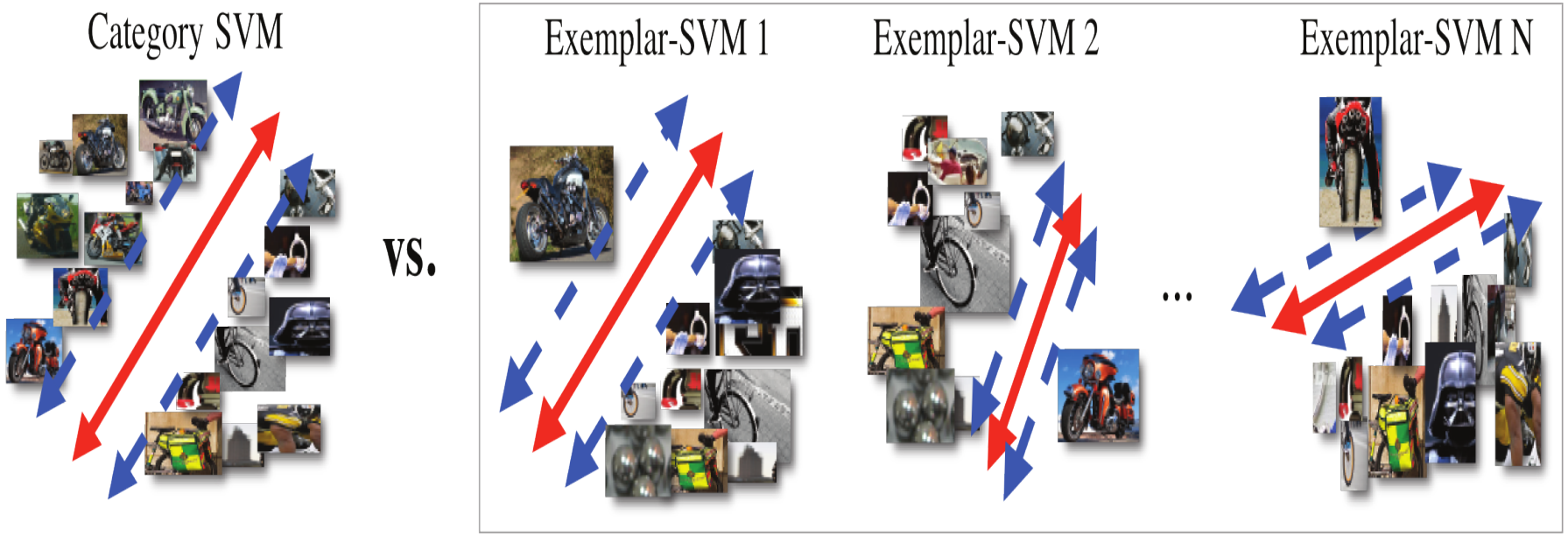
Ensemble of Exemplar-SVMs for Object Detection and Beyond

T. Malisiewicz, A. Gupta, A. Efros,
ICCV 2011.

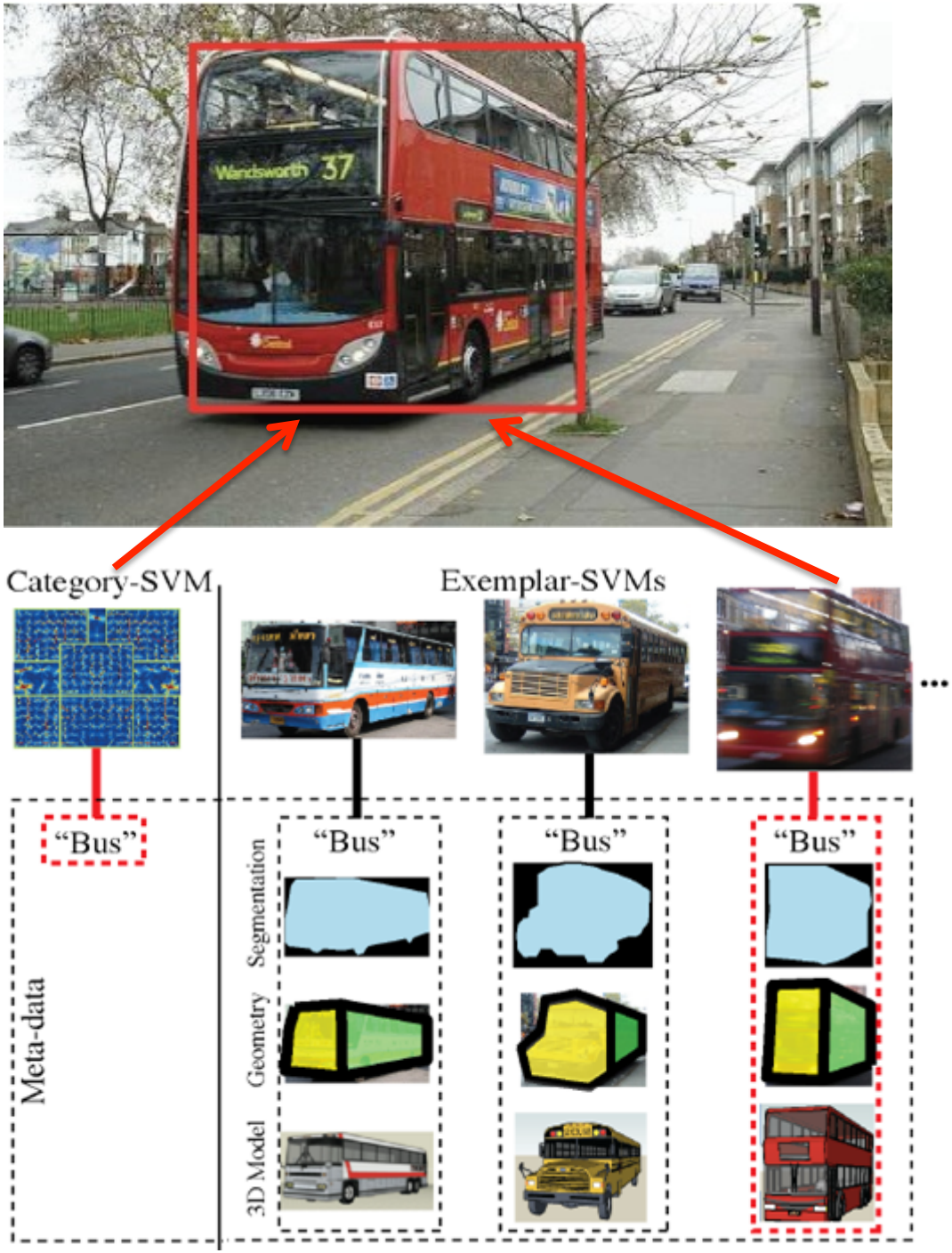
Abstract

- Motivation & Related Work
- Exemplar-SVMs
- Object Detection Results
- Beyond Detection

Ensemble of Exemplar-SVMs



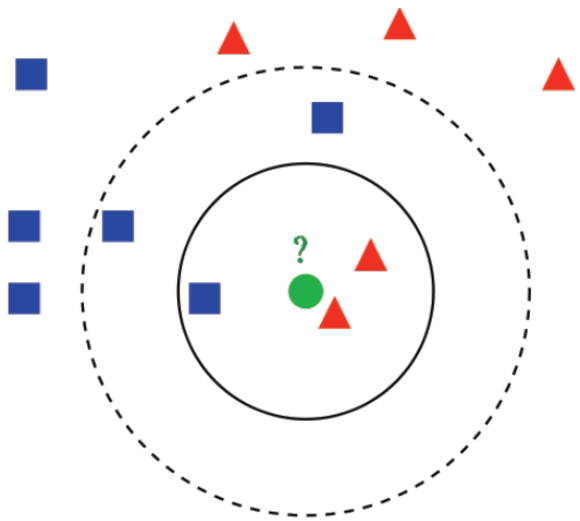
Motivation


Wandsworth 37
Category-SVM
Exemplar-SVMs
...
"Bus"
"Bus"
"Bus"
"Bus"
Meta-data
Segmentation
Geometry
3D Model

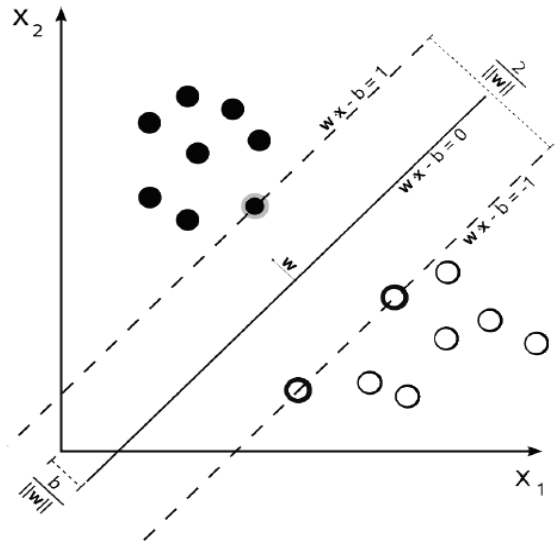
Motivation

- Idea of associating a new instance with something seen in the past
  - Exemplar theory in cognitive psychology
  - Case based reasoning in AI
  - İnstance based methods in ML

Motivation -
Exemplar Reasoning: Non-parametric



KNN: non-parametric



SVM: parametric

Exemplar Theory in CV

- Object Alignment
- Scene Recognition
- Image Parsing
- Object Detection(not competitive)
  - Why?

