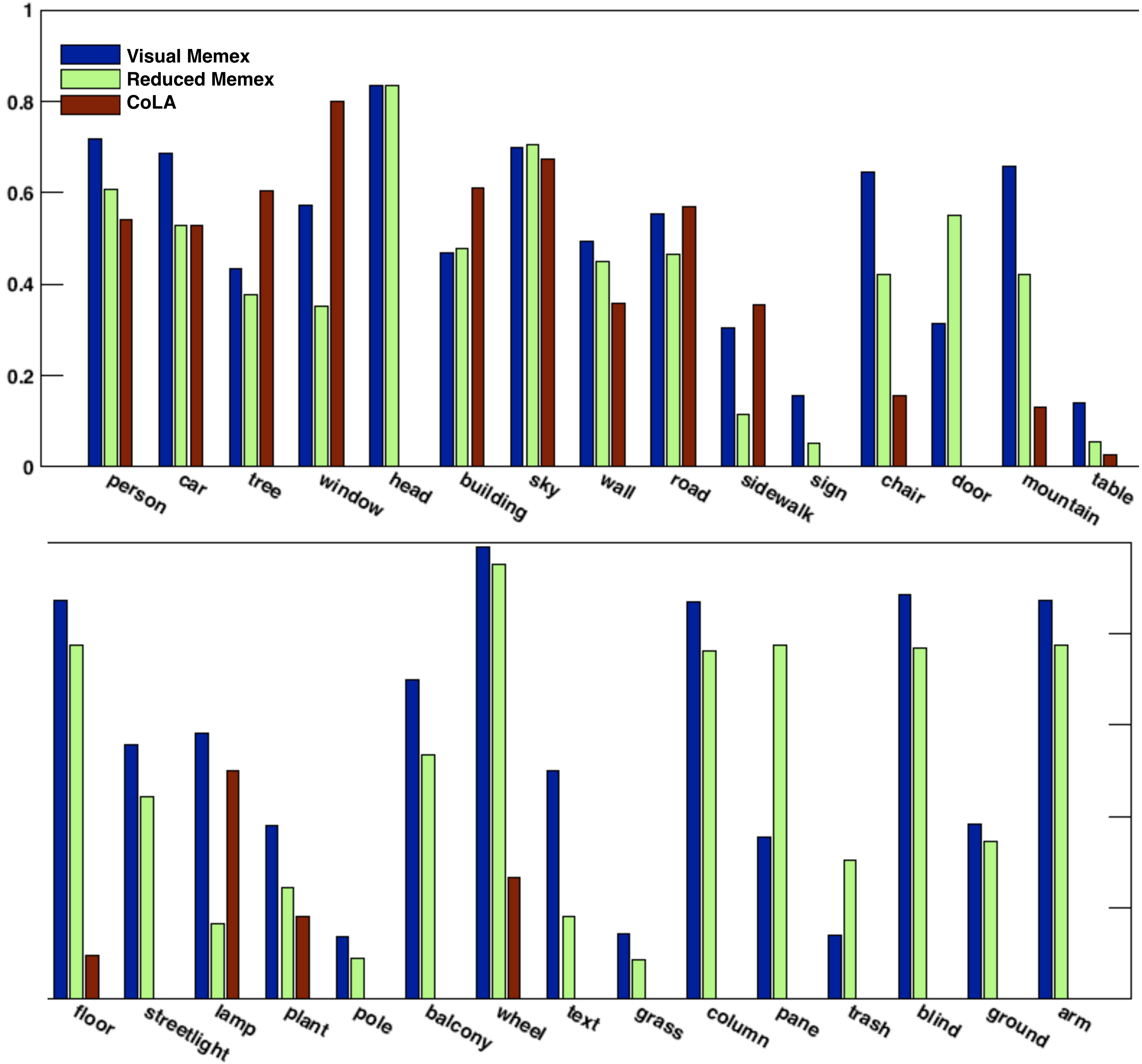
3 Models

- Visual Memex
  - exemplar-based
  - non-parametric object-object relationships
- CoLA*
  - category-based
  - parametric object-object relationships
- Reduced Memex
  - category-based
  - non-parametric object-object relationships

*Galleguillos et al. 2008


Input Image + Hidden Region
Visual Memex Exemplar Predictions
?
?


