# Exemplar-SVMs:
## Visual Object Detection, Label Transfer and Image Retrieval


Tomasz Malisiewicz
(Massachusetts Institute of Technology)

Joint work with:
Abhinav Shrivastava, Abhinav Gupta, and Alexei (Alyosha) Efros
(Carnegie Mellon University)

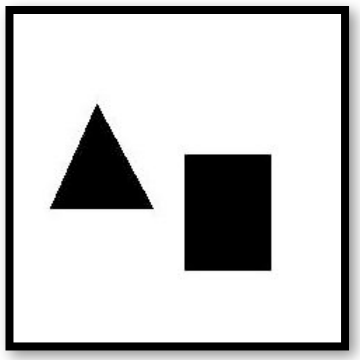
# Talk Overview

- Visual Object Detection
  - Exemplar-SVM Learning
  - Understanding Exemplar-SVMs
- Experimental Results
  - PASCAL VOC Object Detection
  - Label Transfer
  - Cross-domain Image Retrieval
- Concluding remarks and take-home lessons

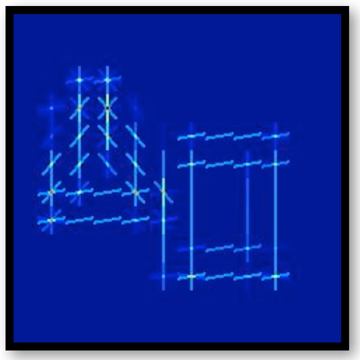
# Object Detectors

# Object Detectors

Dalal et al 2005



Image

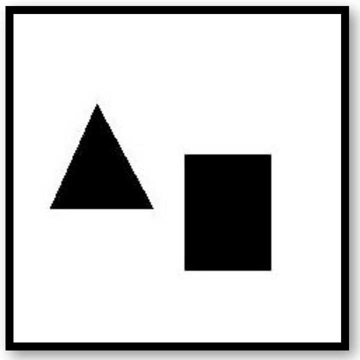


HOG

- Histogram of Oriented Gradients features computed across a multiscale pyramid

# Object Detectors

Dalal et al 2005



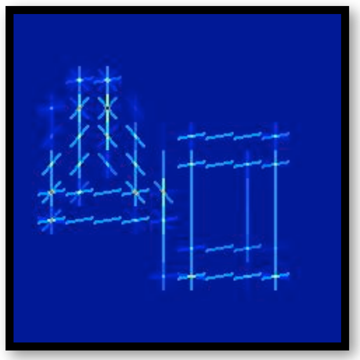
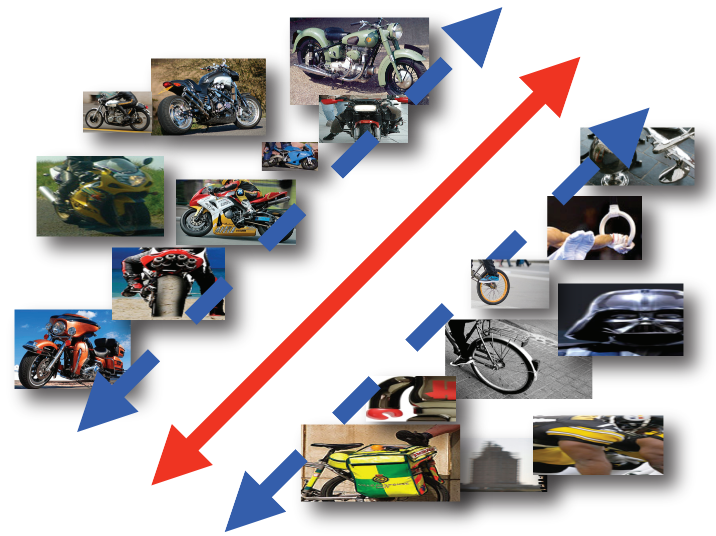


Image HOG



- Histogram of Oriented Gradients features computed across a multiscale pyramid
- Linear SVMs for learning

# Object Detectors

Dalal et al 2005

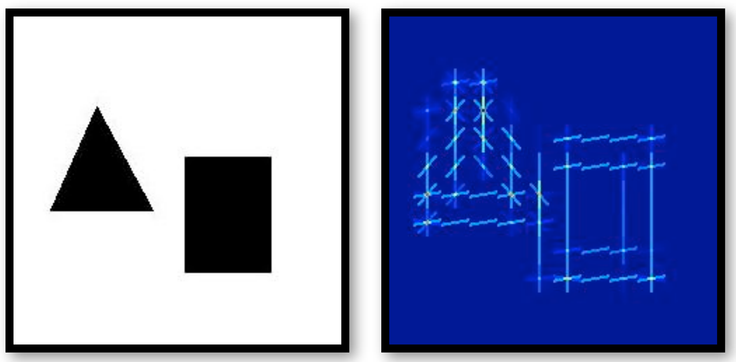
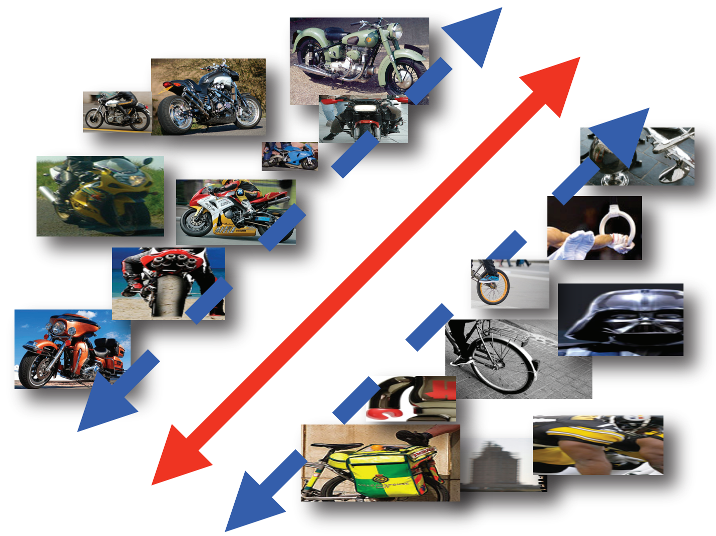


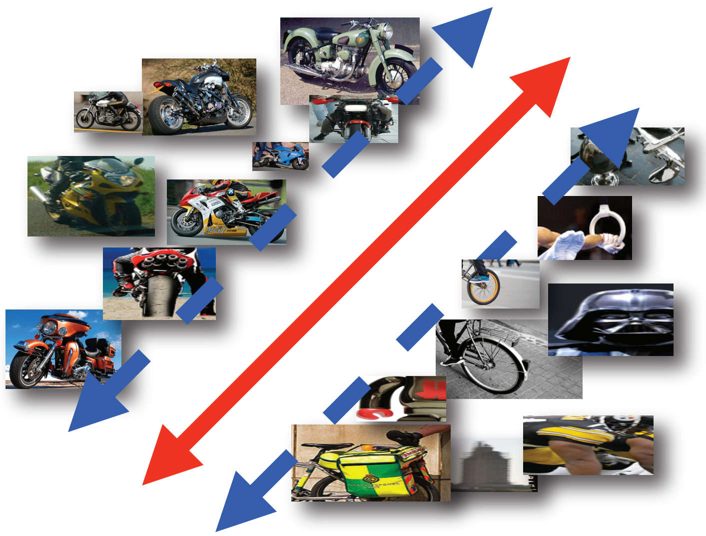
Image HOG



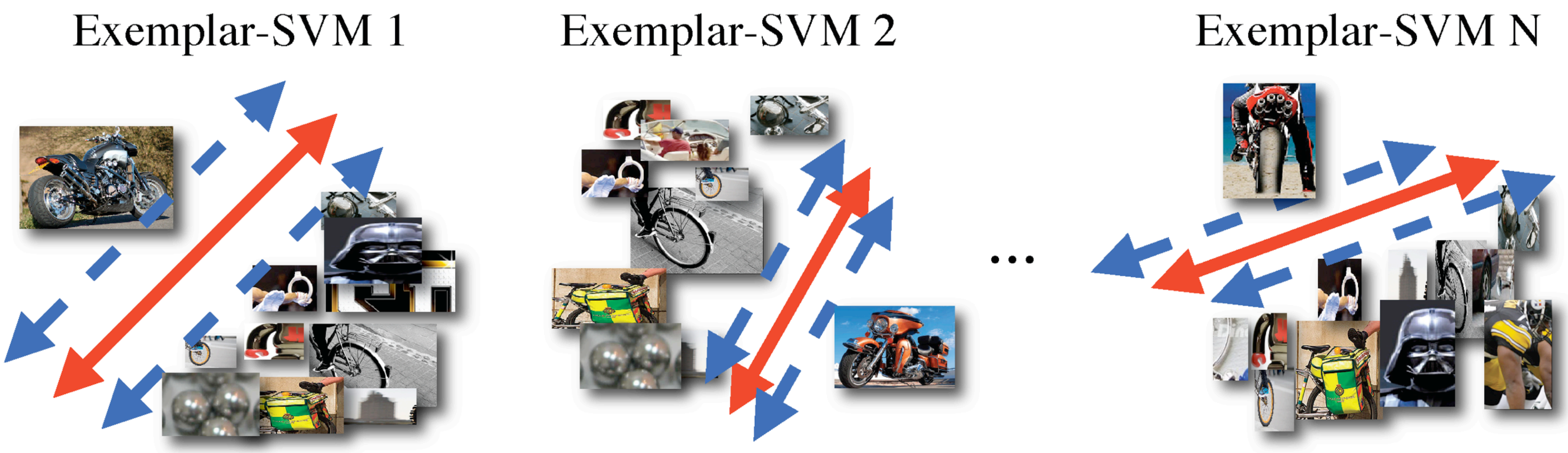
Large
Annotated
Dataset

- Histogram of Oriented Gradients features computed across a multiscale pyramid
- Linear SVMs for learning
- A large dataset such as PASCAL VOC (Everingham et al 2010)

# Object Detectors

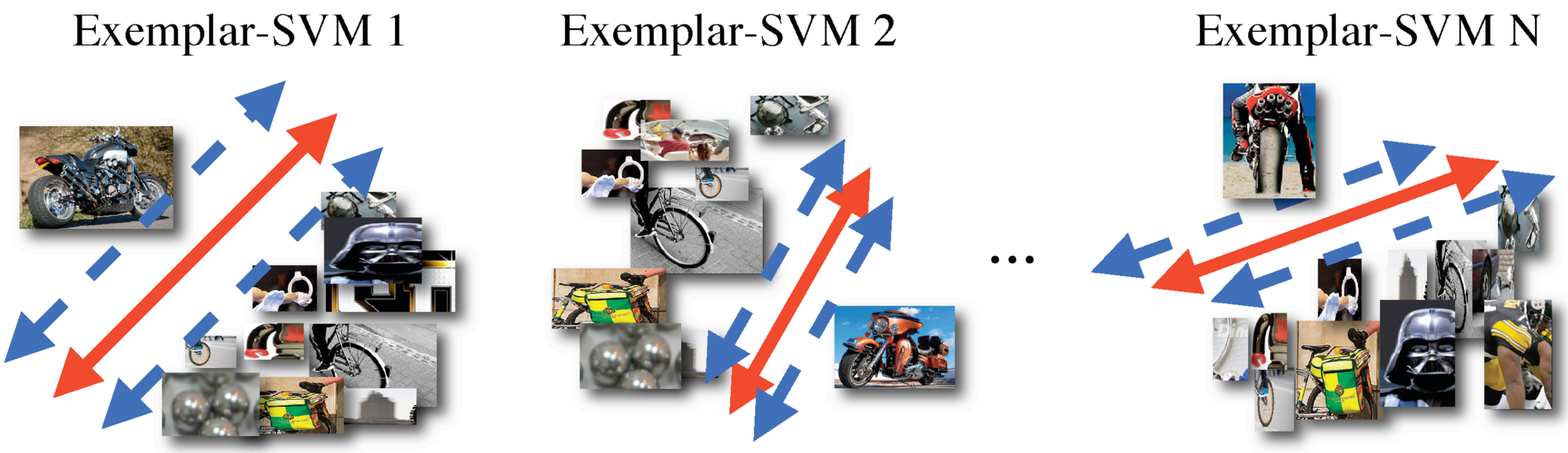


# Exemplar-SVMs




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

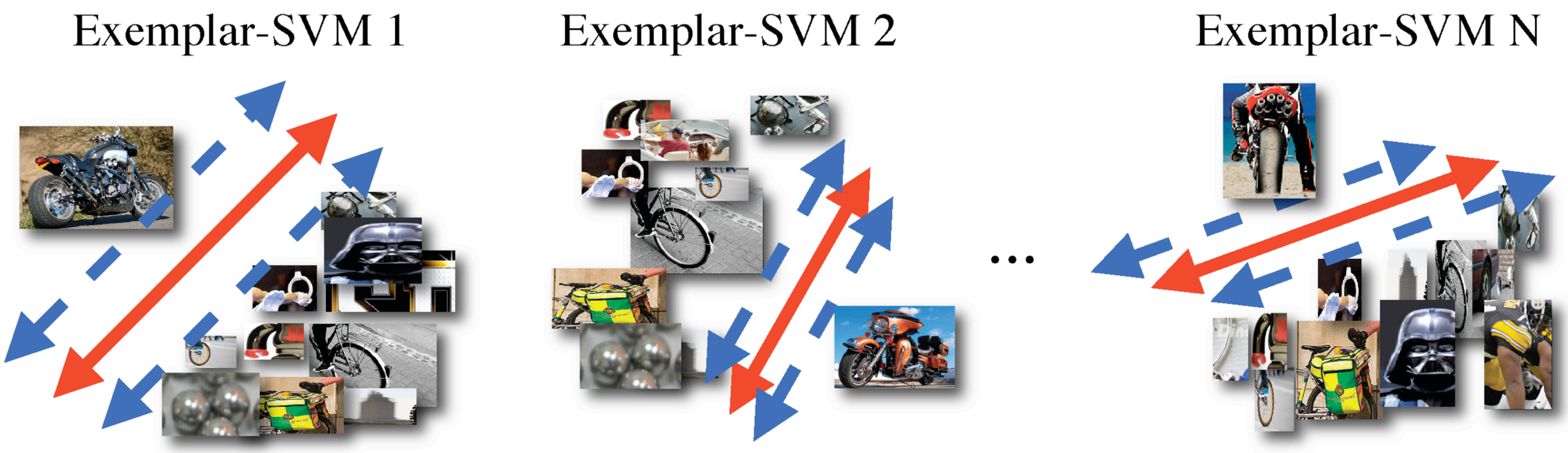
# Exemplar-SVMs



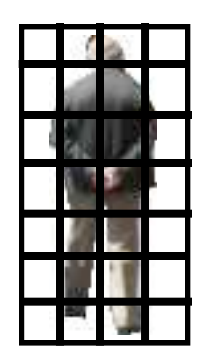

- Best of both worlds:
  - Effectiveness of discriminatively-trained object detectors
  - Explicit correspondence of Nearest Neighbor approaches

