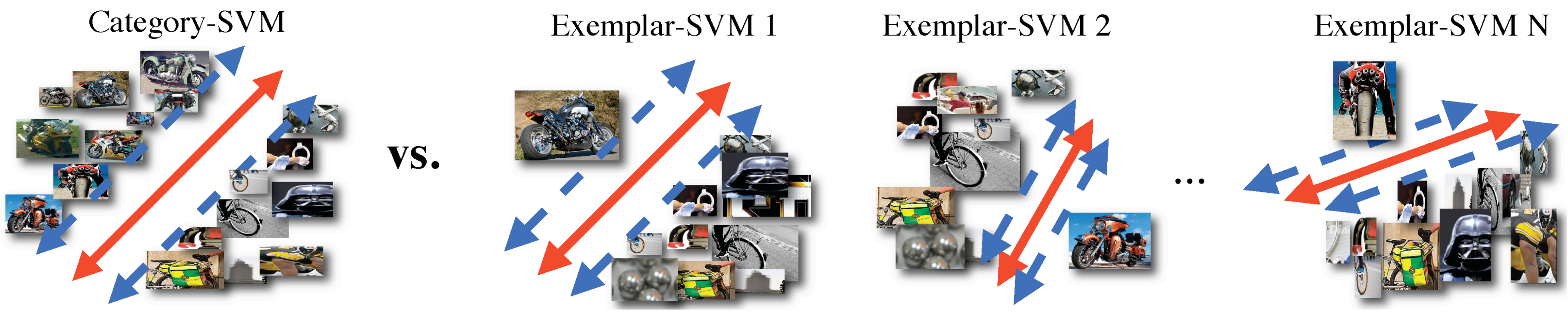
Ensemble of Exemplar-SVMs for Object Detection and Beyond



Tomasz Malisiewicz

September 20, 2011

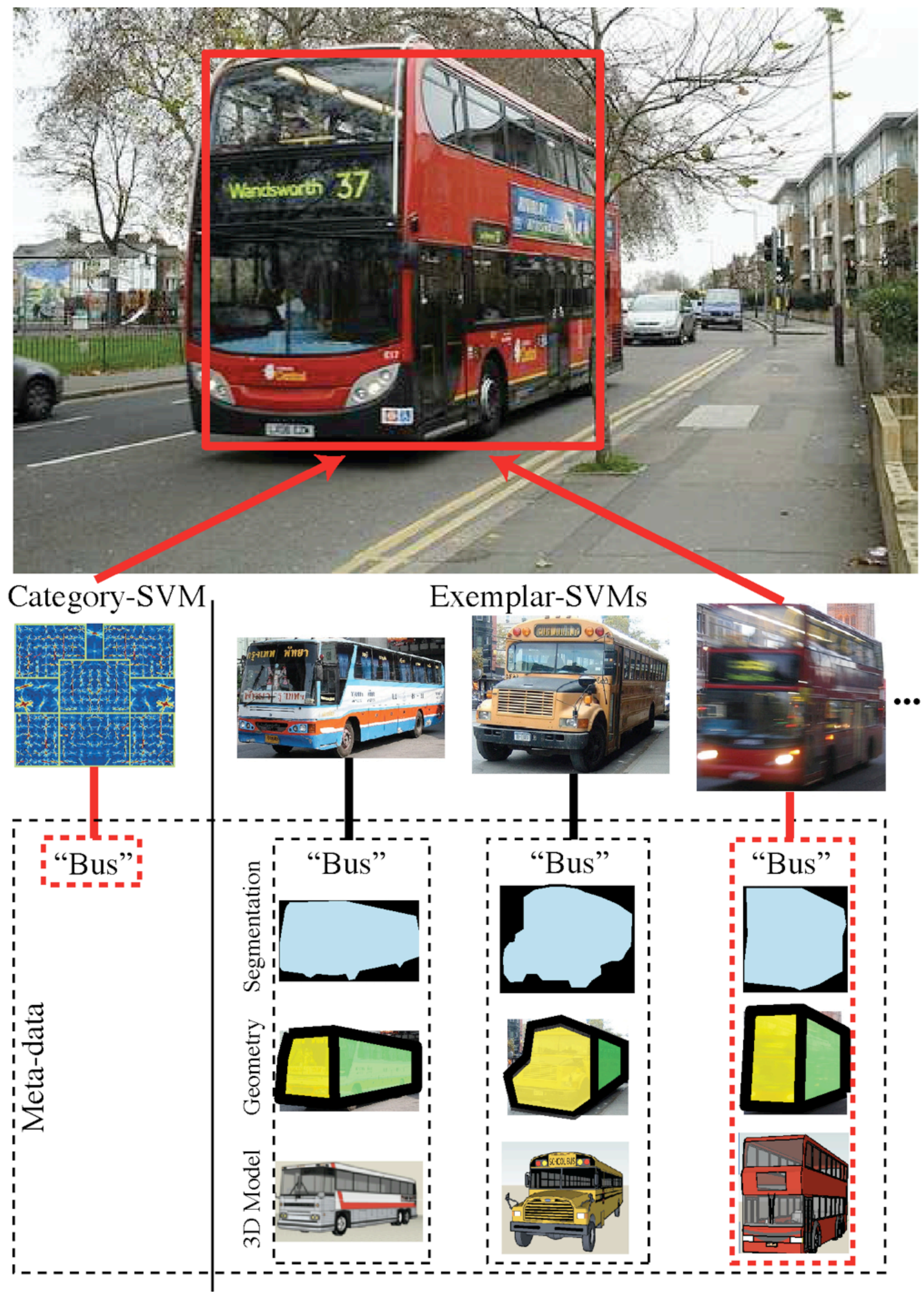
Computer Vision Reading Group@MIT

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

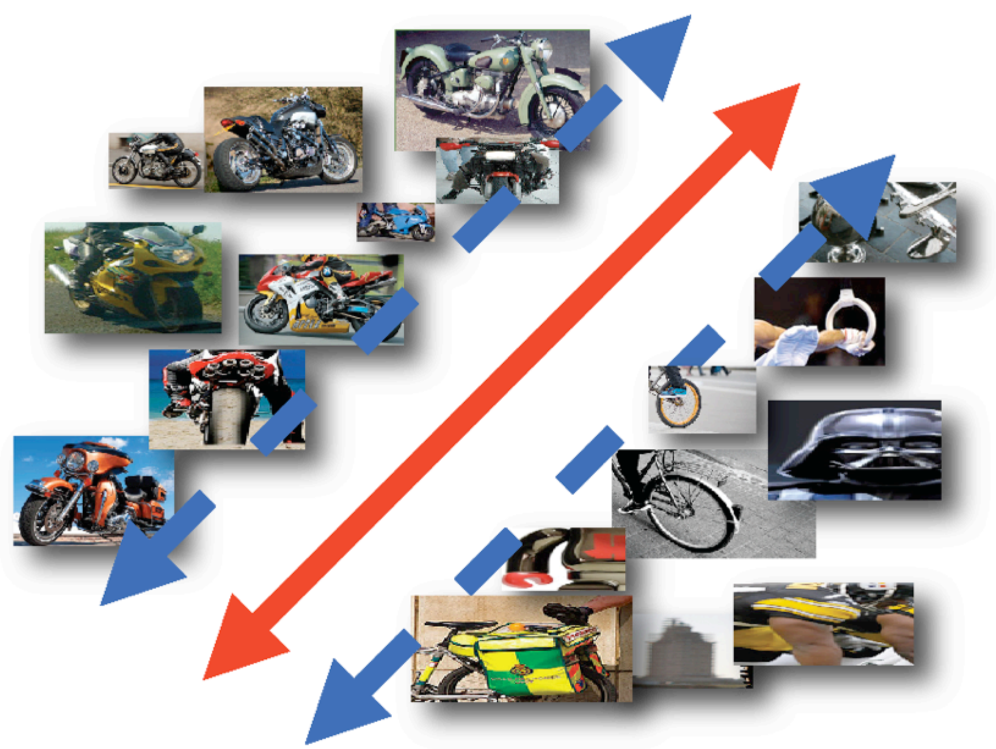
Overview

- Motivation and Related Work
- Learning Exemplar-SVMs
- Results
  - PASCAL VOC Object Detection Results
  - Transfer and Prediction


Wandsworth 37


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

Discriminative Object Detectors

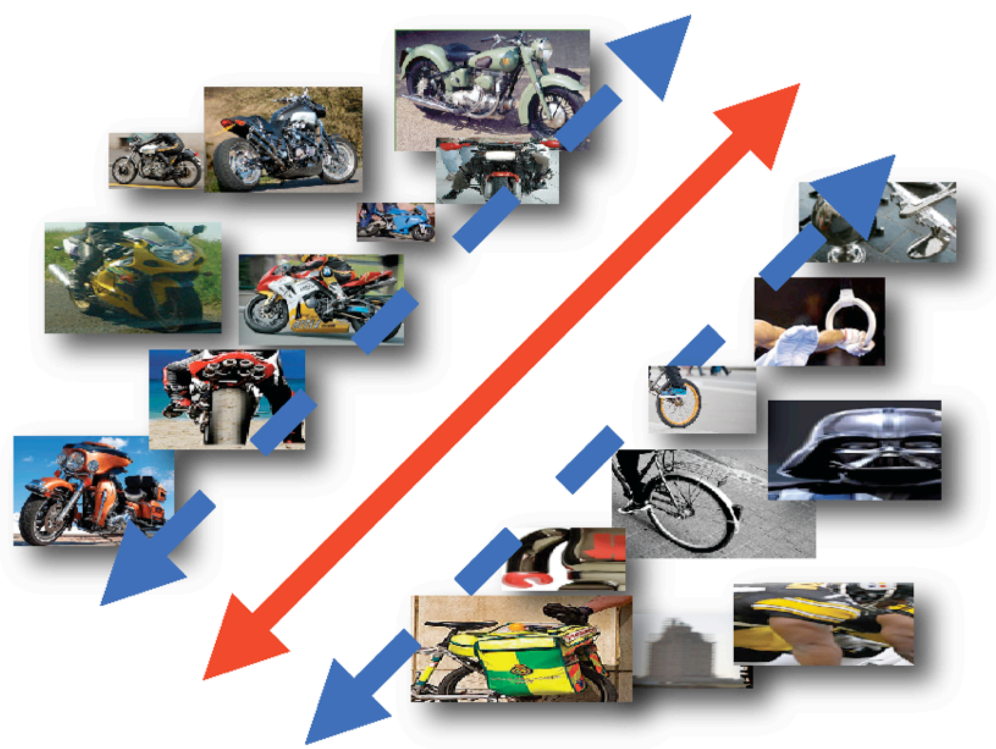


Linear SVM on HOG
Hard-Negative Mining
Sliding Window Detection

DT

Dalal and Triggs 2005

Discriminative Object Detectors

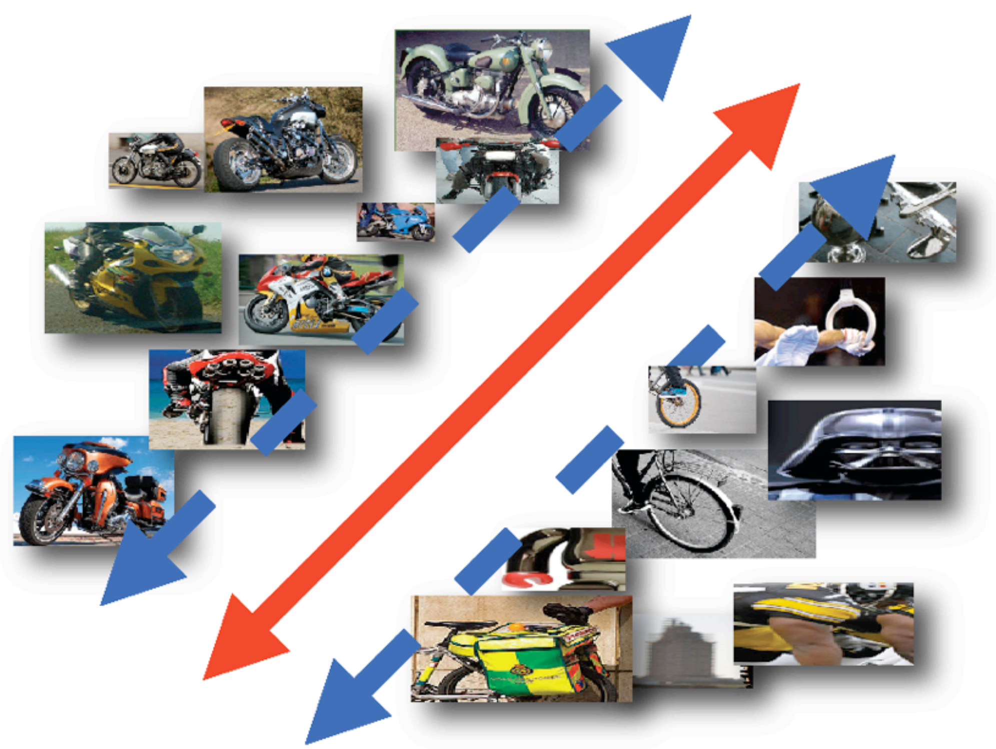


Linear SVM on HOG
Hard-Negative Mining
Sliding Window Detection
Parts
Mixtures

DT LDPM

Dalal and Triggs 2005, Felzenszwalb et al. 2010

Discriminative Object Detectors



Linear SVM on HOG
Hard-Negative Mining
Sliding Window Detection
Parts
Mixtures

Parametric: A fixed number of models per category

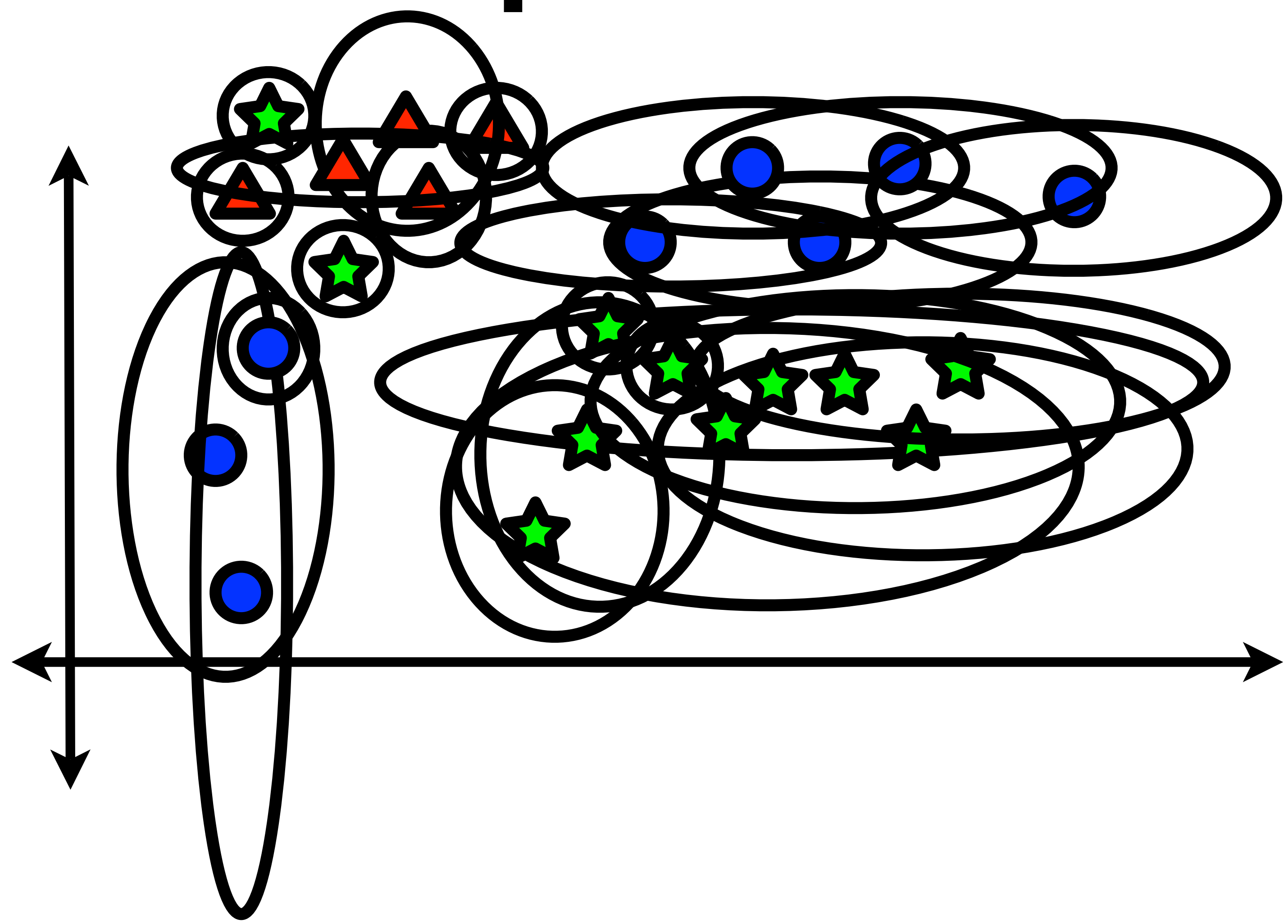
DT LDPM

Dalal and Triggs 2005, Felzenszwalb et al. 2010

Nearest Neighbor Approaches

- Non-parametric: keep all the data around
  - Enables **Label Transfer**
- However
  - No learning implies results depend on features and distance metric
  - Not shown to compete with discriminatively-trained LDPM on Pascal