?
LOEWE
?

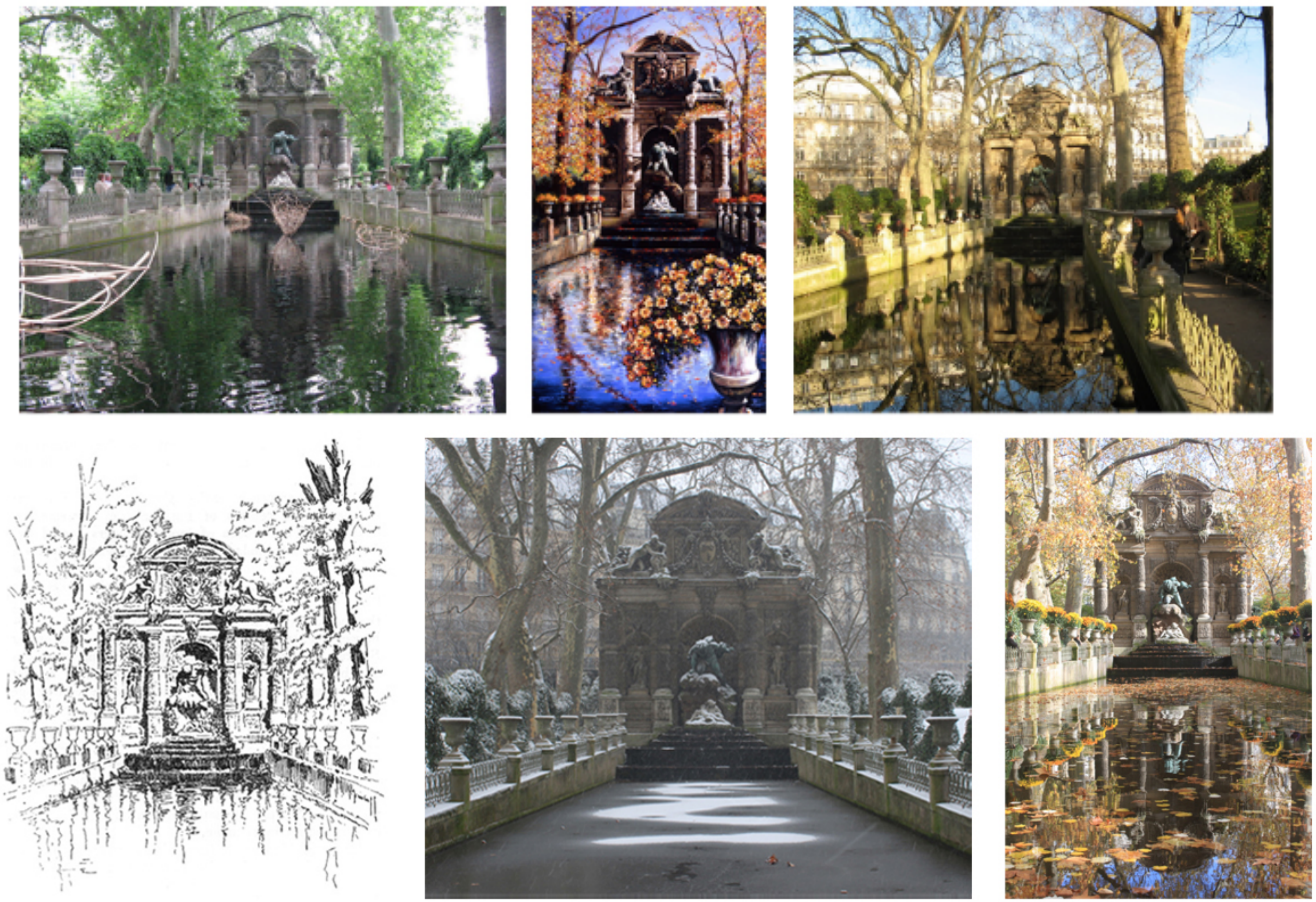


Visual Memex
Reduced Memex
CoLA
1
0.8
0.6
0.4
0.2
0
person
car
tree
window
head
building
sky
wall
road
sidewalk
sign
chair
door
mountain
table
floor
streetlight
lamp
plant
pole
balcony
wheel
text
grass
column
pane
trash
blind
ground
arm

Context Challenge Results

	Overall	Per-Category
Visual Memex	**0.527**	**0.534**
Reduced Memex	0.430	0.454
CoLA	0.457	0.213

