The Context Challenge

The Context Challenge

Contextual Priming for Object Detection
Antonio Torralba, 2003

The Context Challenge

Contextual Priming for Object Detection
Antonio Torralba, 2003

Object detection and localization using local and global features
Kevin Murphy, Antonio Torralba, Daniel Eaton, and William Freeman, 2005

The Context Challenge

GIST
in
CONTEXT

The Context Challenge

Previous models

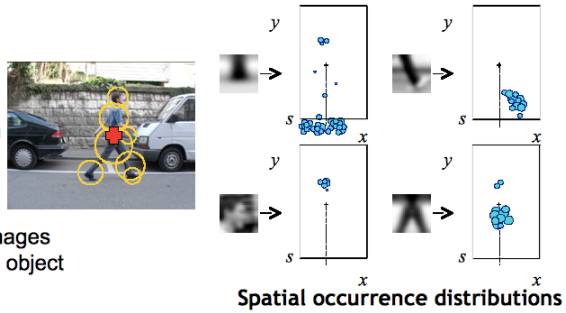
Alyosha's slide

Implicit shape model

- Use Hough space voting to find object
- Leibe and Schiele '03,'05

Learning

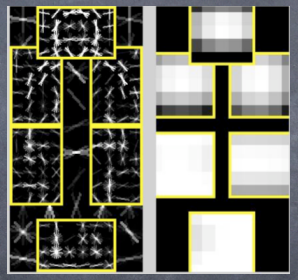
- Learn appearance codebook
  - Cluster over interest points on training images
- Learn spatial distributions
  - Match codebook to training images
  - Record matching positions on object
  - Centroid is given



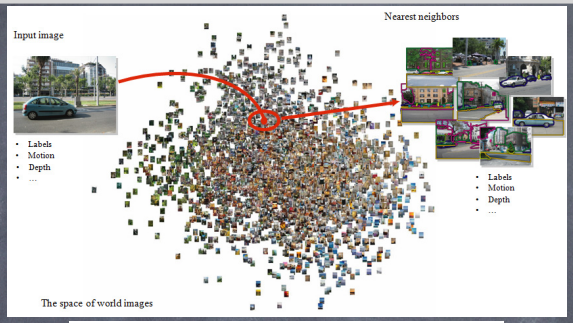
Recognition

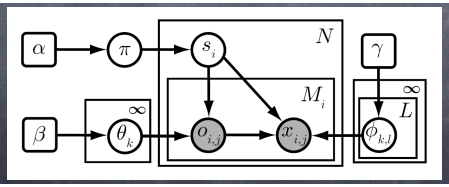
Interest Points | **Matched Codebook Entries** | **Probabilistic Voting**

The Context Challenge



The Context Challenge





$$p(o,x,g|\theta,\phi,\eta) = \prod_{i=1}^{N}\prod_{j=1}^{M_i}\sum_{h_{i,j}=0}^{1} p(o_{i,j}|h_{i,j},\theta)\, p(x_{i,j}|o_{i,j},h_{i,j},\phi)\, p(g_{i,j}|o_{i,j},h_{i,j},\eta)$$

The Context Challenge

How about:

The Context Challenge

How about:

Objects are too small

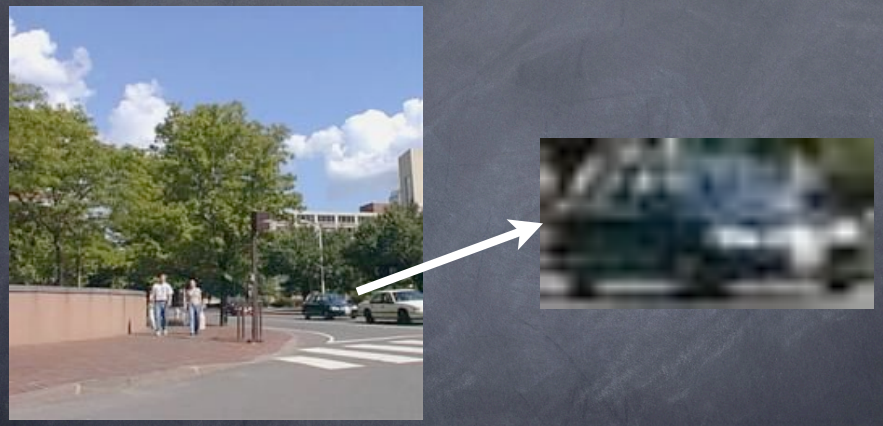
Occlusion

Lack of illumination

Poor resolution

Poor contrast

The Context Challenge



The Context Challenge

The Context Challenge

How far can you go without running an object detector?

The Context Challenge

Context:

P(computer | desk)

The Context Challenge

Context:

P(computer | desk, mouse, keyboard,....)

The Context Challenge

Context:

P(computer | desk, mouse, keyboard,....)

P(keyboard | desk, mouse, computer,....)

The Context Challenge

Context:

P(computer | desk, mouse, keyboard,....)

P(keyboard | desk, mouse, computer,....)

Need an object detector!?!

The Context Challenge

The Context Challenge

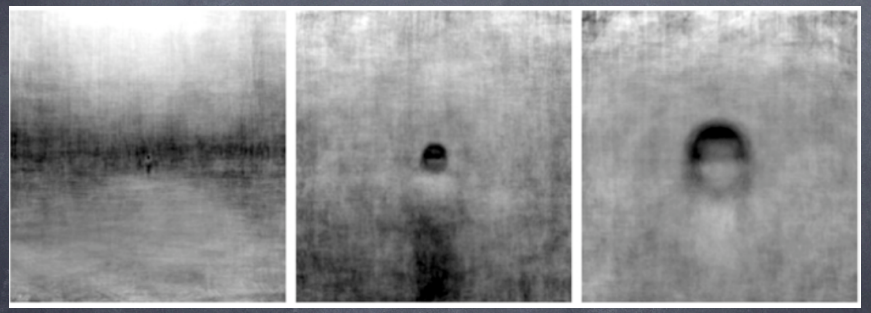
The Context Challenge

Person!

The Context Challenge

What does the average context look like?

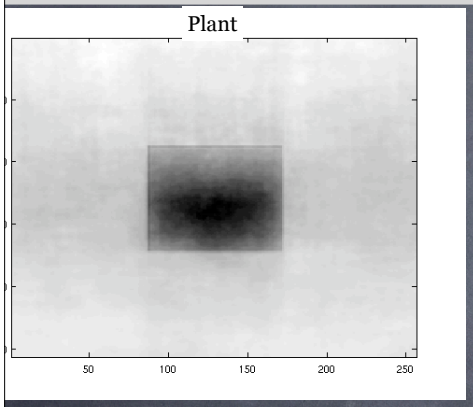
The Context Challenge



The Context Challenge

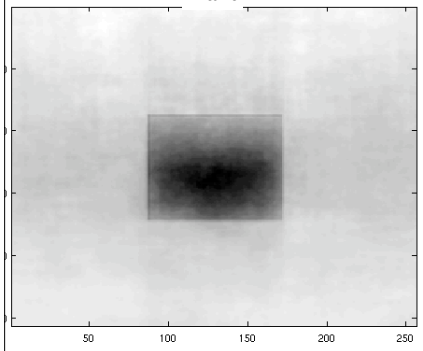
Which categories of objects
have distinct backgrounds?