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

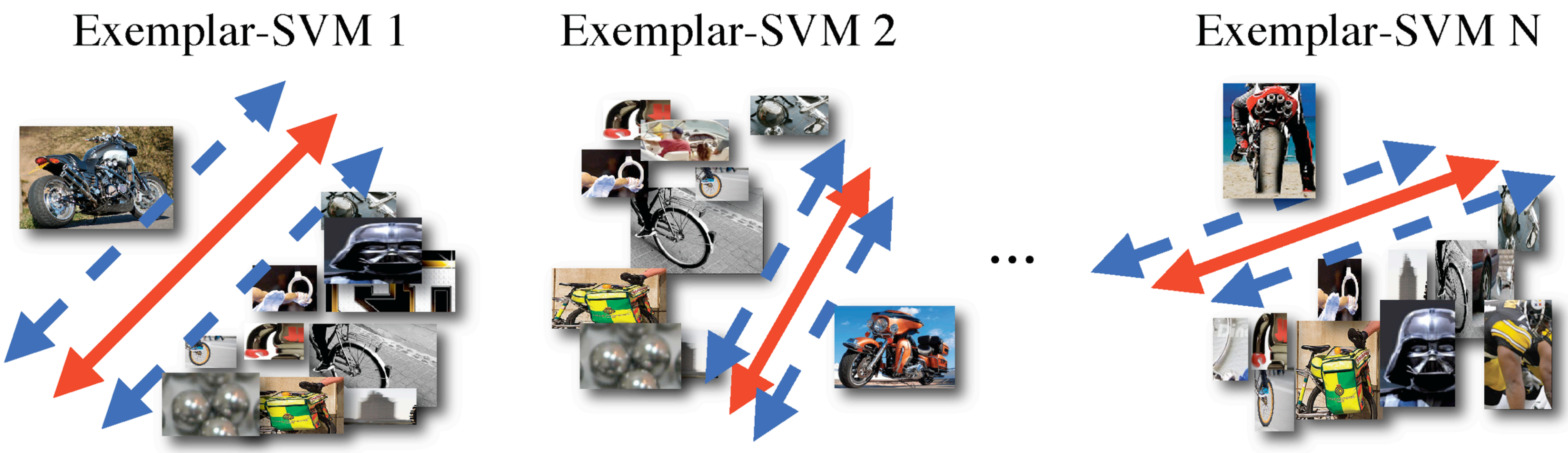
# Exemplar-SVMs




- Because each Exemplar-SVM is defined by a **single** positive instance, we can use different features for each exemplar

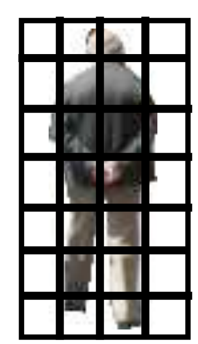



# Exemplar-SVMs




- Because each Exemplar-SVM is defined by a **single** positive instance, we can use different features for each exemplar

- Apply each Exemplar-SVM to test image in a sliding-window fashion

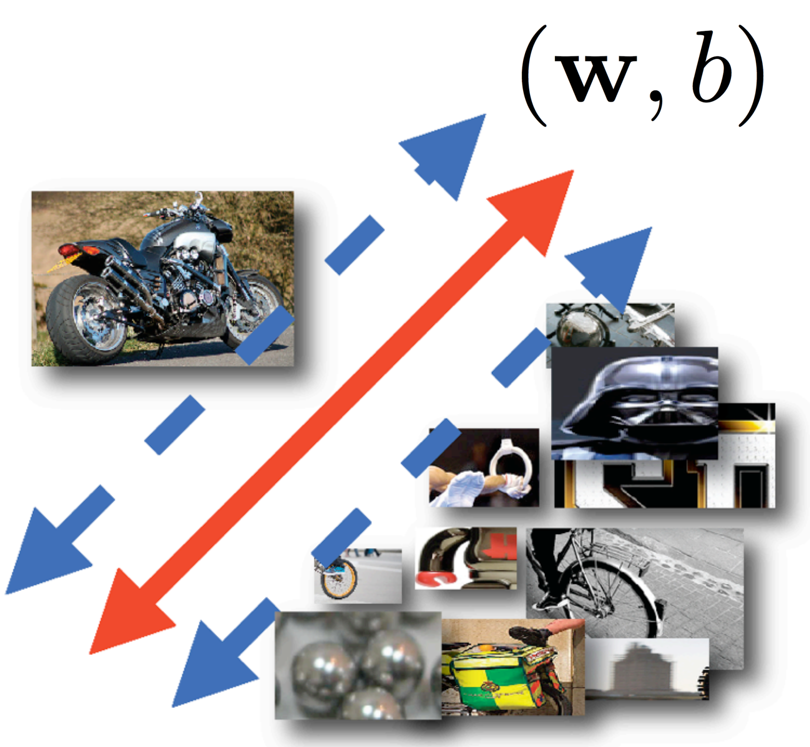



# Exemplar-SVMs

## Exemplar E's Objective Function:

$$\Omega_E(\mathbf{w}, b) = ||\mathbf{w}||^2 + C_1 h(\mathbf{w}^T \mathbf{x}_E + b) + C_2 \sum_{\mathbf{x} \in \mathcal{N}_E} h(-\mathbf{w}^T \mathbf{x} - b)$$

h(x) = max(1-x,0) “hinge-loss”

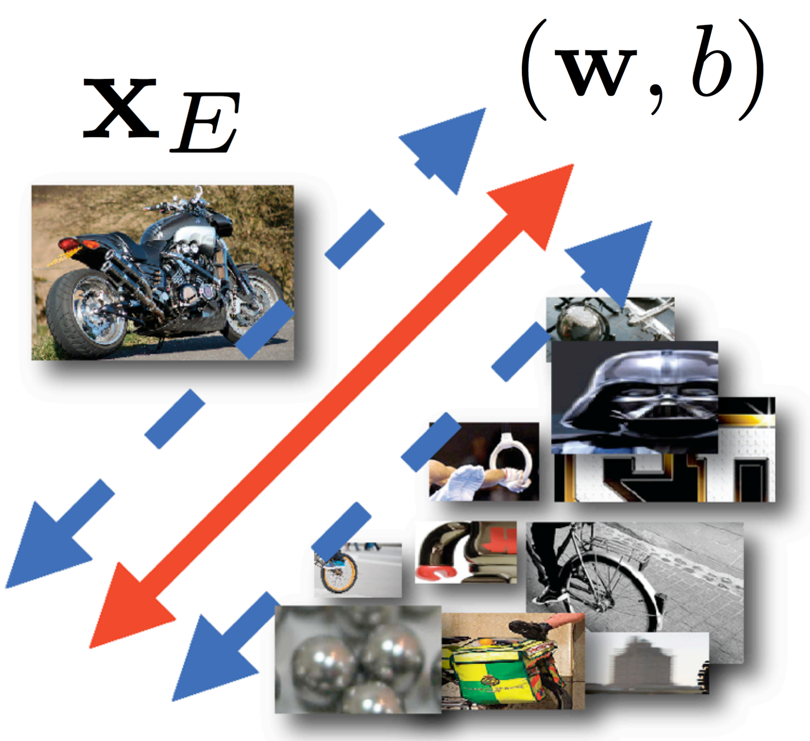


# Exemplar-SVMs

Exemplar E's Objective Function:

$$\Omega_E(\mathbf{w}, b) = ||\mathbf{w}||^2 + C_1 h(\mathbf{w}^T \mathbf{x}_E + b) + C_2 \sum_{\mathbf{x} \in \mathcal{N}_E} h(-\mathbf{w}^T \mathbf{x} - b)$$

h(x) = max(1-x,0) “hinge-loss”



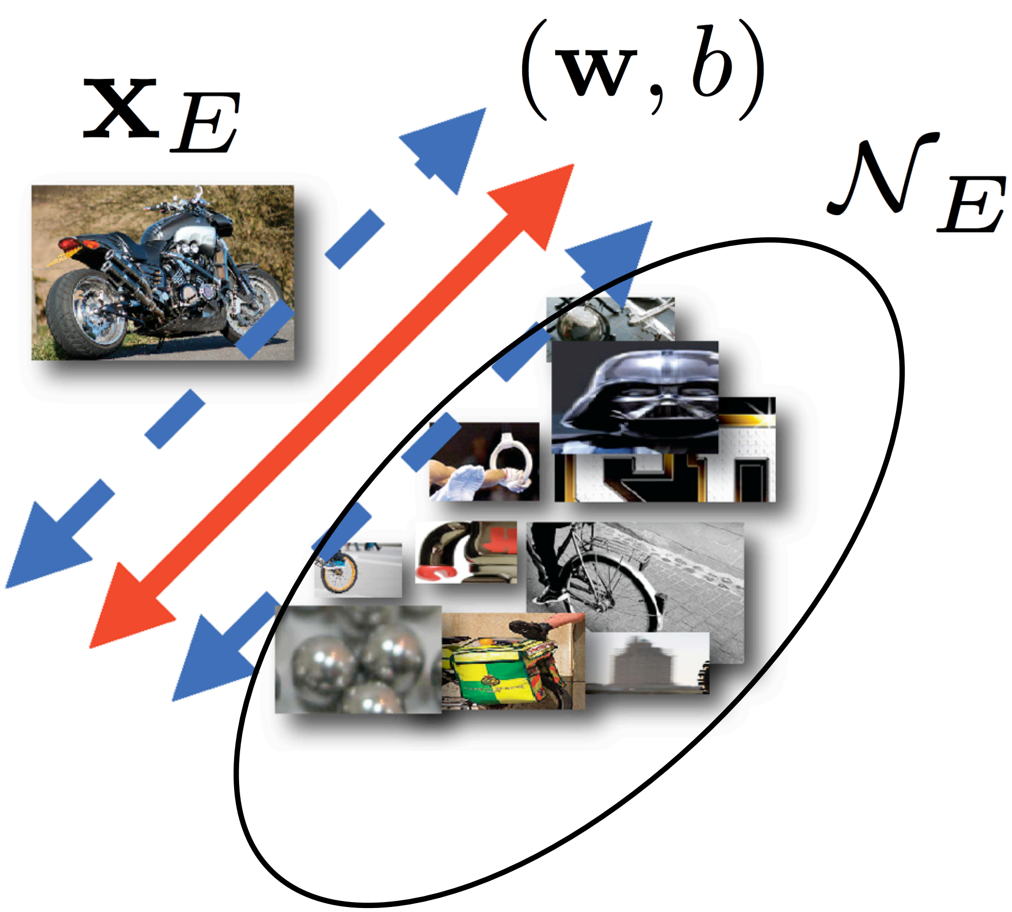

$\mathbf{x}_E$ Exemplar represented by ~100 HOG Cells (~3,000D features)

# Exemplar-SVMs

Exemplar E's Objective Function:

$$\Omega_E(\mathbf{w}, b) = ||\mathbf{w}||^2 + C_1 h(\mathbf{w}^T \mathbf{x}_E + b) + C_2 \sum_{\mathbf{x} \in \mathcal{N}_E} h(-\mathbf{w}^T \mathbf{x} - b)$$

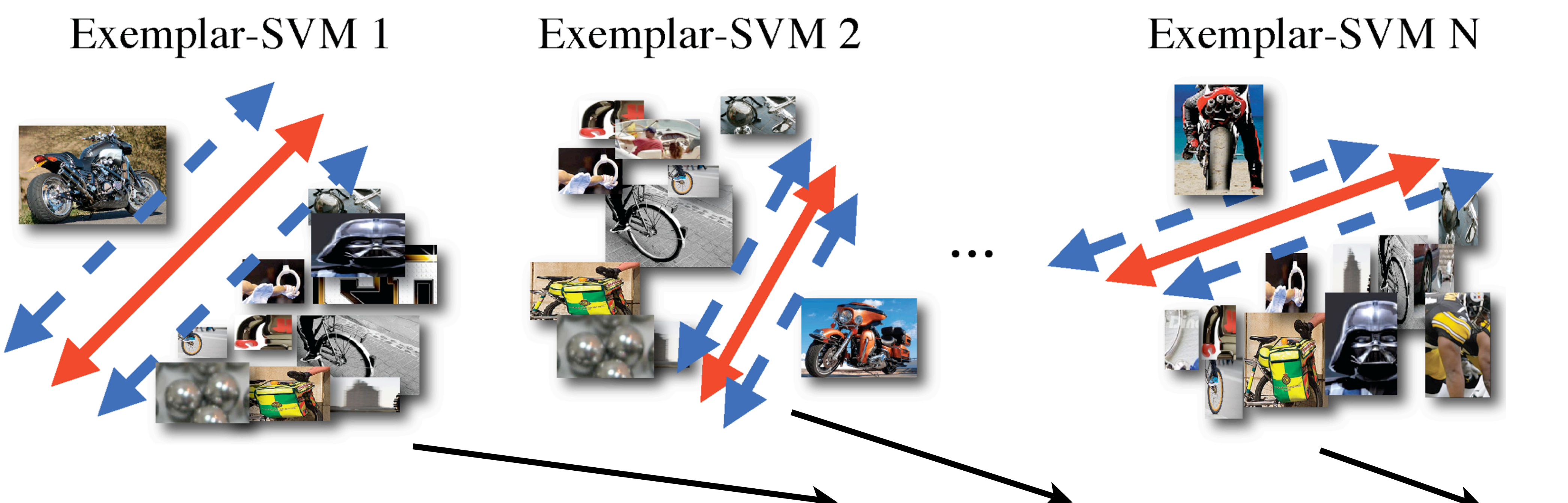
h(x) = max(1-x,0) "hinge-loss"




$\mathbf{x}_E$ Exemplar represented by ~100 HOG Cells (~3,000D features)

$\mathcal{N}_E$ Windows from images not containing any in-class instances (2,000 images x 10,000 windows per image = 20M negatives )

# Embarrassingly Parallel




- Each exemplar performs its own hard negative mining
- Solve many convex learning problems
- Parallel training on cluster

