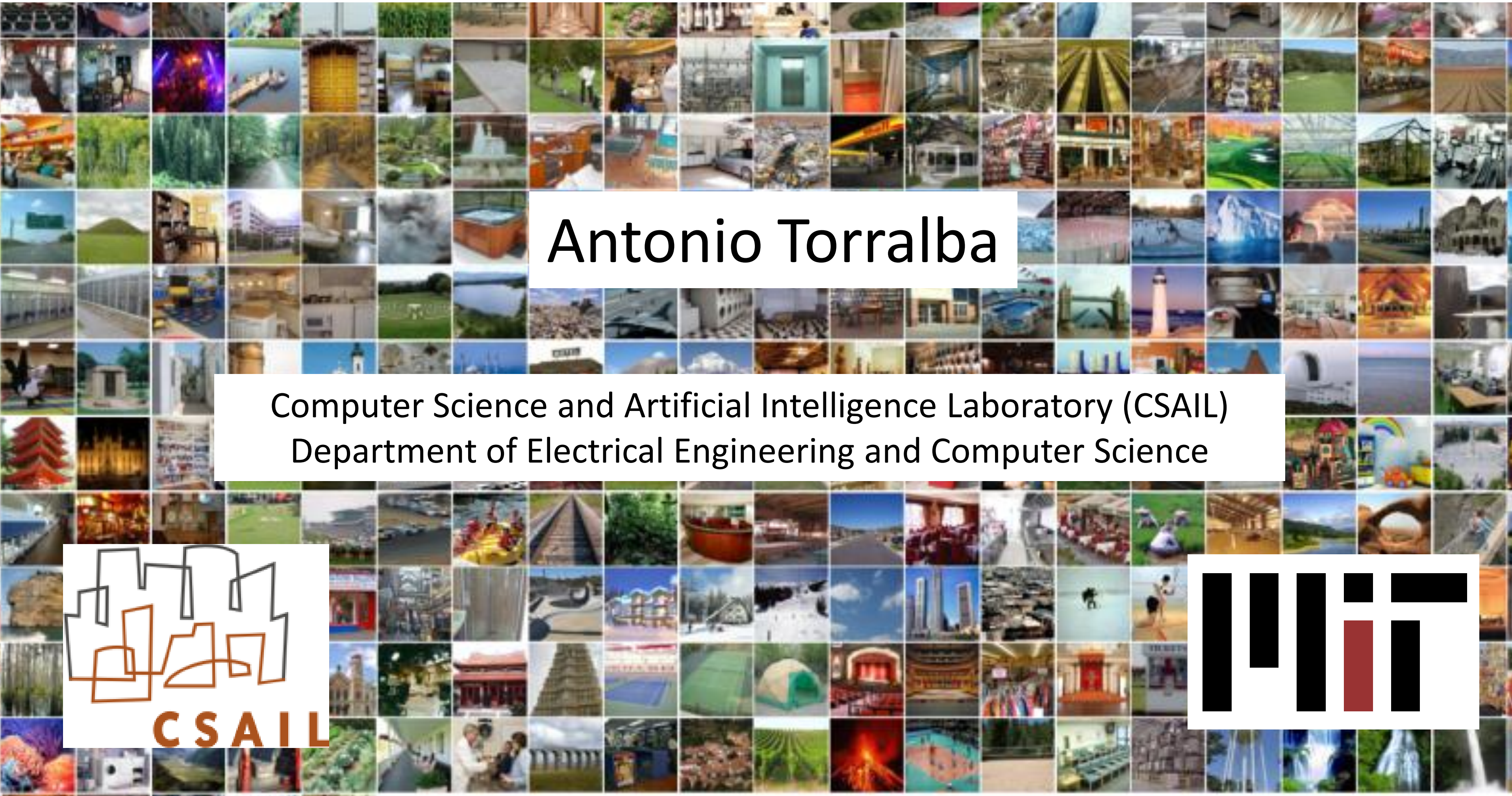
Many classes, context, and miscellaneous



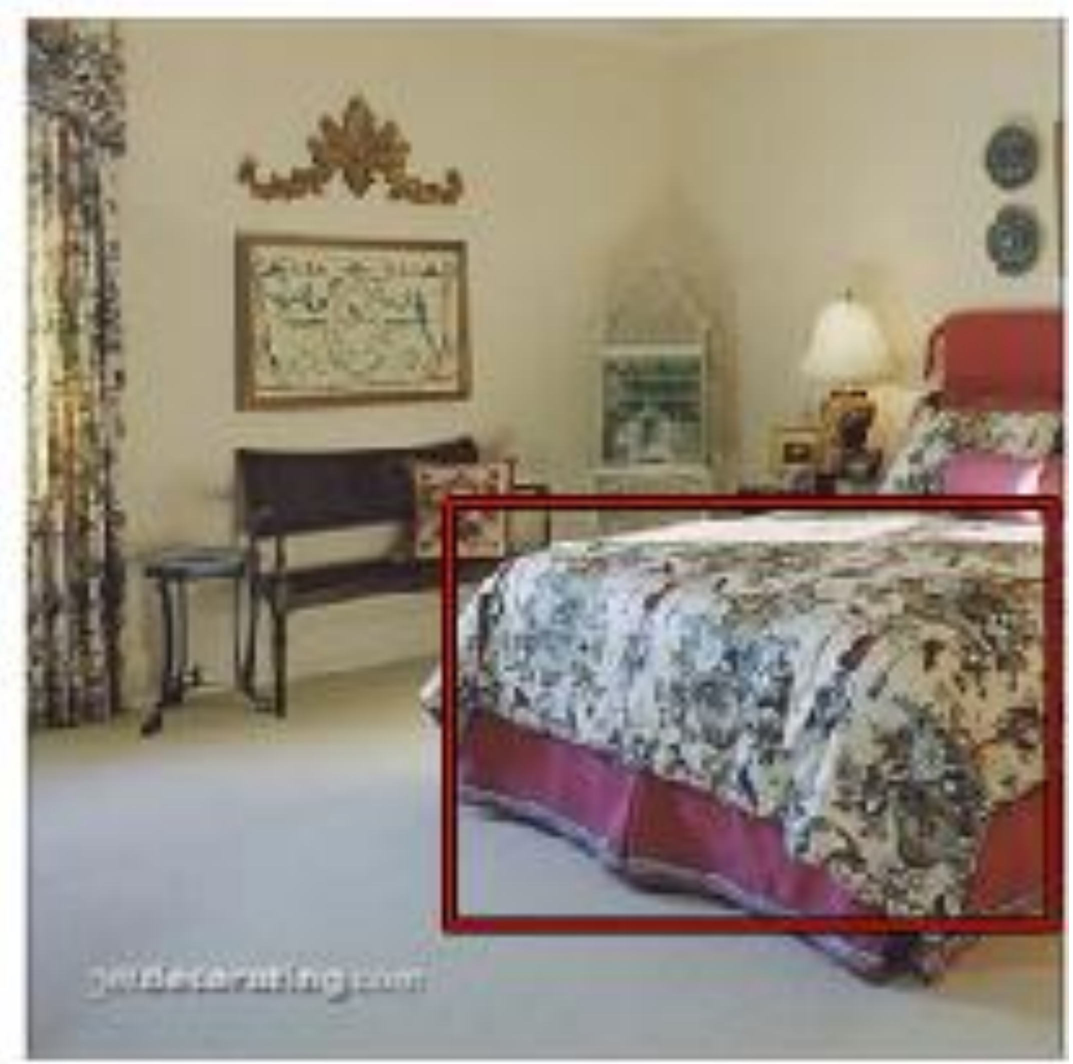
Antonio Torralba

Computer Science and Artificial Intelligence Laboratory (CSAIL)
Department of Electrical Engineering and Computer Science

SNOW EMERGENCY ROUTE
NO PARKING DURING EMERGENCY
you need a zipcar for this.
zipcar

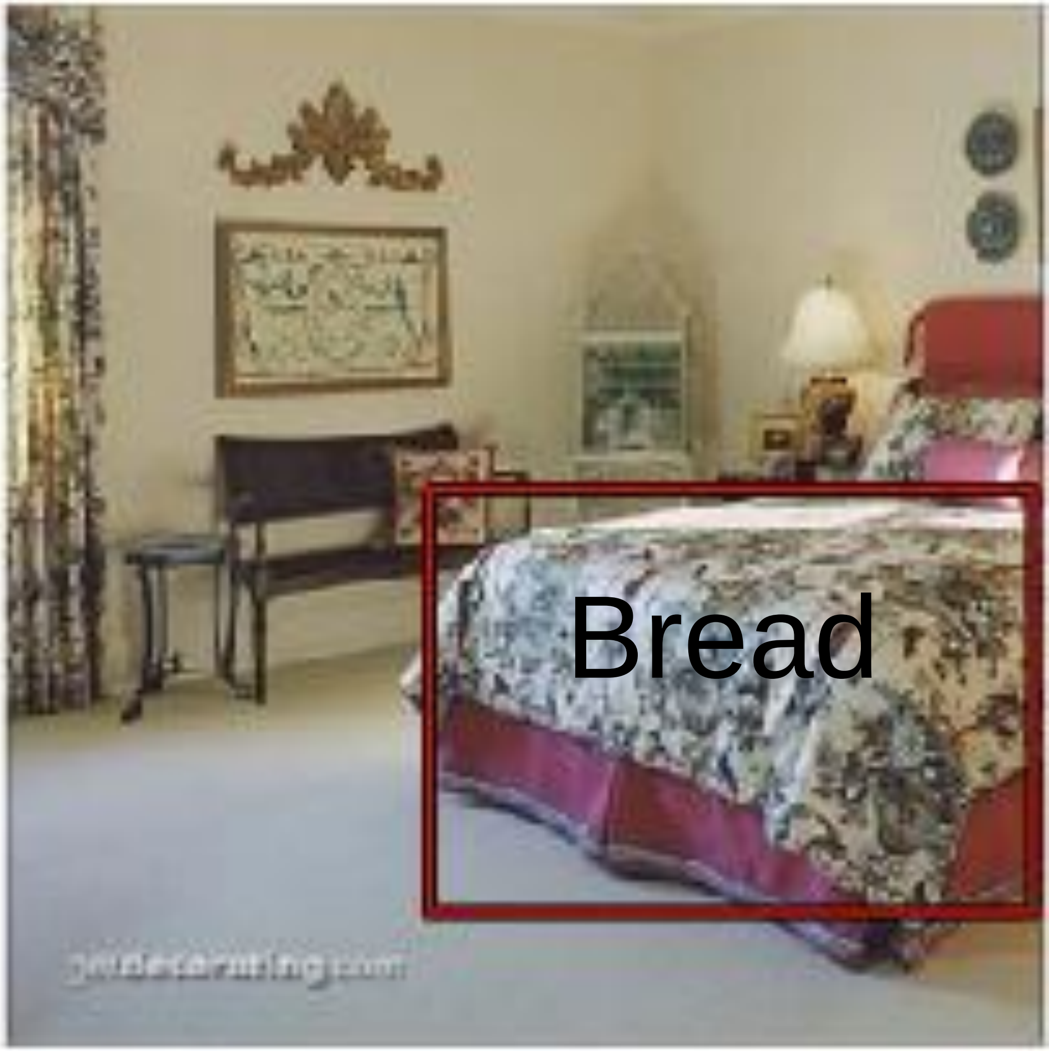
showcase[0.82]
curtain[0.81]
curtain[0.86]
curtain[0.90]
person[0.99]
person[0.96]
person[0.91]
platform[0.83]
gate[0.80]
clothes[0.82]
pillow[0.96]
armchair[0.51]
floor[0.90]
grass[0.86]
car[0.93]
stand[0.84]
chair[0.82]
car[0.83]
car[0.92]
car[0.93]

The detector challenge



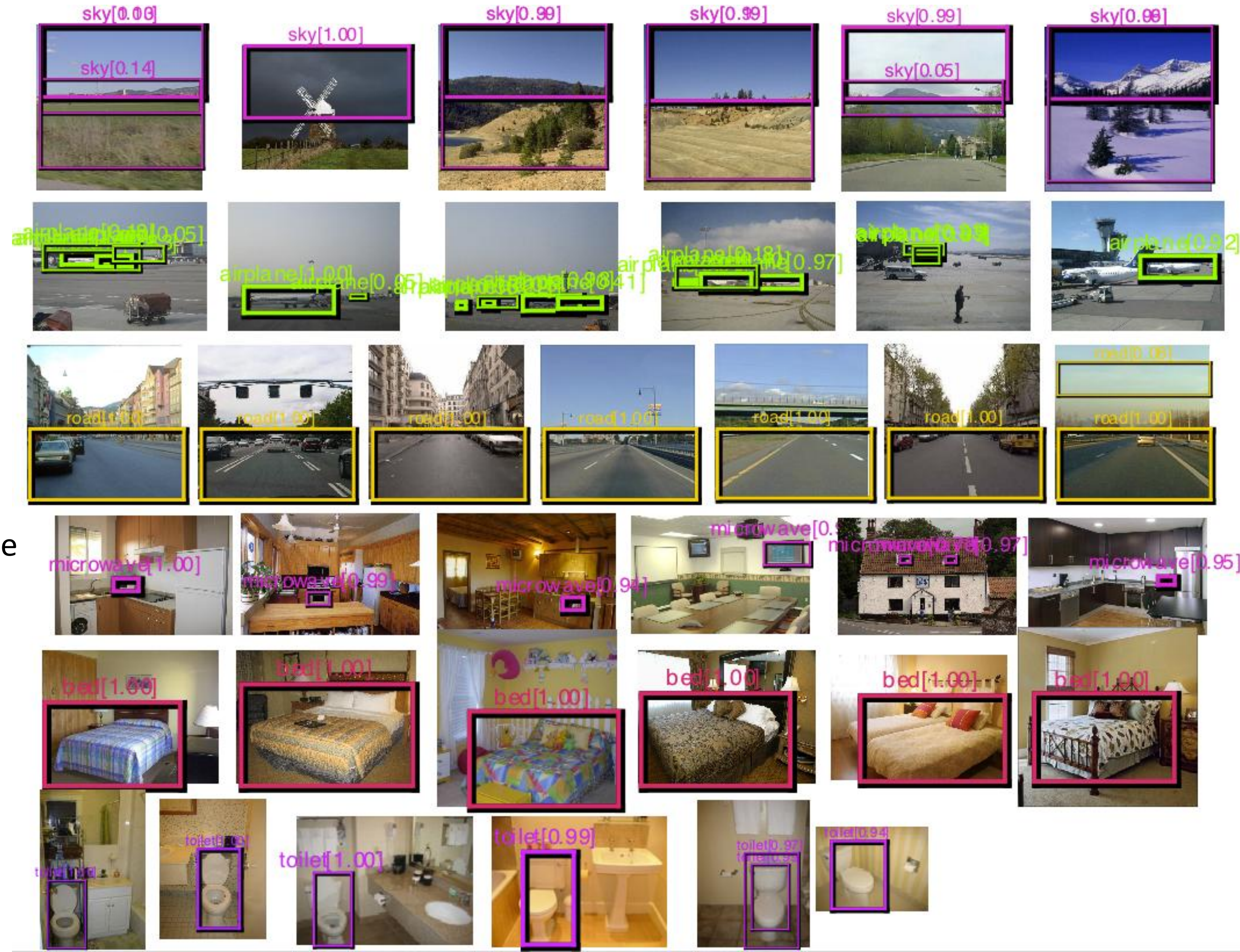
By looking at the output of a detector on a random set
of images, can you guess which object is it trying to detect?

The detector challenge



By looking at the output of a detector on a random set
of images, can you guess which object is it trying to detect?

The best objects/regions



Sky

Airplane

Road

Microwave

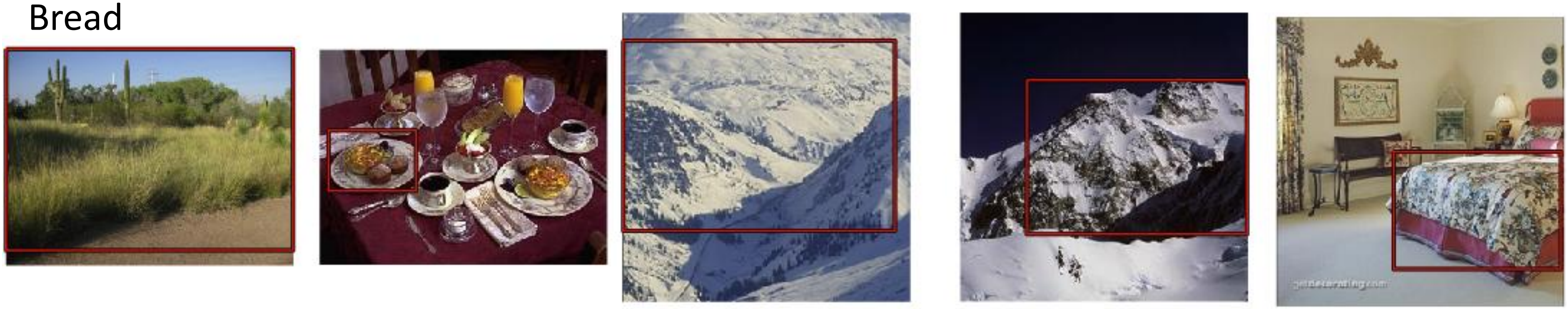
Bed

Toilet

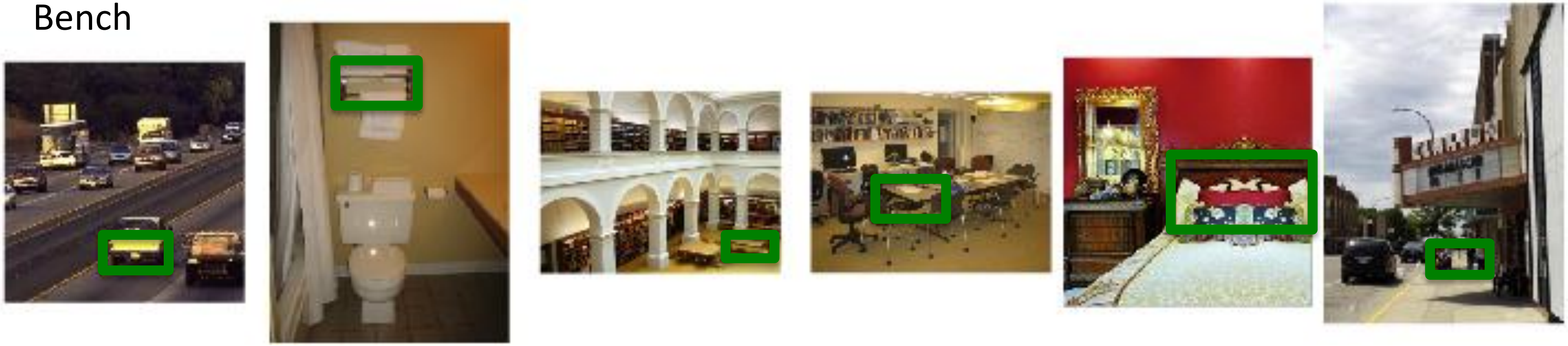
The worst objects/regions

Can you guess what are they trying to detect?

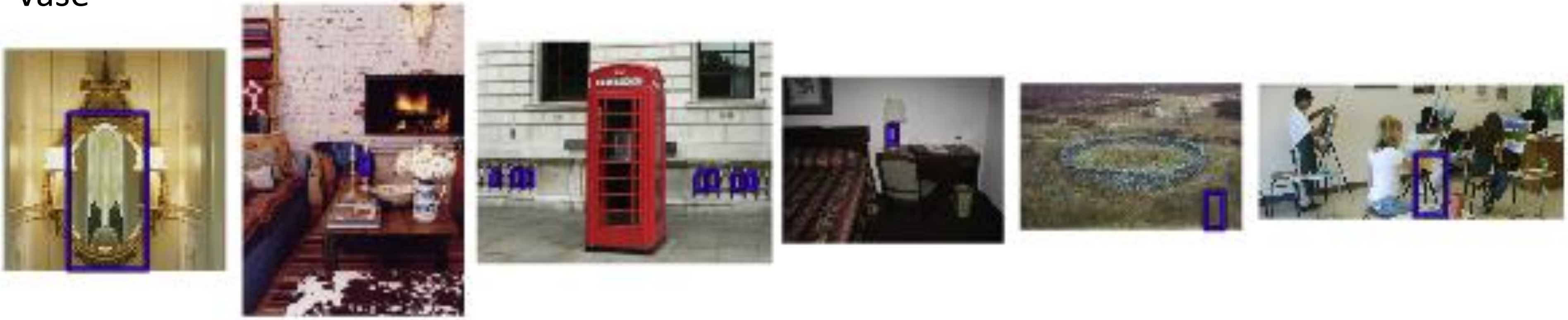
Bread

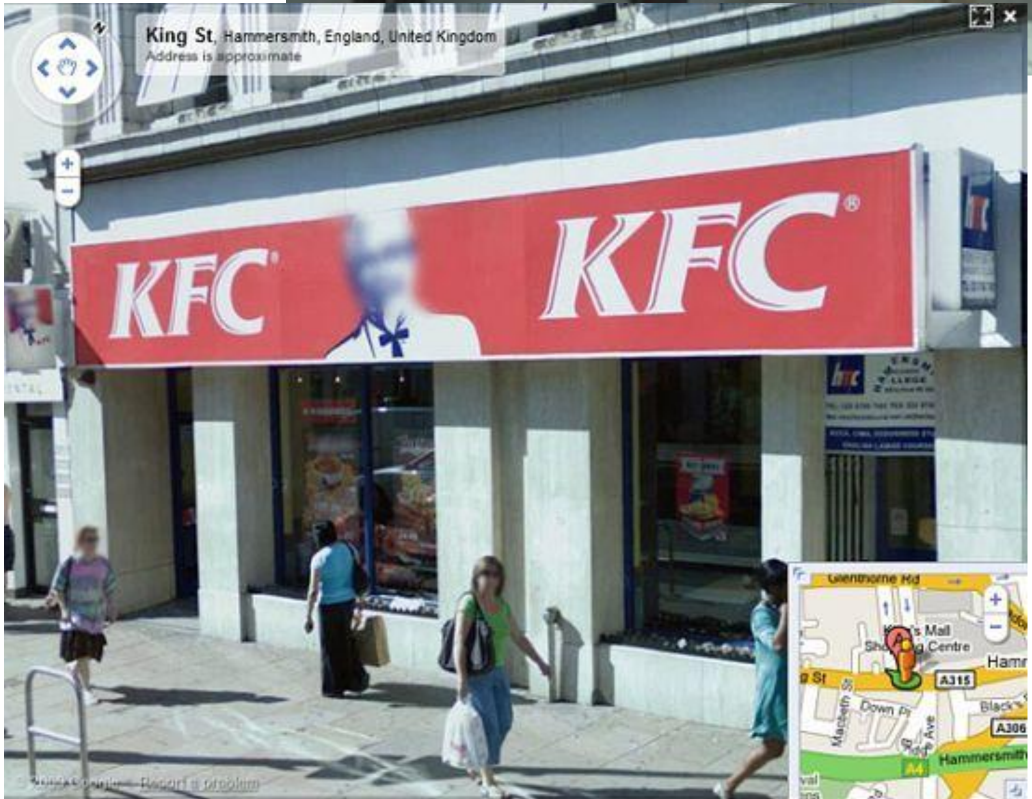


Bench



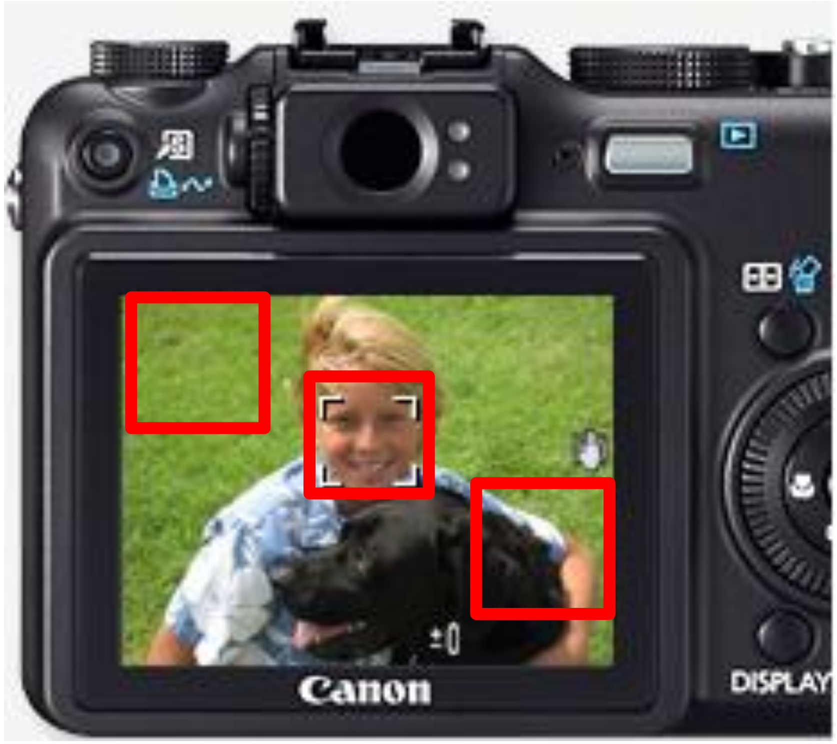
Vase





- The representation and matching of pictorial structures Fischler, Elschlager (1973).
- Face recognition using eigenfaces M. Turk and A. Pentland (1991).
- Human Face Detection in Visual Scenes - Rowley, Baluja, Kanade (1995)
- Graded Learning for Object Detection - Fleuret, Geman (1999)
- Robust Real-time Object Detection - Viola, Jones (2001)
- Feature Reduction and Hierarchy of Classifiers for Fast Object Detection in Video Images - Heisele, Serre, Mukherjee, Poggio (2001)
- ….

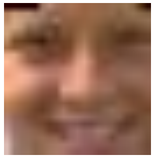
Face detection



Is this a face?

Is this a face?

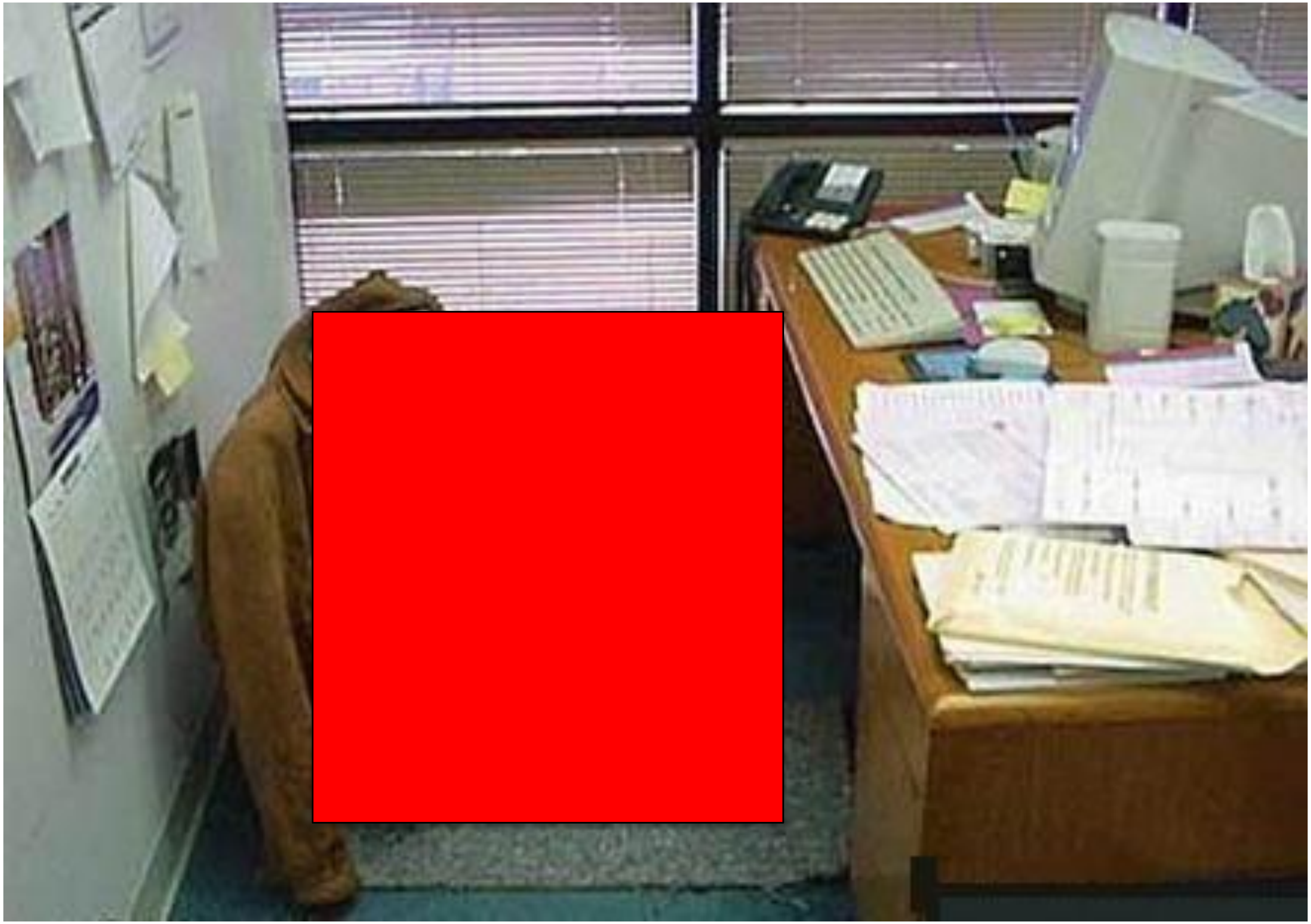
10000 more question later …

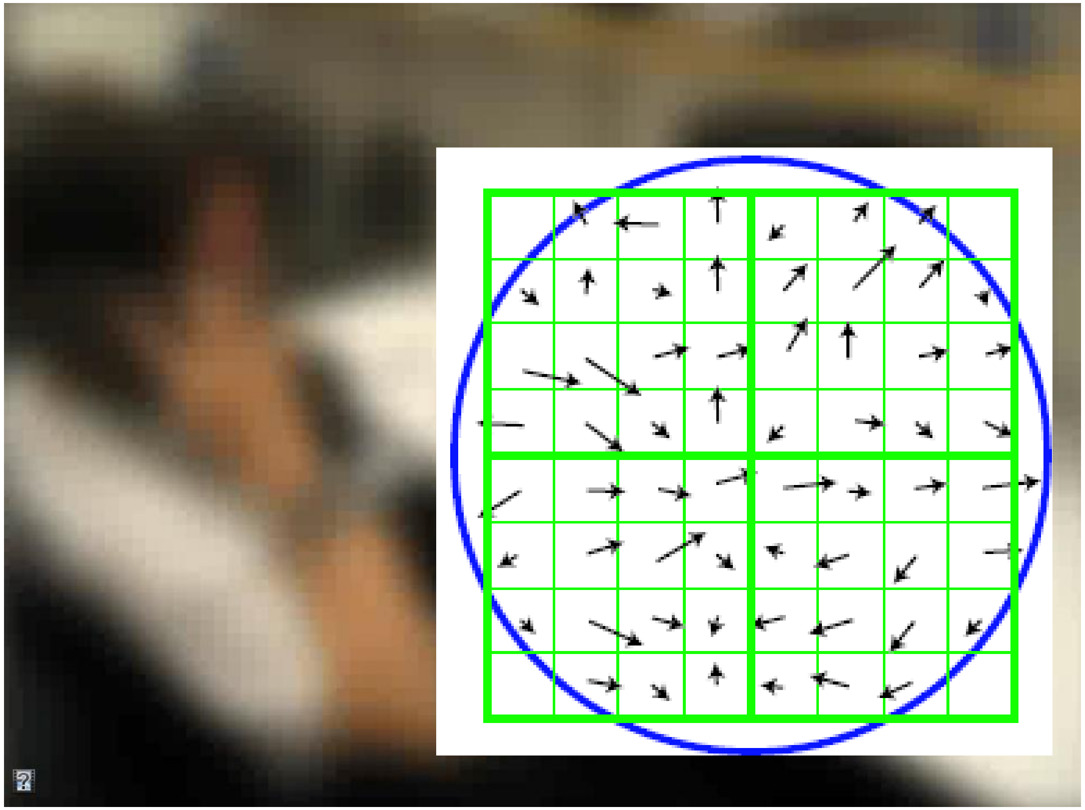


Is this a face?

10000 more questions

What object is hidden behind the red box?



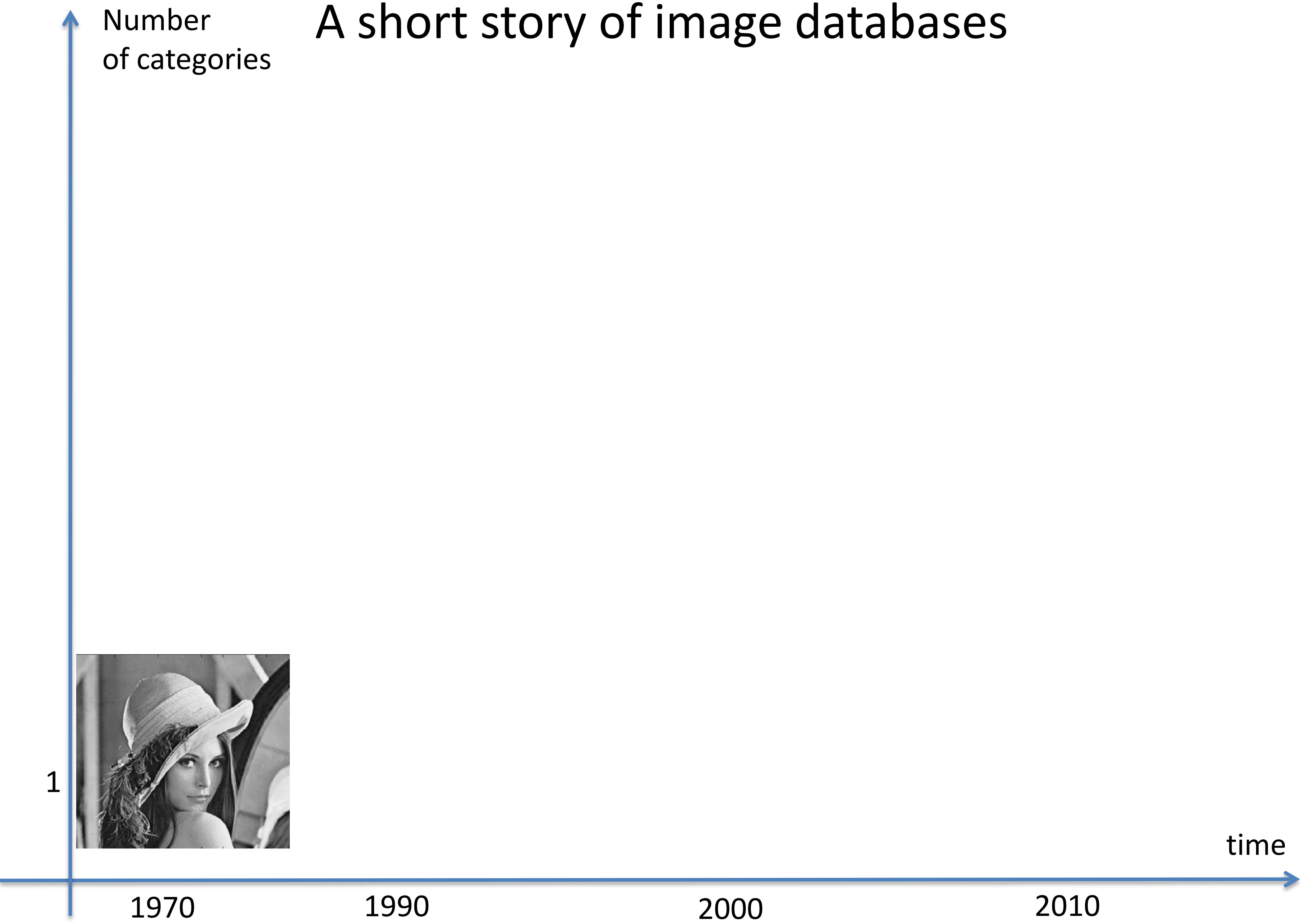


$10^{8\text{-}11}$

images

A short story of image databases

A short story of image databases



The Representation and Matching of Pictorial Structures

MARTIN A. FISCHLER AND ROBERT A. ELSCHLAGER

Abstract—The primary problem dealt with in this paper is the following. Given some description of a visual object, find that object in an actual photograph. Part of the solution to this problem is the

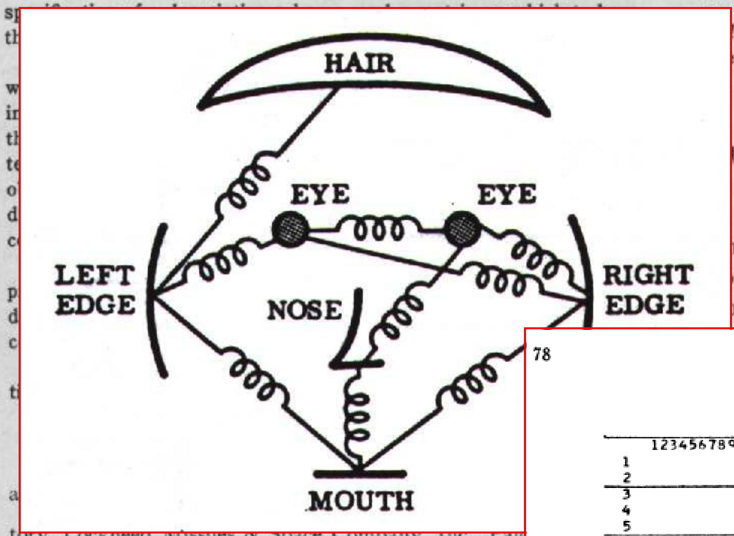
stereo compilation, and image change detection. In fact, the abili to describe, match, and register scenes is basic for almost a image processing task.

ex Terms—Dynamic programming, heuristic optimizatio description, picture matching, picture processing, represe

tory, Lockheed Missiles & Space Company, Inc., Pal
94304.

Introduction

HE PRIMARY PROBLEM dealt with in th paper is the following. Given some description a visual object, find that object in an actual phot . The object might be simple, such as a line, plicated, such as an ocean wave, and the descriptio e linguistic, pictorial, procedural, etc. The actu



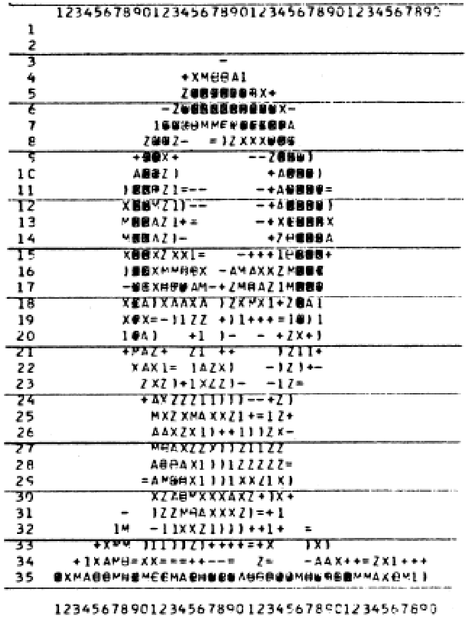
New York, N. Y., on February 15, 1932. He received the B.E.E. degree from the City College of New York, New York, in 1954 and the M.S. and Ph.D. degrees in electrical engineering from Stanford University, Stanford, Calif., in 1958 and 1962, respectively.

He served in the U. S. Army for two years and held positions at the National Bureau of Standards and at Hughes Aircraft Corporation during the period 1954 to 1958. In 1958 he joined the technical staff of the Lockheed Missiles & Space Company, Inc., at the Lockheed Palo Alto Research Laboratory, Palo Alto, Calif., and currently holds the title of Staff Scientist. He has conducted research and published in the areas of artificial intelligence, picture processing, switching theory, computer organization, and information theory.

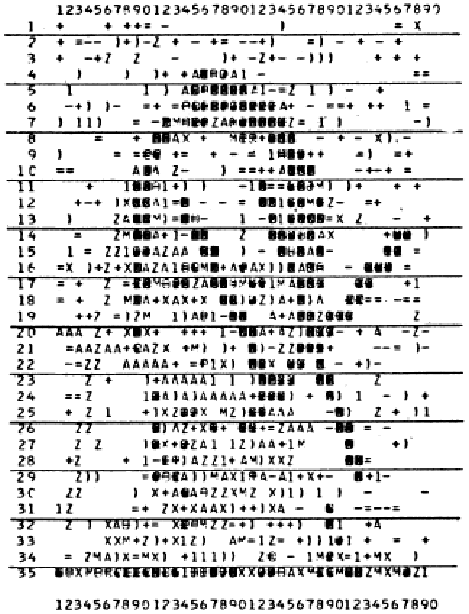
Dr. Fischler is a member of the Association for Computing Machinery, the Pattern Recognition Society, the Mathematical Association of America, Tau Beta Pi, and Eta Kappa Nu. He is currently an Associate Editor of the journal *Pattern Recognition* and is a past Chairman of the San Francisco Chapter of the IEEE Society on Systems, Man, and Cybernetics.

Robert A. Elschlager was born in Chicago, Ill., on May 25, 1943. He received the B.S. degree in mathematics from the University of Illinois, Urbana, in 1964, and the M.S. degree in mathematics from the University of California, Berkeley, in 1969.

as been an Associate
kheed Missiles & Space
Lockheed Palo Alto Re-
lto, Calif. His current
processing, operating
understanding.
merican Mathematical
erica, and the Associa-



Original picture.



Noisy picture (sensed scene) as used in experiment.

A short story of image databases

Number
of categories

1

COIL-20
1996.

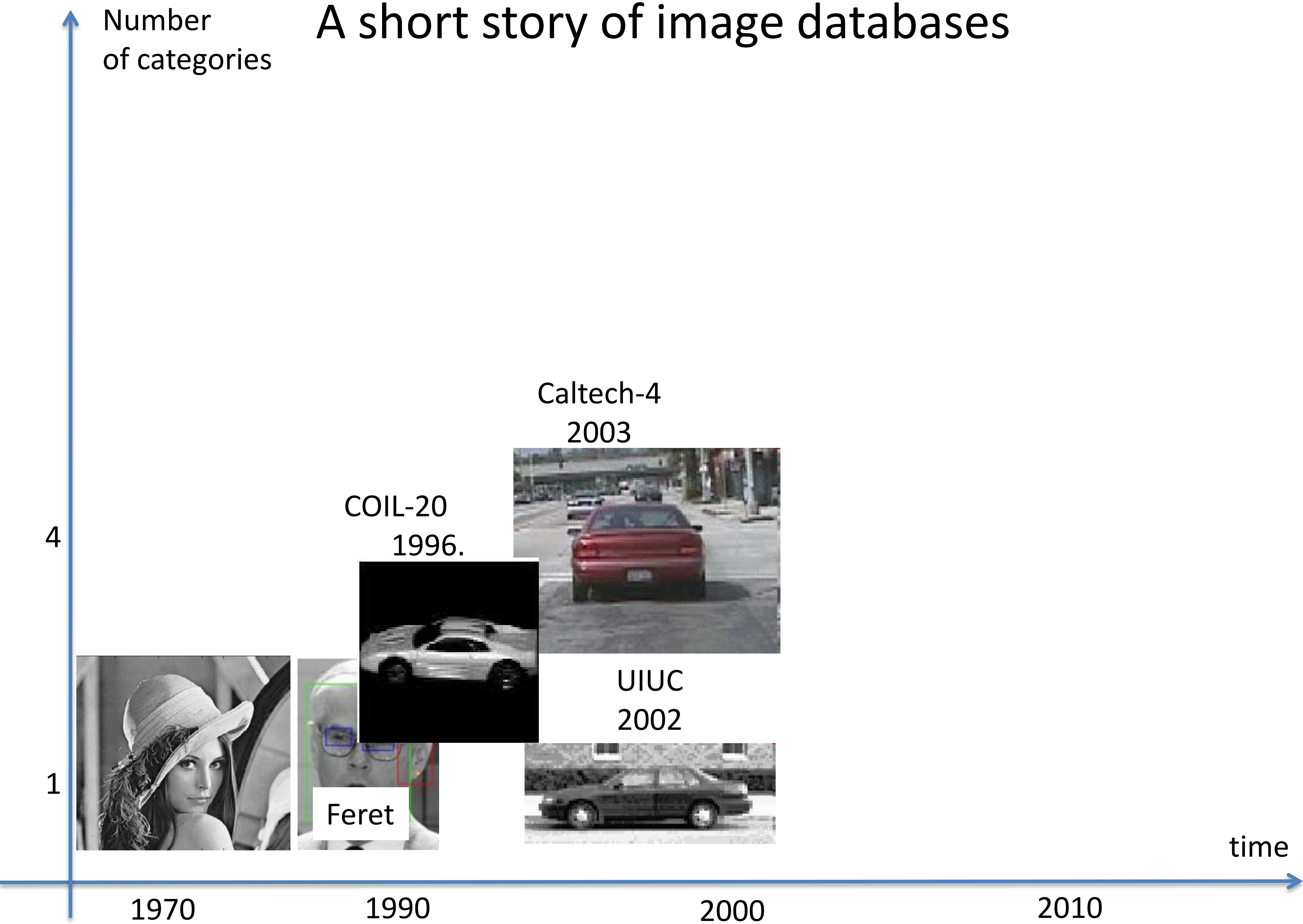
Feret

UIUC
2002

time

1970 1990 2000 2010

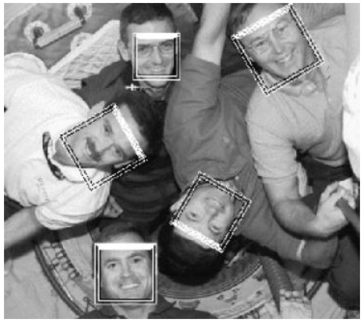
A short story of image databases

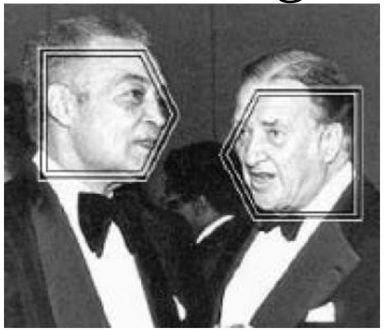


Multiclass object detection

Using a set of independent binary classifiers was a common strategy:

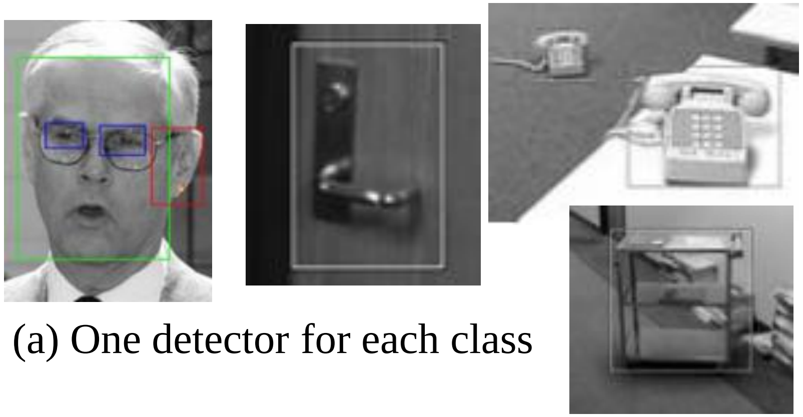
- Viola-Jones extension for dealing with rotations



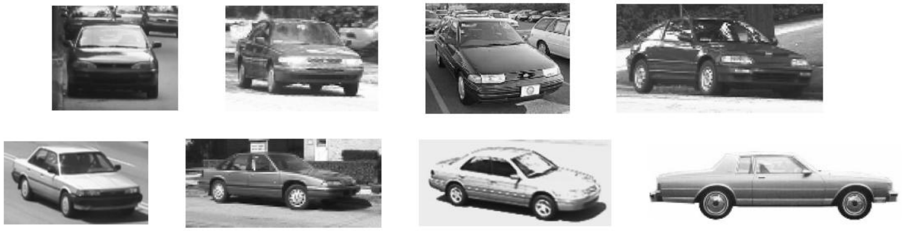


- two cascades for each view

- Schneiderman-Kanade multiclass object detection

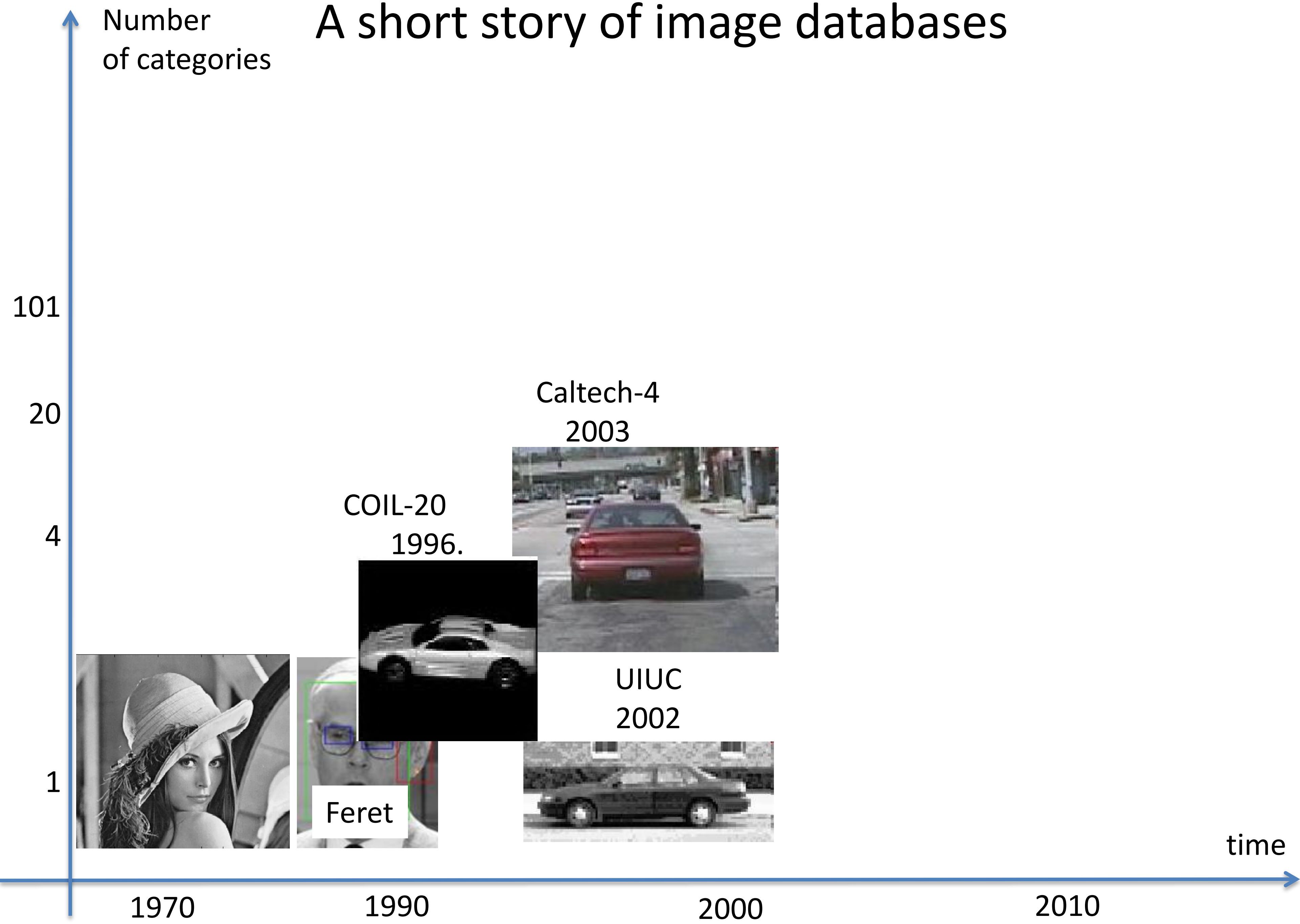


(a) One detector for each class



(b) For cars, classifiers are trained on 8 viewpoints

A short story of image databases



Number of categories

101

20

4

1

Caltech-4 2003

COIL-20 1996.

