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

GIST in CONTEXT
The Context Challenge

Previous models

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

Spatial occurrence distributions
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

The Context Challenge

The Context Challenge

The Context Challenge

The Context Challenge

The Context Challenge

The Context Challenge

The Context Challenge
The Context Challenge

How far can you go without running an object detector?

Context:

$P(\text{computer} \mid \text{desk})$

Context:

$P(\text{computer} \mid \text{desk, mouse, keyboard,....})$

Context:

$P(\text{keyboard} \mid \text{desk, mouse, computer,....})$
The Context Challenge

Context:

\[
P(\text{computer} \mid \text{desk, mouse, keyboard,....})
\]
\[
P(\text{keyboard} \mid \text{desk, mouse, computer,....})
\]

Need an object detector!?!
What does the average context look like?

Which categories of objects have distinct backgrounds?
The Context Challenge

Plant

Mountain

Road

Sky

The Context Challenge

Movie...
The Context Challenge

Person

Sign

Window

Street light
\[ P(O|v) = \frac{P(v|O)P(O)}{P(v)} = \frac{P(v_L|O,v_C)P(O|v_C)}{P(v_L|v_C)}P(O|v_C) \]

\[ O = \{o, x, \sigma, \ldots\} \]
### The Context Challenge

**Object priming**

\[
P(O|v_C) = P(\sigma|x, o, v_C)P(x|o, v_C)P(o|v_C)
\]

**Object location**

\[
P(O|v_C) = P(\sigma|x, o, v_C)P(x|o, v_C)P(o|v_C)
\]

**Object scale**

\[
P(O|v_C) = P(\sigma|x, o, v_C)P(x|o, v_C)P(o|v_C)
\]

**No objects in the representation!**

\[
P(O|v_C) = P(\sigma|x, o, v_C)P(x|o, v_C)P(o|v_C)
\]
The Context Challenge

Why would this work?

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

Learning the models

\[
P(o|v_C) \approx P(v_C|o) = \sum_{i=1}^{M} b_i G(v_C; \mu_i, \Sigma_i)
\]
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

Gist

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

PCA

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

32 descriptors

...
The Context Challenge

Computing $P(O|v_C)$

GMM

Object priming

$P(O|v_C)$

Focus of attention

$P(x|o, v_C)$

$o_1 =$ 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 $P(x|o, v_C)$

Mixture of Gaussians

The Context Challenge $P(x|o, v_C)$

The Context Challenge $P(x|o, v_C)$
Conditional average of $v(x, k)$ for images that contain people at three different scales 5, 25 and 125 pixels.
The Context Challenge

All we are doing...

GMM

... ... ...

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

The Context Challenge

LabelMe statistics for training categories

The Context Challenge

LabelMe Database

The model...
The Context Challenge

Evaluation statistics

For each test image, find using the GMM model

\[ P(O|v_C) \]

Observe:

Total objects, positives, false positives, missed

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

Precision - Recall

Airplane, test, AP = 0.384

Car, test, AP = 0.305

Person, test, AP = 0.528
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

Hold it!

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

Evaluation statistics, v2.0

For each test image, find

\[ P(O|v_C) \]

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

Threshold = 0.05

Images that have multiple object categories
The Context Challenge

Combine the databases?
Combine the class models?

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

Misclassified as car
The Context Challenge

Correct: sofa
Misclassified: car, monitor, person

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