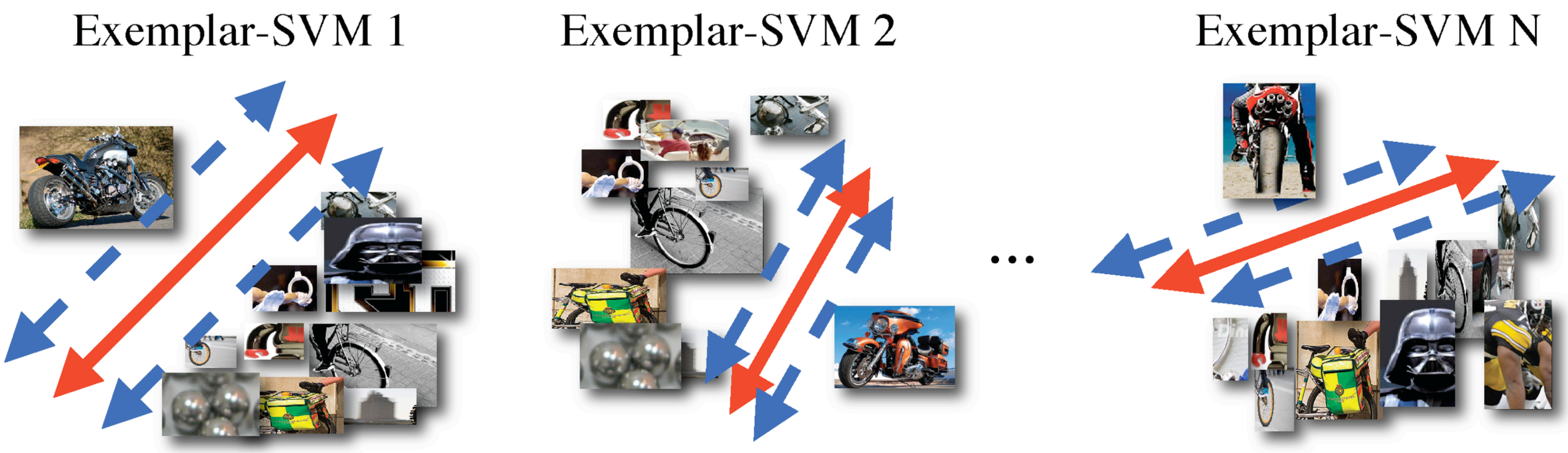
Per-Exemplar Methods



- NN-method, where each exemplar has its own distance "similarity" function
- Better than using a single similarity measure across all exemplars

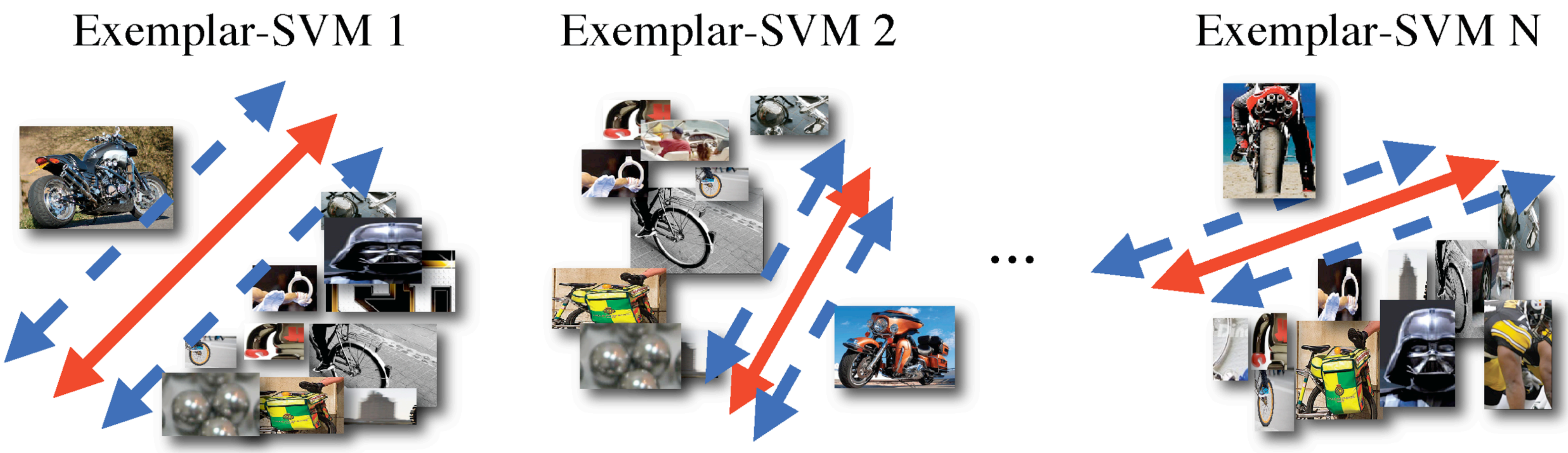
Frome et al. 2007, Malisiewicz et al. 2008

Exemplar-SVMs



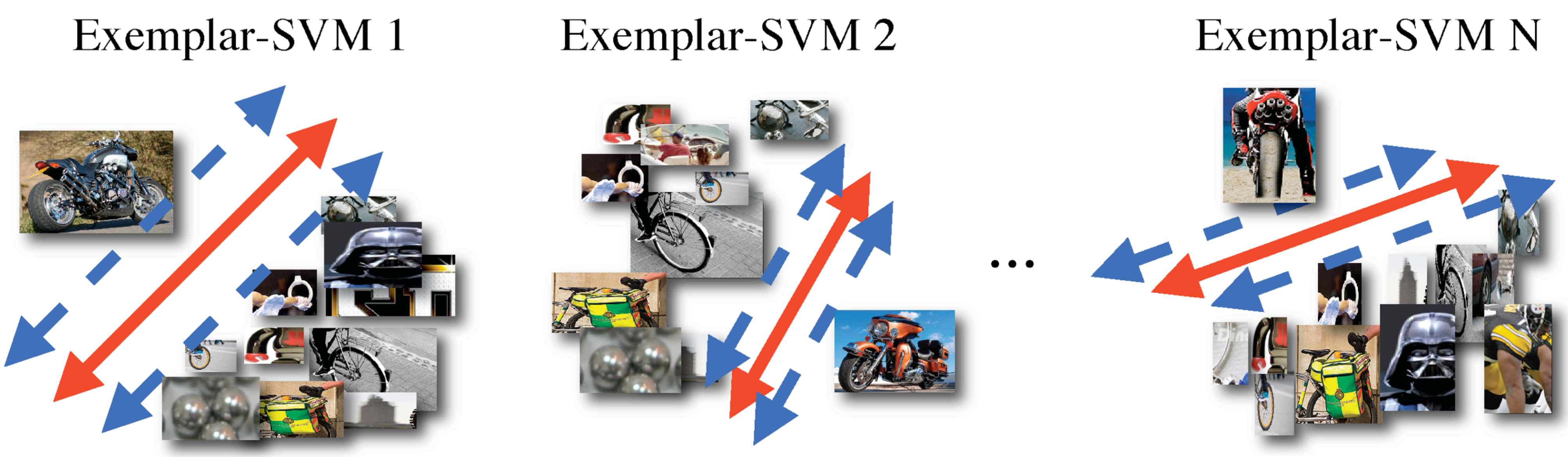
- Combine
  - Effectiveness of discriminatively-trained object detectors
  - Explicit correspondence of Nearest Neighbor approaches

Exemplar-SVMs



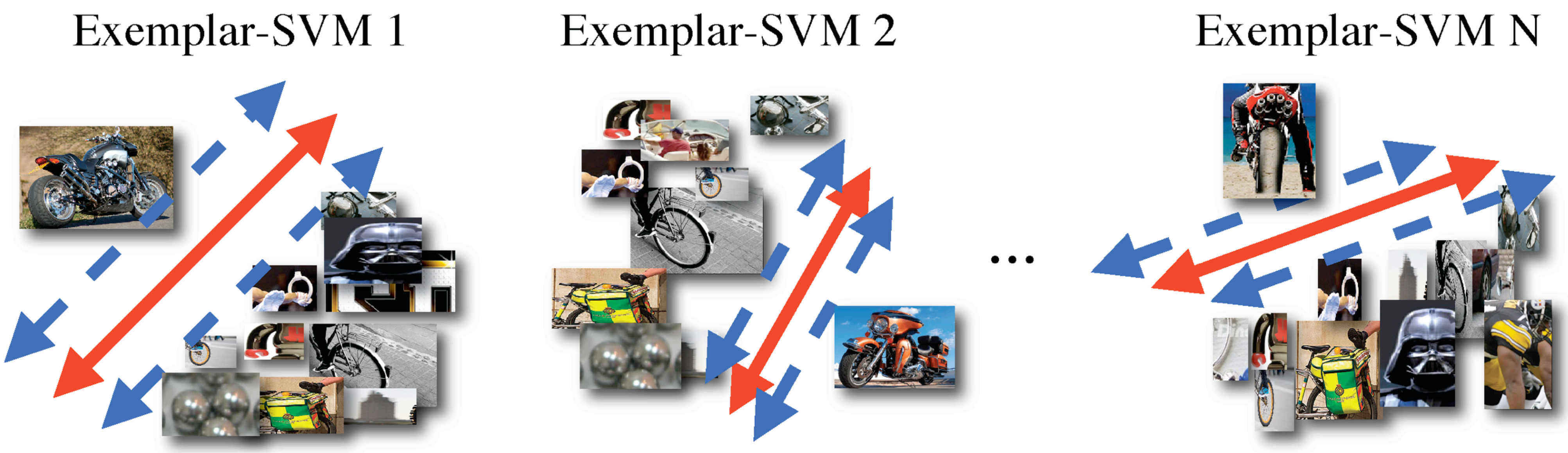
- Learn a separate linear SVM for each instance (exemplar) in the dataset (PASCAL VOC)

Exemplar-SVMs



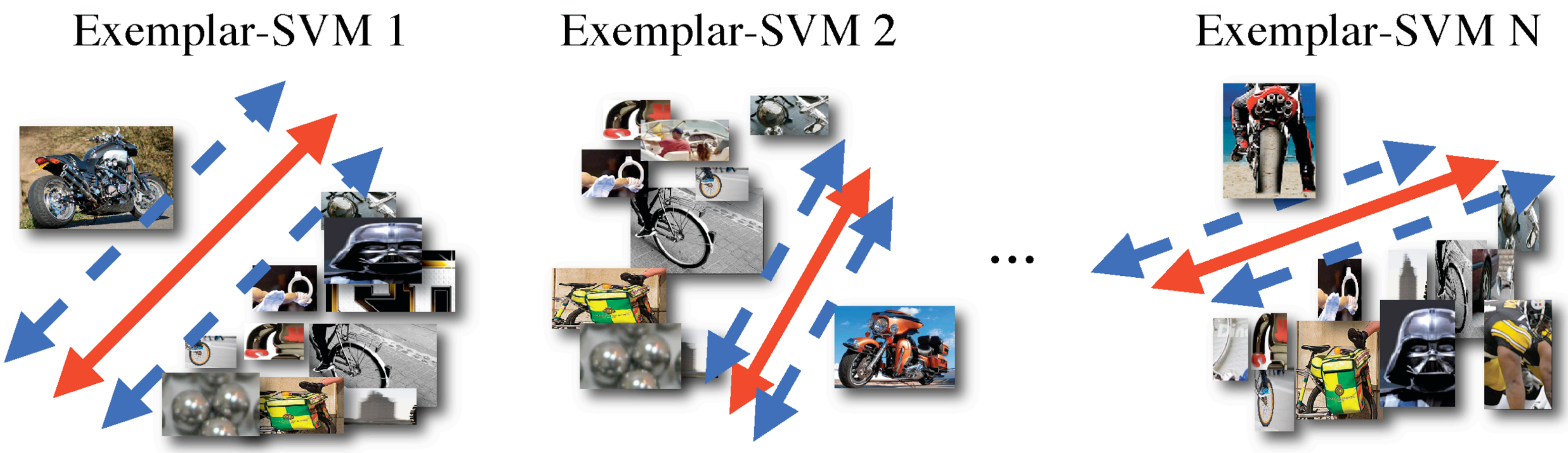
- Learn a separate linear SVM for each instance (exemplar) in the dataset (PASCAL VOC)
- Each Exemplar-SVM is trained with a **single** positive instance

Exemplar-SVMs



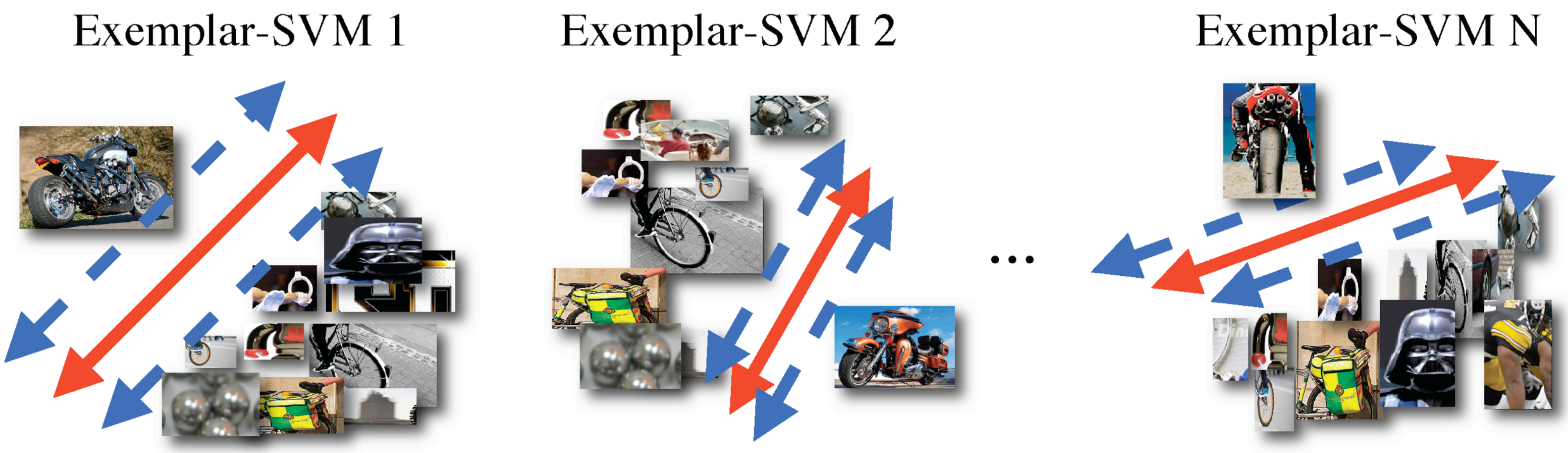
- Learn a separate linear SVM for each instance (exemplar) in the dataset (PASCAL VOC)
- Each Exemplar-SVM is trained with a **single** positive instance
- Each Exemplar-SVM is more defined by "*what it is not*" vs. "*what it is similar to*"

Exemplar-SVMs



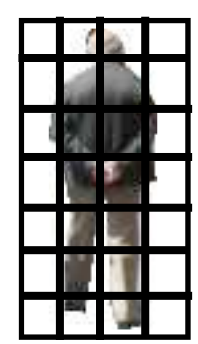
- Because each Exemplar-SVM is defined by a **single** positive instance, we can use different features for each exemplar

Exemplar-SVMs



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

- Adapt features to each exemplar's aspect ratio



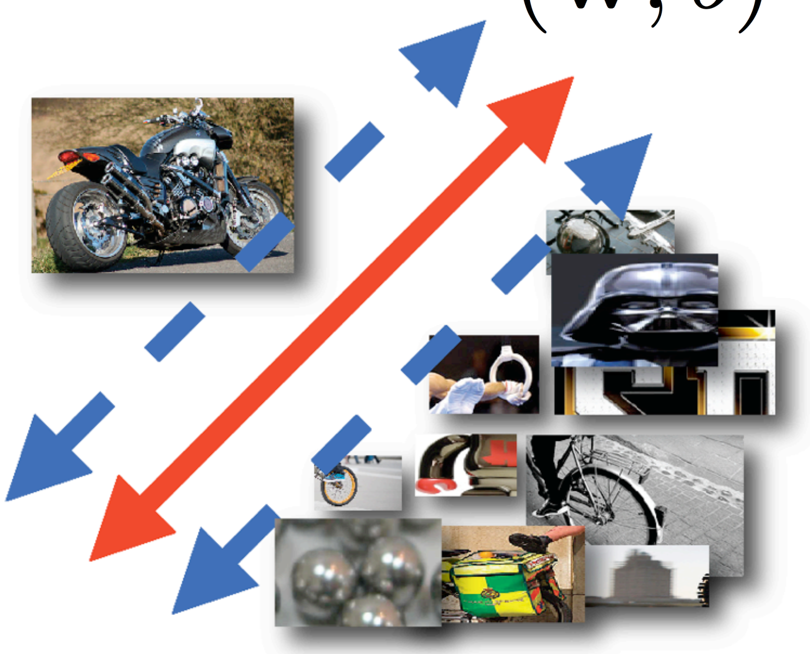
Exemplar-SVMs

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”