Parametric Approaches

- Can handle negative data well
- No need to explicitly store negative data(vs KNN)
- Positive Data?
  - Implicit assumption that all positivie examples are visually related, which
  - Results in over generalized models


Not that similar!
car
car
car
car
car
car
car
car
car
car
car
car
car
cars
car
car
car
car
car
car
car
car
car
car suv
car
car

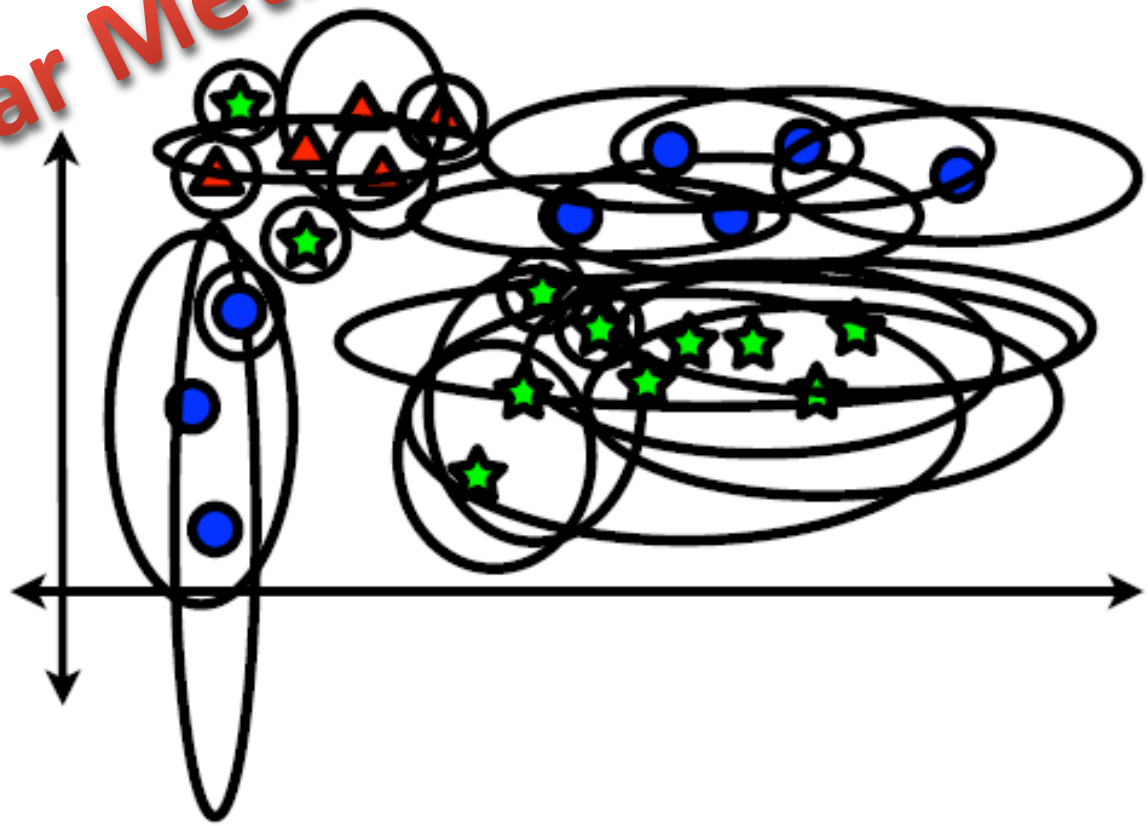
Motivation-Desirable Approach

- Strengths of Dalal/Trigs/Felzenswalb/Ramanan style detector (Hog/DPM)
  - Powerful descriptor
  - Discriminative framework
  - Handle massive amount of negatives
- Not rigidly representing positives
- Good association for meta-data transfer

parametric negatives

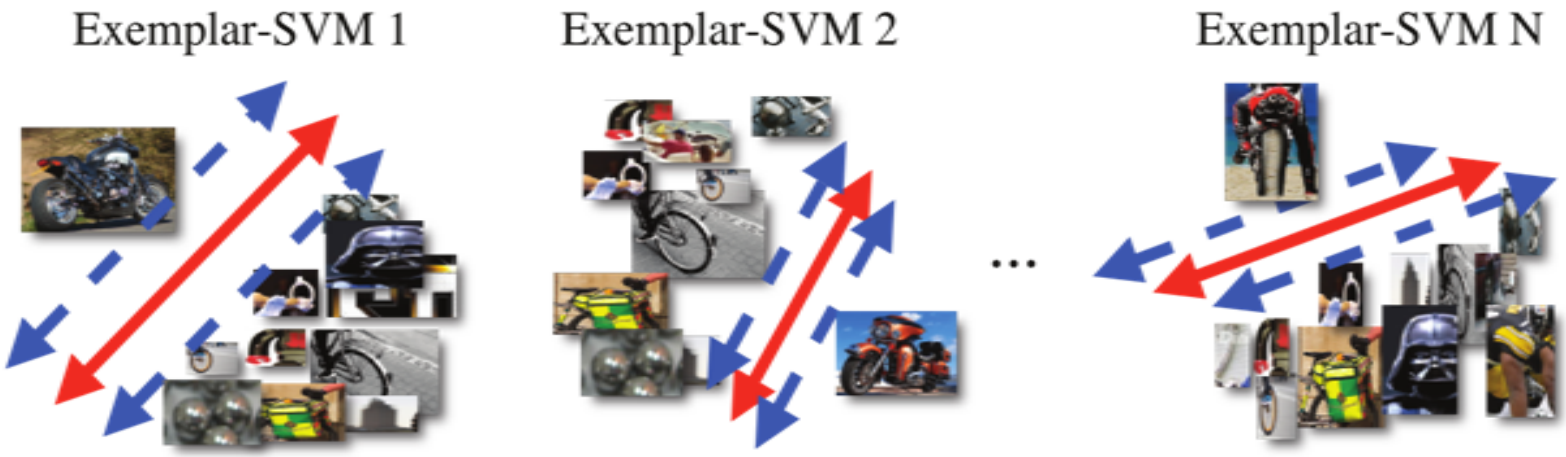
non-parametric positives

Per-Exemplar Methods



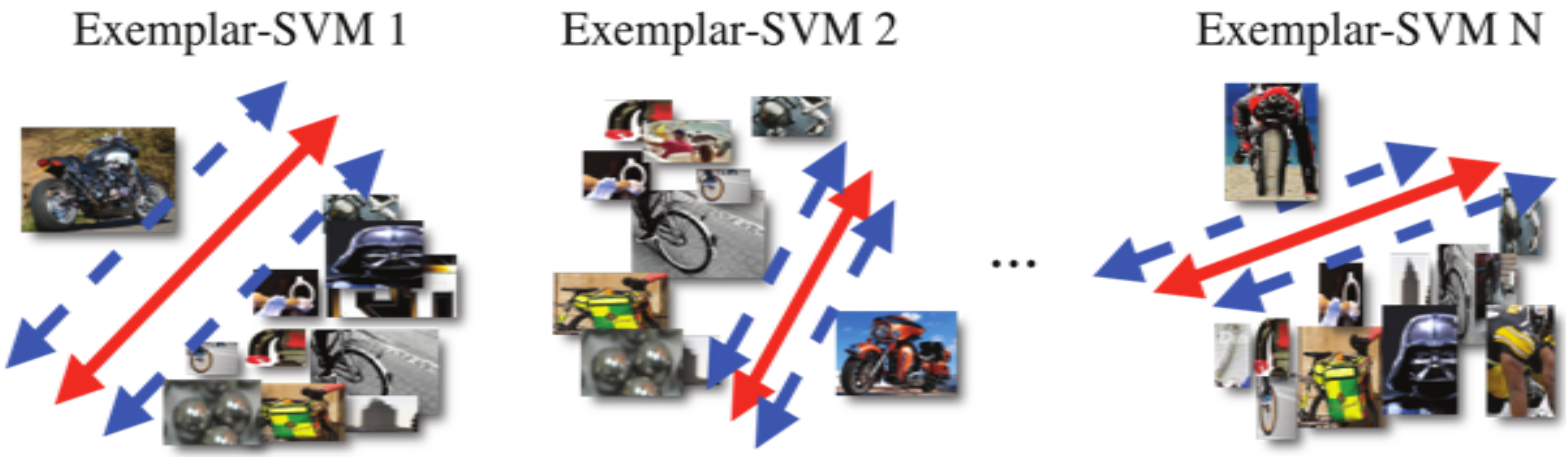
- NN-method: distance similarity function for each exemplar
- Better than a single similarity measure accross all exemplars

Exemplar-SVMs



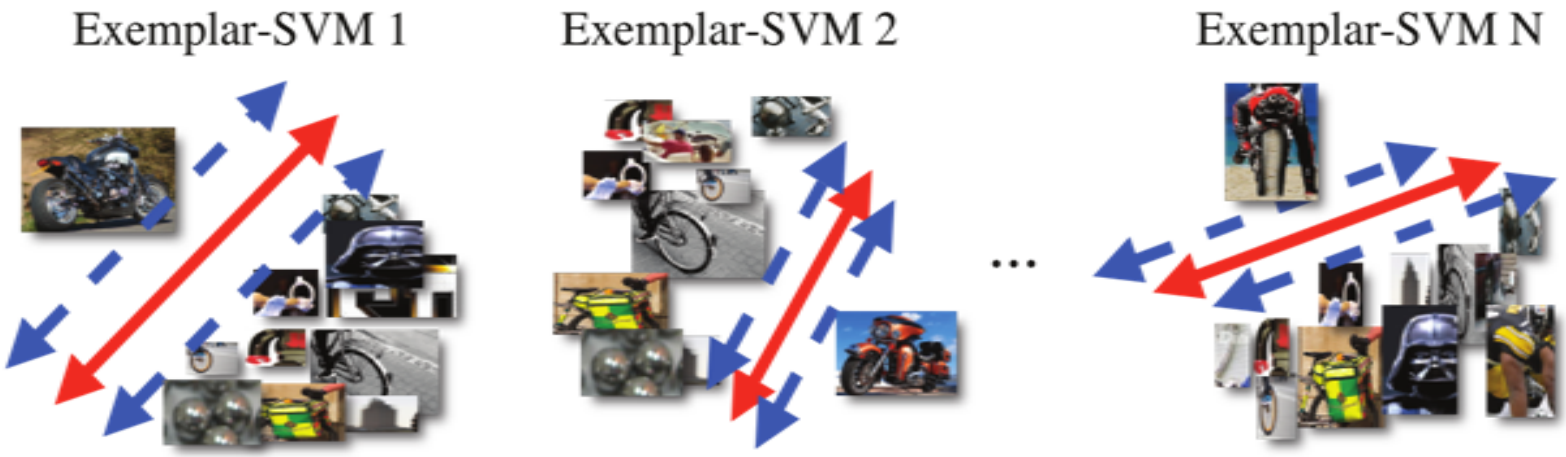
- Effectiveness of discriminatively trained object detectors
- Explicit correspondence of Nearest Neighbor approaches

Exemplar-SVMs



- Learn a model for each positive with Linear SVM
- Each Exemplar-SVM is trained with single positive instance
- Exemplar's decision boundary is defined, by what it's not!