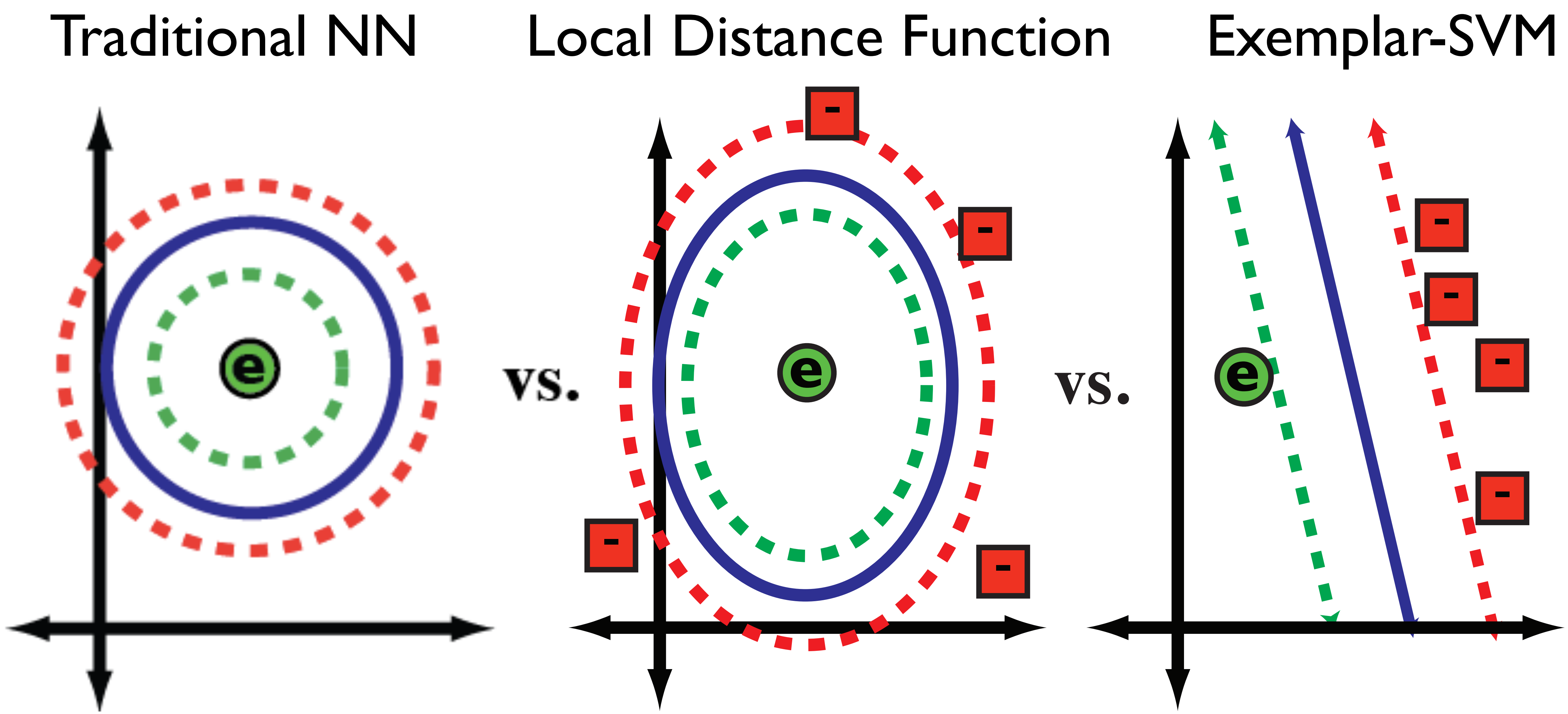
# Visualizing Exemplar-SVMs

# Visualizing Exemplar-SVMs

Exemplar-SVMs

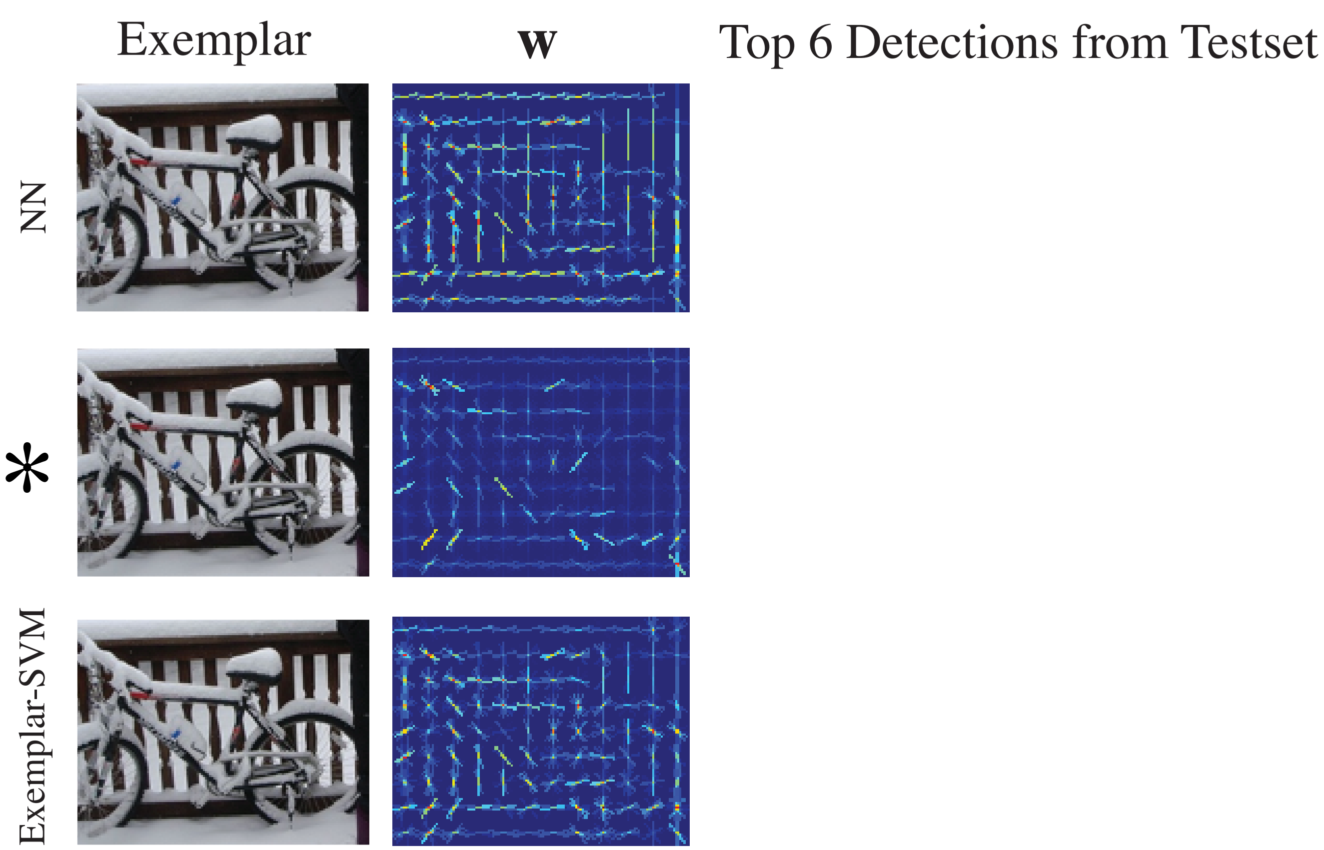
Top Detections in Test Set

# Understanding Exemplar-SVMs



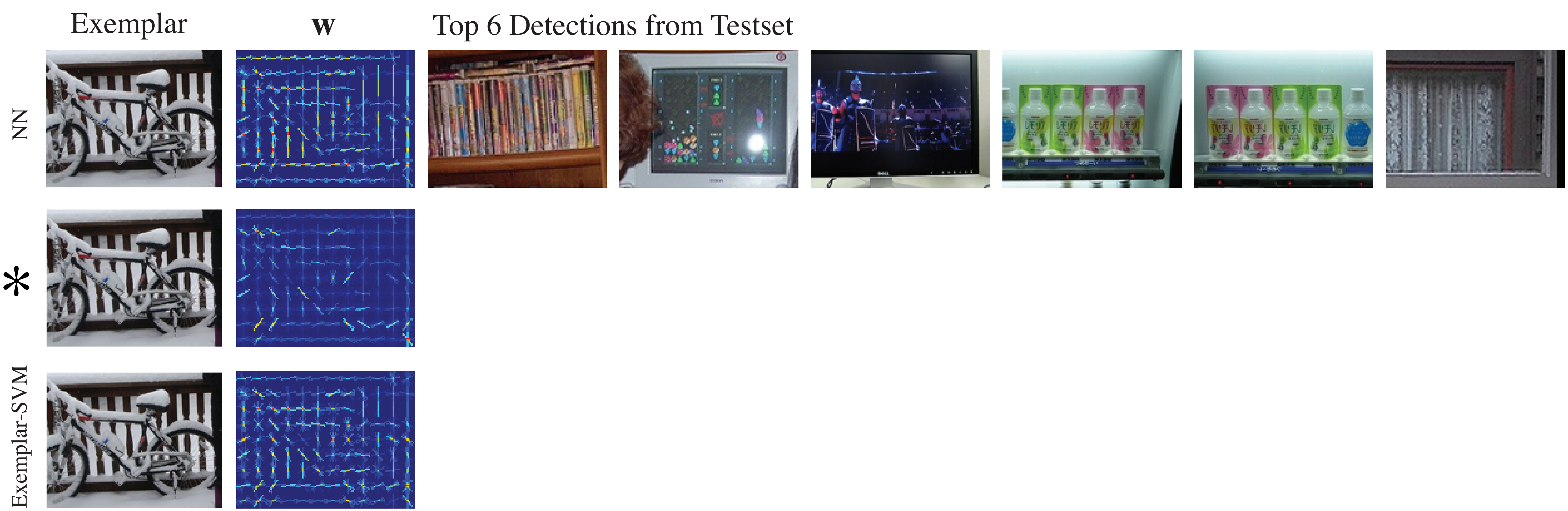

Frome et al, NIPS 2006

# Understanding Exemplar-SVMs




*Learned Distance Function

# Understanding Exemplar-SVMs




*Learned Distance Function

# Understanding Exemplar-SVMs


*Learned Distance Function

# Understanding Exemplar-SVMs


*Learned Distance Function

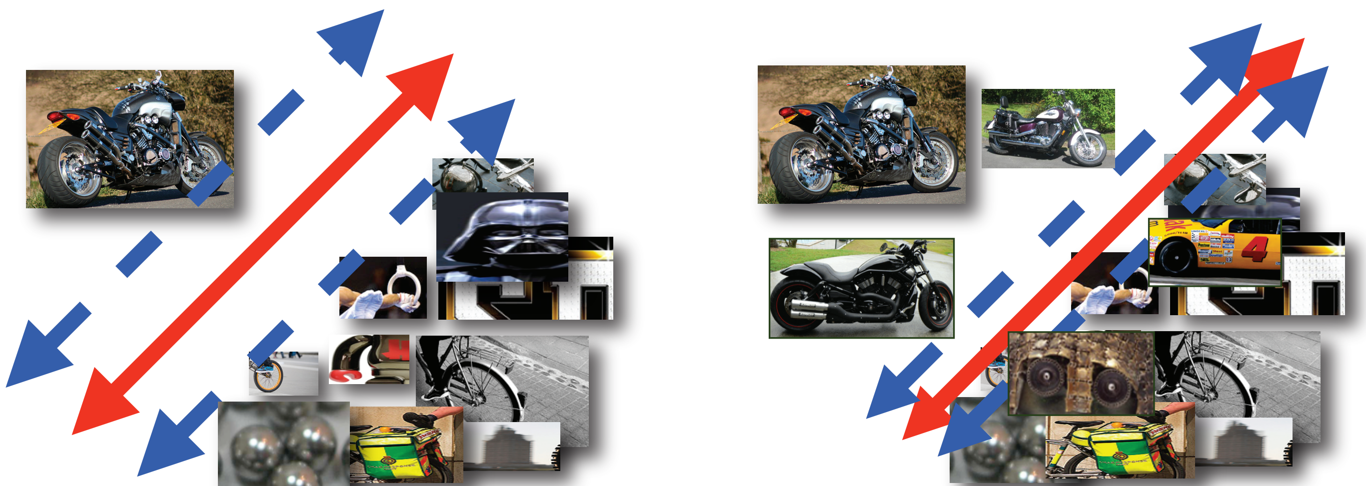
# Ensemble of Exemplar-SVMs

# Ensemble of Exemplar-SVMs

Platt Calibration
(Platt 1999)