$(\mathbf{w}, b)$

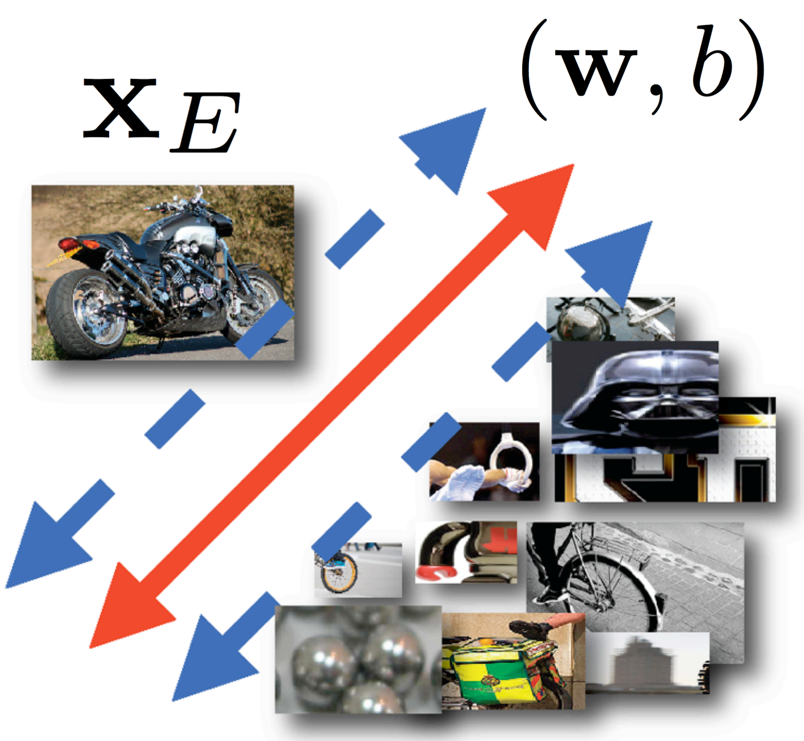


Exemplar-SVMs

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”



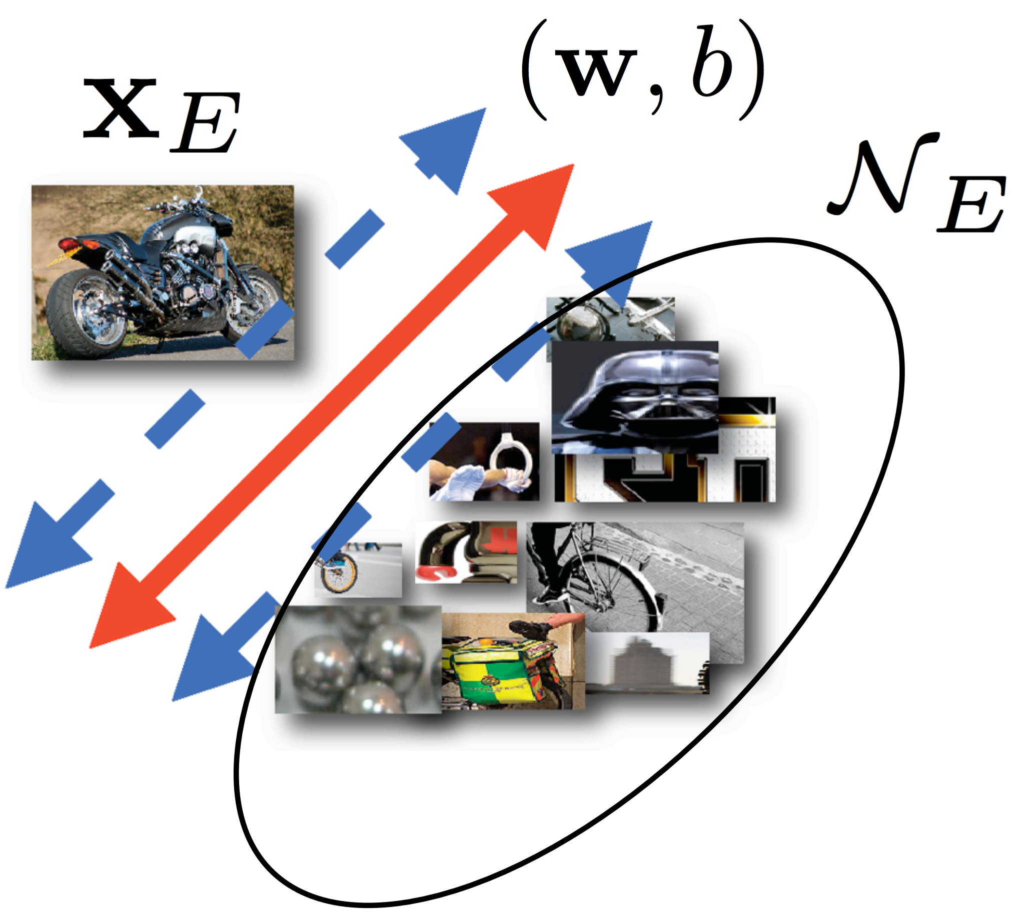
$\mathbf{x}_E$ Exemplar represented by ~100 HOG Cells (~3,100 features)

Exemplar-SVMs

Exemplar E's Objective Function:

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

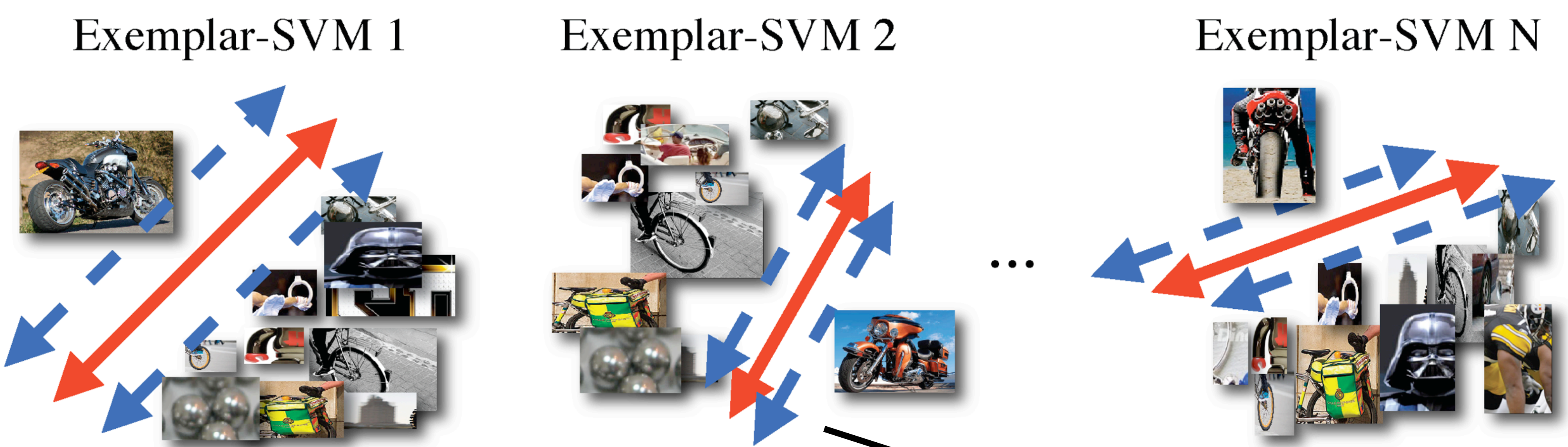
h(x) = max(1-x,0) "hinge-loss"

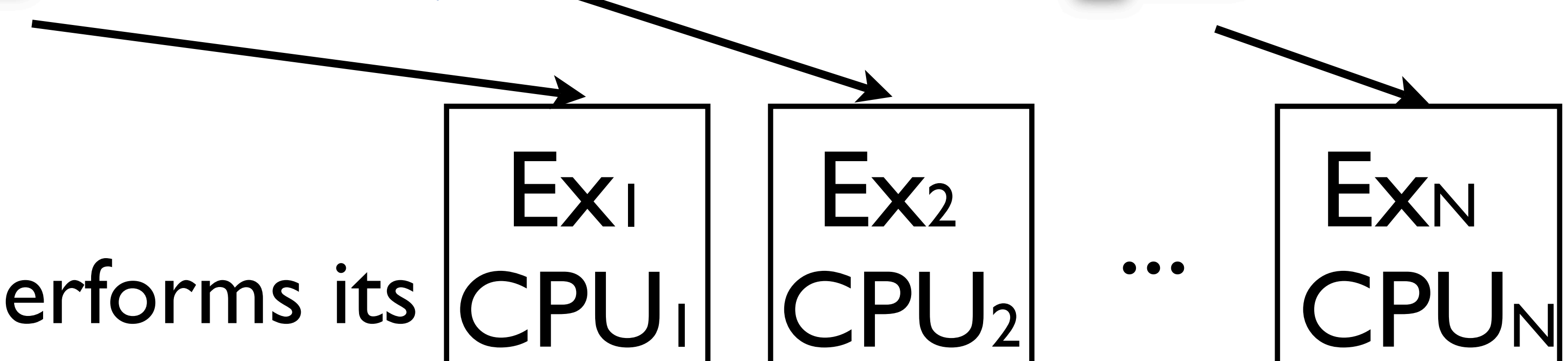


$\mathbf{x}_E$ Exemplar represented by ~100 HOG Cells (~3,100 features)

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

Large-scale training

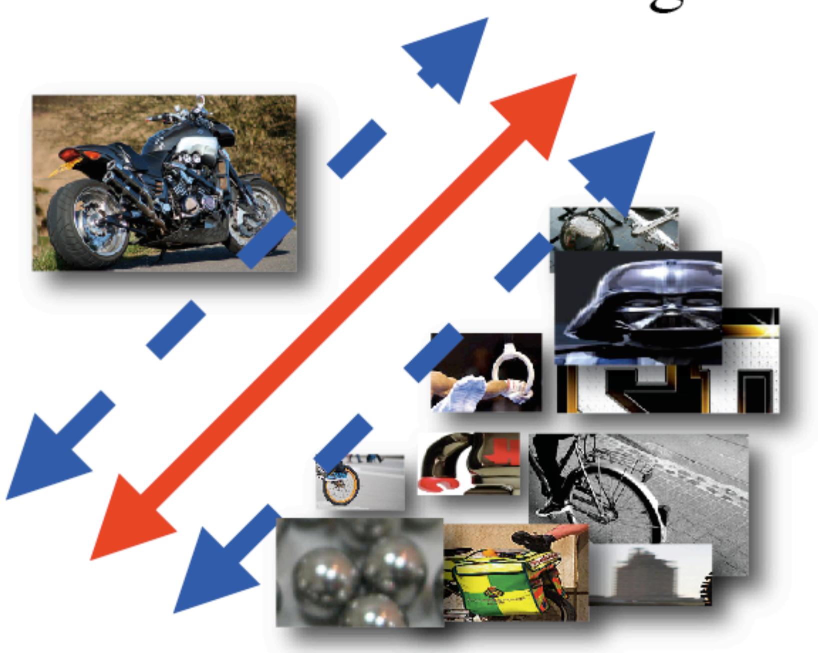




- Each exemplar performs its own hard negative mining
- Solve many convex learning problems
- Parallel training on cluster

Exemplar-SVM Calibration

SVM after training



Exemplar-SVM Calibration

1) Apply ExemplarSVM to held-out negative images and all positive images

Exemplar-SVM Calibration

1) Apply ExemplarSVM to held-out negative images and all positive images

2) Fit sigmoid to responses [Platt 1999]

$$f(\mathbf{x}|\mathbf{w}_E, \alpha_E, \beta_E) = \frac{1}{1 + e^{-\alpha_E(\mathbf{w}_E^T \mathbf{x} - \beta_E)}}$$