The Context Challenge

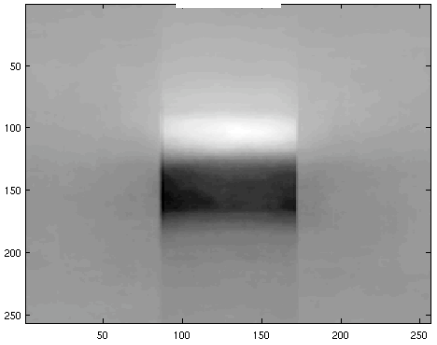


The Context Challenge

Plant

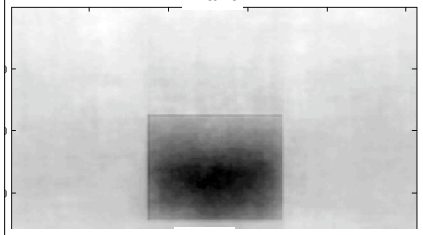


Mountain

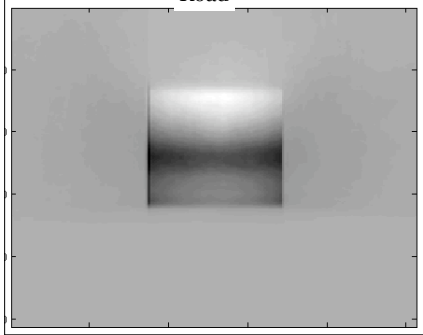


The Context Challenge

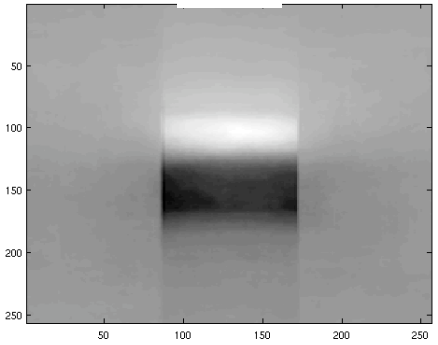
Plant



Road

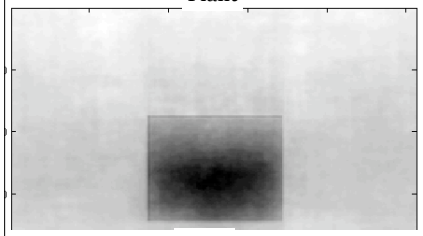


Mountain

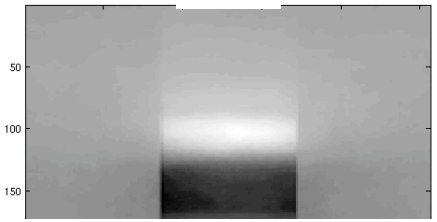


The Context Challenge

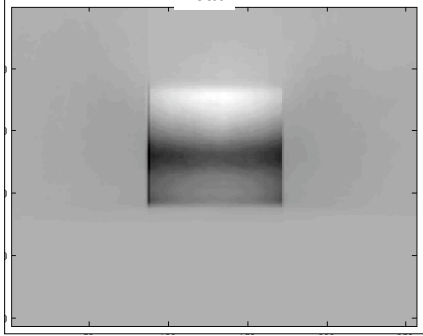
Plant



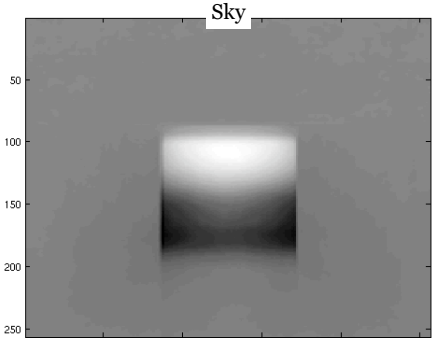
Mountain



Road



Sky

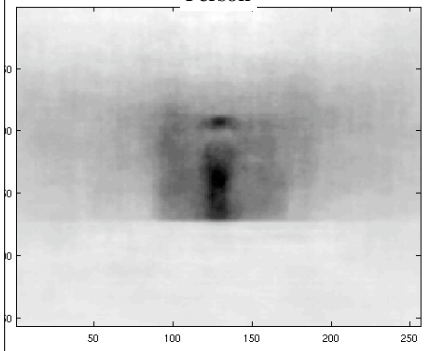


The Context Challenge

Movie...