Exemplar-SVMs



- Because each exemplar is defined by a single positive instance, different features can be used for each exemplar
- Adapt features to each exemplar's aspect ratio

7x4 HOG

4x8 HOG

Exemplar-SVMs

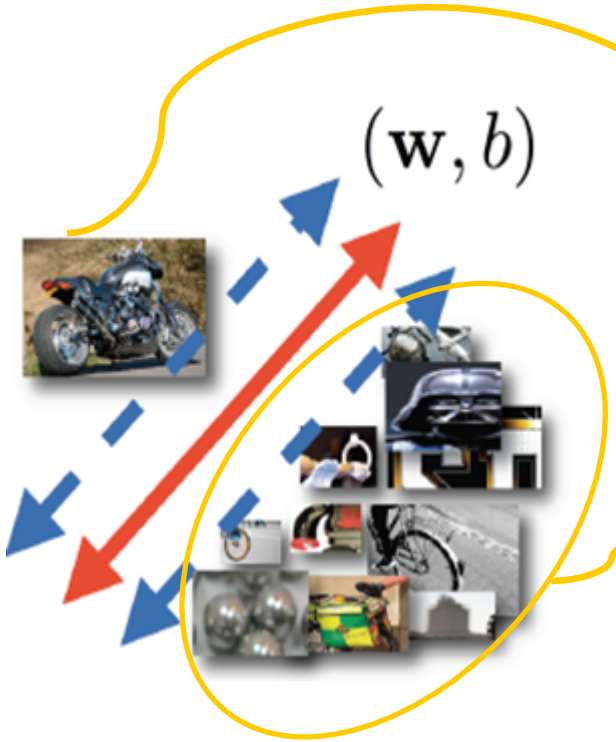
Exemplar E's Objective function:

C1 = 0.5

C2 = 0.1

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

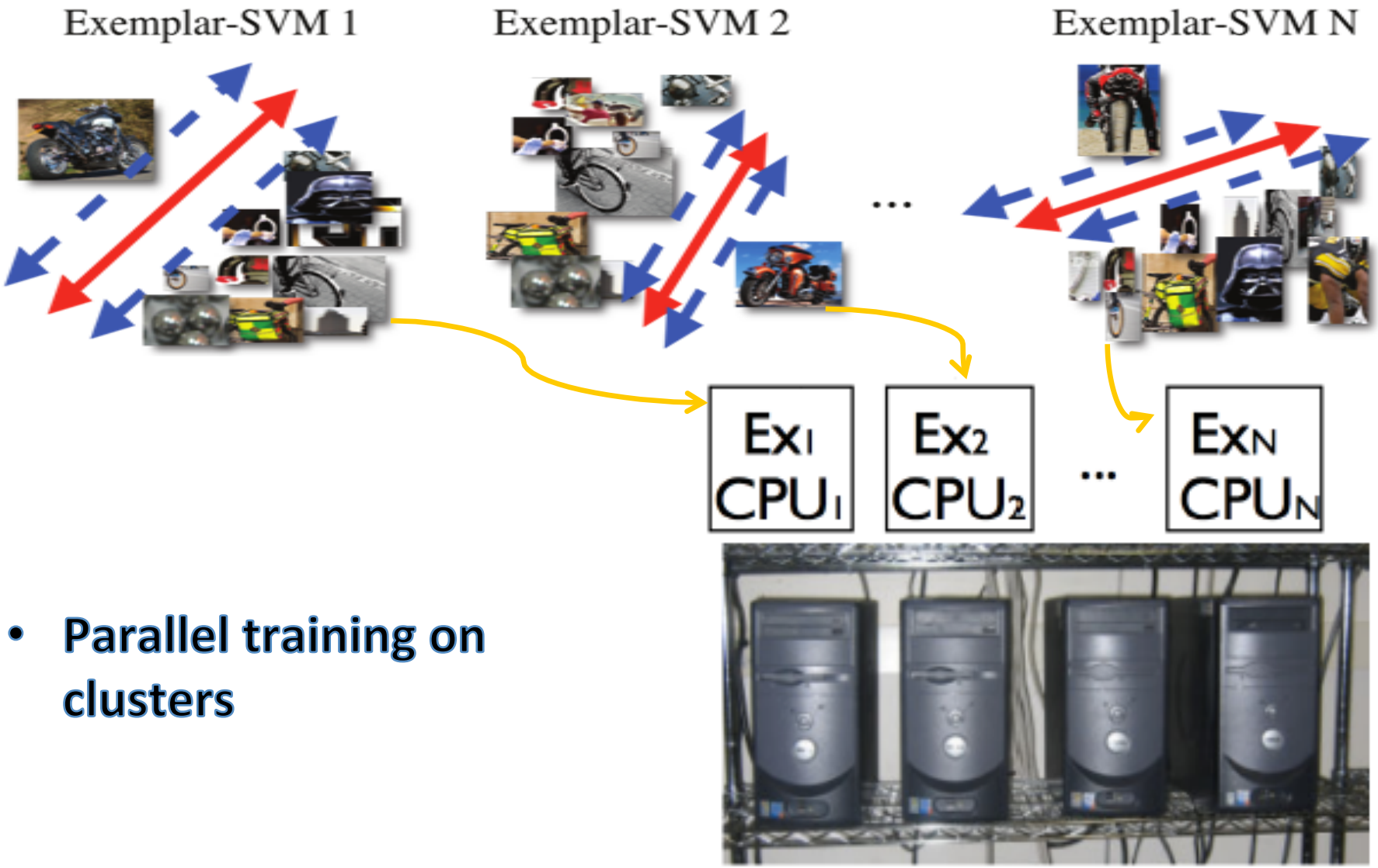
h(x) = max(1-x,0) "hinge-loss"



$\mathbf{x}_E$ **Exemplar represented by ~100 HOG cells**

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

Large-Scale Training



- **Parallel training on clusters**

Exemlar-SVM Calibration

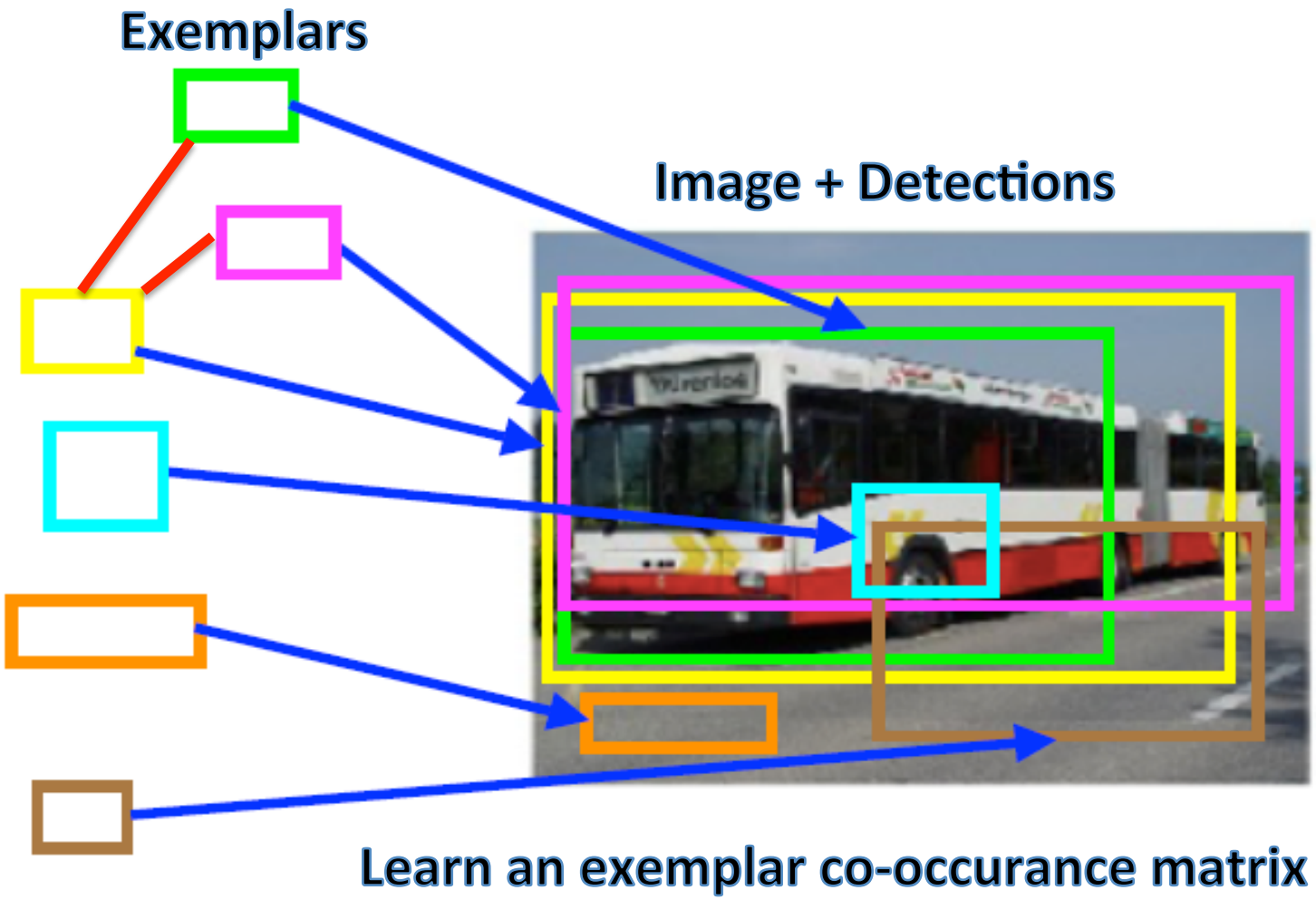
Svm After Training

- Apply exemplar-SVM to held out negative images and all positive images
- Fit sigmoid to responses [Plat 1999]