Before Calibration

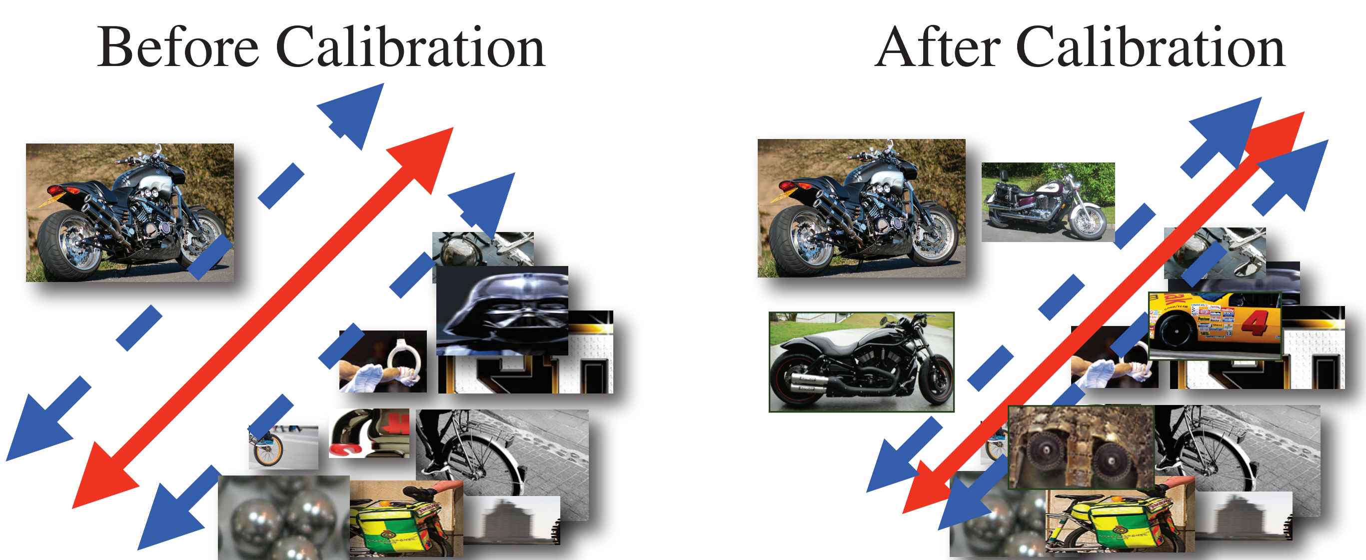
After Calibration



Exemplars **Compete**

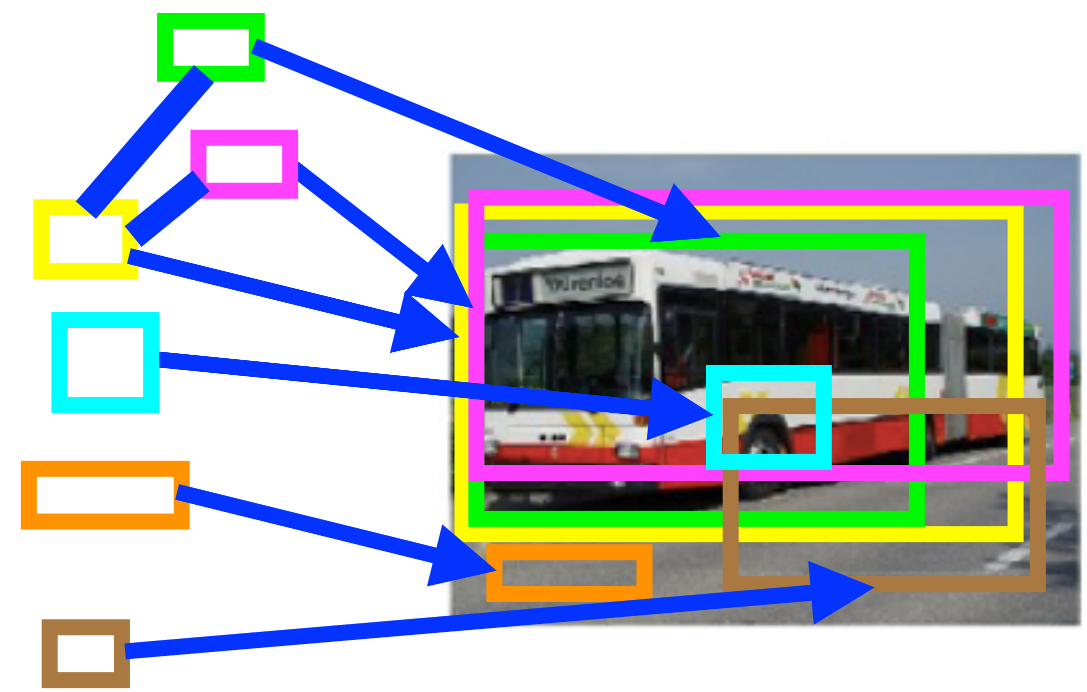
# Ensemble of Exemplar-SVMs

Platt Calibration
(Platt 1999)




Exemplars **Compete**

Learning Exemplar
Co-occurrence Matrix



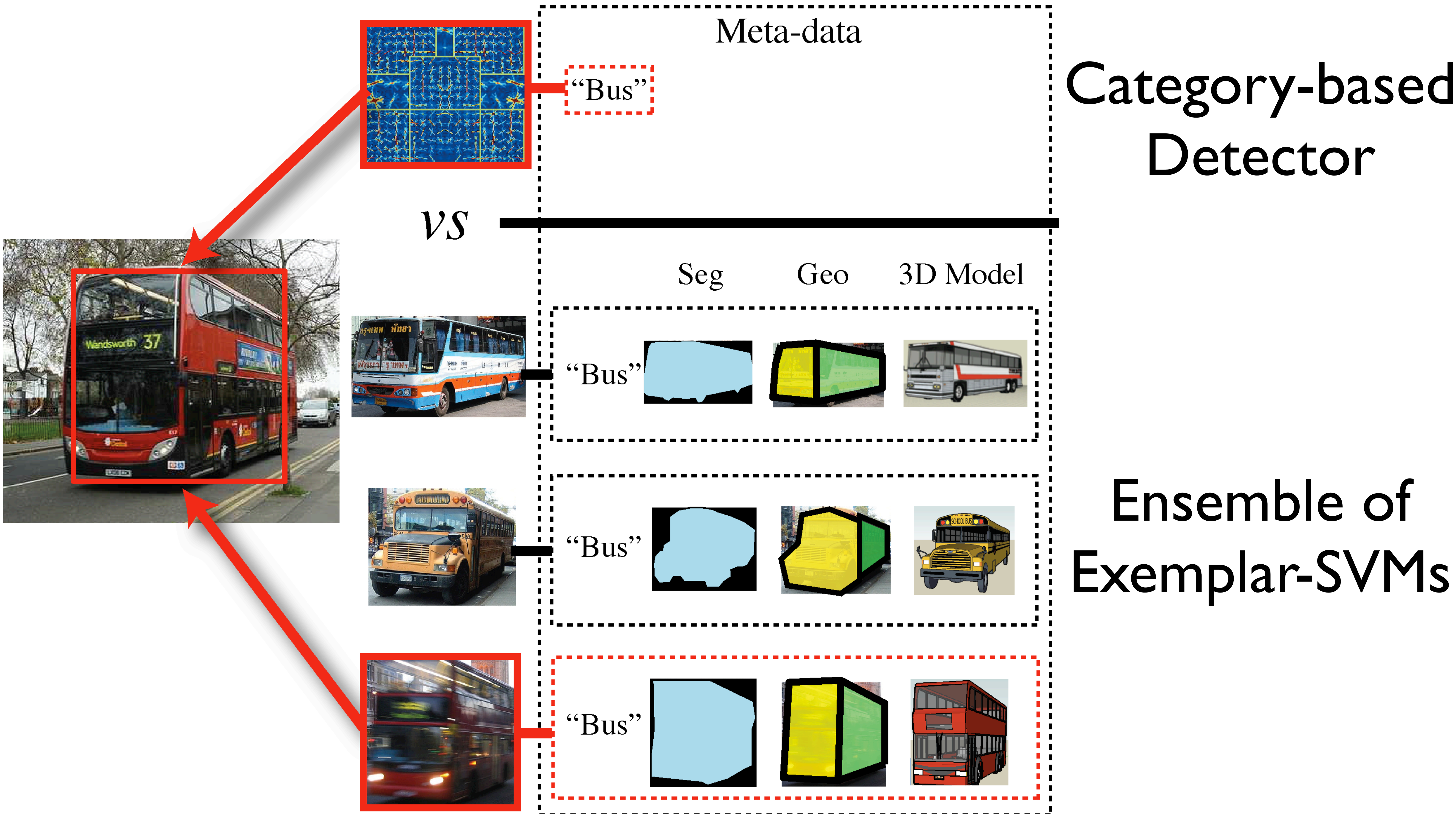
Exemplars are **Combined**

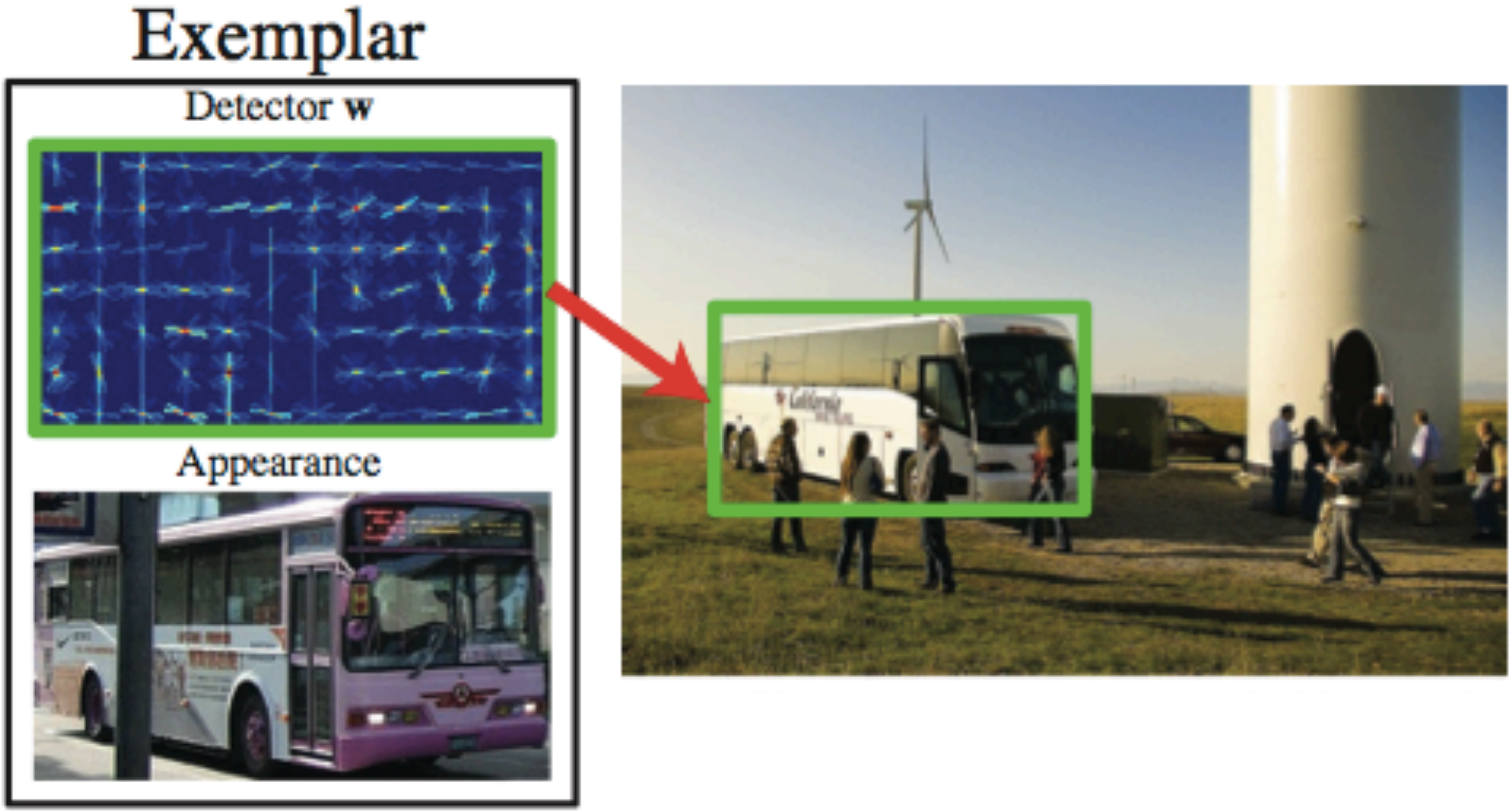
# Object Category Detection

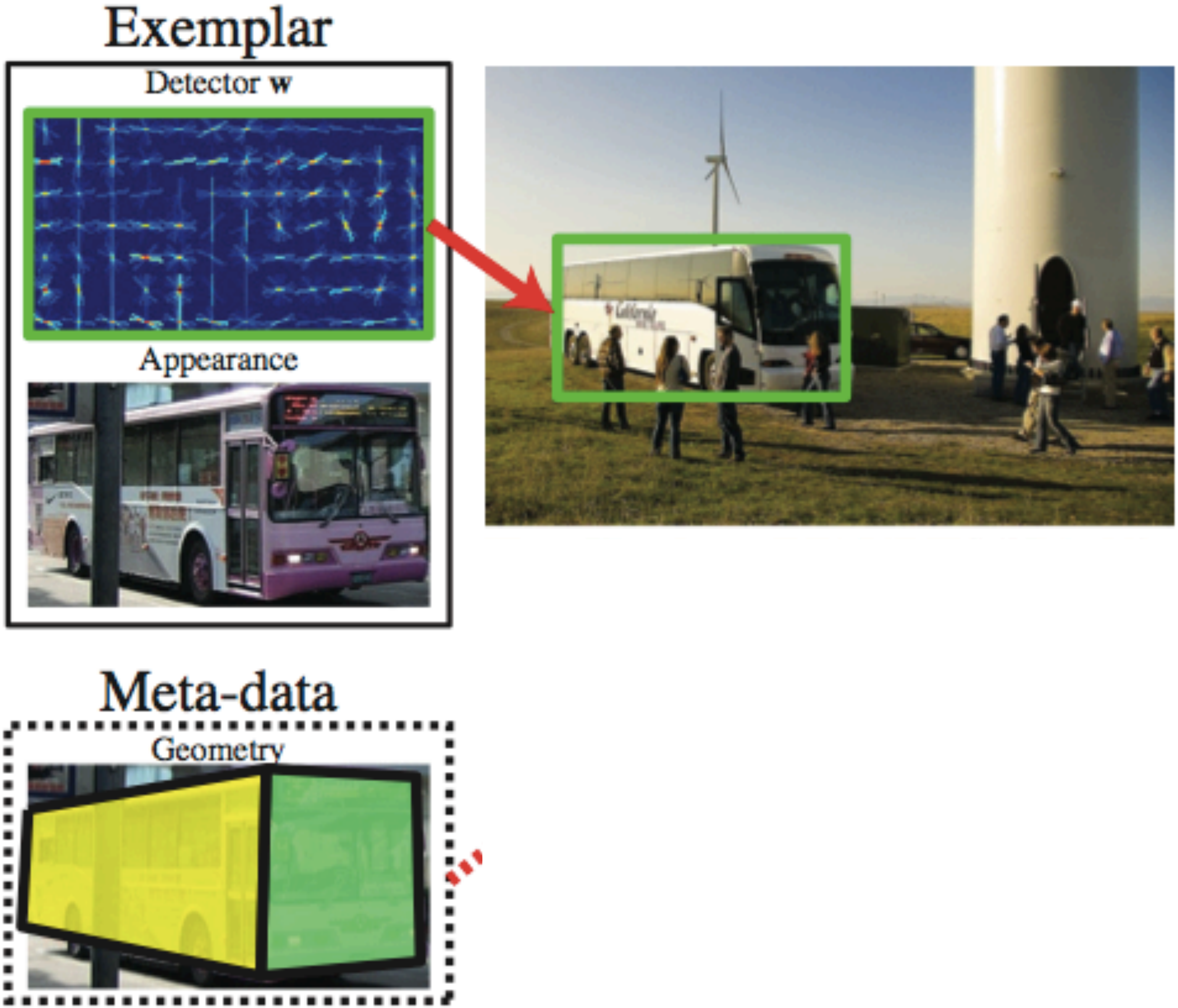
mAP averaged across 20 object categories on the PASCAL VOC 2007 object detection task