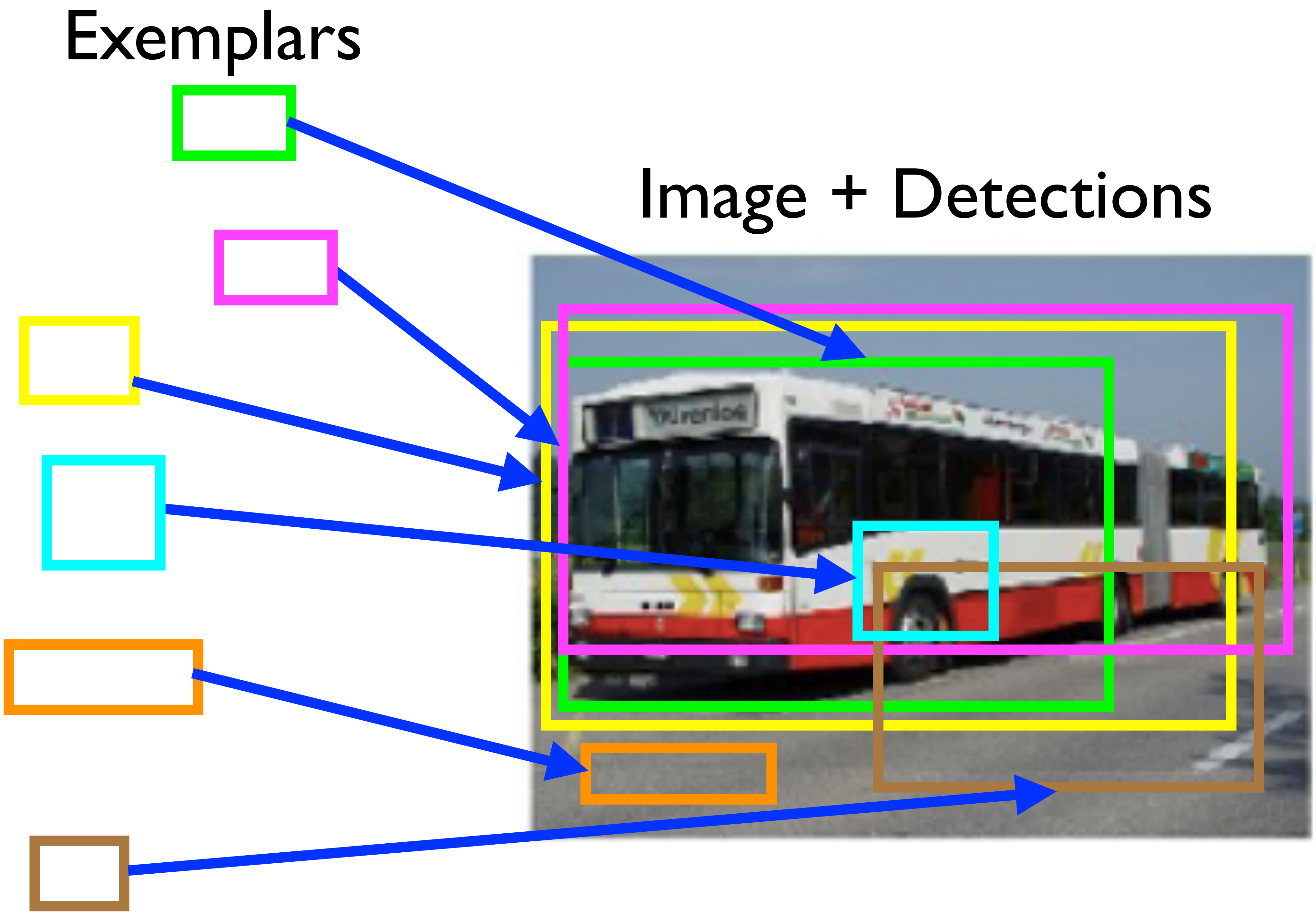
Exemplar-SVM Calibration

1) Apply ExemplarSVM to held-out negative images and all positive images

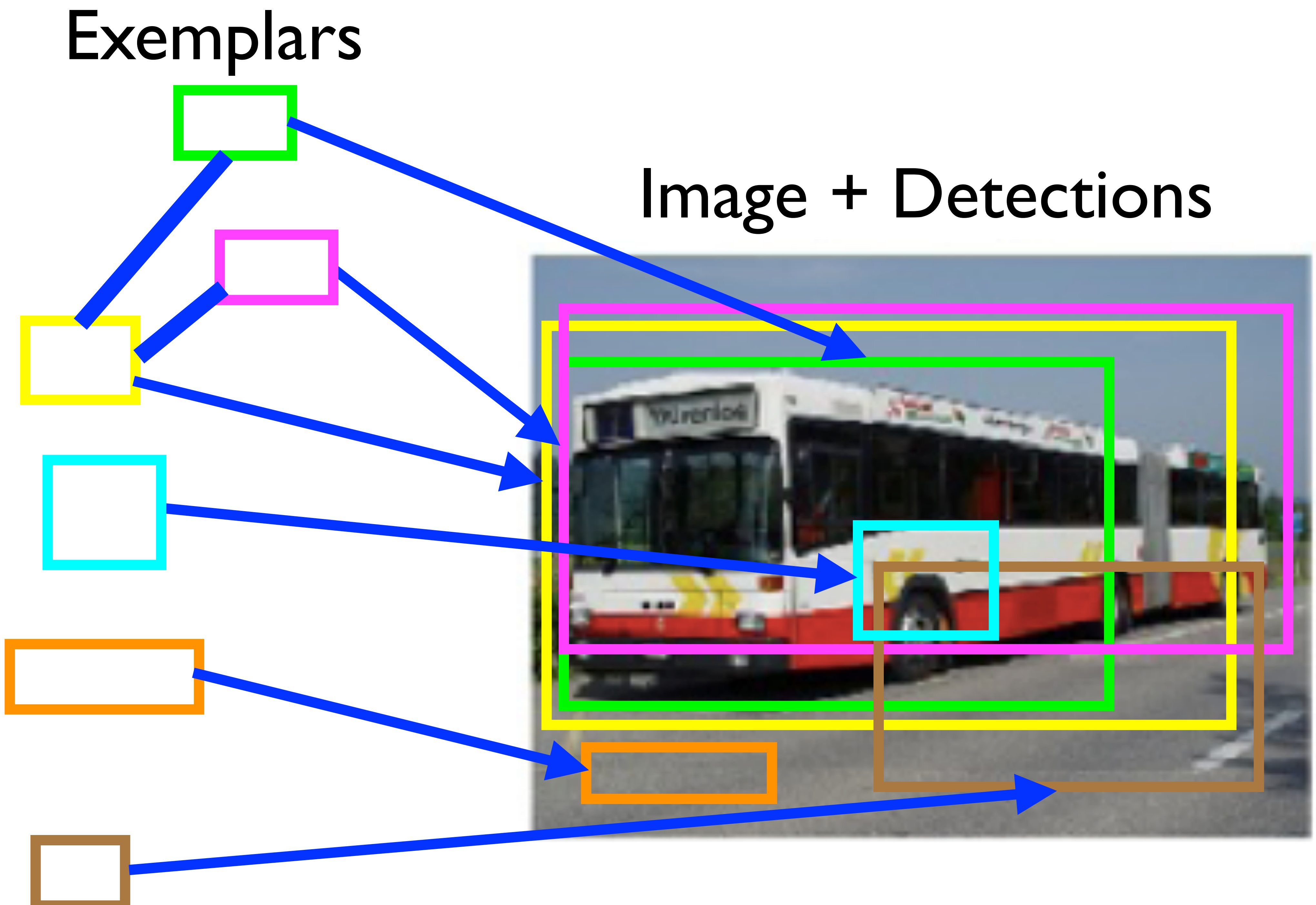
2) Fit sigmoid to responses [Platt 1999]

$$f(\mathbf{x}|\mathbf{w}_E, \alpha_E, \beta_E) = \frac{1}{1 + e^{-\alpha_E(\mathbf{w}_E^T\mathbf{x} - \beta_E)}}$$

Ensemble of Exemplar-SVMs



Ensemble of Exemplar-SVMs

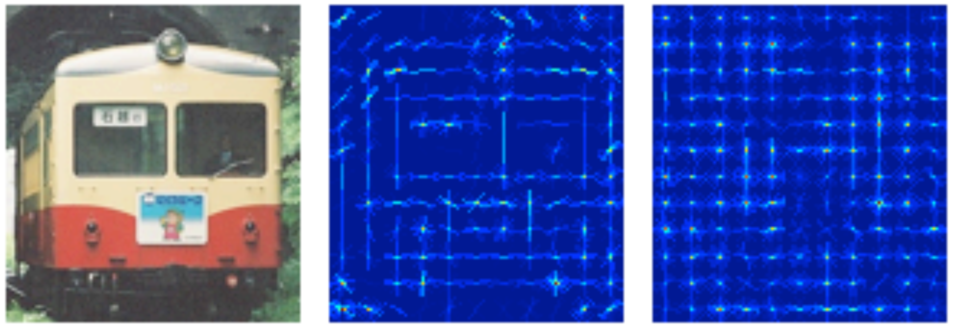


Learn an exemplar **co-occurence matrix**

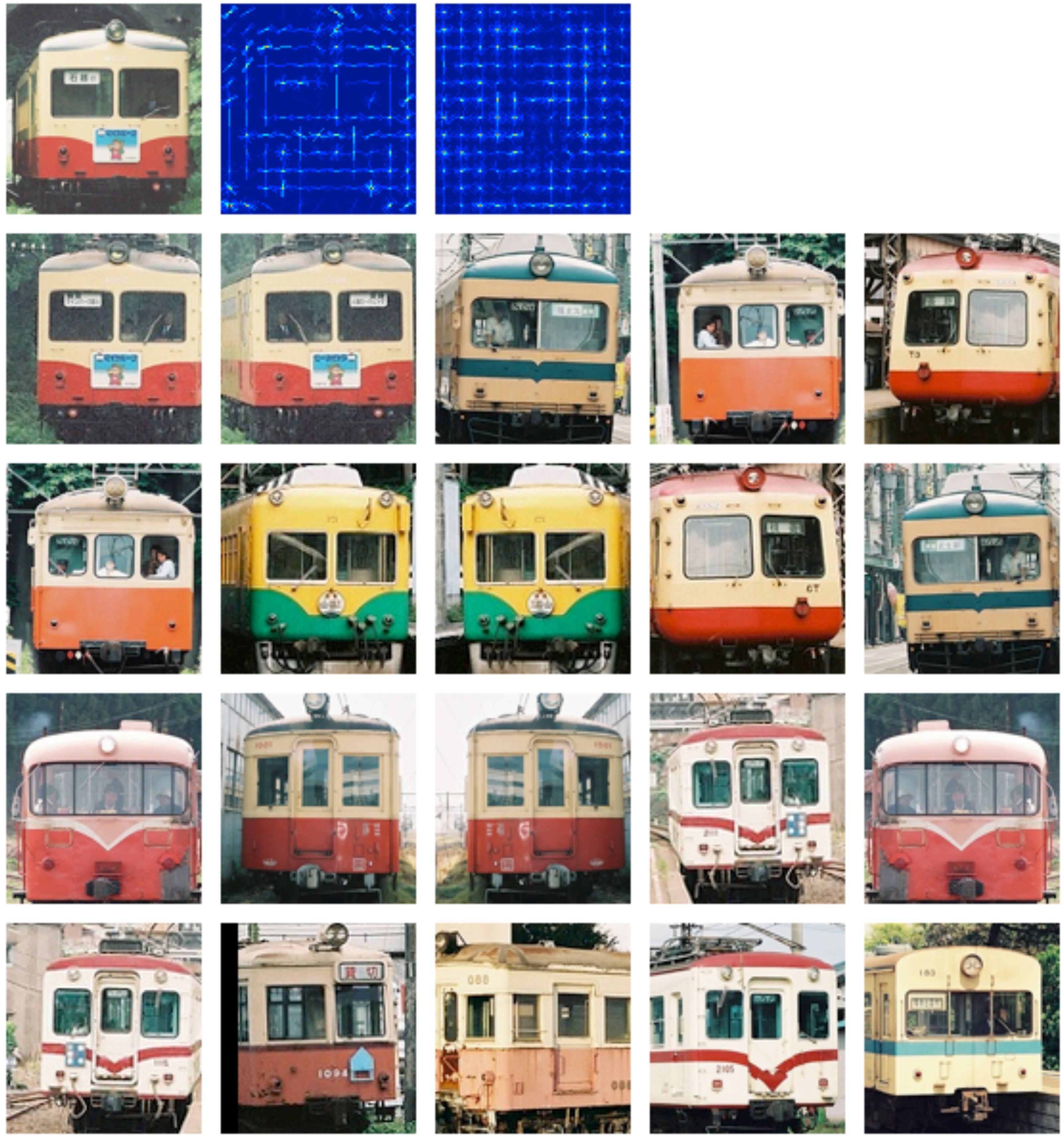
Qualitative Results

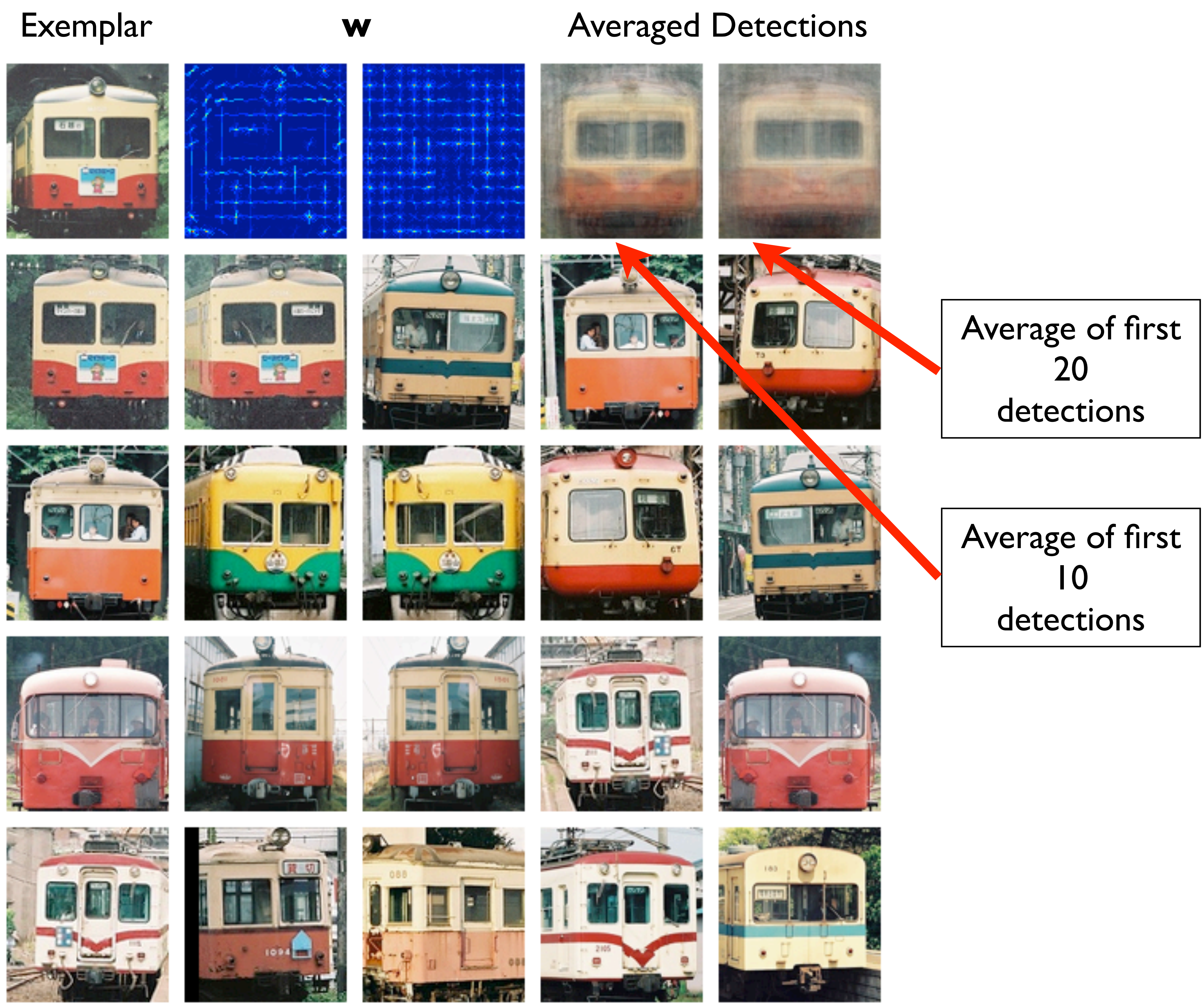
- Let's take a look at some Exemplar-SVM results in PASCAL VOC dataset