Cross-domain Image Matching



w/ Abhinav Shrivastava

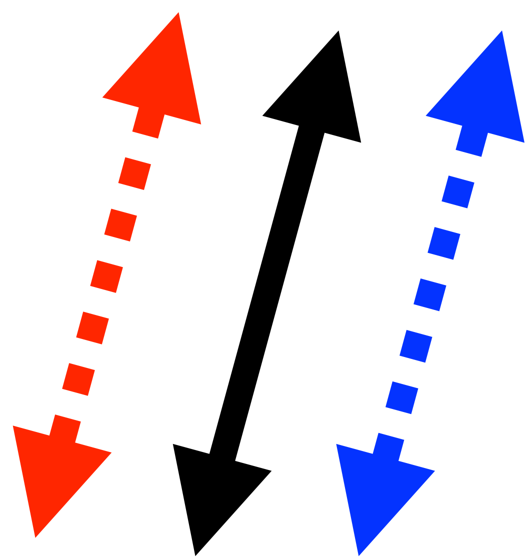
SIGGRAPH ASIA 2011

Learn Exemplar-SVM for query image

Query Image

Learn Exemplar-SVM for query image

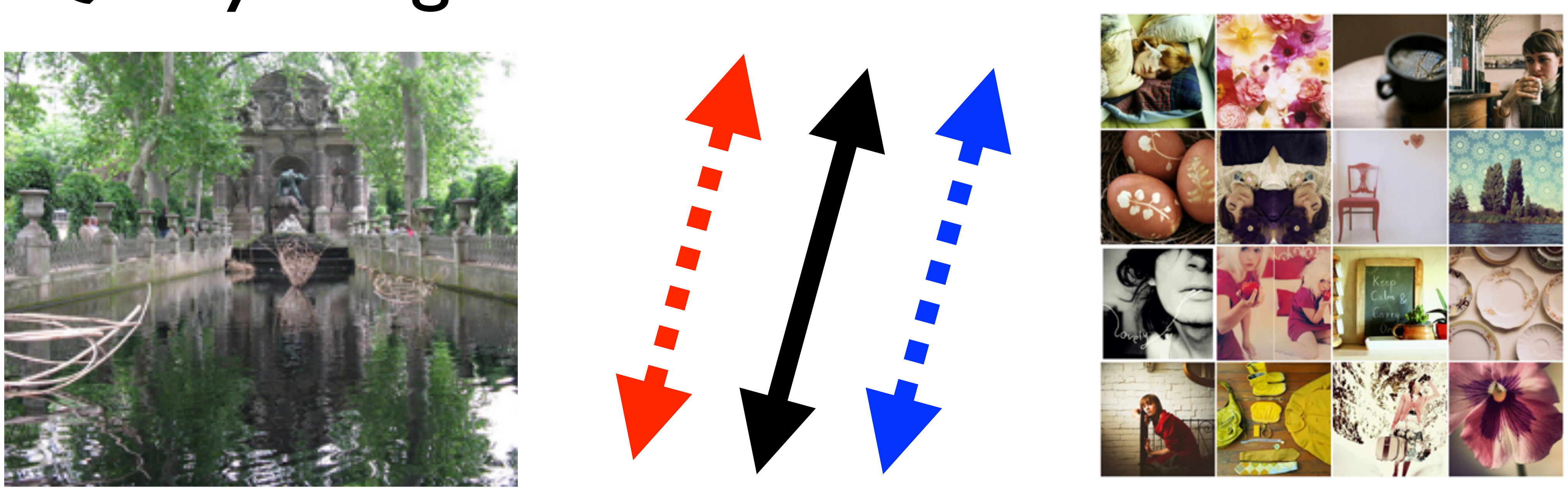