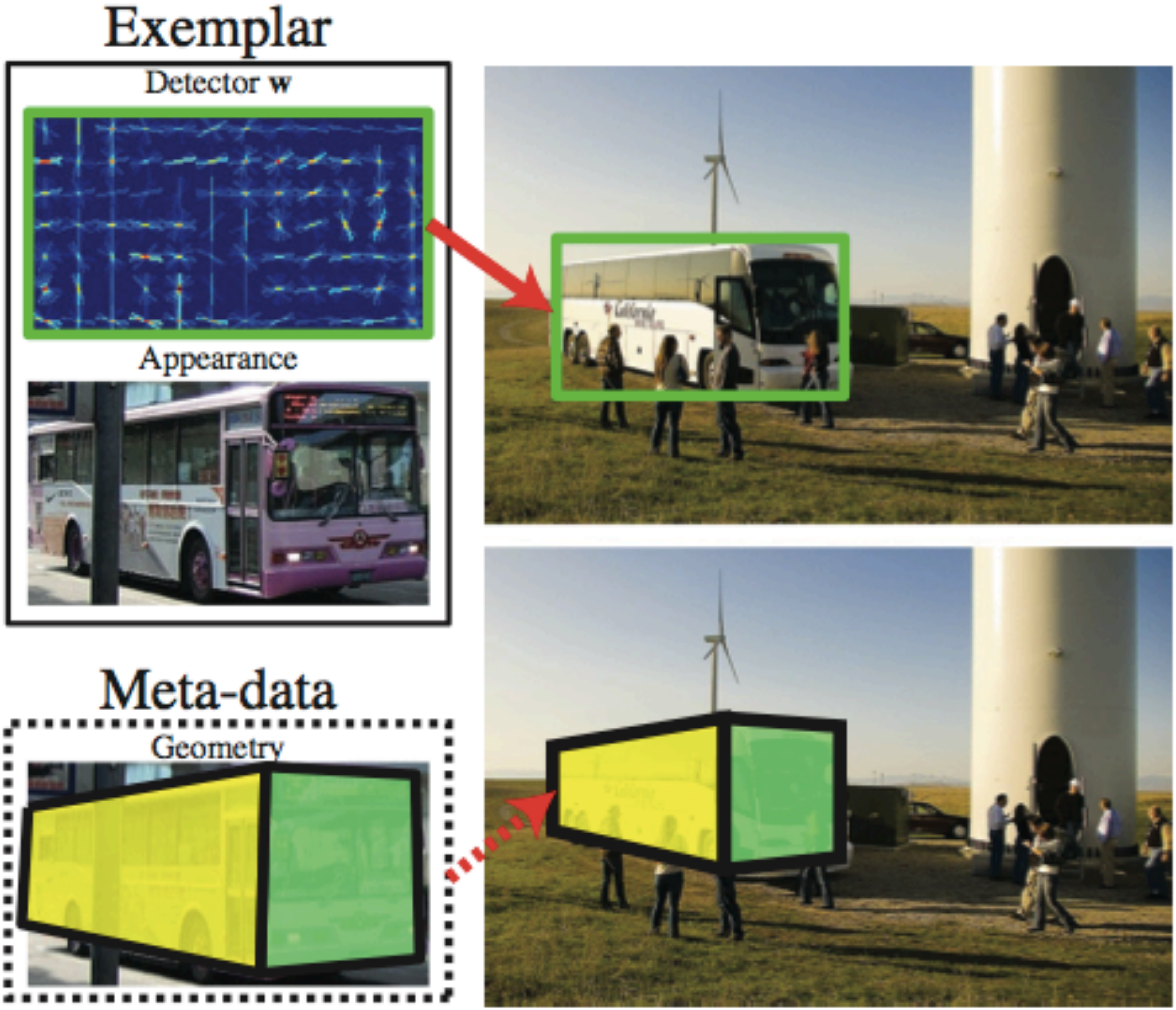
| | |
|---|---|
| Traditional NN + Calibration | 0.110 |
| Local Distance Function + Calibration | 0.157 |
| **Exemplar-SVMs + Calibration** | **0.198** |
| **Exemplar-SVMs + Co-occurrence** | **0.227** |
| One SVM per category (Dalal and Triggs 2005) | 0.097 |
| Deformable Part Model (Felzenszwalb et al 2010) | 0.266 |