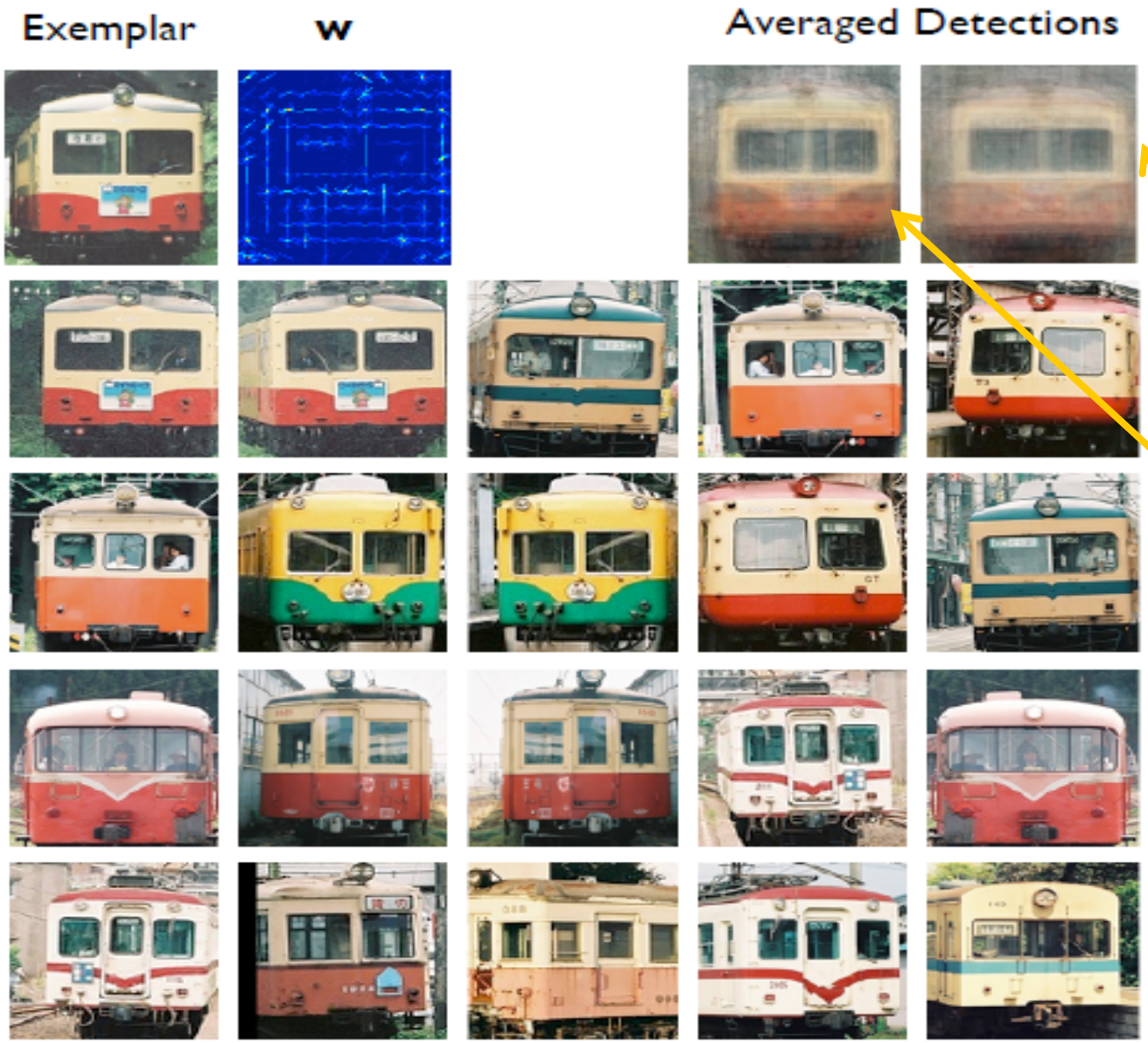
Svm After Calibration

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

Ensemble of Exemlar-SVMs

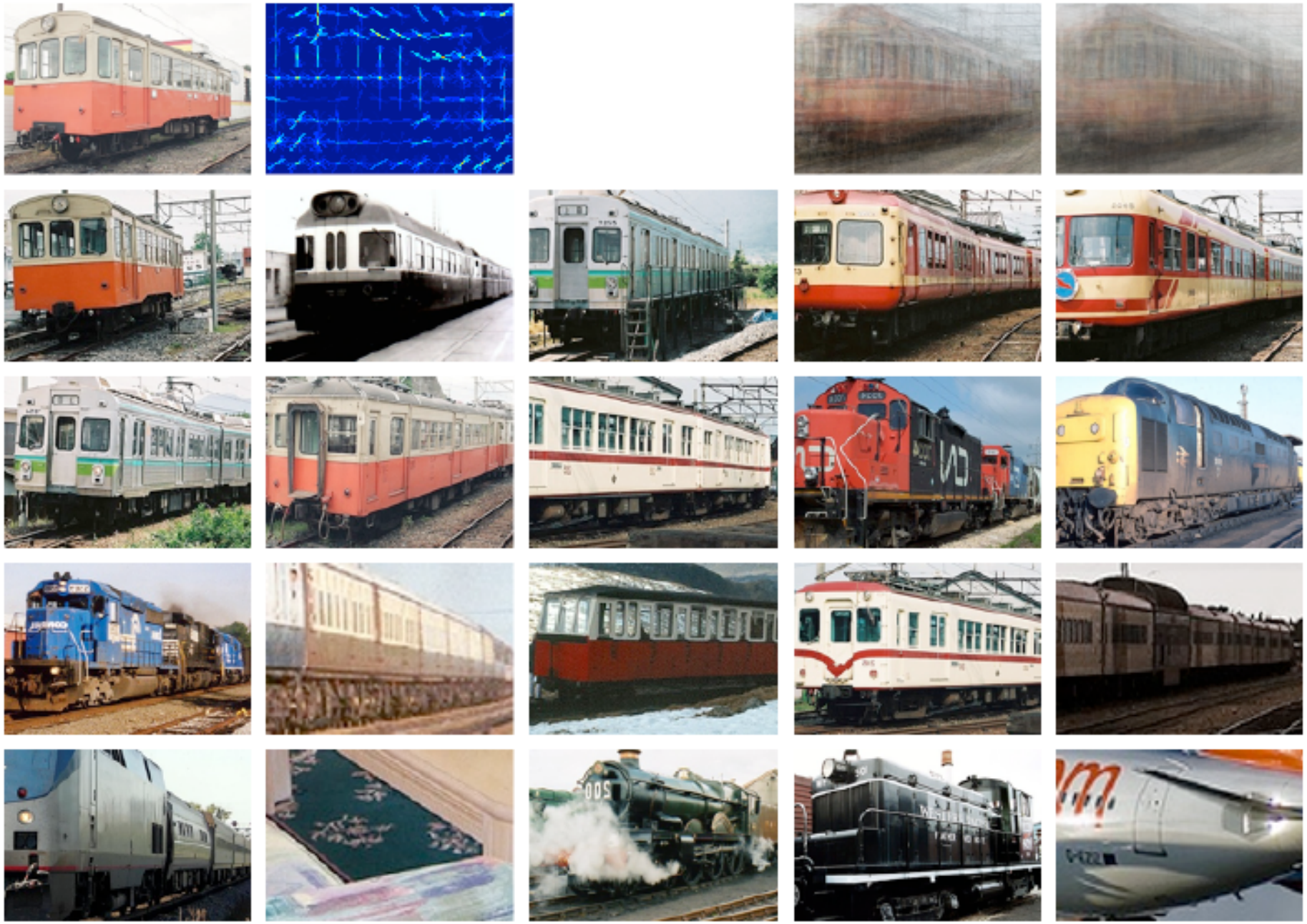


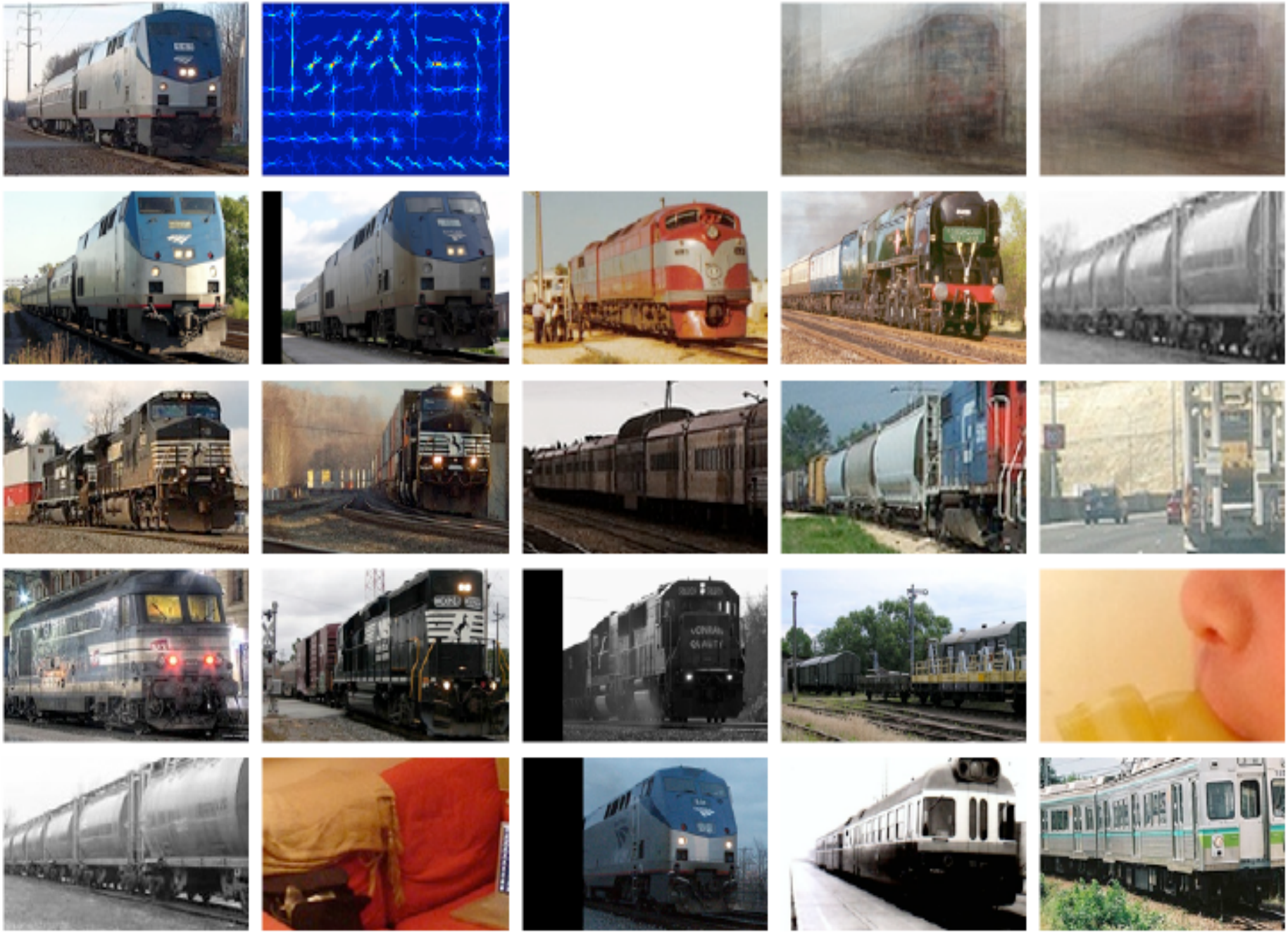
Qualitative Results



Average of first 20 detections

Average of first 10 detections






NREX
BNSF
8008
CN
0808

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

Quantitative Results

- Pascal VOC 2007 dataset
- A standard computer vision object detection benchmark
- 20 object categories
- Machine performance is far below human