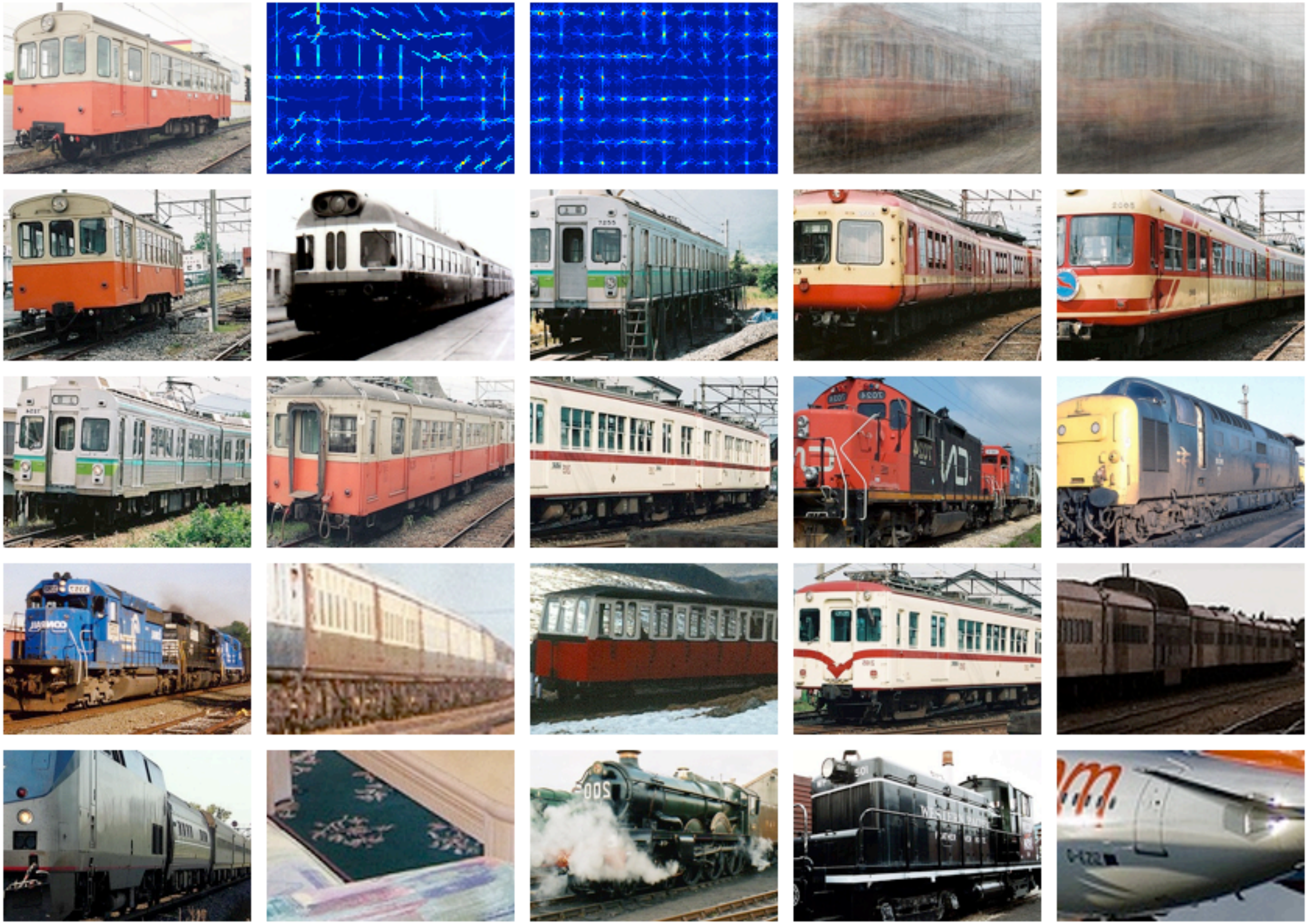
Exemplar **w**

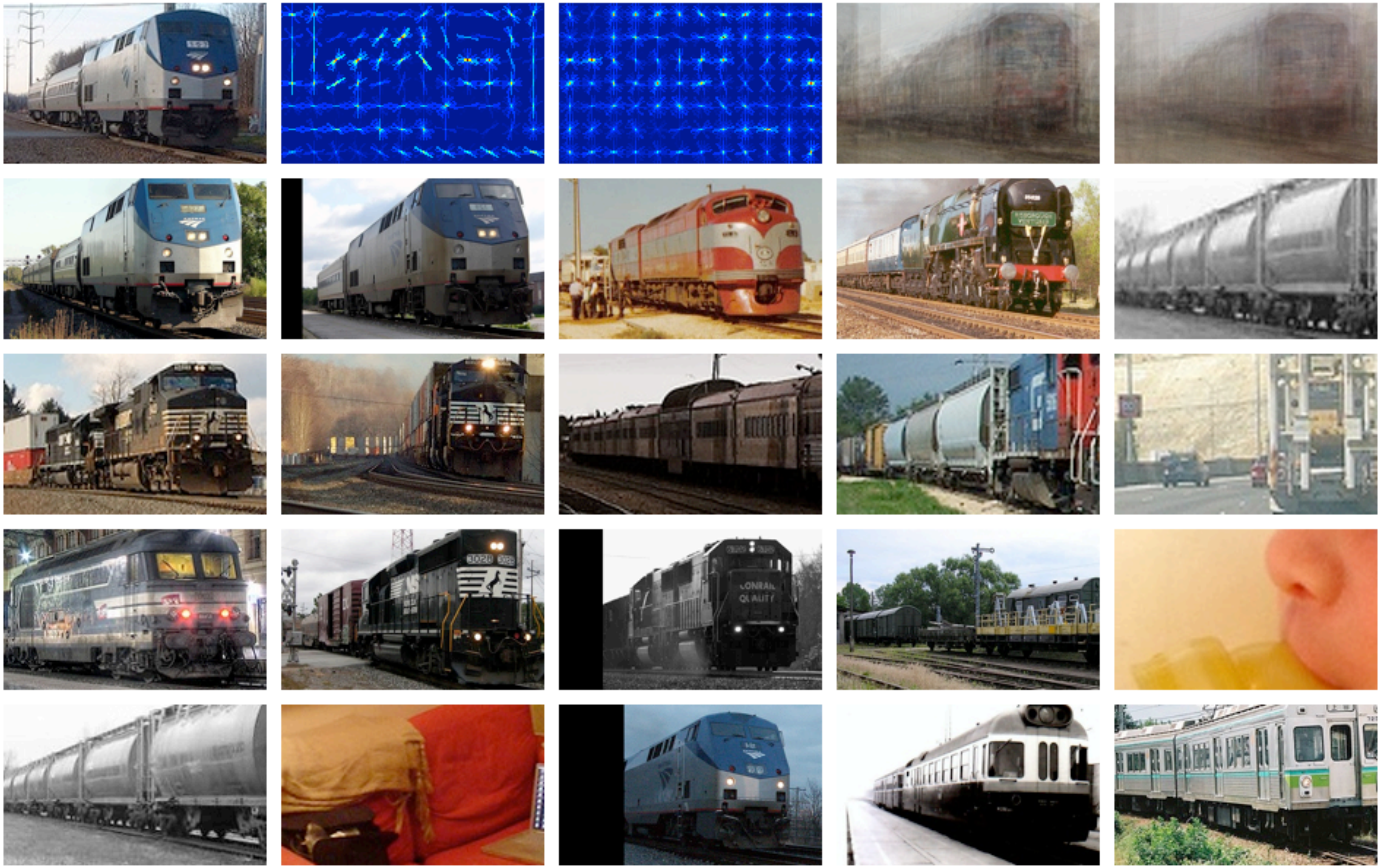


Exemplar **w**




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



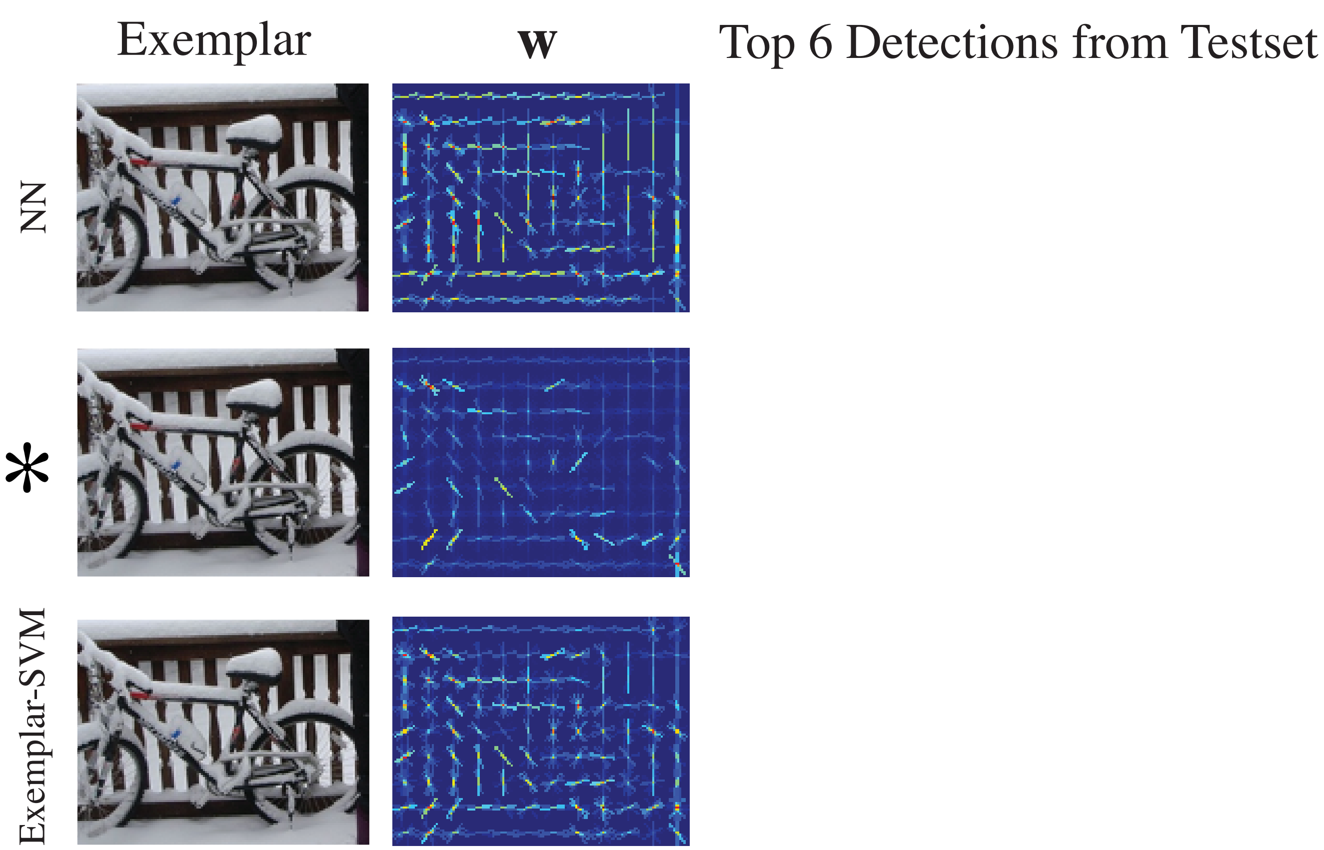



NREX
Pan Am Railways
8006
BNSF
DANGER
0806

Evaluating Exemplar-SVMs

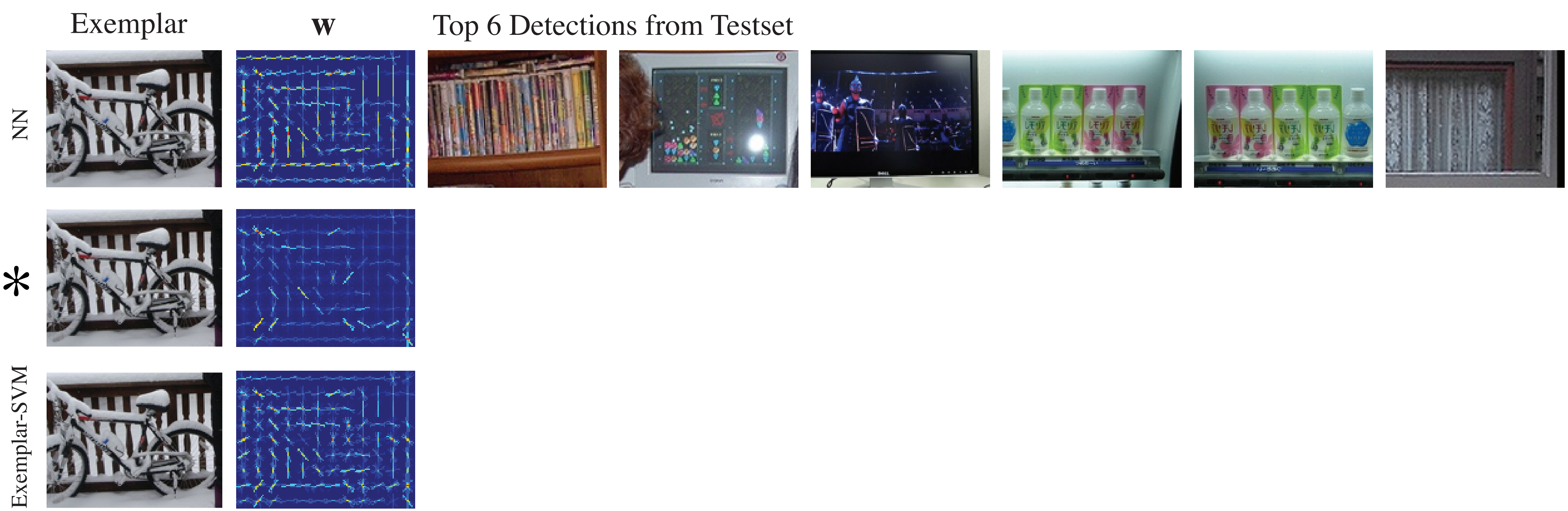
- **Nearest Neighbor**
  - No Learning
- **Per-Exemplar Distance Functions**
  - Learning in distance-to-exemplar space [Malisiewicz et al. 2008]
- **Exemplar-SVMs**

Comparison of 3 methods



*Learned Distance Function

Comparison of 3 methods



*Learned Distance Function

Comparison of 3 methods

*Learned Distance Function

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

PASCAL VOC 2007 Object Category Detection Results

Approach	aeroplane	bicycle	bird	boat	bottle	bus	car	cat	chair	cow	diningtable	dog	horse	motorbike	person	pottedplant	sheep	sofa	train	tvmonitor	mAP
NN	.006	.094	.000	.005	.000	.006	.010	.092	.001	.092	.001	.004	.096	.094	.005	.018	.009	.008	.096	.144	.039
NN+Cal	.056	.293	.012	.034	.009	.207	.261	.017	.094	.111	.004	.033	.243	.188	.114	.020	.129	.003	.183	.195	.110
DFUN+Cal	.162	.364	.008	.096	.097	.316	.366	.092	.098	.107	.002	.093	.234	.223	.109	.037	.117	.016	.271	.293	.155
E-SVM+Cal	.204	.407	.093	.100	.103	.310	.401	.096	.104	.147	.023	.097	.384	.320	.192	.096	.167	.110	.291	.315	.198
E-SVM+Co-occ	.208	.480	.077	.143	.131	.397	.411	.052	.116	.186	.111	.031	.447	.394	.169	.112	.226	.170	.369	.300	.227
CZ [6]	.262	.409	–	–	–	.393	.432	–	–	–	–	–	–	.375	–	–	–	–	.334	–	–
DT [7]	.127	.253	.005	.015	.107	.205	.230	.005	.021	.128	.014	.004	.122	.103	.101	.022	.056	.050	.120	.248	.097
LDPM [9]	.287	.510	.006	.145	.265	.397	.502	.163	.165	.166	.245	.050	.452	.383	.362	.090	.174	.228	.341	.384	.266

Table 1. **PASCAL VOC 2007 object detection results.** We compare our full system (ESVM+Co-occ) to four different exemplar based baselines including NN (Nearest Neighbor), NN+Cal (Nearest Neighbor with calibration), DFUN+Cal (learned distance function with calibration) and ESVM+Cal (Exemplar-SVM with calibration). We also compare our approach against global methods including our implementation of Dalal-Triggs (learning a single global template), LDPM [9] (Latent deformable part model), and Chum et al. [6]'s exemplar-based method. [The NN, NN+Cal and DFUN+Cal results for person category are obtained using 1250 exemplars]

Object Category Detection

mAP on PASCAL VOC 2007 detection task

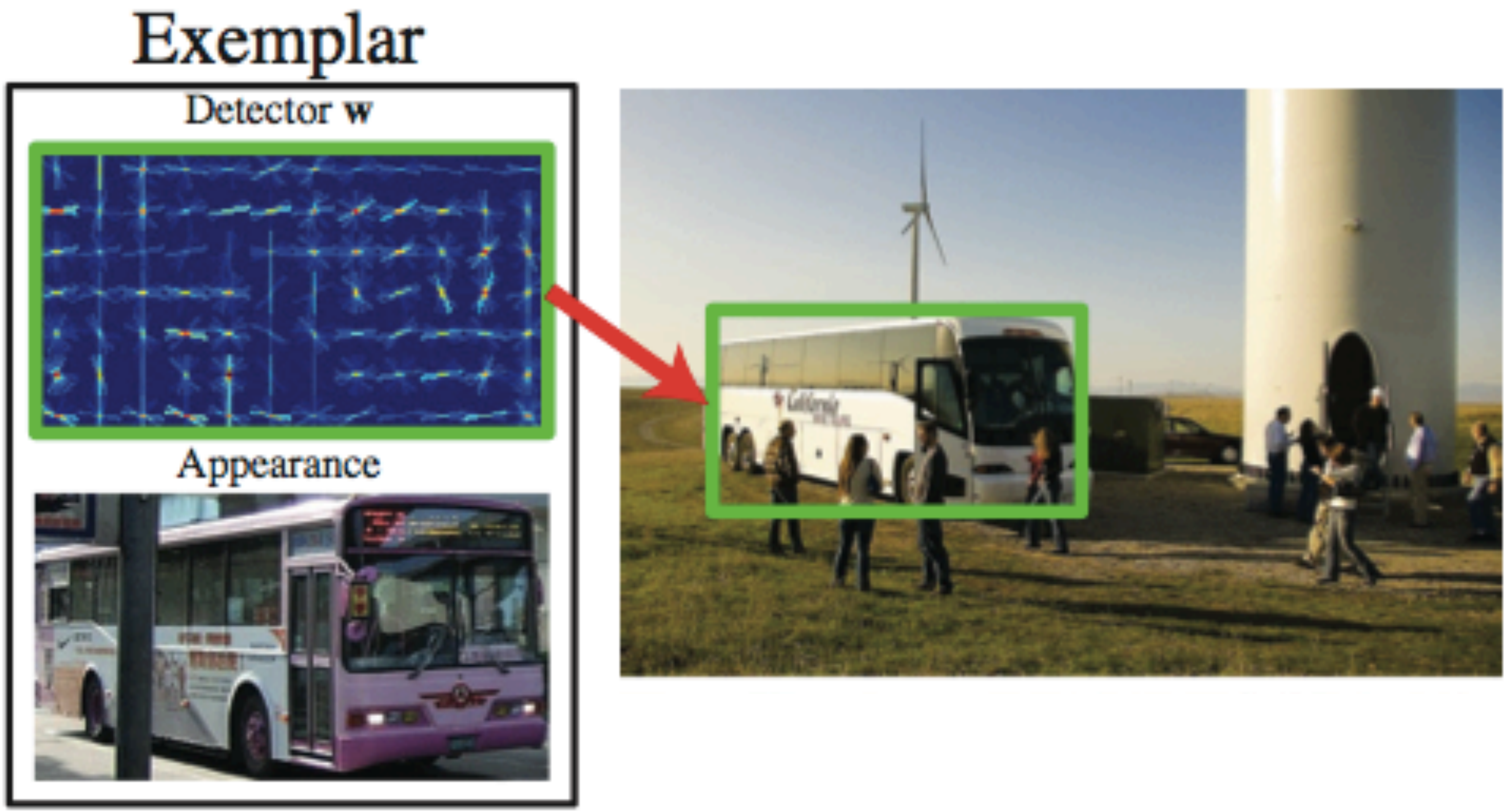
NN + Cal	0.110
DFUN + Cal	0.155
Exemplar-SVMs + Cal	0.198
Exemplar-SVMs + Co-occ	0.227
DT*	0.097
LDPM**	0.266