Query Image



Learn Exemplar-SVM for query image

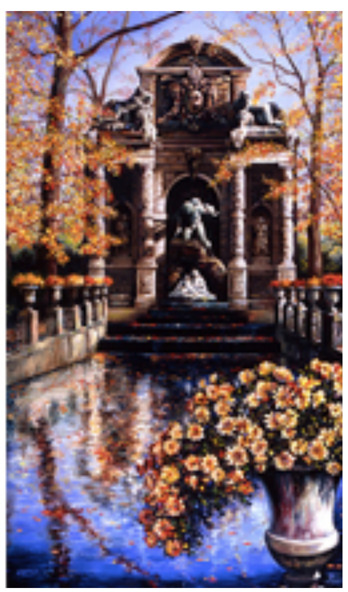
Query Image

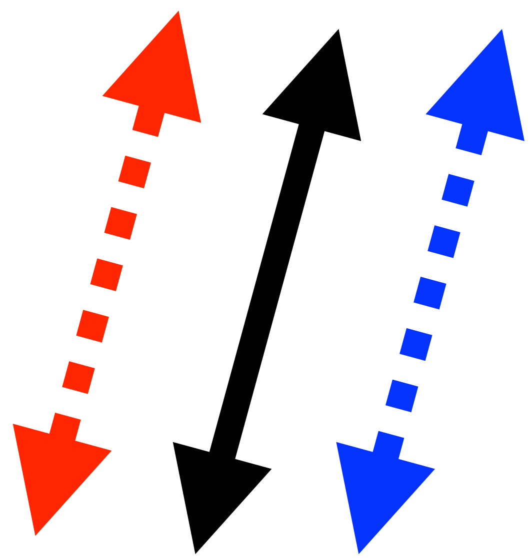
Random Flickr Images



Learn Exemplar-SVM for query painting

Query Painting