Pascal VOC: 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

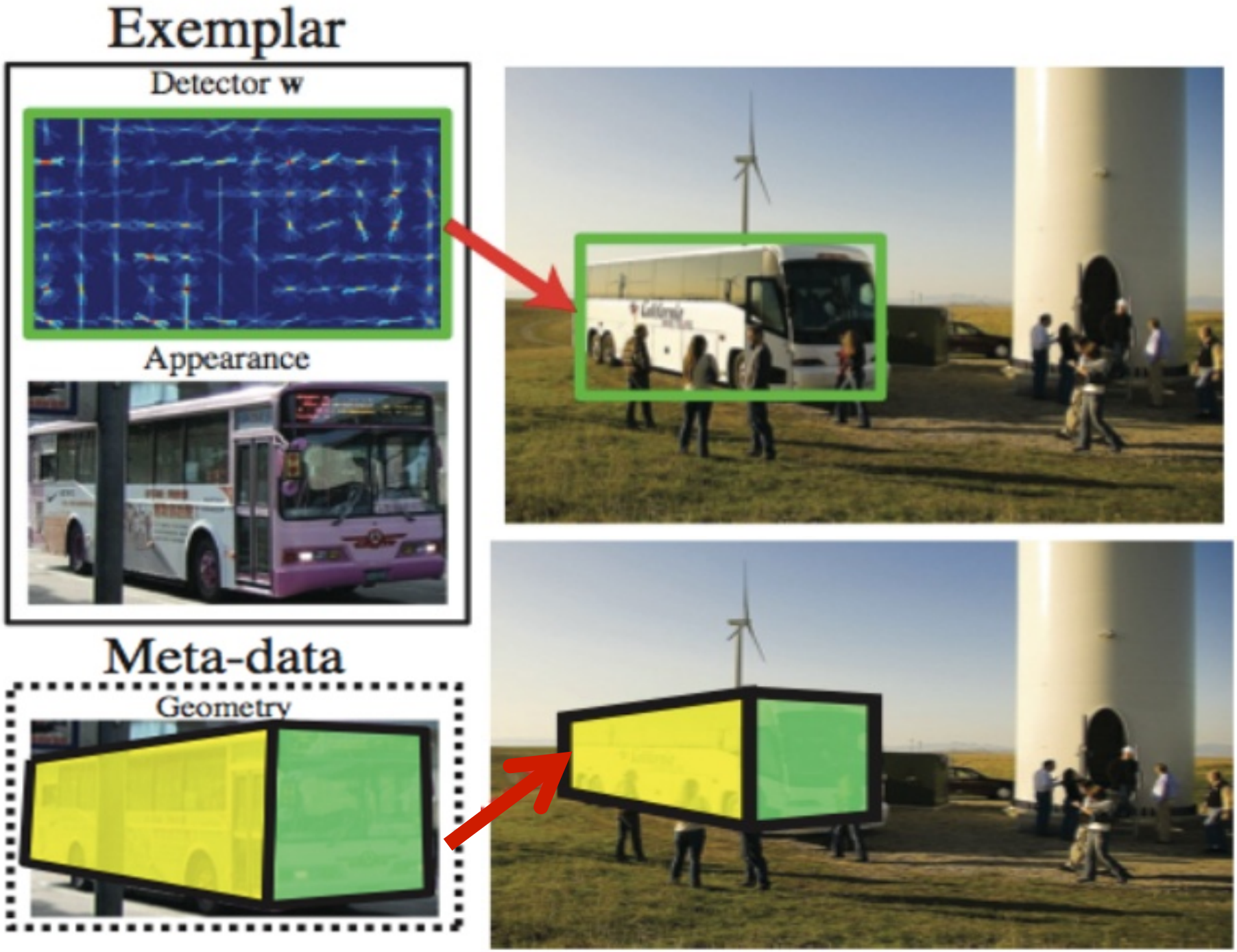
mAP on PASCAL VOC 2007 detection task

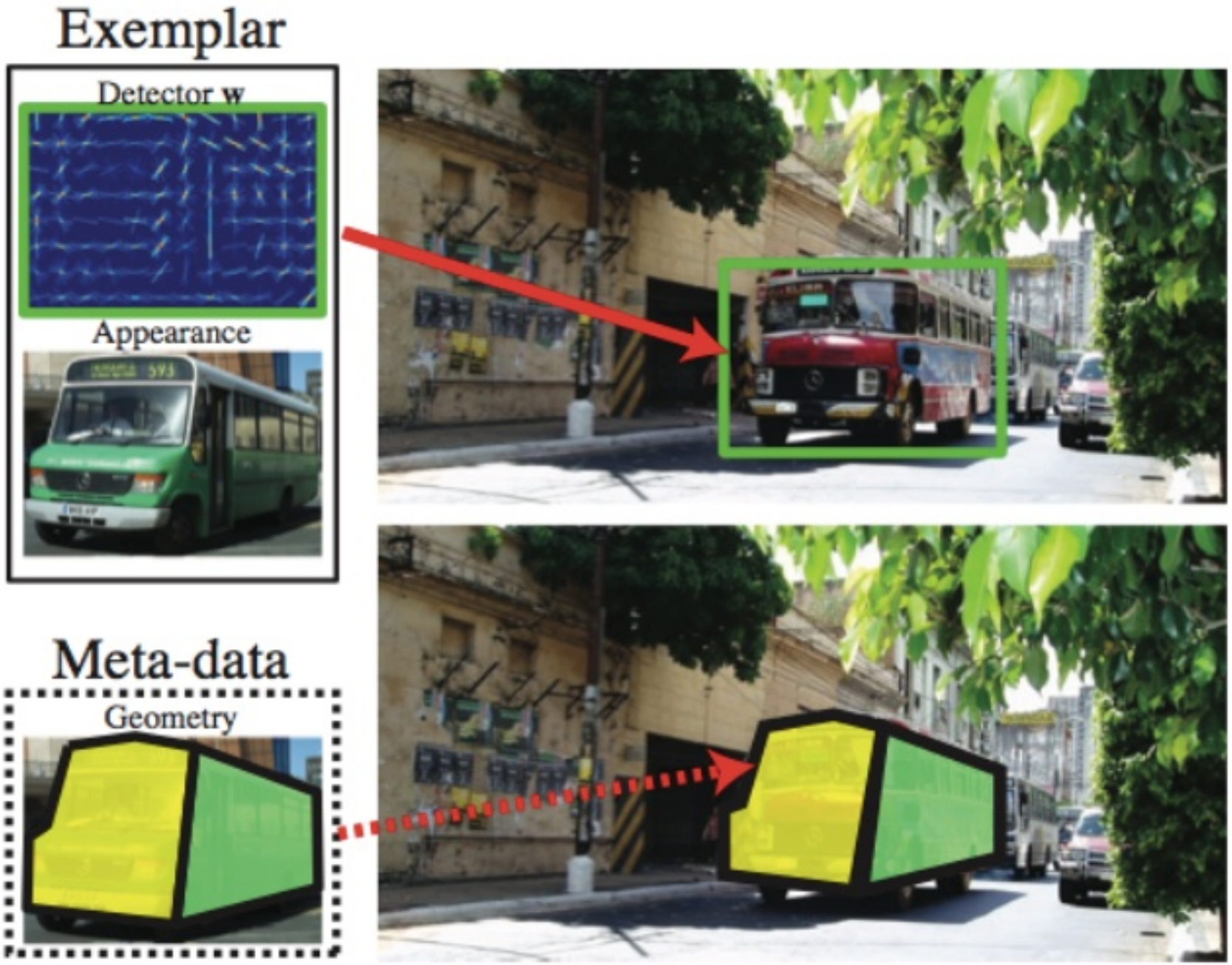
NN + Cal	0.110
DFUN + Cal	0.155
Exemplar-SVMs + Cal	0.198
Exemplar-SVMs + Co-occ	0.227
DT*	0.097
LDPM**	0.266

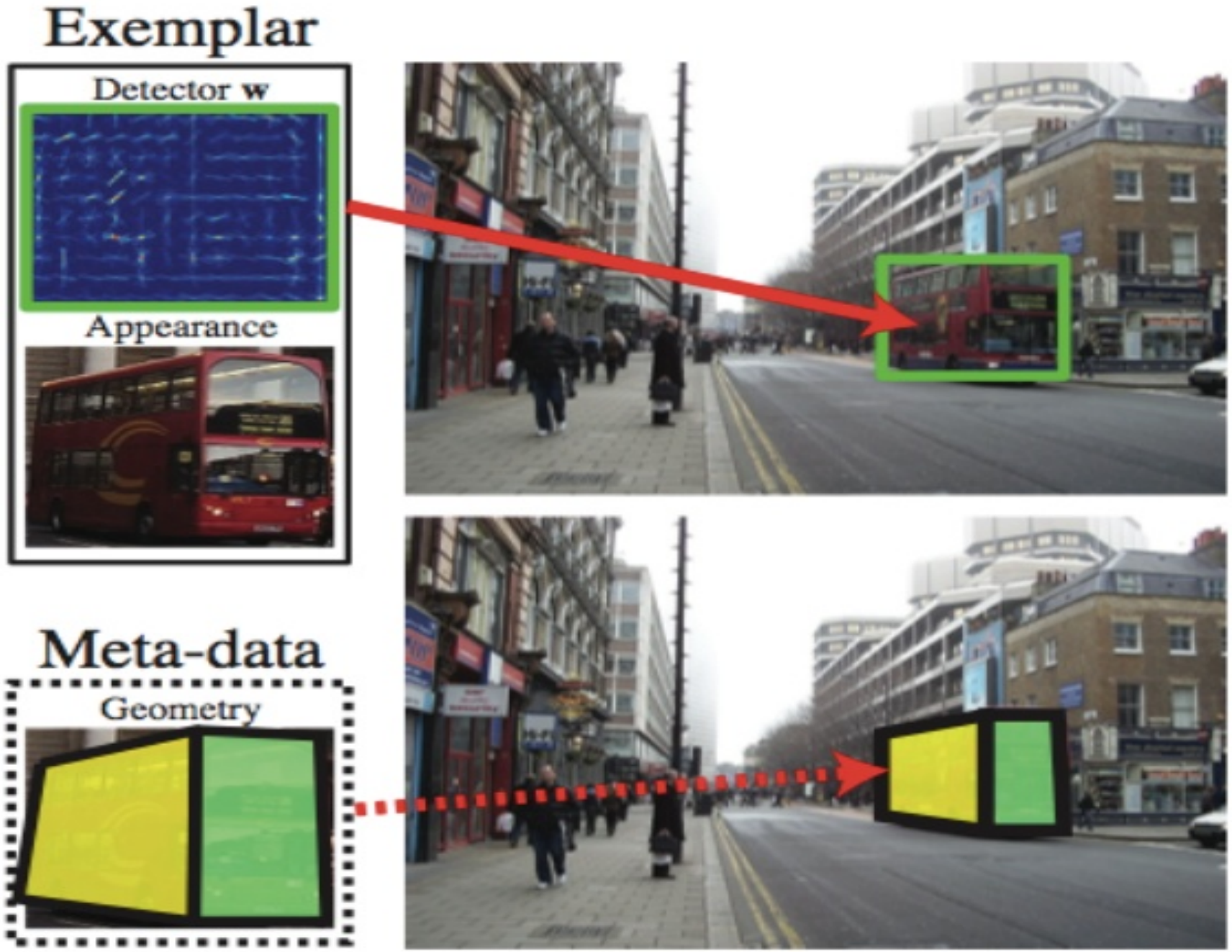
*Dalal et al. 2005 **Felzenszwalb et al. 2010