UIUC 2002

Feret

time

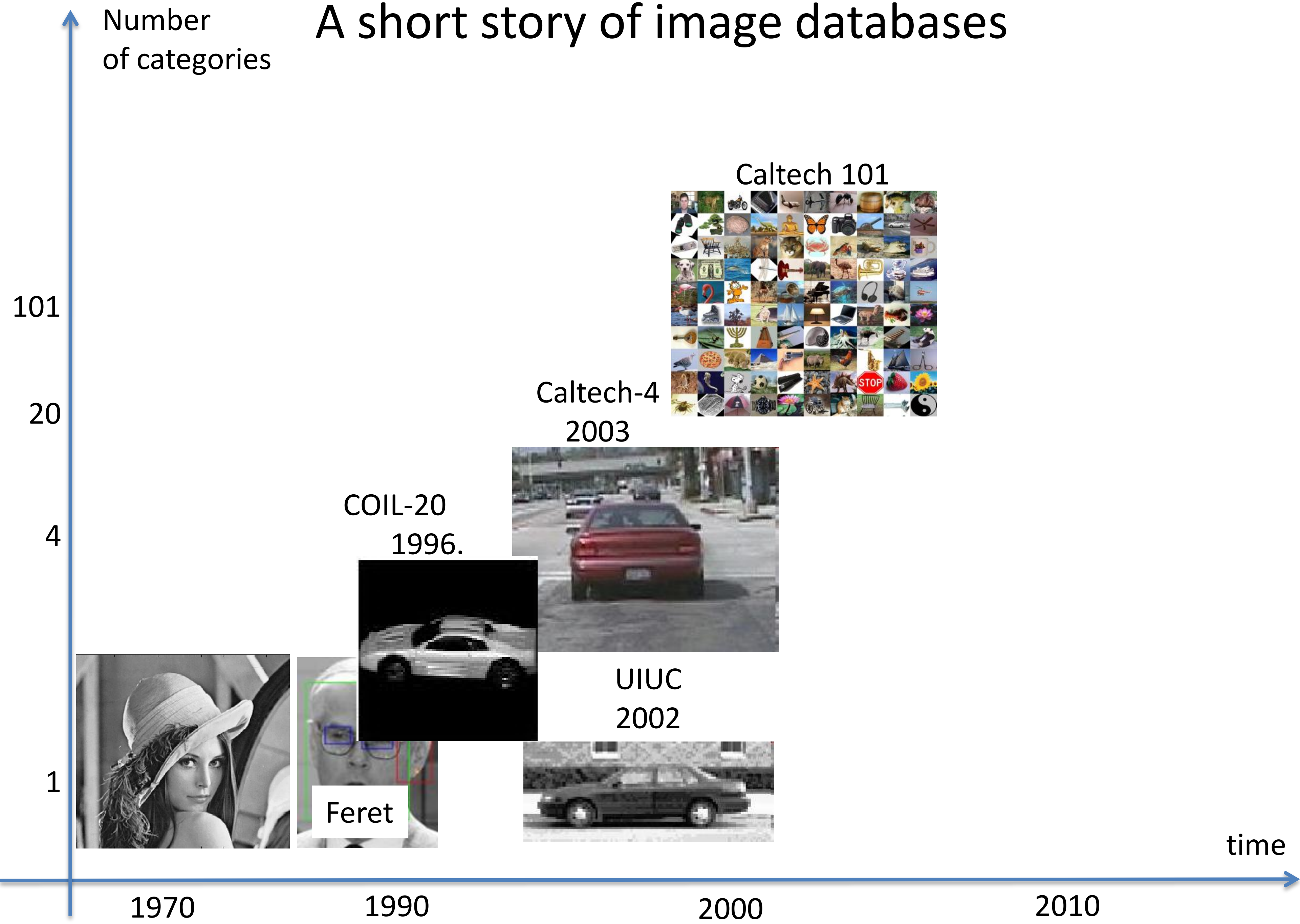
1970

1990

2000

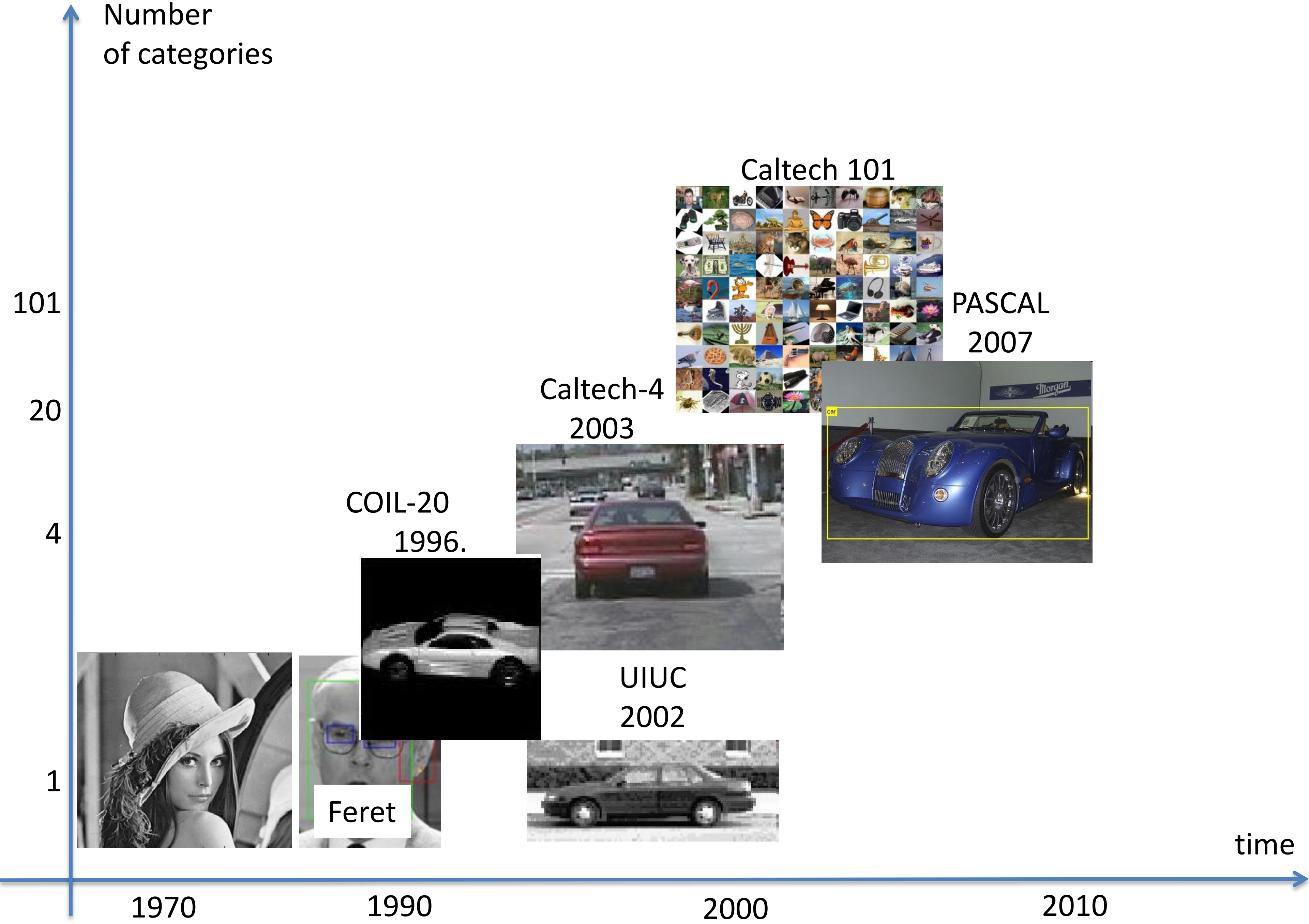
2010

A short story of image databases



- Sharing global model parameters
  - L. Fei-Fei, R. Fergus, and P. Perona, 2004
  - T. Deselaers, B. Alexe, and V. Ferrari, 2010
- Sharing parts
  - S. Krempp, D. Geman, and Y. Amit, 2002
  - A. Torralba, K. P. Murphy, and W. T. Freeman, 2004
  - E. Bart and S. Ullman, 2005
  - A. Opelt, A. Pinz, and A. Zisserman, 2006
  - E. Bart, I. Porteous, P. Perona, and M. Welling, 2008
  - E. Sudderth, A. Torralba, W. T. Freeman, and W. Willsky, 2005
  - S. Fidler, M. Boben, and A. Leonardis, 2009
- Sharing attributes
  - C. Lampert, H. Nickisch, and S. Harmeling, 2009
- Sharing transformations
  - E. Miller, N. Matsakis, and P. Viola, 2000
- Sharing classifier parameters
  - B. Shahbaba and R. M. Neal, 2007
  - M. Marszalek and C. Schmid, 2007
  - A. Quattoni,M. Collins, and T. Darrell, 2008
  - R. Fergus, H. Bernal, Y. Weiss, and A. Torralba, 2010
  - T. Tommasi, F. Orabona, and B. Caputo, 2010

…

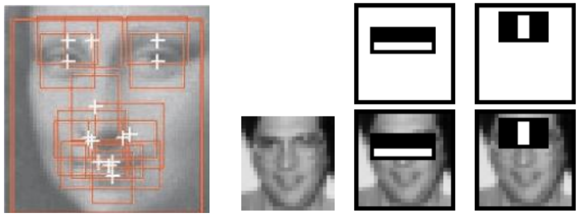

Number
of categories
101
20
4
1
Caltech 101
PASCAL
2007
Caltech-4
2003
COIL-20
1996.
UIUC
2002
Feret
time
1970
1990
2000
2010

Back to isolated models

Bag of words models

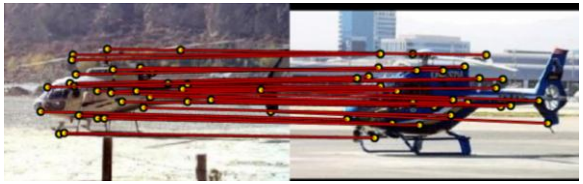
Csurka, Dance, Fan, Willamowski, and Bray 2004
Sivic, Russell, Freeman, Zisserman, ICCV 2005

Voting models



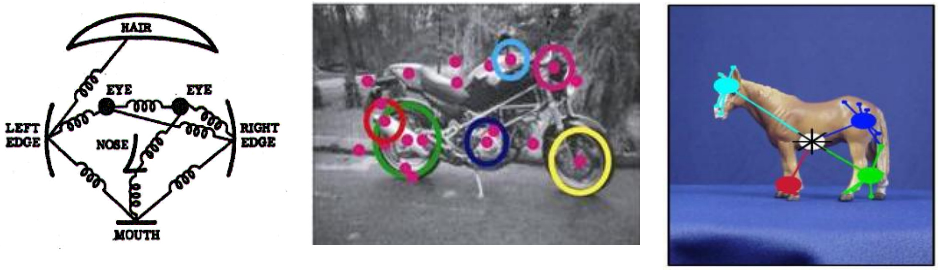
Viola and Jones, ICCV 2001
Heisele, Poggio, et. al., NIPS 01
Schneiderman, Kanade 2004
Vidal-Naquet, Ullman 2003

Shape matching Deformable models



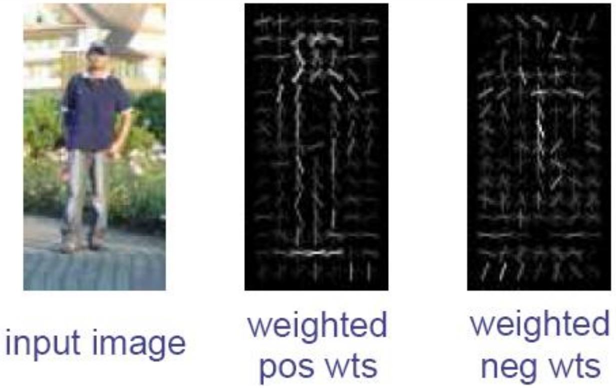
Berg, Berg, Malik, 2005
Cootes, Edwards, Taylor, 2001

Constellation models

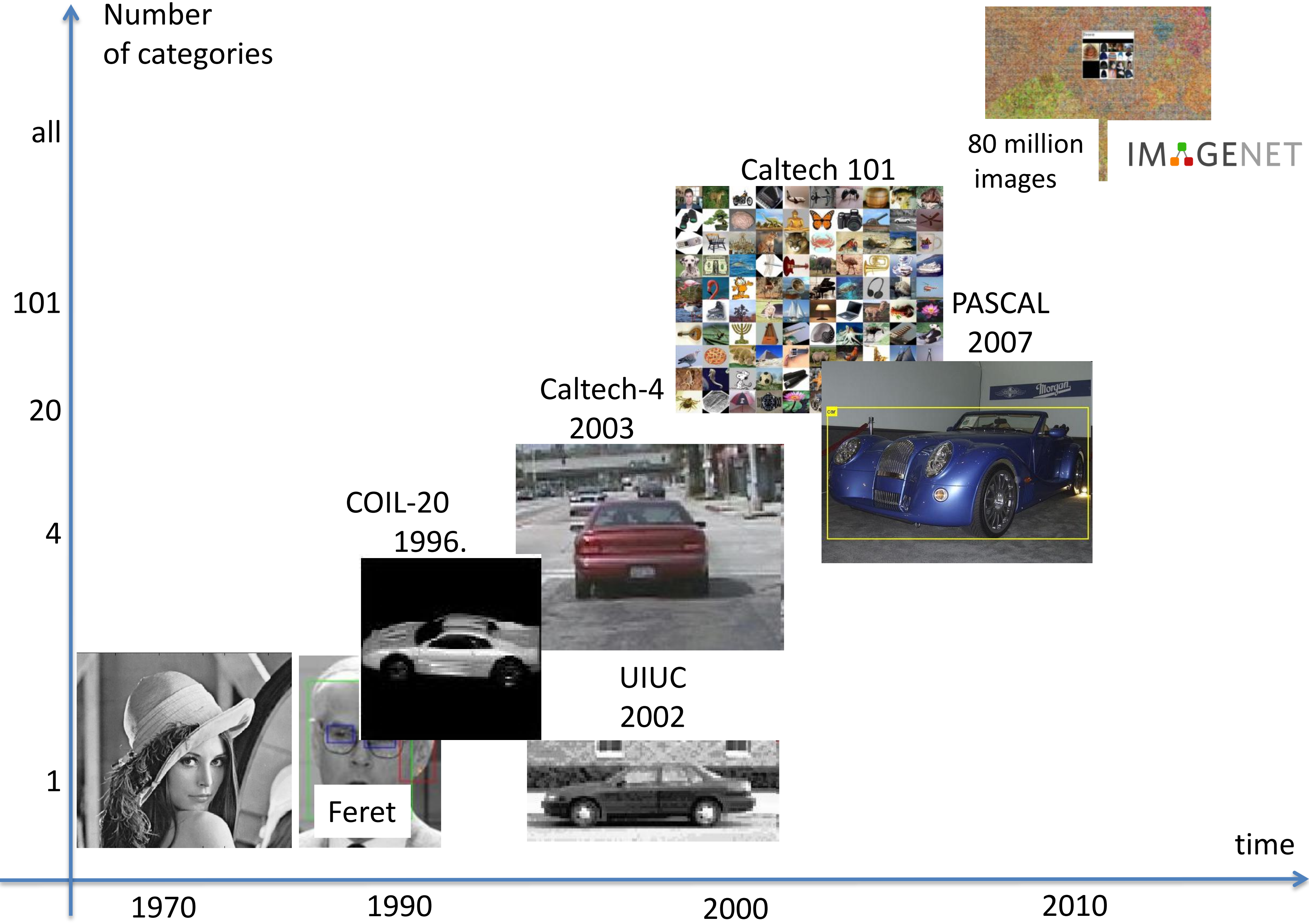


Fischler and Elschlager, 1973
Burl, Leung, and Perona, 1995
Weber, Welling, and Perona, 2000
Fergus, Perona, & Zisserman, CVPR 2003

Rigid template models



Sirovich and Kirby 1987
Turk, Pentland, 1991
Dalal & Triggs, 2006
Felzenszwalb, McAllester & Ramanan, 2008



Number
of categories

all

101

20

4

1

Feret

COIL-20
1996.

UIUC
2002

Caltech-4
2003

Caltech 101

PASCAL
2007

80 million
images

IMAGENET

time

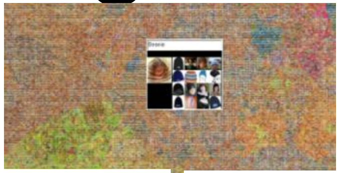
1970

1990

2000

2010

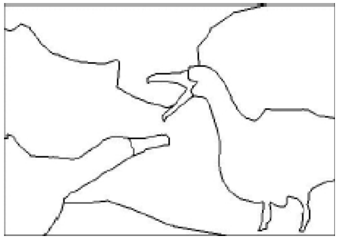
Big data collection efforts



80 million images

IMAGENET

Berkeley segmentation database



Caltech 101

SUN database

Pascal

Caltech-4

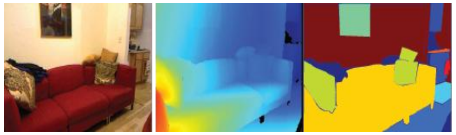
UIUC

UIUC

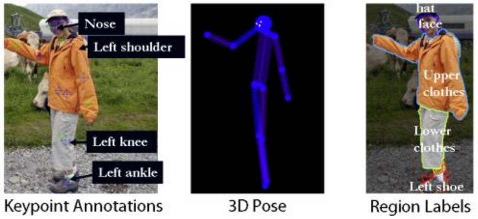
Attributes database

Has Horn
Has leg
Has Head
Has Wool

NYU Depth Dataset



H3D Dataset



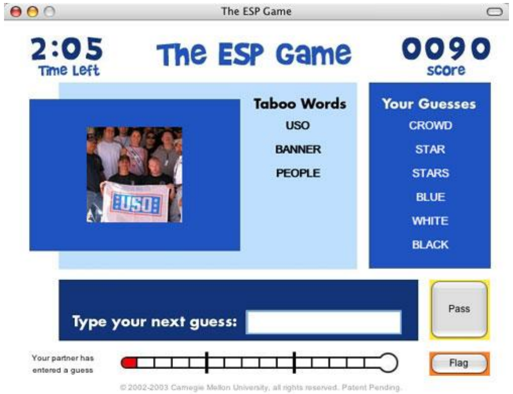
Segments ↔ Framed objects ↔ Scenes ↔ Parts & attributes ↔ 3D ↔ ?

A short history of image annotation

Labeling to get a Ph.D.

Labeling for fun

Luis Von Ahn and Laura Dabbish 2004



Labeling for money

(Sorokin, Forsyth, 2008)

amazonmechanical turk
beta
Artificial Artificial Intelligence

Labeling because it gives you added value

Just for labeling

(Russell et al 2005)

LabelMe

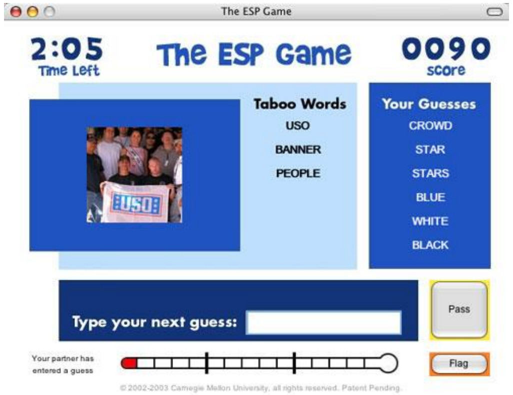
Visipedia

(Belongie, Perona, et al, 2011)

A short history of image annotation

Labeling for fun

Luis Von Ahn and Laura Dabbish 2004



The ESP Game

2:05 Time Left

The ESP Game

0090 score

Taboo Words
USO
BANNER
PEOPLE

Your Guesses
CROWD
STAR
STARS
BLUE
WHITE
BLACK

Type your next guess:

Your partner has entered a guess

© 2002-2003 Carnegie Mellon University, all rights reserved. Patent Pending.

Labeling to get a Ph.D.

Labeling for money

(Sorokin, Forsyth, 2008)

amazonmechanical turk
beta
Artificial Artificial Intelligence

Just for labeling

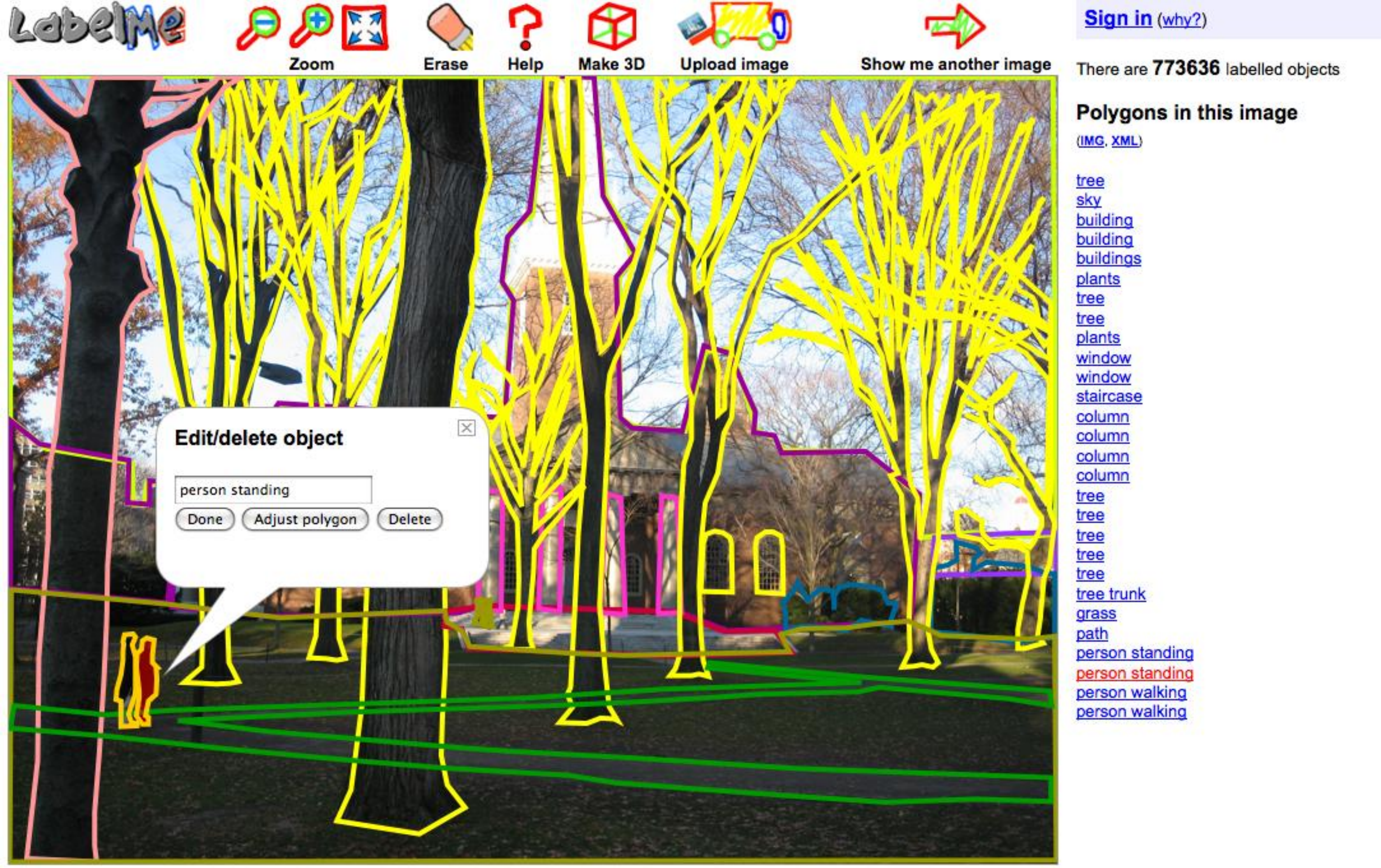
(Russell et al 2005)

LabelMe

Labeling because it gives you added value

Visipedia

(Belongie, Perona, et al, 2011)



There are 773636 labelled objects

Polygons in this image

(IMG, XML)

tree
sky
building
building
buildings
plants
tree
tree
plants
window
window
staircase
column
column
column
column
tree
tree
tree
tree
tree
tree trunk
grass
path
person standing
person standing
person walking
person walking

Edit/delete object

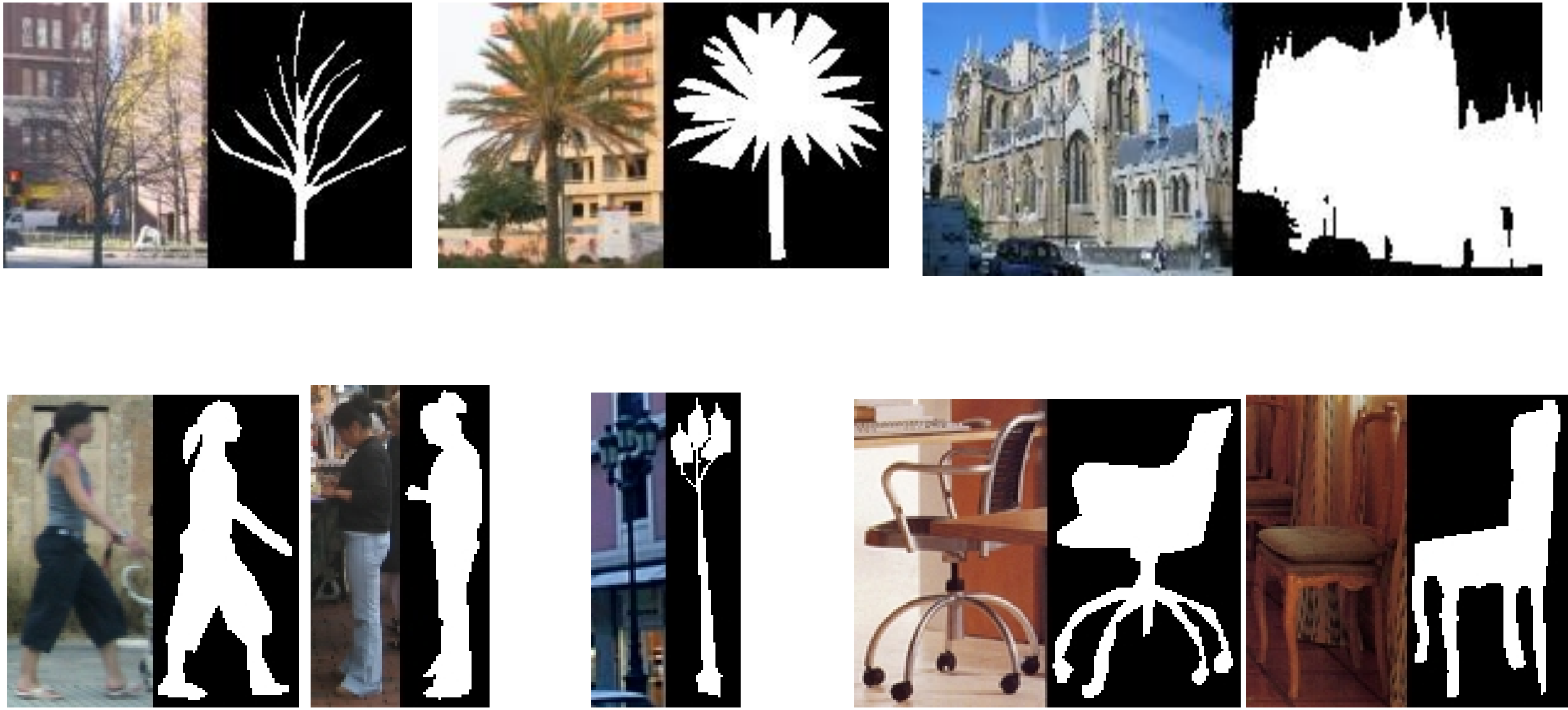
person standing

Tool went online July 1st, 2005

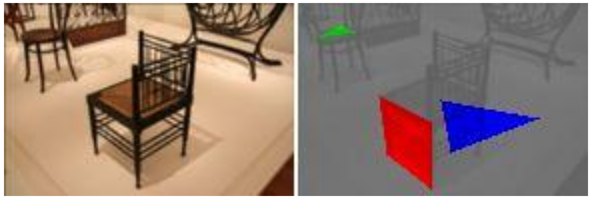
Labelme.csail.mit.edu

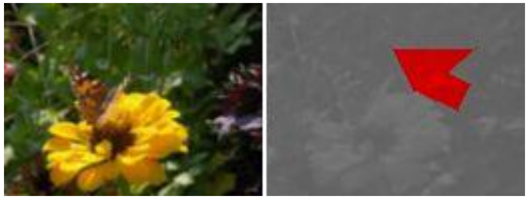
B. Russell, A. Torralba, K. Murphy, W.T. Freeman. IJCV 2008

Extreme labeling



Testing





Most common labels:

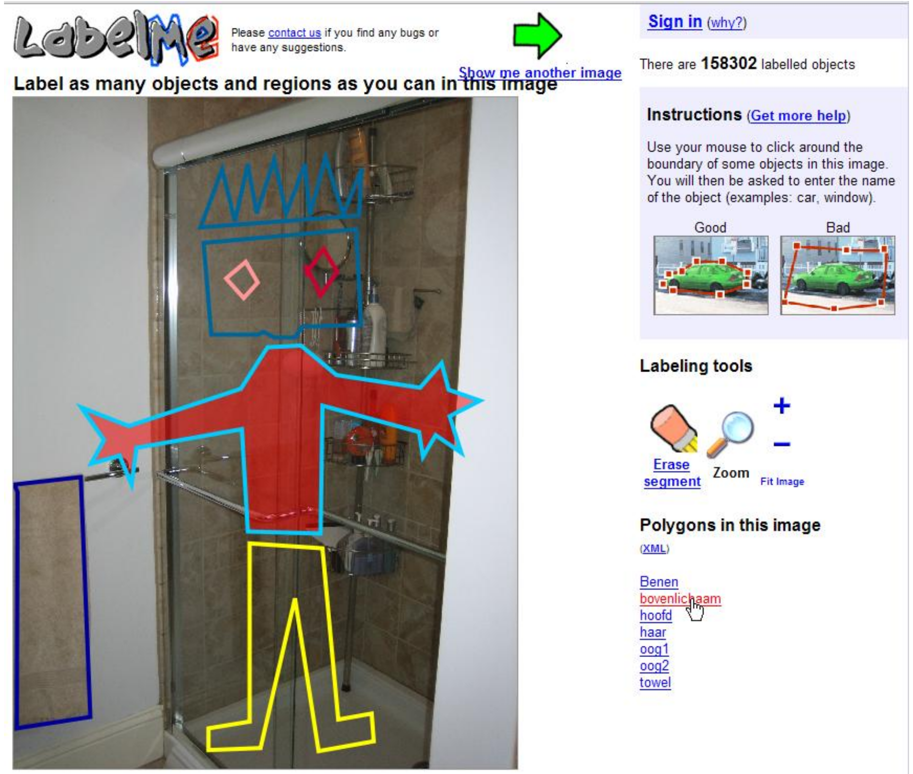
test

adksdsa

woiieiie

…

Do not try this at home



Please contact us if you find any bugs or have any suggestions.

Label as many objects and regions as you can in this image

Sign in (why?)

There are **158302** labelled objects

Instructions (Get more help)

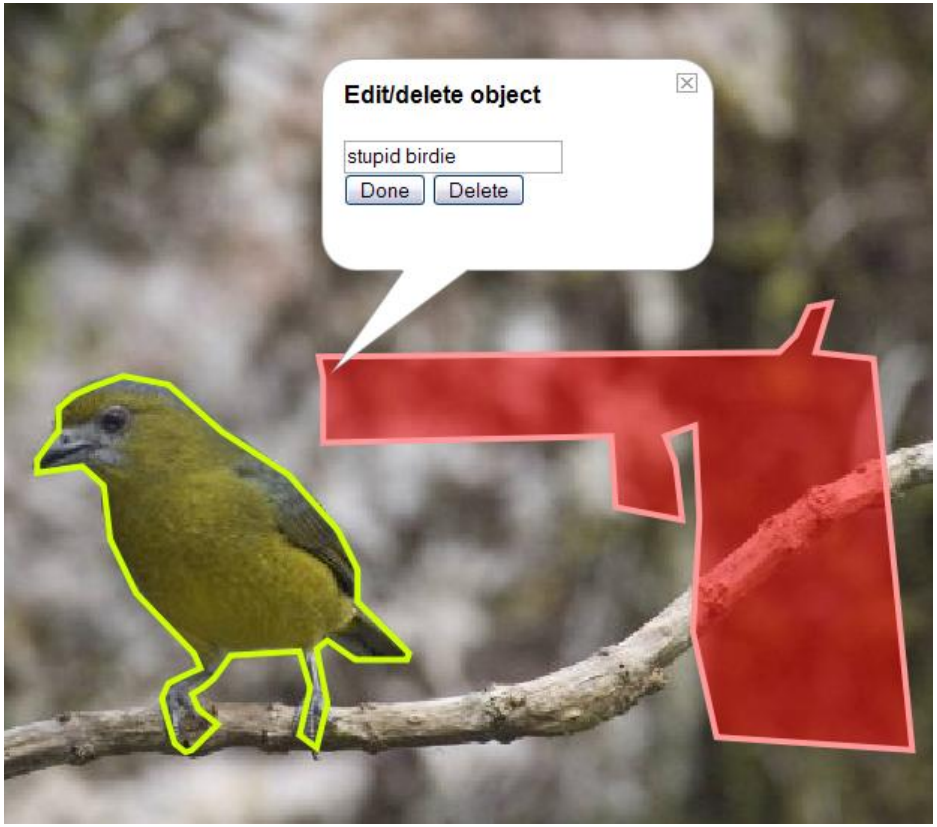
Use your mouse to click around the boundary of some objects in this image. You will then be asked to enter the name of the object (examples: car, window).

Labeling tools

Polygons in this image

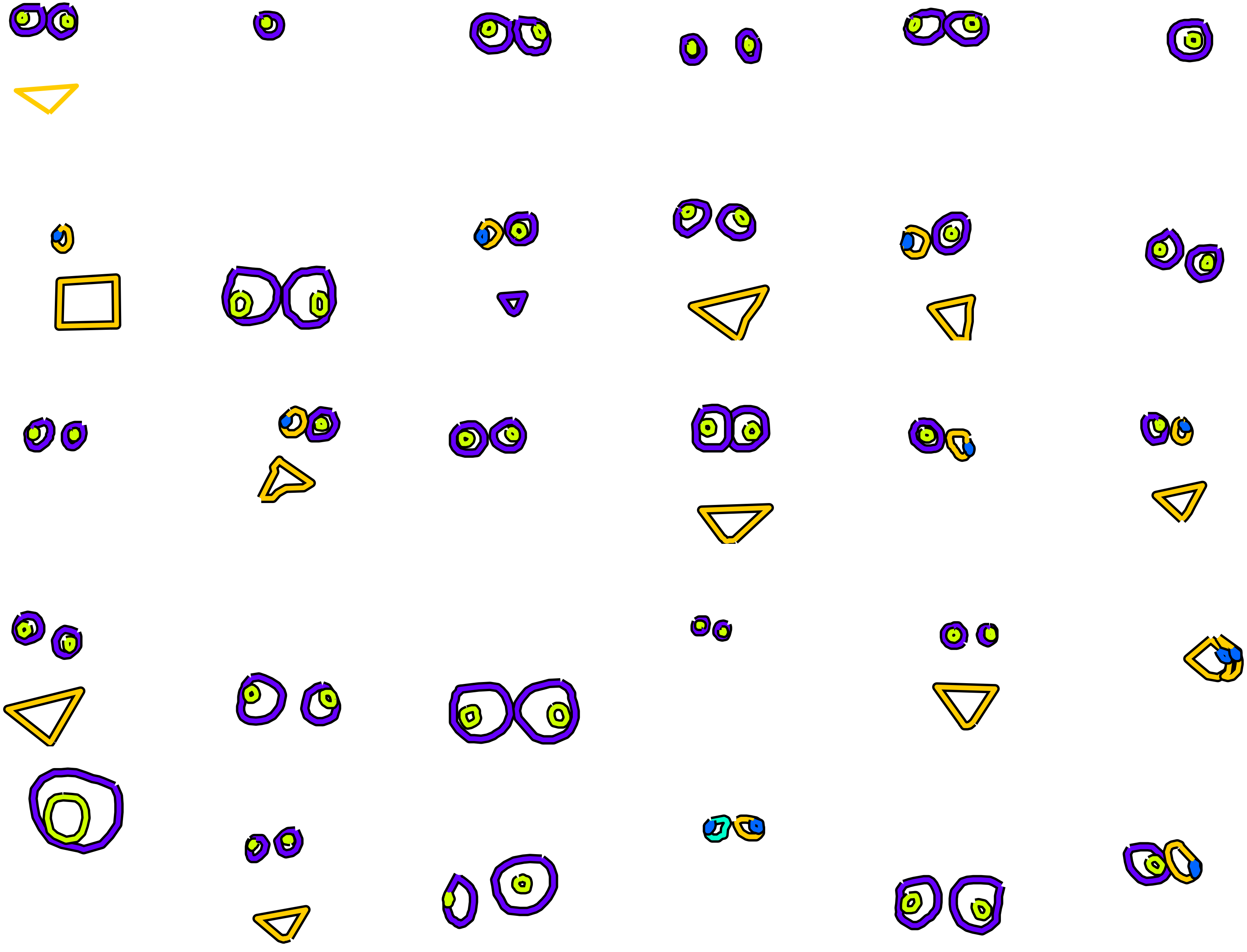
(XML)

Benen
bovenlichaam
hoofd
haar
oog1
oog2
towel



Edit/delete object

stupid birdie

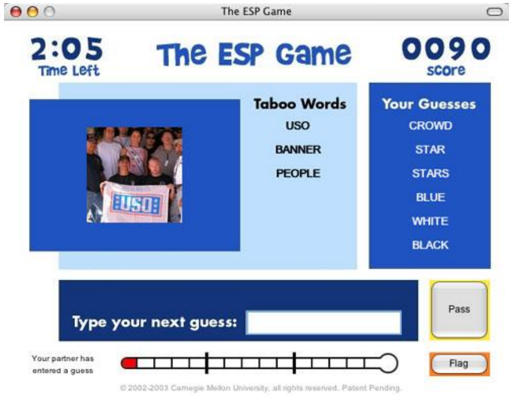


…and many more images

A short history of image annotation

Labeling for fun

Luis Von Ahn and Laura Dabbish 2004



Labeling to get a Ph.D.

Labeling for money

(Sorokin, Forsyth, 2008)

amazonmechanical turk

Artificial Artificial Intelligence

beta

Just for labeling

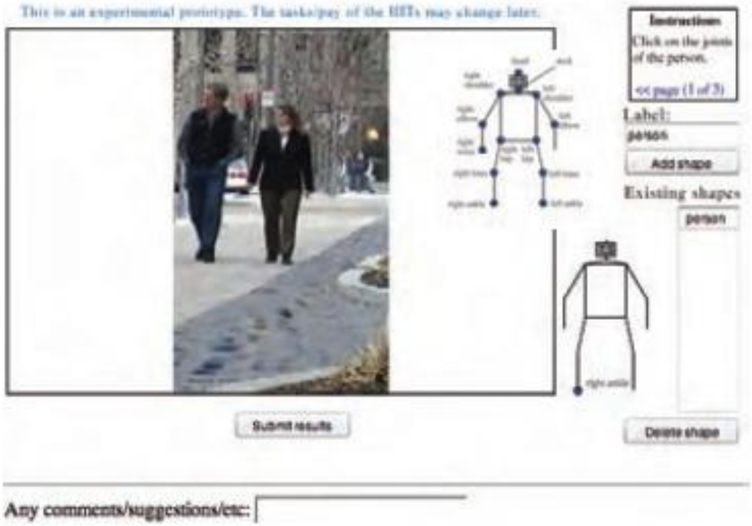
(Russell et al 2005)

LabelMe

Labeling because it gives you added value

Visipedia

(Belongie, Perona, et al, 2011)



This is an experimental prototype. The tasks/pay of the HITs may change later.

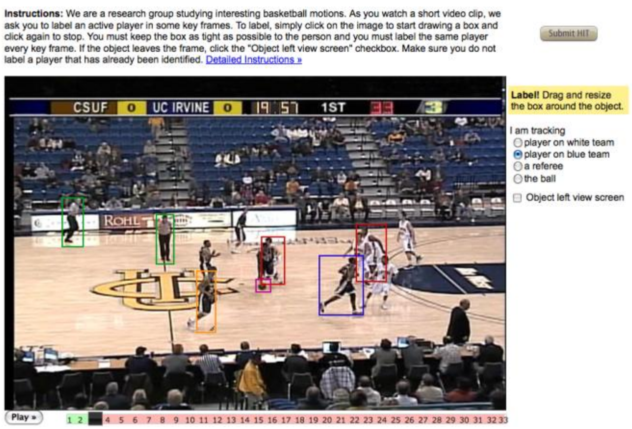
Instruction
Click on the joints of the person.
<< page (1 of 3)

Label:
person

Existing shapes
person

Any comments/suggestions/etc:

Sorokin, Forsyth, 2008



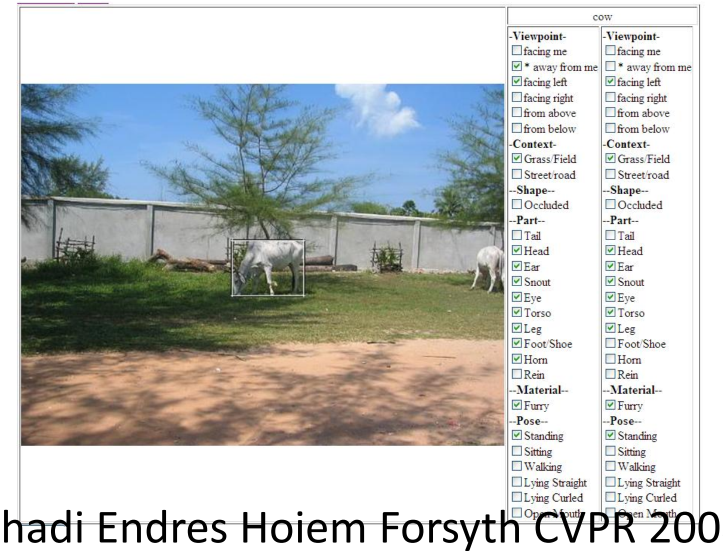
Instructions: We are a research group studying interesting basketball motions. As you watch a short video clip, we ask you to label an active player in some key frames. To label, simply click on the image to start drawing a box and click again to stop. You must keep the box as tight as possible to the person and you must label the same player every key frame. If the object leaves the frame, click the "Object left view screen" checkbox. Make sure you do not label a player that has already been identified. Detailed Instructions »

Label! Drag and resize the box around the object.

I am tracking
○ player on white team
● player on blue team
○ a referee
○ the ball

□ Object left view screen

Carl Vondrick, Deva Ramanan, Don Patterson

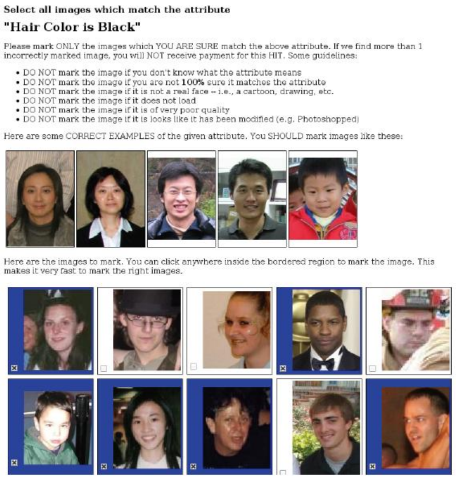


cow

-Viewpoint-
□ facing me
☑ * away from me
☑ facing left
□ facing right
□ from above
□ from below
-Context-
☑ Grass/Field
□ Street/road
--Shape--
□ Occluded
--Part--
□ Tail
☑ Head
☑ Ear
☑ Snout
☑ Eye
☑ Torso
☑ Leg
☑ Foot/Shoe
☑ Horn
□ Rein
--Material--
☑ Furry
--Pose--
☑ Standing
□ Sitting
□ Walking
□ Lying Straight
□ Lying Curled
□ Open Mouth

-Viewpoint-
□ facing me
□ * away from me
☑ facing left
□ facing right
□ from above
□ from below
-Context-
☑ Grass/Field
□ Street/road
--Shape--
□ Occluded
--Part--
□ Tail
☑ Head
☑ Ear
☑ Snout
☑ Eye
☑ Torso
☑ Leg
□ Foot/Shoe
□ Horn
□ Rein
--Material--
☑ Furry
--Pose--
☑ Standing
□ Sitting
□ Walking
□ Lying Straight
□ Lying Curled
□ Open Mouth

Farhadi Endres Hoiem Forsyth CVPR 2008



Select all images which match the attribute

"Hair Color is Black"

Please mark ONLY the images which YOU ARE SURE match the above attribute. If we find more than 1 incorrectly marked image, you will NOT receive payment for this HIT. Some guidelines:

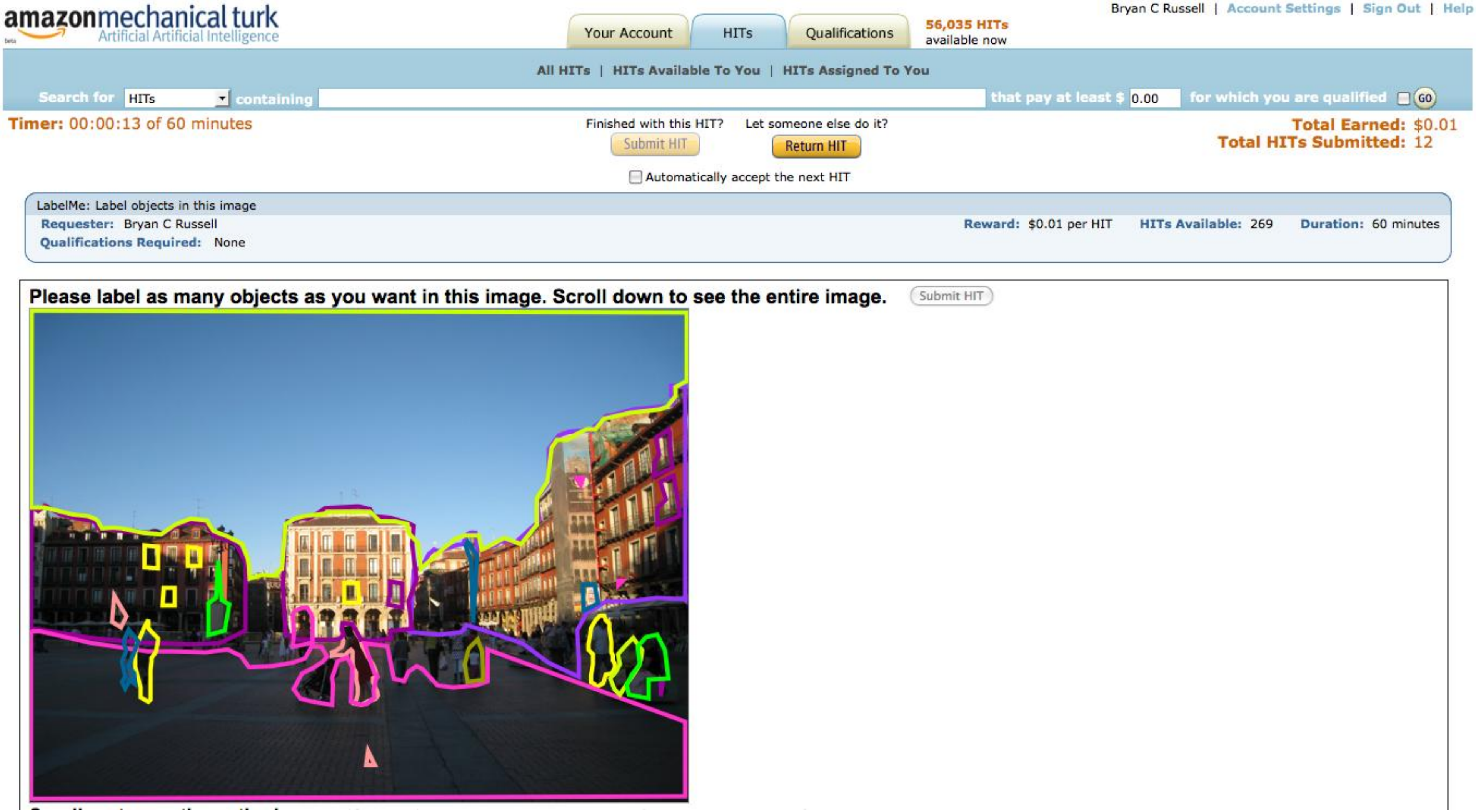
- DO NOT mark the image if you don't know what the attribute means
- DO NOT mark the image if you are not **100%** sure it matches the attribute
- DO NOT mark the image if it is not a real face -- i.e., a cartoon, drawing, etc.
- DO NOT mark the image if it does not load
- DO NOT mark the image if it is of very poor quality
- DO NOT mark the image if it is looks like it has been modified (e.g. Photoshopped)

Here are some CORRECT EXAMPLES of the given attribute. You SHOULD mark images like these:

Here are the images to mark. You can click anywhere inside the bordered region to mark the image. This makes it very fast to mark the right images.

N. Kumar, A. C. Berg,
P. N. Belhumeur, and S. K. Nayar, ICCV 2009

And many more…



amazonmechanical turk
Artificial Artificial Intelligence
Your Account
HITs
Qualifications
56,035 HITs available now
Bryan C Russell | Account Settings | Sign Out | Help
All HITs | HITs Available To You | HITs Assigned To You
Search for HITs containing that pay at least $ 0.00 for which you are qualified □ GO

Timer: 00:00:13 of 60 minutes

Finished with this HIT? Submit HIT

Let someone else do it? Return HIT

□ Automatically accept the next HIT

Total Earned: $0.01

Total HITs Submitted: 12

LabelMe: Label objects in this image

Requester: Bryan C Russell

Reward: $0.01 per HIT

HITs Available: 269

Duration: 60 minutes

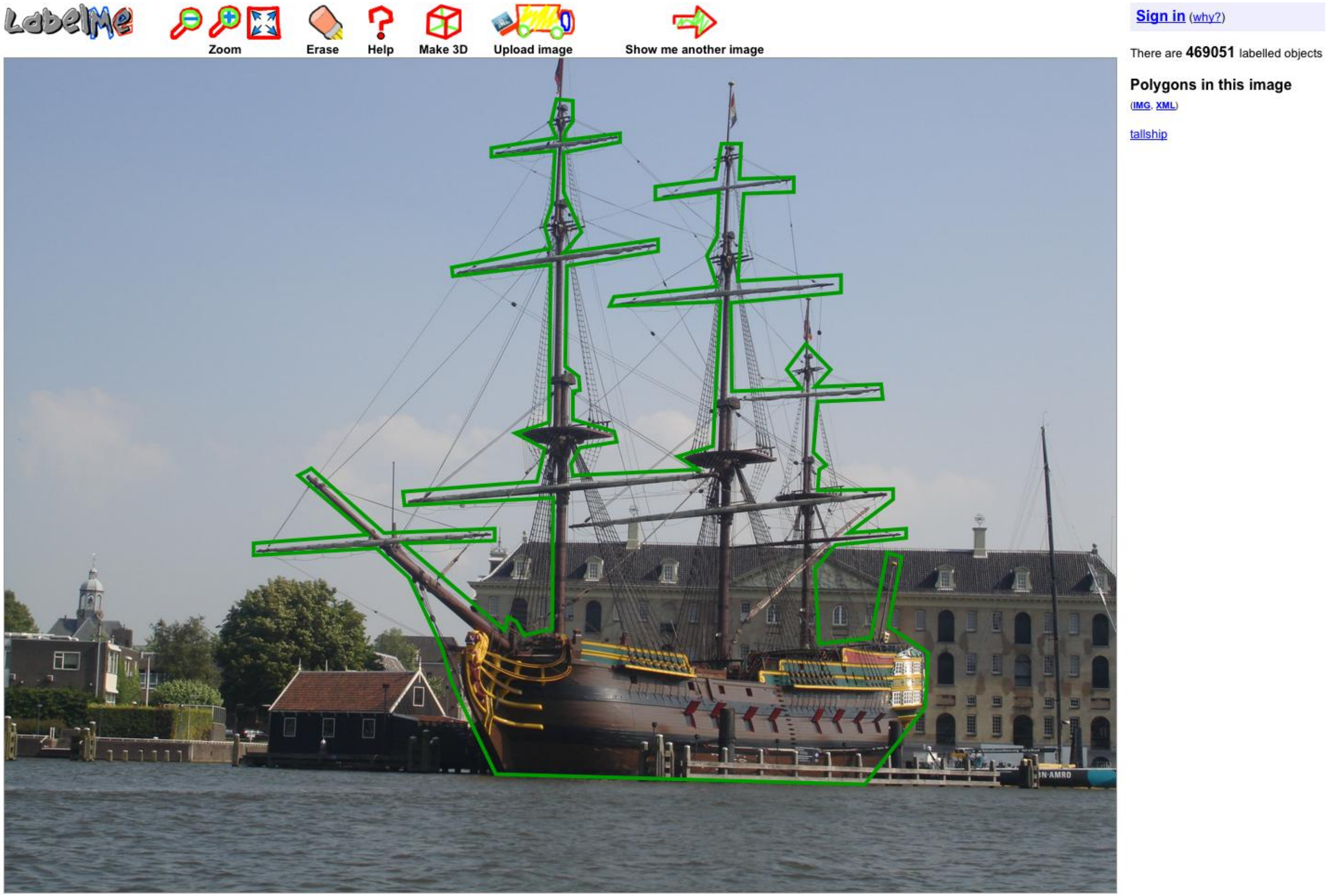
Qualifications Required: None

Please label as many objects as you want in this image. Scroll down to see the entire image.

With Bryan Russell

1 cent

Task: Label one object in this image



Sign in (why?)

There are 469051 labelled objects

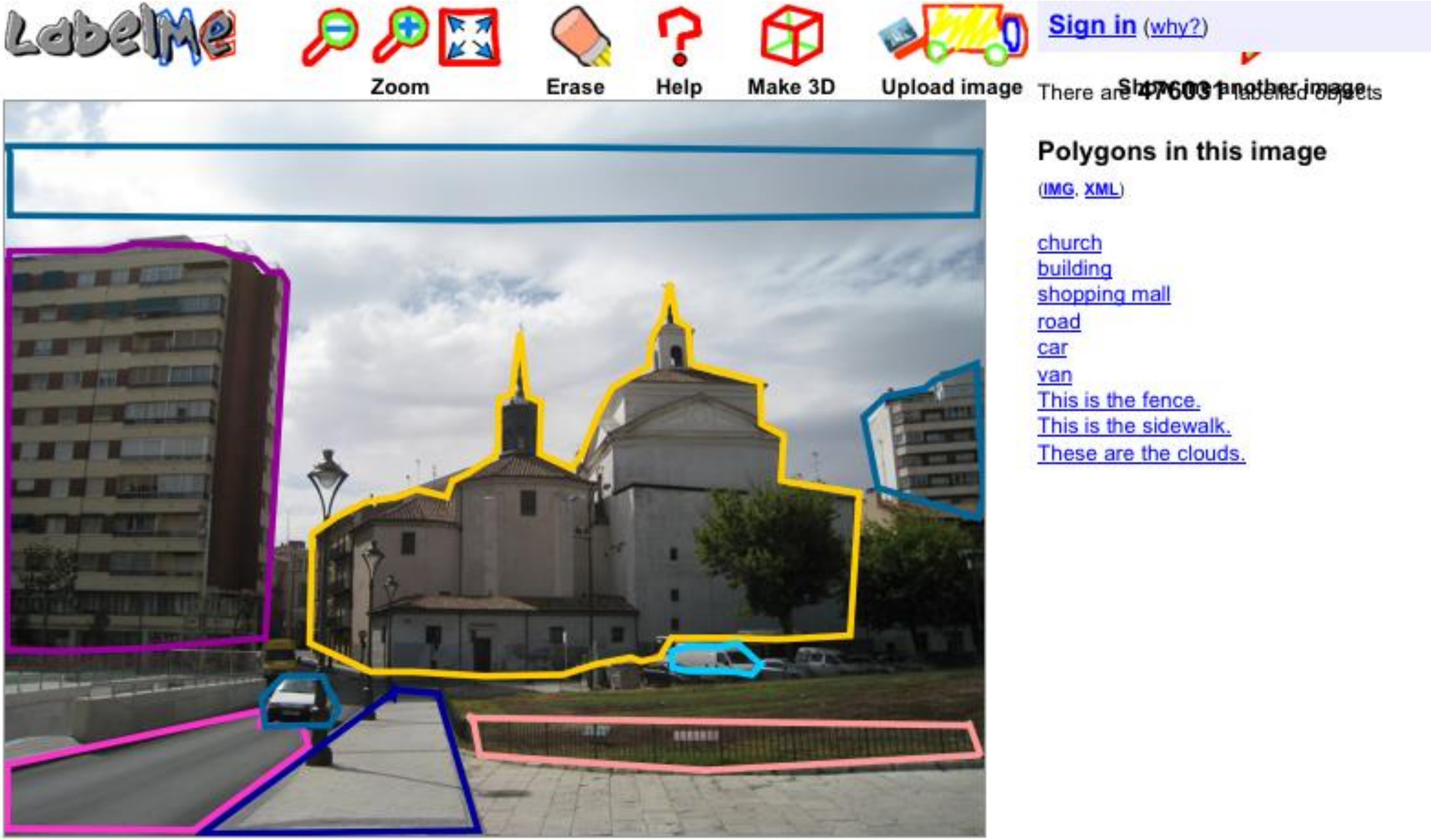
Polygons in this image

(IMG, XML)

tallship

1 cent

Task: Label as many objects in this image as you can



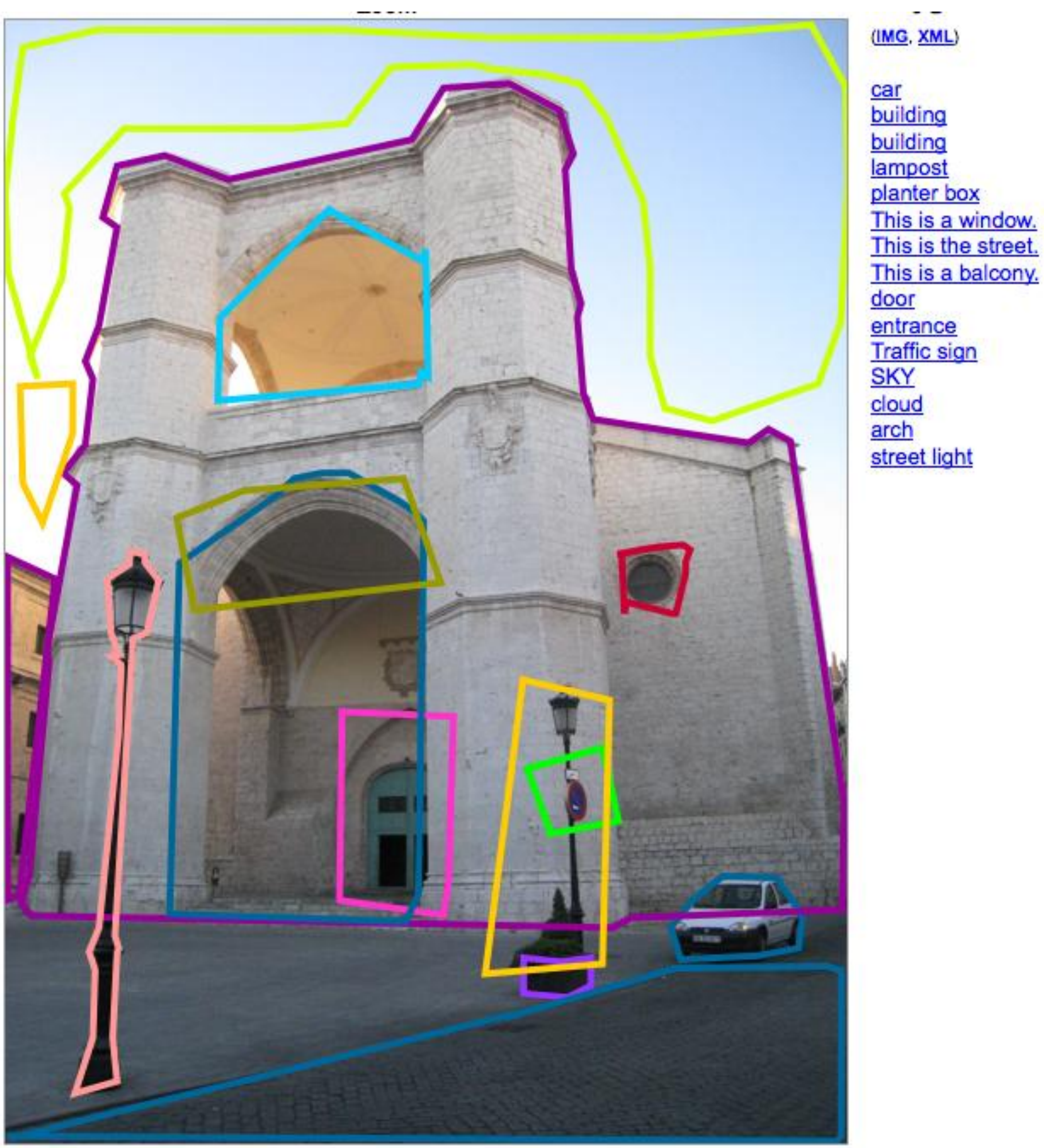
Sign in (why?)