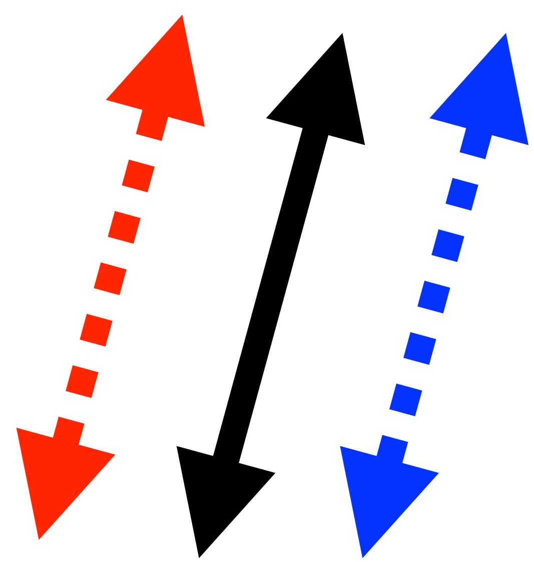




Random Flickr Images

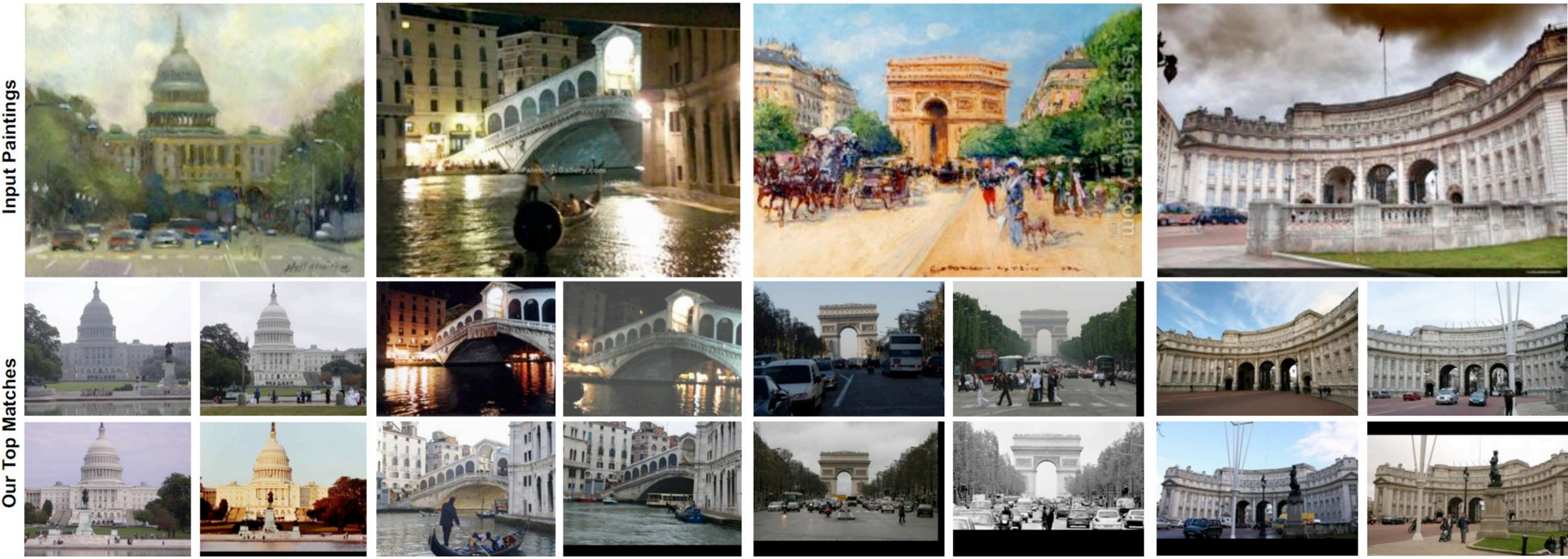
Learn Exemplar-SVM for query sketch

Query Sketch

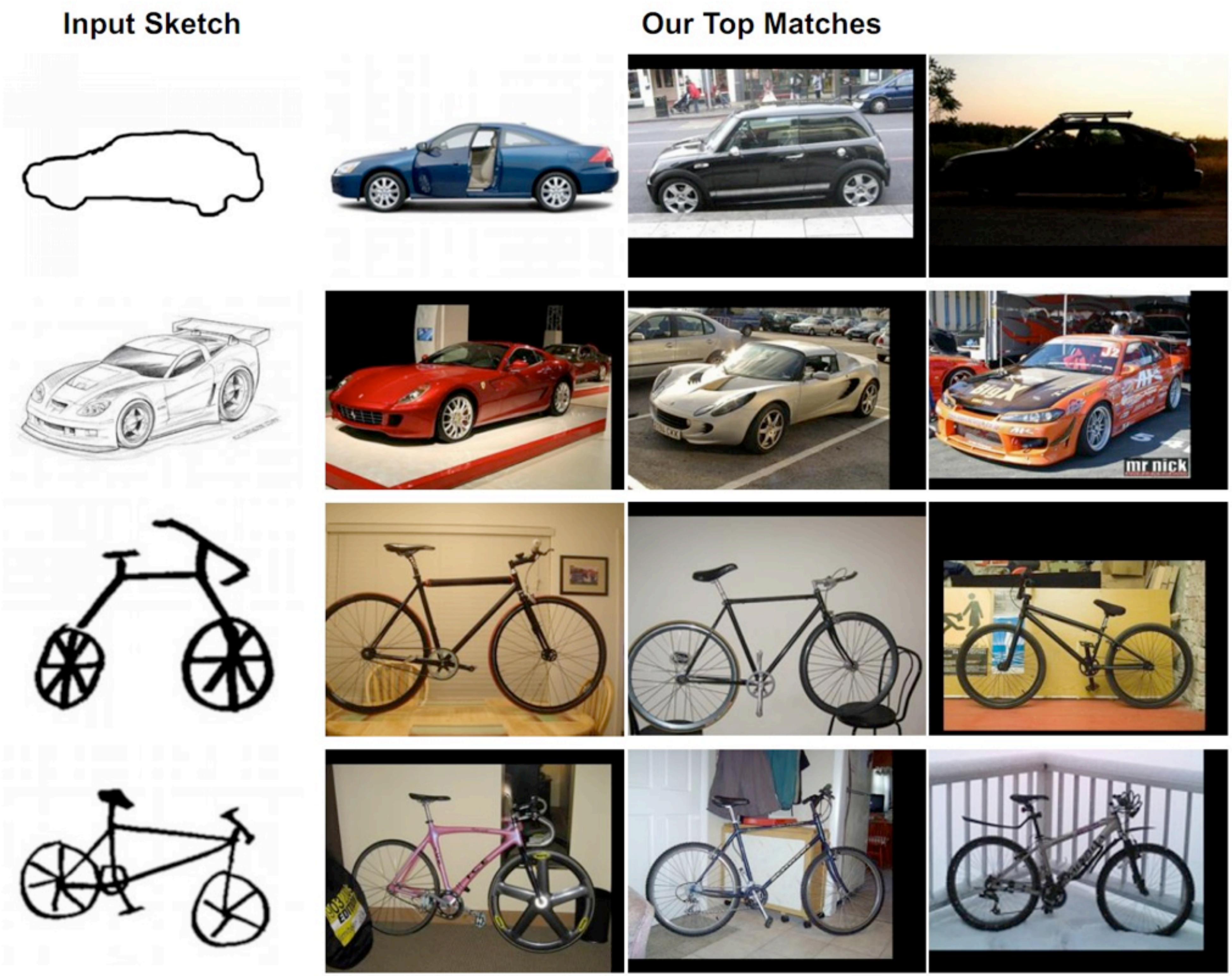


Random Flickr Images

Painting to Image



Sketch to Image