# Beyond Detection: Label Transfer

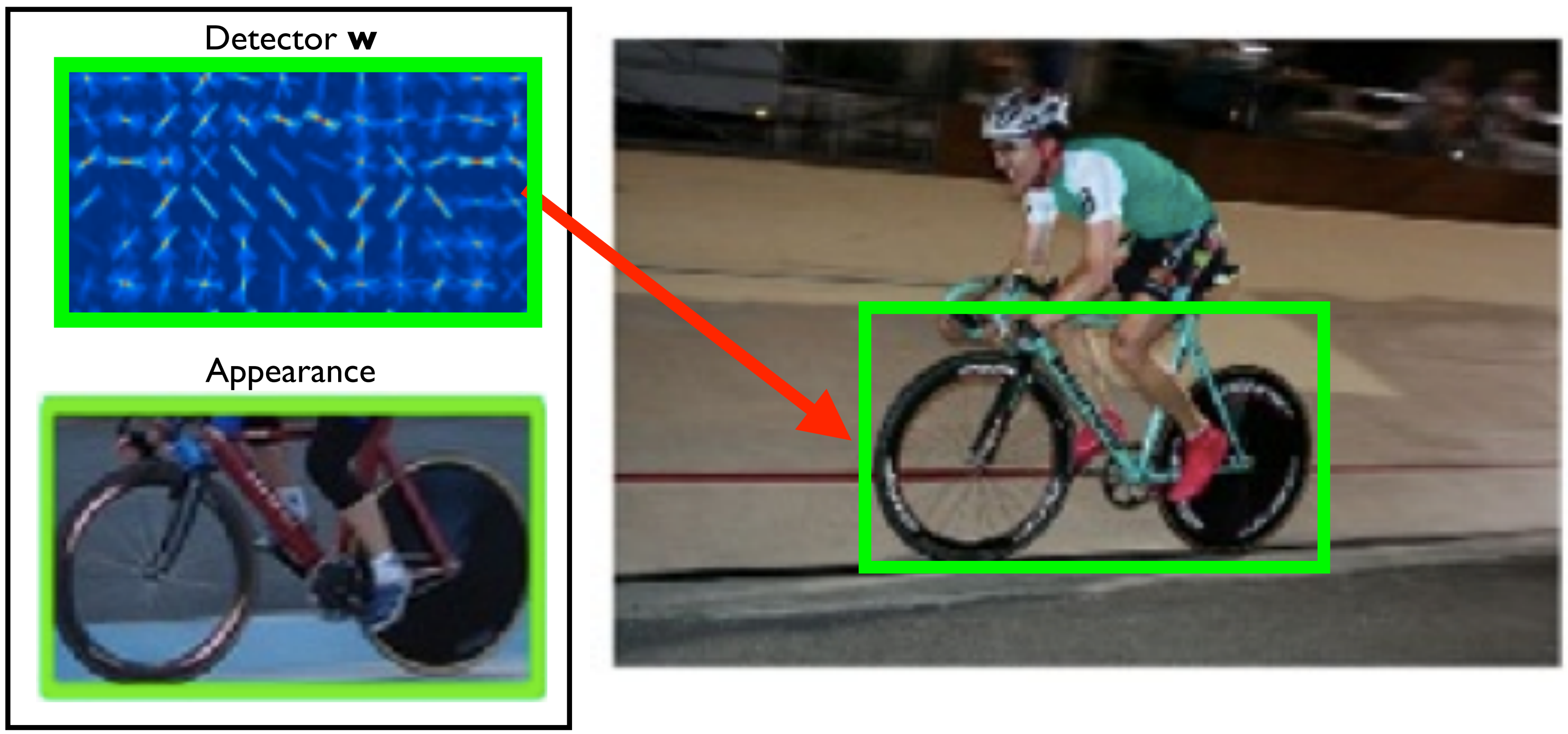


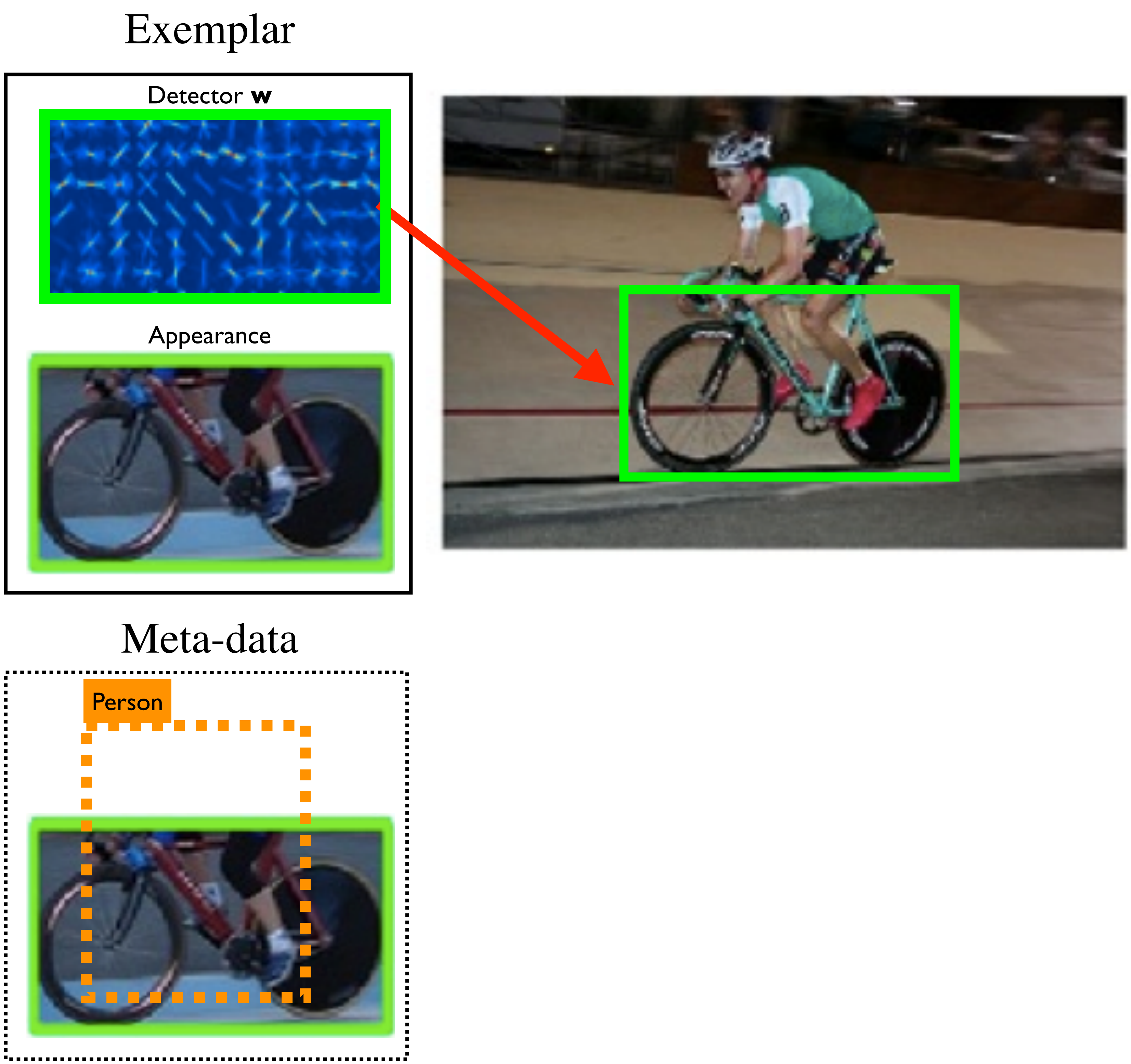

Exemplar
Detector w
Appearance

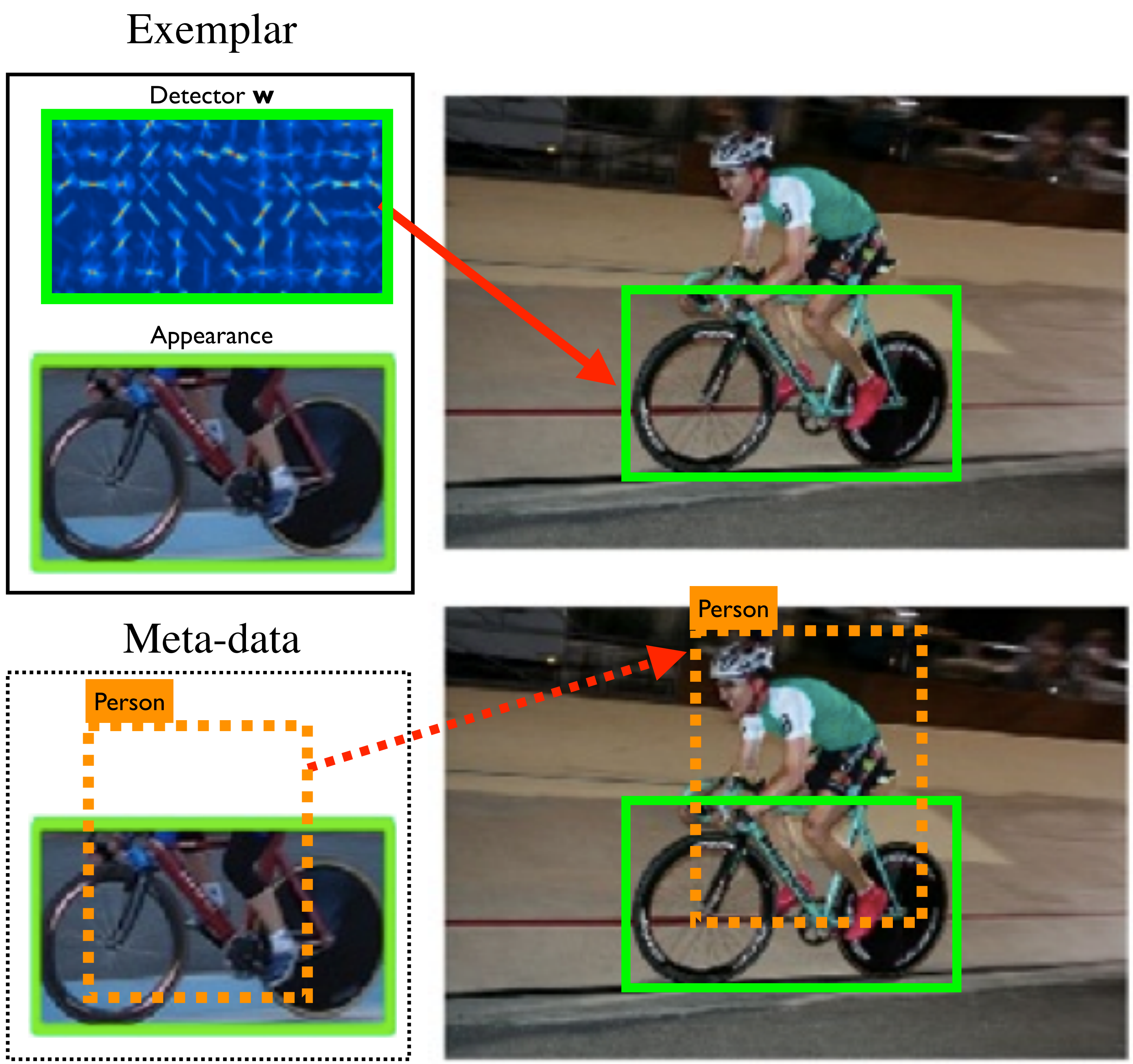

Exemplar
Detector w
Appearance
Meta-data
Geometry


Exemplar
Detector w
Appearance
Meta-data
Geometry

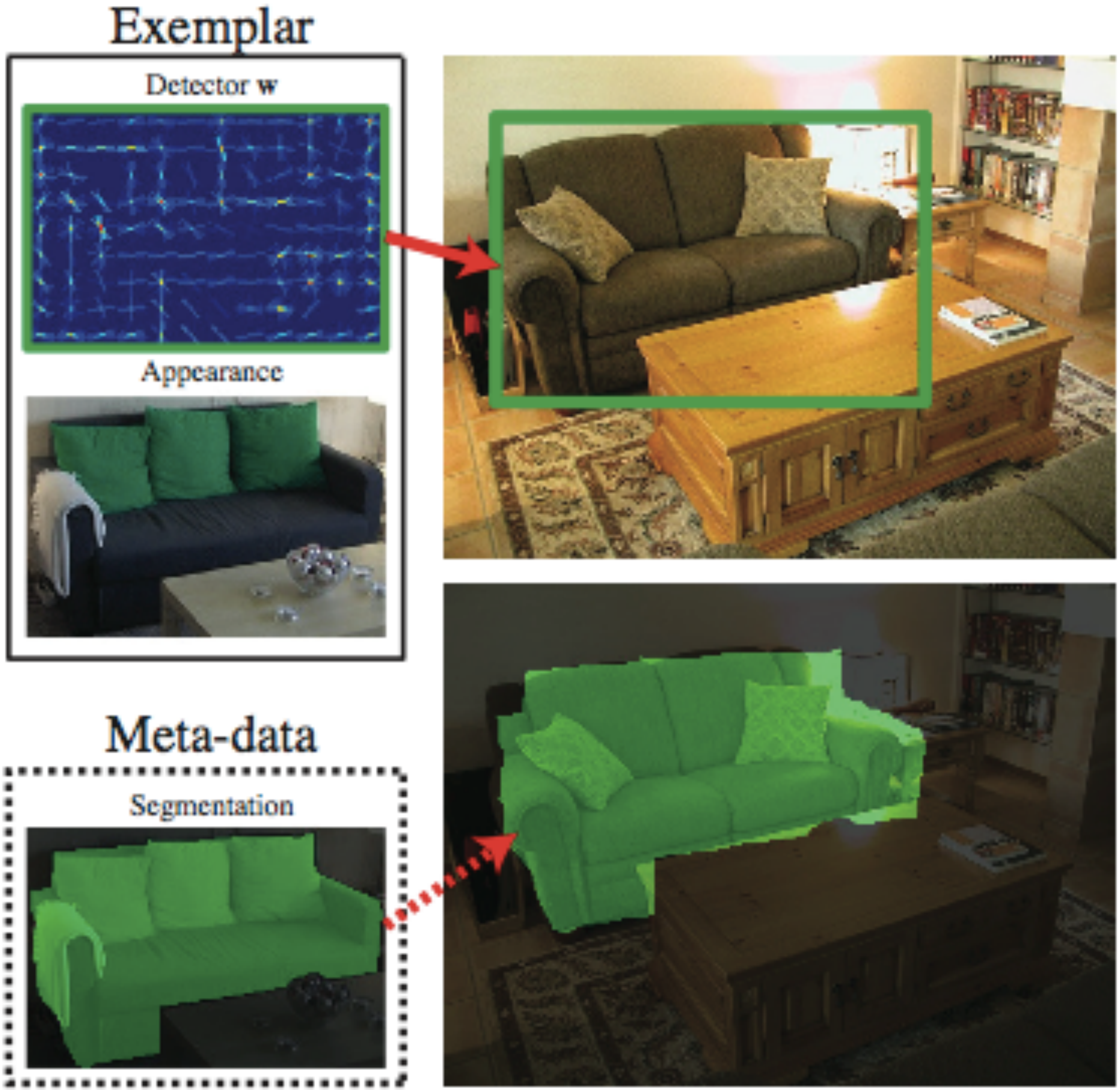
Exemplar




Exemplar
Detector w
Appearance
Meta-data
Person



Exemplar
Detector w
Appearance
Meta-data
Person
Person


Exemplar
Detector w
Appearance
Meta-data
Segmentation


Exemplar
Detector w
Appearance
Meta-data
3D Model

# Talk Overview

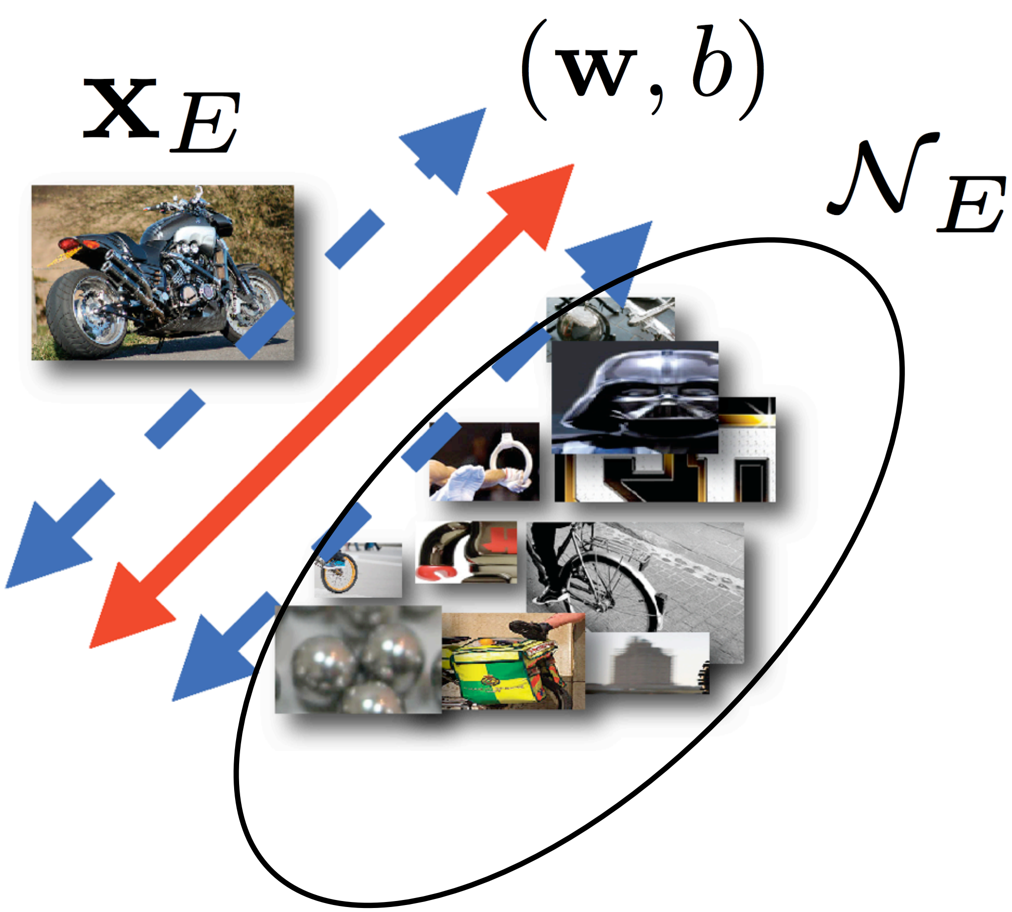
- Visual Object Detection
  - Exemplar-SVM Learning
  - Understanding Exemplar-SVMs
- Experimental Results
  - PASCAL VOC Object Detection
  - Label Transfer
  - **Cross-domain Image Retrieval**
- Concluding remarks and take-home lessons

# Exemplar-SVMs

Exemplar E's Objective Function:

$$\Omega_E(\mathbf{w}, b) = ||\mathbf{w}||^2 + C_1 h(\mathbf{w}^T \mathbf{x}_E + b) + C_2 \sum_{\mathbf{x} \in \mathcal{N}_E} h(-\mathbf{w}^T \mathbf{x} - b)$$

h(x) = max(1-x,0) "hinge-loss"




$\mathbf{x}_E$ Exemplar represented by ~100 HOG Cells (~3,000D features)

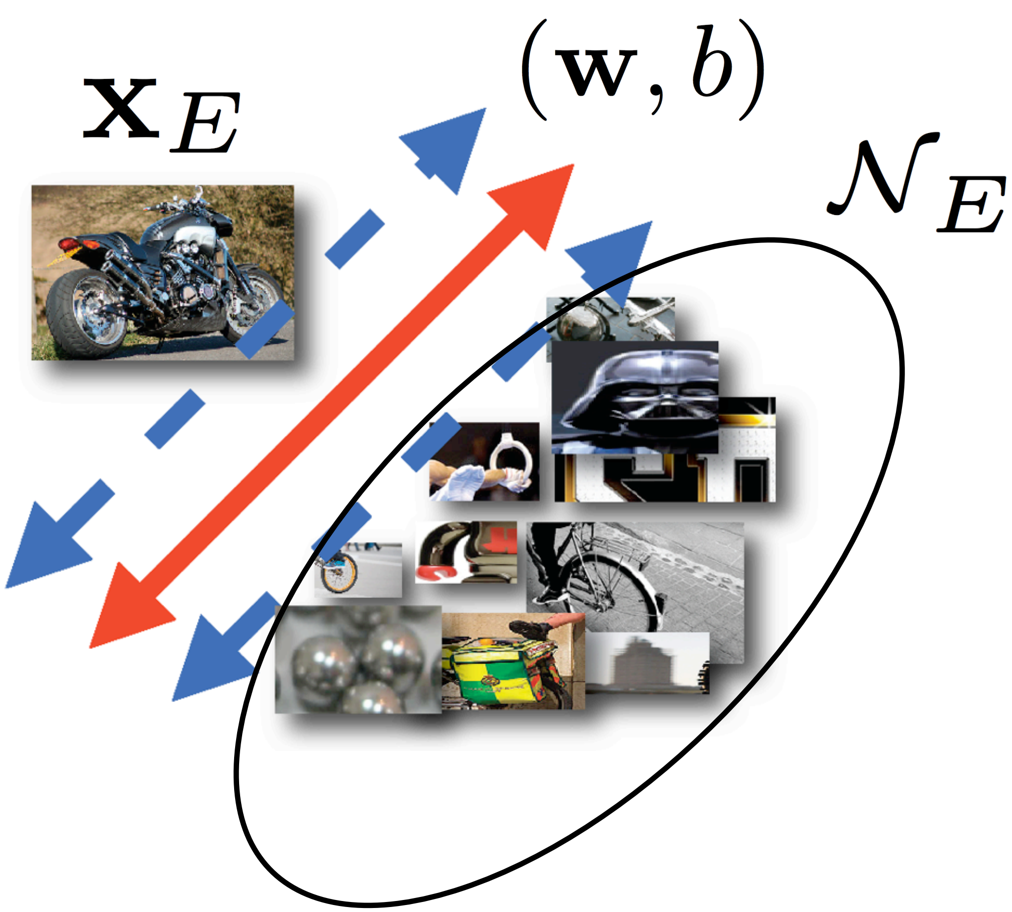
$\mathcal{N}_E$ Windows from images not containing any in-class instances (2,000 images x 10,000 windows per image = 20M negatives )

# Exemplar-SVMs

Exemplar E's Objective Function:

$$\Omega_E(\mathbf{w}, b) = ||\mathbf{w}||^2 + C_1 h(\mathbf{w}^T \mathbf{x}_E + b) + C_2 \sum_{\mathbf{x} \in \mathcal{N}_E} h(-\mathbf{w}^T \mathbf{x} - b)$$

h(x) = max(1-x,0) "hinge-loss"




$\mathbf{x}_E$ Exemplar represented by ~100 HOG Cells (~3,000D features)

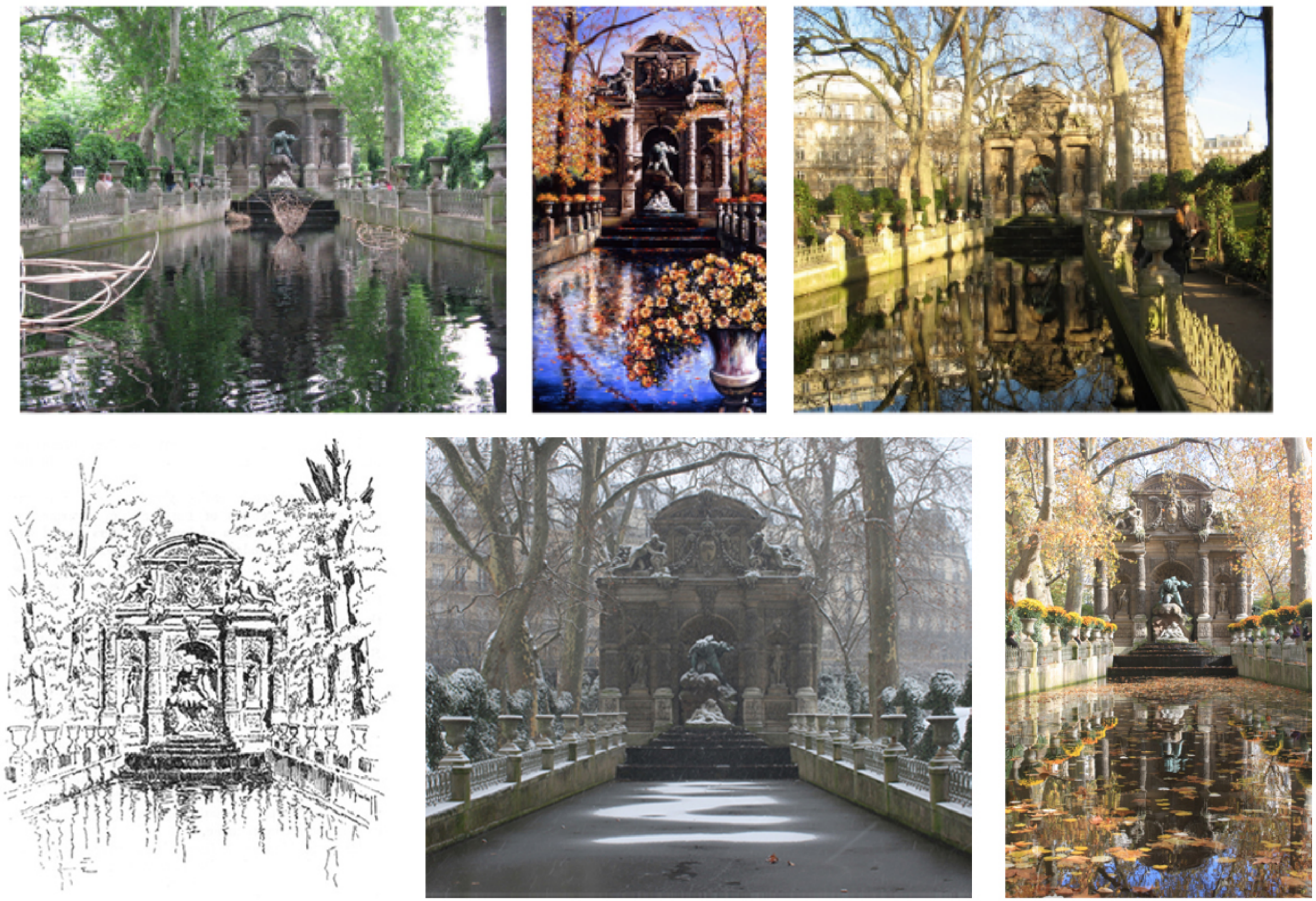
$\mathcal{N}_E$ Windows from images not containing any in-class instances (2,000 images x 10,000 windows per image = 20M negatives )