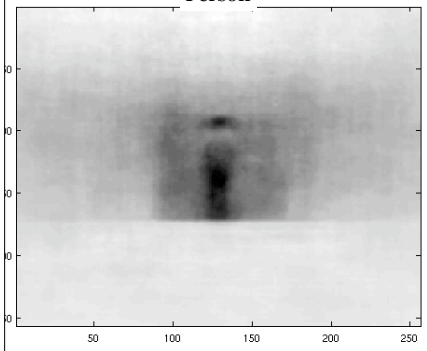
The Context Challenge

Person

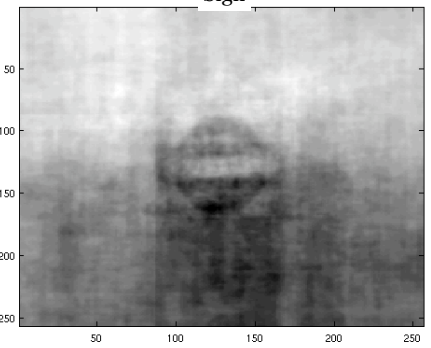


The Context Challenge

Person

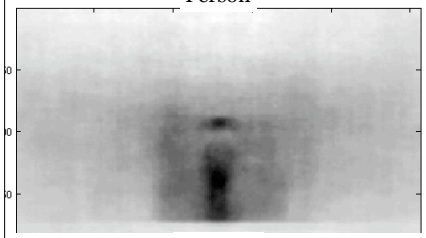


Sign

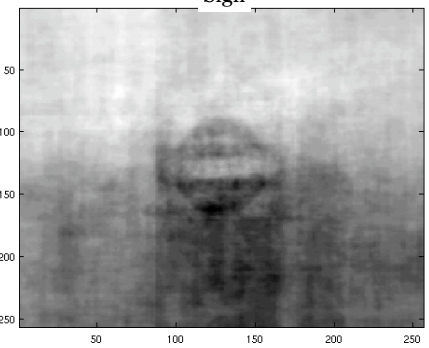


The Context Challenge

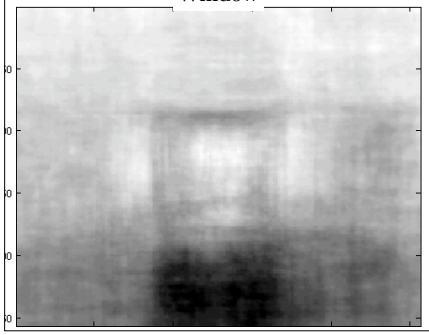
Person



Sign

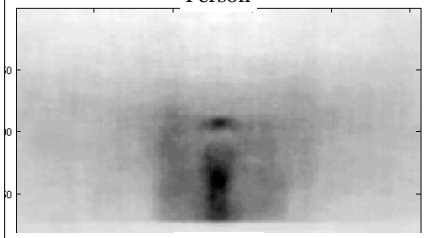


Window

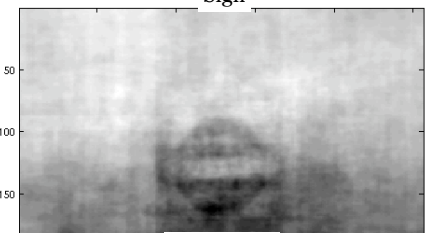


The Context Challenge

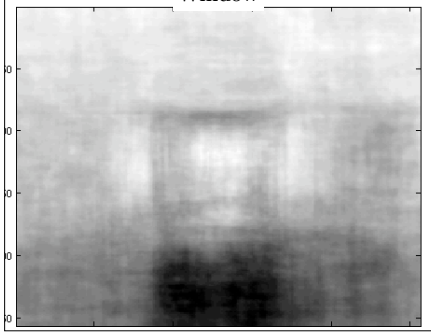
Person



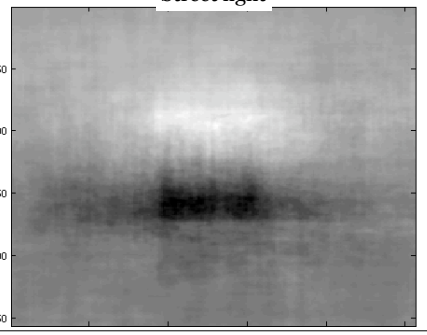
Sign



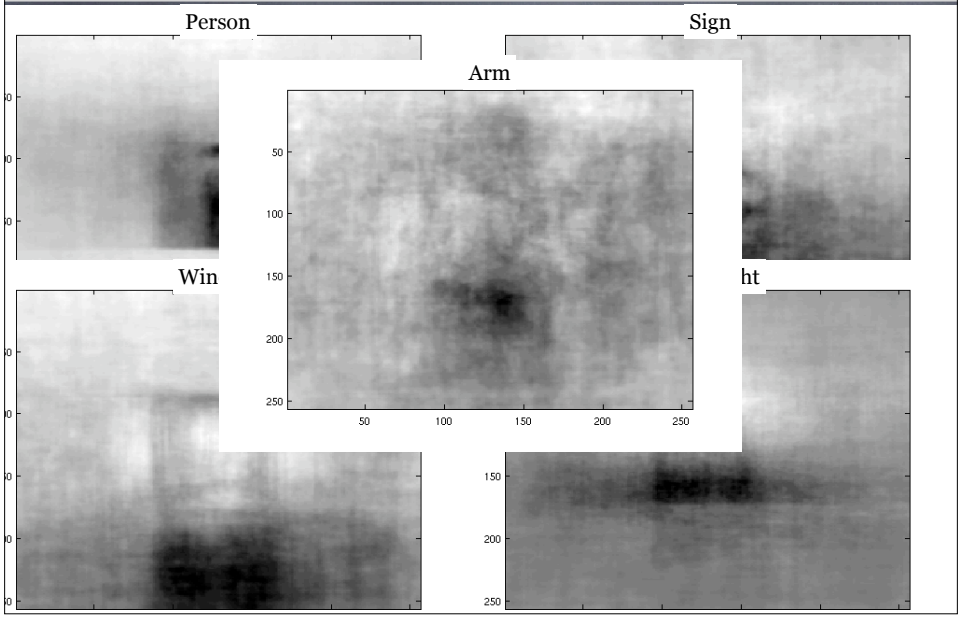
Window



Street light



The Context Challenge



The Context Challenge

Model

The Context Challenge

Model

$$P(O|\mathbf{v}) = \frac{P(\mathbf{v}|O)}{P(\mathbf{v})}P(O) = \frac{P(\mathbf{v}_L|O,\mathbf{v}_C)}{P(\mathbf{v}_L|\mathbf{v}_C)}P(O|\mathbf{v}_C)$$

$$O = \{o, \mathbf{x}, \sigma, \ldots\}$$

The Context Challenge

Model

$$P(O|\mathbf{v}) = \frac{P(\mathbf{v}|O)}{P(\mathbf{v})}P(O) = \frac{P(\mathbf{v}_L|O,\mathbf{v}_C)}{P(\mathbf{v}_L|\mathbf{v}_C)}P(O|\mathbf{v}_C)$$

$$O = \{o, \mathbf{x}, \sigma, \ldots\}$$

$$P(O|\mathbf{v}_C) = P(\sigma|\mathbf{x}, o, \mathbf{v}_C)P(\mathbf{x}|o, \mathbf{v}_C)P(o|\mathbf{v}_C)$$

The Context Challenge

$$P(O|\mathbf{v}_C) = P(\sigma|\mathbf{x}, o, \mathbf{v}_C)P(\mathbf{x}|o, \mathbf{v}_C)\boxed{P(o|\mathbf{v}_C)}$$

Object priming

The Context Challenge

$$P(O|\mathbf{v}_C) = P(\sigma|\mathbf{x}, o, \mathbf{v}_C)\boxed{P(\mathbf{x}|o, \mathbf{v}_C)}P(o|\mathbf{v}_C)$$

Object location

The Context Challenge

$$P(O|\mathbf{v}_C) = \boxed{P(\sigma|\mathbf{x}, o, \mathbf{v}_C)}P(\mathbf{x}|o, \mathbf{v}_C)P(o|\mathbf{v}_C)$$

Object scale

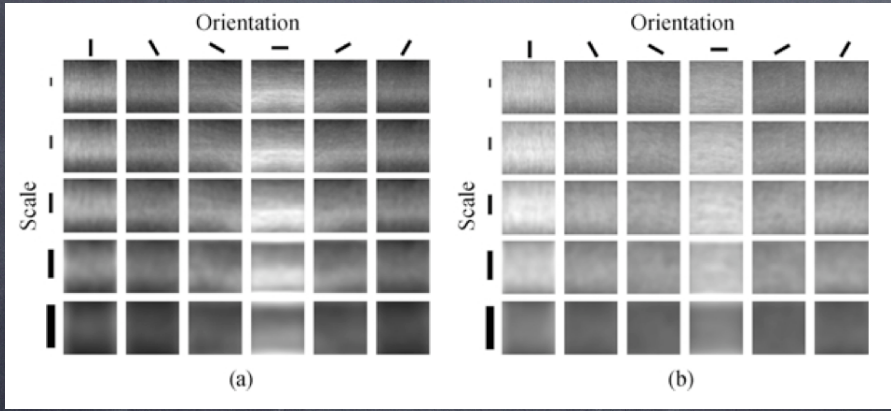
The Context Challenge

No objects in the representation!