Painting to GPS

IM2GPS: Hays et al. 2008

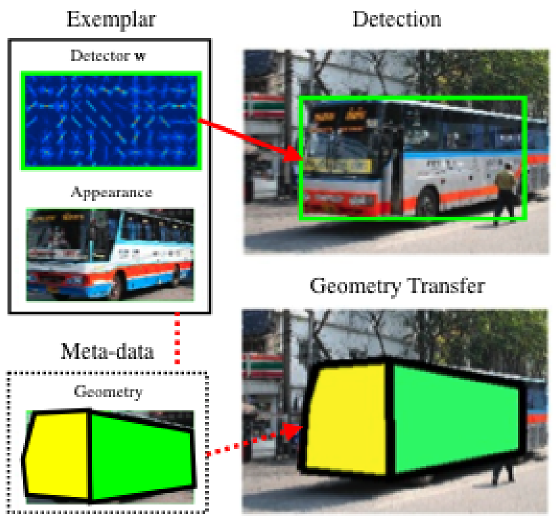
Painting to GPS

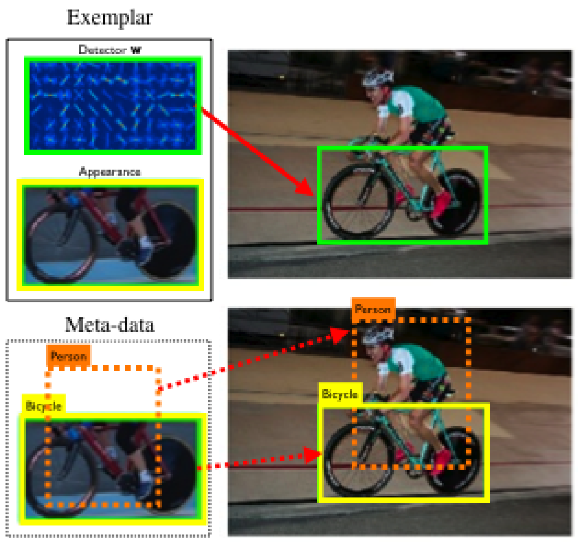
IM2GPS: Hays et al. 2008

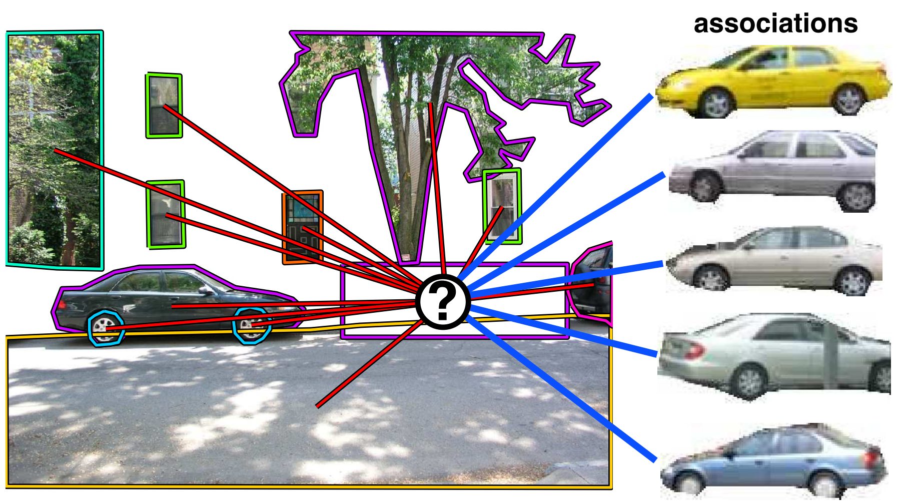
Thesis Conclusions

Thesis Conclusions