# Object Category Detection

Exemplar-SVMs* = Exemplar-SVMs with random negatives

| | |
|---|---|
| Traditional NN + Calibration | 0.110 |
| Local Distance Function + Calibration | 0.157 |
| Exemplar-SVMs + Calibration | 0.198 |
| Exemplar-SVMs + Co-occurrence | 0.227 |
| **Exemplar-SVMs* + Calibration** | **0.142** |
| **Exemplar-SVMs* + Co-occurrence** | **0.197** |

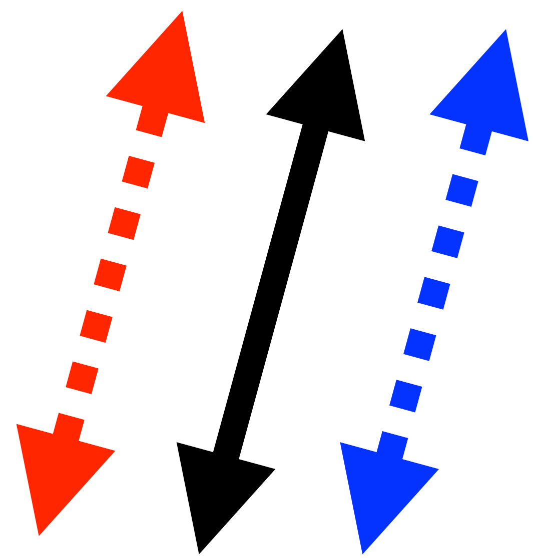
# Cross-domain Image Matching



Abhinav Shrivastava, Tomasz Malisiewicz, Abhinav Gupta, Alexei A. Efros. **Data-driven Visual Similarity for Cross-domain Image Matching.** In SIGGRAPH ASIA, 2011.