The Context Challenge

Why would this work?

The Context Challenge



Conditional average of $v(\mathbf{x}, k)$ with respect to the presence or absence of different objects. (a) $E[v(\mathbf{x}, k) | \neg people, car]$ and (b) $E[v(\mathbf{x}, k) | people, \neg car]$.

The Context Challenge

Learning the models

The Context Challenge

Learning the models

$$P(o|\mathbf{v}_C) \approx P(\mathbf{v}_C|o) = \sum_{i=1}^{M} b_i G(\mathbf{v}_C; \mu_i, \Sigma_i)$$

The Context Challenge

Computing P(O|v_c)....

The Context Challenge

Computing P(O|v_c)....

Gist →

8192 descriptor for
4 frequencies and 8
orientations, 16x16 block

The Context Challenge

Computing P(O|v_c)....

8192 descriptor for
4 frequencies and 8
orientations, 16x16 block

...

8192 descriptor for
4 frequencies and 8
orientations, 16x16 block

The Context Challenge

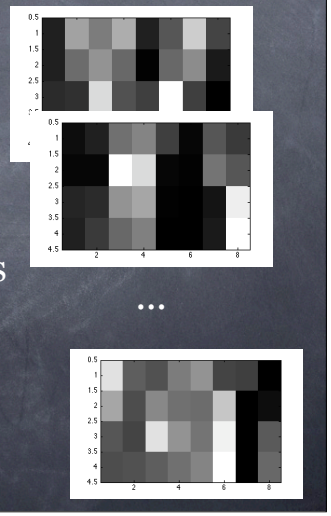
Computing P(O|v_c)....

8192 descriptor for
4 frequencies and 8
orientations, 16x16 block

...

8192 descriptor for
4 frequencies and 8
orientations, 16x16 block

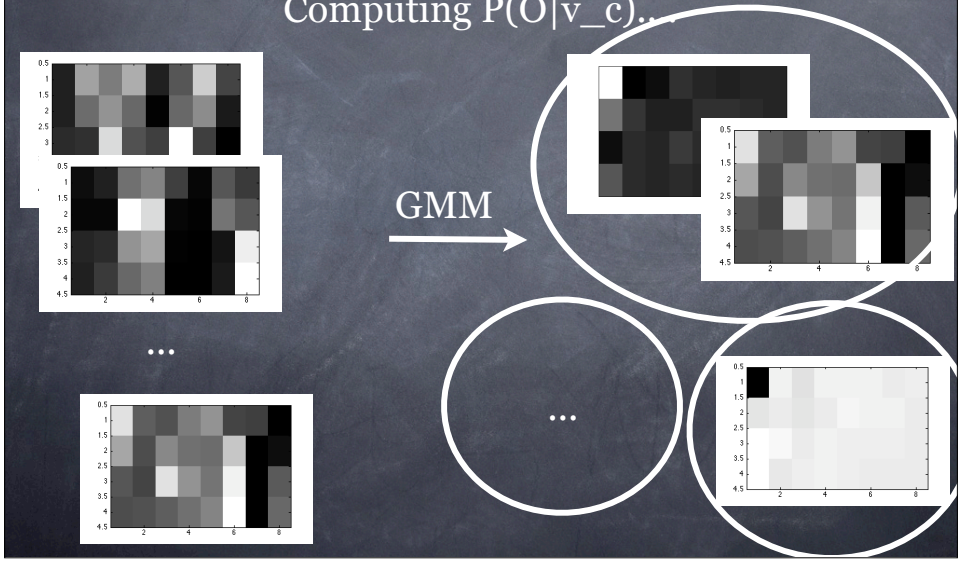
PCA →
32 descriptors



...

The Context Challenge

Computing P(O|v_c)...



The Context Challenge

Results

Object priming

$$P(O|\mathbf{v}_C)$$

The Context Challenge $P(O|\mathbf{v}_C)$

*o*1 = people, *o*2 = furniture, *o*3 = vehicles and *o*4 = trees. The bars at the right-hand of each picture represent the probability *P* (*o* | **v***C*).

The Context Challenge

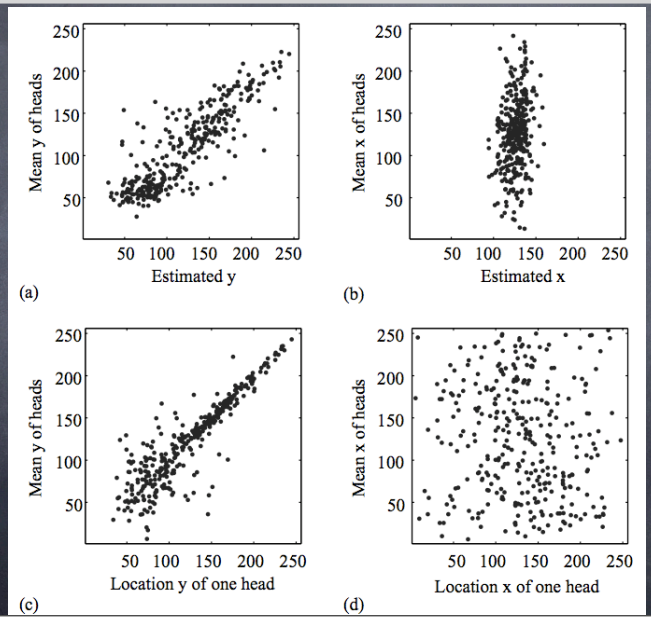
Focus of attention

$$P(\mathbf{x}|o, \mathbf{v}_C)$$

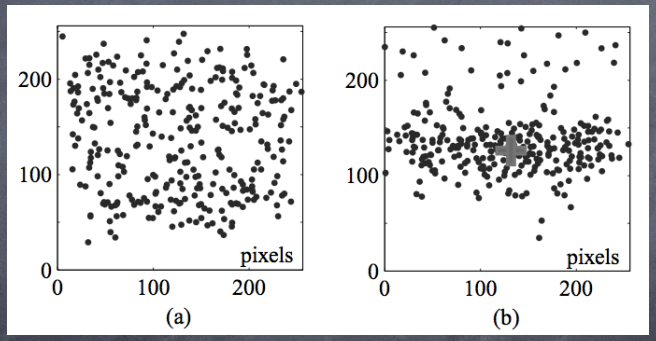
The Context Challenge $P(\mathbf{x}|o, \mathbf{v}_C)$