- Visual Memex can be used for recognition, interpretation, and prediction

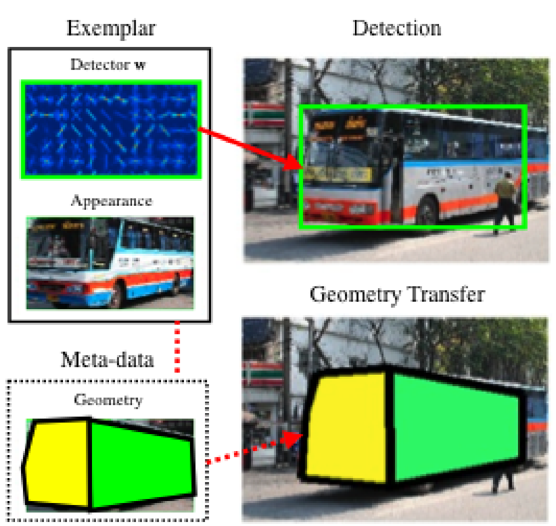


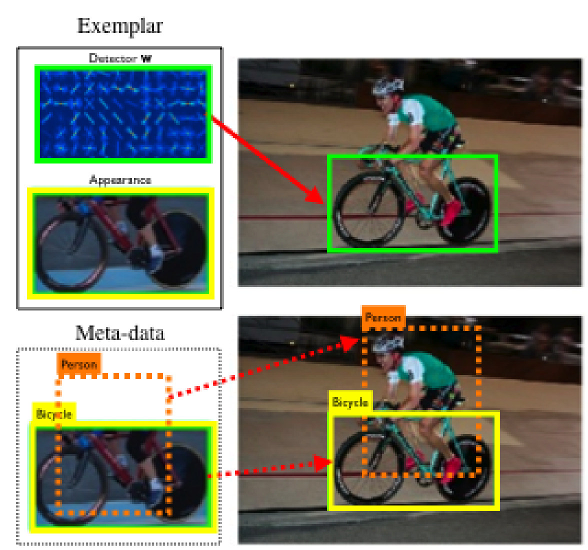


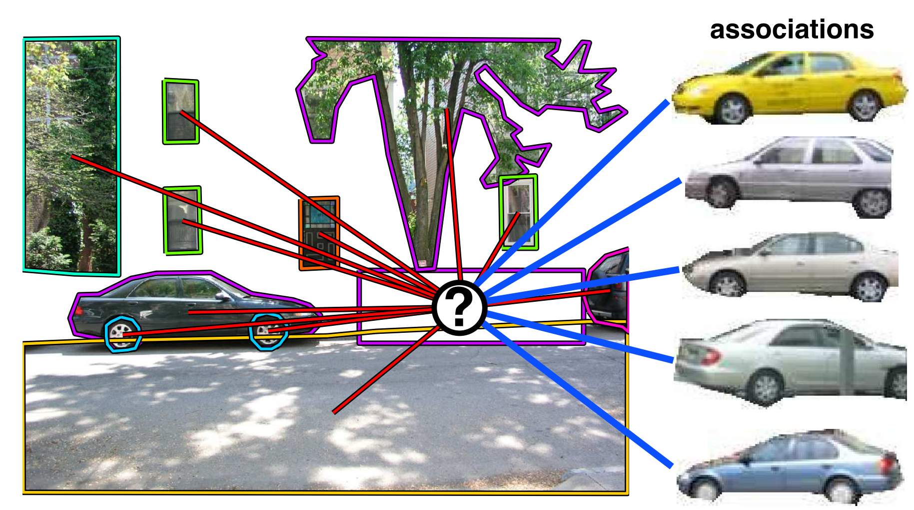


Thesis Conclusions