There are 476031 labelled objects

Polygons in this image

(IMG, XML)

church
building
shopping mall
road
car
van
This is the fence.
This is the sidewalk.
These are the clouds.



(IMG, XML)

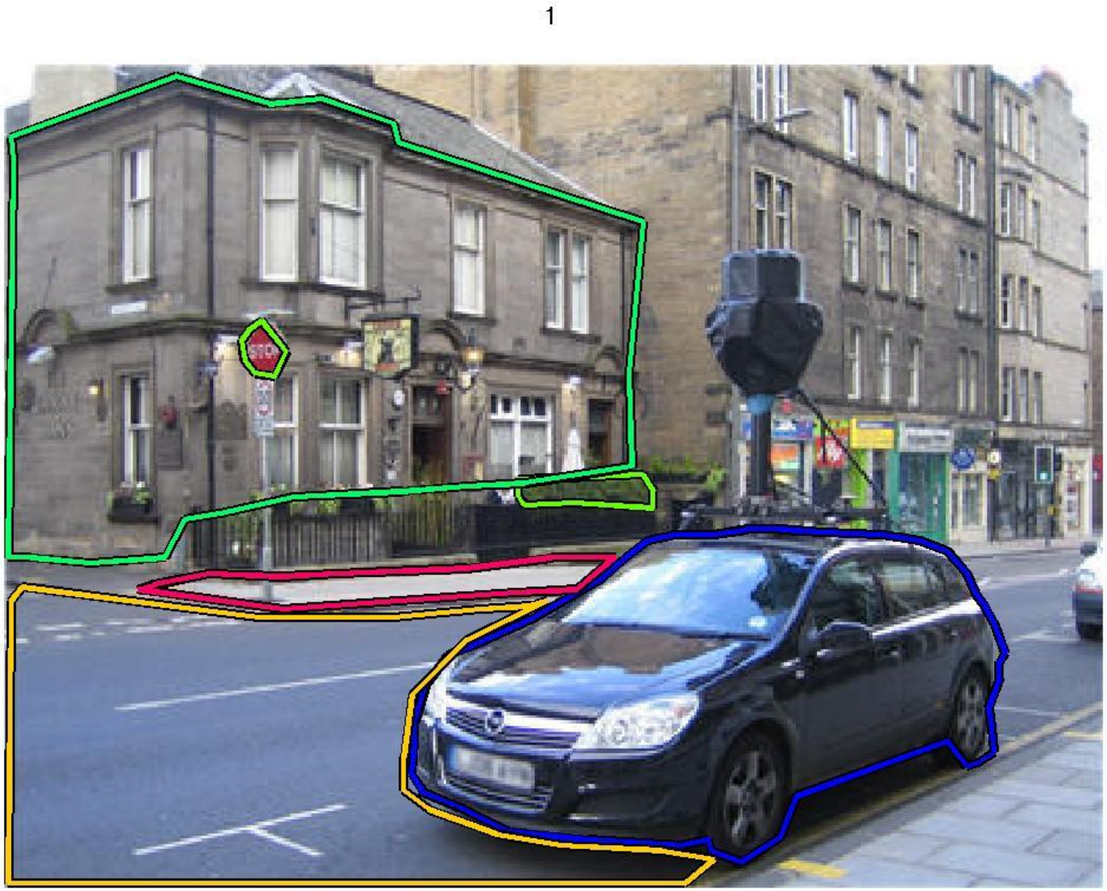
car
building
building
lampost
planter box
This is a window.
This is the street.
This is a balcony.
door
entrance
Traffic sign
SKY
cloud
arch
street light

LabelMe iterations

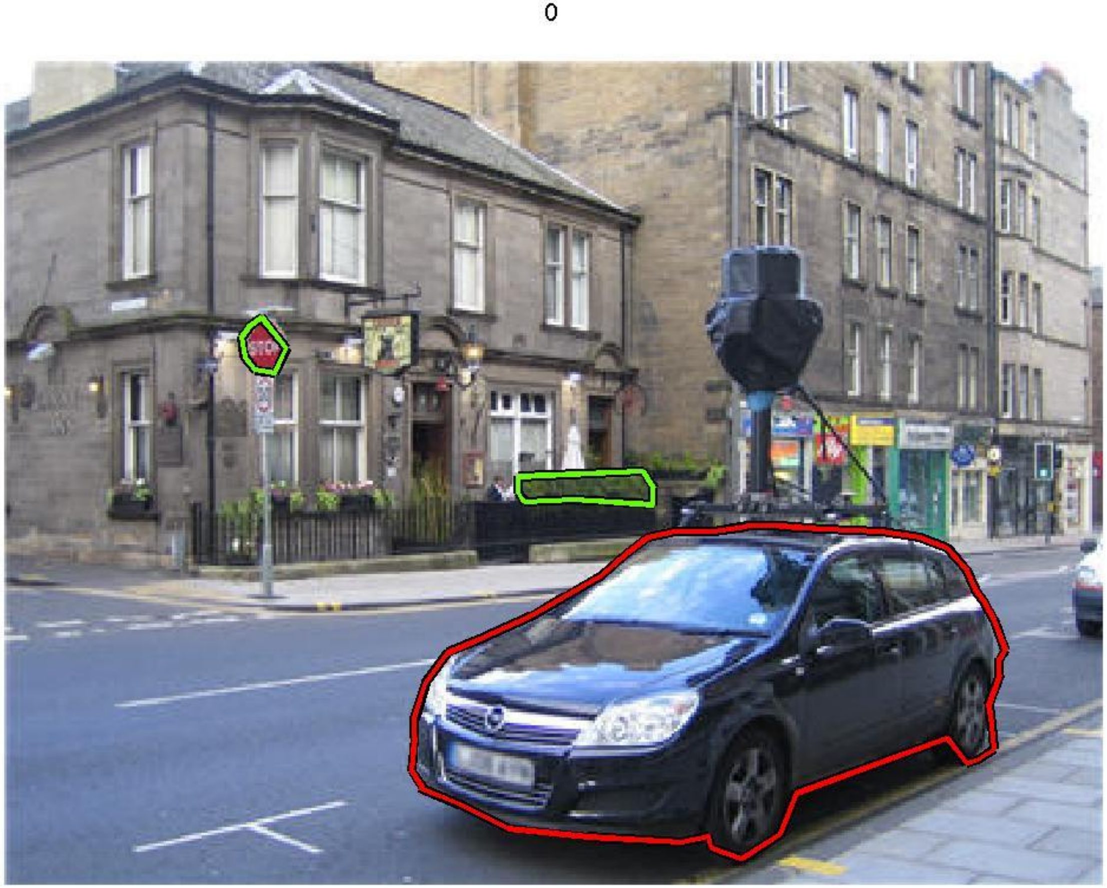
1) Label as many objects as you can
2) Delete any wrong polygon
3) Go to 1

1

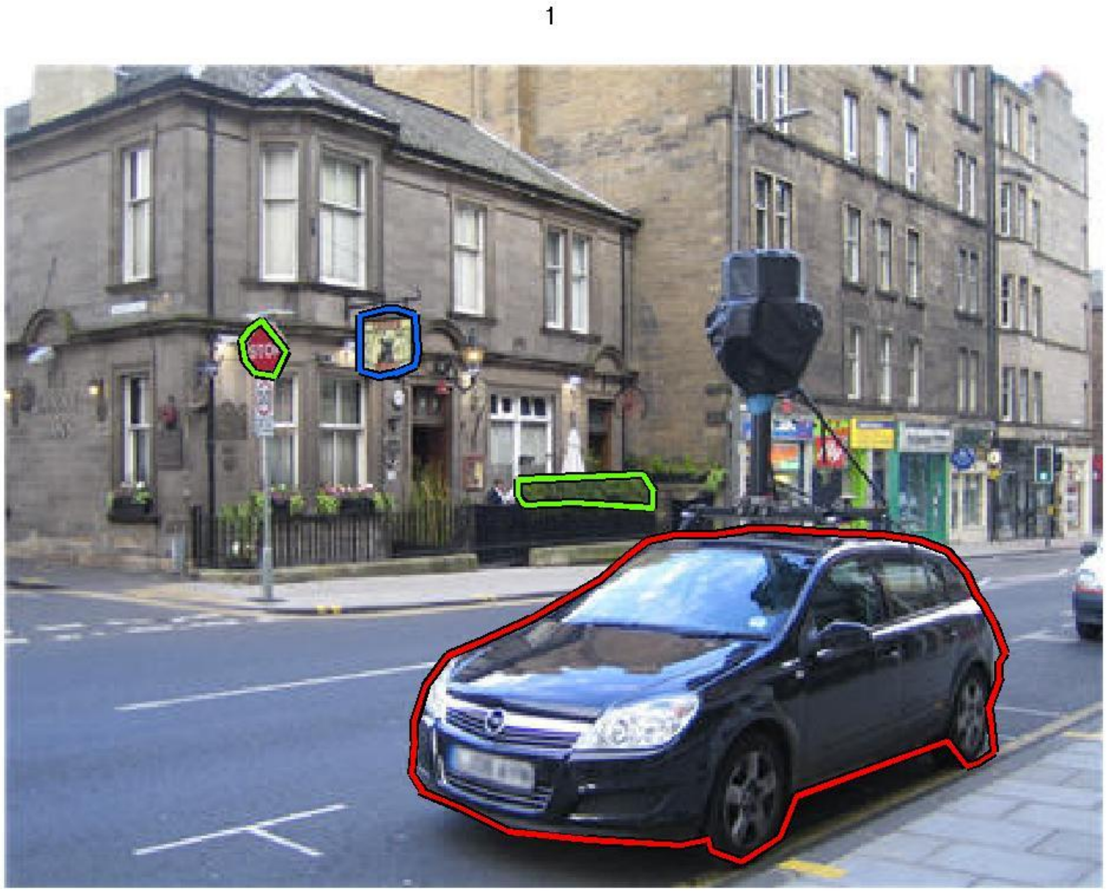
Label some objects



Delete any wrong polygons

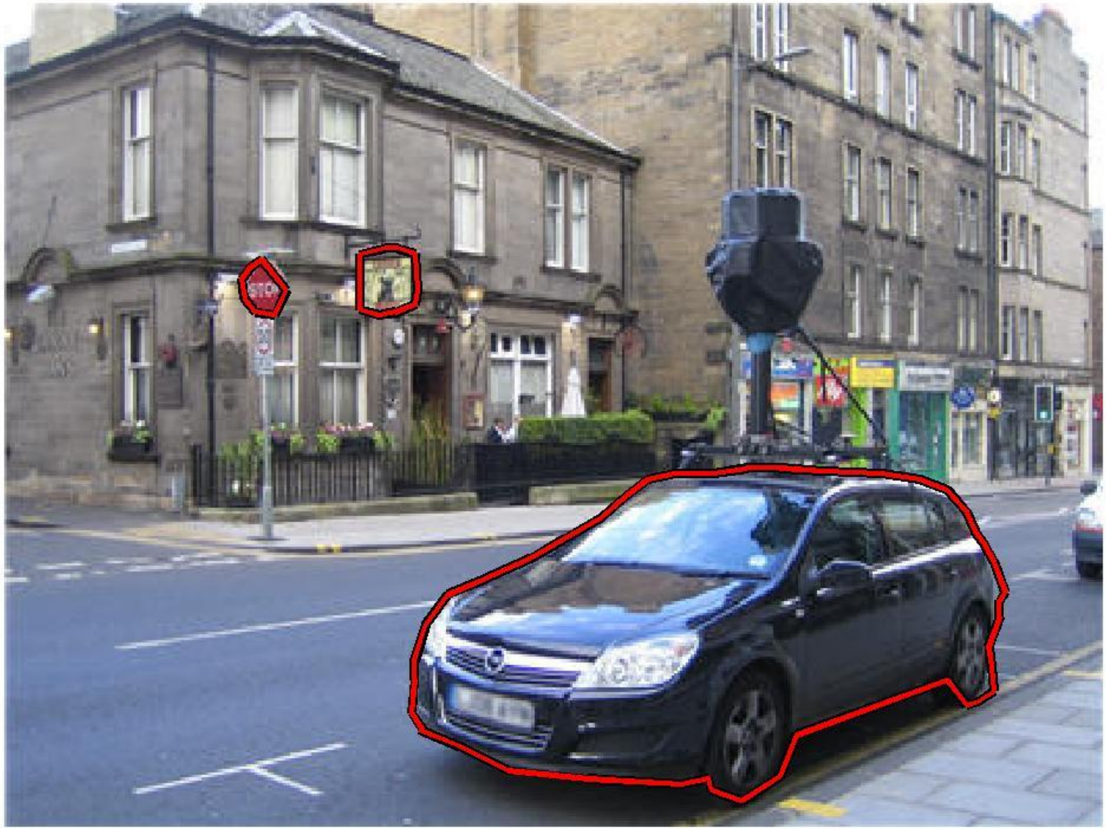


Label some objects

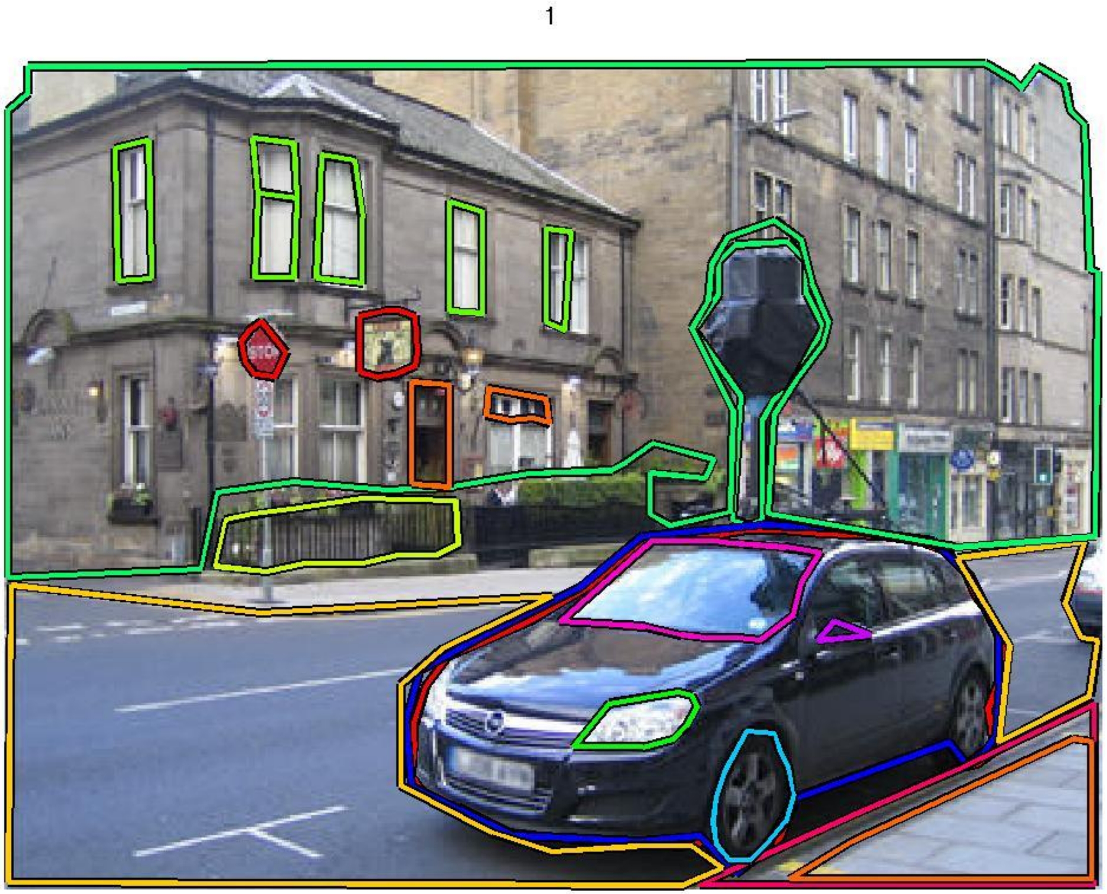


Delete any wrong polygons

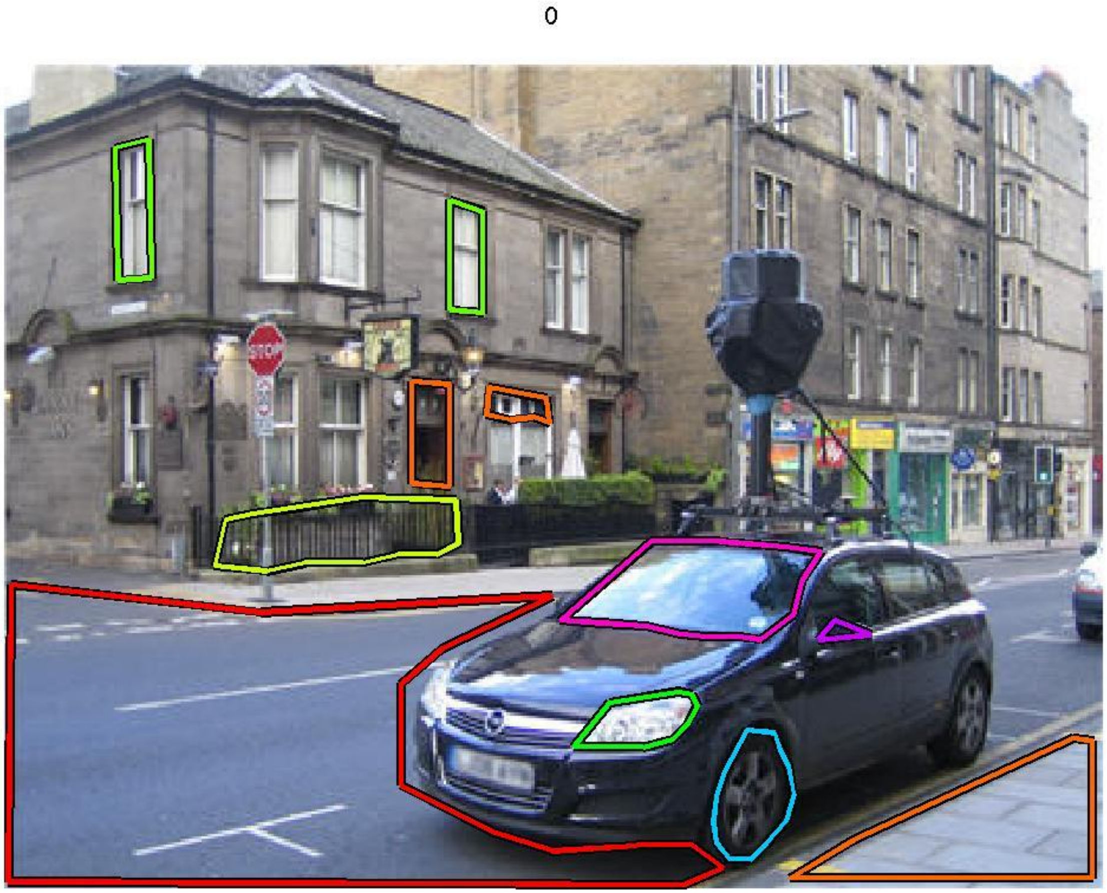
0



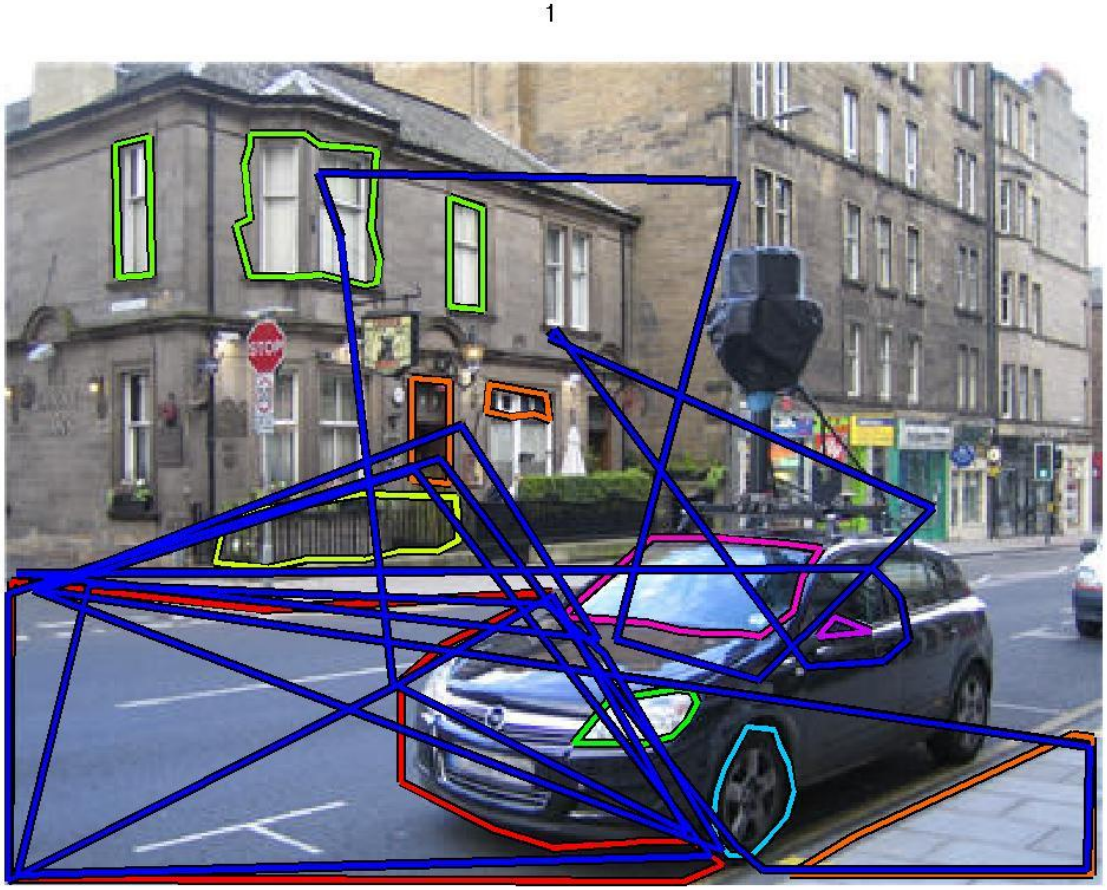
Label some objects



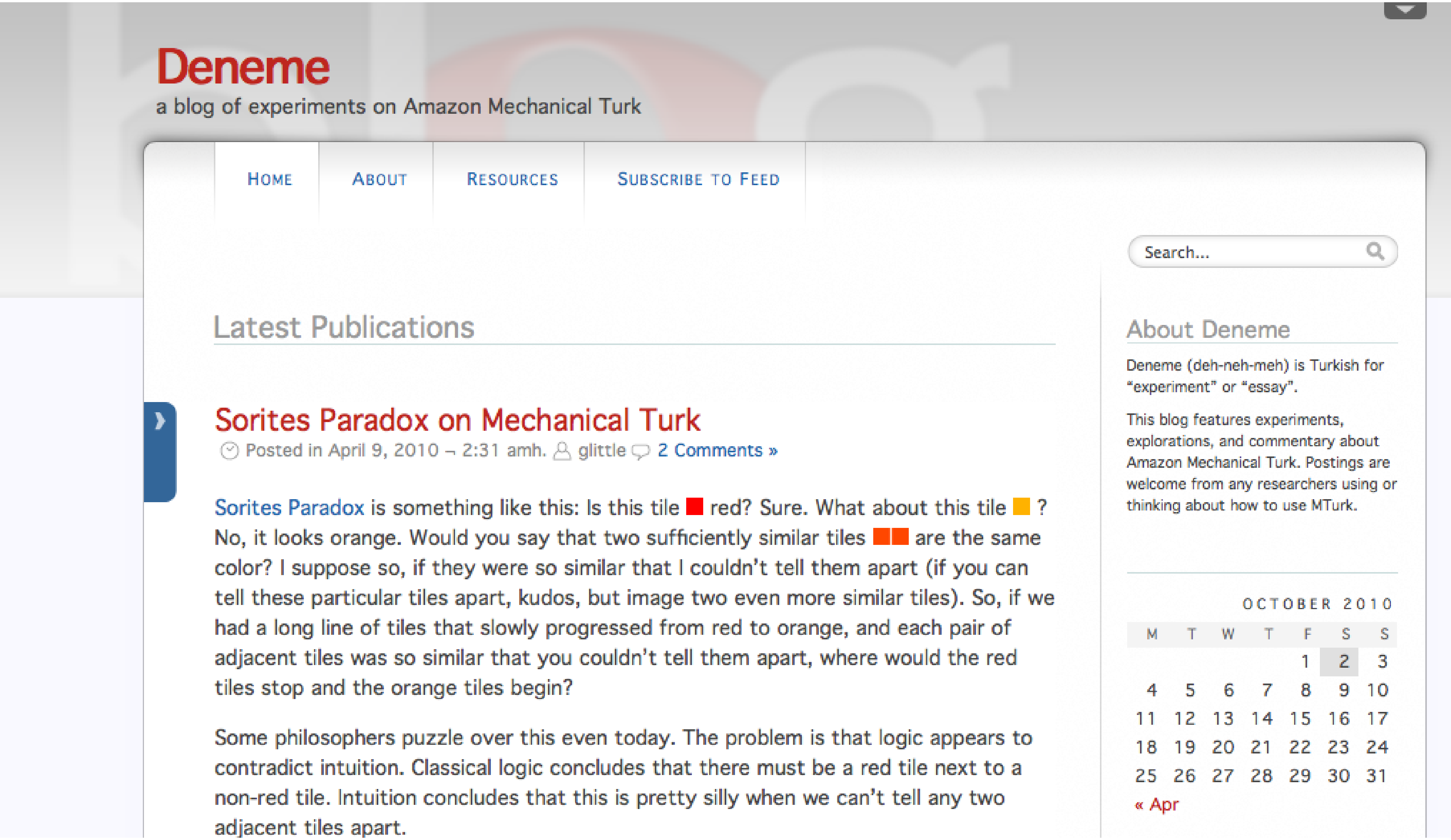
Delete any wrong polygons



Label some objects



http://groups.csail.mit.edu/uid/deneme/



Deneme

a blog of experiments on Amazon Mechanical Turk

Latest Publications

Sorites Paradox on Mechanical Turk

Posted in April 9, 2010 ‒ 2:31 amh. glittle 2 Comments »

Sorites Paradox is something like this: Is this tile 🟥 red? Sure. What about this tile 🟧 ? No, it looks orange. Would you say that two sufficiently similar tiles 🟥🟥 are the same color? I suppose so, if they were so similar that I couldn't tell them apart (if you can tell these particular tiles apart, kudos, but image two even more similar tiles). So, if we had a long line of tiles that slowly progressed from red to orange, and each pair of adjacent tiles was so similar that you couldn't tell them apart, where would the red tiles stop and the orange tiles begin?

Some philosophers puzzle over this even today. The problem is that logic appears to contradict intuition. Classical logic concludes that there must be a red tile next to a non-red tile. Intuition concludes that this is pretty silly when we can't tell any two adjacent tiles apart.

About Deneme

Deneme (deh-neh-meh) is Turkish for "experiment" or "essay".

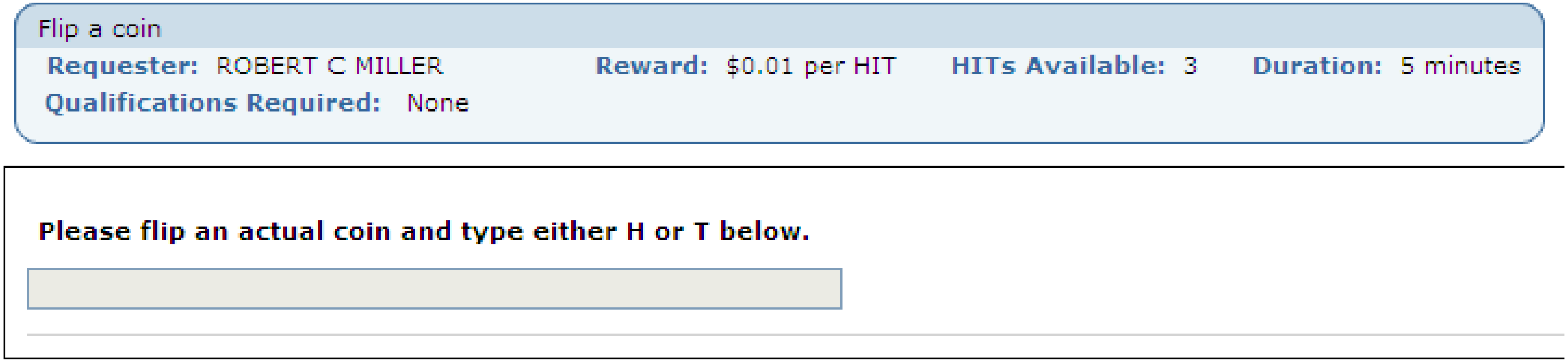
This blog features experiments, explorations, and commentary about Amazon Mechanical Turk. Postings are welcome from any researchers using or thinking about how to use MTurk.

OCTOBER 2010

M	T	W	T	F	S	S
				1	2	3
4	5	6	7	8	9	10
11	12	13	14	15	16	17
18	19	20	21	22	23	24
25	26	27	28	29	30	31

« Apr

Do humans do what you ask for?



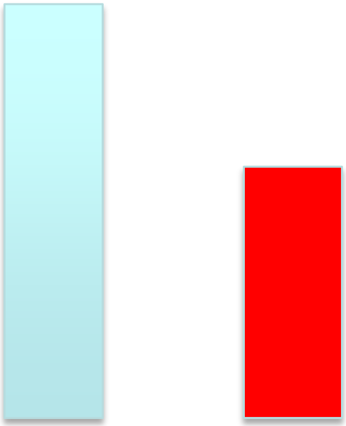
Flip a coin

Requester: ROBERT C MILLER **Reward:** $0.01 per HIT **HITs Available:** 3 **Duration:** 5 minutes

Qualifications Required: None

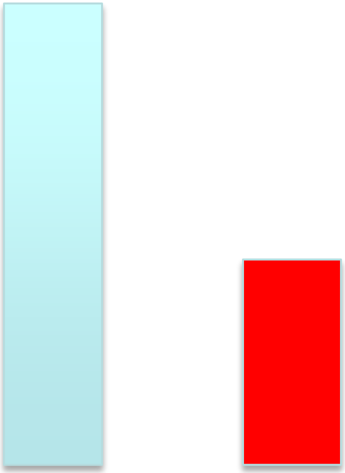
Please flip an actual coin and type either H or T below.

After 50 HITS:



31 heads, 19 tails

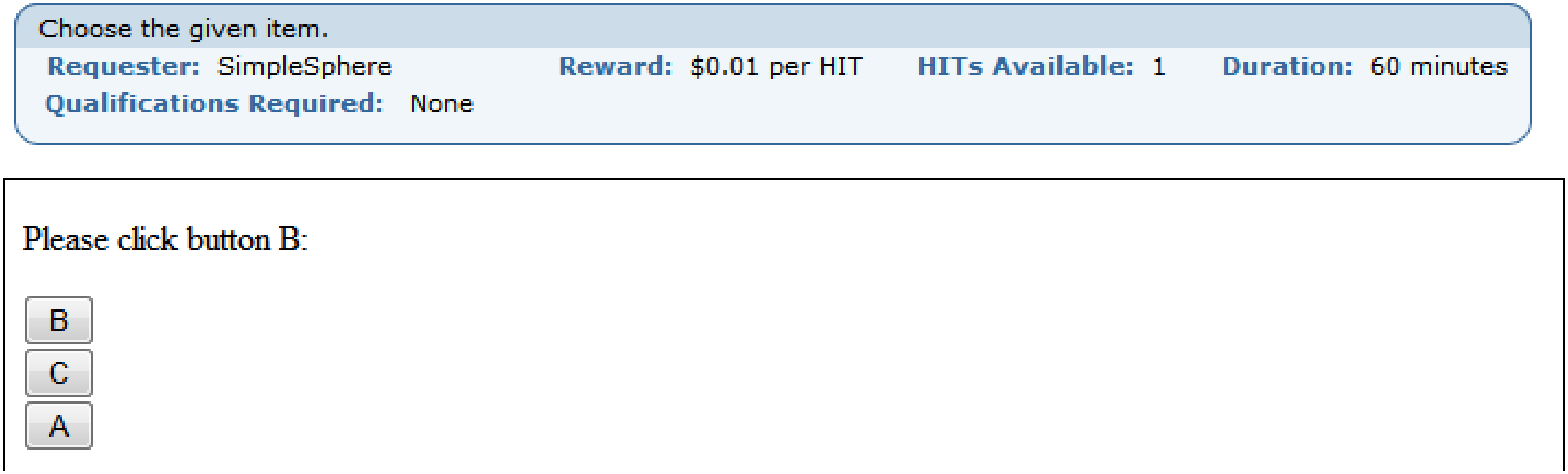
And 50 more:



34 heads, 16 tails

Experiment by Rob Miller
From http://groups.csail.mit.edu/uid/deneme/

Are humans reliable even in simple tasks?



Choose the given item.

Requester: SimpleSphere **Reward:** $0.01 per HIT **HITs Available:** 1 **Duration:** 60 minutes

Qualifications Required: None

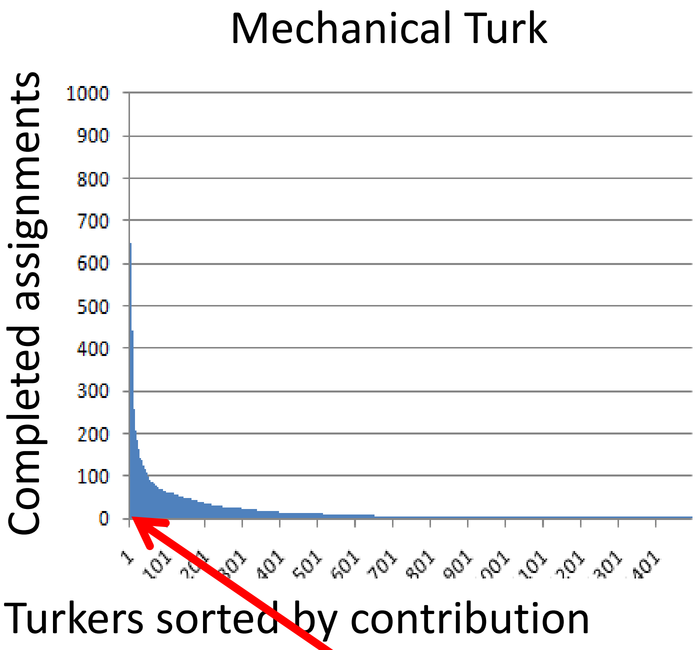
Please click button B:

Results of 100 HITS

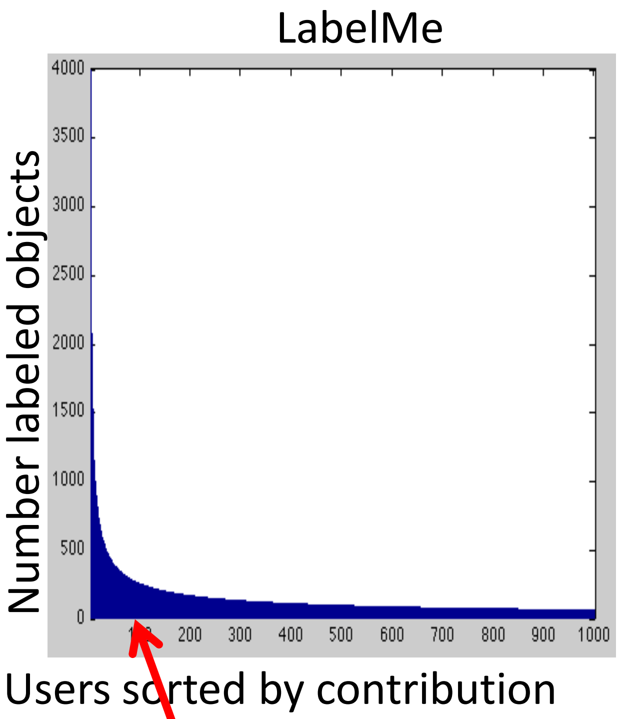
A:	2
B:	96
C:	2

Experiment by Greg Little
From http://groups.csail.mit.edu/uid/deneme/

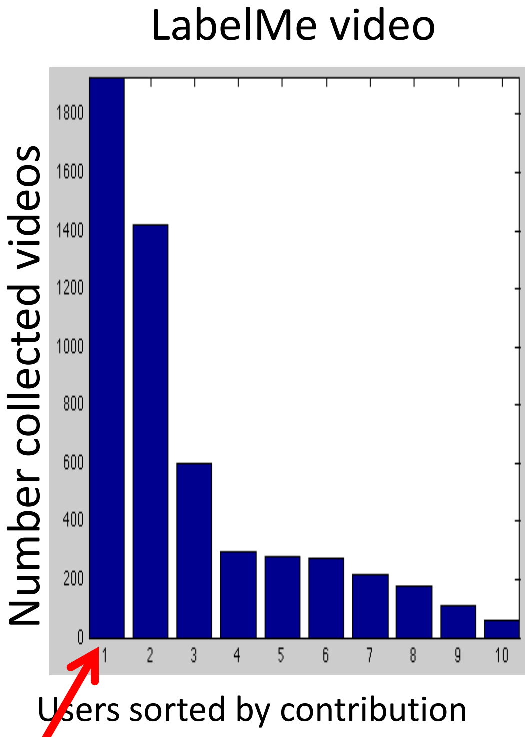
Who does the work?



From http://groups.csail.mit.edu/uid/deneme/



(from July 7th, 2008 through March 19th, 2009)



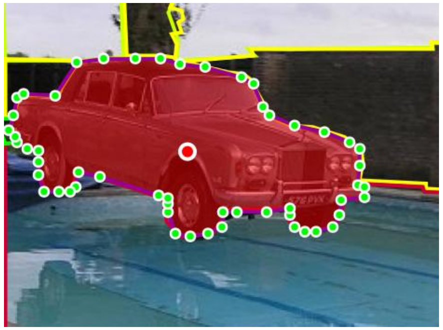
Let's hire that one

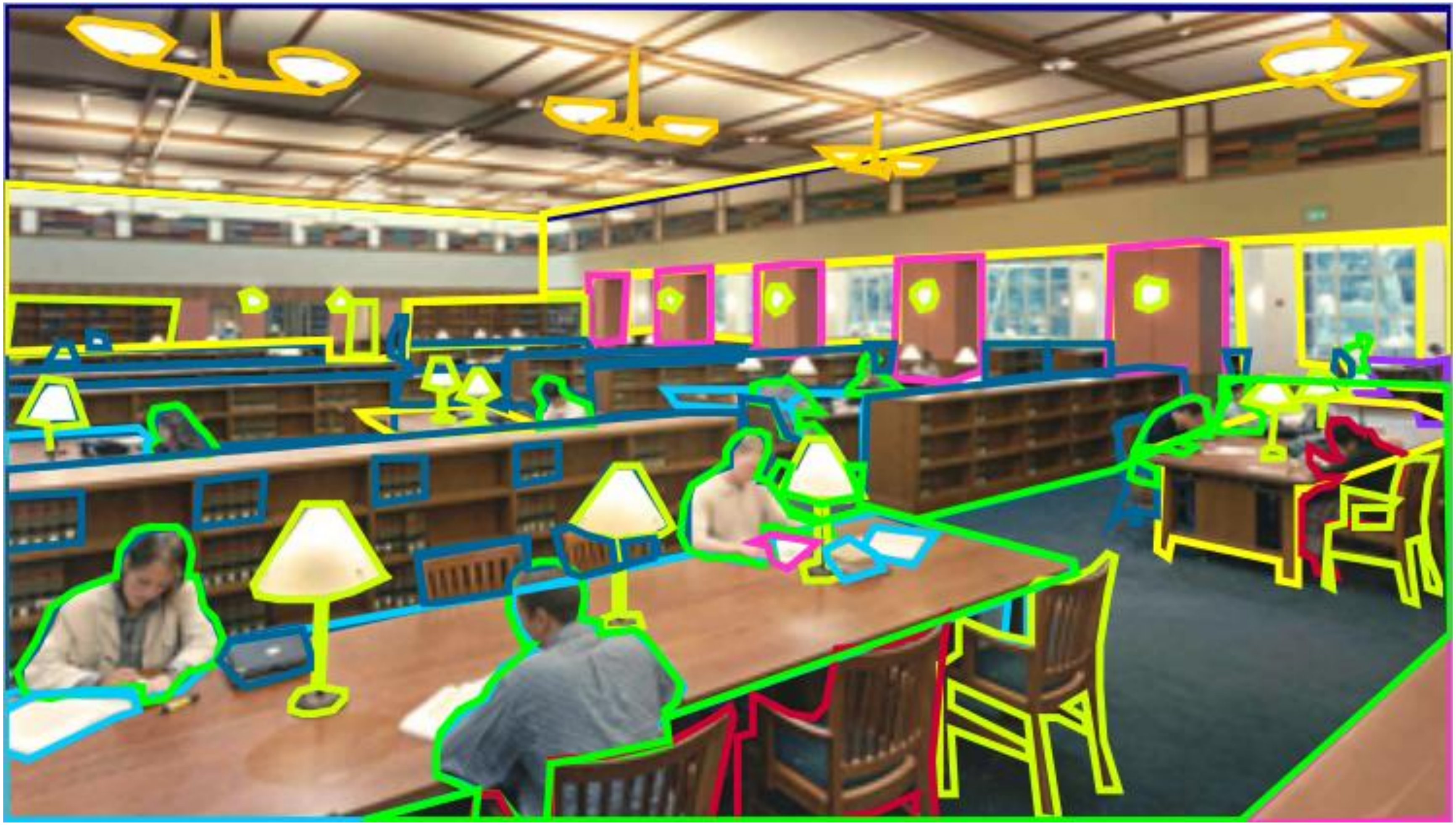
My mother's work in context

- PASCAL 11:
  - > 10? workers
  - 27.374 bounding boxes

- ImageNet:
  - >25.000 workers
  - 11.231.732 images labeled with one word

- My mother:
  - 213.841 segmented objects
  - Job offer: I am looking for more parents





C XI
C XII
C XIII
C XIV
C XV
C XVI
C XIX
C XX
C XXI
C XXII
C XXIII
C XXIV
LEITZ
D III, 1-3
LEITZ

CXI
CXII
CXIII
CXIV
CXV
CXVI
CXXI
CXXII
CXXIII
CXXIV
LEITZ
LEITZ

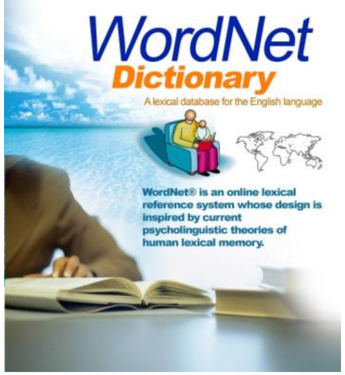
…and 15000 more images

SUN Dataset Project

We want:

- Large variety of scene categories (we want them all)
- Lots of objects categories
- Multi-object scenes

1. We take all scene words from a dictionary



WordNet
Dictionary
A lexical database for the English language

WordNet® is an online lexical reference system whose design is inspired by current psycholinguistic theories of human lexical memory.

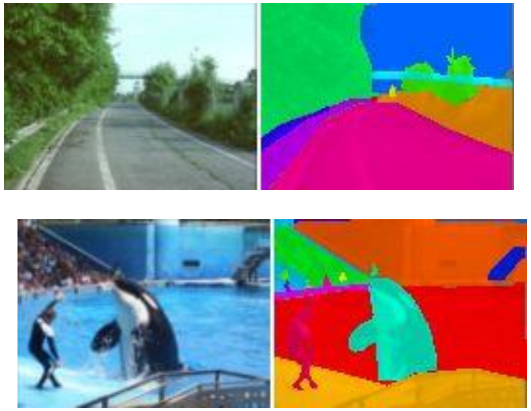
2. We download images and clean the categories

Google Image Search

altavista

flickr GAMMA

3. We segment all the images



Krista Ehinger

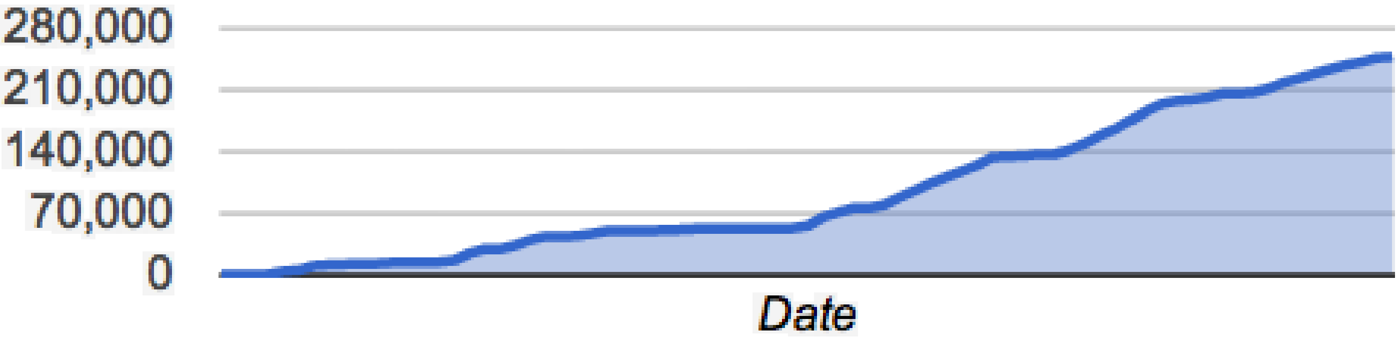
Jianxiong Xiao

Xiao, Hays, Ehinger, Oliva, Torralba; CVPR 2010

SUN Database, update

Dataset and Source Code: http://sundatabase.mit.edu

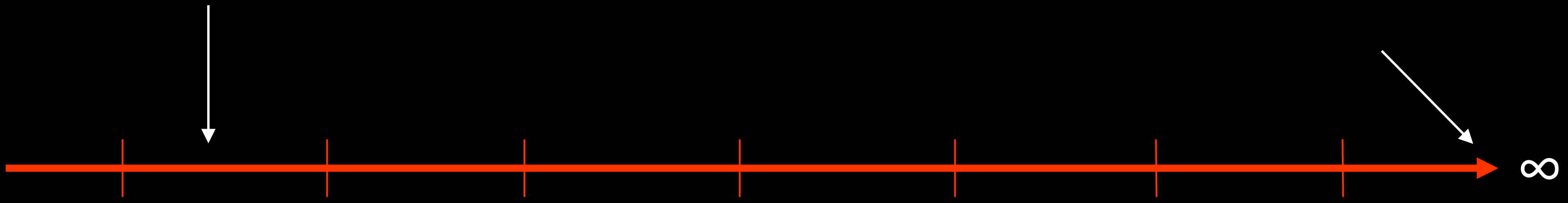
- 908 scene categories
- 131,072 images
- 3,819 object categories
- 249,522 segmented objec



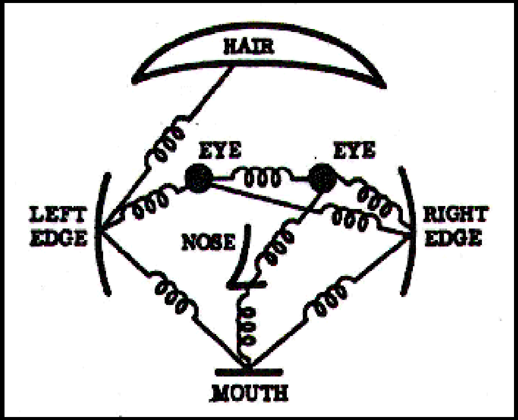
The two extremes of learning

Extrapolation problem
Generalization
Diagnostic features

Interpolation problem
Correspondence
Finding the differences

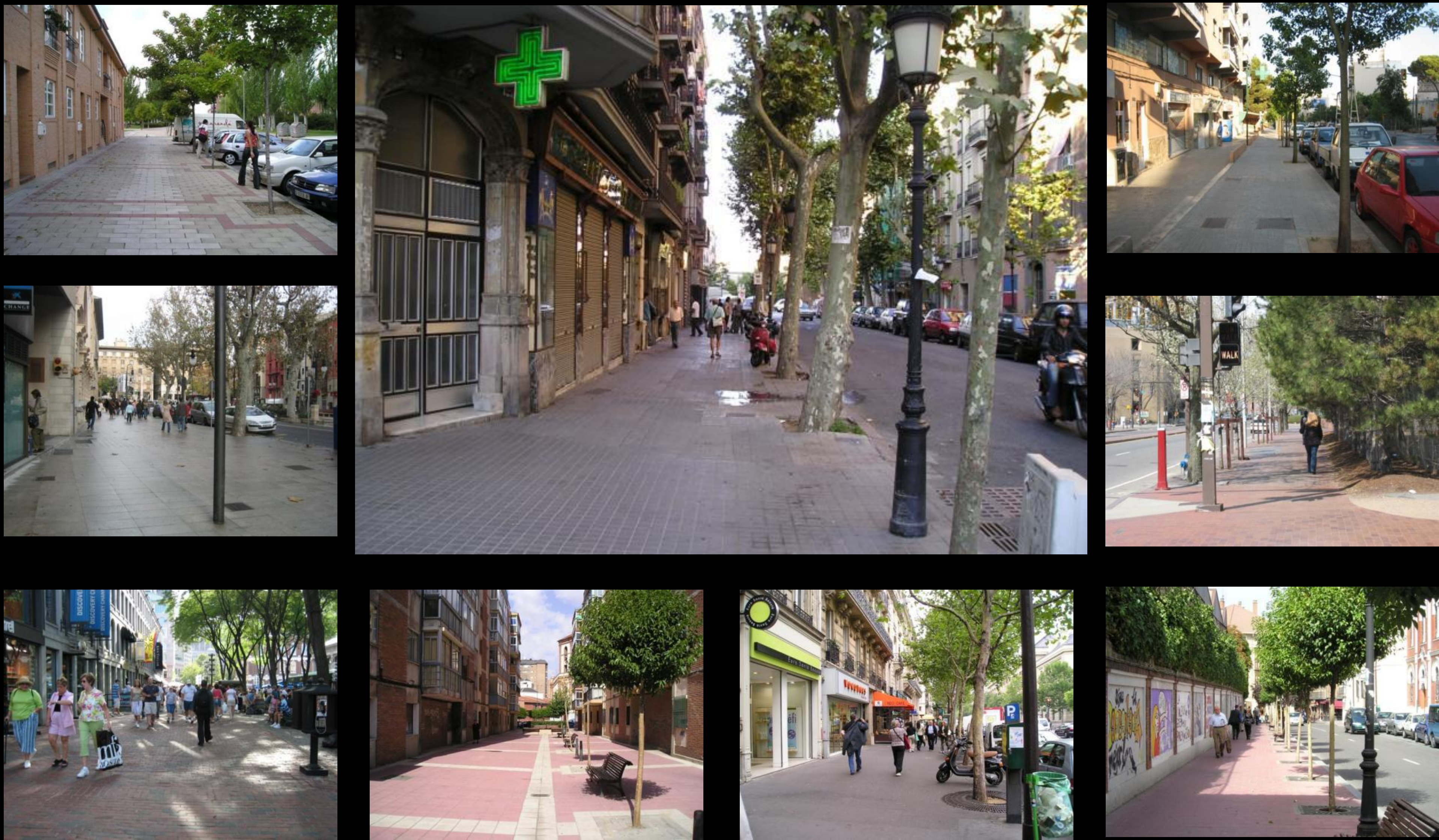


1 10 10^2 10^3 10^4 10^5 10^6 ∞ Number of training samples



Why is scene understanding hard?
Scenes are unique

But not all scenes are so original



But not all scenes are so original

But not all scenes are so original

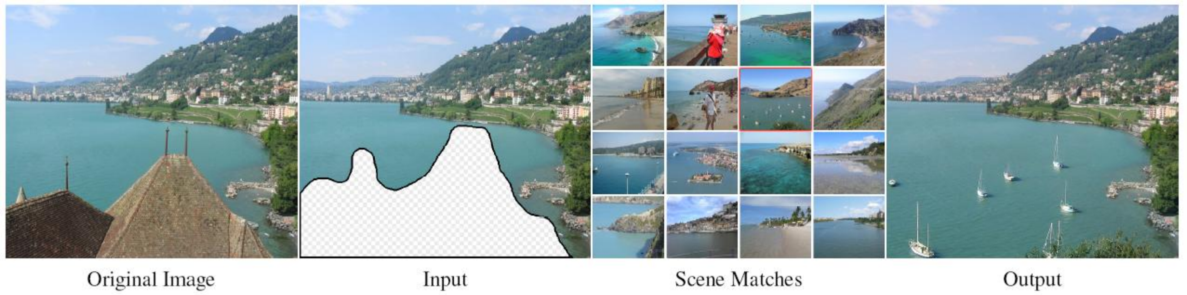
But not all scenes are so original

Large databases

PhotoSynth, Snavely et al. 2006

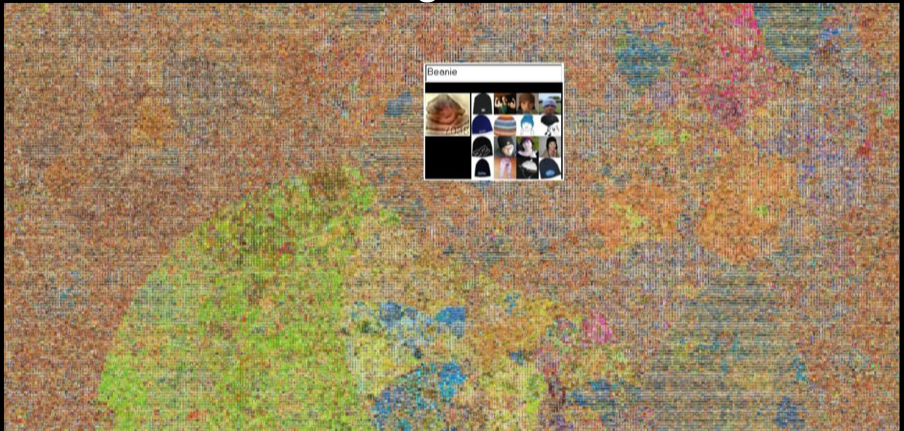
Image completion using Flickr images

Hays and Efros, 2007



Recognition: 80 million images

A. Torralba, R. Fergus, W.T. Freeman. 2008



Event prediction

What can happen here?