Mixture of Gaussians

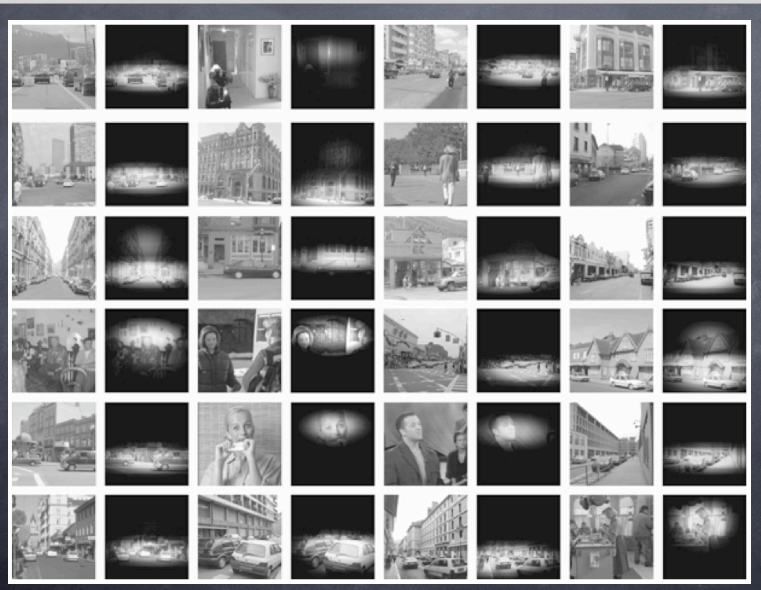
The Context Challenge $P(\mathbf{x}|o, \mathbf{v}_C)$



The Context Challenge $P(\mathbf{x}|o, \mathbf{v}_C)$



The Context Challenge $P(\mathbf{x}|o, \mathbf{v}_C)$



The Context Challenge

Scale selection

$P(\sigma|o, \mathbf{v}_C)$

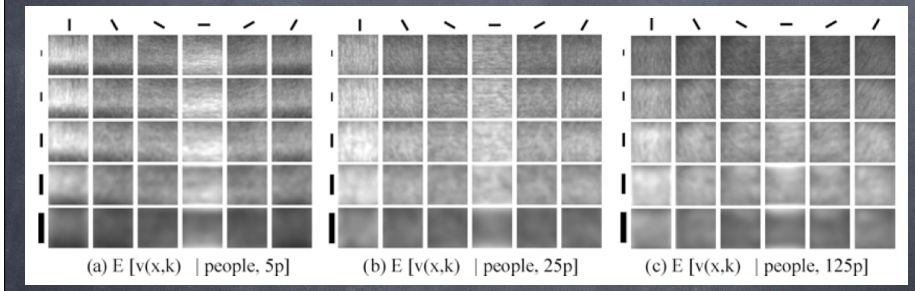
The Context Challenge $P(\sigma|o, \mathbf{v}_C)$

Scale selection

Mixture of Gaussians

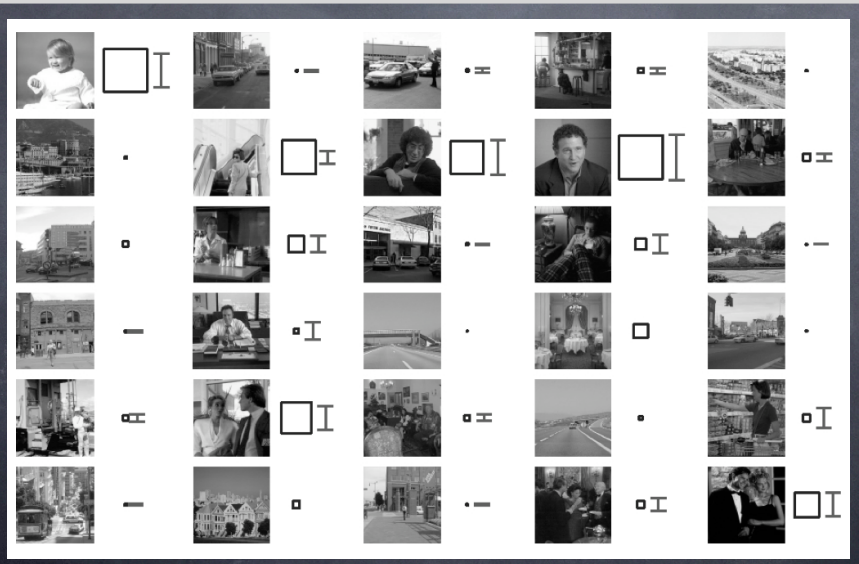
The Context Challenge $P(\sigma|o, \mathbf{v}_C)$

Scale selection



Conditional average of $\mathbf{v}(\mathbf{x}, k)$ for images that contain people at three different scales 5, 25 and 125 pixels.