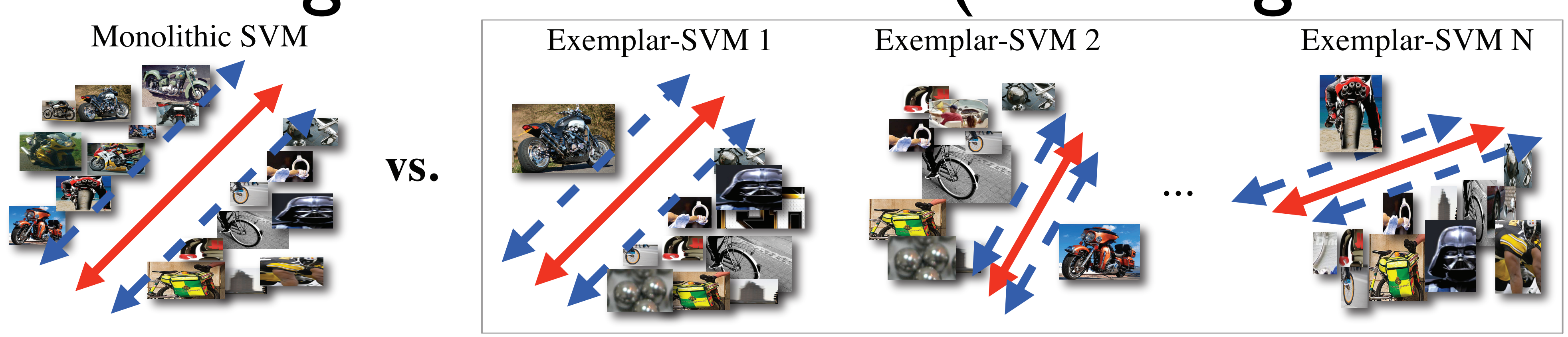
- Visual Memex can be used for recognition, interpretation, and prediction







- Learning visual associations is the **key** to building a Visual Memex (and image matching)



Thank You

*Wordle from dissertation