Liu, Yuen, Torralba. CVPR 2009. Yuen, Torralba. ECCV 2010

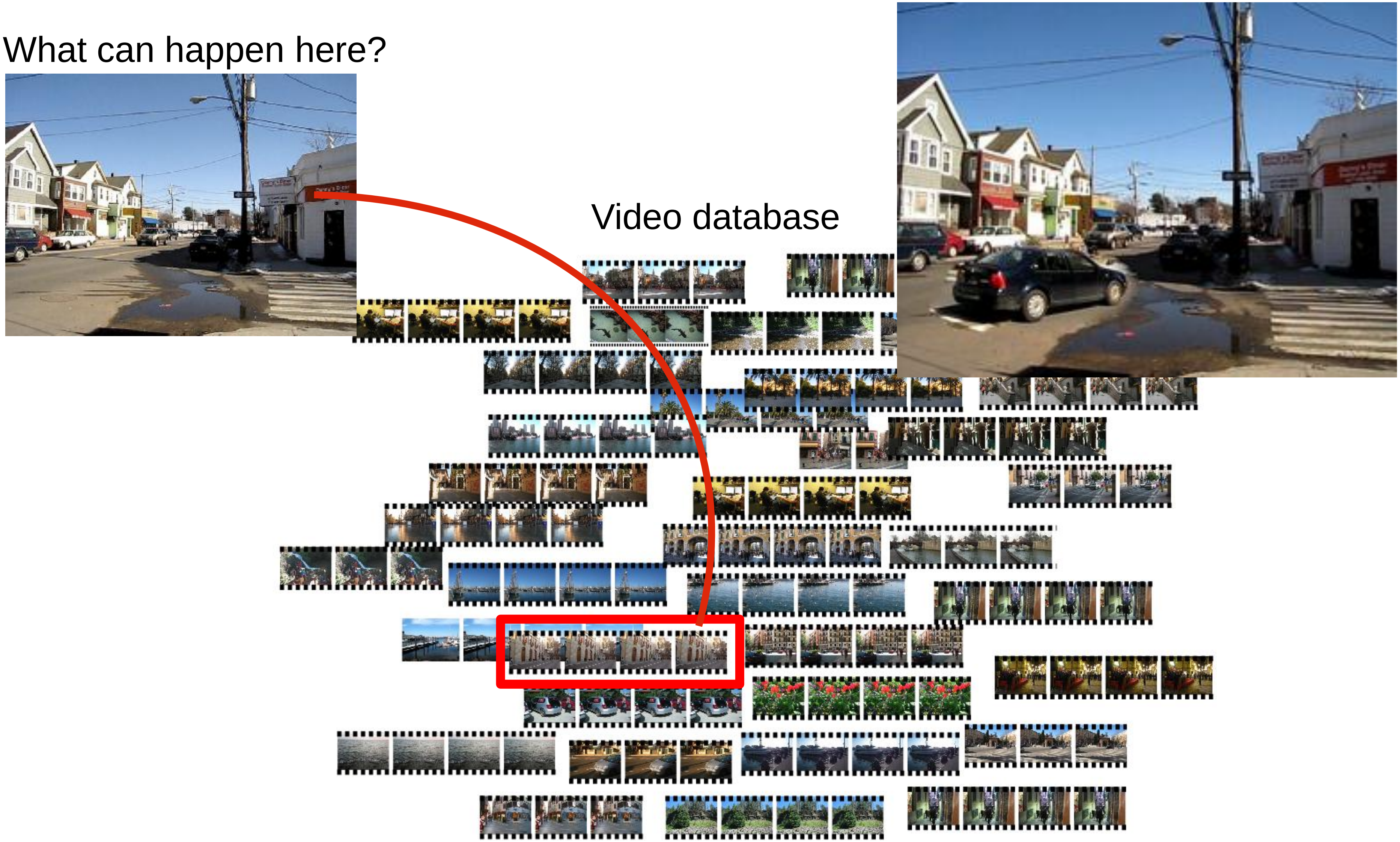
Event prediction

What can happen here?

Video database

Liu, Yuen, Torralba. CVPR 2009. Yuen, Torralba. ECCV 2010

Event prediction

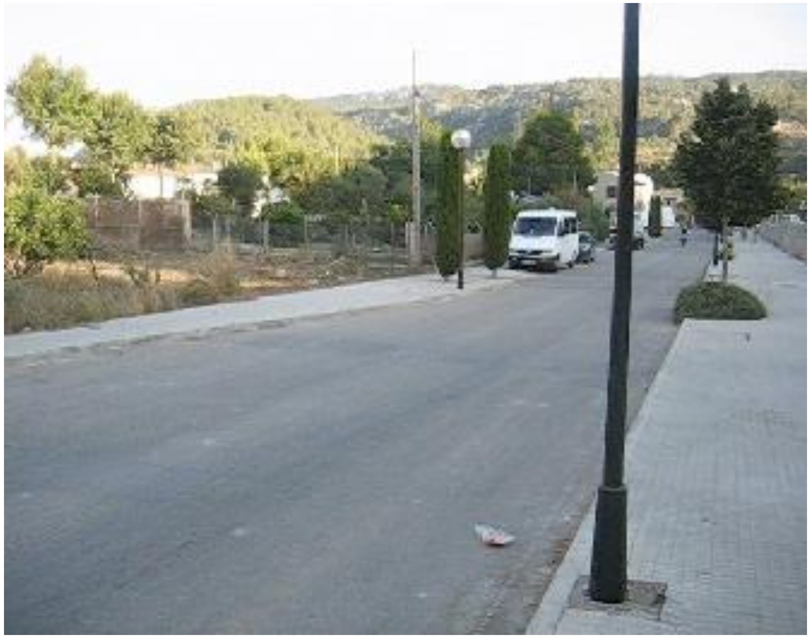


What can happen here?

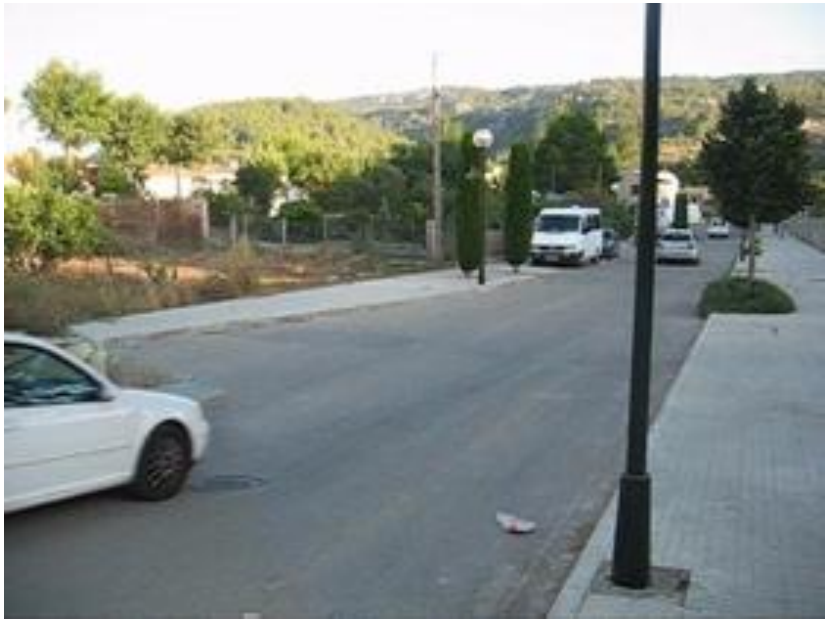
Video database

Liu, Yuen, Torralba. CVPR 2009; Yuen, Torralba. ECCV 2010

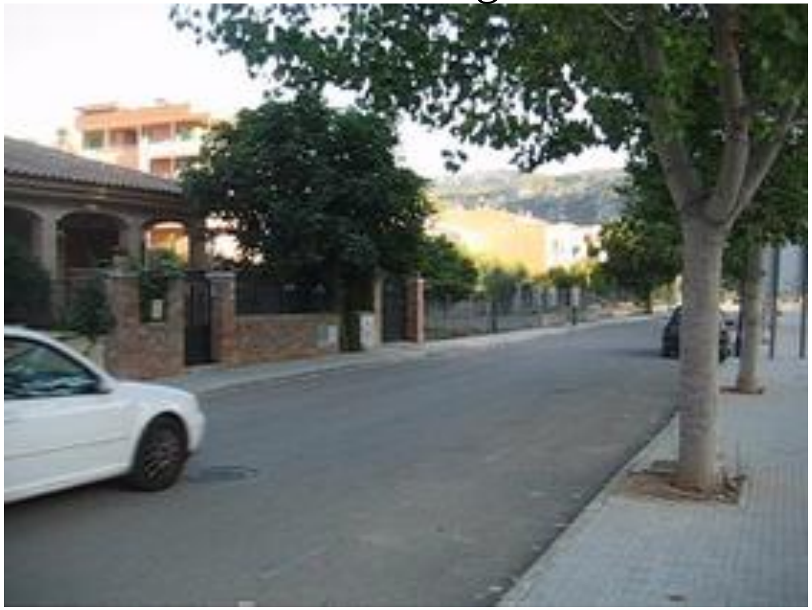
What can happen here?



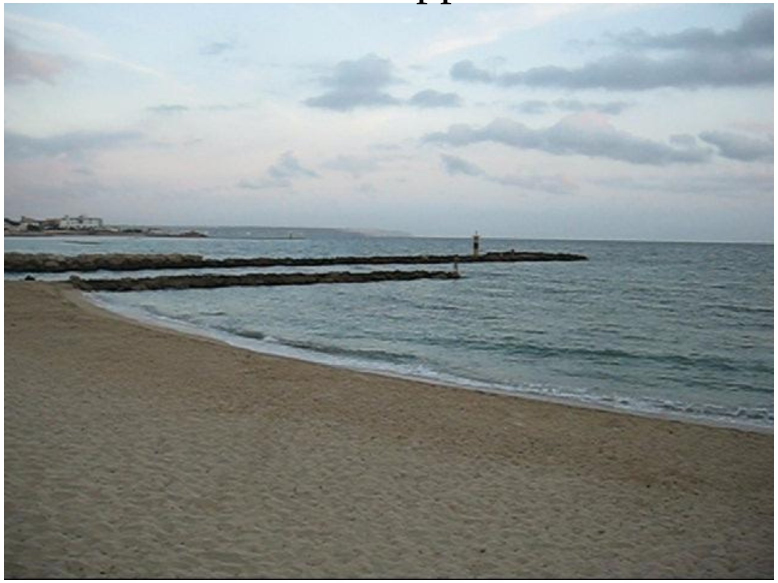
Prediction



Nearest neighbor

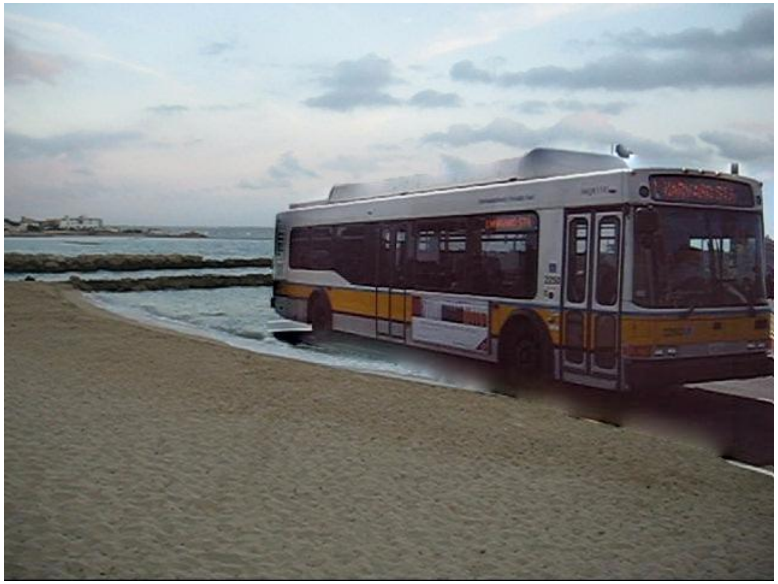


What can happen here?

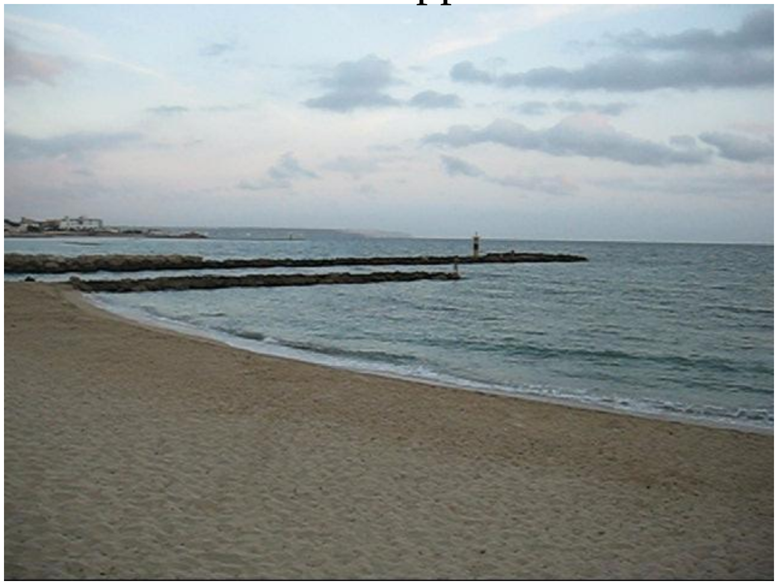


What can happen here?

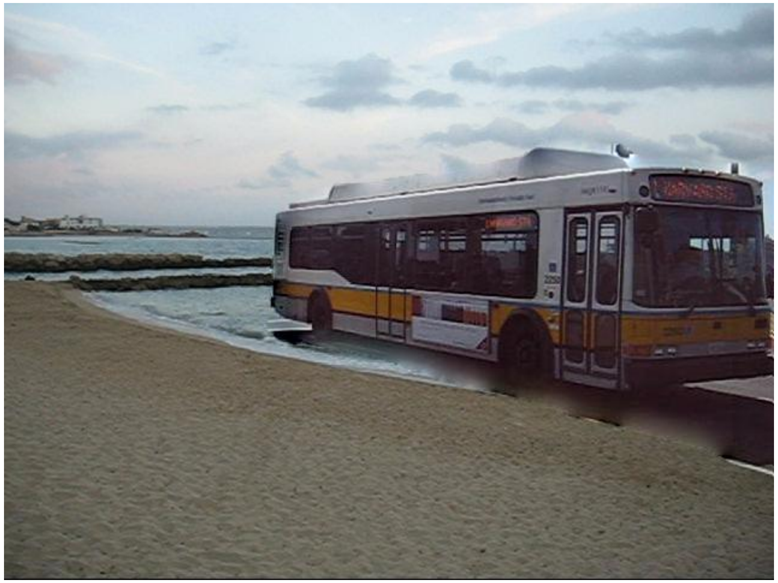
Prediction



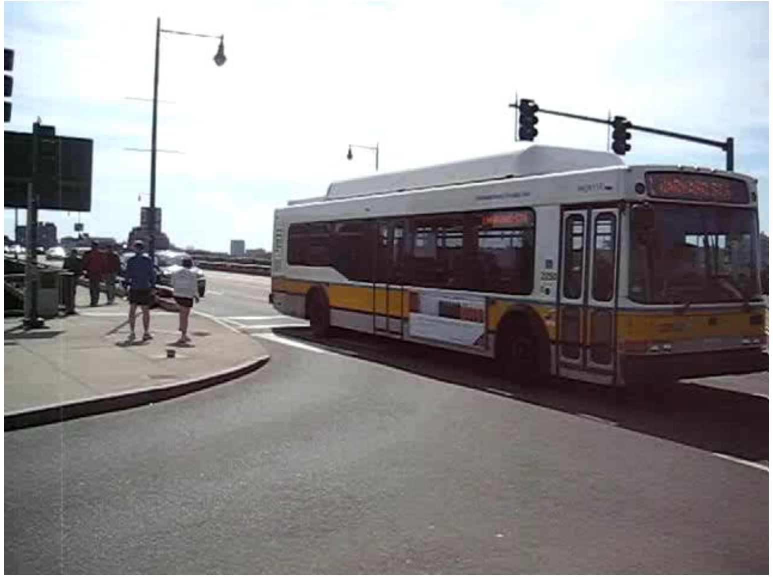
What can happen here?



Prediction



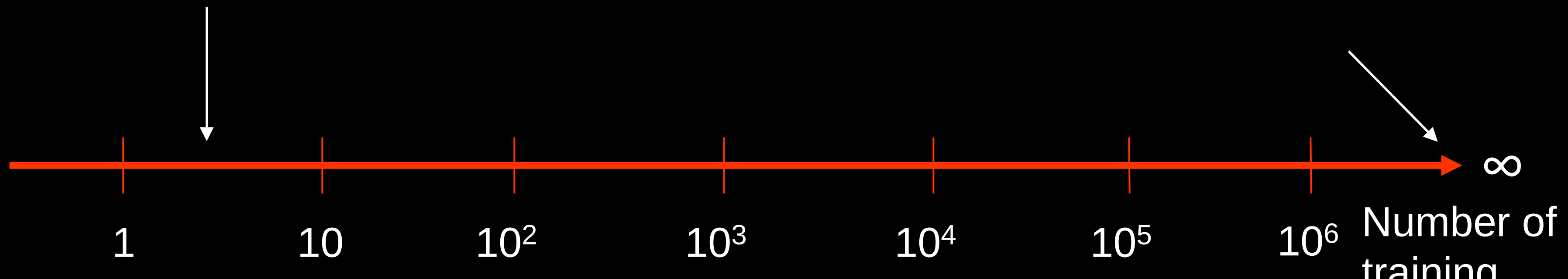
Nearest neighbor



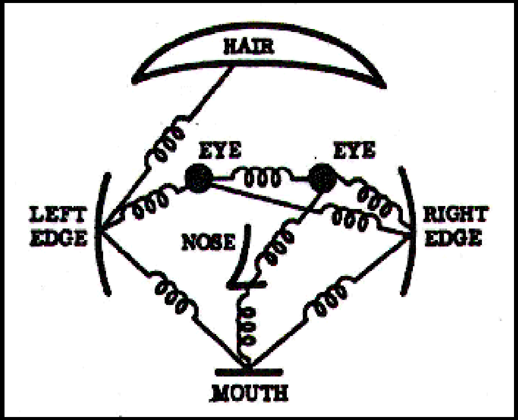
The two extremes of learning

Extrapolation problem
Generalization
Diagnostic features

Interpolation problem
Correspondence
Finding the differences

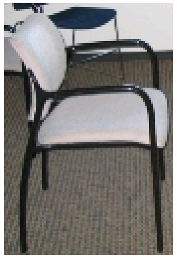


Number of training samples



Shared features

- Is learning the object class 1000 easier than learning the first?



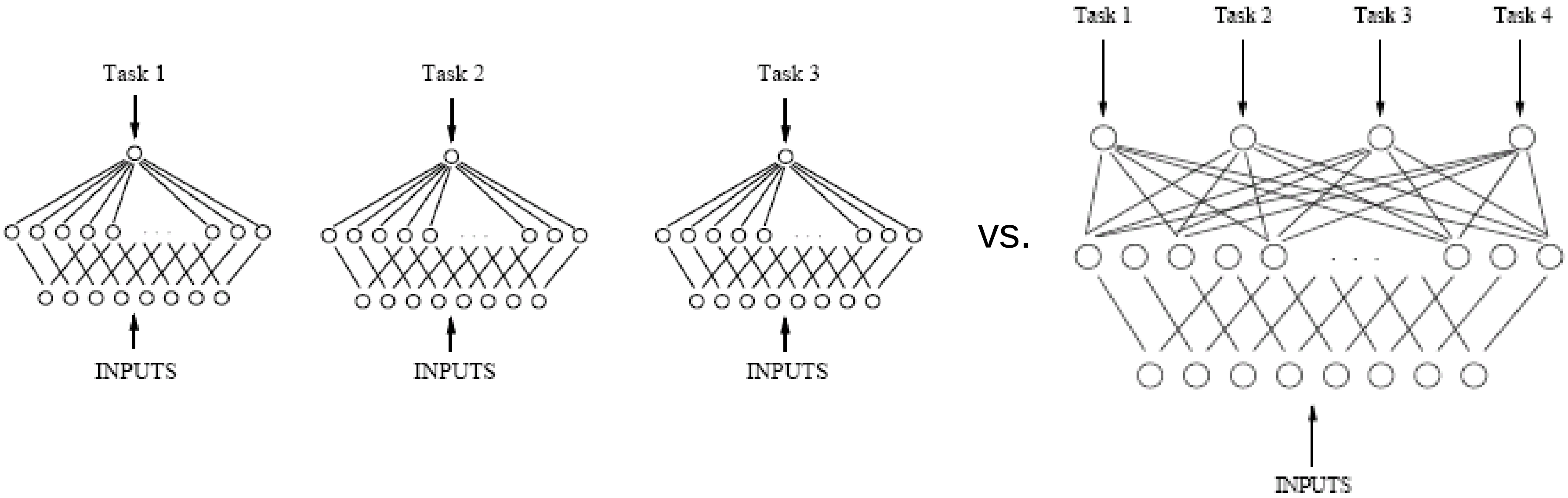
…

- Can we transfer knowledge from one object to another?
- Are the shared properties interesting by themselves?

Multitask learning

R. Caruana. Multitask Learning. ML 1997

“MTL improves generalization by leveraging the domain-specific information contained in the training signals of *related* tasks. It does this by training tasks in parallel while using a shared representation”.



Sejnowski & Rosenberg 1986; Hinton 1986; Le Cun et al. 1989; Suddarth & Kergosien 1990; Pratt et al. 1991; Sharkey & Sharkey 1992; …

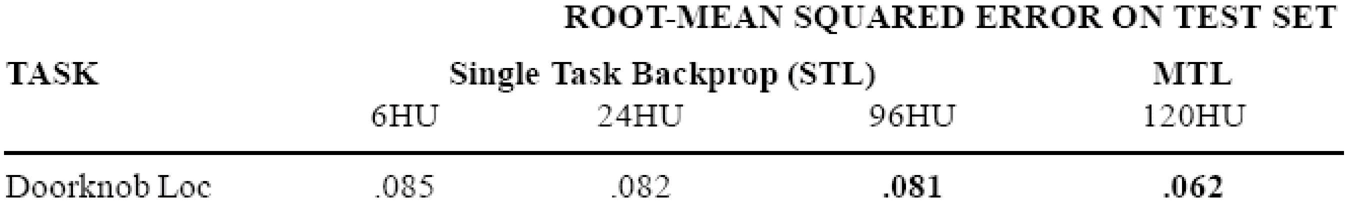
Multitask learning

R. Caruana. Multitask Learning. ML 1997

Primary task: detect door knobs

Tasks used:

- horizontal location of doorknob
- single or double door
- horizontal location of doorway center
- width of doorway
- horizontal location of left door jamb
- horizontal location of right door jamb
- width of left door jamb
- width of right door jamb
- horizontal location of left edge of door
- horizontal location of right edge of door

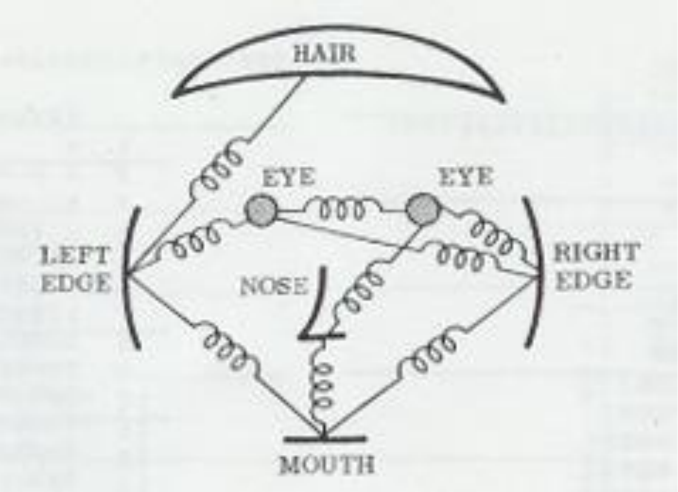


ROOT-MEAN SQUARED ERROR ON TEST SET

TASK	Single Task Backprop (STL)			MTL
	6HU	24HU	96HU	120HU
Doorknob Loc	.085	.082	**.081**	**.062**

Sharing in constellation models

(next Wednesday)

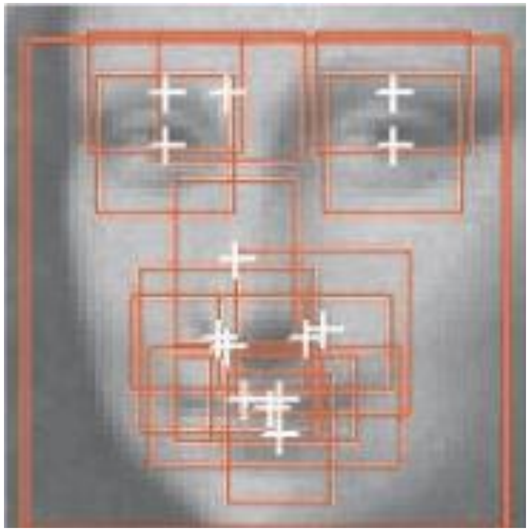


Pictorial Structures

Fischler & Elschlager, IEEE Trans. Comp. 1973

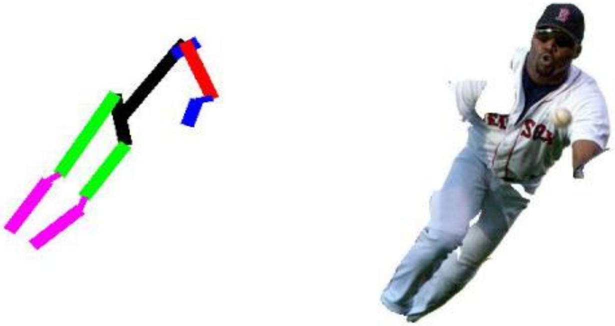
Constellation Model

Fergus, Perona, & Zisserman, CVPR 2003



SVM Detectors

Heisele, Poggio, et. al., NIPS 2001



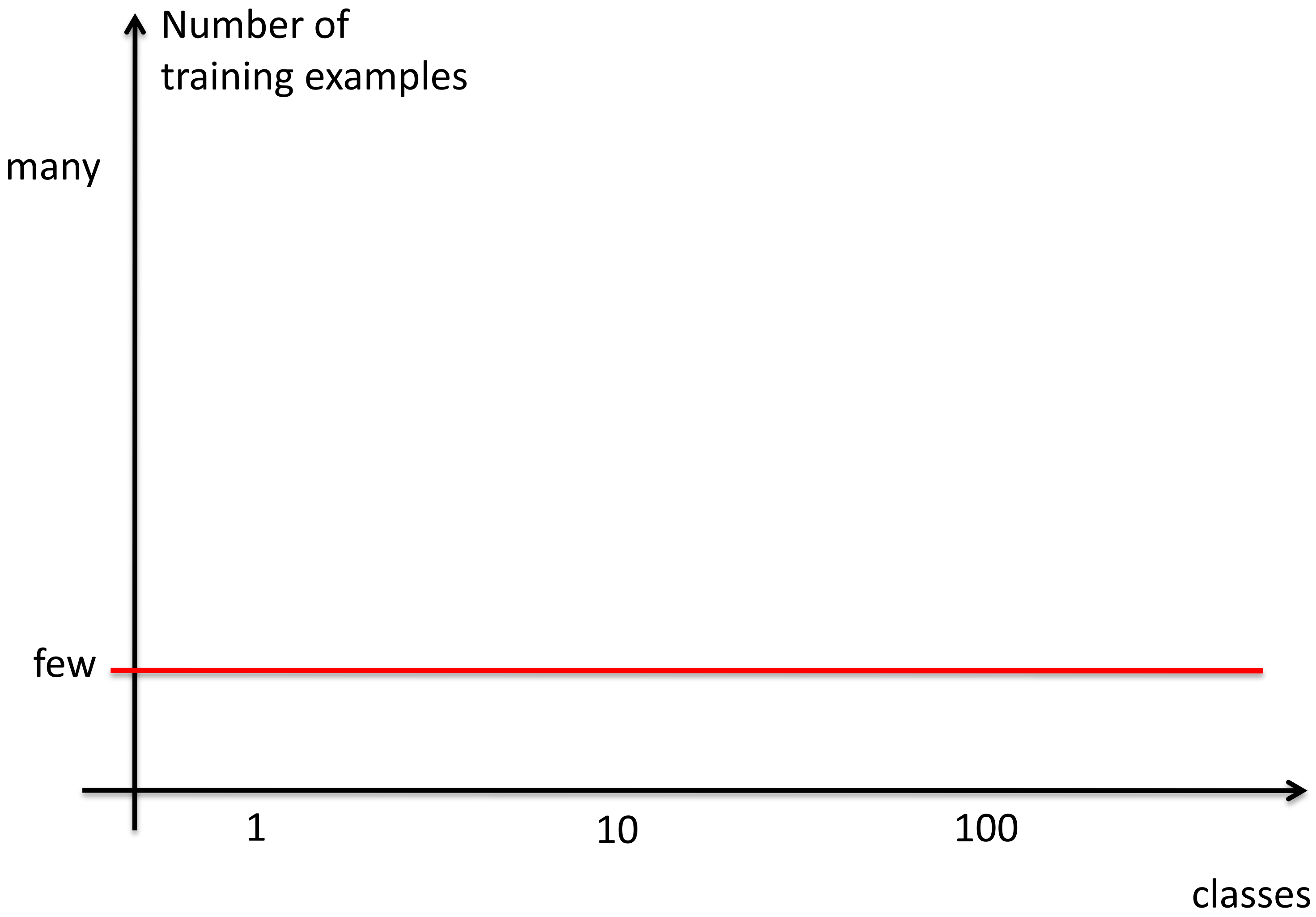
Model-Guided Segmentation

Mori, Ren, Efros, & Malik, CVPR 2004

Some more references

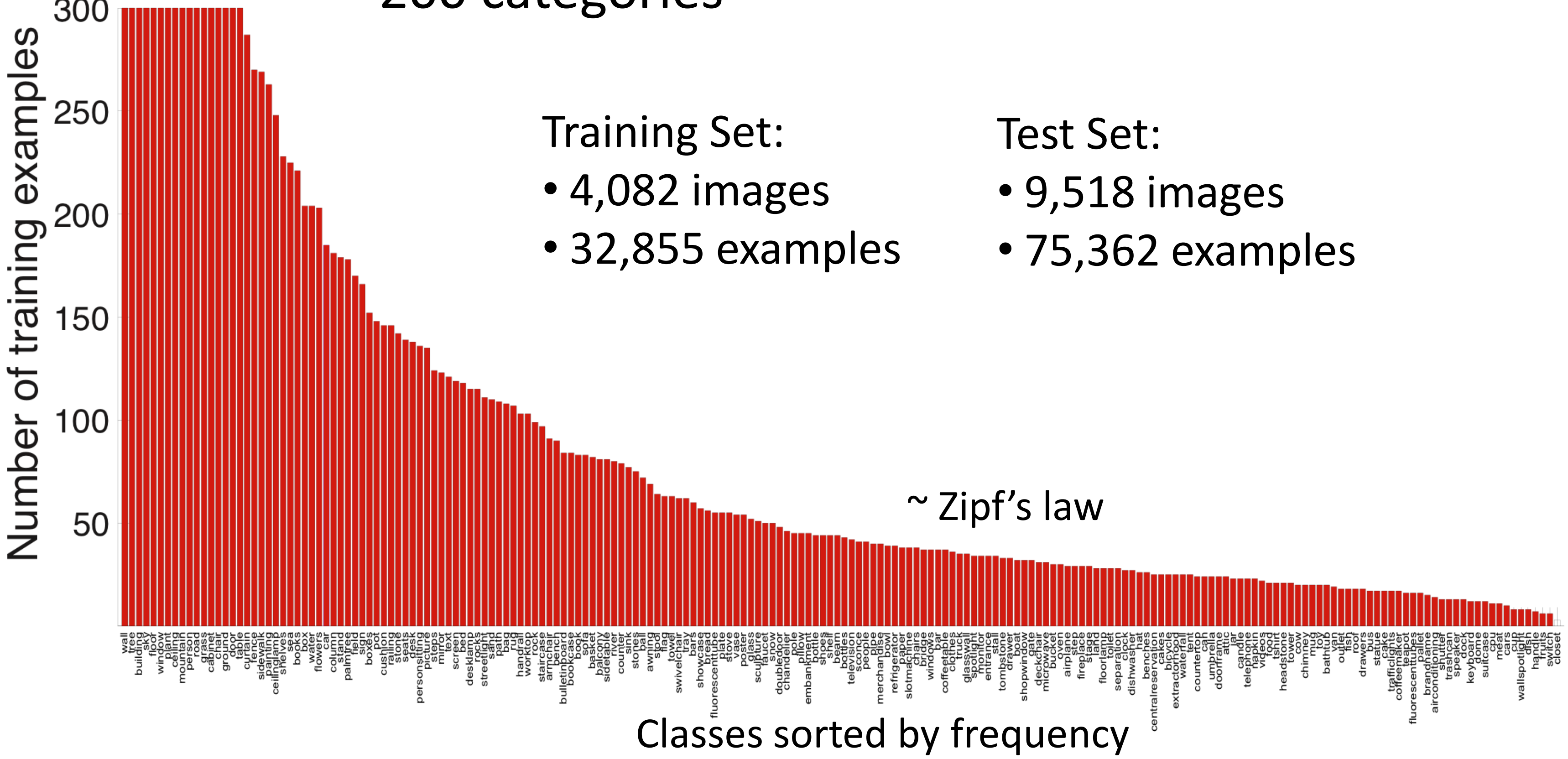
- Baxter 1996
- Caruana 1997
- Schapire, Singer, 2000
- Thrun, Pratt 1997
- Krempp, Geman, Amit, 2002
- E.L.Miller, Matsakis, Viola, 2000
- Mahamud, Hebert, Lafferty, 2001
- Fink et al. 2003, 2004
- LeCun, Huang, Bottou, 2004
- Holub, Welling, Perona, 2005
- …

Current training settings for learning from few training examples



SUN database

200 categories

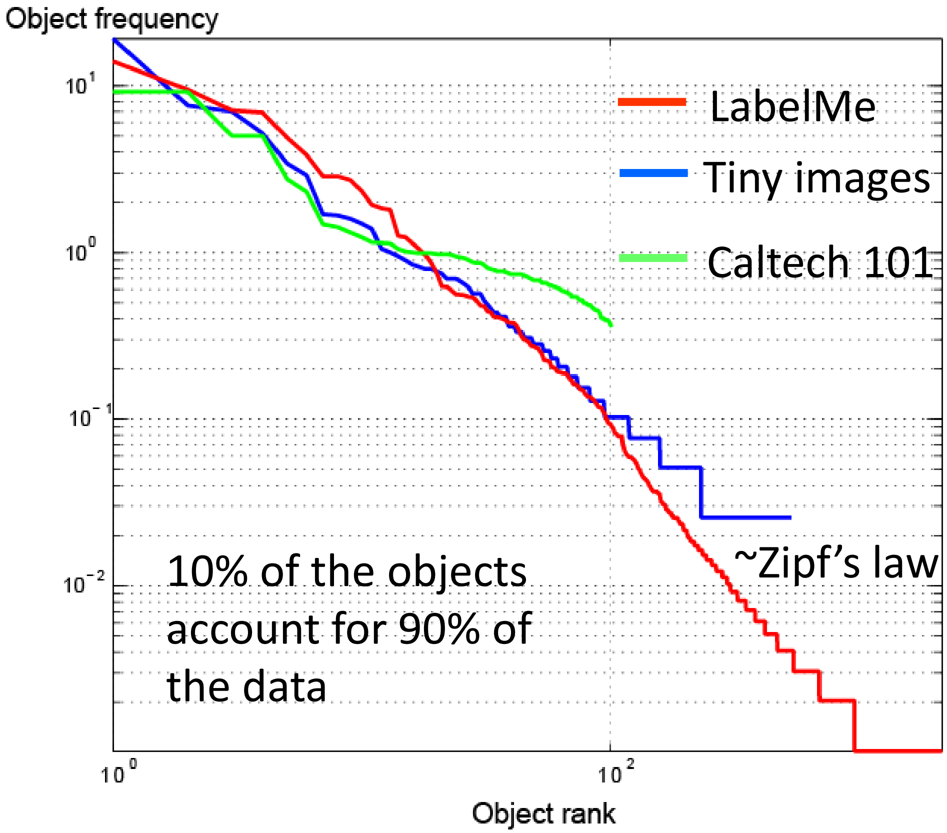


Training Set:
- 4,082 images
- 32,855 examples

Test Set:
- 9,518 images
- 75,362 examples

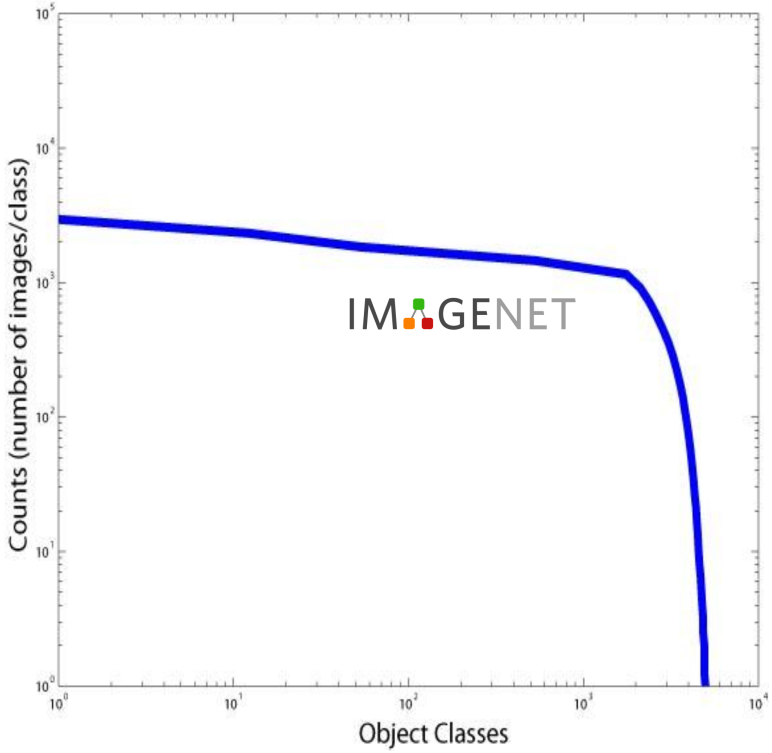
The first 9 objects account for 50% of all training examples
17 classes with more than 300 examples
109 classes with less than 50 examples

Object distributions



10% of the objects account for 90% of the data

~Zipf's law



The two extremes of learning

Extrapolation pr

Generalizatio

Diagnostic feat

erpolation problem

Correspondence

ding the differences

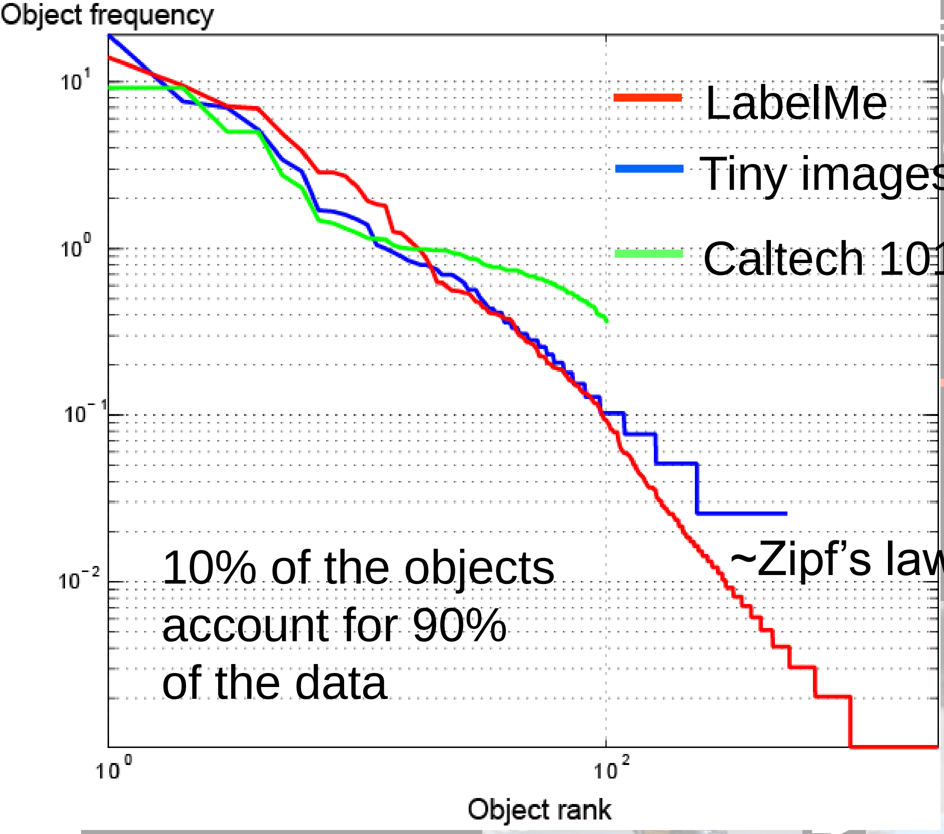
1

10

10⁶

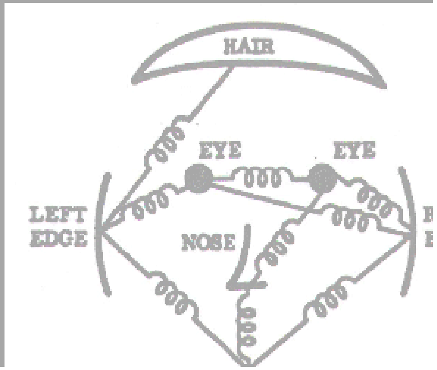
∞

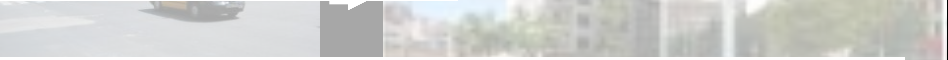
Number of training samples



10% of the objects account for 90% of the data

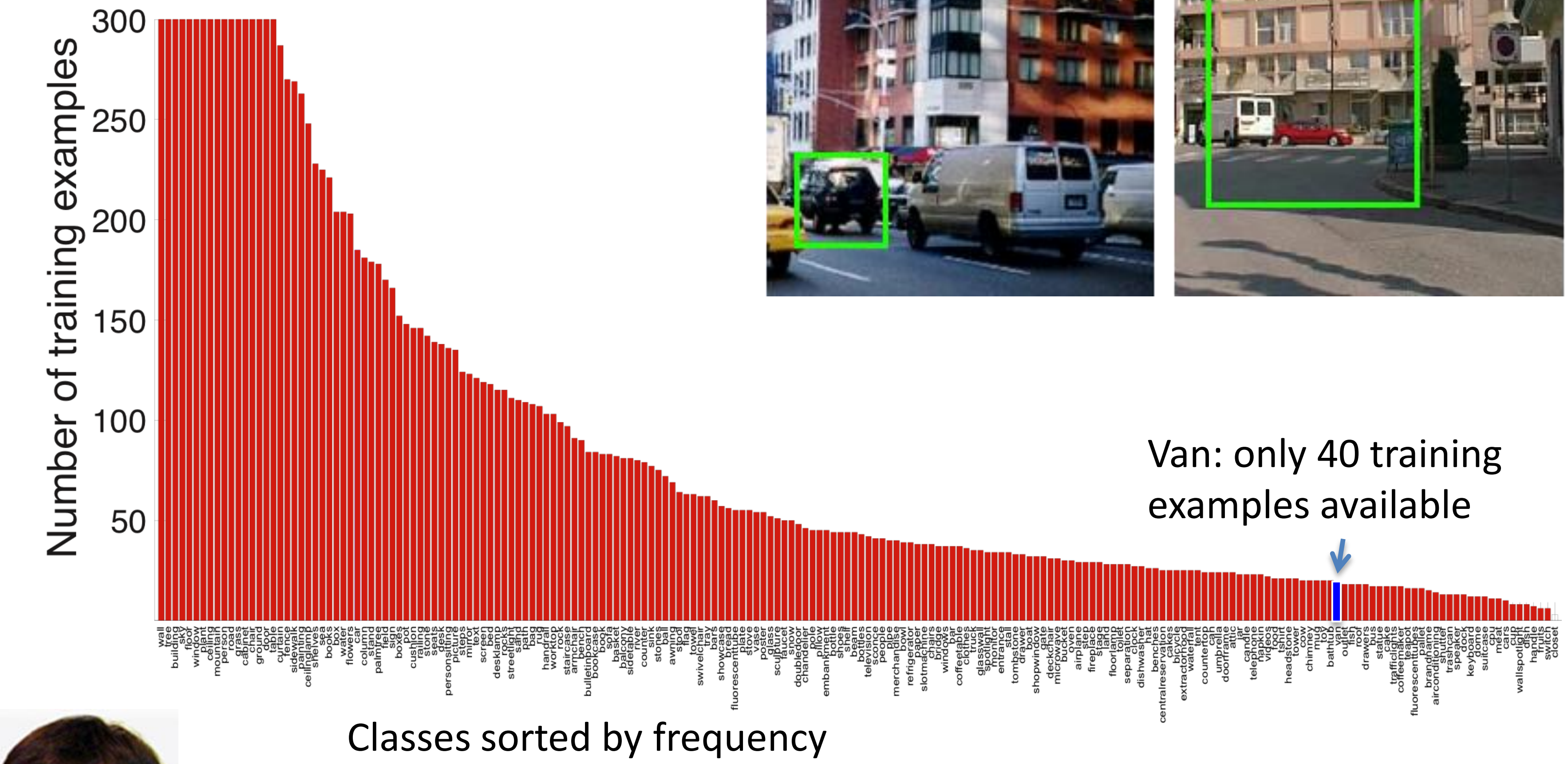
~Zipf's law





The two extremes of learning co-exist

SUN database

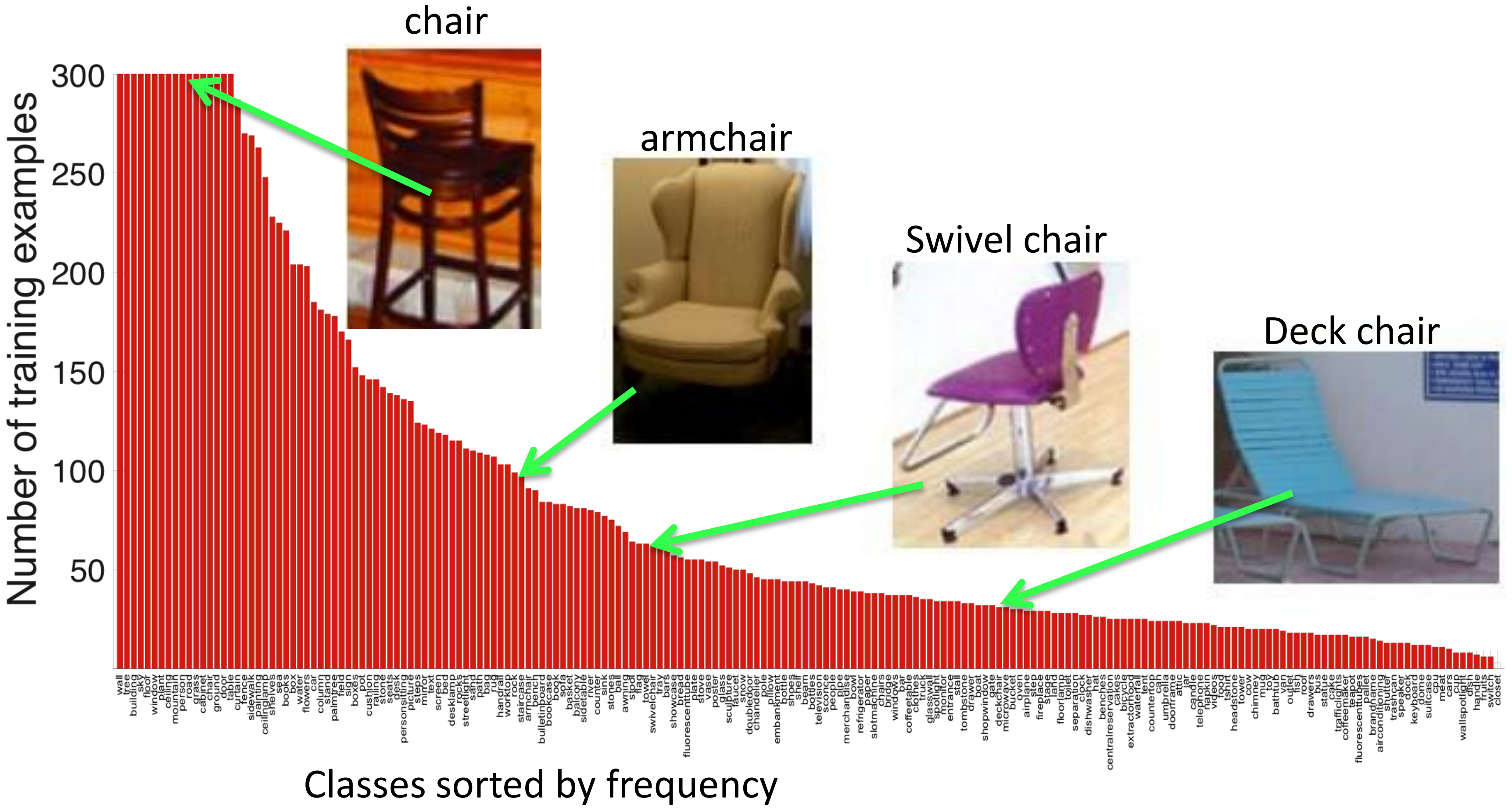


Van: only 40 training examples available

Ruslan Salakhutdinov

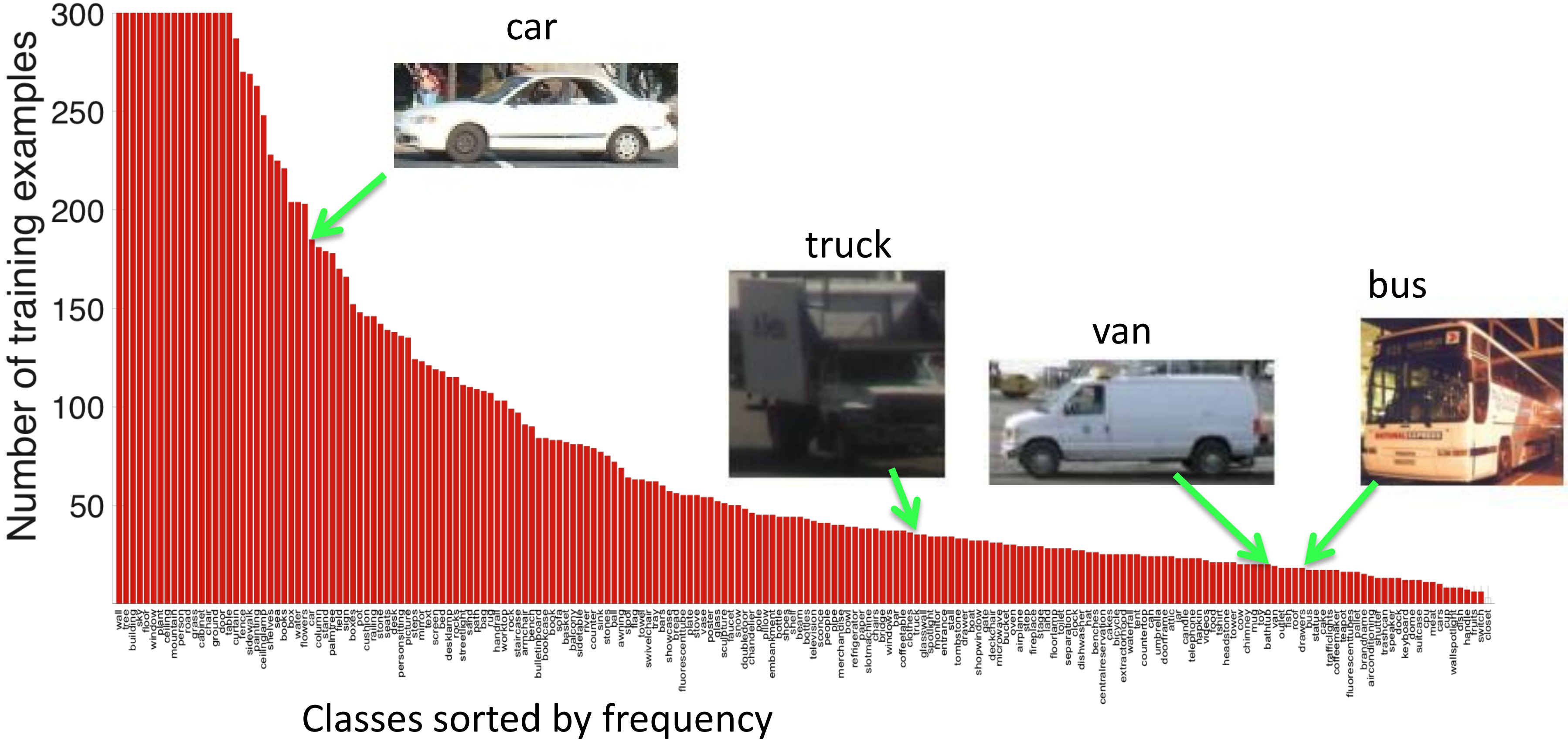
Salakhutdinov, Torralba, and Tenenbaum, MIT Technical Report, 2010