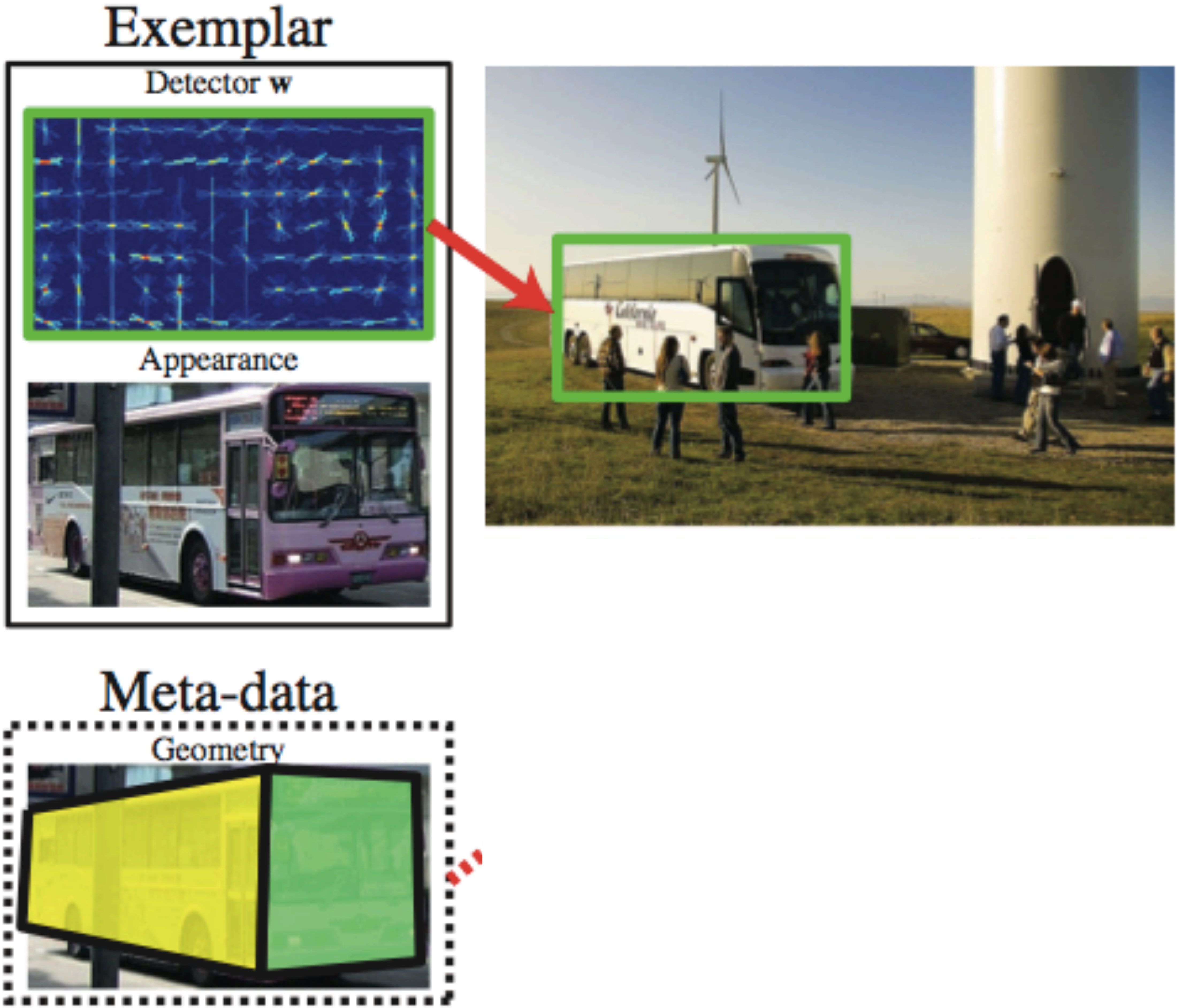
*Dalal et al. 2005 **Felzenszwalb et al. 2010

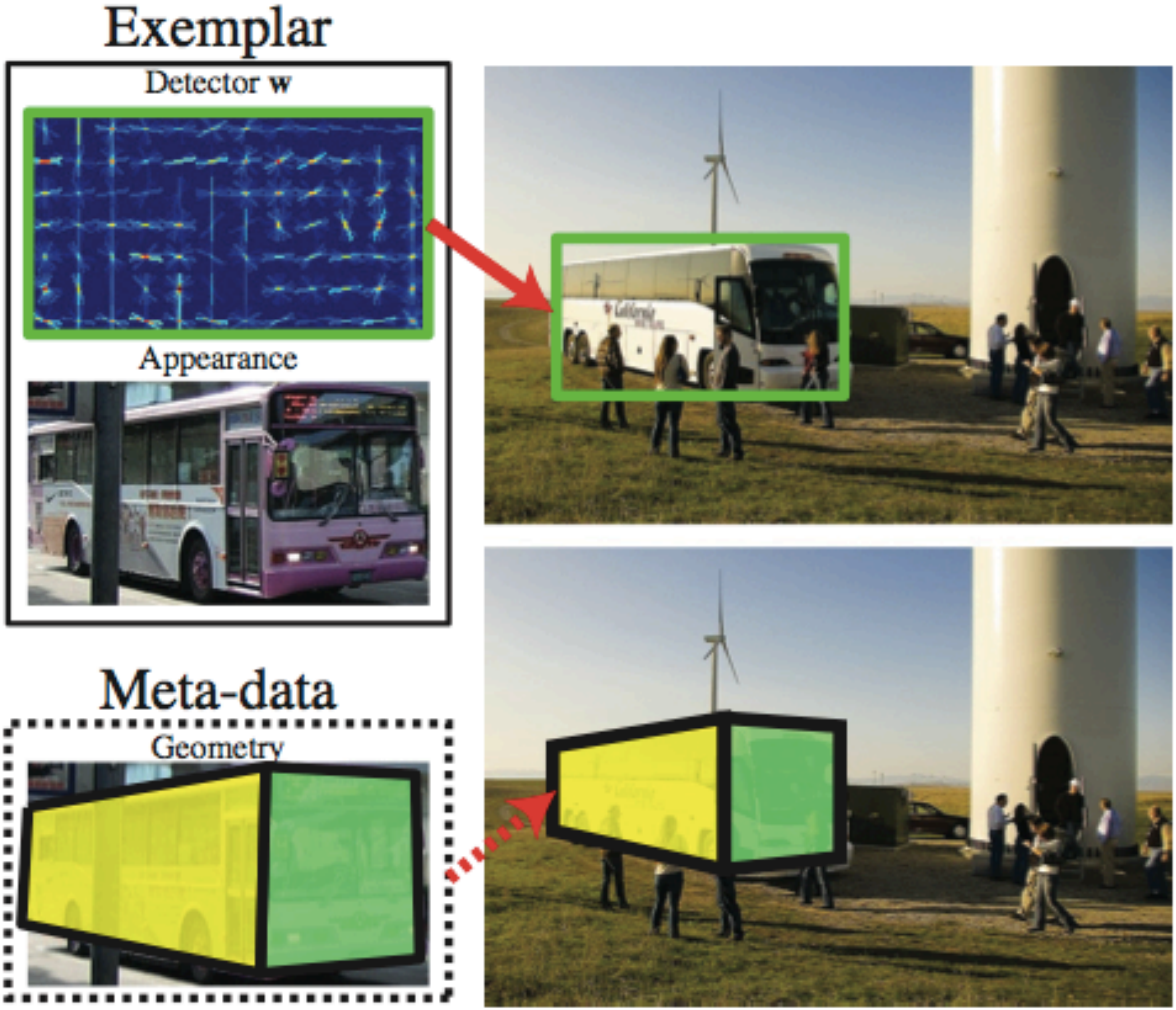
Beyond Object Category Detection

- Based on the idea of label transfer, ExemplarSVMs can be used for tasks which go beyond object category detection