Beyond Object Detection

- Geometry Transfer
- 3D Model Transfer
- Relating Object Priming


Exemplar
Detector w
Appearance
Meta-data
Geometry


Exemplar
Detector w
Appearance
Meta-data
Geometry


Exemplar
Detector w
Appearance
Meta-data
Geometry

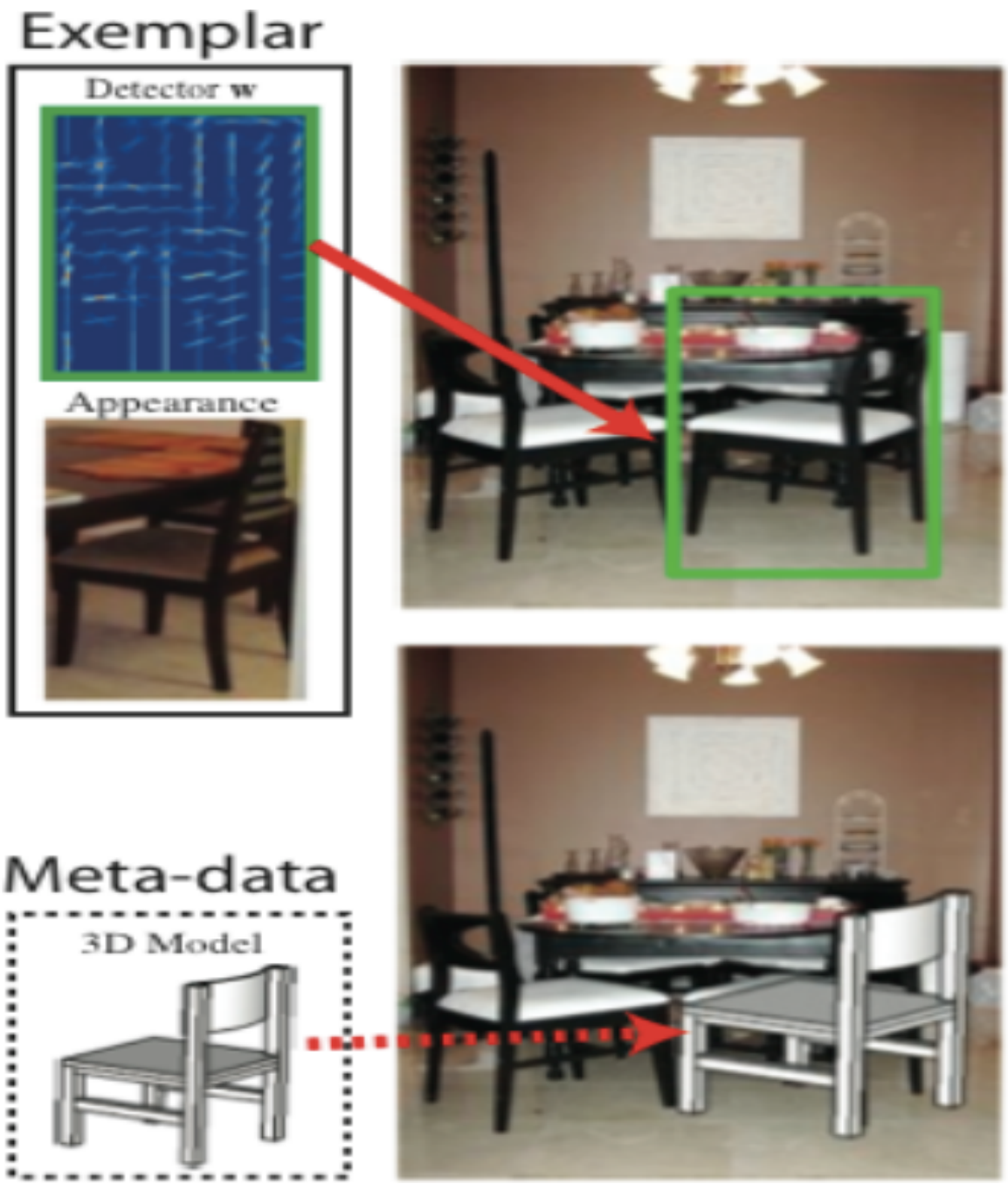
Geometry Transfer: Evaluation on Buses

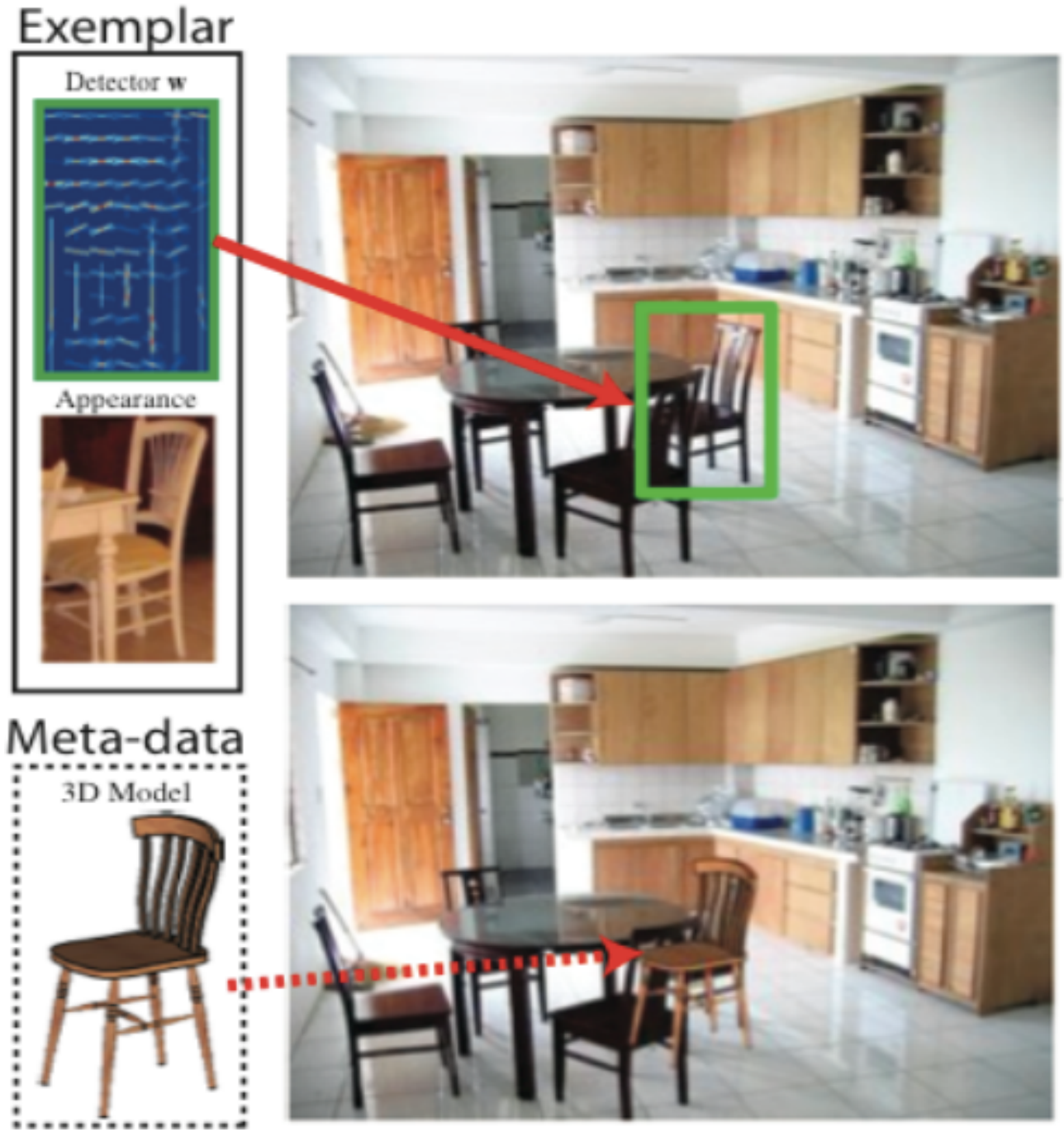
- Measure pixelwise accuracy on the 3-class geometric-labeling problem:
  "left", "front", "right" -facing
- 43.0% Hoeim et al. 2005
- 51.0% Category-SVM* + NN
- 62.3% Exemplar-SVMs