Rare objects are similar to frequent objects



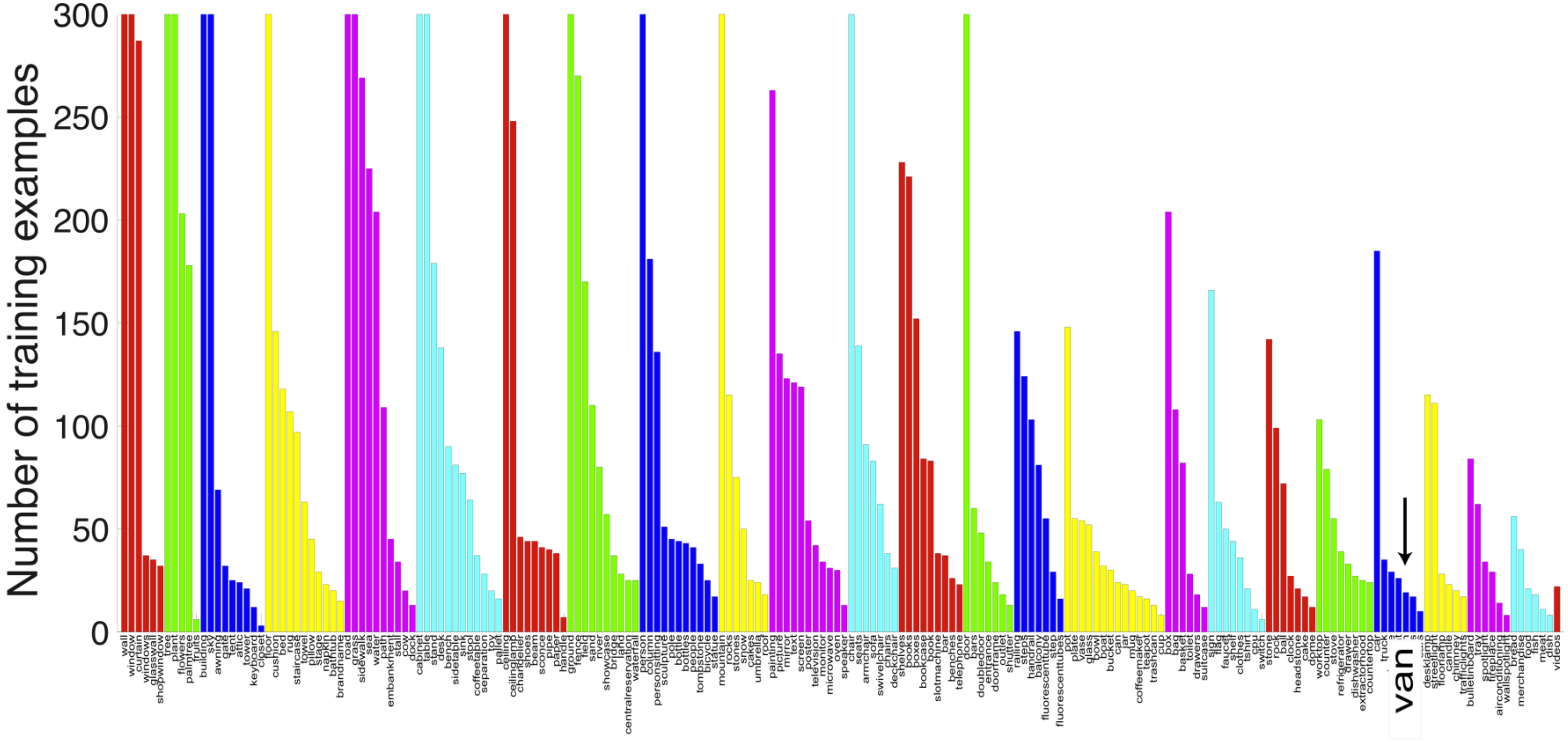
Salakhutdinov, Torralba, and Tenenbaum, MIT Technical Report, 2010

Rare objects are similar to frequent objects



Salakhutdinov, Torralba, and Tenenbaum, MIT Technical Report, 2010

Rare objects are similar to frequent objects

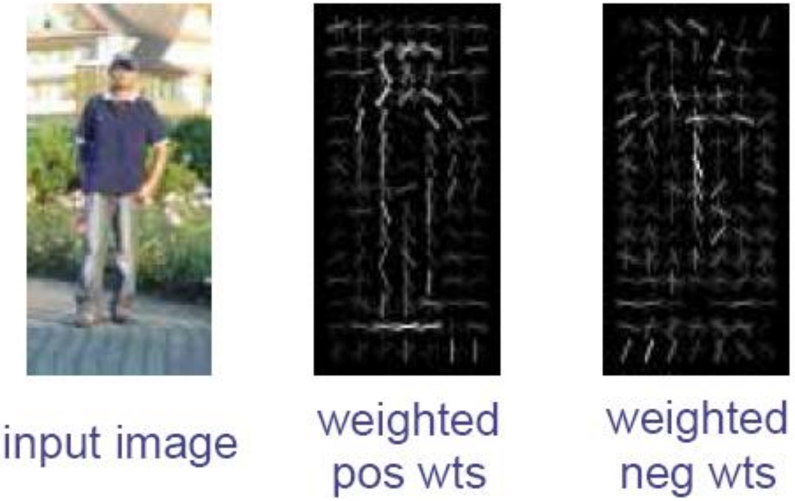


Classes sorted by similarity and frequency

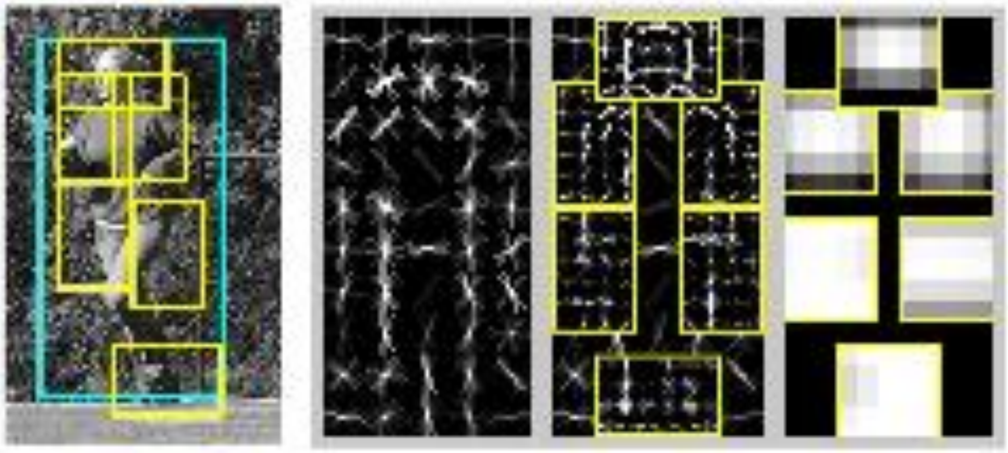
Salakhutdinov, Torralba, and Tenenbaum, MIT Technical Report, 2010

Detector

Dalal & Triggs, 2006



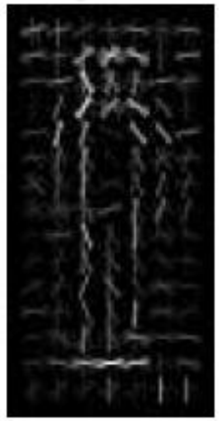
Felzenszwalb, McAllester & Ramanan, 2008

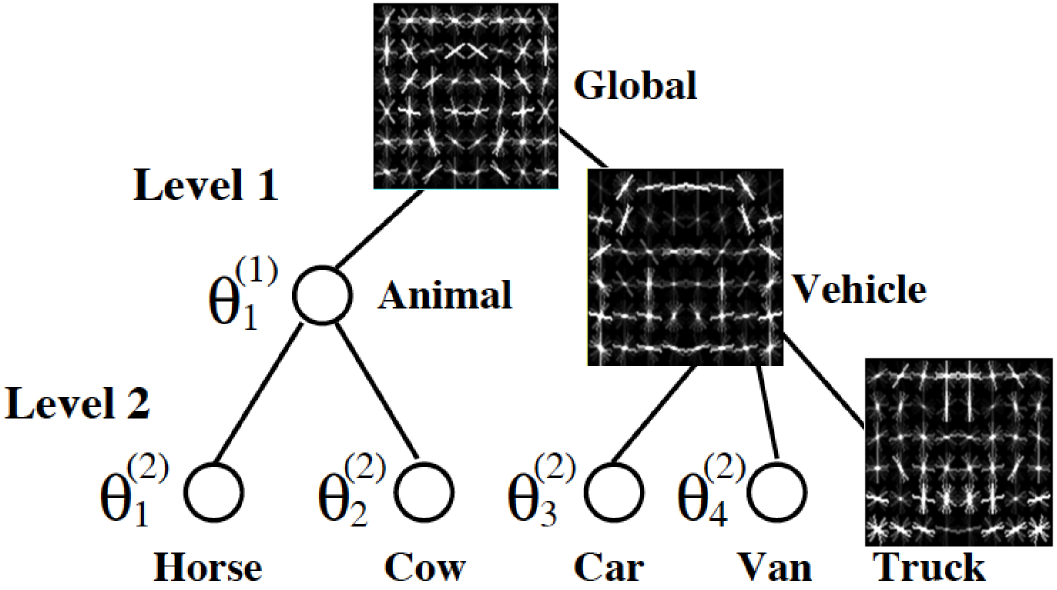


Salakhutdinov, Torralba, and Tenenbaum, MIT Technical Report, 2010

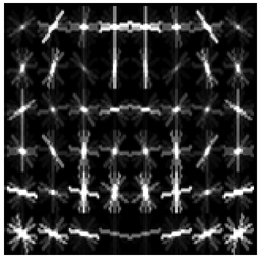
Generative model of classifier parameters

$y = \beta\ \phi(\text{patch})$



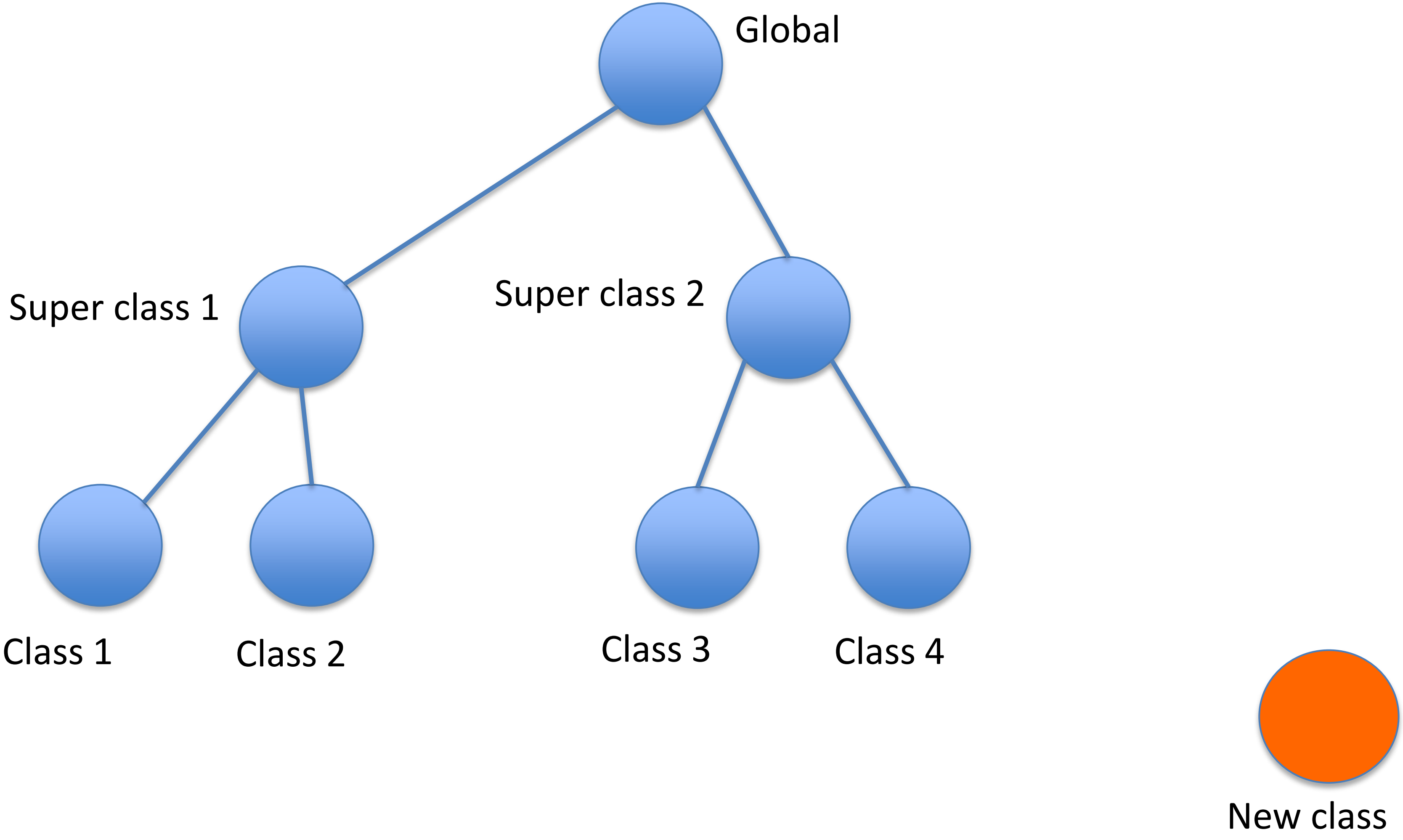


$$\beta = \theta^2 + \theta^1 + \theta^0 =$$



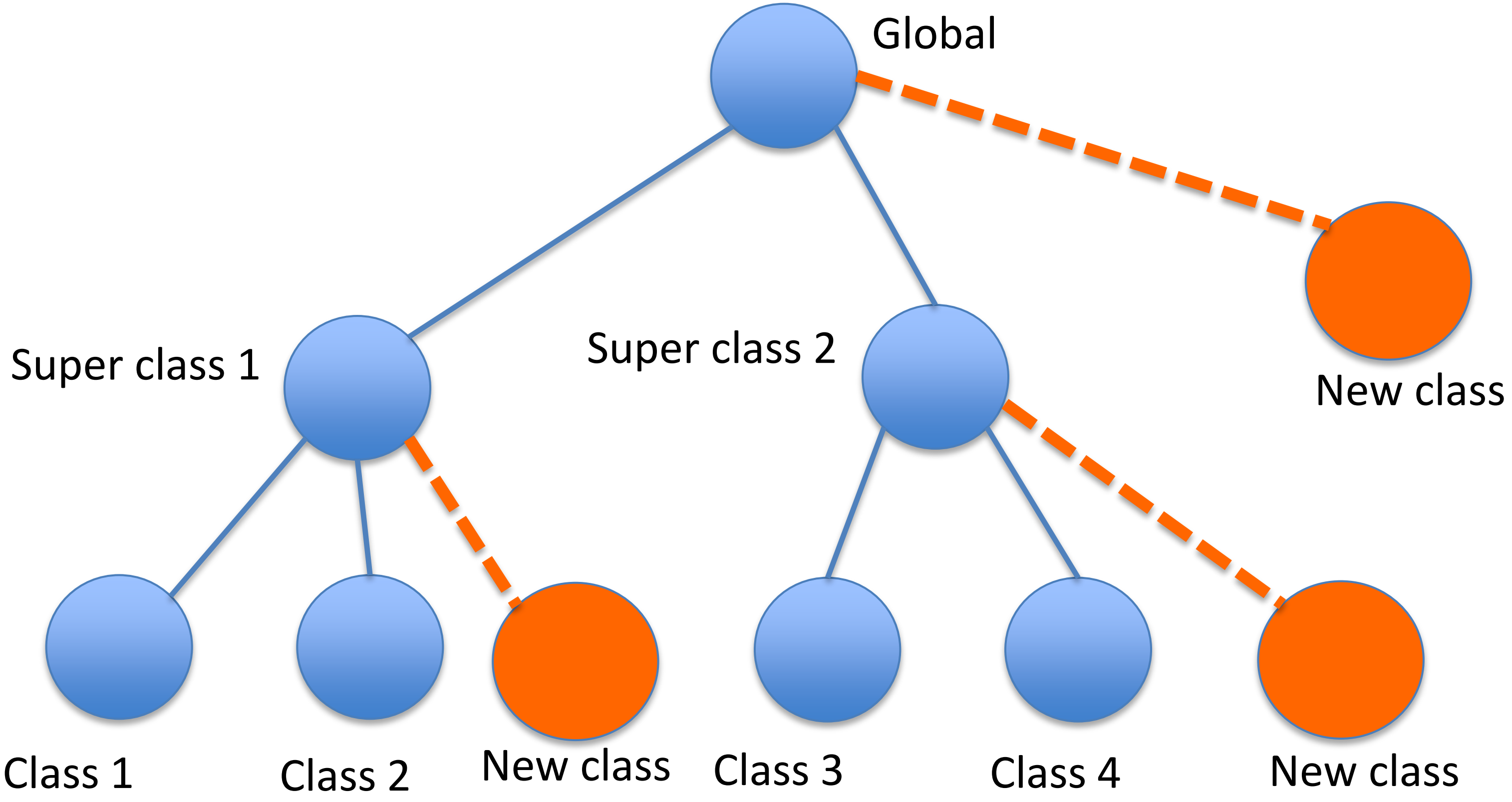
Salakhutdinov, Torralba, and Tenenbaum, MIT Technical Report, 2010

Building the tree

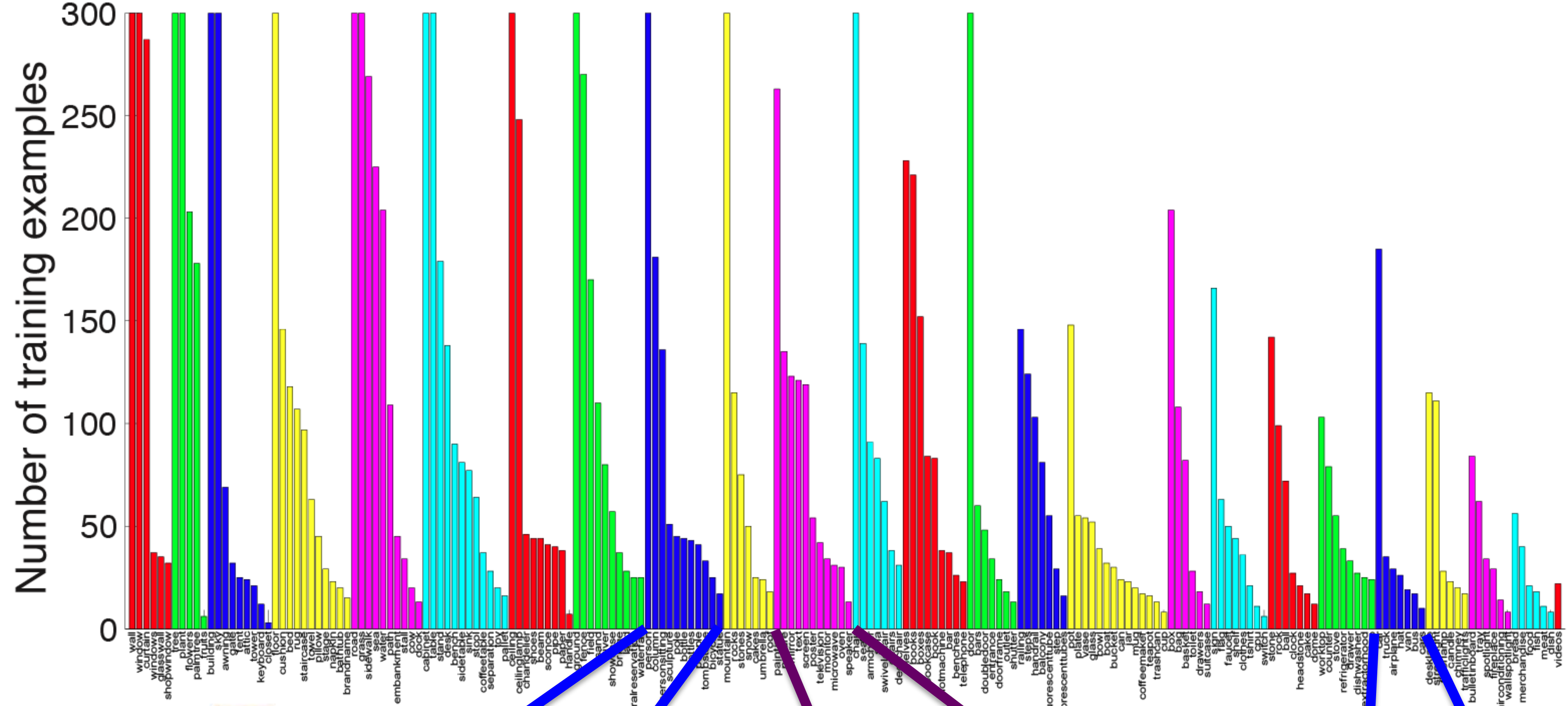


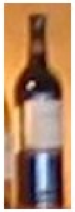
Salakhutdinov, Torralba, and Tenenbaum, CVPR, 2010

Building the tree



Salakhutdinov, Torralba, and Tenenbaum, MIT Technical Report, 2010

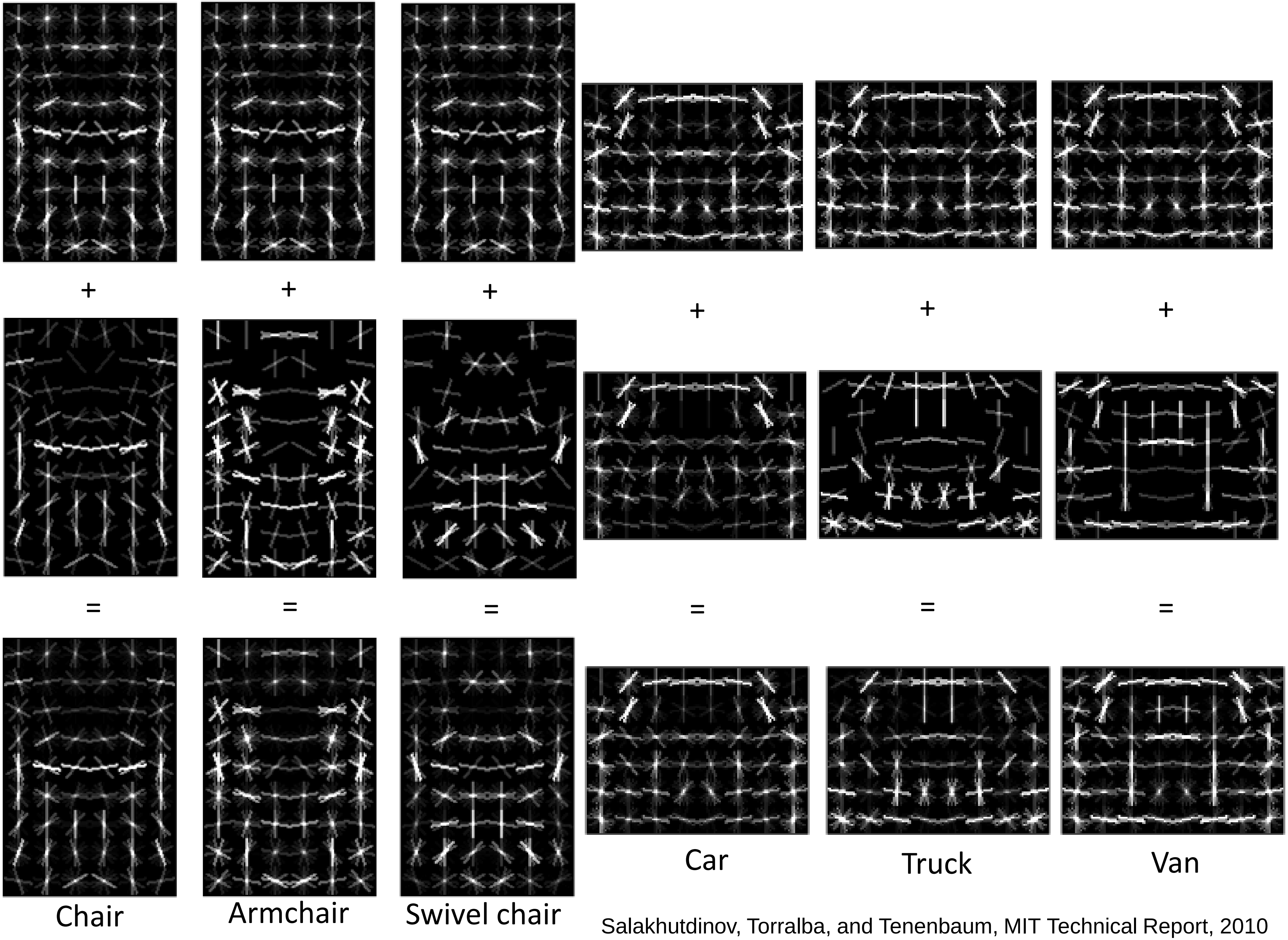




person
column
personsitting
sculpture
pole
bottle
bottles
people
tombstone
bicycle
statue

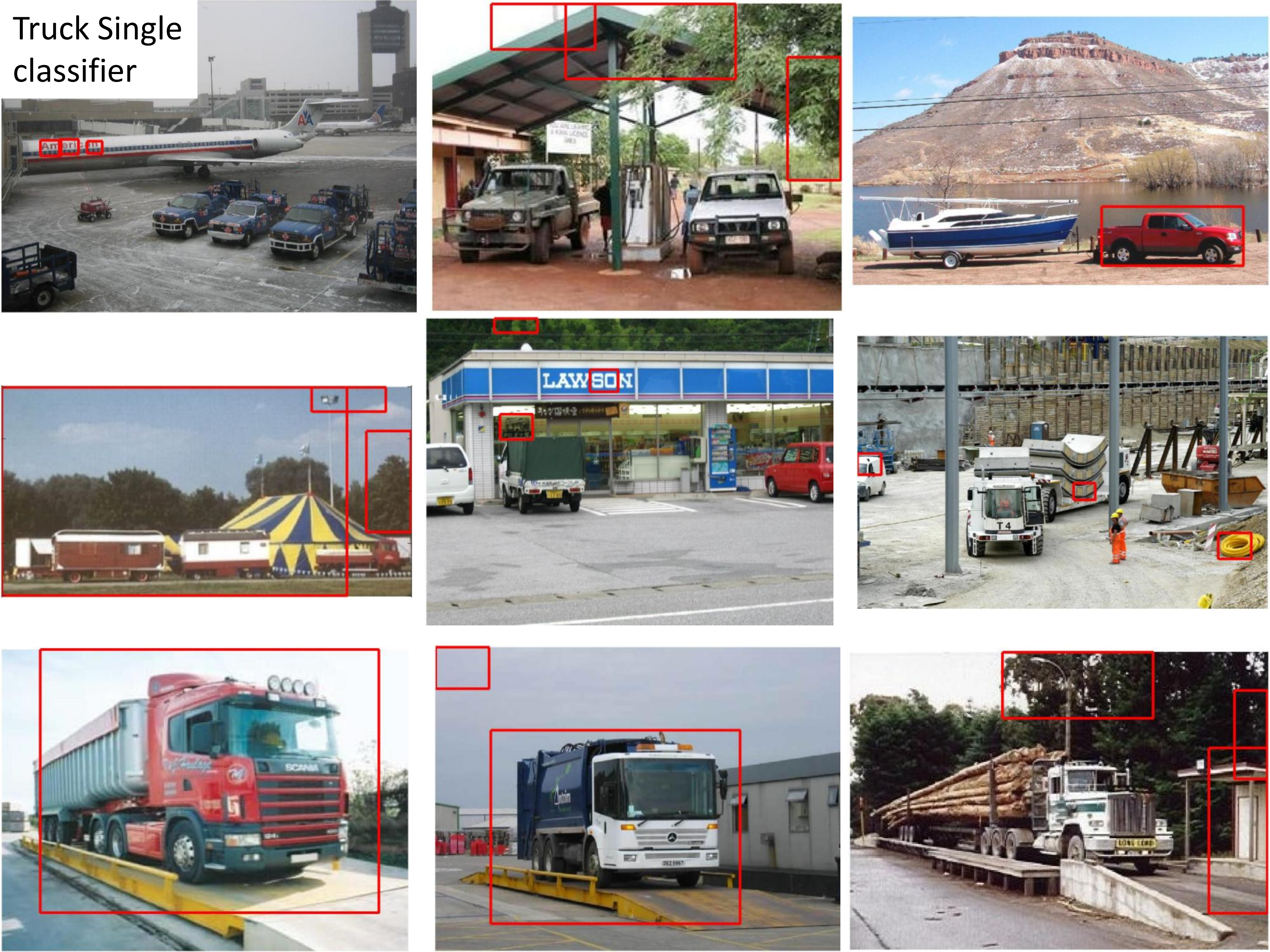
painting
picture
mirror
text
screen
poster
television
monitor
microwave
oven
speaker

car
truck
airplane
hat
van
bus
cars

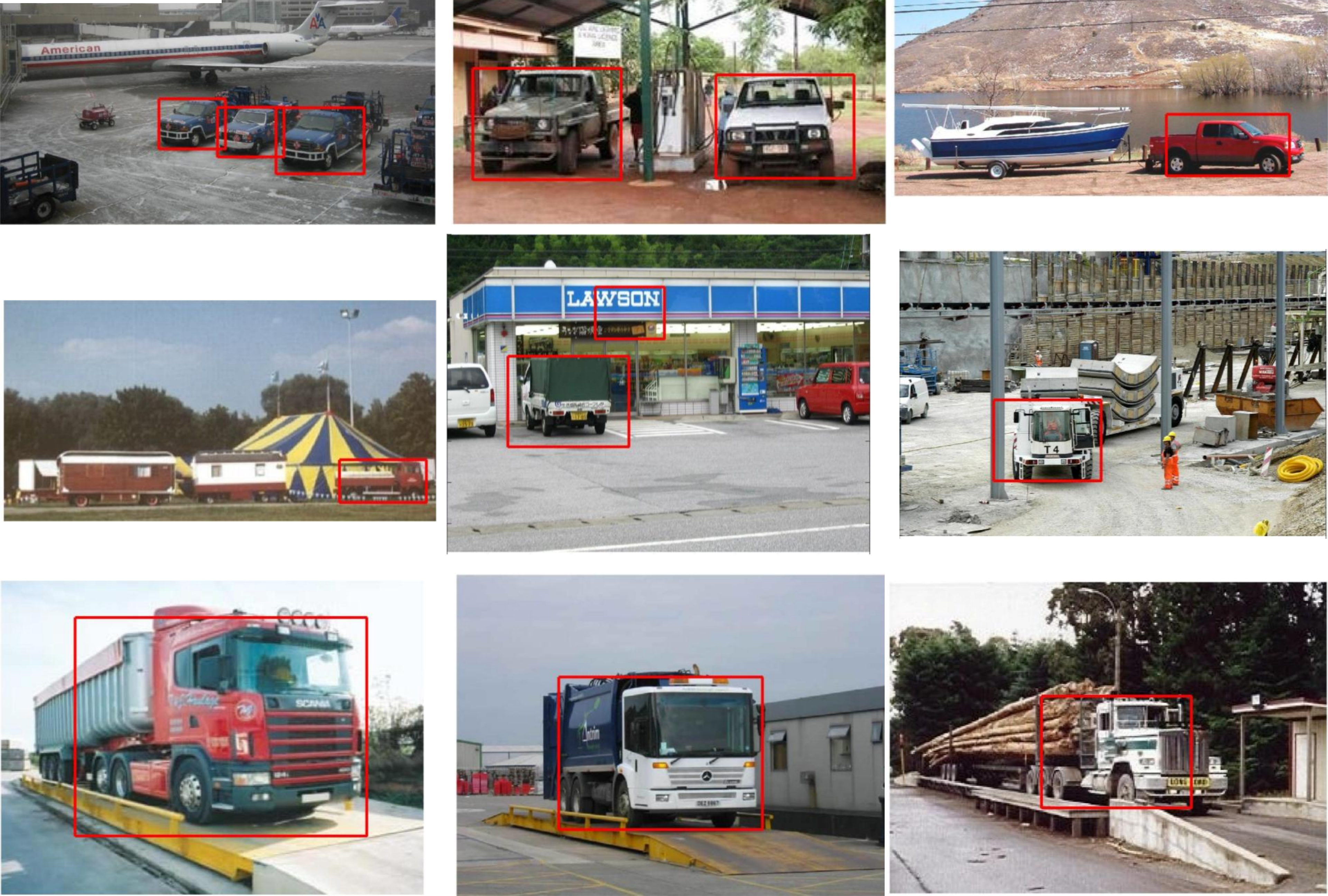


Salakhutdinov, Torralba, and Tenenbaum, MIT Technical Report, 2010

Truck Single classifier

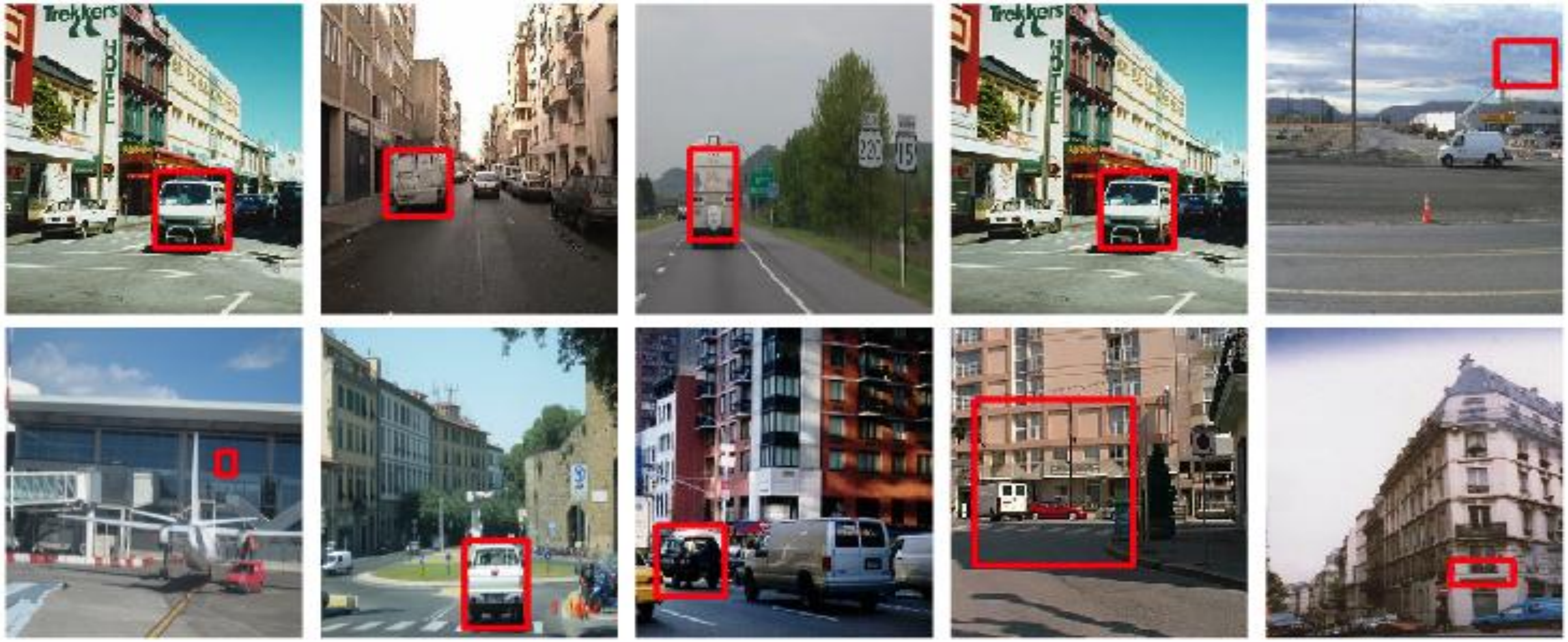


Truck Shared classifier

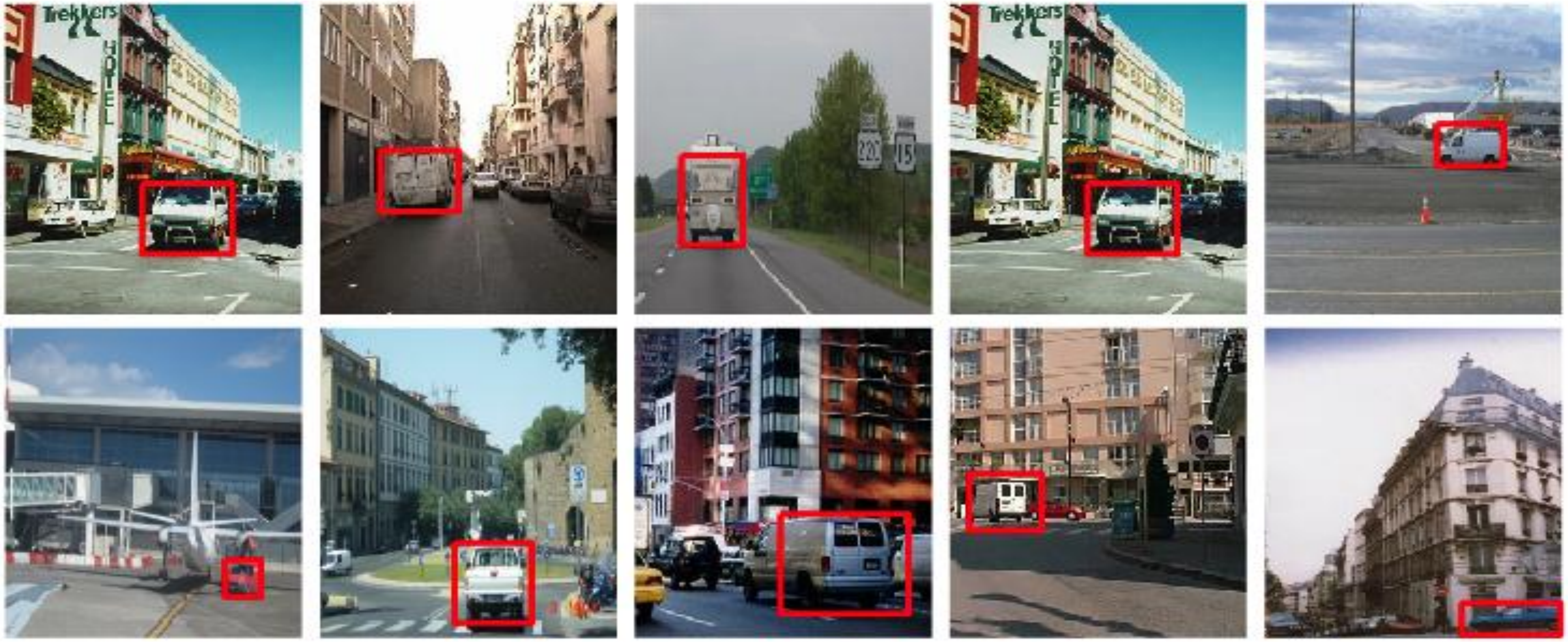


Vans

Single classifier

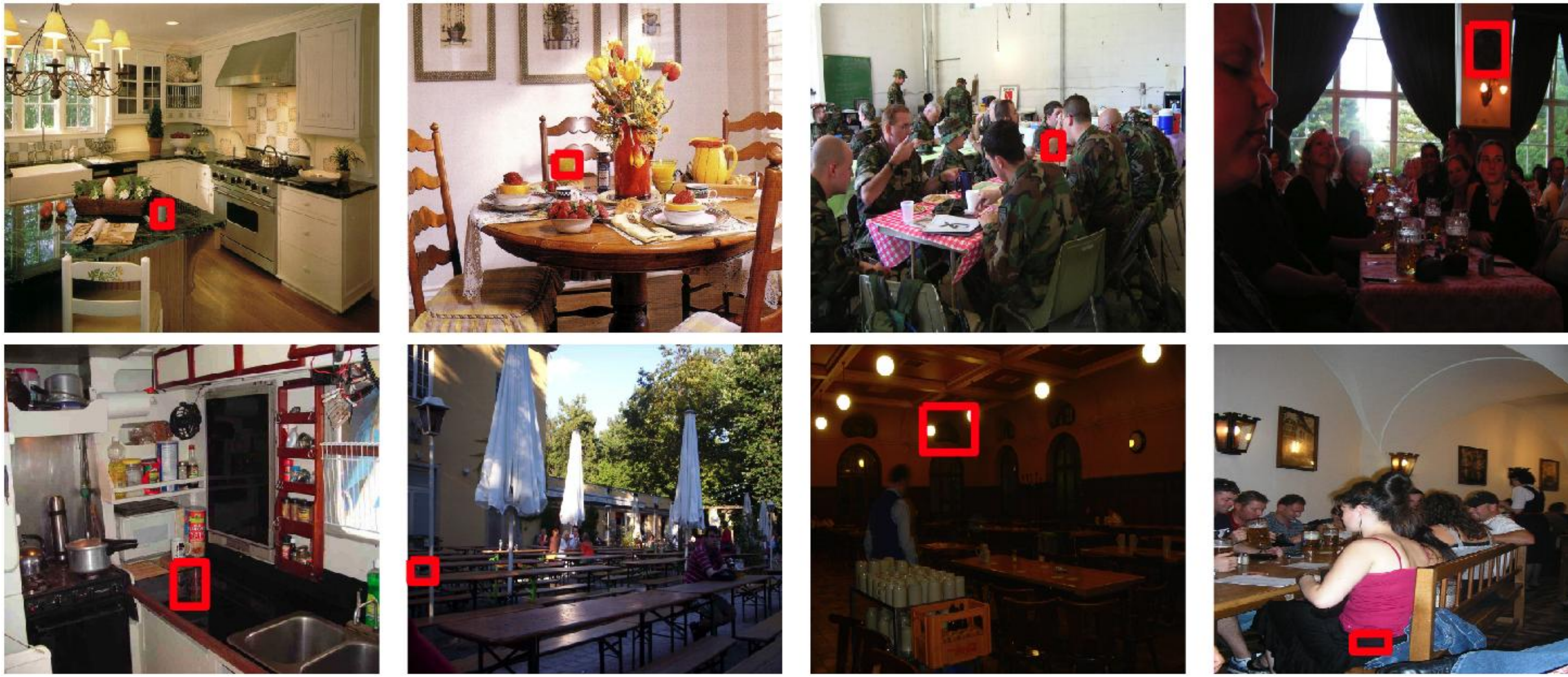


Shared classifier

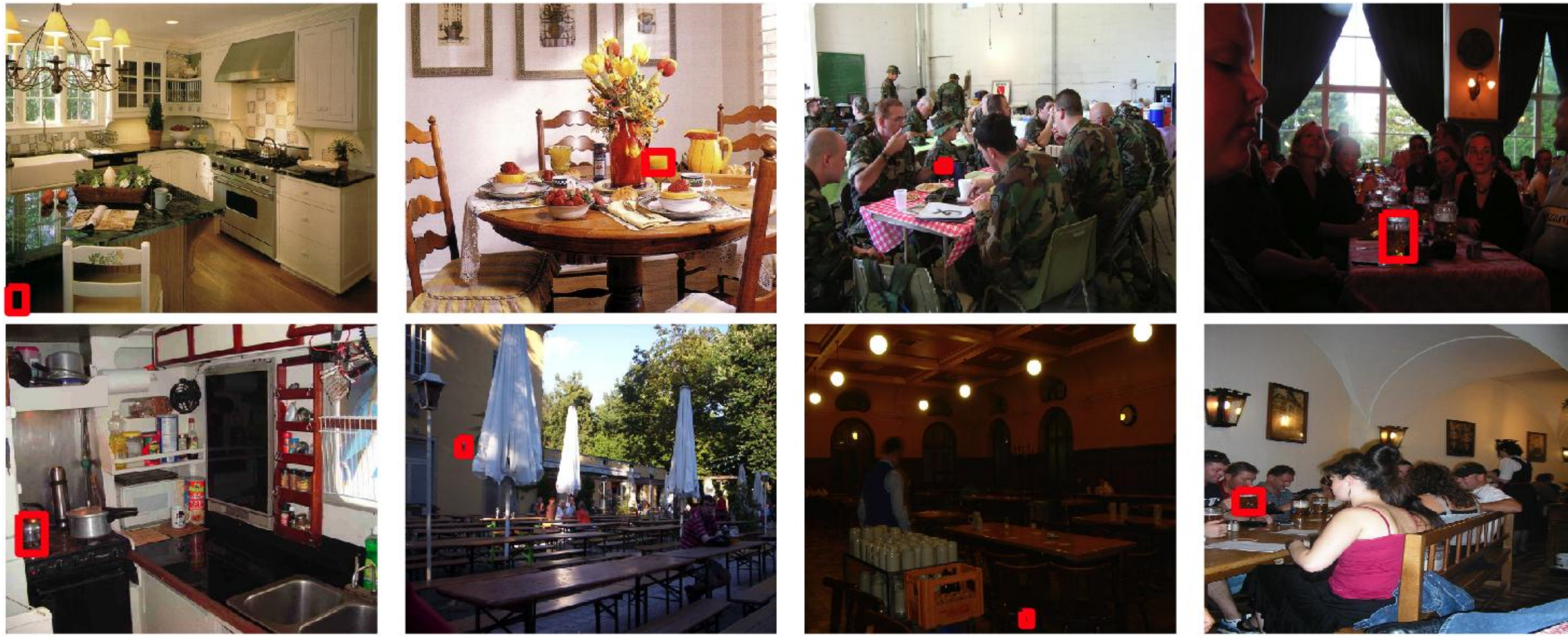


Mugs

Single classifier

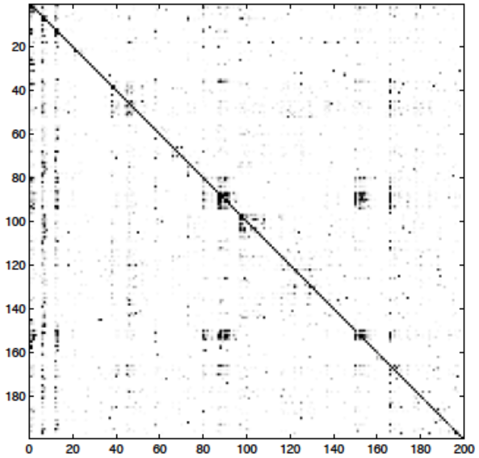


Shared classifier



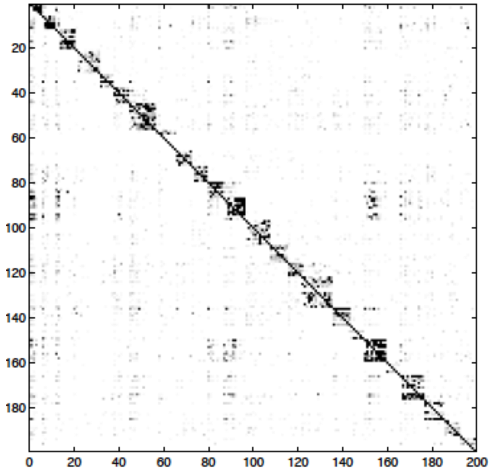
Confusions

Single classifier



Object Category	Three Most Confused Objects		
car (54.70)	van (4.92)	truck (2.43)	bus (1.02)
bus (10.54)	ceiling (3.03)	seats (1.82)	building (1.01)
truck (19.15)	sky (9.41)	building (4.81)	wall (1.07)
van (17.11)	car (9.21)	staircase (1.23)	building (0.87)
chair (22.84)	armchair (3.49)	stool (1.53)	deck chair (1.51)
deck chair (1.59)	ceiling (1.02)	sky (0.21)	wall (0.18)
armchair (19.77)	chair (2.15)	car (1.32)	wall (1.21)
table (18.61)	stool (9.11)	desk (6.63)	coffee table (1.85)
coffee table (2.38)	cakes (1.02)	chair (0.98)	bucket (0.48)
desk (11.62)	floor (4.55)	table (1.02)	wall (0.97)

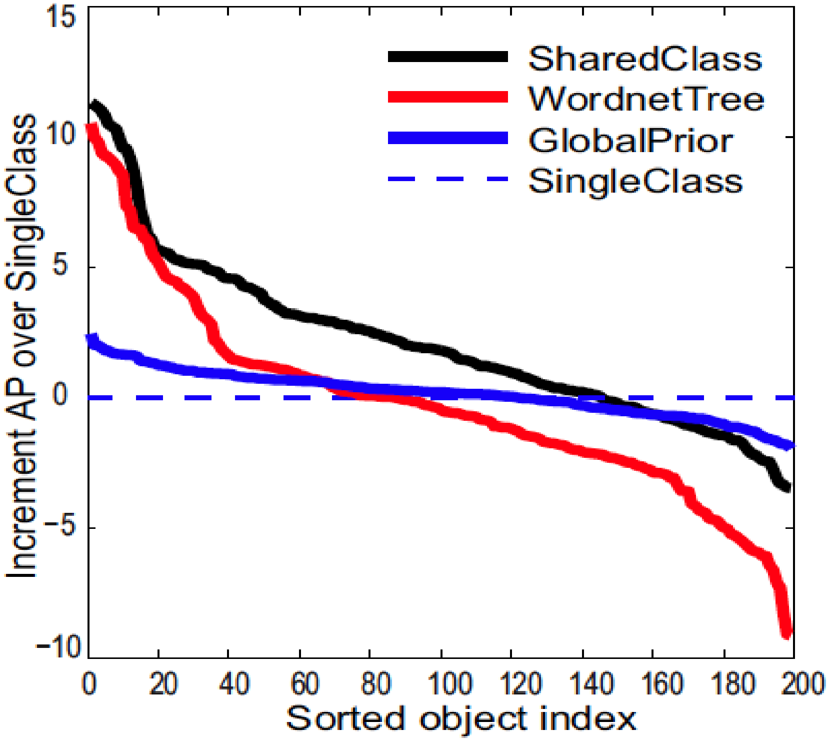
Shared classifier



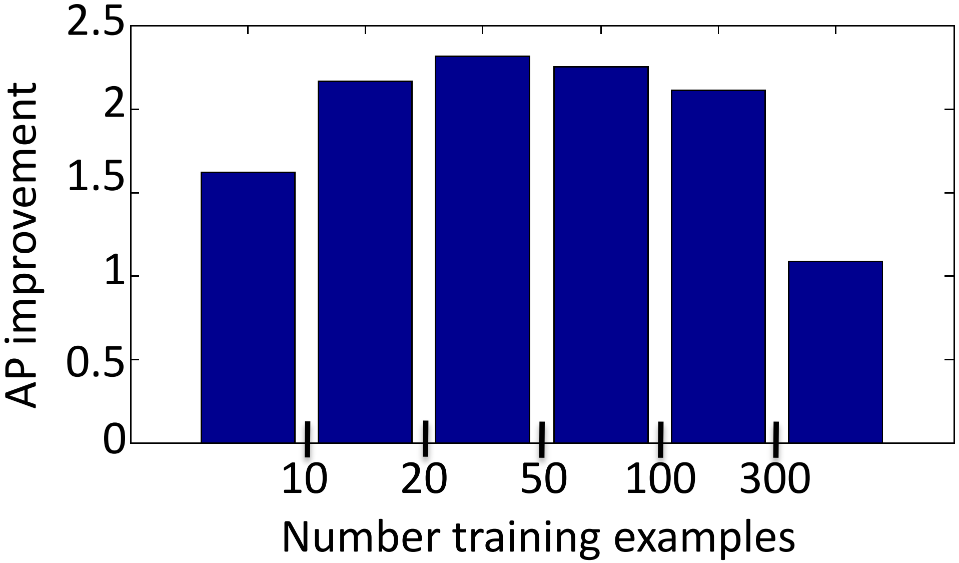
Object Category	Three Most Confused Objects		
car (55.20)	van (4.99)	truck (2.41)	bus (1.05)
bus (19.54)	car (5.03)	van (3.82)	truck (2.01)
truck (29.54)	car (4.41)	van (2.87)	bus (1.23)
van (28.09)	car (4.02)	truck (1.31)	bus (1.24)
chair (23.65)	armchair (3.29)	stool (1.59)	deck chair (1.64)
deck chair (12.78)	chair (1.38)	armchair (0.97)	table (0.17)
armchair (26.78)	chair (3.32)	deck chair (2.08)	sofa (1.21)
table (19.03)	stool (9.34)	desk (2.63)	coffee table (2.13)
coffee table (13.16)	table (3.06)	side table (0.98)	stand (0.79)
desk (18.07)	stand (2.55)	table (1.54)	armchair (1.21)

Salakhutdinov, Torralba, and Tenenbaum, MIT Technical Report, 2010

Improvement over baseline

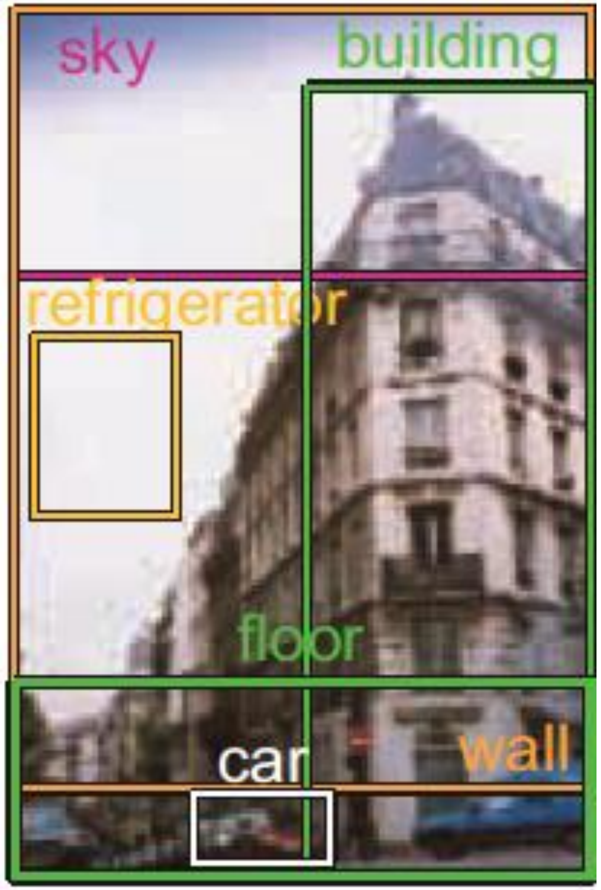


Improvement as a function of amount of training data

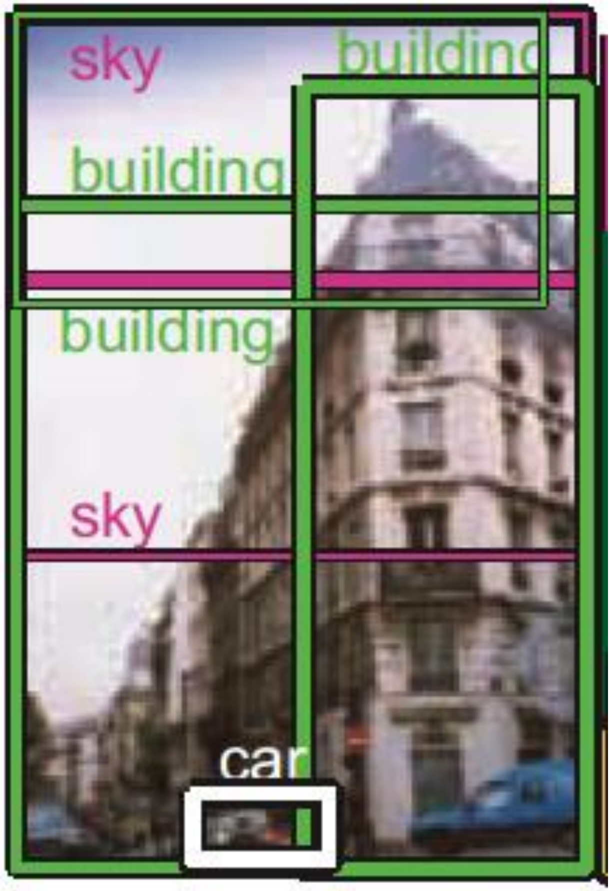


Salakhutdinov, Torralba, and Tenenbaum, MIT Technical Report, 2010

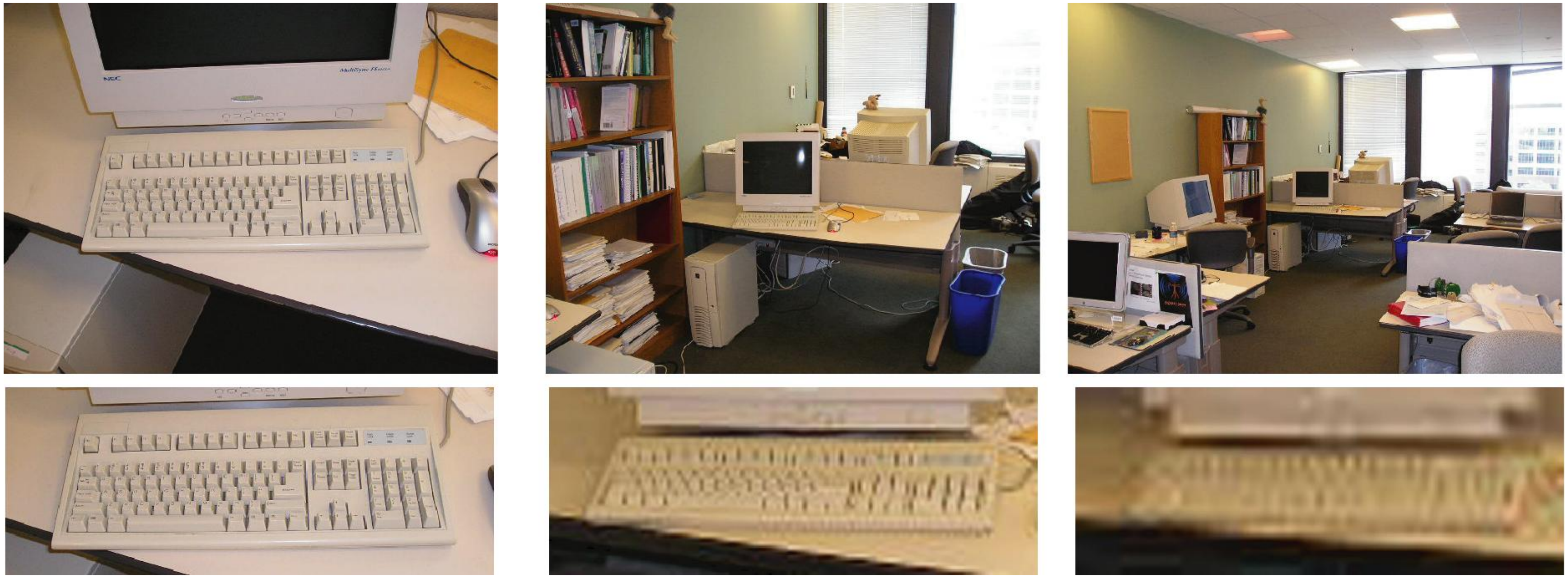
Detector output

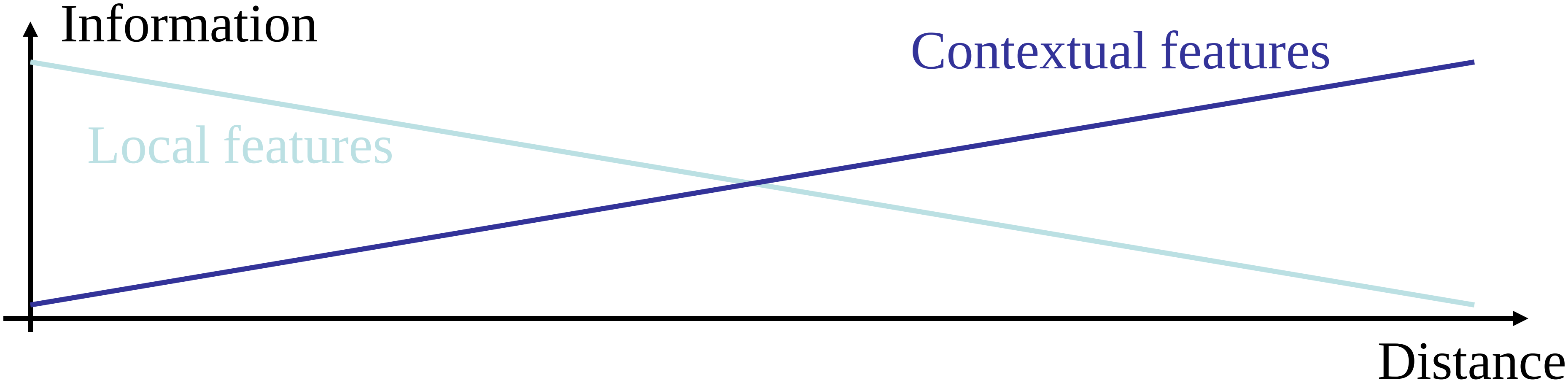


Improved with context reasoning



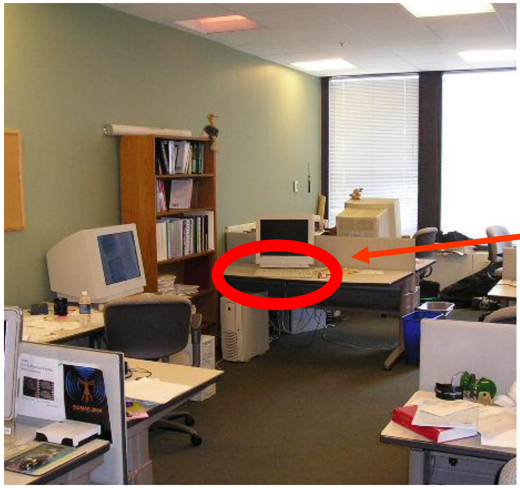
Is local information even enough?