The Context Challenge $P(\sigma|o, \mathbf{v}_C)$

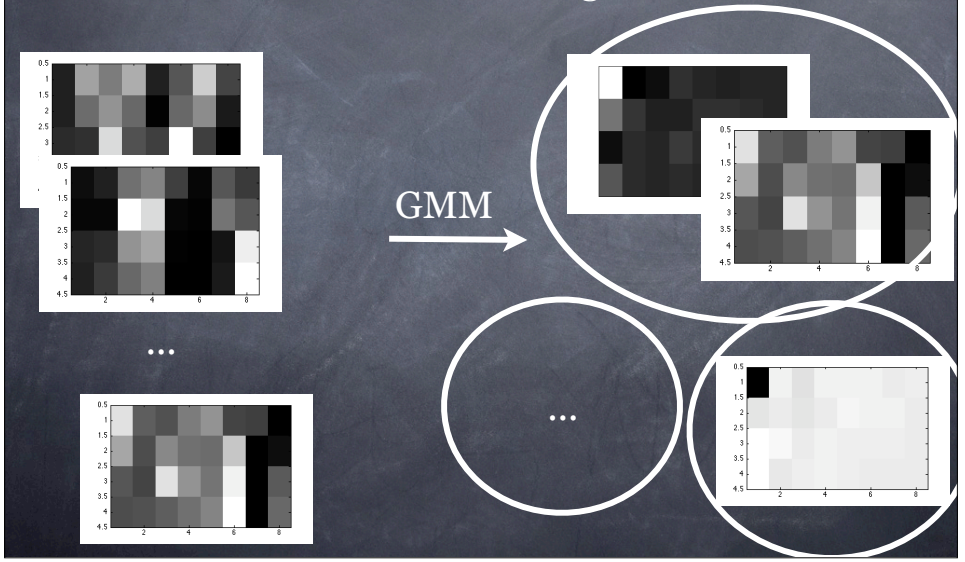


The Context Challenge

All we are doing…

The Context Challenge

All we are doing…



The Context Challenge

Evaluation

The Context Challenge

Evaluation

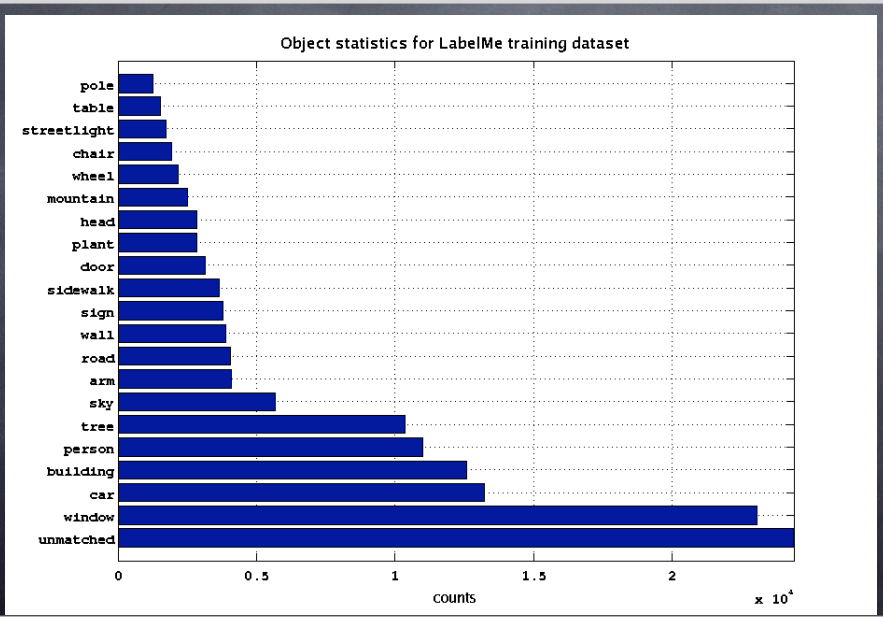
Testing on more object categories…

Top 20 categories for number of training samples, LabelMe database, as in Scene alignment paper

The Context Challenge

LabelMe statistics for training categories

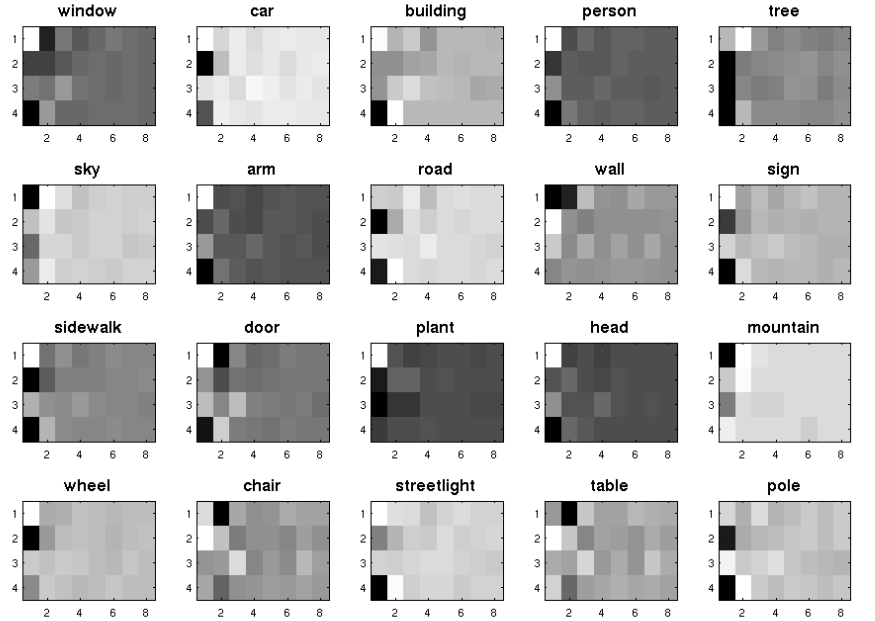
The Context Challenge



The Context Challenge

LabelMe Database

The model…



The Context Challenge

Evaluation statistics

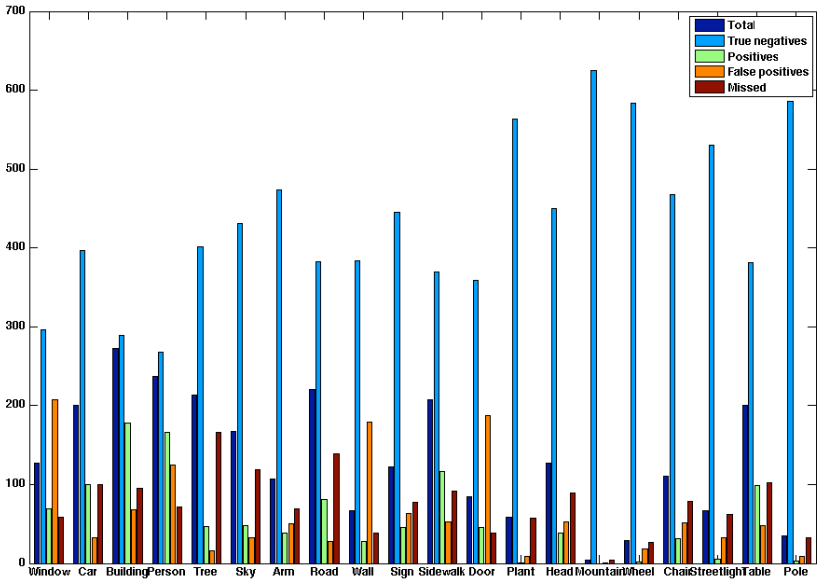
For each test image, find using the GMM model

$$P(O|\mathbf{v}_C)$$

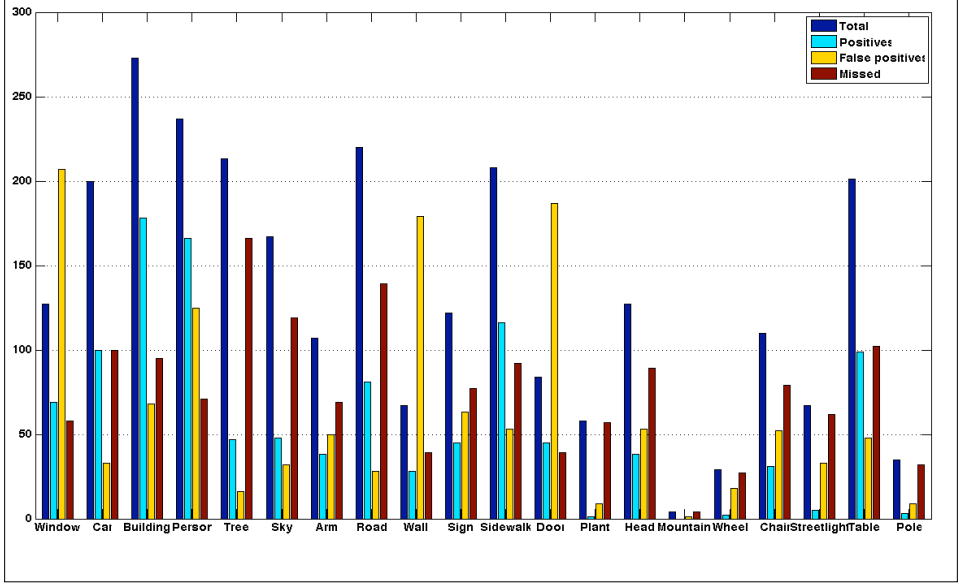
Observe:

Total objects, positives, false positives, missed

The Context Challenge



The Context Challenge

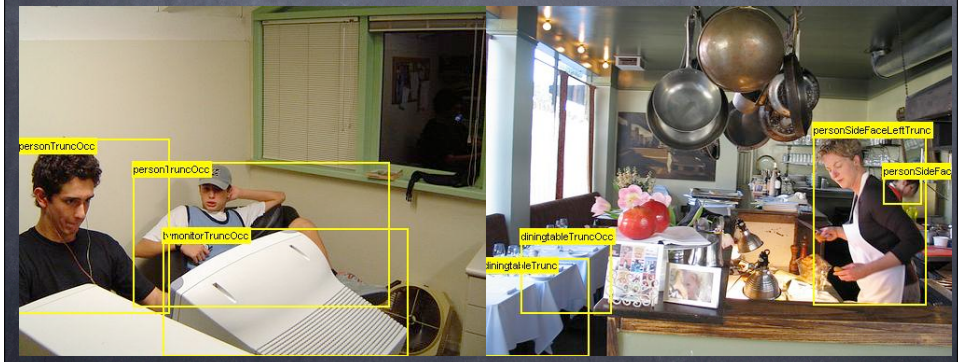


The Context Challenge

PASCAL Database

The Context Challenge

PASCAL Database



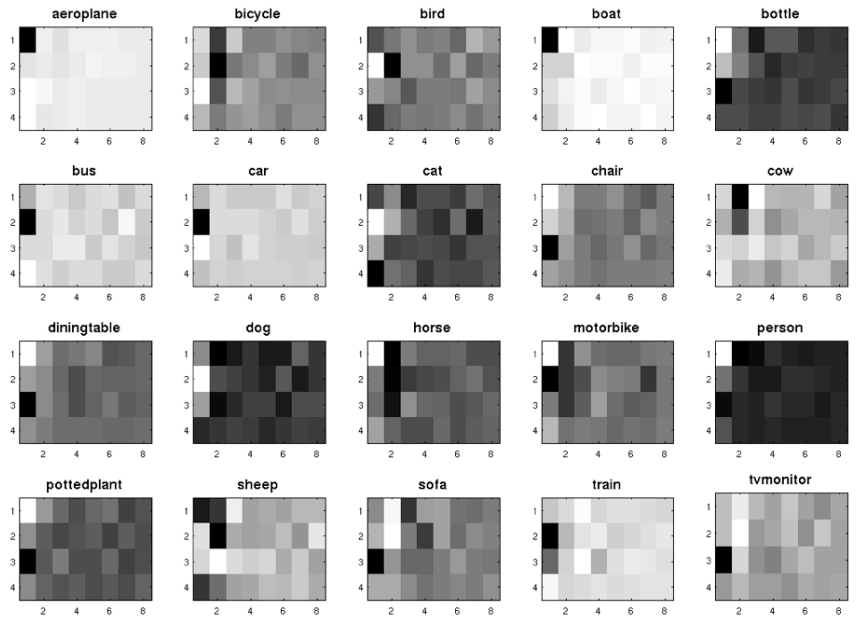
The Context Challenge

PASCAL Database

The Context Challenge

PASCAL Database

The model...



The Context Challenge

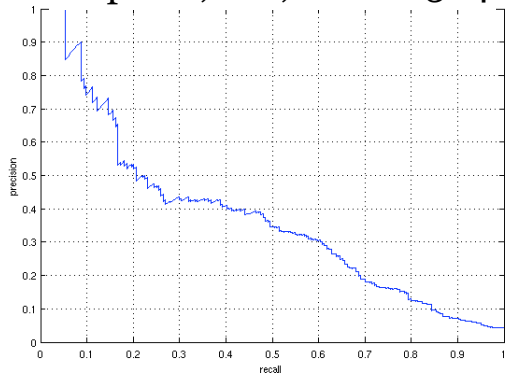
PASCAL Database

Using evaluation method from competition

Object probability from each class is used for the confidence

The Context Challenge

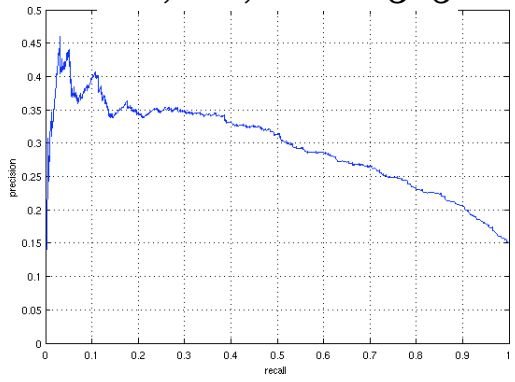
Airplane, test, AP = 0.384



Precision -Recall

The Context Challenge

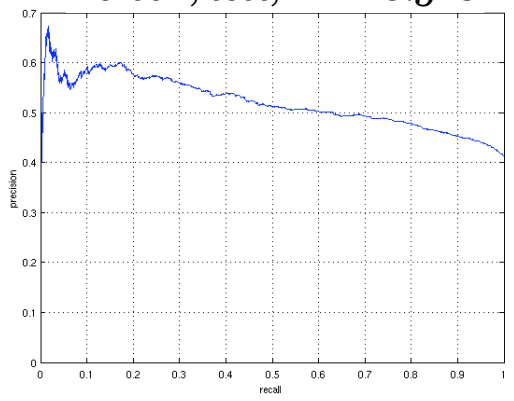
Car, test, AP = 0.305



Precision -Recall

The Context Challenge

Person, test, AP = 0.528



Precision -Recall

The Context Challenge

PASCAL Database

Using evaluation method from competition

Airplane, test, AP = 0.384 vs 0.775 best

Person, test, AP = 0.528 vs 0.859 best

Car, test, AP = 0.305 vs 0.78 best

The Context Challenge

PASCAL Database

Using evaluation method from competition

Airplane, test, AP ... best

Perso... ...59 best

C... ...305 vs 0.78 best

Hold it!

The Context Challenge

Evaluation statistics, v2.0

For each test image, find

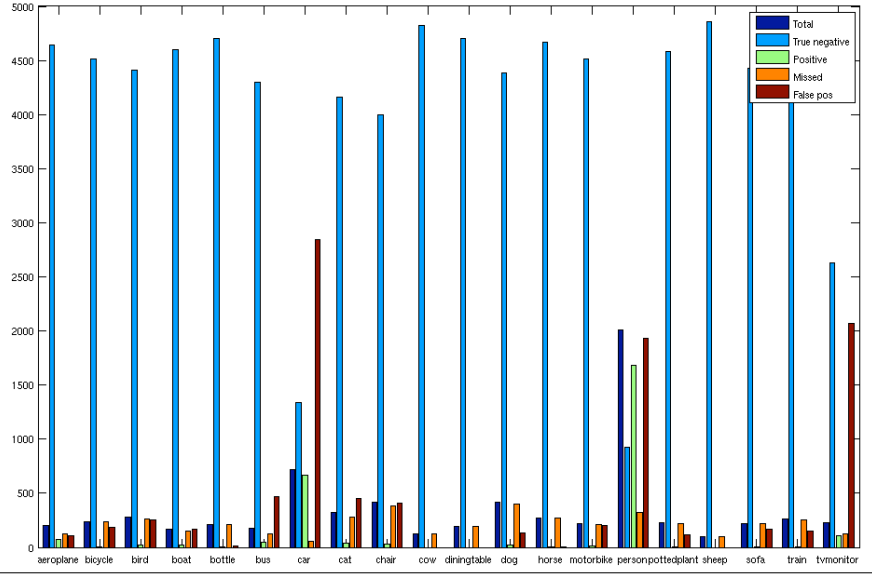
$$P(O|\mathbf{v}_C)$$

Observe:

Total objects, positives, false positives, missed, given a threshold for the probability of an object being in the image.