# Learn Exemplar-SVM for query image

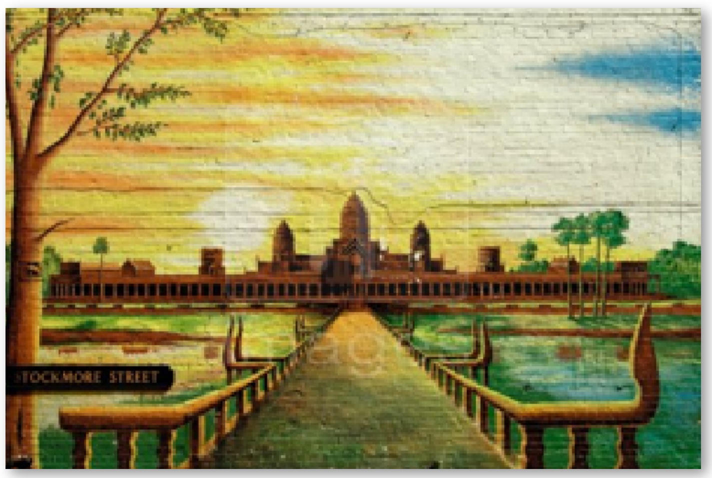
Query Image



Negatives mined from random Flickr images

# Image Retrieval

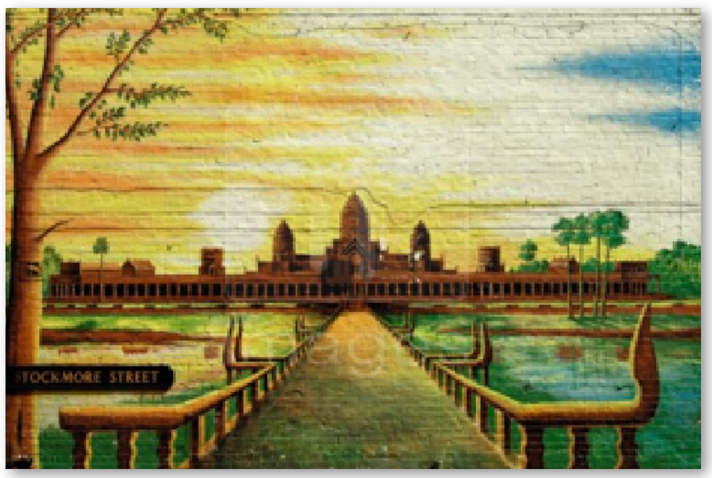
Query Image




Random Flickr Images

# Image Retrieval

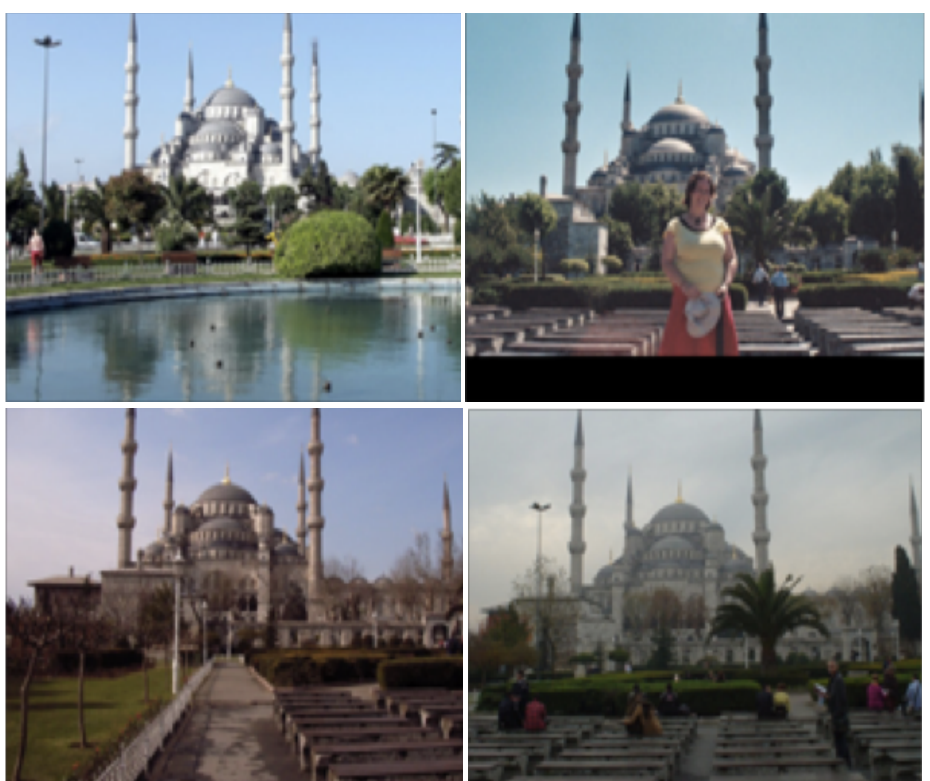
Query Image




Random Flickr Images

# Search using Paintings

**Painting**

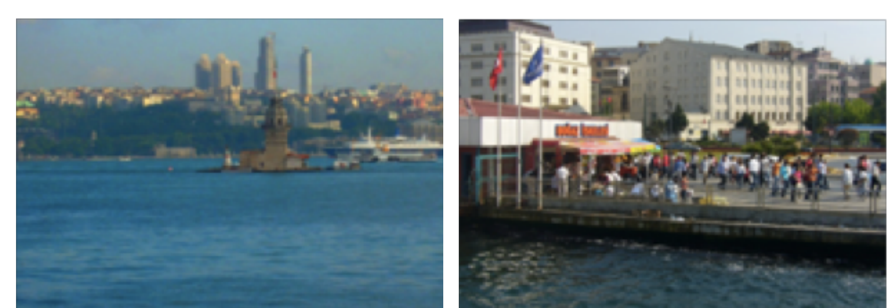


Our Approach

GIST

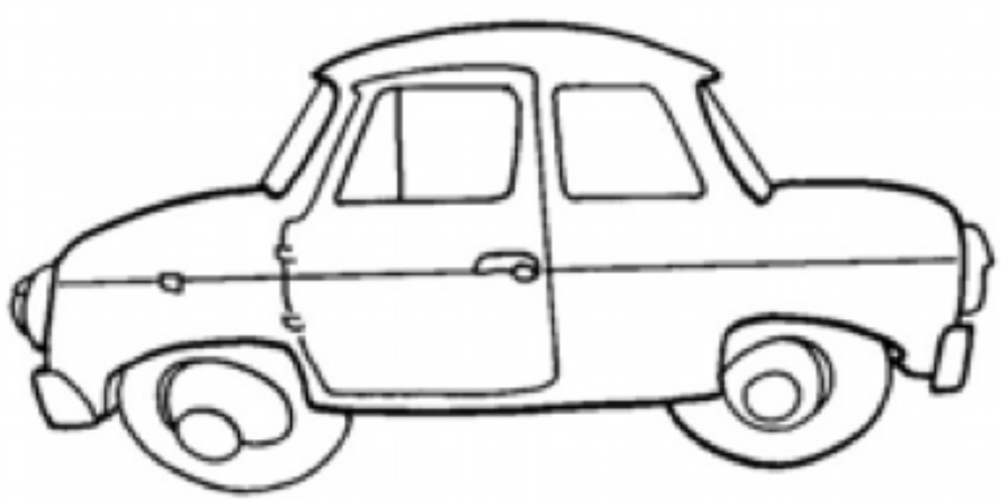
Bag-of-Words

Tiny Images



HOG

# Search Using Sketches



Input Sketch

Tiny Images

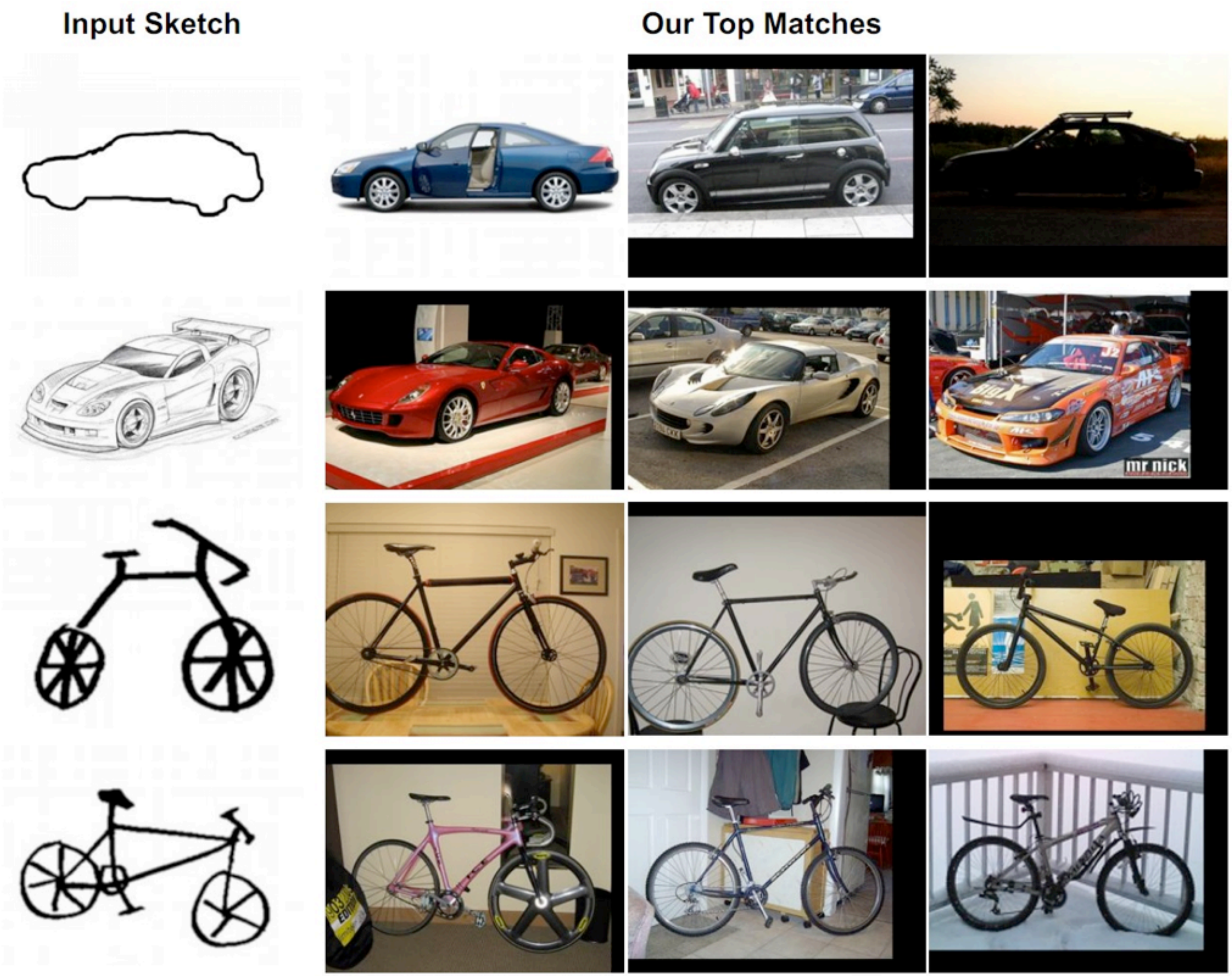

GIST

Our Approach

Bag-of-Words

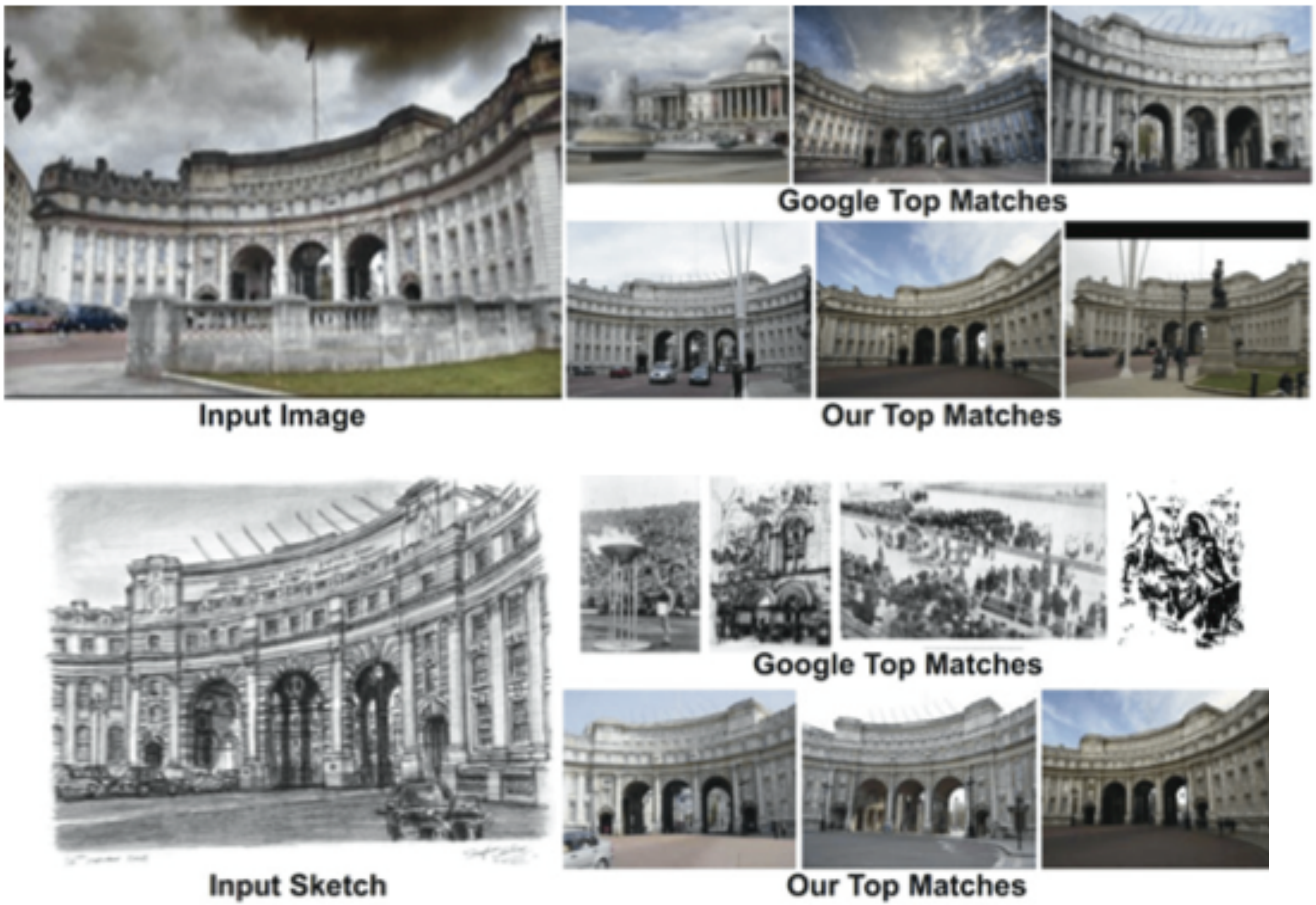
HOG

# Sketch to Image

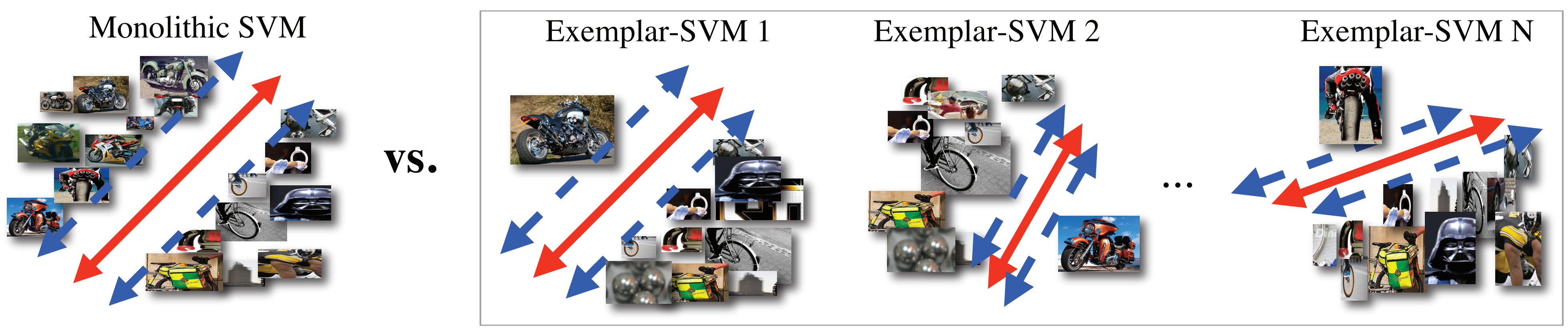


# Exemplar-SVM vs. Google

# Exemplar-SVM vs. Google



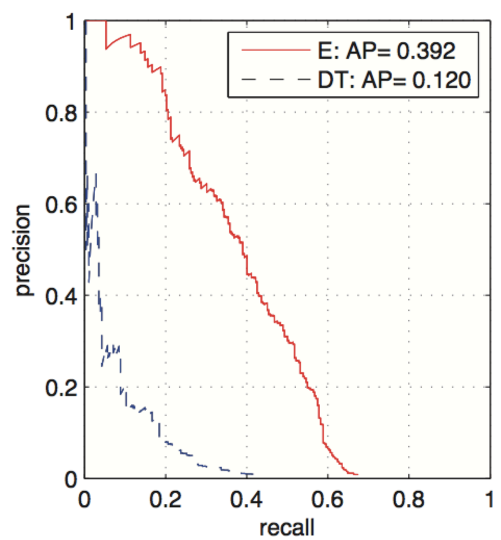
# Concluding Remarks

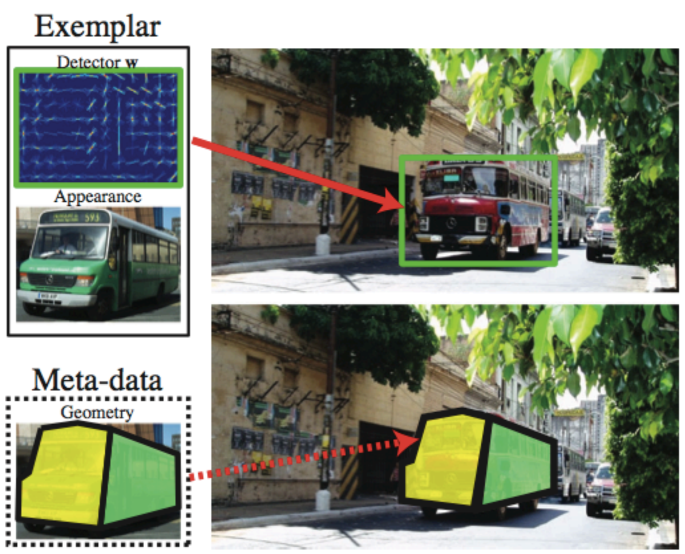


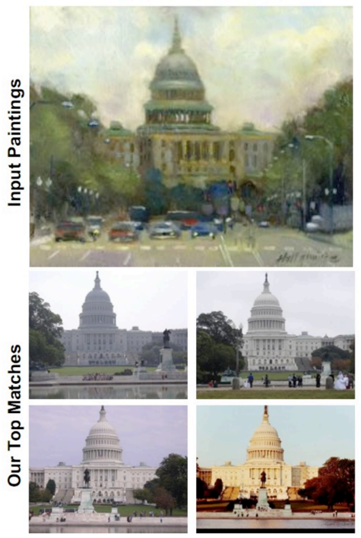

- A mixture model with N mixture components
- The positives are represented non-parametrically and the negatives are represented parametrically

# Concluding Remarks

- Exemplar-SVMs can be used for detection, label transfer, as well as cross-domain image matching

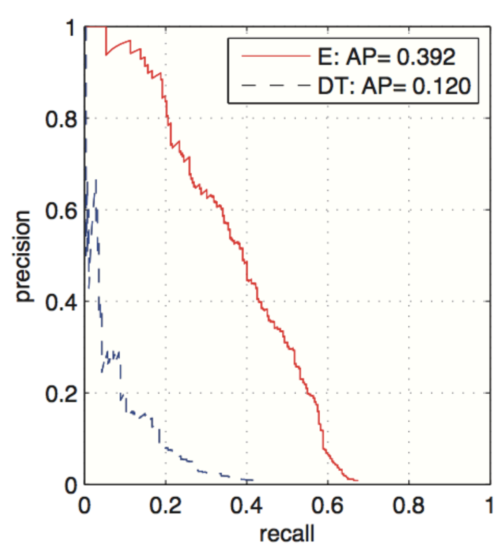


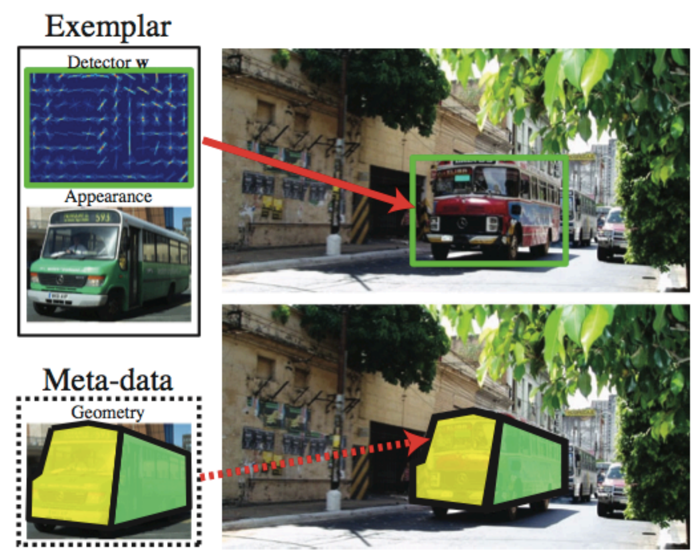



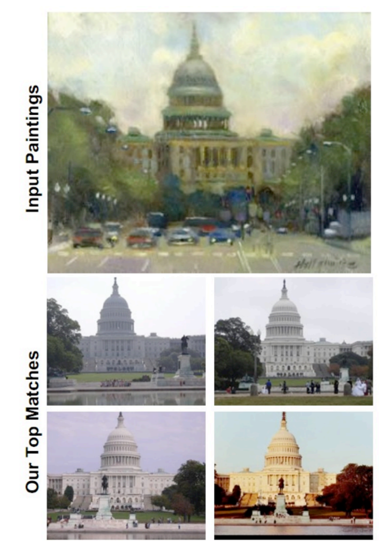



# Concluding Remarks

- Exemplar-SVMs can be used for detection, label transfer, as well as cross-domain image matching



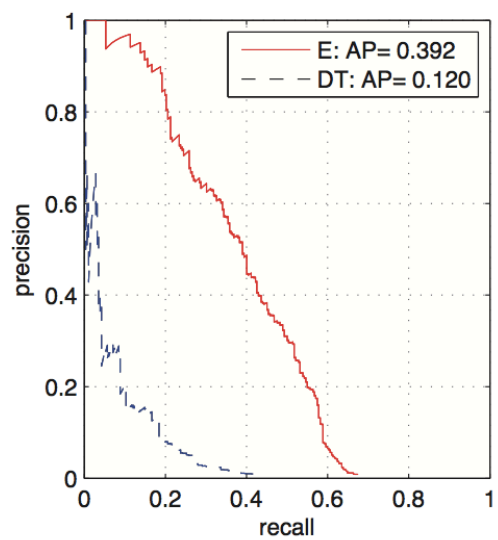



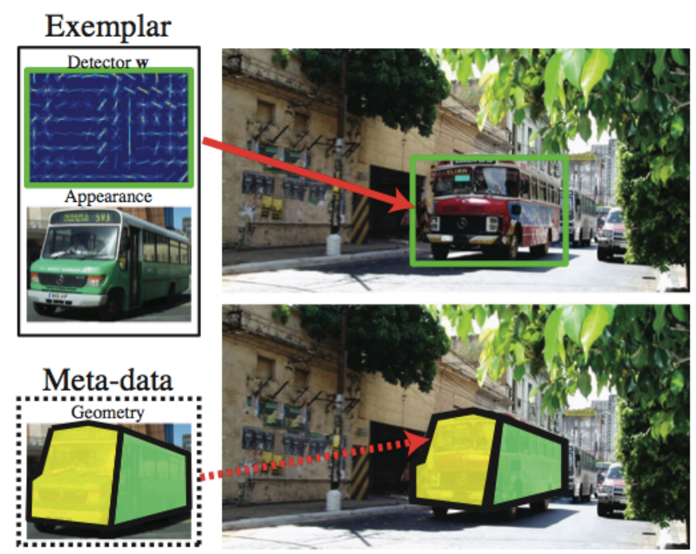



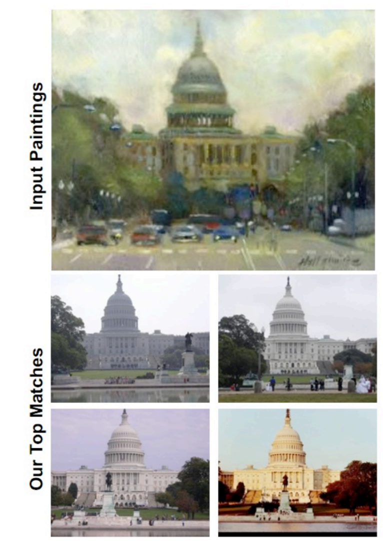

- Goods news: Results surprisingly nice, embarrassingly parallel learning

# Concluding Remarks

- Exemplar-SVMs can be used for detection, label transfer, as well as cross-domain image matching




- Goods news: Results surprisingly nice, embarrassingly parallel learning

- Bad news: Computationally Expensive

# Thank you

# Thank you

Come visit poster #30 in the Informatics Forum
or Google "exemplar svm" to find papers and code