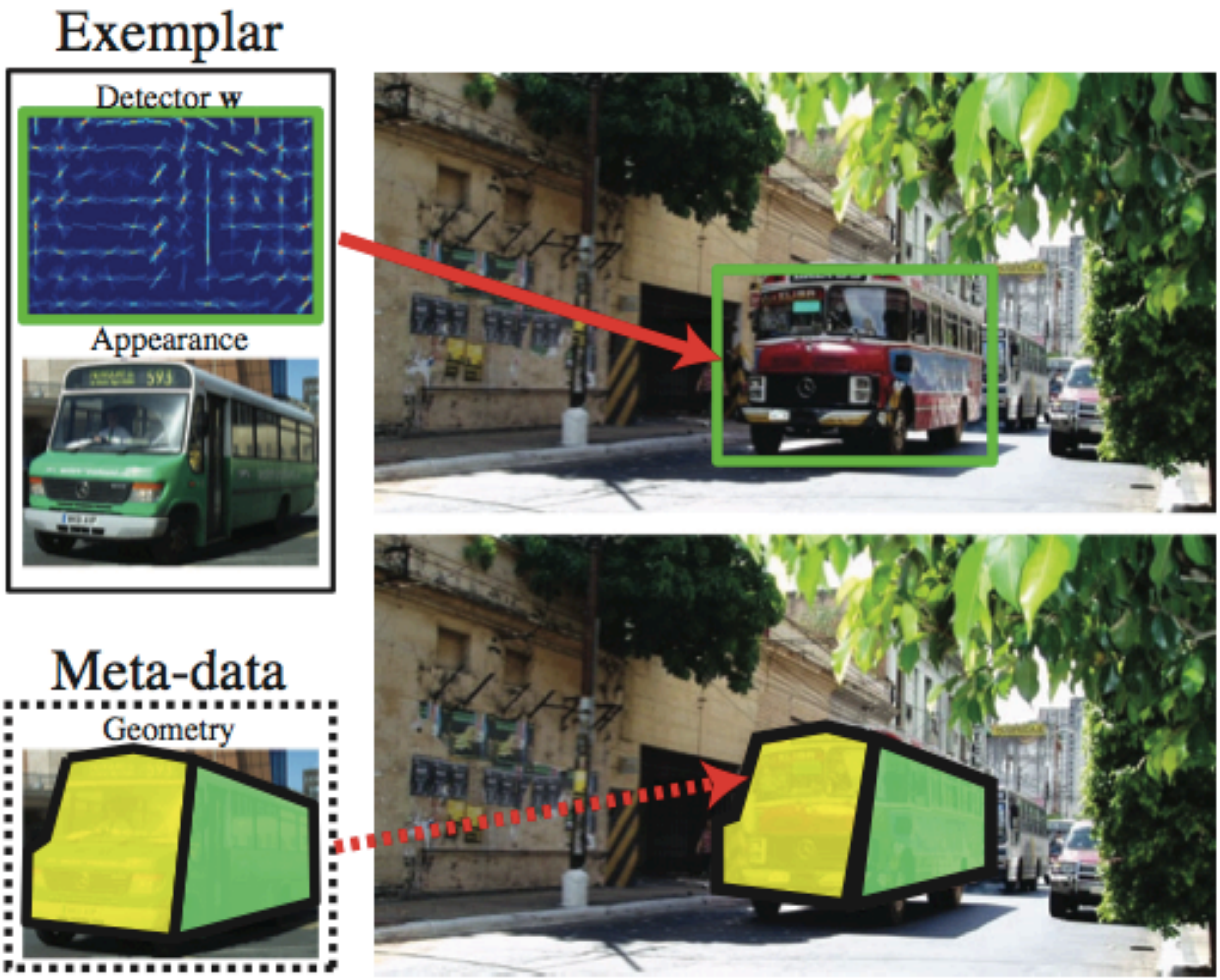
Task 1: Geometry Transfer


Exemplar
Detector w
Appearance


Exemplar
Detector w
Appearance
Meta-data
Geometry



Exemplar
Detector w
Appearance
Meta-data
Geometry



Exemplar
Detector w
Appearance
Meta-data
Geometry


Exemplar
Detector w
Appearance
Meta-data
Geometry

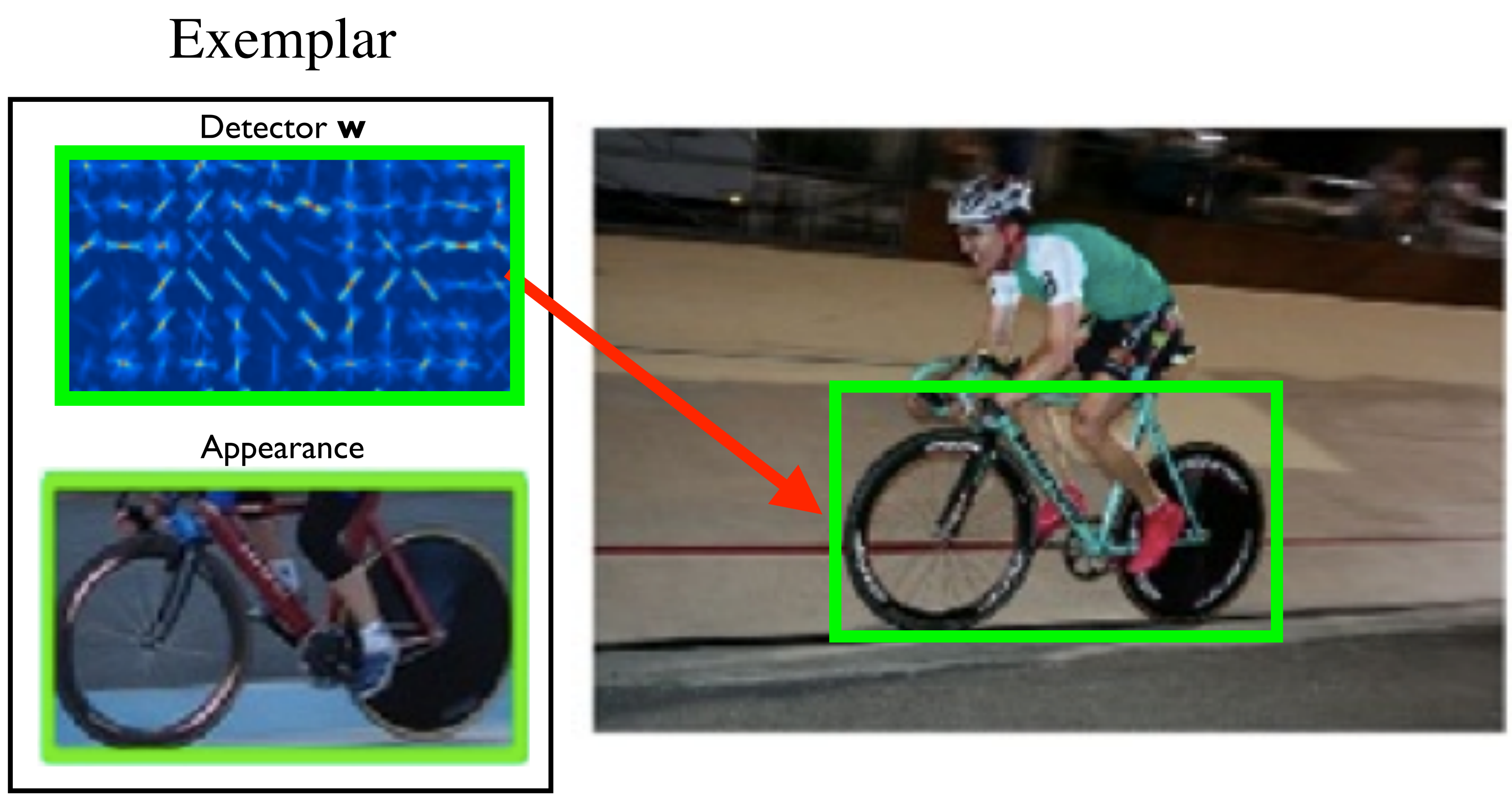
Task 1: Evaluation on Buses

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

- 43.0% Hoiem et al. 2005
- 51.0% Category-SVM* + NN
- **62.3**% Exemplar-SVMs

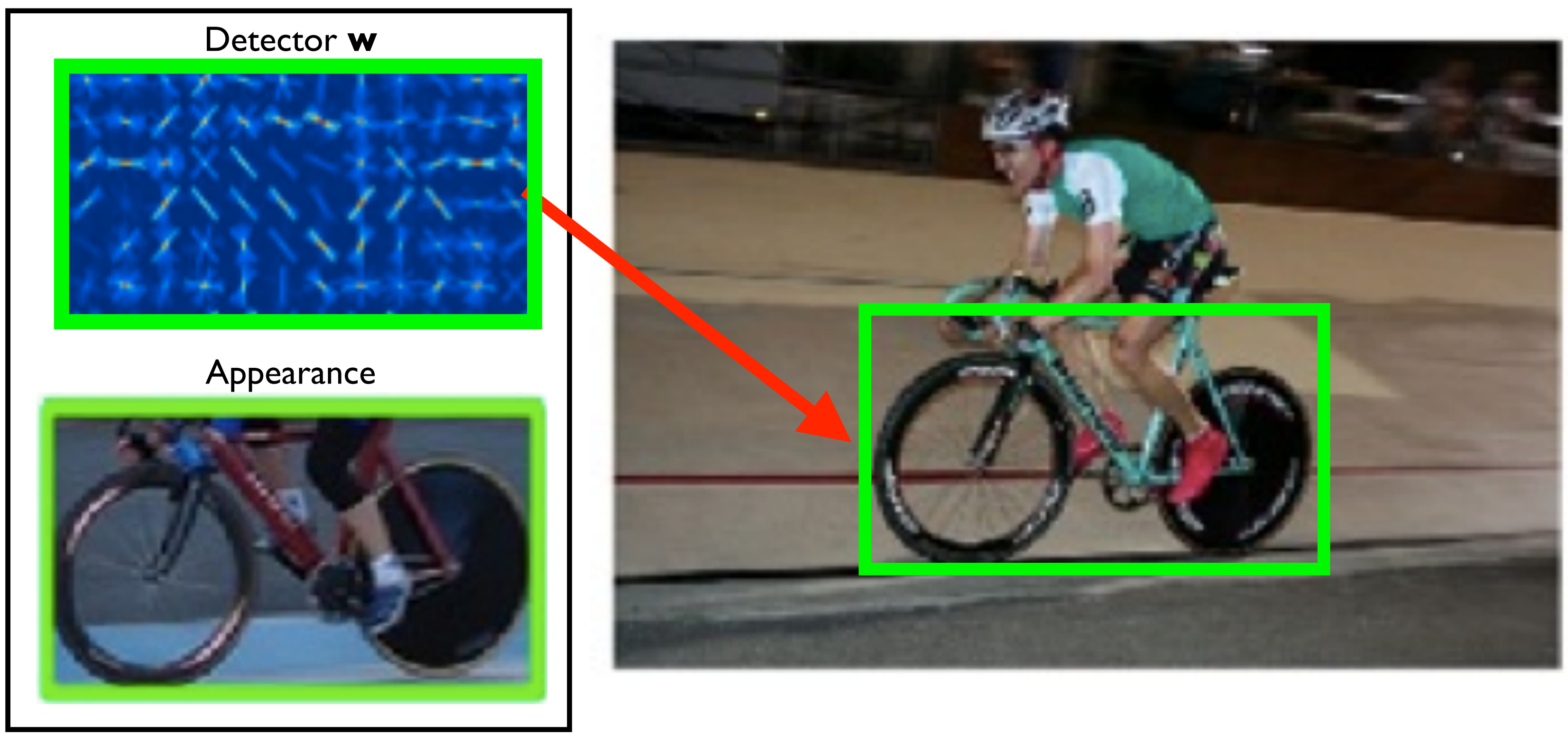
*Felzenszwalb et al. 2010

Task II: Person Prediction

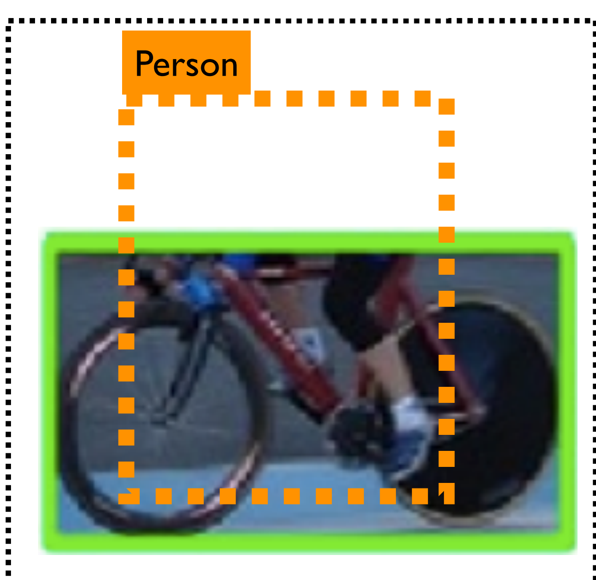


Task II: Person Prediction

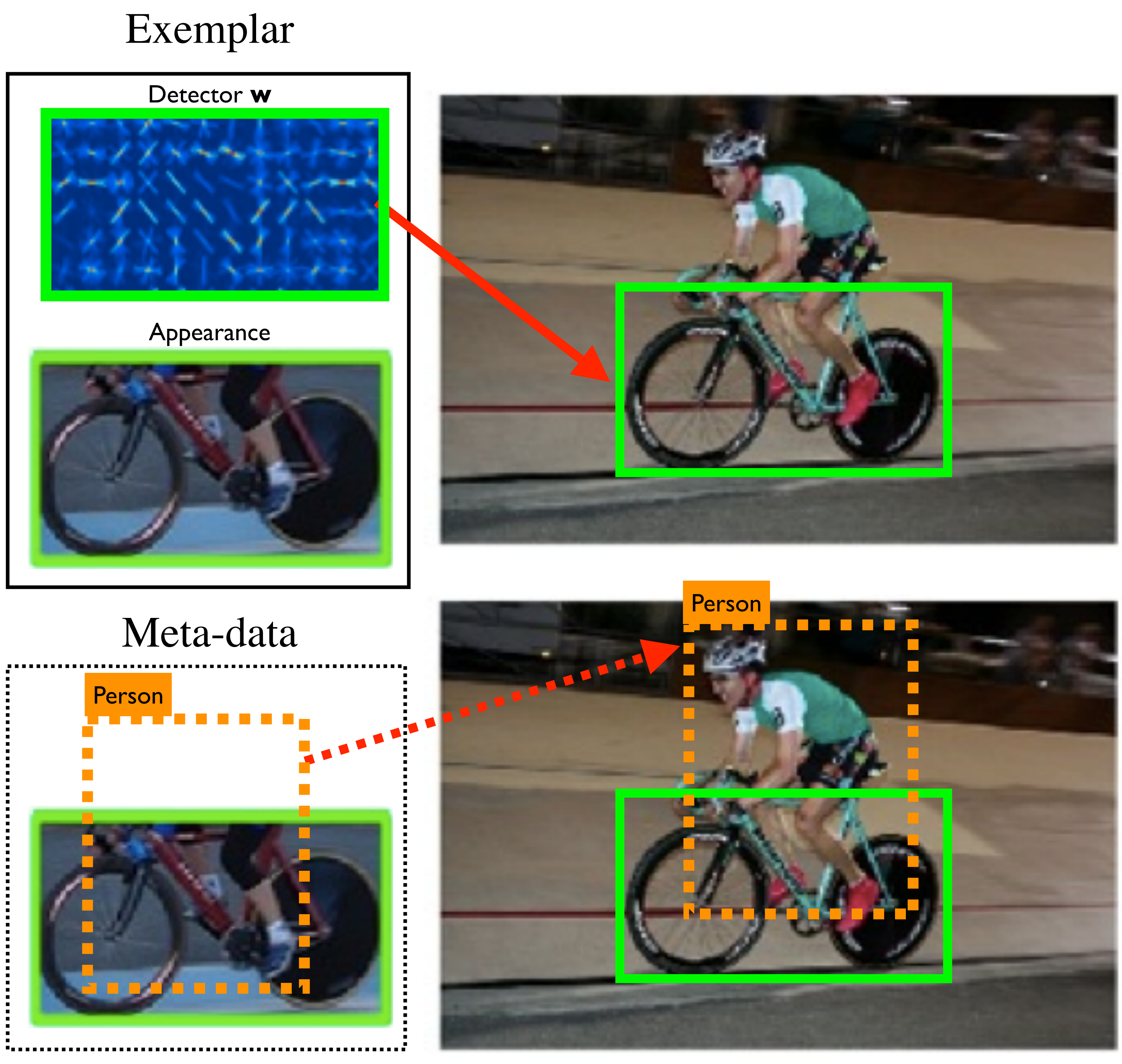
Exemplar

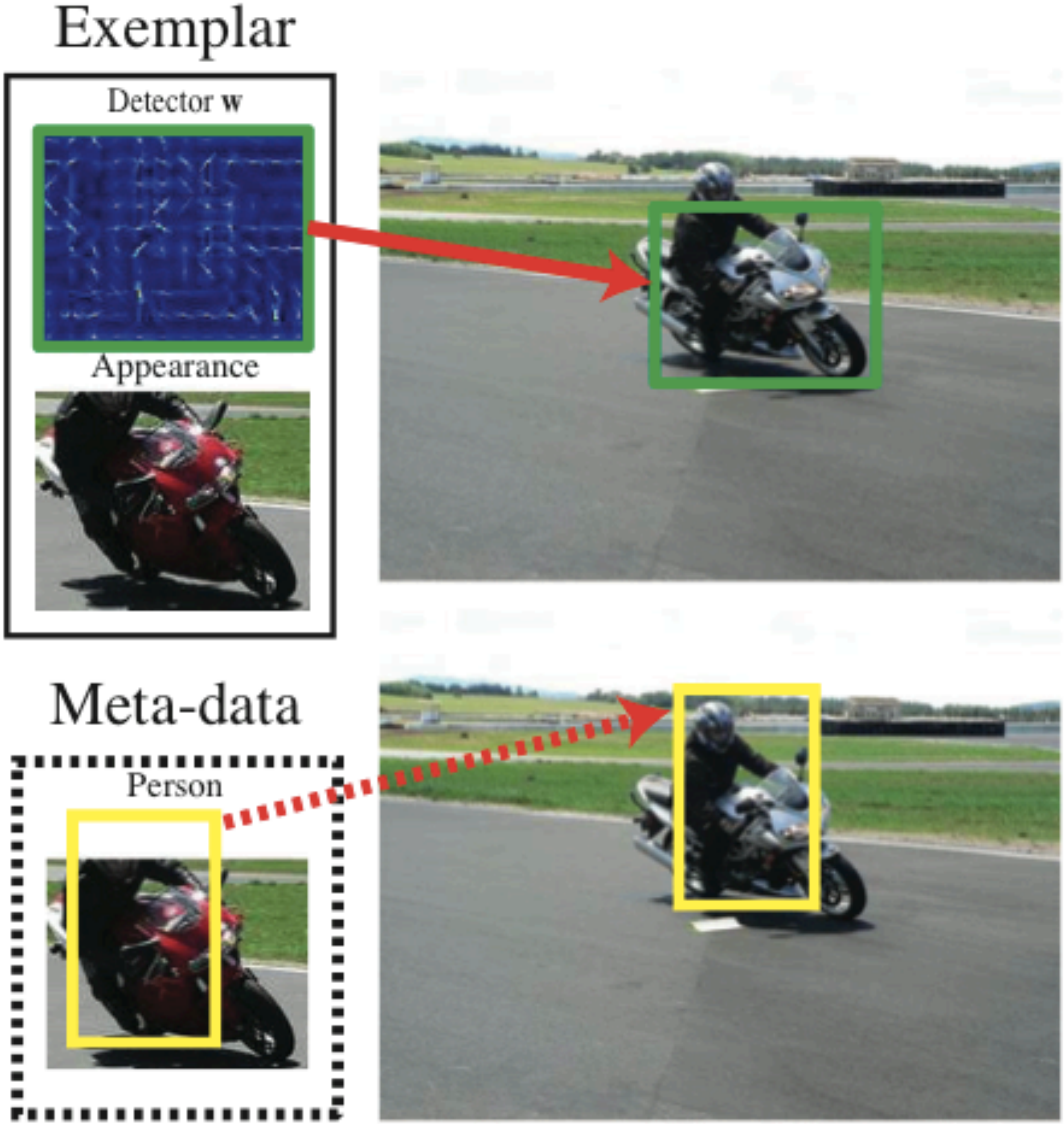


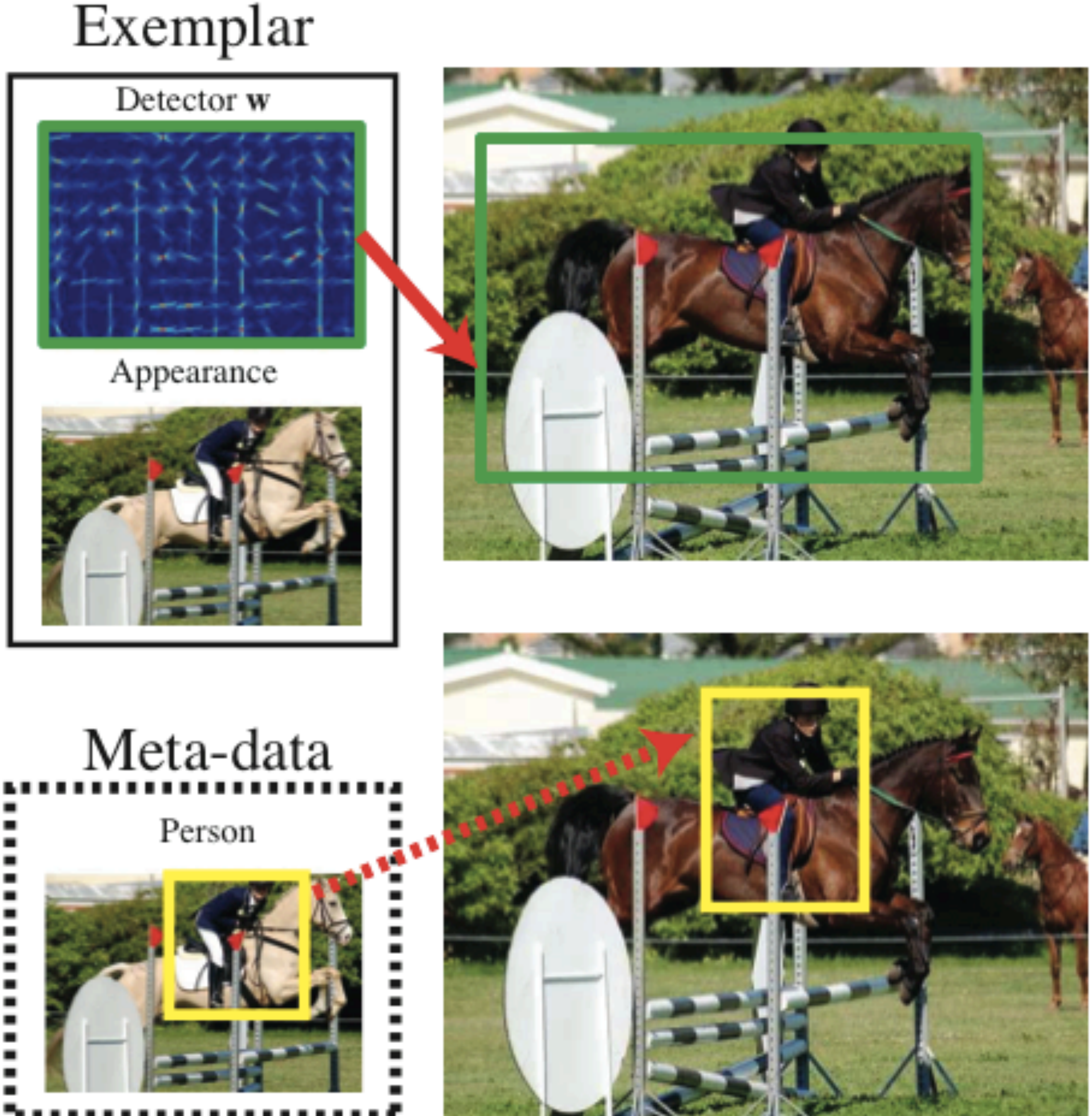
Meta-data



Task II: Person Prediction




Exemplar
Detector w
Appearance
Meta-data
Person



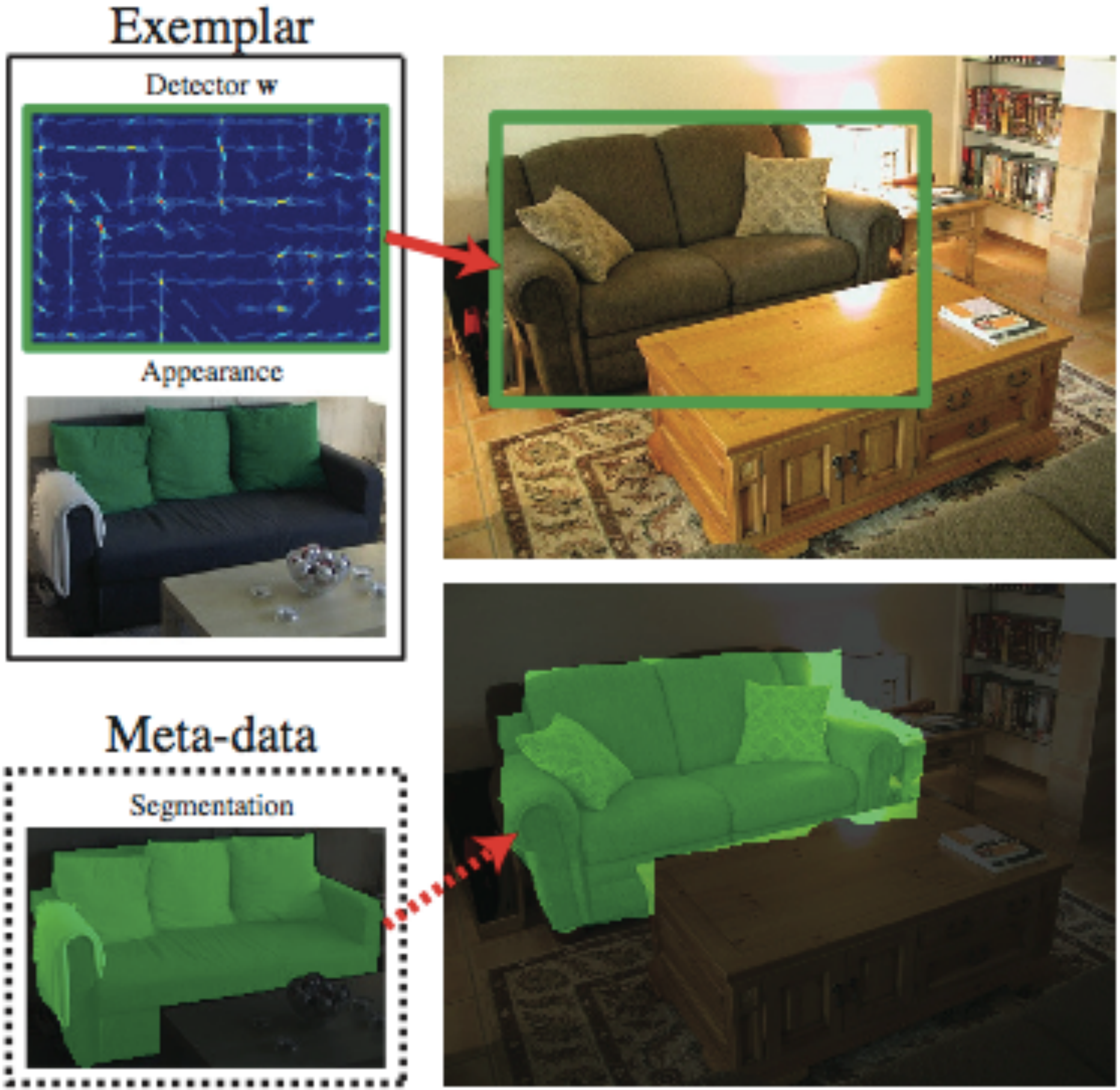
Exemplar
Detector w
Appearance
Meta-data
Person