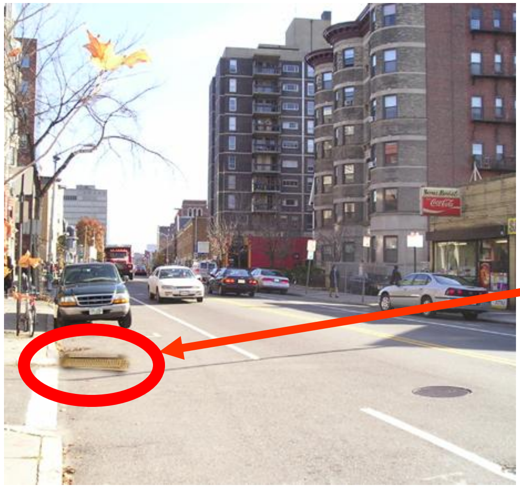


The system does not care about the scene, but we do…

We know there is a keyboard present in this scene even if we cannot see it clearly.



We know there is no keyboard present in this scene

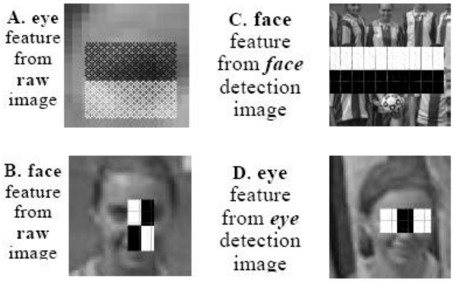


… even if there is one indeed.

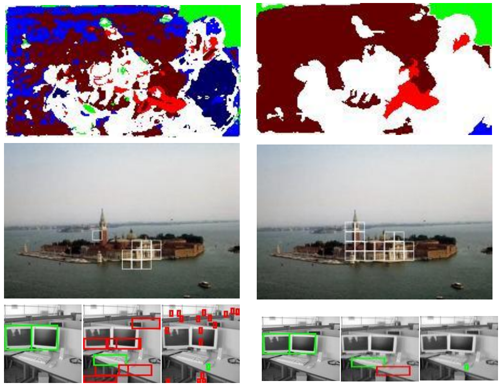
Objects in context

Torralba, Sinha (2001)

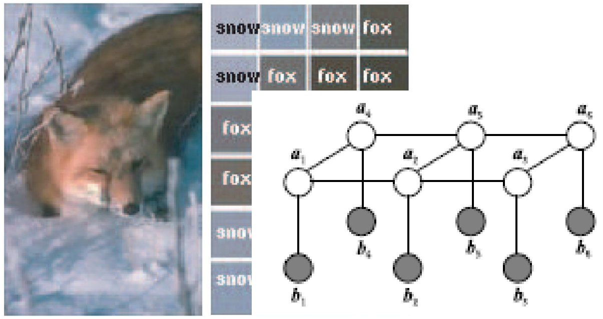
Fink & Perona (2003)


A. eye feature from raw image
C. face feature from *face* detection image
B. face feature from raw image
D. eye feature from *eye* detection image

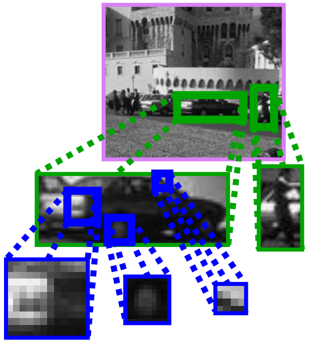
Kumar, Hebert (2005)



Carbonetto, de Freitas & Barnard (2004)

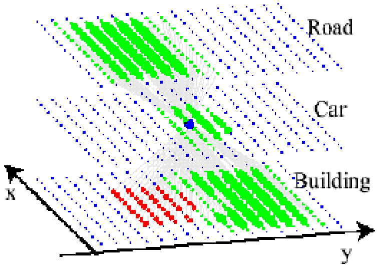


Sudderth, Torralba, Wilsky, Freeman (2005)



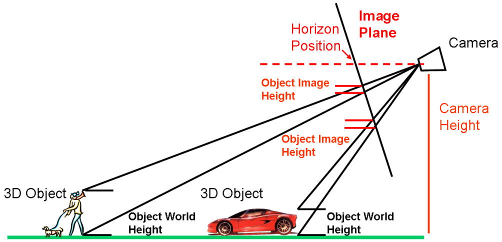
Heitz and Koller (2008)

Torralba Murphy Freeman (2004)

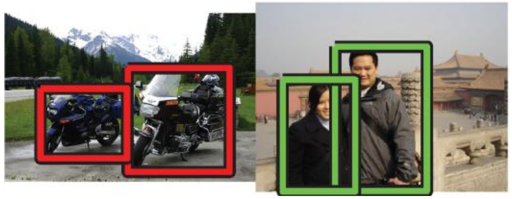


Rabinovich et al (2007)

Hoiem, Efros, Hebert (2005)



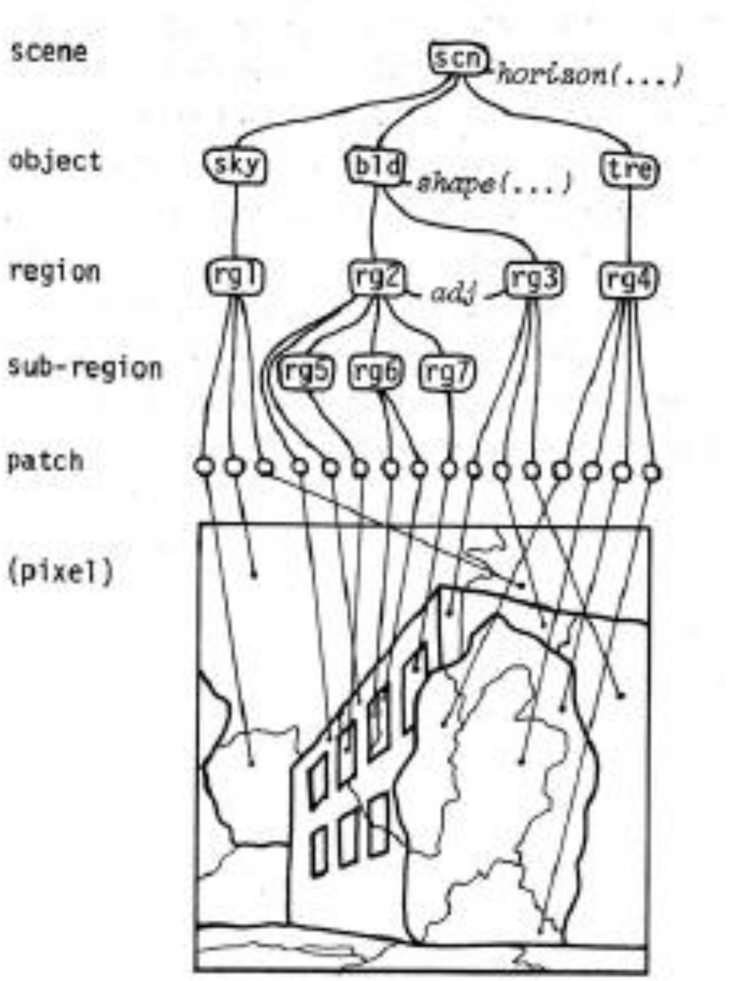
Desai, Ramanan, and Fowlkes (2009)



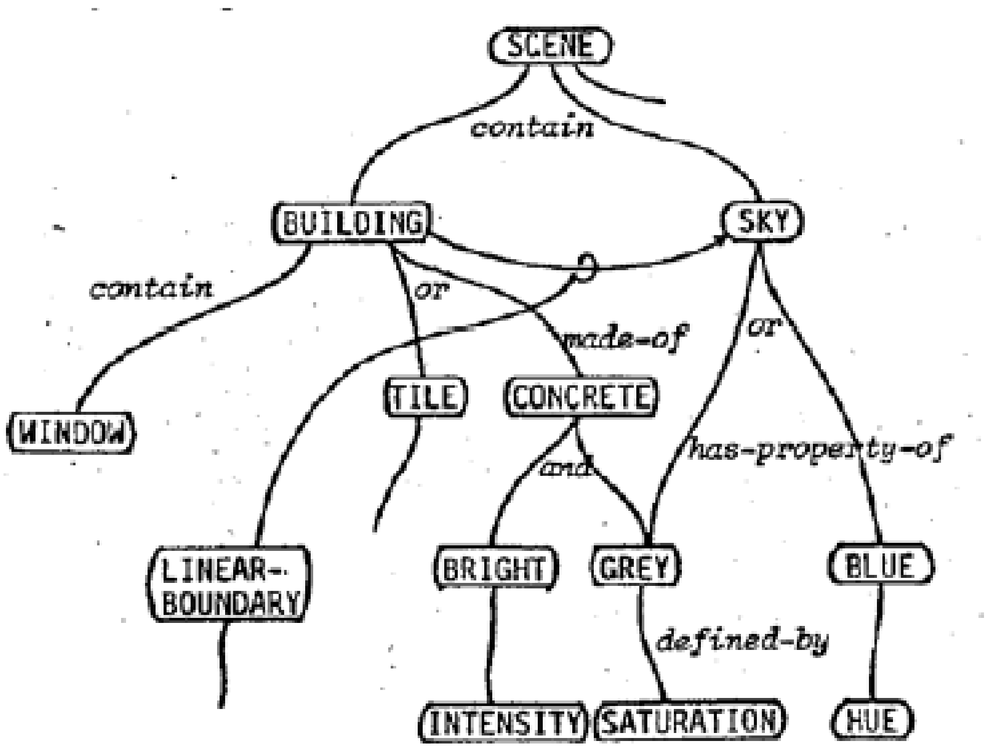
Issues:

- Lack of a good benchmark
- Focus on improving detection

Grammars



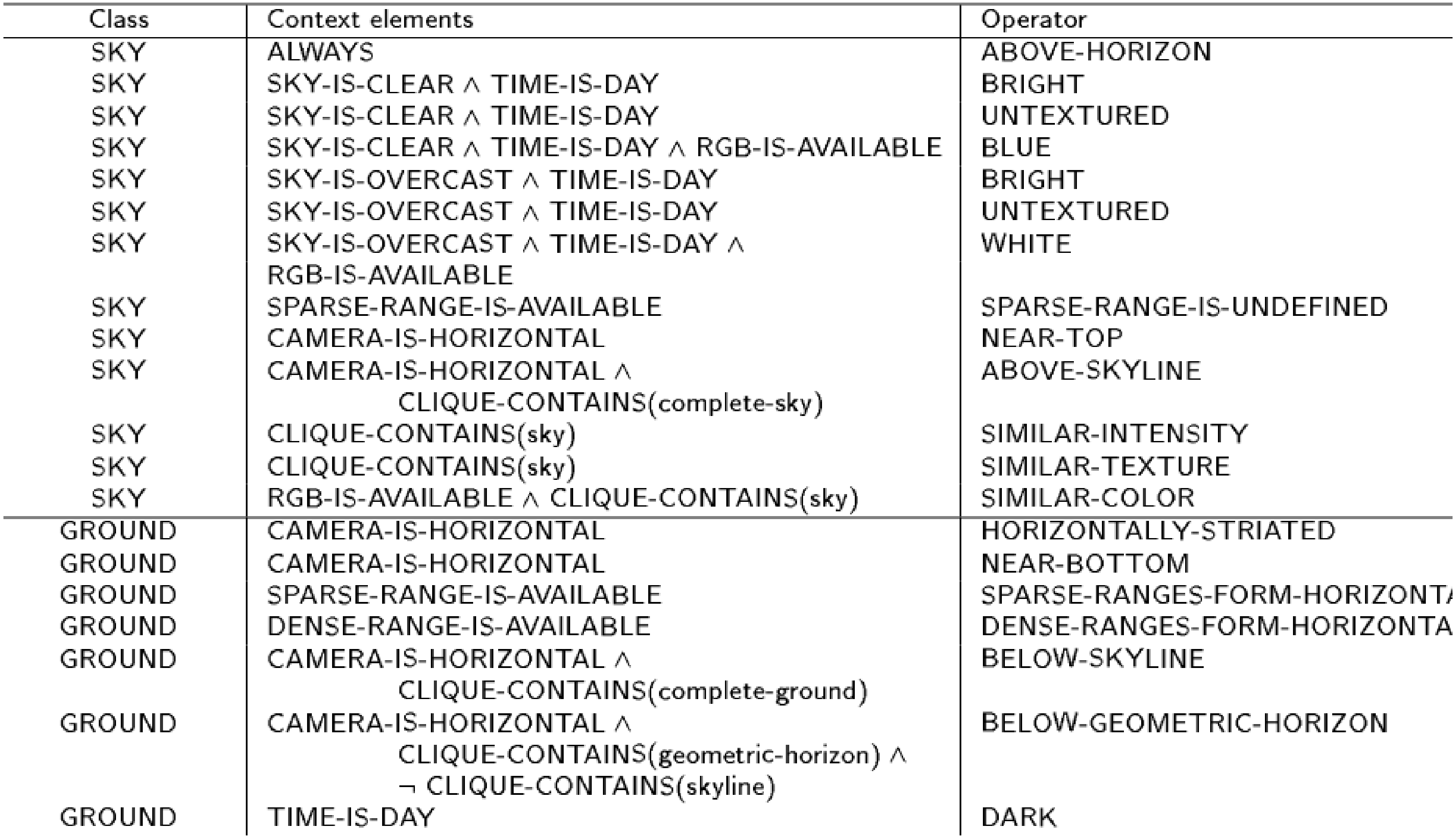
[Ohta & Kanade 1978]



- Guzman (*SEE*), 1968
- Noton and Stark 1971
- Hansen & Riseman (*VISIONS*), 1978
- Barrow & Tenenbaum 1978
- Brooks (*ACRONYM*), 1979
- Marr, 1982
- Yakimovsky & Feldman, 1973

CONDOR system

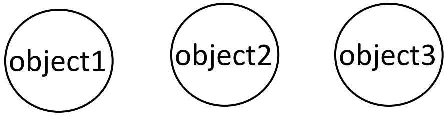
Strat and Fischler (1991)



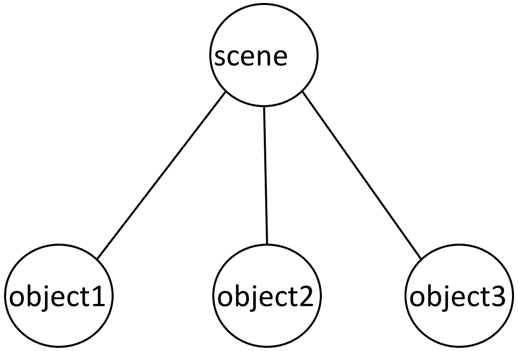
Class	Context elements	Operator
SKY	ALWAYS	ABOVE-HORIZON
SKY	SKY-IS-CLEAR ∧ TIME-IS-DAY	BRIGHT
SKY	SKY-IS-CLEAR ∧ TIME-IS-DAY	UNTEXTURED
SKY	SKY-IS-CLEAR ∧ TIME-IS-DAY ∧ RGB-IS-AVAILABLE	BLUE
SKY	SKY-IS-OVERCAST ∧ TIME-IS-DAY	BRIGHT
SKY	SKY-IS-OVERCAST ∧ TIME-IS-DAY	UNTEXTURED
SKY	SKY-IS-OVERCAST ∧ TIME-IS-DAY ∧ RGB-IS-AVAILABLE	WHITE
SKY	SPARSE-RANGE-IS-AVAILABLE	SPARSE-RANGE-IS-UNDEFINED
SKY	CAMERA-IS-HORIZONTAL	NEAR-TOP
SKY	CAMERA-IS-HORIZONTAL ∧ CLIQUE-CONTAINS(complete-sky)	ABOVE-SKYLINE
SKY	CLIQUE-CONTAINS(sky)	SIMILAR-INTENSITY
SKY	CLIQUE-CONTAINS(sky)	SIMILAR-TEXTURE
SKY	RGB-IS-AVAILABLE ∧ CLIQUE-CONTAINS(sky)	SIMILAR-COLOR
GROUND	CAMERA-IS-HORIZONTAL	HORIZONTALLY-STRIATED
GROUND	CAMERA-IS-HORIZONTAL	NEAR-BOTTOM
GROUND	SPARSE-RANGE-IS-AVAILABLE	SPARSE-RANGES-FORM-HORIZONT
GROUND	DENSE-RANGE-IS-AVAILABLE	DENSE-RANGES-FORM-HORIZONTA
GROUND	CAMERA-IS-HORIZONTAL ∧ CLIQUE-CONTAINS(complete-ground)	BELOW-SKYLINE
GROUND	CAMERA-IS-HORIZONTAL ∧ CLIQUE-CONTAINS(geometric-horizon) ∧ ¬ CLIQUE-CONTAINS(skyline)	BELOW-GEOMETRIC-HORIZON
GROUND	TIME-IS-DAY	DARK

- Guzman (*SEE*), 1968
- Noton and Stark 1971
- Hansen & Riseman (*VISIONS*), 1978
- Barrow & Tenenbaum 1978
- Brooks (*ACRONYM*), 1979
- Marr, 1982
- Ohta & Kanade, 1978
- Yakimovsky & Feldman, 1973

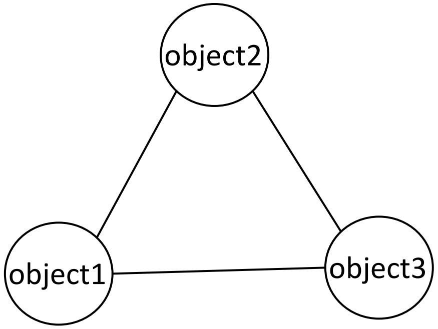
Context models



Independent model

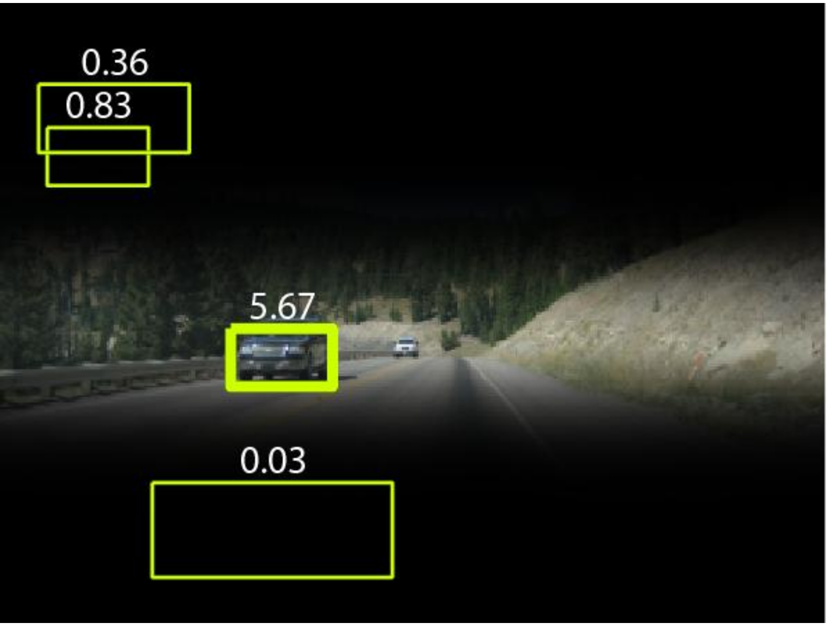


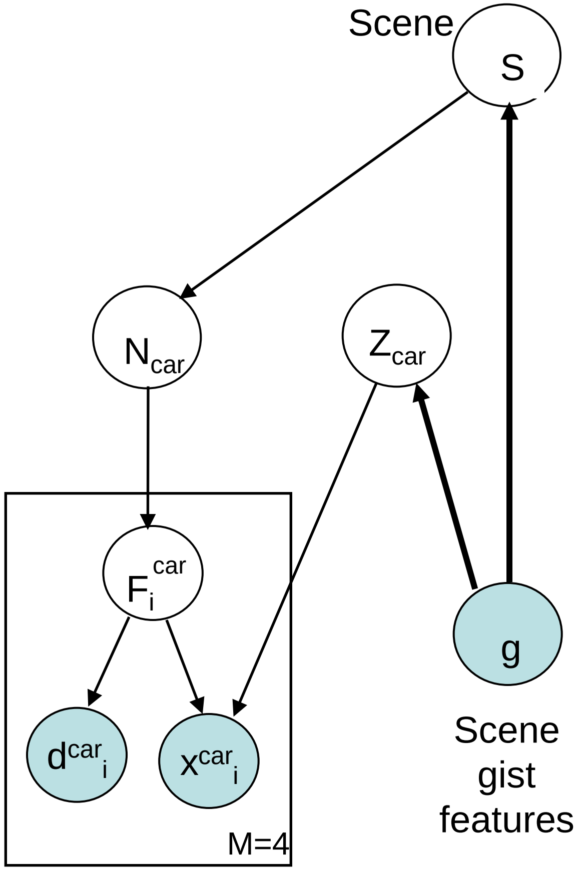
Objects are correlated via the scene



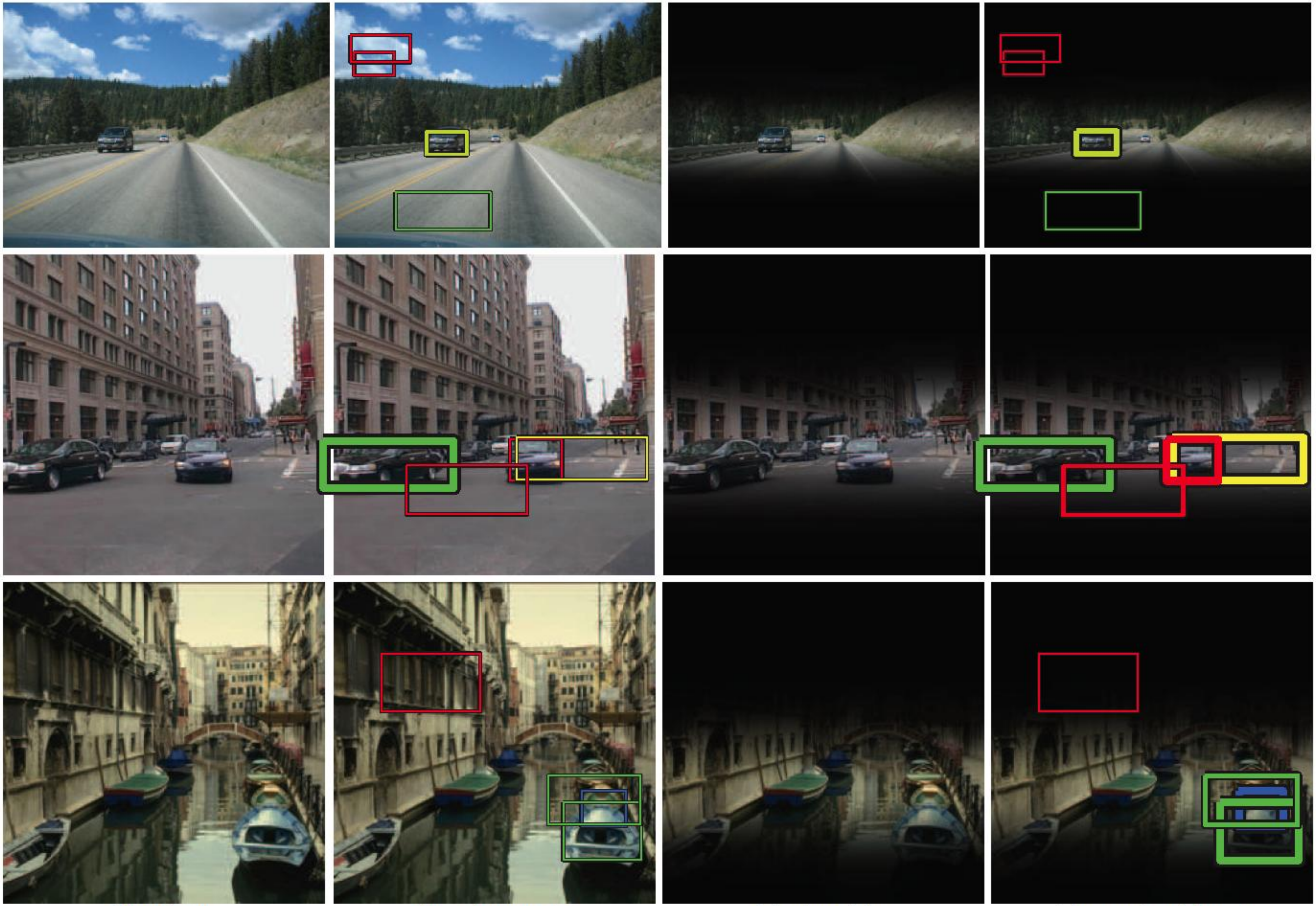
Dependencies among objects

An integrated model of Scenes, Objects, and Parts





Torralba, Sinha (ICCV 2001), Torralba, Murphy, Freeman 2010



a) input image

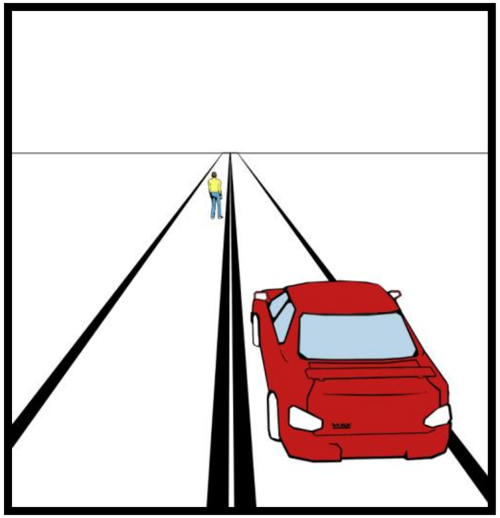
b) car detector output

c) location priming

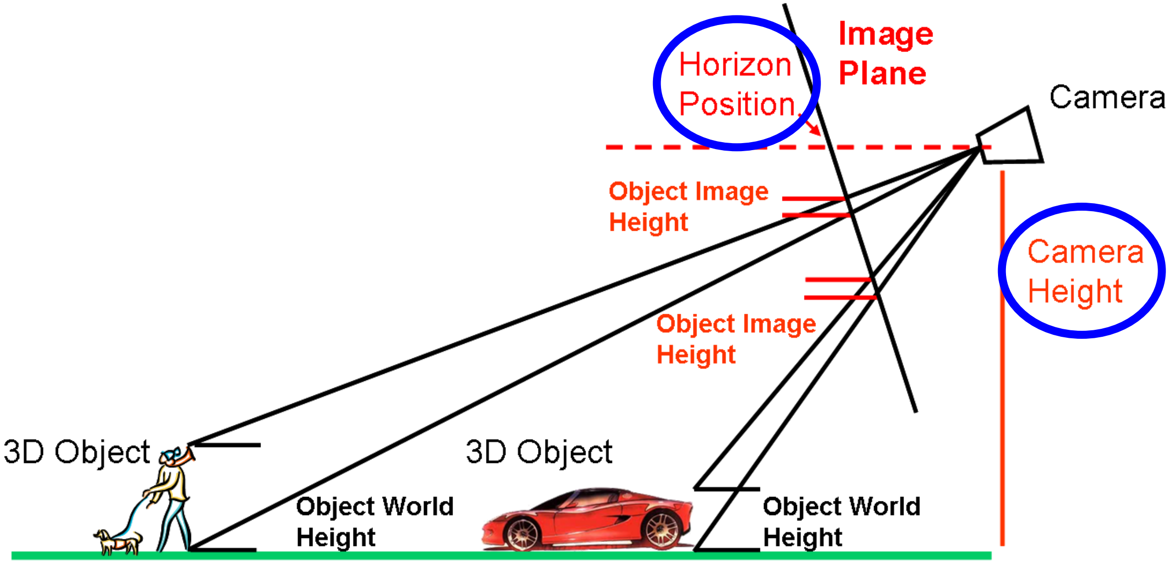
c) integrated model output

Torralba, Sinha (ICCV 2001), Torralba, Murphy, Freeman 2010

3d Scene Context



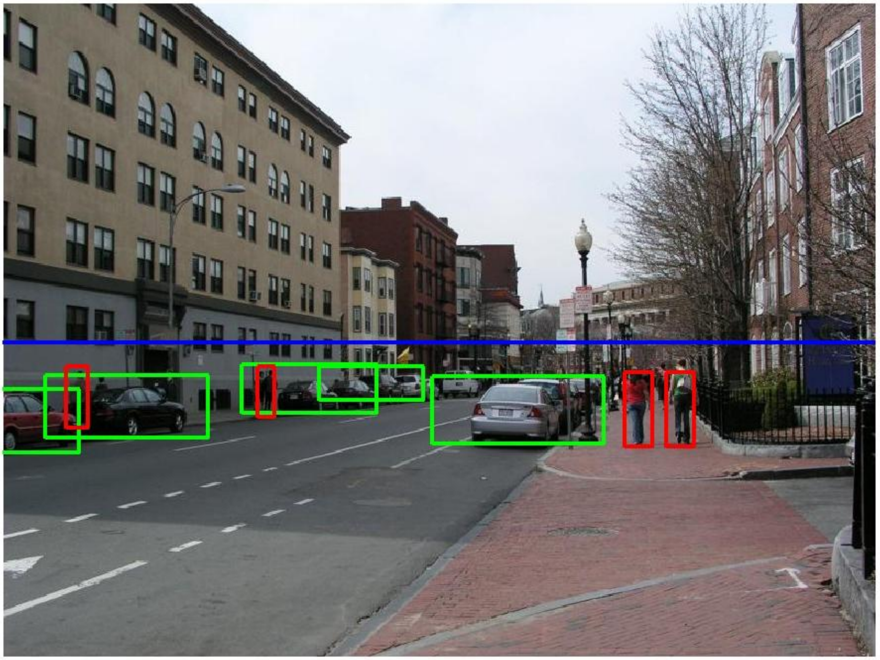
Image

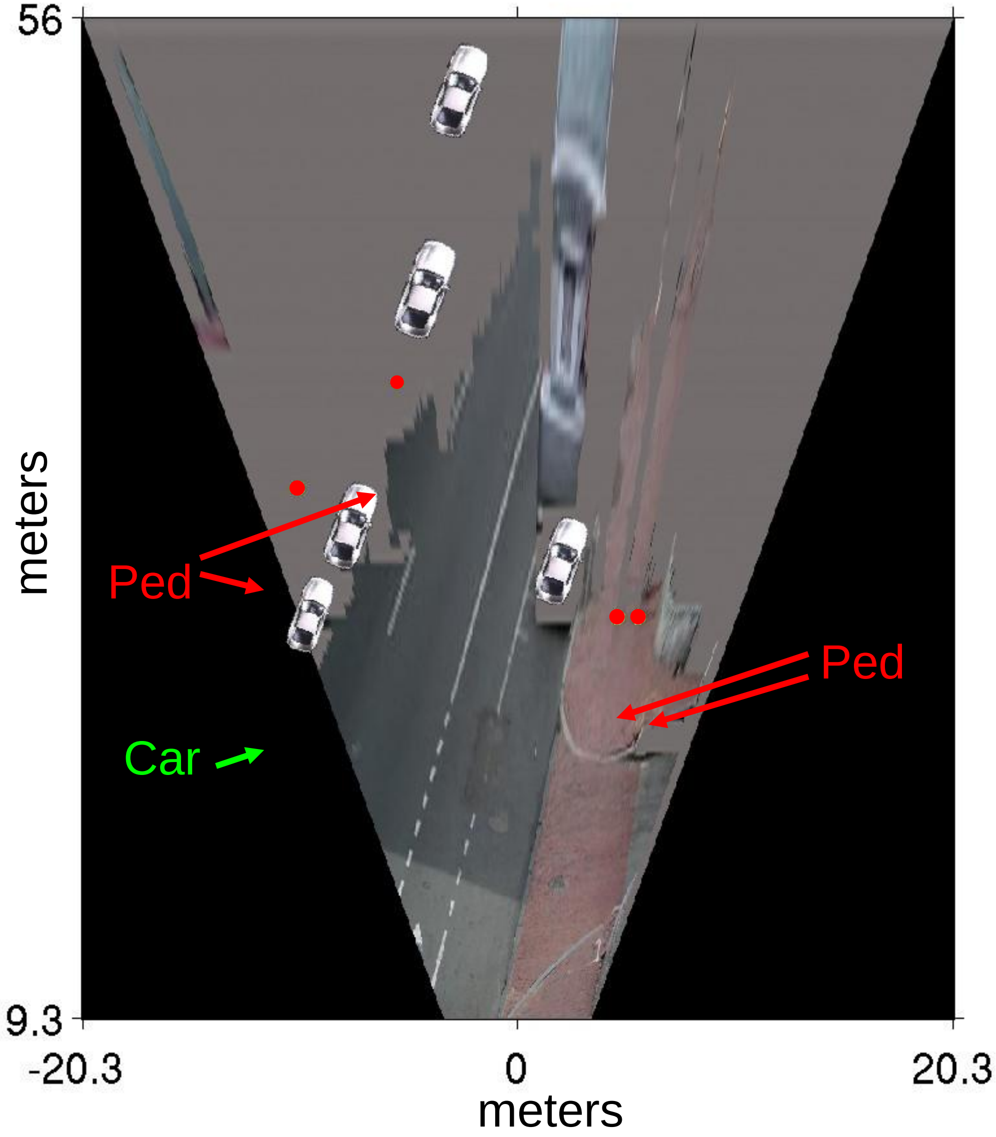


World

Hoiem, Efros, Hebert ICCV 2005

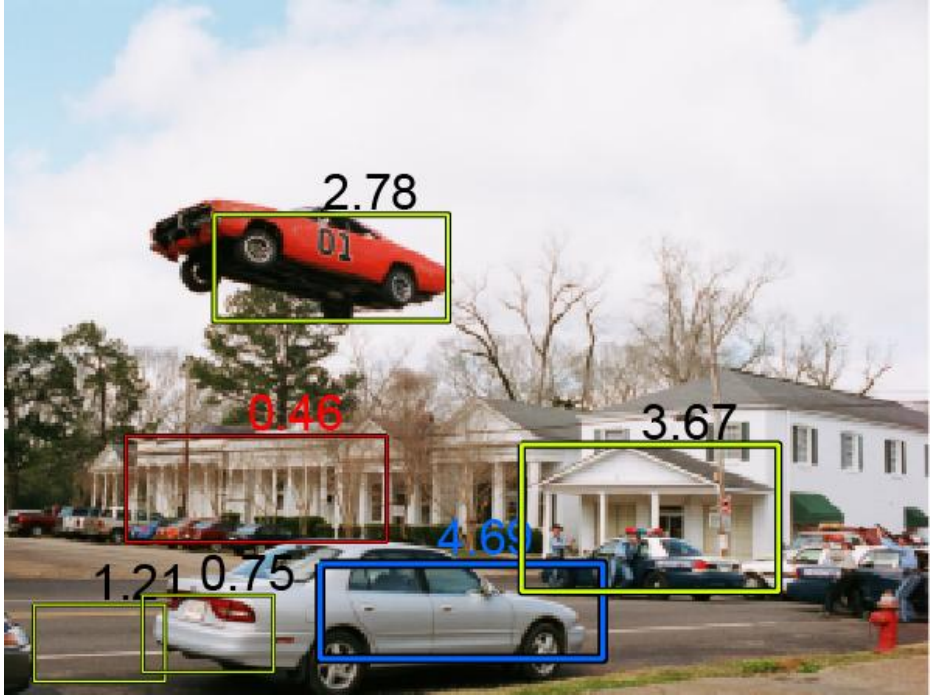
3d Scene Context





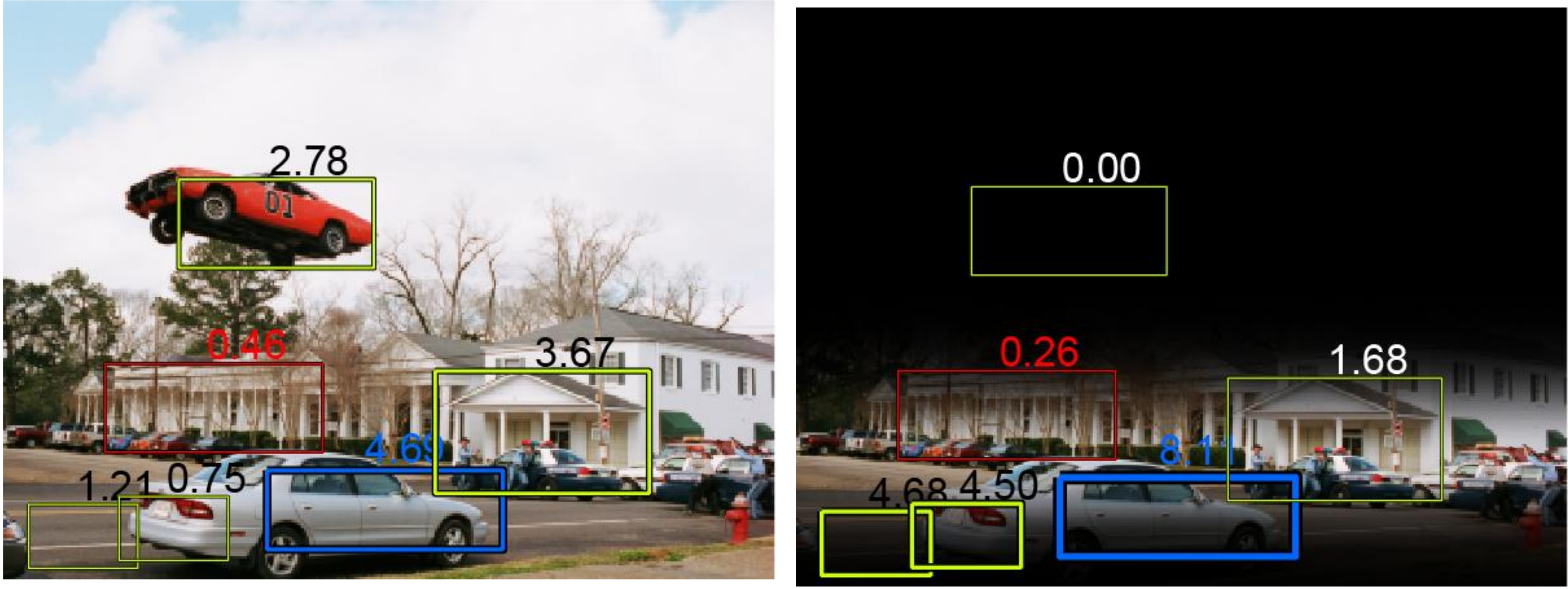
Hoiem, Efros, Hebert ICCV 2005

A car out of context …



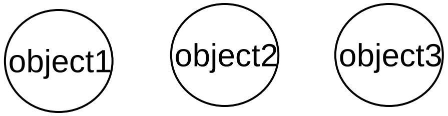
Torralba, Sinha (ICCV 2001), Torralba, Murphy, Freeman 2010

A car out of context …

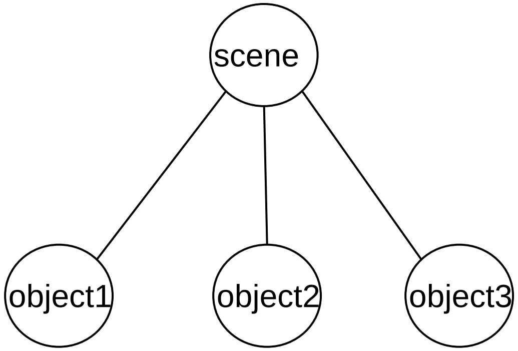


Torralba, Sinha (ICCV 2001), Torralba, Murphy, Freeman 2010

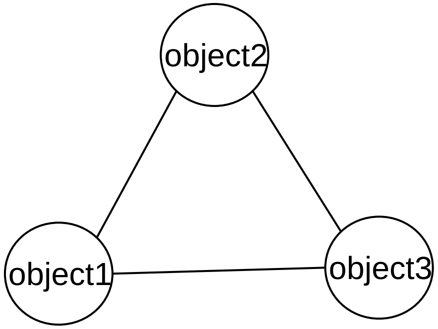
Context models



Independent model



Objects are correlated via the scene

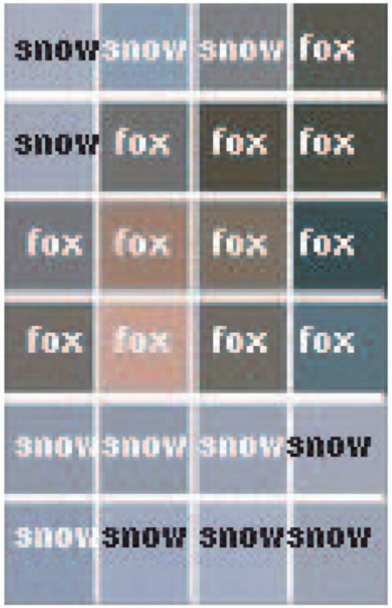


Dependencies among objects

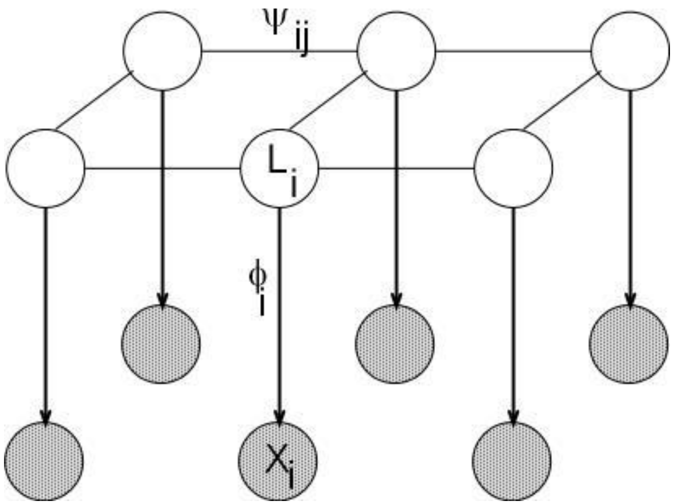
Pixel labeling using MRFs

Enforce consistency between neighboring labels, and between labels and pixels

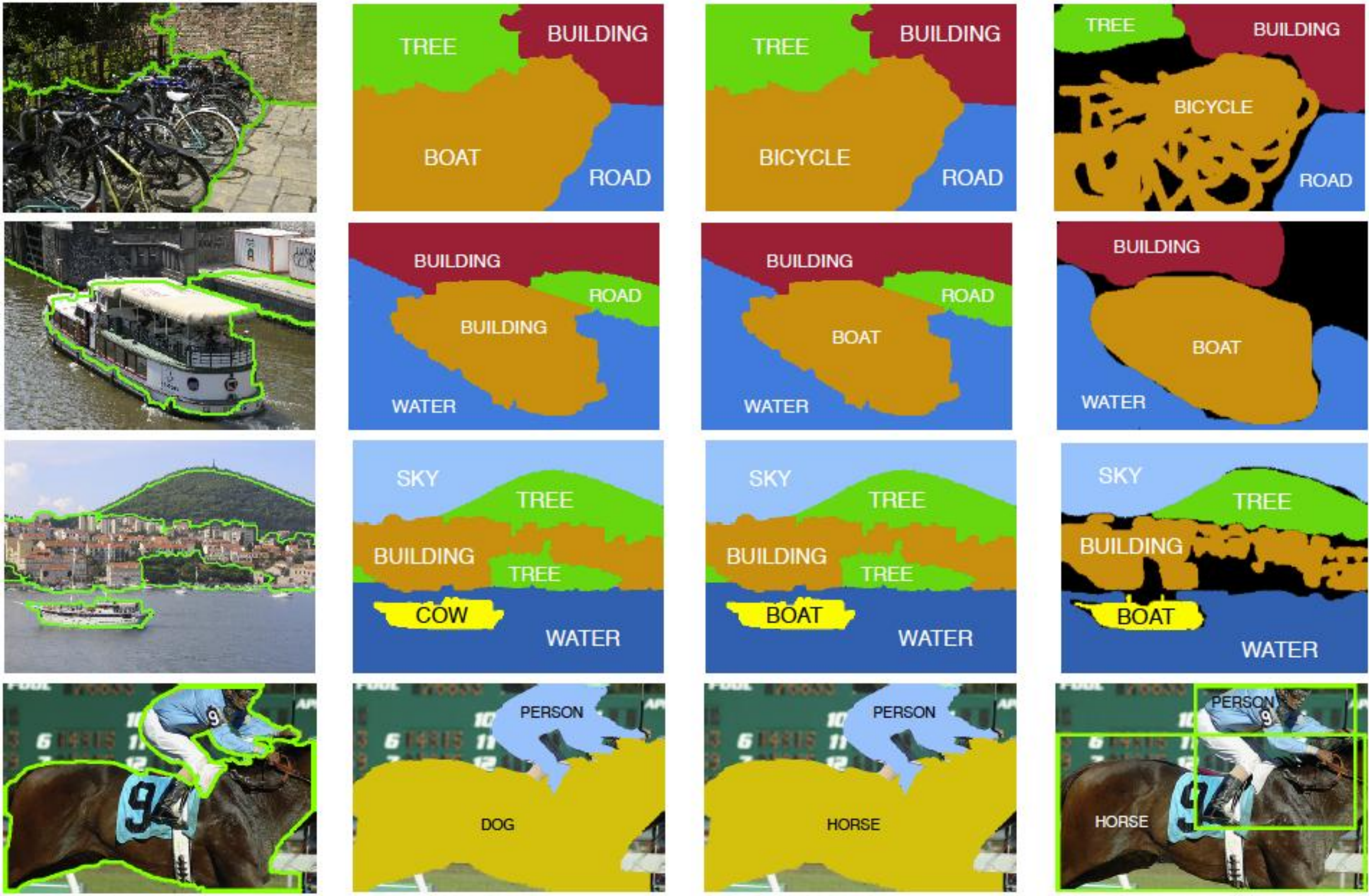
$$P(L,x) = P(L)P(x|L) = [\frac{1}{Z}\prod_i \prod_{j\in N_i} \psi_{ij}(L_i, L_j)][\prod_i P(x_i|L_i)]$$



snow	snow	snow	fox
snow	fox	fox	fox
fox	fox	fox	fox
fox	fox	fox	fox
snow	snow	snow	snow
snow	snow	snow	snow

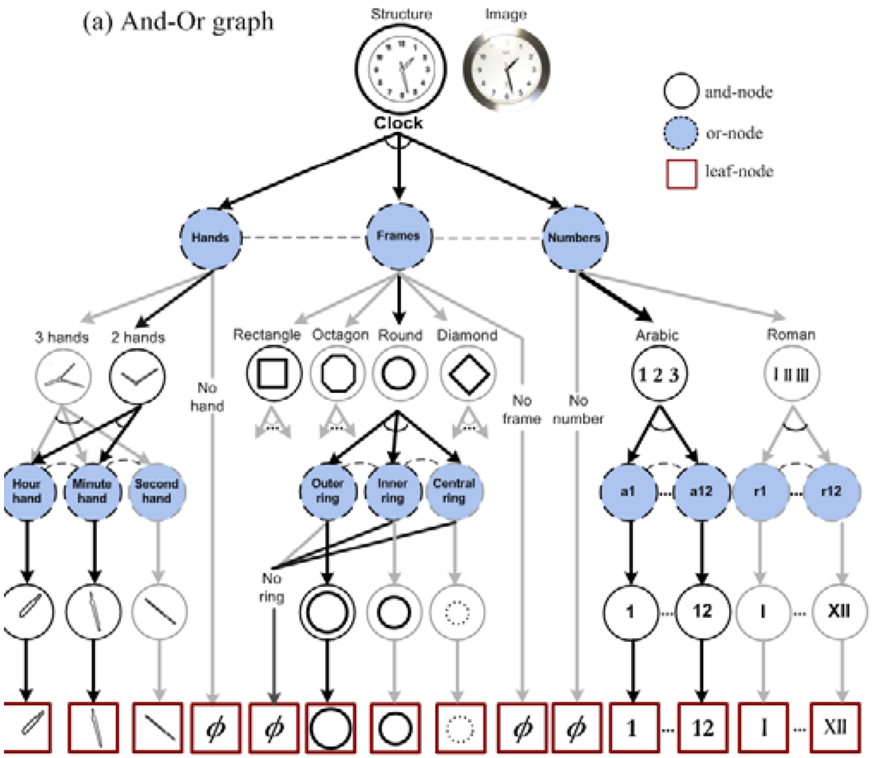


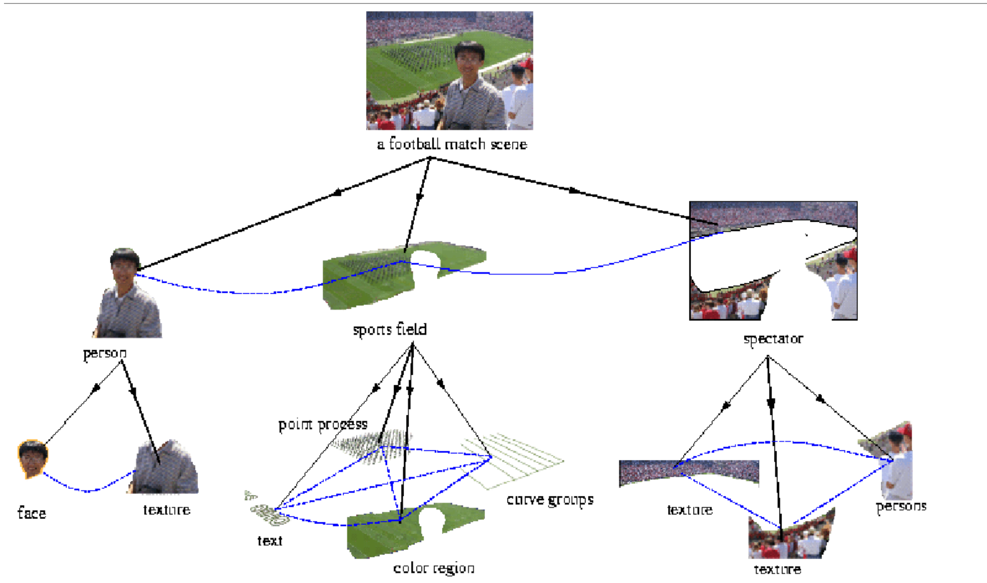
Carbonetto, de Freitas & Barnard, ECCV’04



A. Rabinovich, A. Vedaldi, C. Galleguillos, E. Wiewiora and S. Belongie. Objects in Context. ICCV 2007

Grammars for objects and scenes

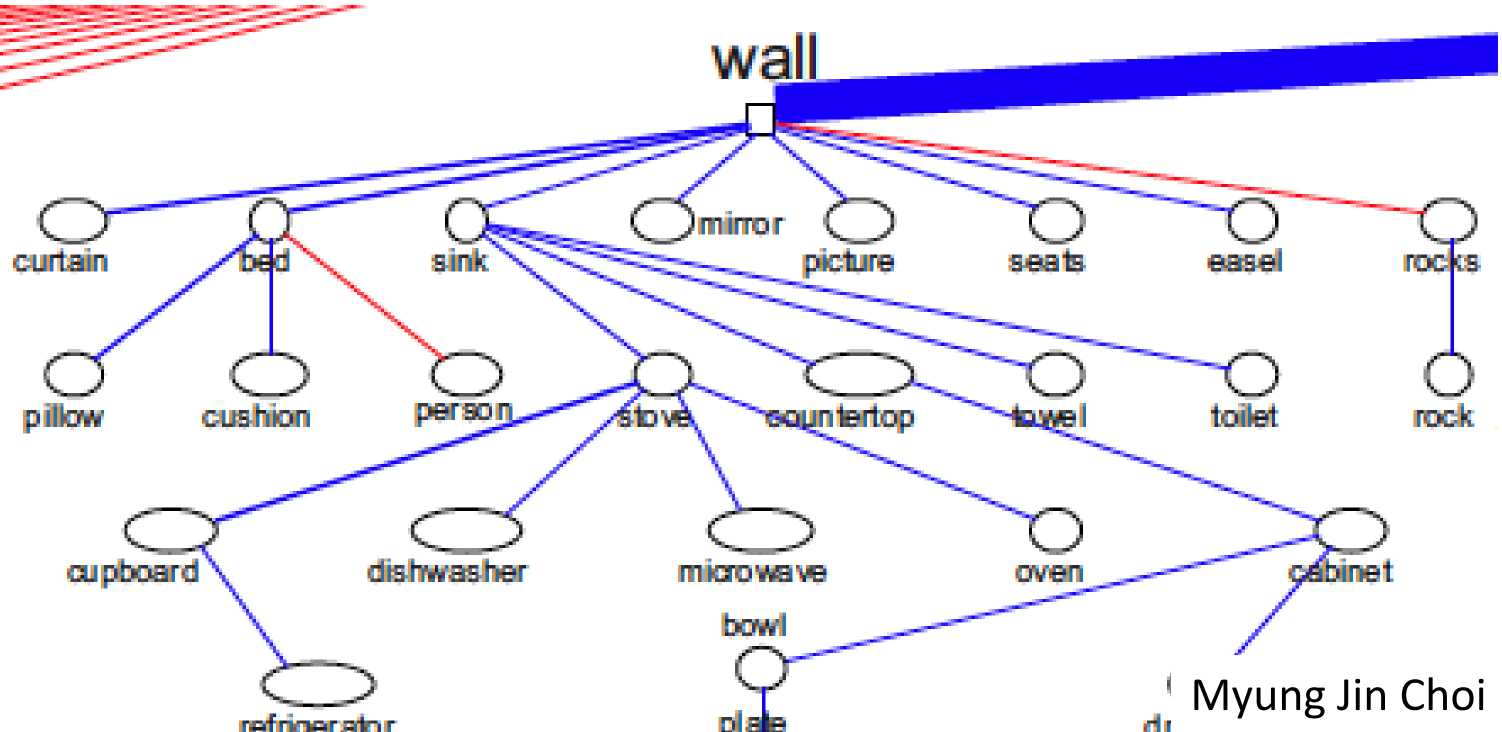




Example: parsing (Tu et al, 2000-2004)

S.C. Zhu and D. Mumford. A Stochastic Grammar of Images. Foundations and Trends in Computer Graphics and Vision, 2006.