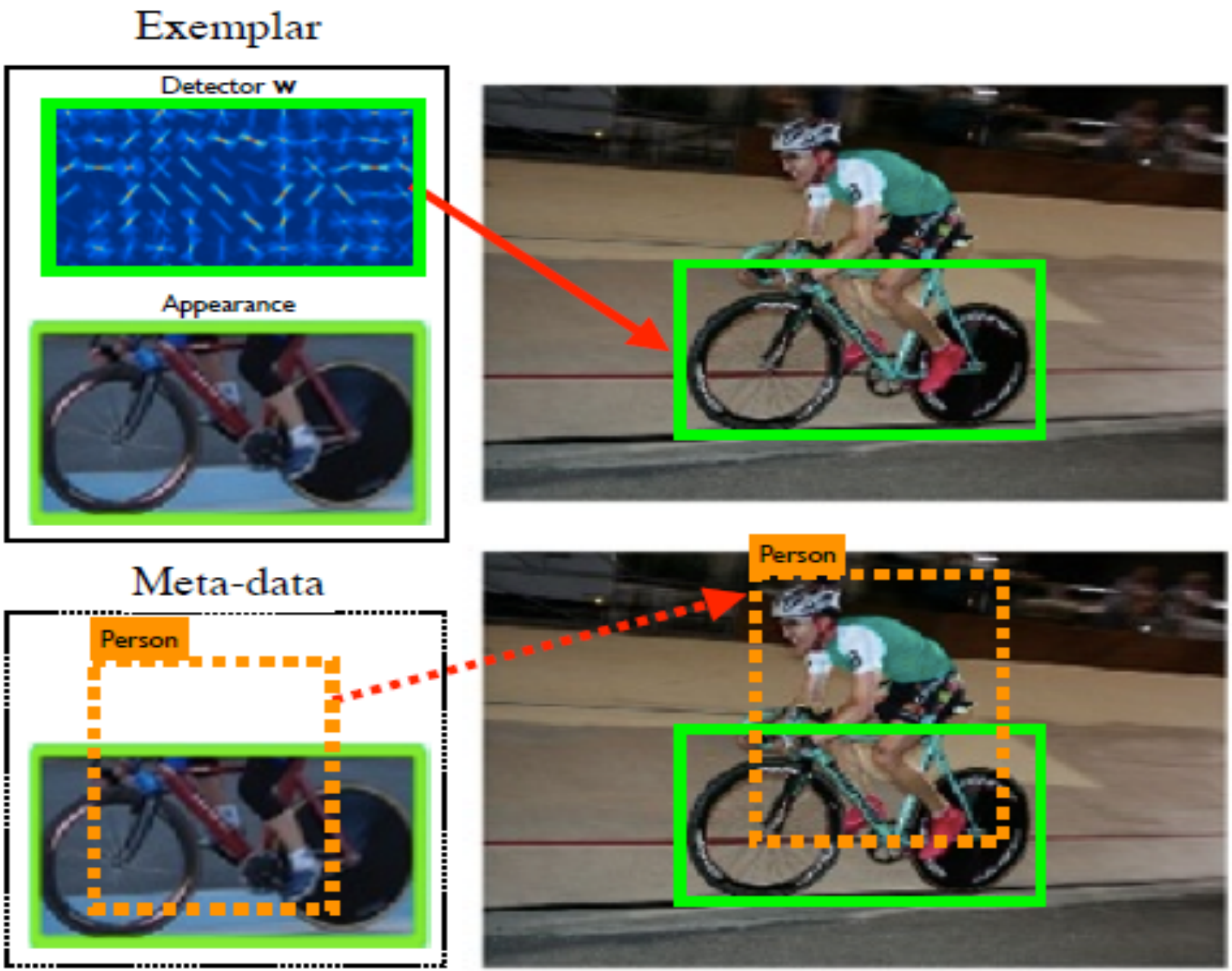
*felzenswalb et al. 2010

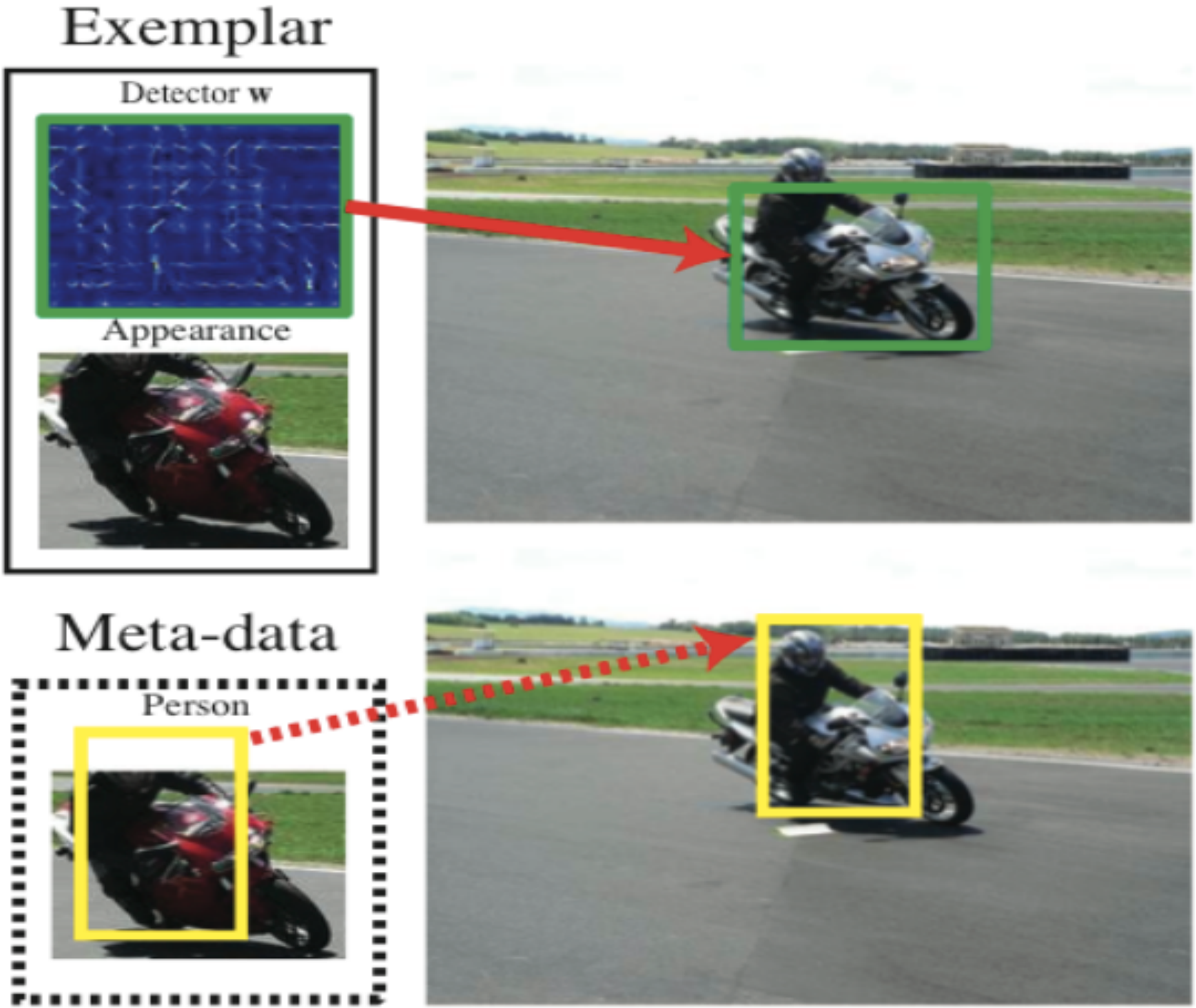
3D Model Transfer

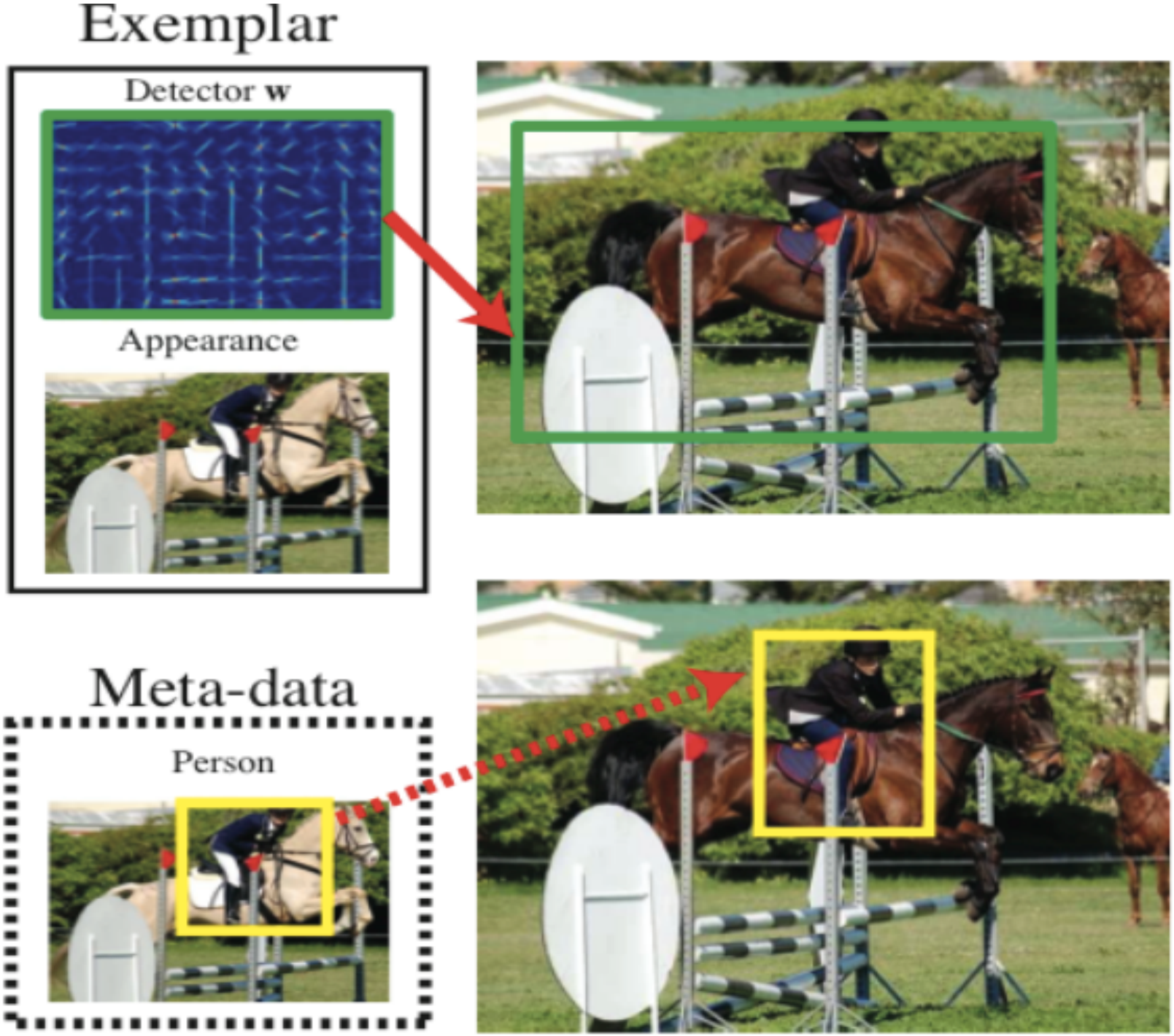
Manually align 3D model from Google 3D Warehouse with a subset of PASCAL VOC "chair" exemplars