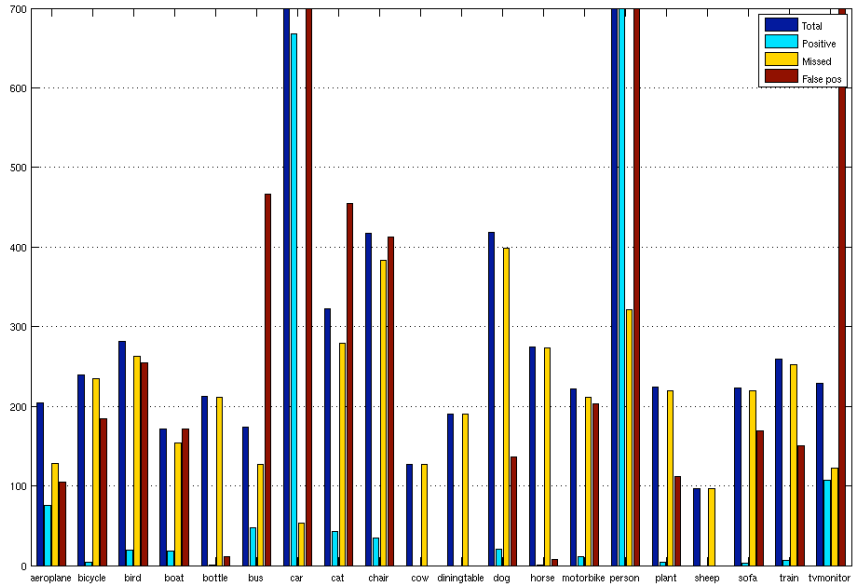
The Context Challenge

Threshold = 0.1



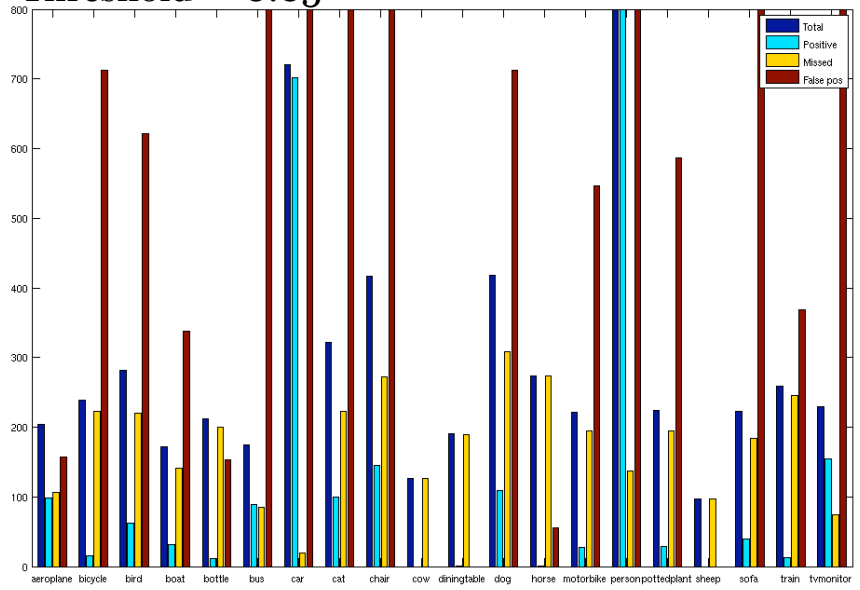
The Context Challenge

Threshold = 0.1



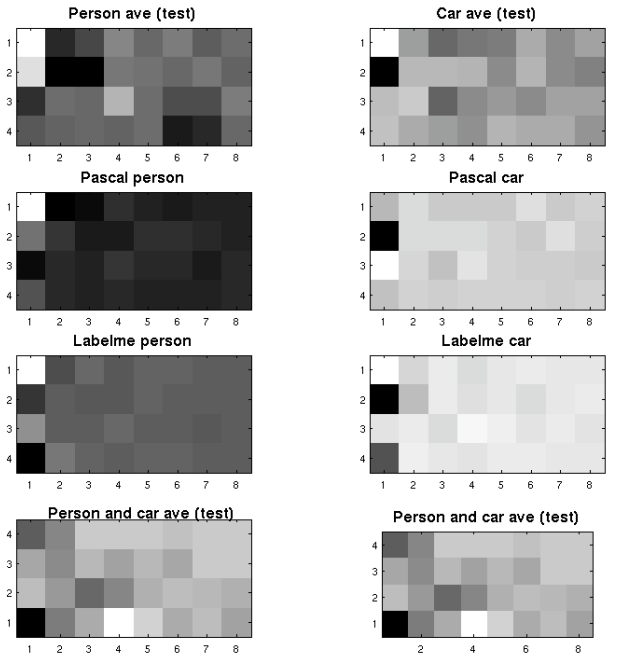
The Context Challenge

Threshold = 0.05



The Context Challenge

Images that have multiple object categories

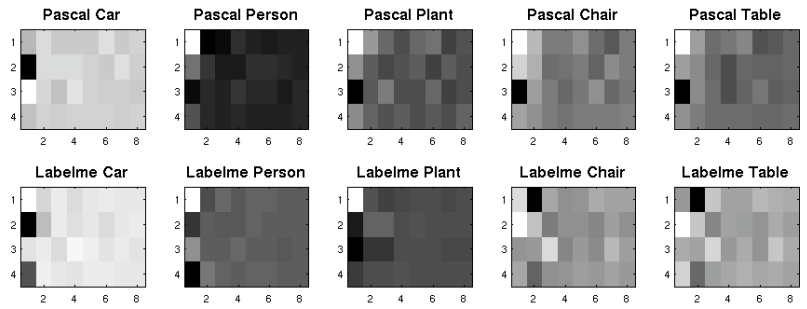


The Context Challenge

Combine the databases?

Combine the class models?

The Context Challenge



The Context Challenge

More evaluation:

More object categories
More gist
Compare/combine with HOG/local detector

The Context Challenge

Misclassified as car

The Context Challenge

Correct: sofa
Misclassified: car, monitor, person

The Context Challenge

Correct: car
Misclassified: sofa, person, monitor

The Context Challenge

Misclassified: bicycle, car, cat, monitor
Missed: chair, table, plant

The Context Challenge