Exploiting Hierarchical Context on a Large Database of Object Categories

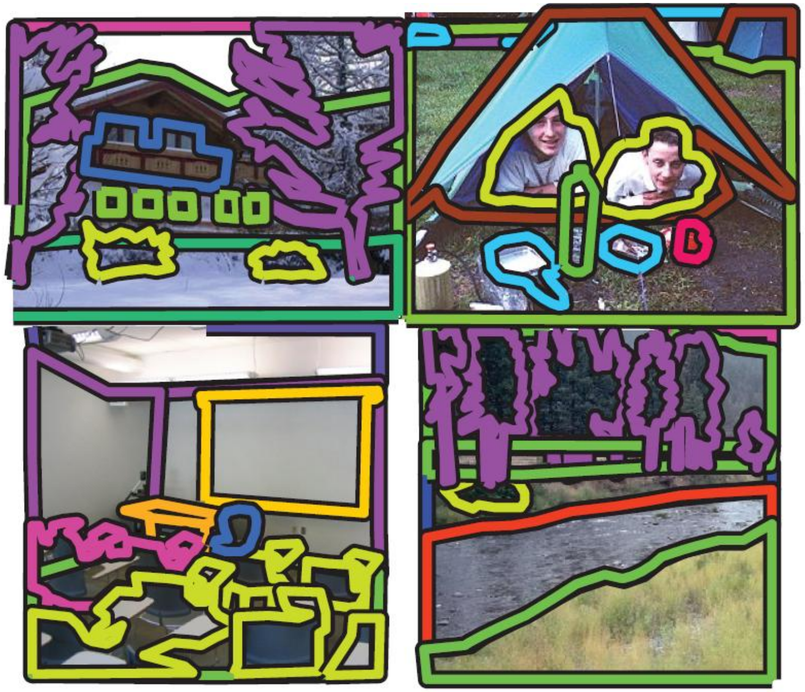


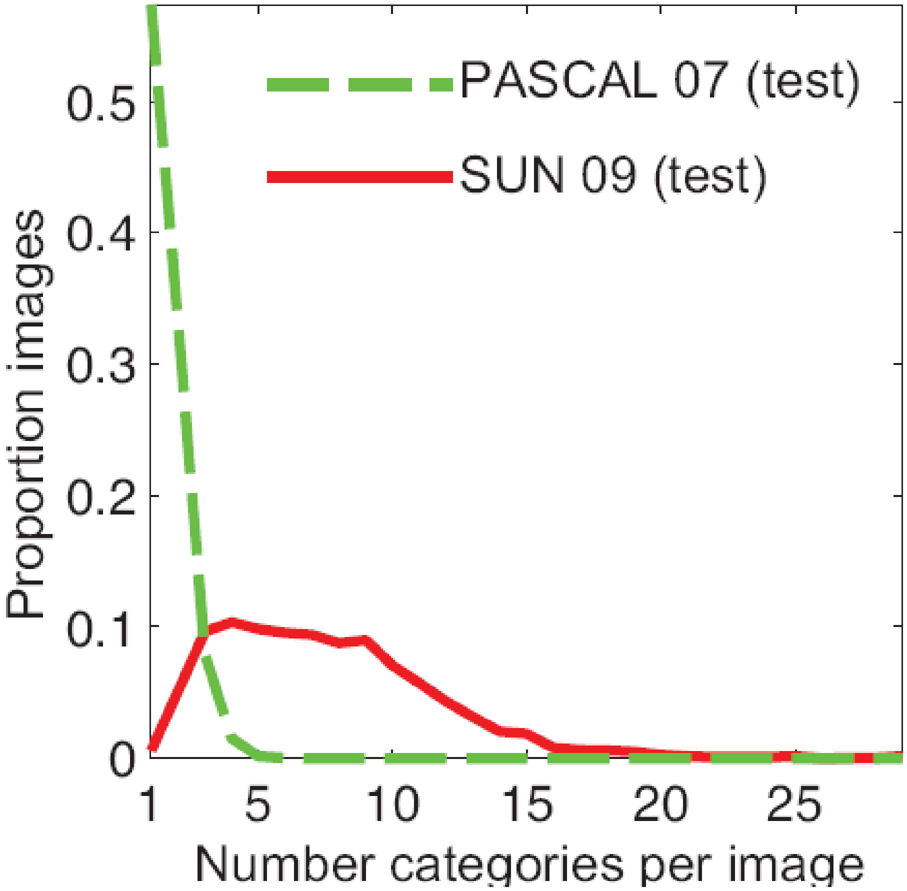
Myung Jin Choi

Joseph Lim

Myung Jin Choi, Joseph Lim, Antonio Torralba, and Alan S. Willsky. CVPR 2010

SUN database

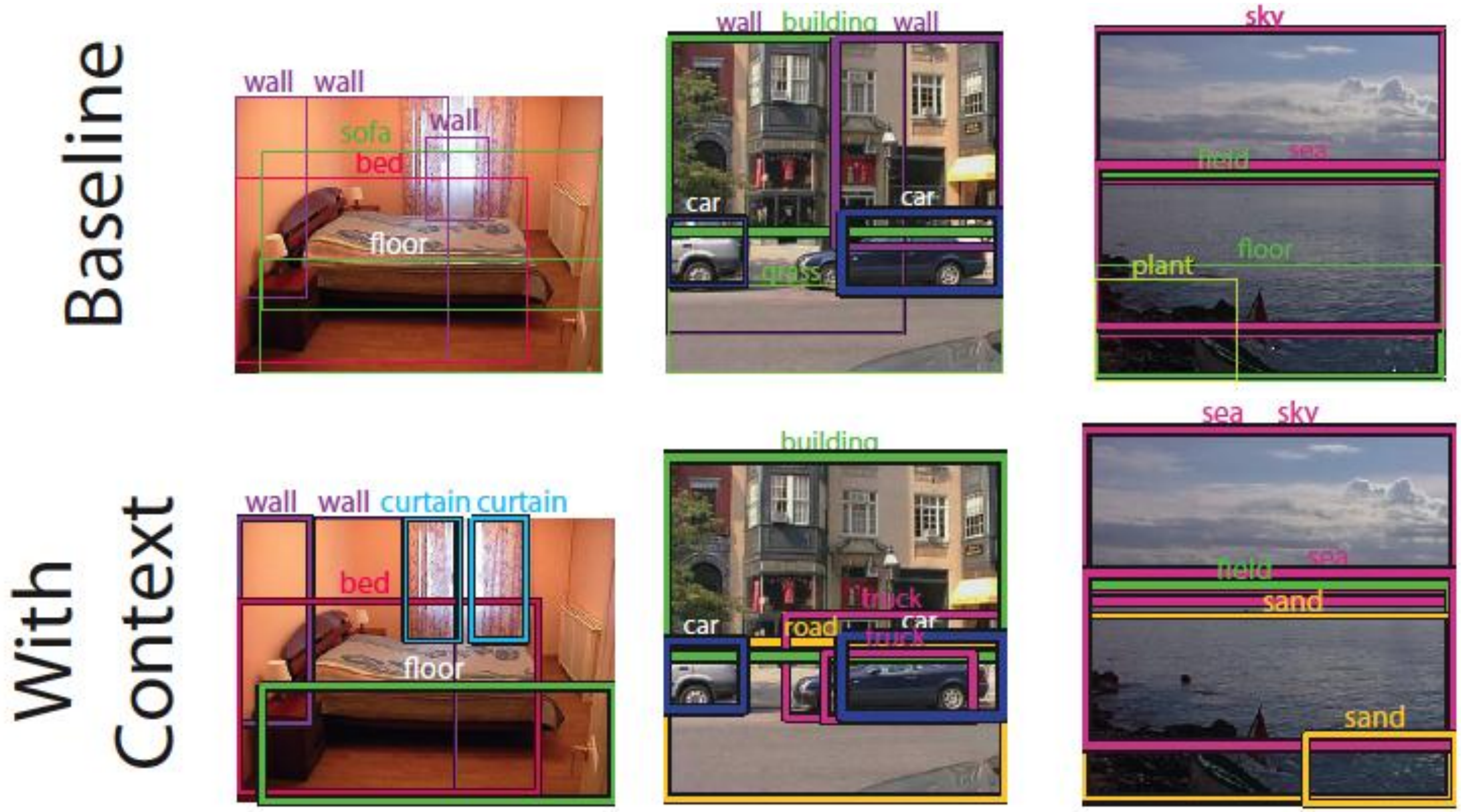


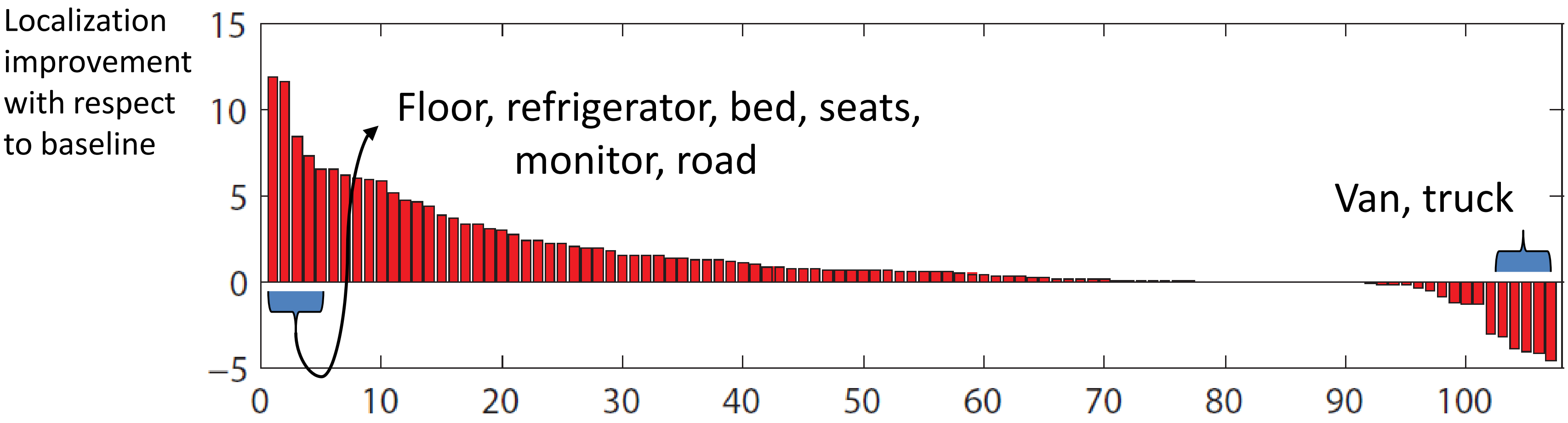


12,000 annotated images

107 object categories

152,000 annotated object instances



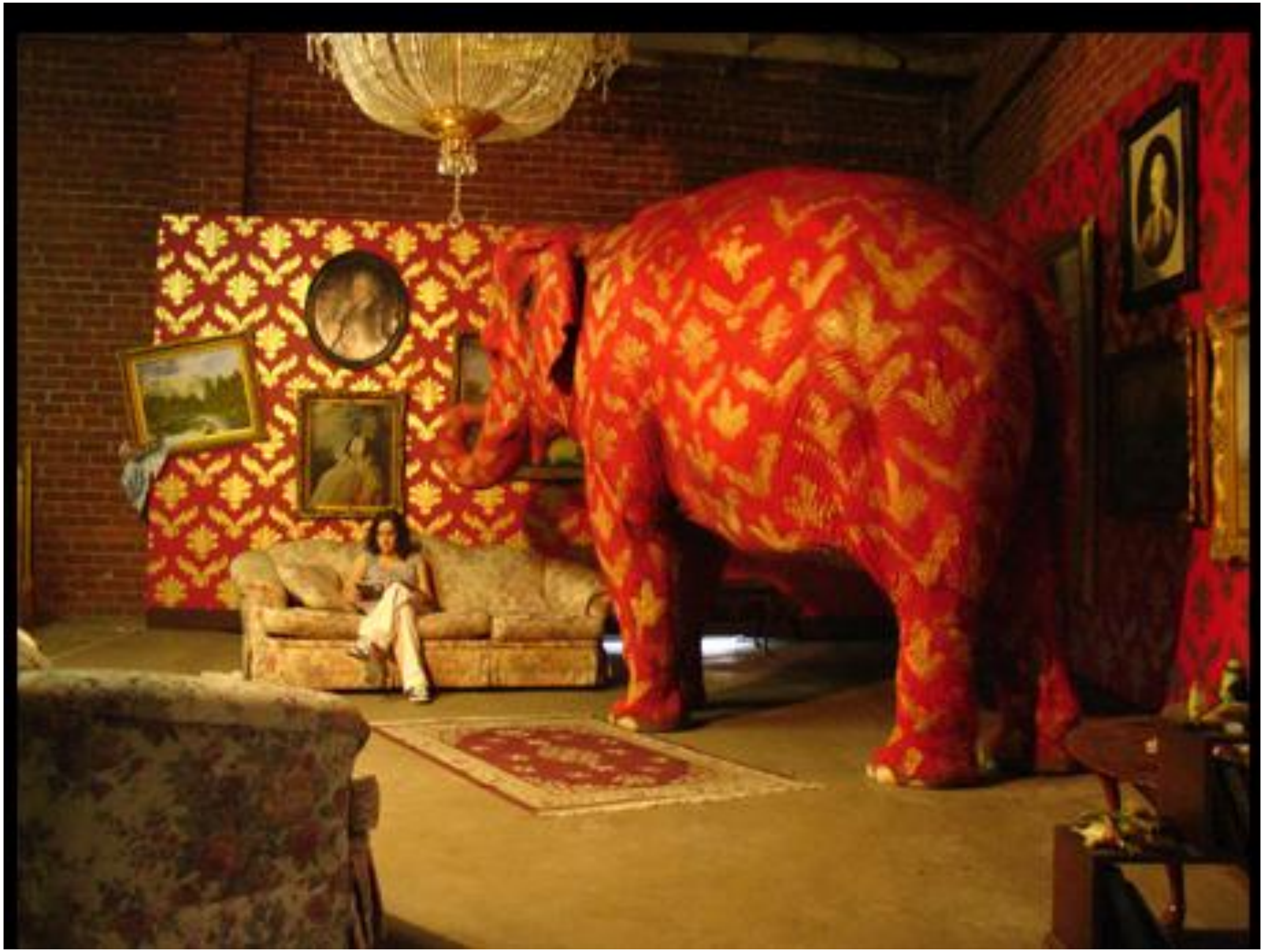


Floor, refrigerator, bed, seats, monitor, road

Van, truck

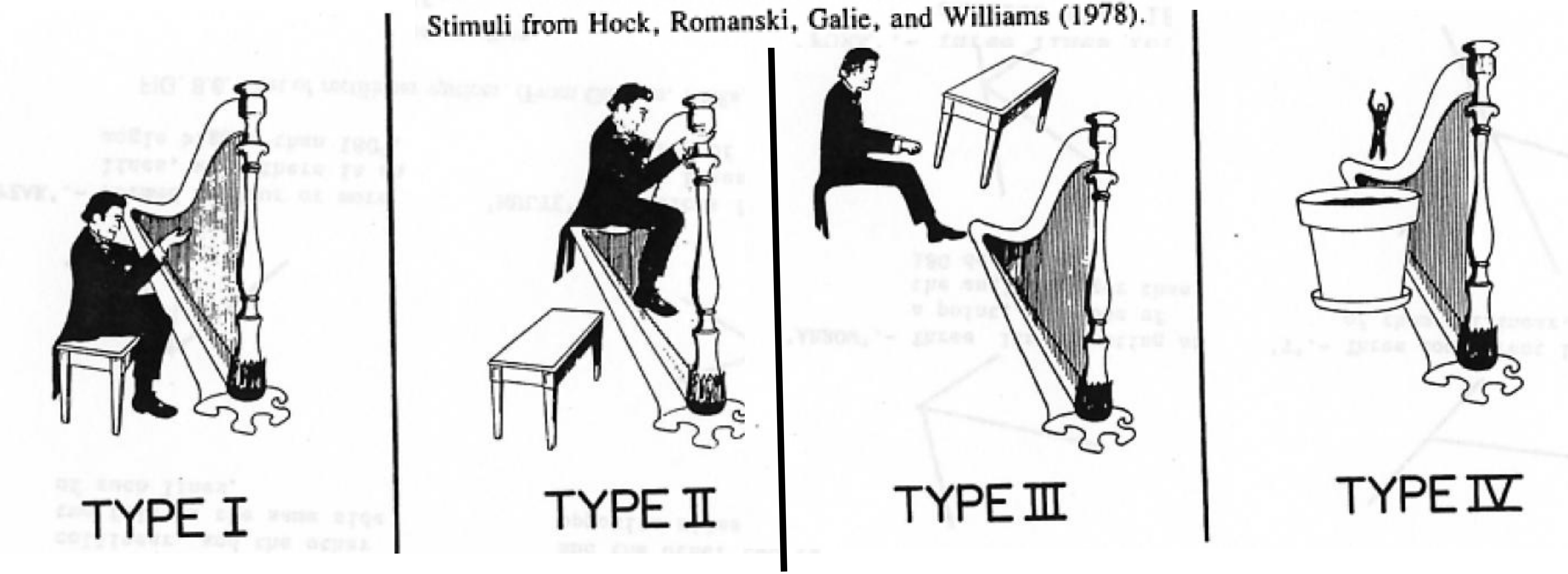
Who needs context anyway?

We can recognize objects even out of context

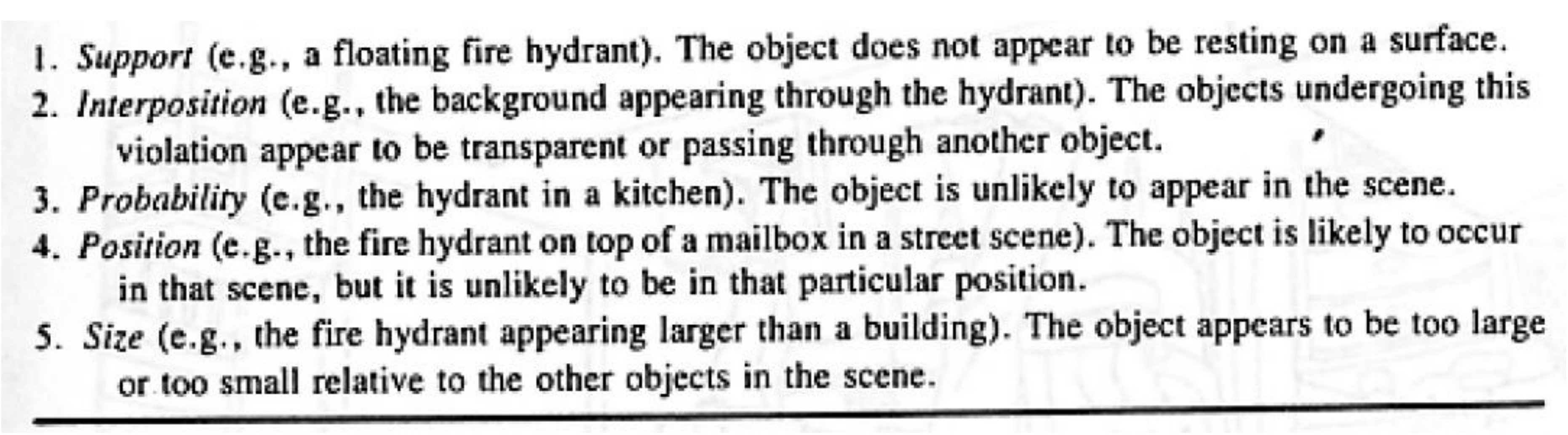


Banksy

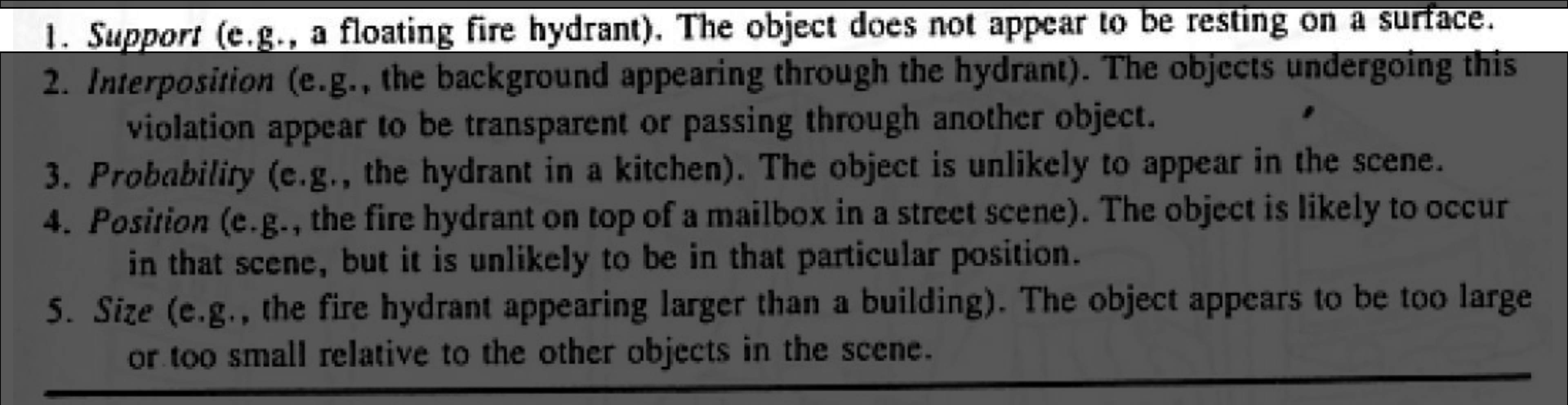
Biederman’s violations (1981)



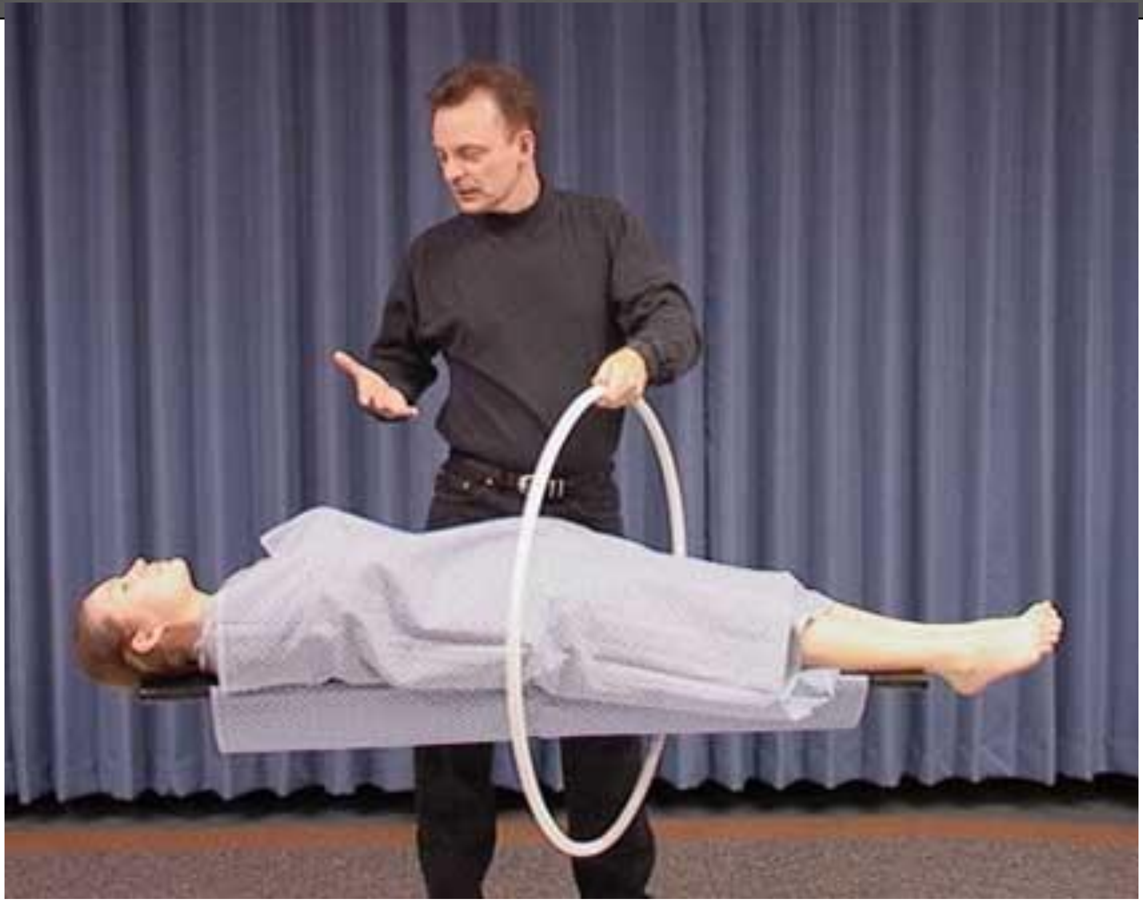
Stimuli from Hock, Romanski, Galie, and Williams (1978).

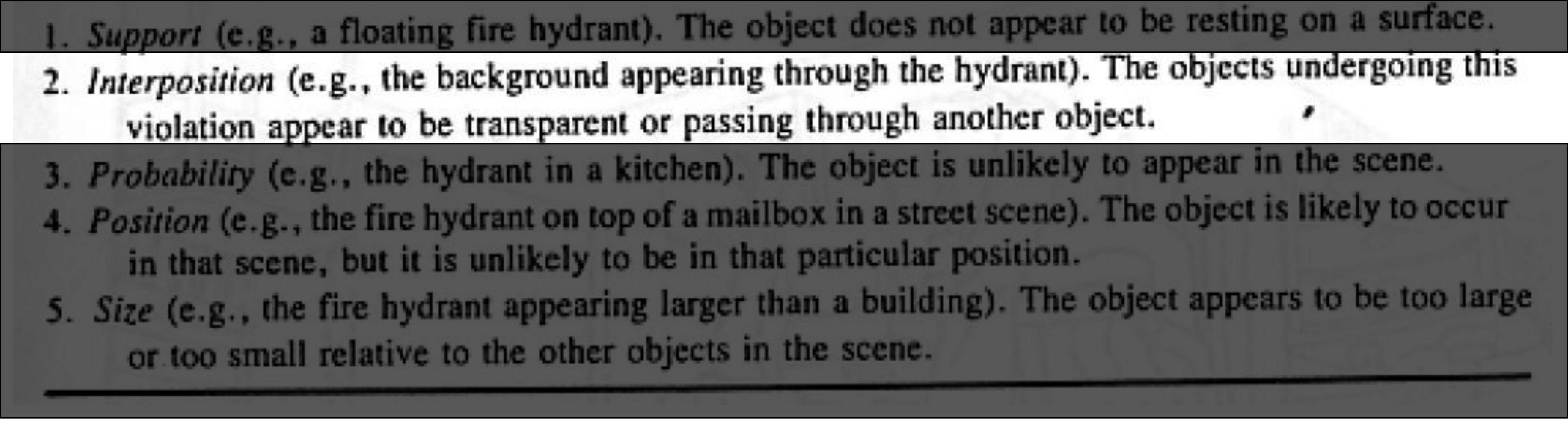


1. *Support* (e.g., a floating fire hydrant). The object does not appear to be resting on a surface.
2. *Interposition* (e.g., the background appearing through the hydrant). The objects undergoing this violation appear to be transparent or passing through another object.
3. *Probability* (e.g., the hydrant in a kitchen). The object is unlikely to appear in the scene.
4. *Position* (e.g., the fire hydrant on top of a mailbox in a street scene). The object is likely to occur in that scene, but it is unlikely to be in that particular position.
5. *Size* (e.g., the fire hydrant appearing larger than a building). The object appears to be too large or too small relative to the other objects in the scene.

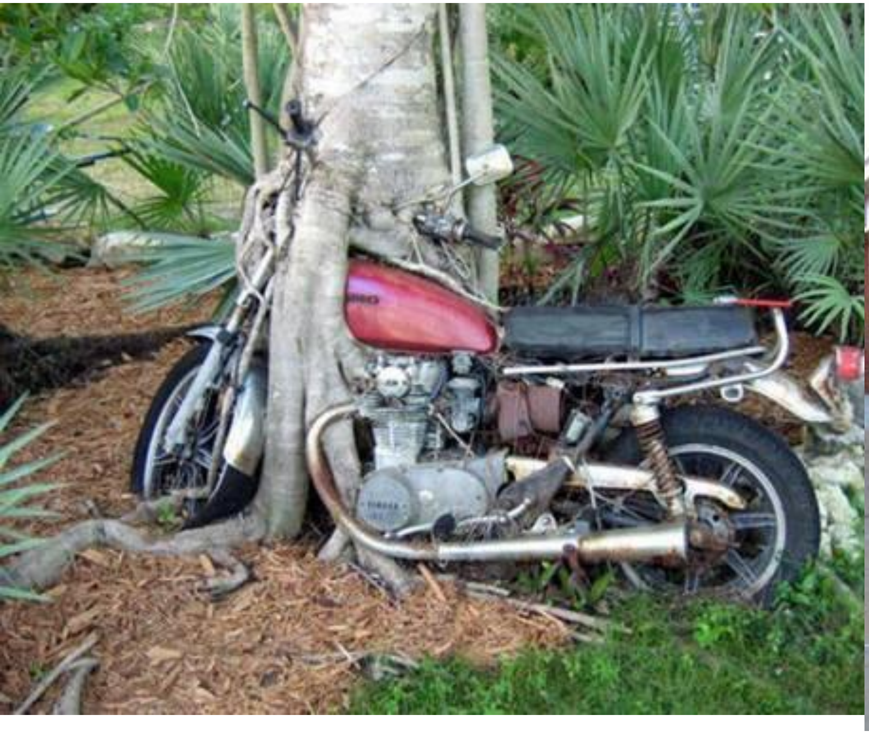


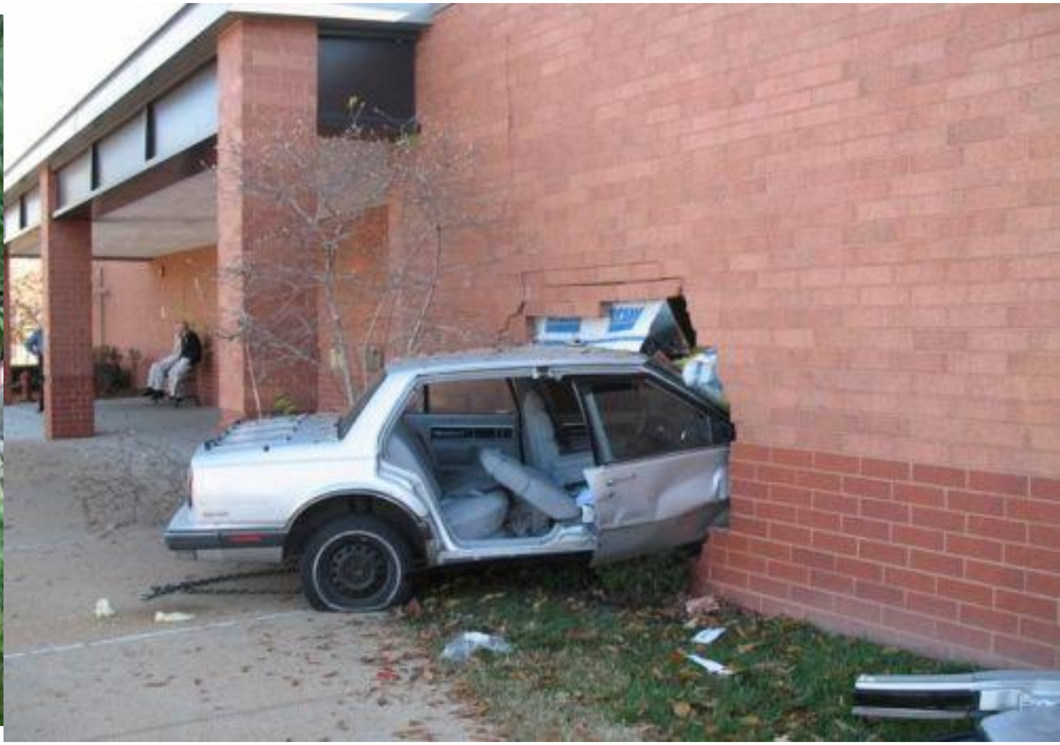
1. *Support* (e.g., a floating fire hydrant). The object does not appear to be resting on a surface.
2. *Interposition* (e.g., the background appearing through the hydrant). The objects undergoing this violation appear to be transparent or passing through another object.
3. *Probability* (e.g., the hydrant in a kitchen). The object is unlikely to appear in the scene.
4. *Position* (e.g., the fire hydrant on top of a mailbox in a street scene). The object is likely to occur in that scene, but it is unlikely to be in that particular position.
5. *Size* (e.g., the fire hydrant appearing larger than a building). The object appears to be too large or too small relative to the other objects in the scene.

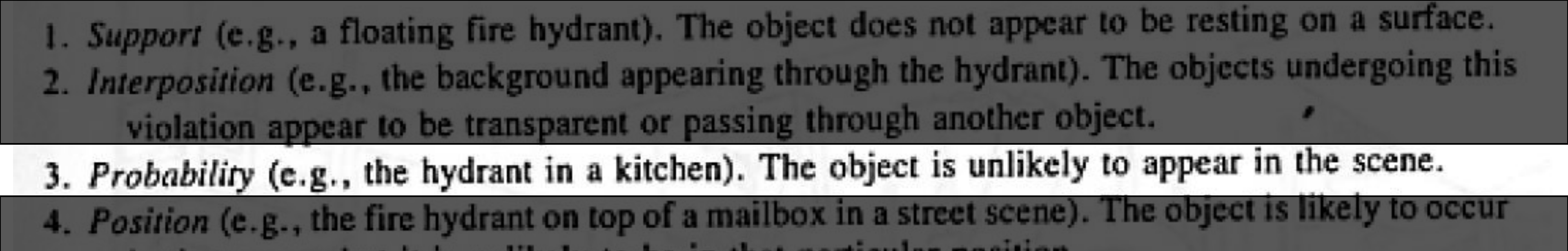




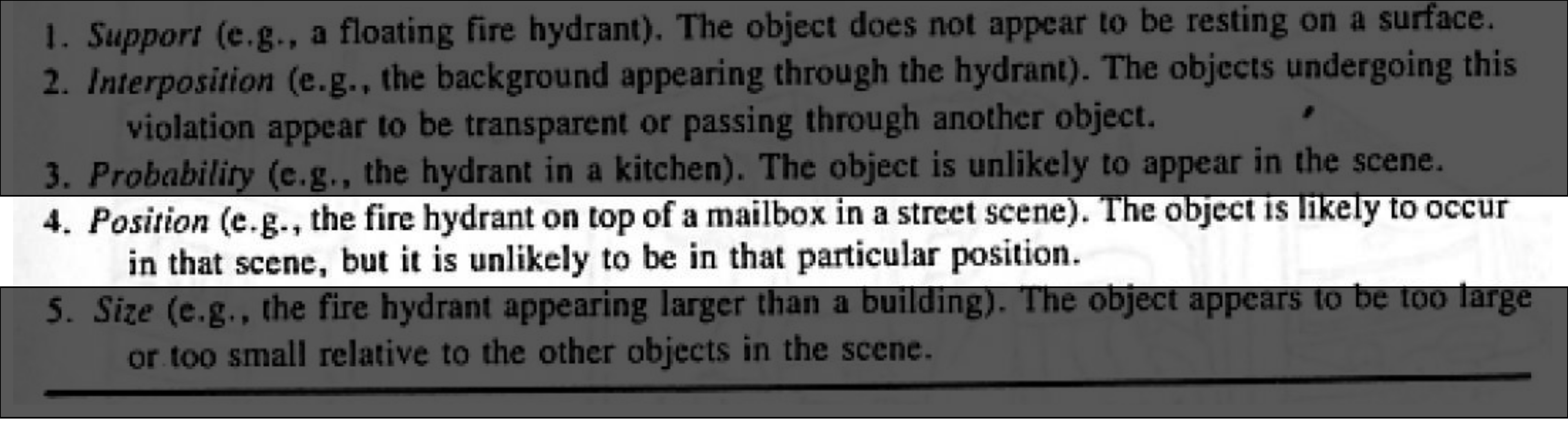
1. *Support* (e.g., a floating fire hydrant). The object does not appear to be resting on a surface.
2. *Interposition* (e.g., the background appearing through the hydrant). The objects undergoing this violation appear to be transparent or passing through another object.
3. *Probability* (e.g., the hydrant in a kitchen). The object is unlikely to appear in the scene.
4. *Position* (e.g., the fire hydrant on top of a mailbox in a street scene). The object is likely to occur in that scene, but it is unlikely to be in that particular position.
5. *Size* (e.g., the fire hydrant appearing larger than a building). The object appears to be too large or too small relative to the other objects in the scene.



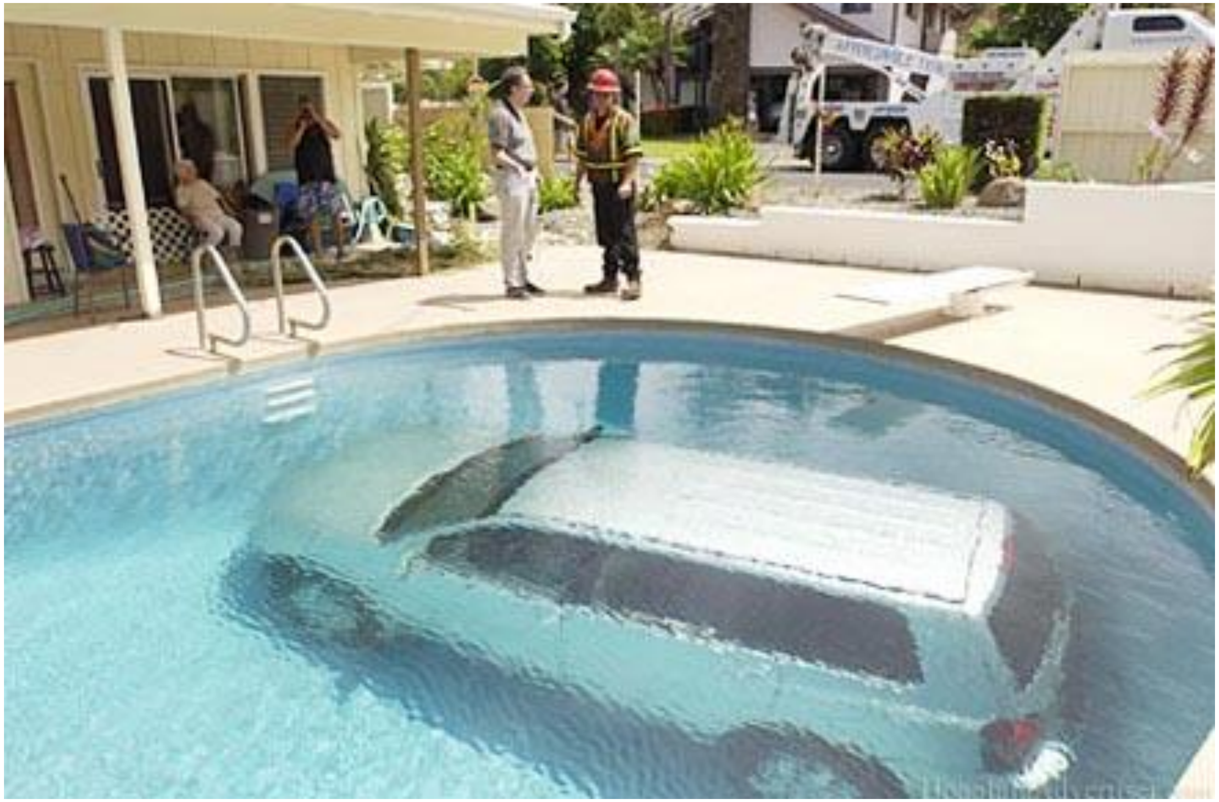


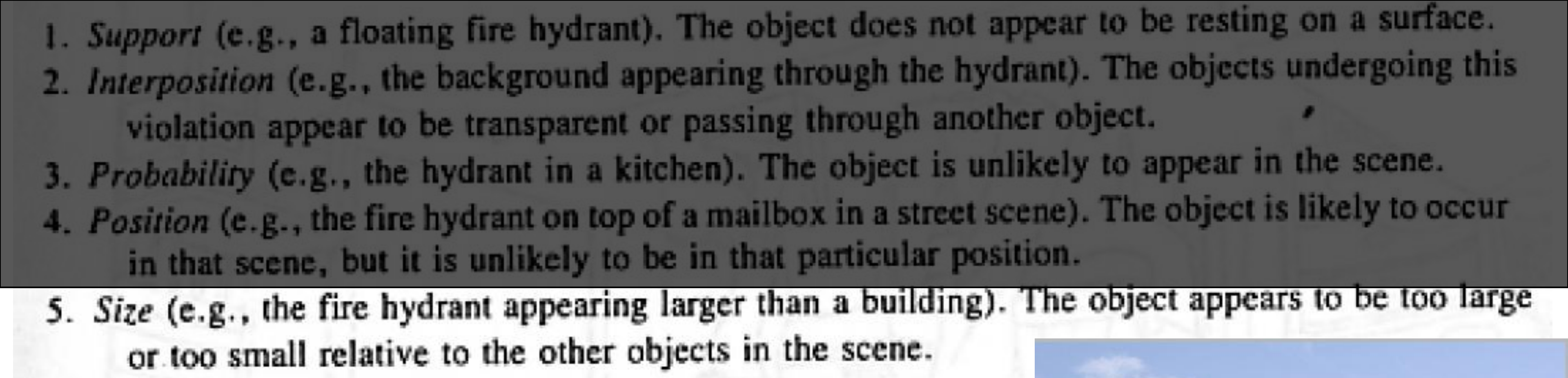


1. *Support* (e.g., a floating fire hydrant). The object does not appear to be resting on a surface.
2. *Interposition* (e.g., the background appearing through the hydrant). The objects undergoing this violation appear to be transparent or passing through another object.
3. *Probability* (e.g., the hydrant in a kitchen). The object is unlikely to appear in the scene.
4. *Position* (e.g., the fire hydrant on top of a mailbox in a street scene). The object is likely to occur in that scene, but it is unlikely to be in that p
5. *Size* (e.g., the fire hydrant appearing larger thar or too small relative to the other objects in th

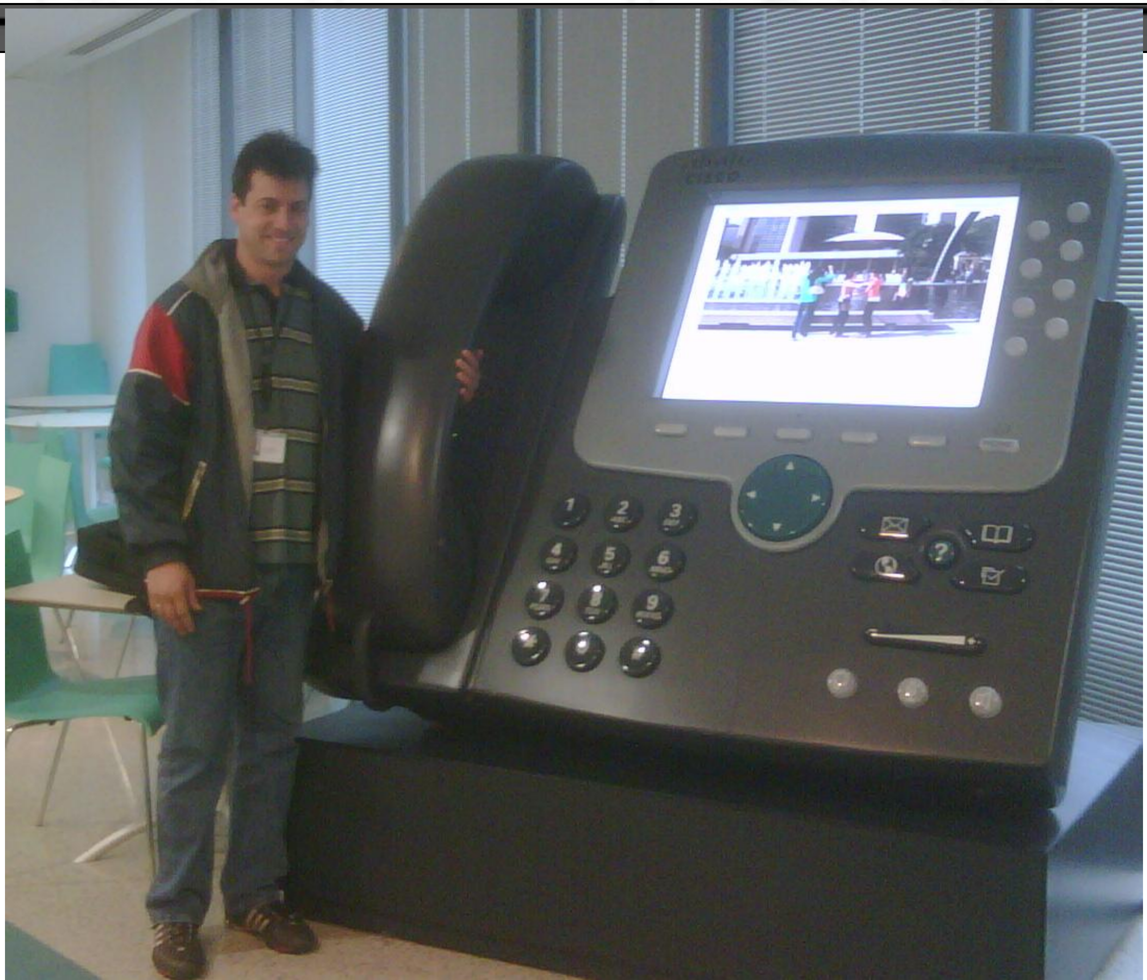


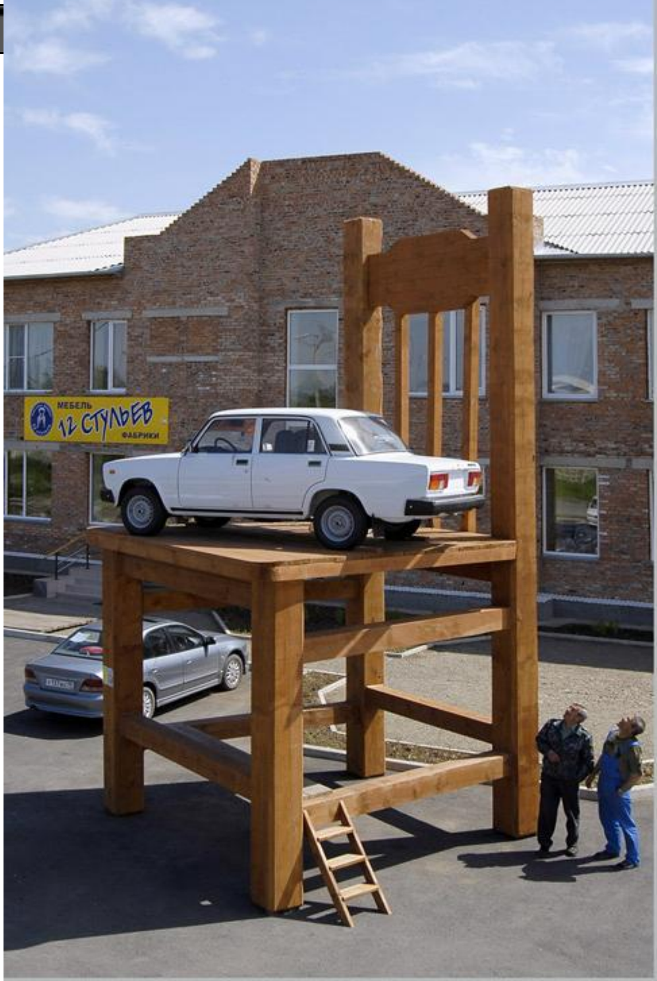
1. *Support* (e.g., a floating fire hydrant). The object does not appear to be resting on a surface.
2. *Interposition* (e.g., the background appearing through the hydrant). The objects undergoing this violation appear to be transparent or passing through another object.
3. *Probability* (e.g., the hydrant in a kitchen). The object is unlikely to appear in the scene.
4. *Position* (e.g., the fire hydrant on top of a mailbox in a street scene). The object is likely to occur in that scene, but it is unlikely to be in that particular position.
5. *Size* (e.g., the fire hydrant appearing larger than a building). The object appears to be too large or too small relative to the other objects in the scene.





1. *Support* (e.g., a floating fire hydrant). The object does not appear to be resting on a surface.
2. *Interposition* (e.g., the background appearing through the hydrant). The objects undergoing this violation appear to be transparent or passing through another object.
3. *Probability* (e.g., the hydrant in a kitchen). The object is unlikely to appear in the scene.
4. *Position* (e.g., the fire hydrant on top of a mailbox in a street scene). The object is likely to occur in that scene, but it is unlikely to be in that particular position.
5. *Size* (e.g., the fire hydrant appearing larger than a building). The object appears to be too large or too small relative to the other objects in the scene.