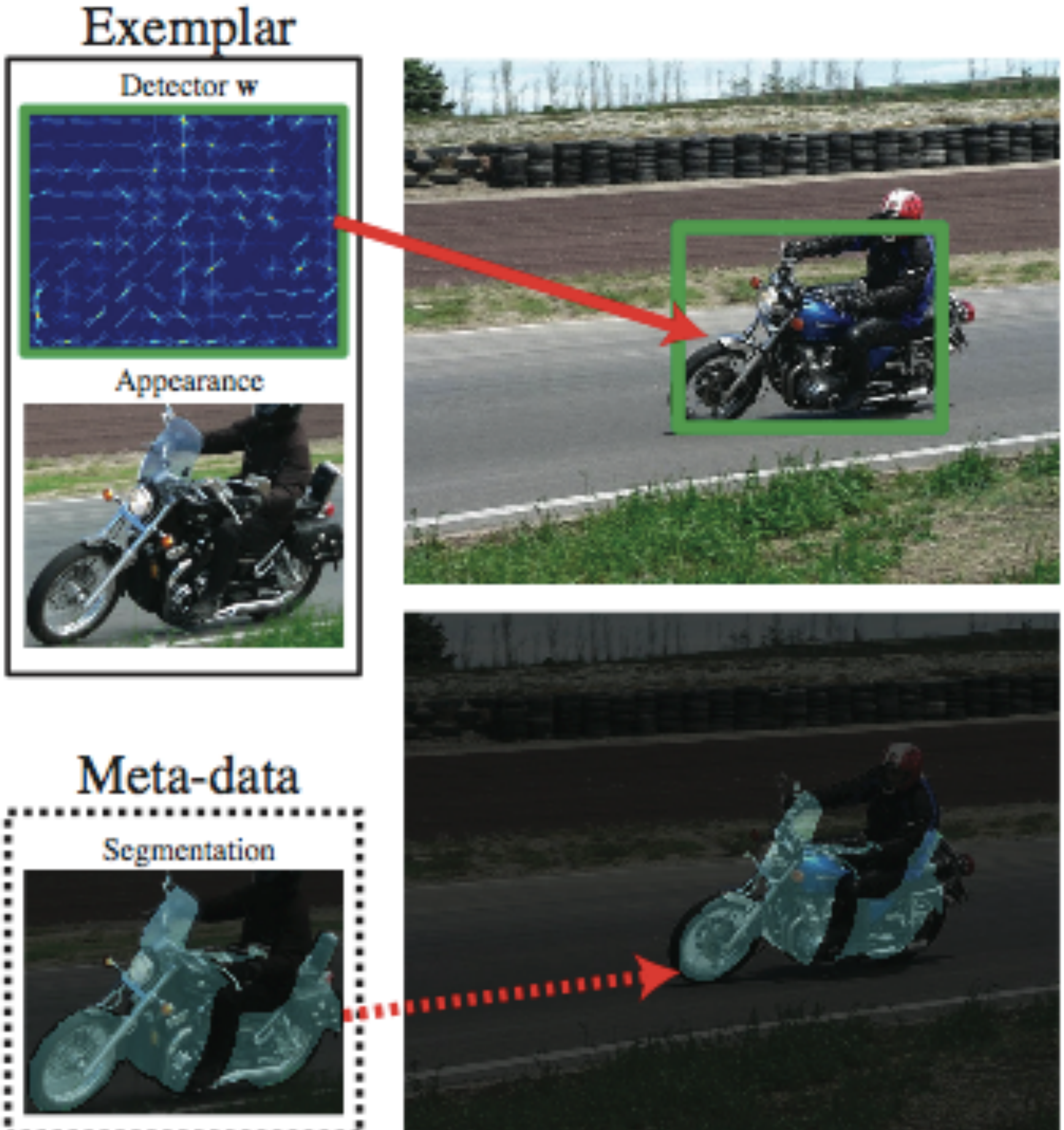
Task II: Evaluation

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.

More Transfer Examples

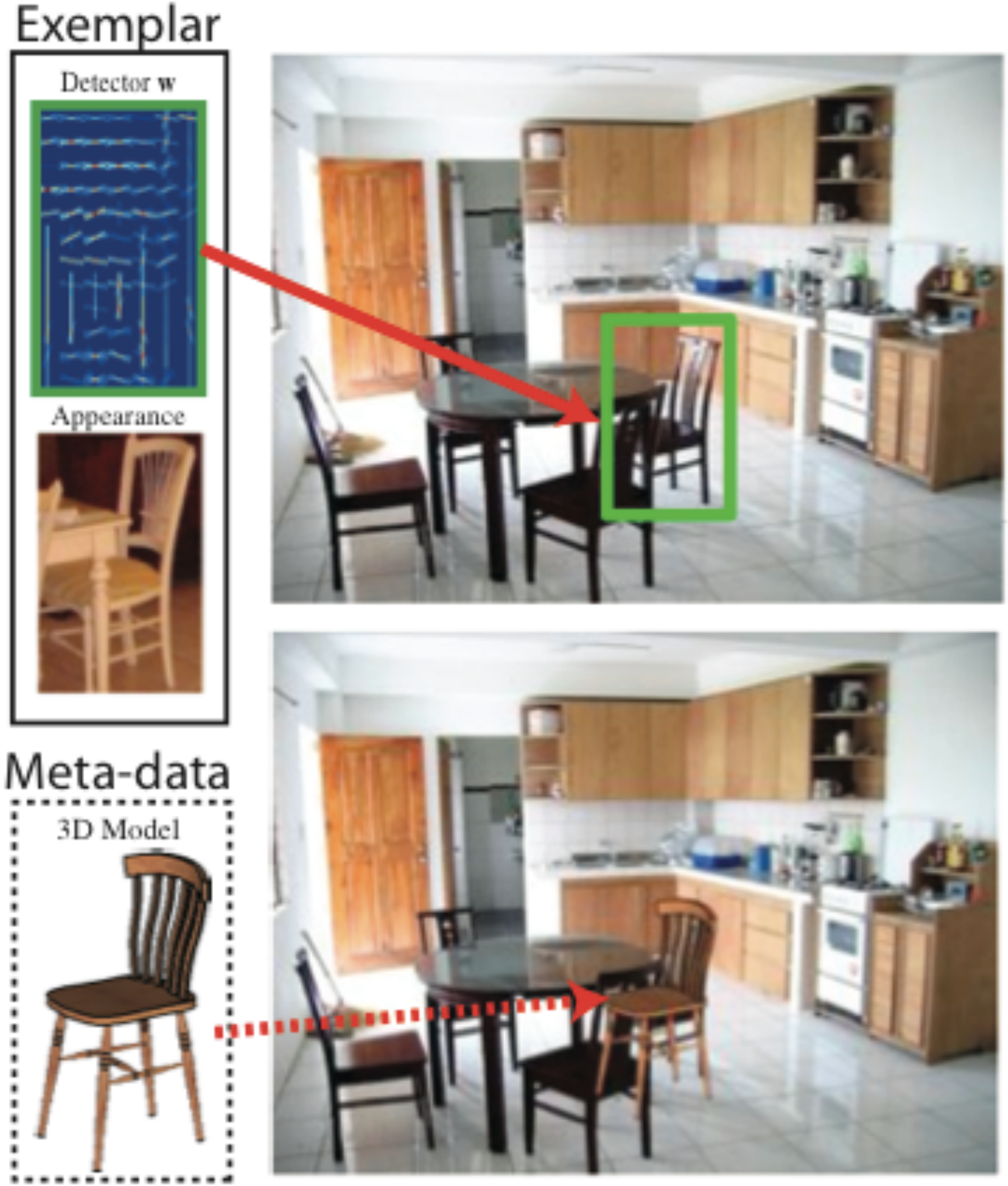

Exemplar
Detector w
Appearance
Meta-data
Segmentation


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

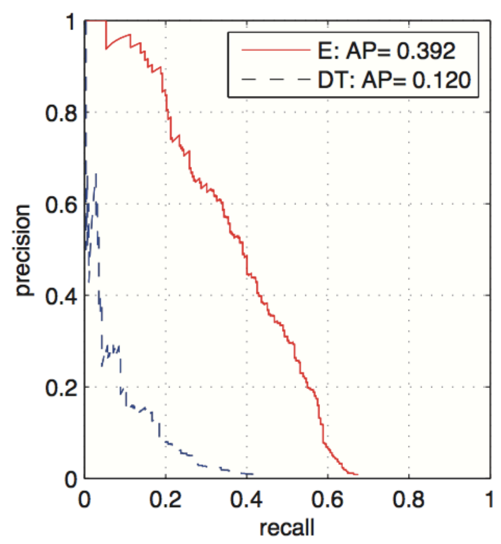

Exemplar
Detector w
Appearance
Meta-data
3D Model

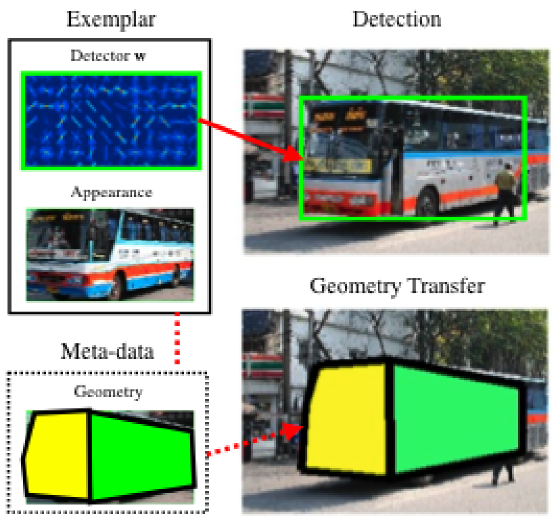

Exemplar
Detector w
Appearance
Meta-data
3D Model

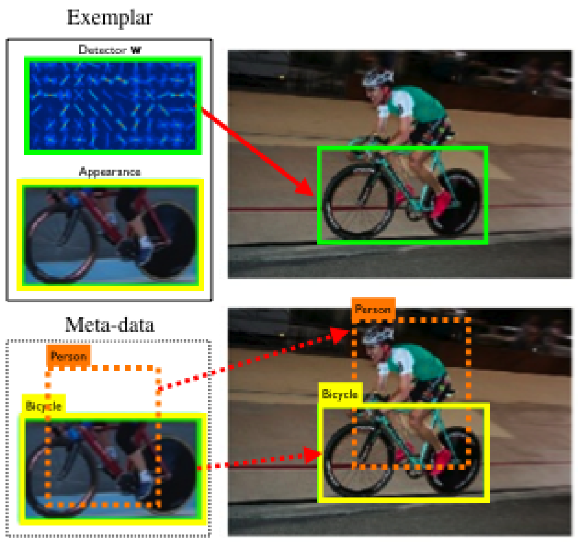
Conclusion

Conclusion

- ExemplarSVMs can be used for recognition, label transfer, and complementary object prediction

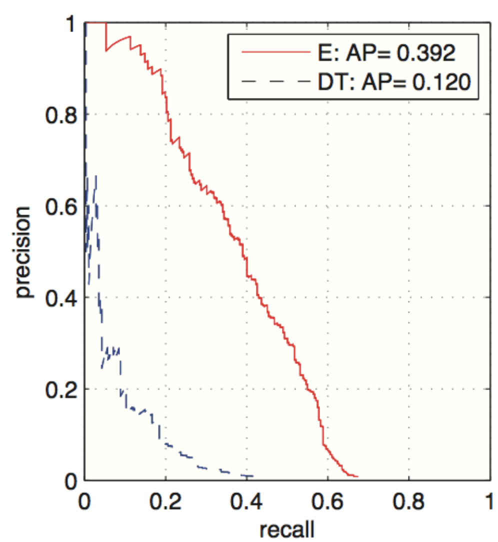


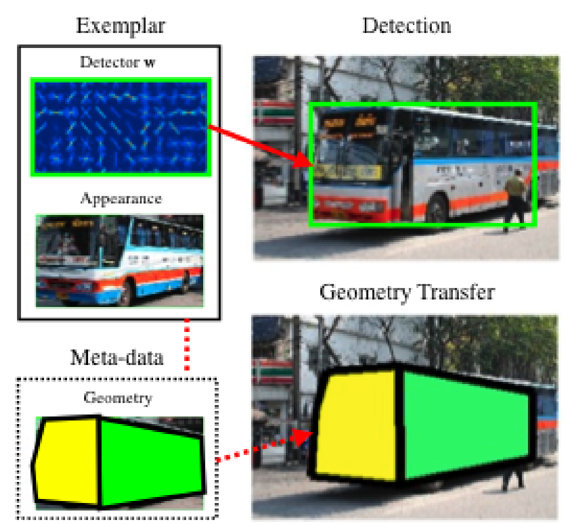


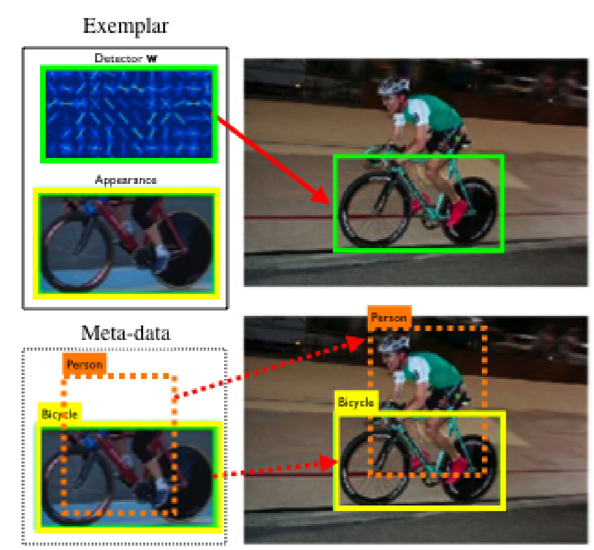


Conclusion

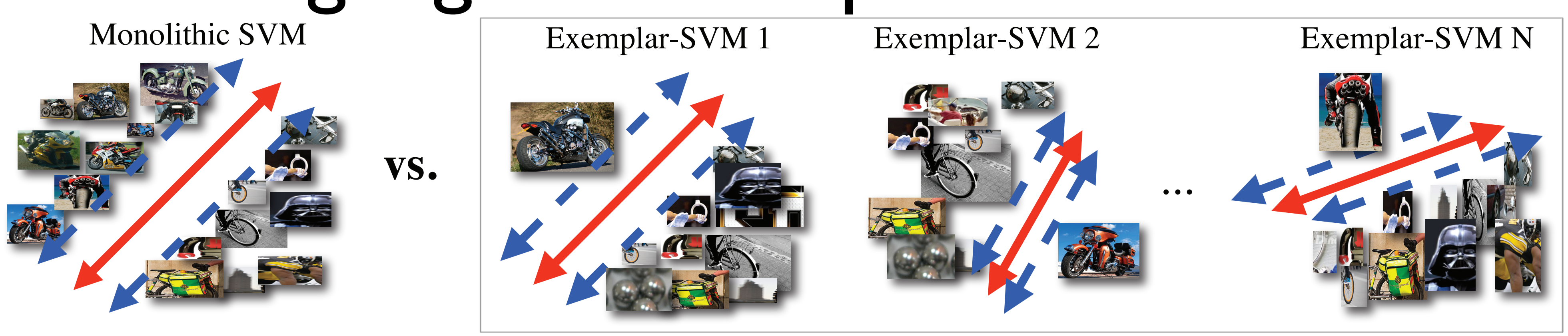
- ExemplarSVMs can be used for recognition, label transfer, and complementary object prediction







- Large-scale negative mining is the **key** to learning a good ExemplarSVM



Thank You

Questions?