Exemplar
Detector w
Appearance
Meta-data
3D Model


Exemplar
Detector w
Appearance
Meta-data
3D Model


Exemplar
Detector w
Appearance
Meta-data
Person
Person


Exemplar
Detector w
Appearance
Meta-data
Person

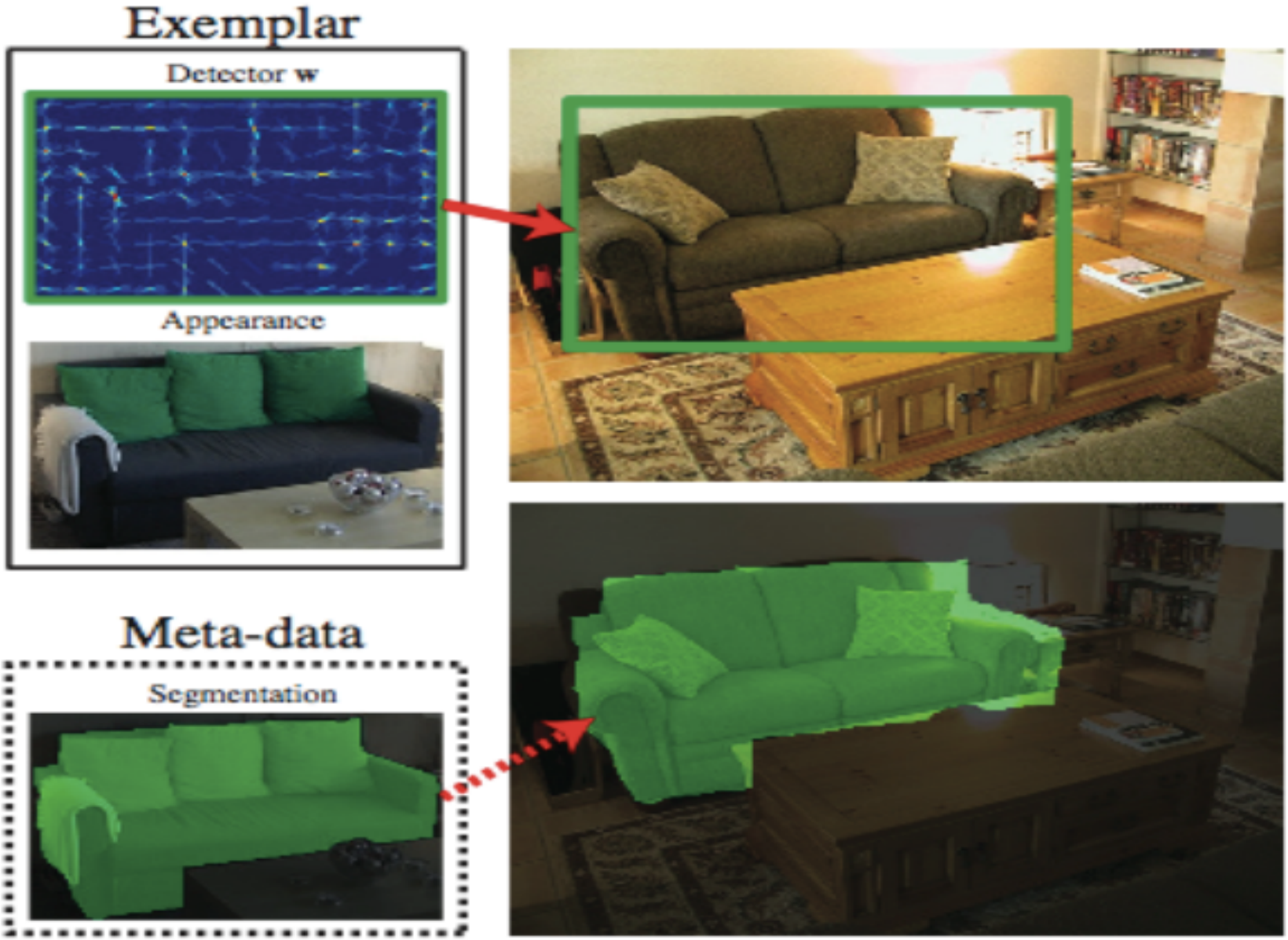

Exemplar
Detector w
Appearance
Meta-data
Person

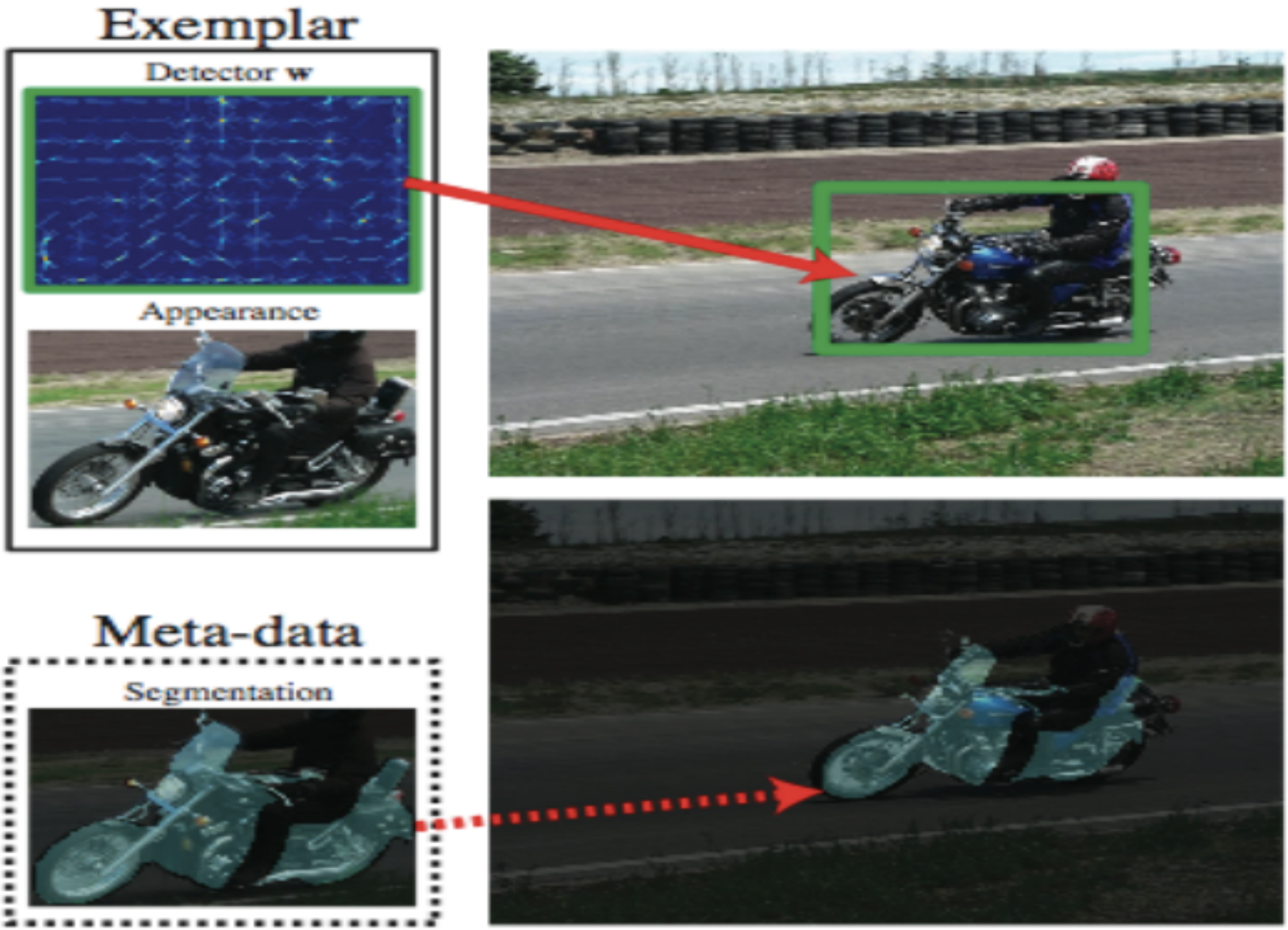
Related Object Priming: Evaluation on Person Prediction

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.

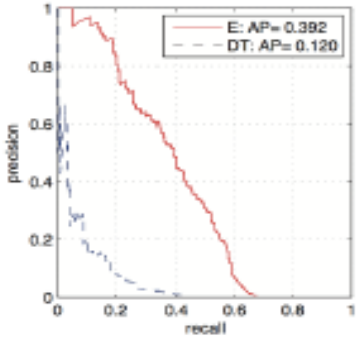
More Transfer Examples

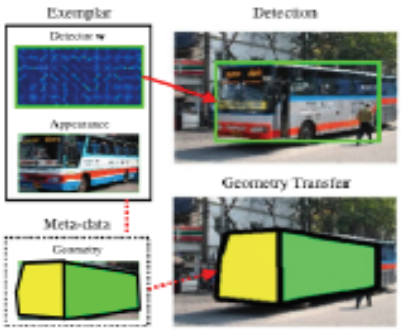


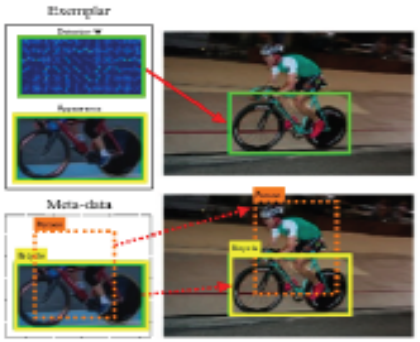

Exemplar
Detector w
Appearance
Meta-data
Segmentation

Conclusion

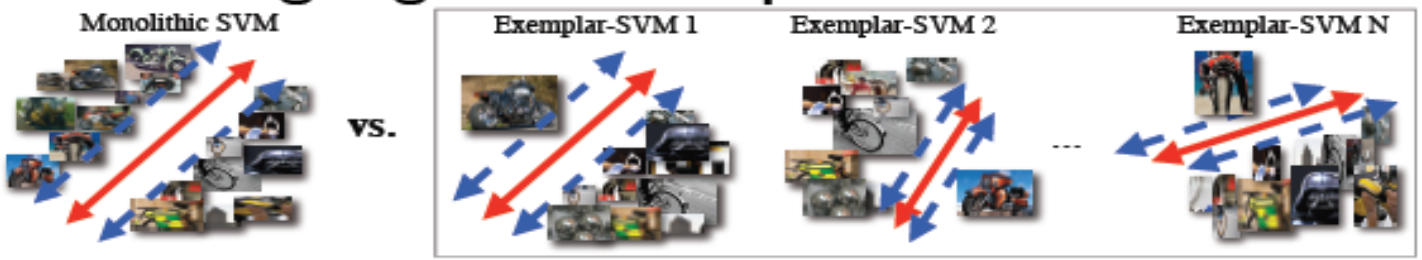
- ExemplarSVMs can be used for recognition, label transfer, and complementary object prediction







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



- Generalization is possible from a single positive example and a vast set of negatives

References

- **Ensemble of Exemplar-SVMs for Object Detection and Beyond, T. Malisiewicz, A. Gupta, A. Efros, ICCV 2011.**
- https://speakerd.s3.amazonaws.com/presentations/11f1d950241d013023151231381d9c14/slides.pdf
- http://www.cs.cmu.edu/~tmalisie/projects/iccv11/malisiewicz_icml2012_talk.pdf
- http://www.cs.cmu.edu/~tmalisie/projects/iccv11/exemplarsvm-mit-talk.pdf