Unusual quantities
Unusual pose

…

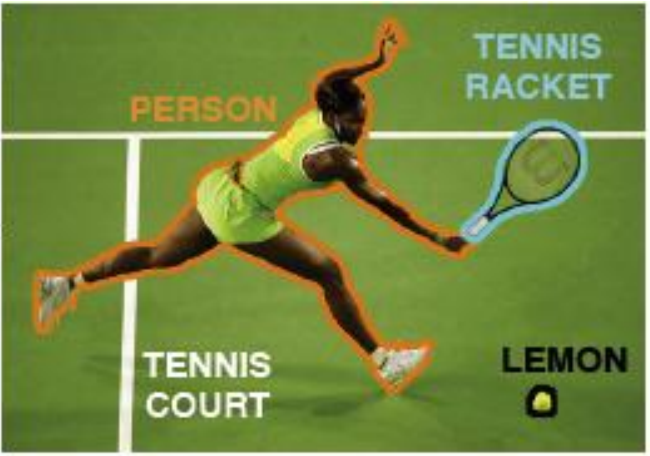
Out-of-context database

Among 161 images, there are:

- 13 co-occurrence violation
- 77 support violation
- 65 with both co-occurrence and support violations
- 6 with other violations (e.g. scale)

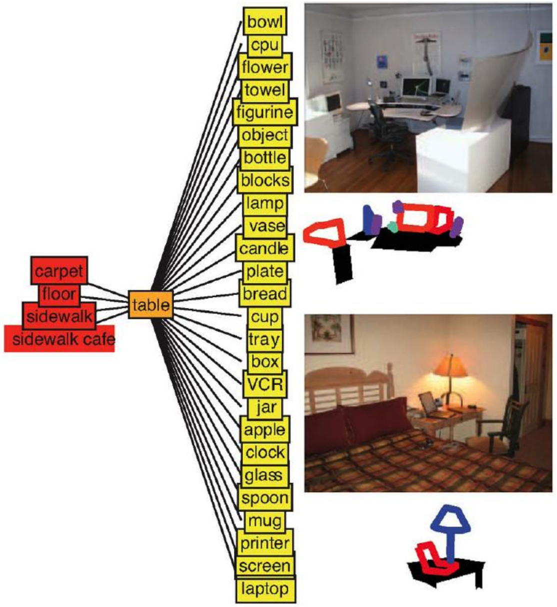
Context models and out-of-context objects

Co-occurrences



Rabinovich et al (2007)
Felzenszwalb, et al (2009)

Qualitative spatial relations

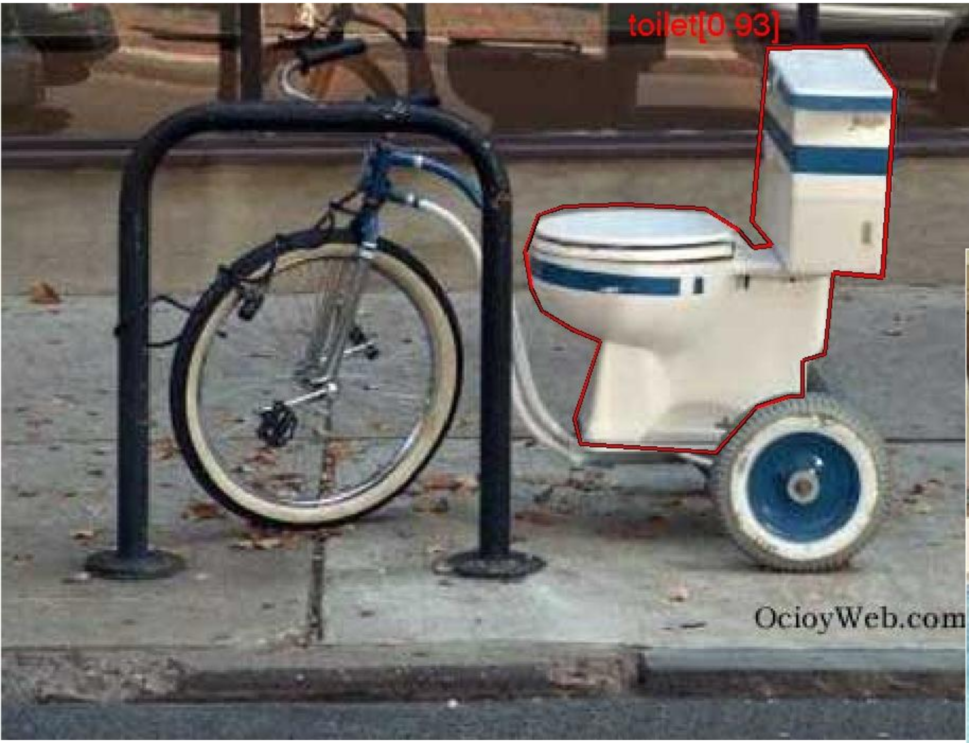


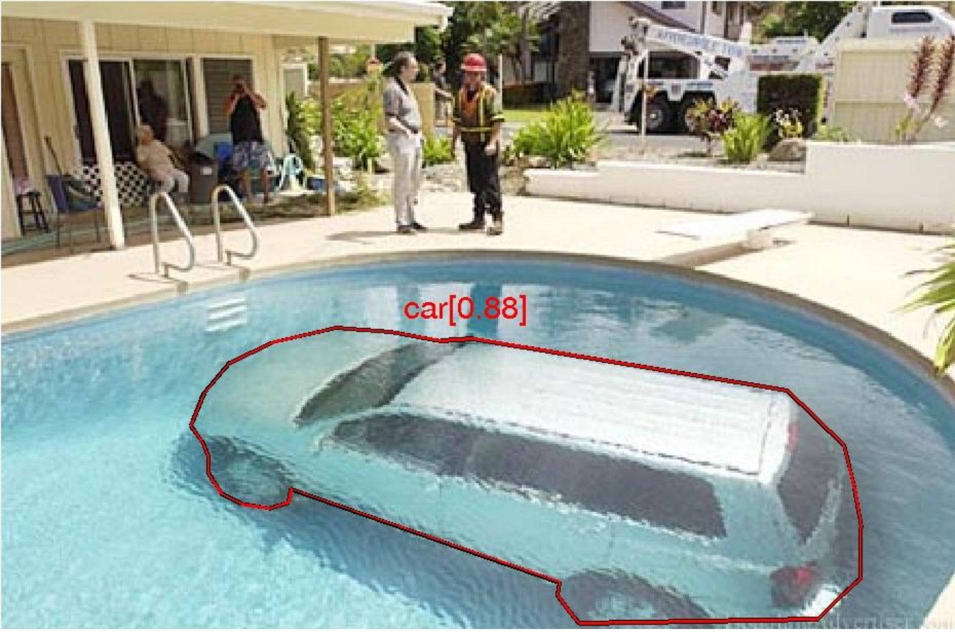
Galleguillos et al (2008)
Desai et al (2009)
Russell, Torralba (2010)
Abinav et al (2010)

2D/3D geometry

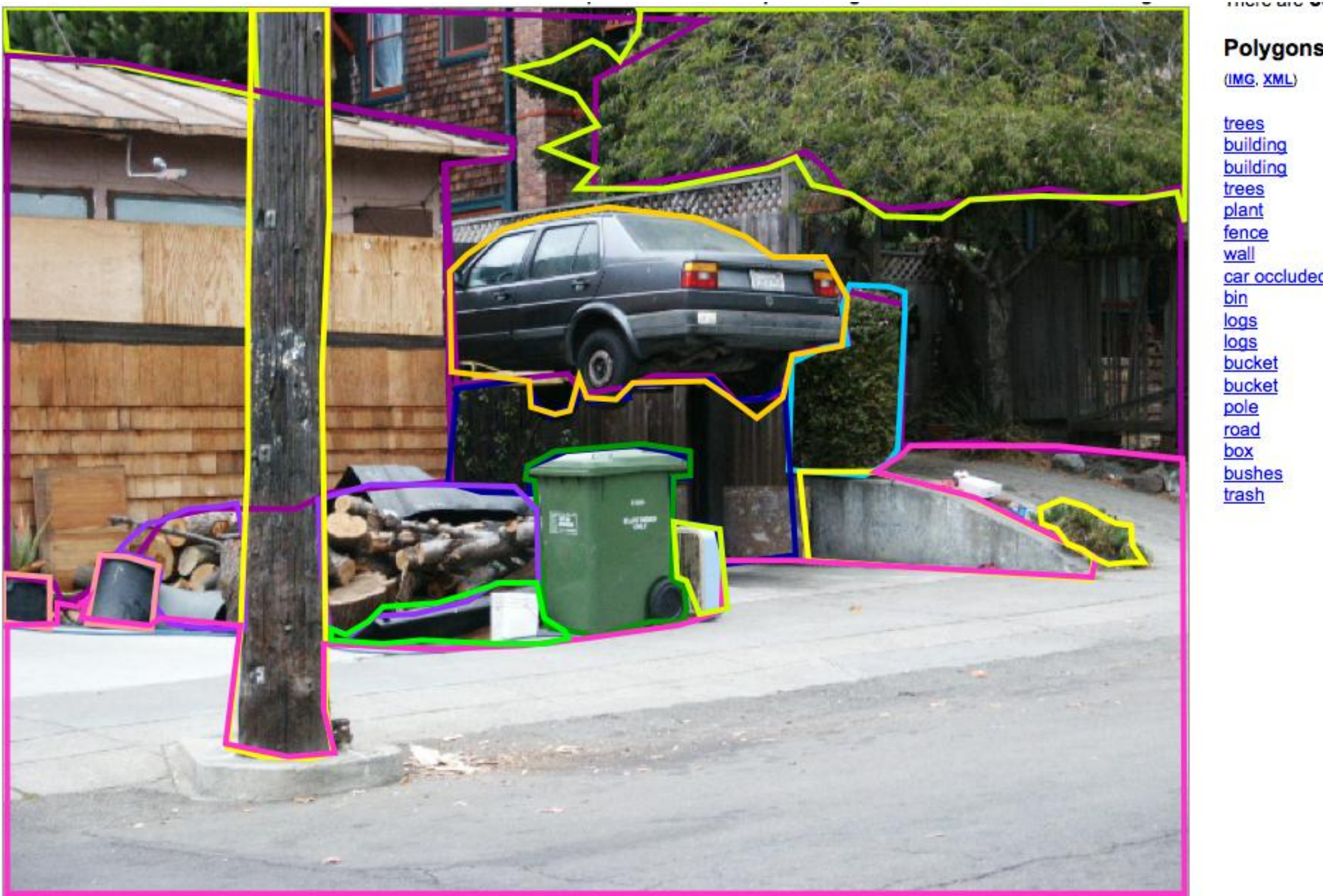
Torralba, Sinha (2001)
Fink, Perona (2003)
Murphy et al (2003)
Torralba et al (2004)
Hoiem, et al (2005)
Kumar, Hebert (2005)
Gould et al (2007)
Heitz and Koller (2008)

Some images are easy

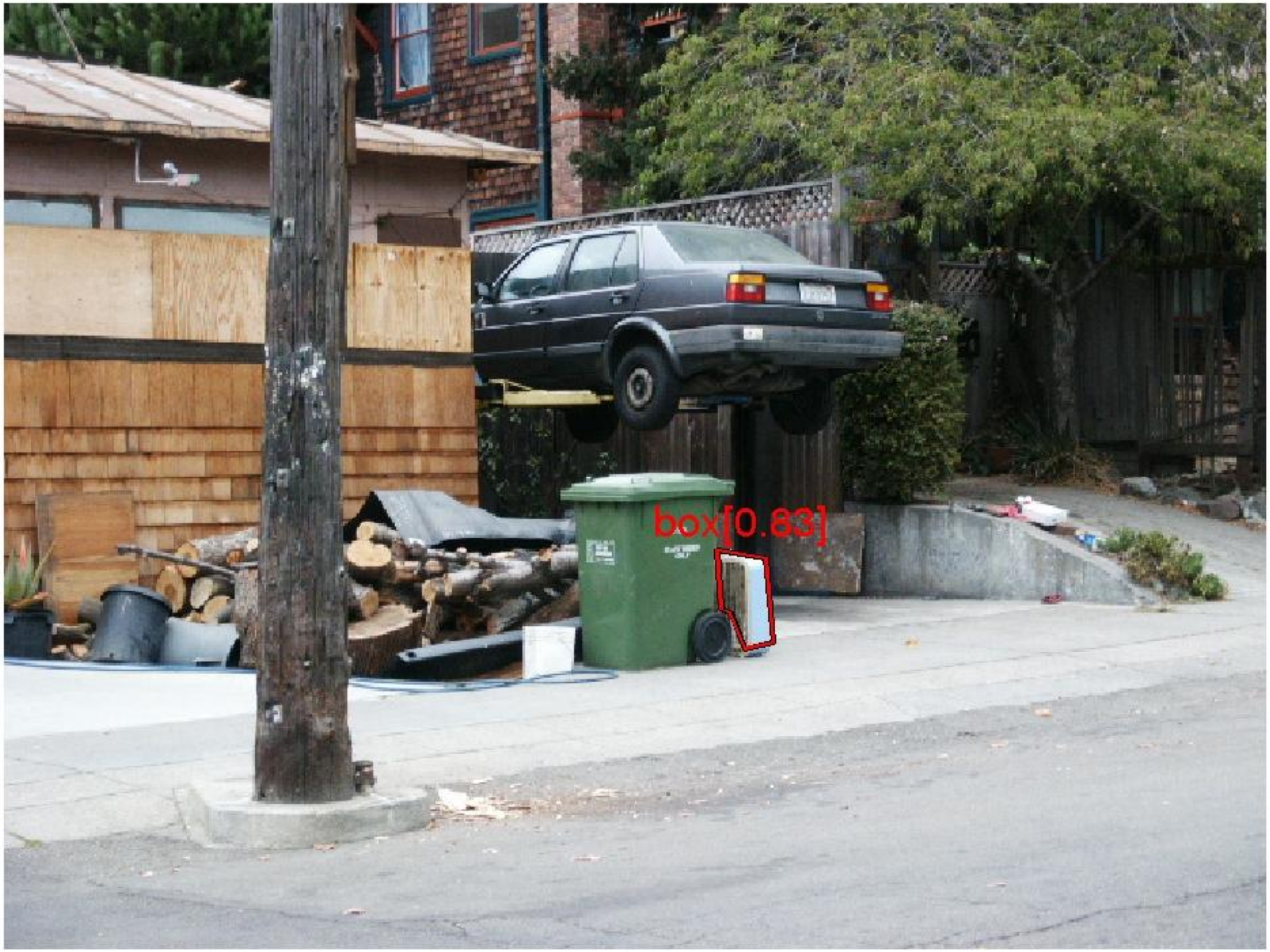




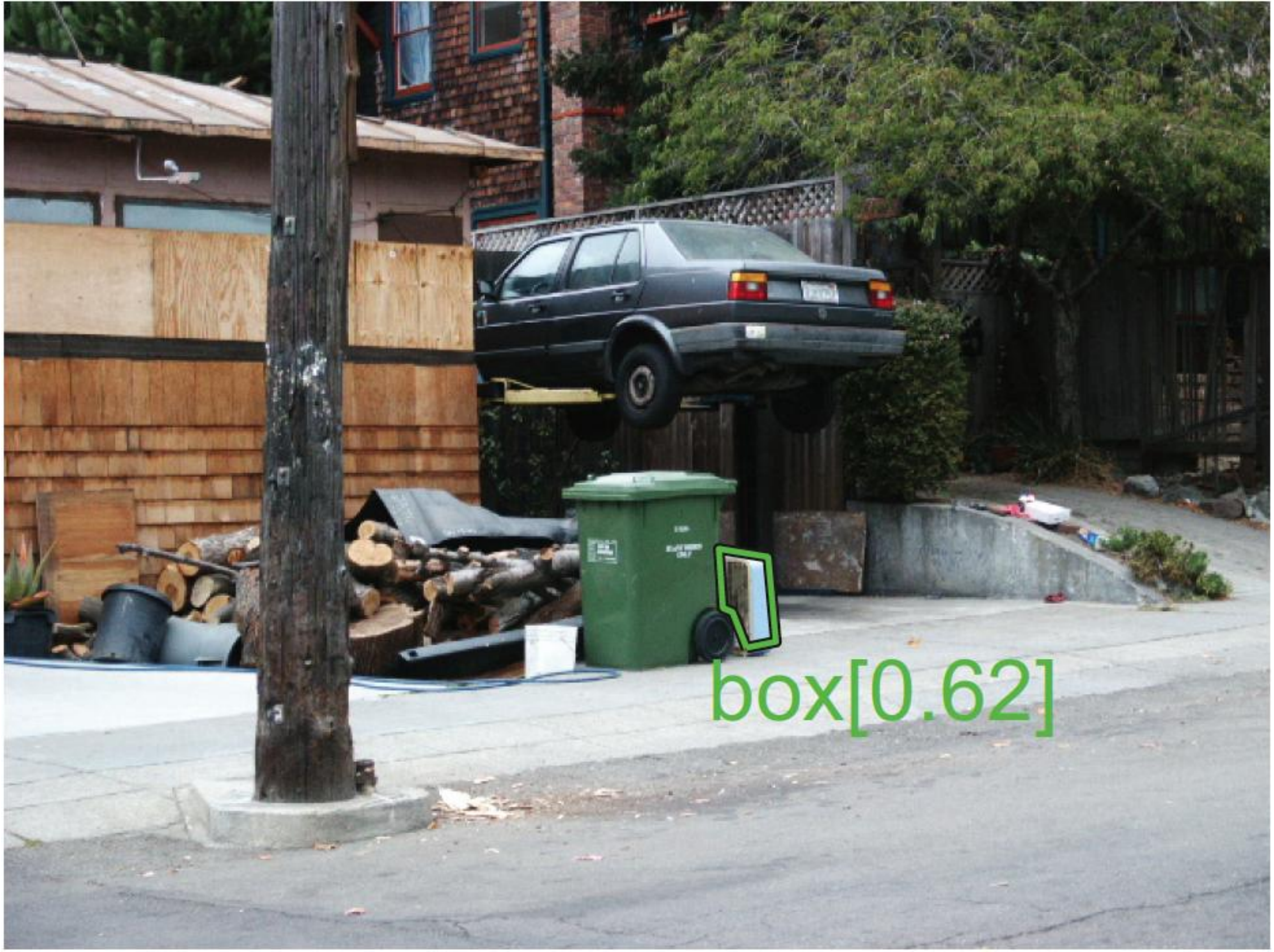
Locate the out of context object



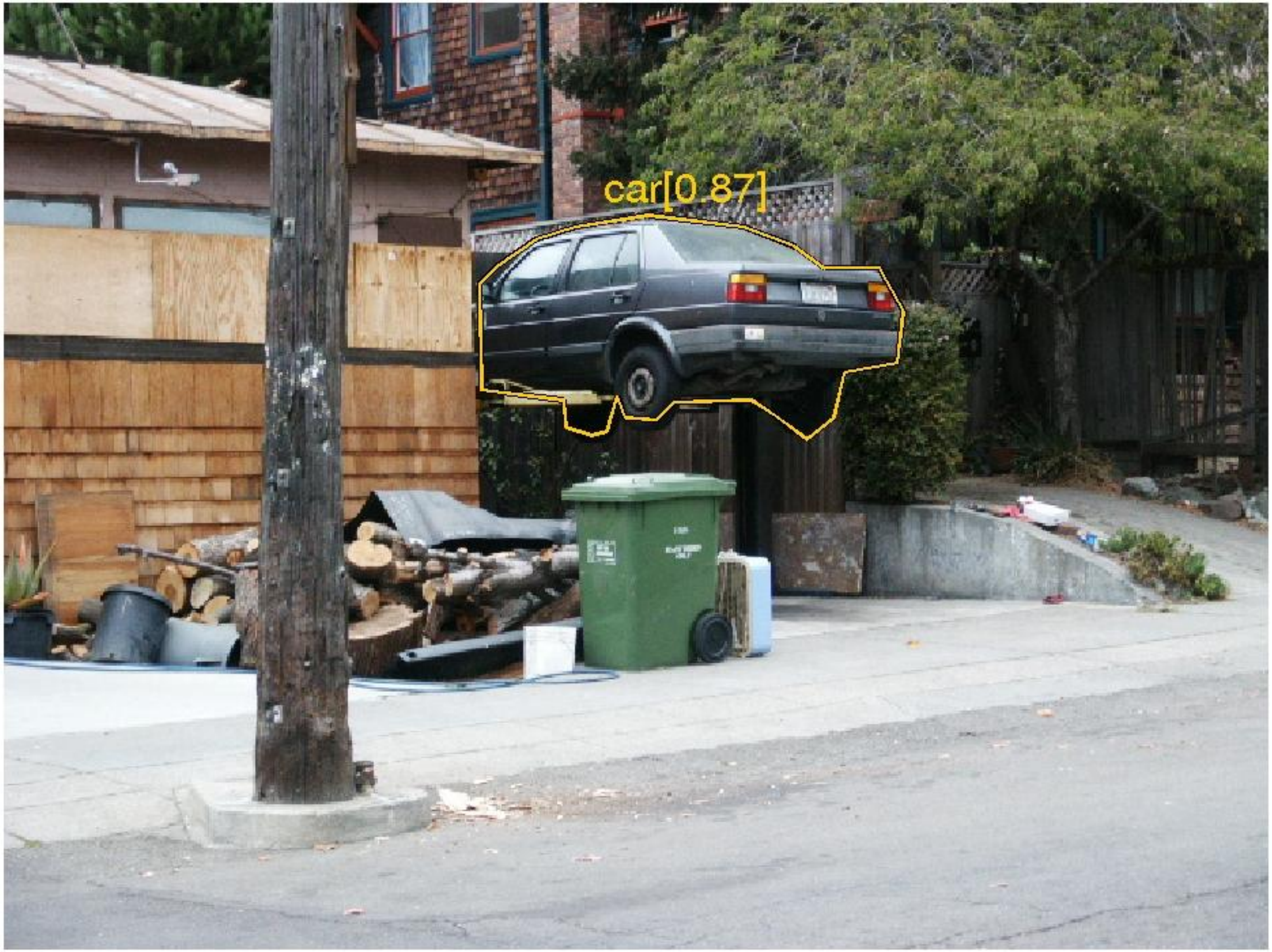
Co-occurrences only model



Co-occurrences and location model (Gaussian)

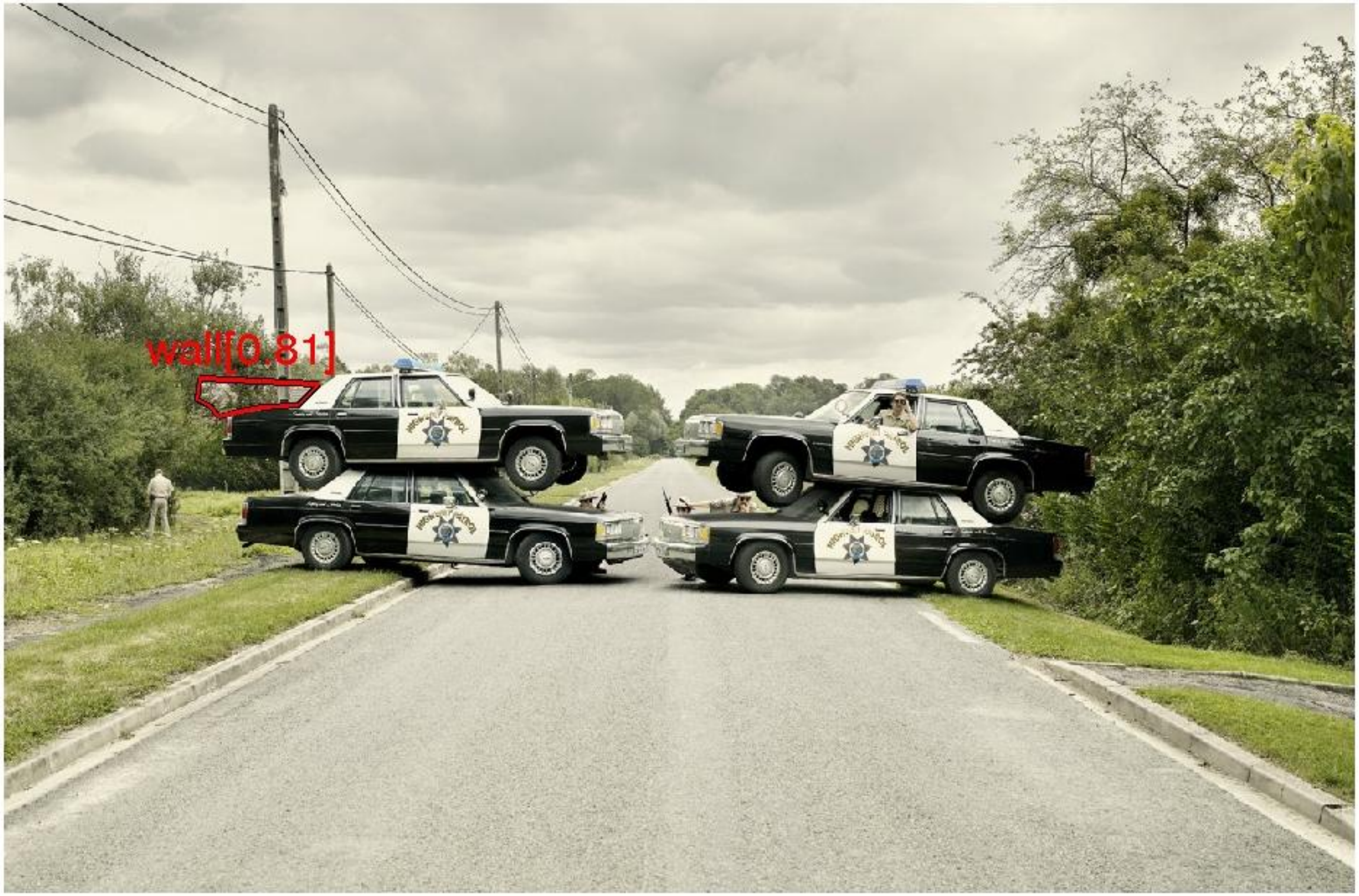


Co-occurrences and support model

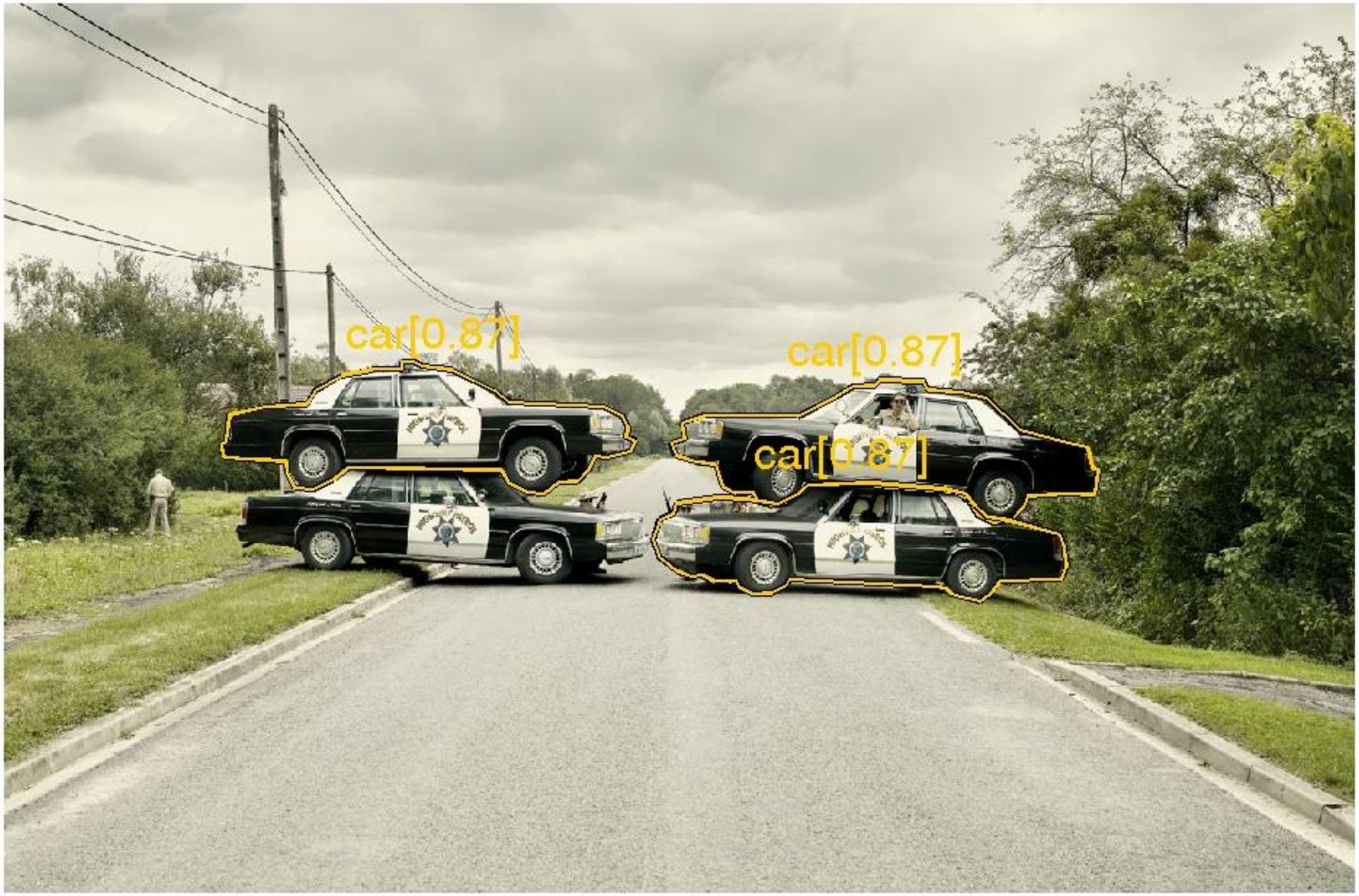


HIGHWAY PATROL
HIGHWAY PATROL
HIGHWAY PATROL
HIGHWAY PATROL

Co-occurrences only model

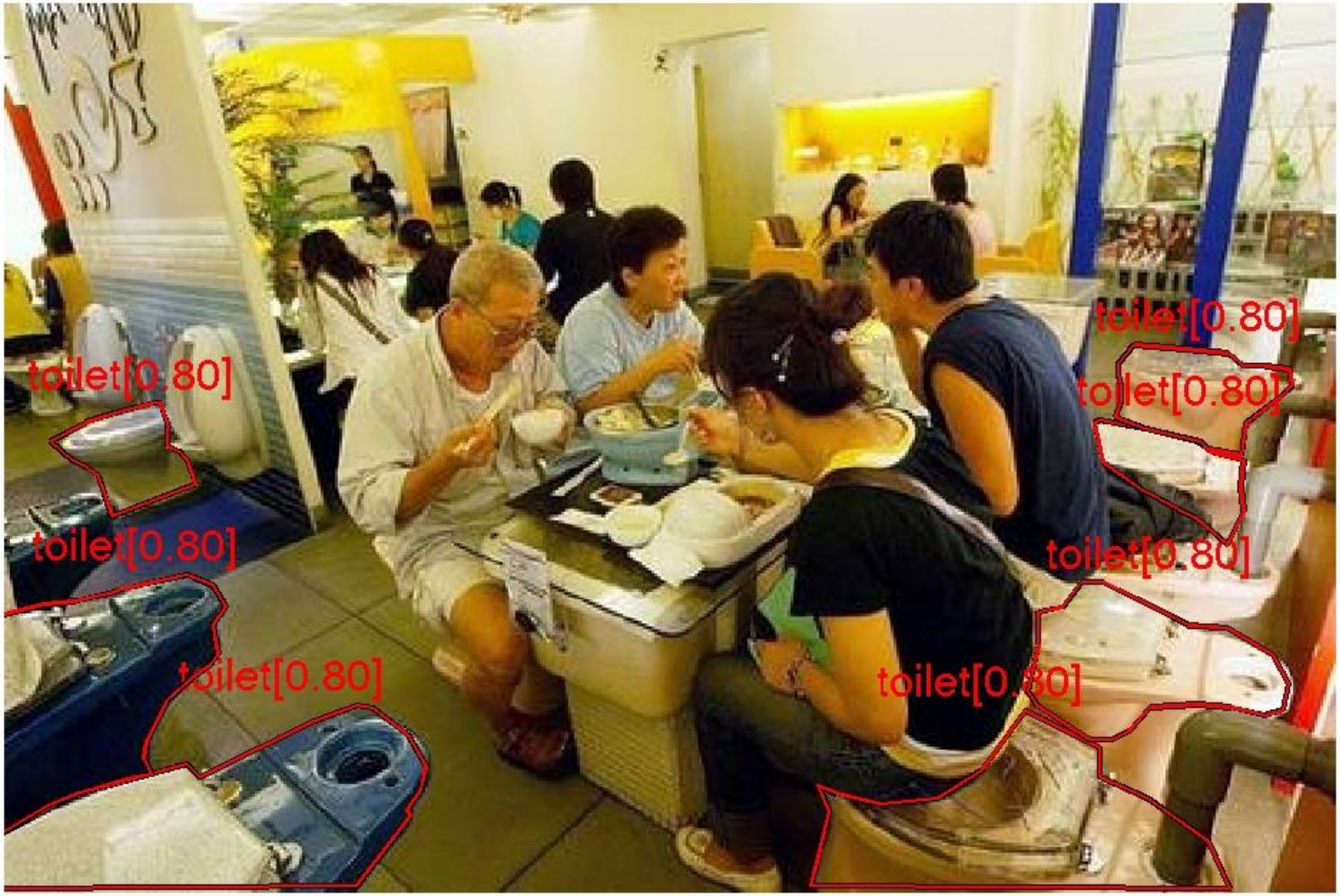


Co-occurrences and support model

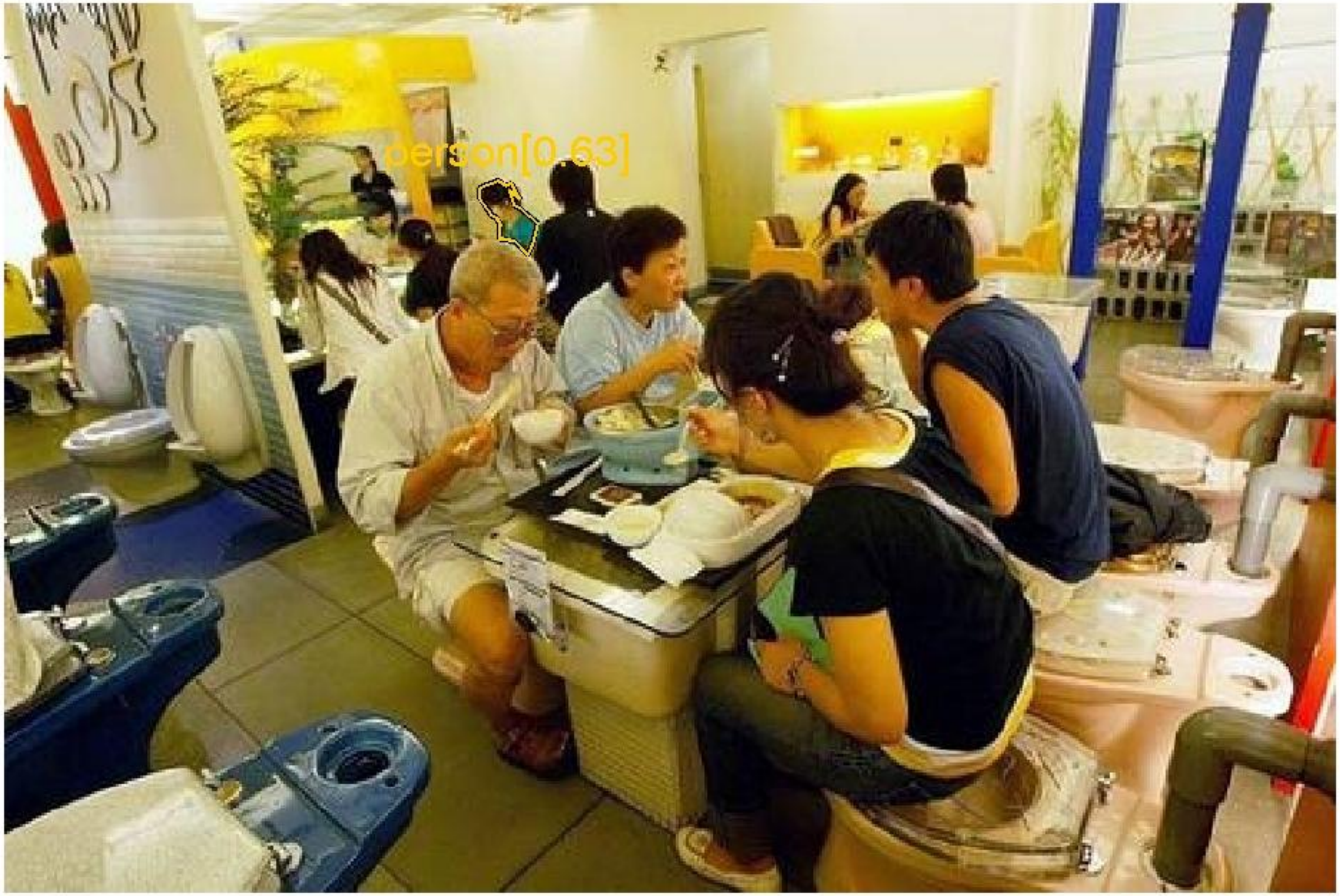


Image

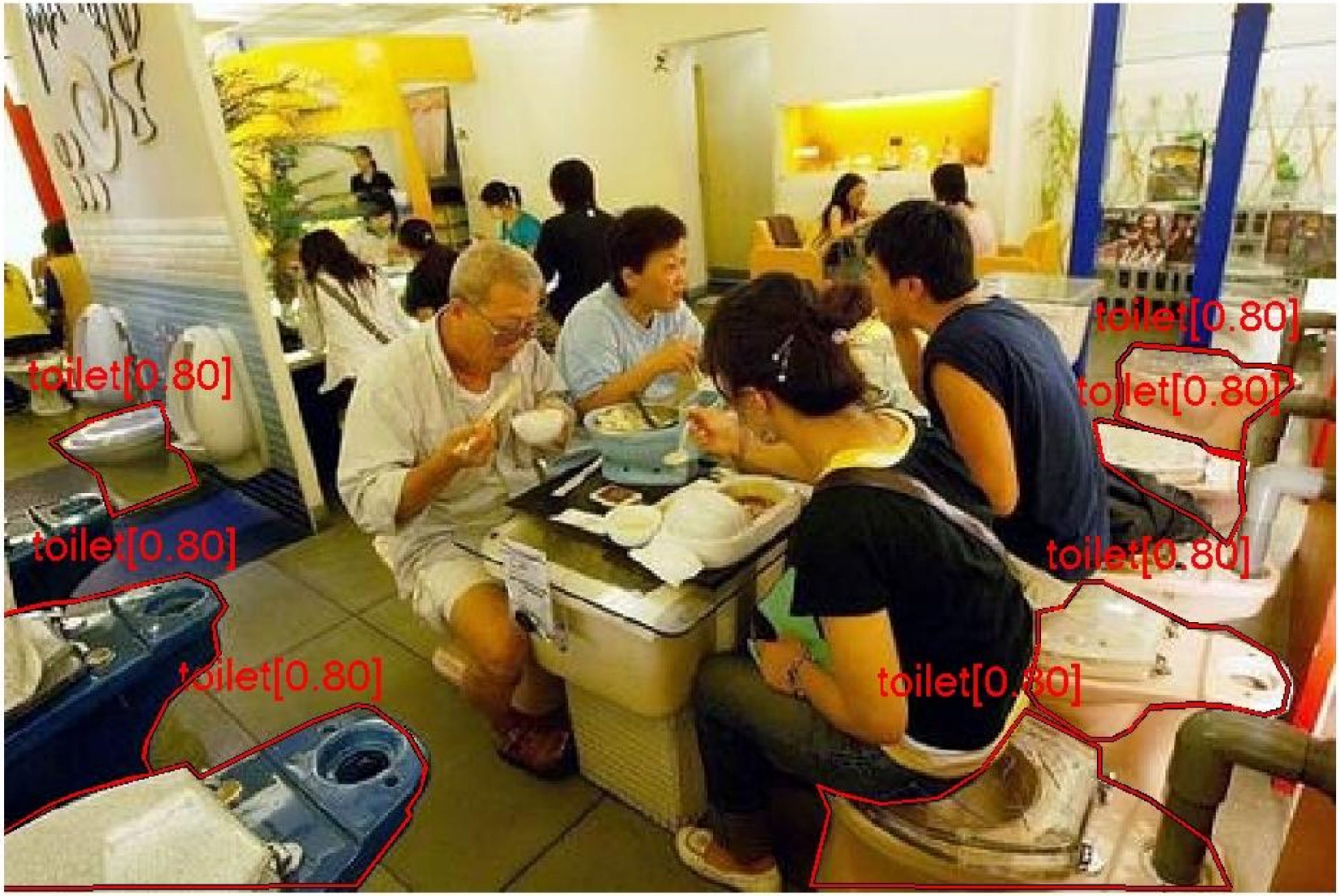
Co-occurrences only model



Support only model



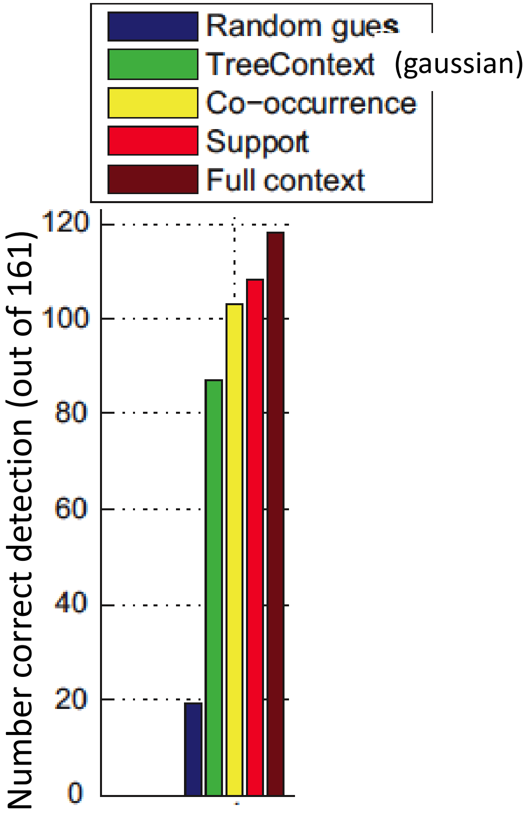
Co-occurrences and support model



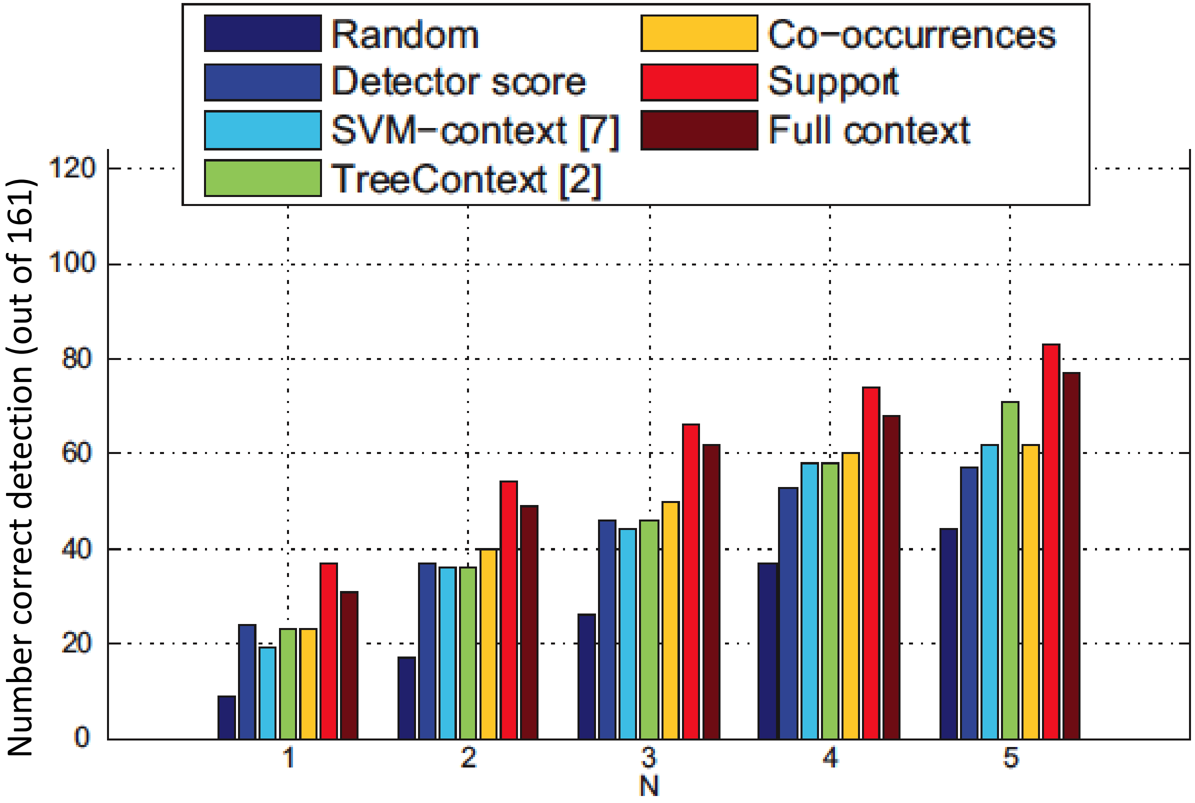
Detecting out of context objects

Out of 161 images

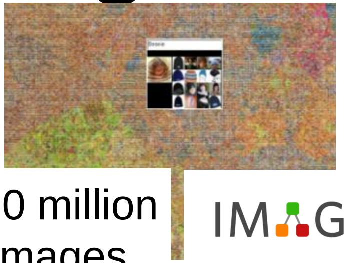
Ground-truth labels



From detector outputs



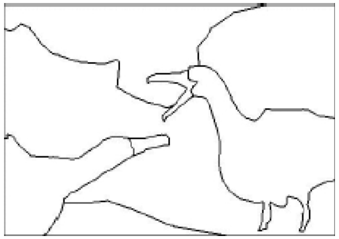
Big data collection efforts



80 million images

IMAGENET

Berkeley segmentation database



Caltech 101

Caltech-4

SUN database

Pascal

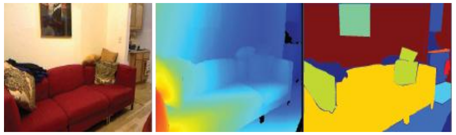
UIUC

UIUC

Attributes database

Has Horn
Has leg
Has Head
Has Wool

NYU Depth Dataset



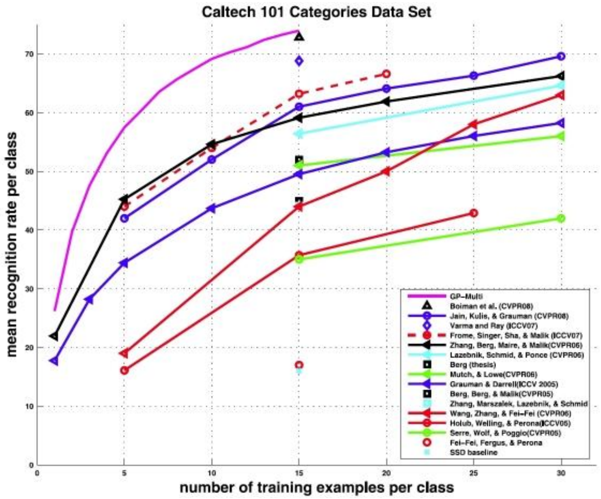
H3D Dataset

Keypoint Annotations 3D Pose Region Labels

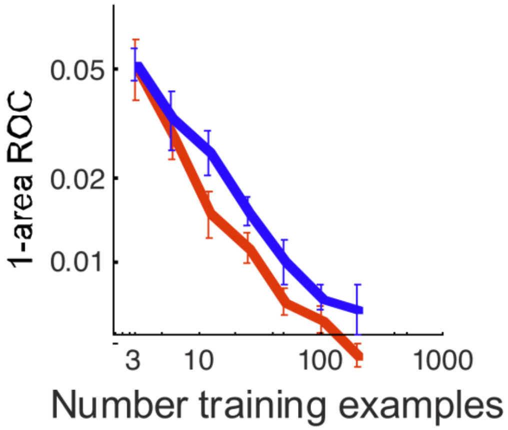
Segments Framed objects Scenes Parts & attributes 3D ?

The more data, the better

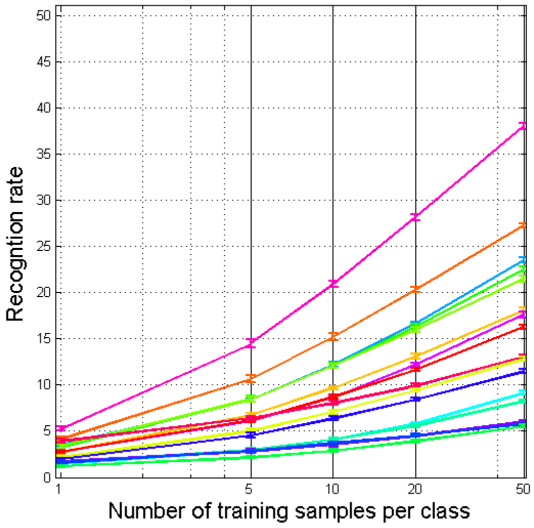
Classification (Caltech 101)



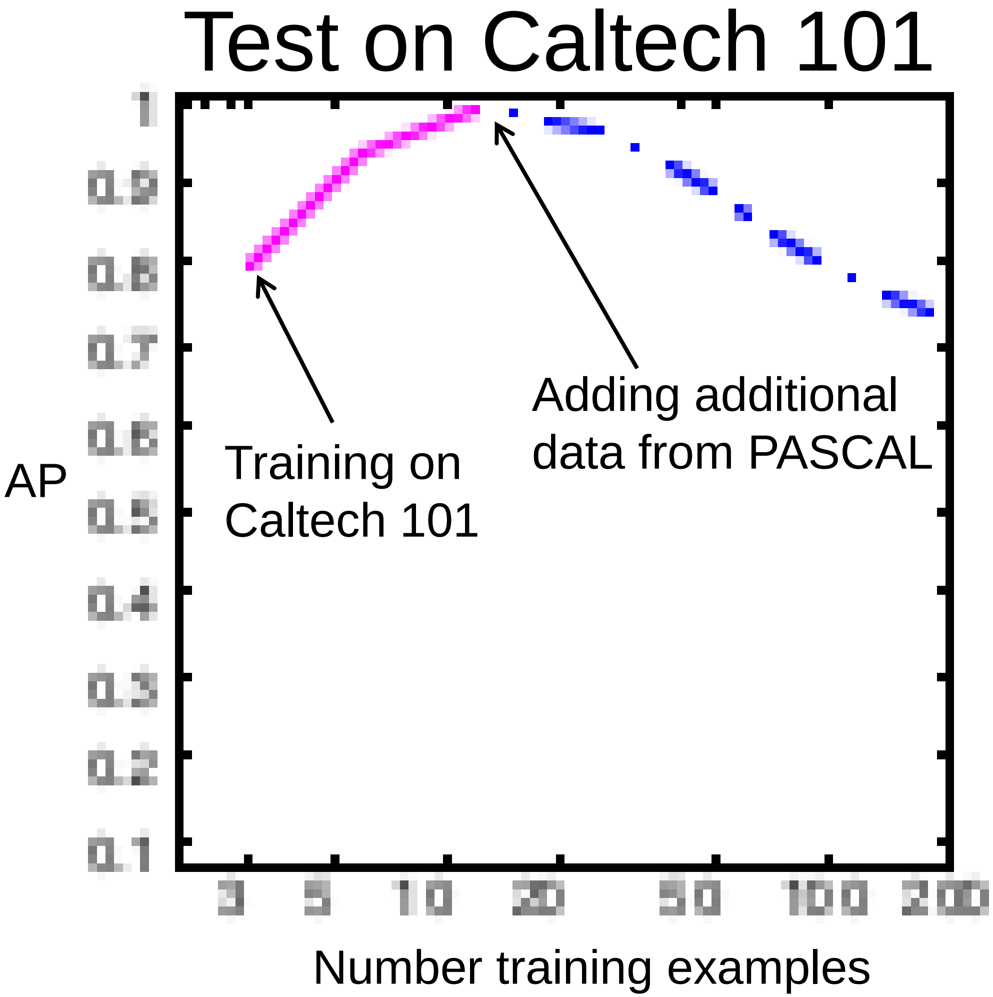
Car detection (PASCAL07, SUN09)



Scene recognition (SUN)



The benefits of getting more data



Test on Caltech 101

Training on Caltech 101

Adding additional data from PASCAL

Task: car detection
Features: HOG

Generalization across datasets

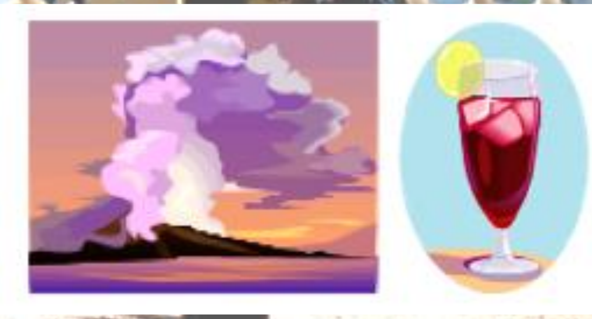
- A. Bergamo, L. Torresani and A. Fitzgibbon. PICODES: Learning a Compact Code for Novel-Category Recognition. NIPS 2011.
- F. Perronnin, J. Sánchez and Y. Liu, Large-Scale Image Categorization with Explicit Data Embedding. CVPR 2010.
- F. Perronnin, J. Sánchez and T. Mensink, Improving the Fisher Kernel for Large-Scale Image Classification. ECCV 2010.
- P. Dollar, C. Wojek, B. Schiele and P. Perona, Pedestrian Detection: A Benchmark. CVPR 2009.
- …

Unbiased Look at Dataset Bias

Alyosha Efros (CMU)

Antonio Torralba (MIT)

CSAIL THE ROBOTICS INSTITUTE



Disclaimer: no graduate students have been harmed in the production of this paper

Are datasets measuring the right thing?

- In Machine Learning:

  Dataset is The World

- In Recognition

  Dataset is a *representation* of The World

- ML solution: domain transfer
- Vision question: Do datasets provide a good representation?

Visual Data is Inherently Biased

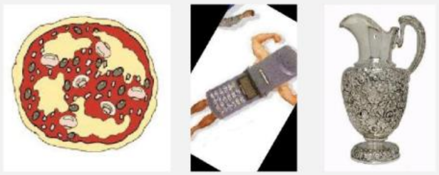
- Internet is a tremendous repository of visual data (Flickr, YouTube, Picassa, etc)
- But it’s not random samples of visual world

Our Question

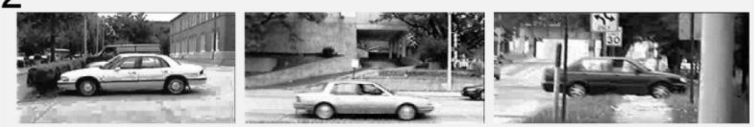
- How much does this bias affect standard datasets used for object recognition?

"*Name That Dataset!*" game

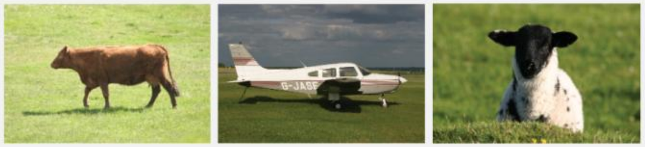
1



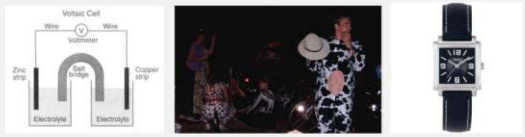
2



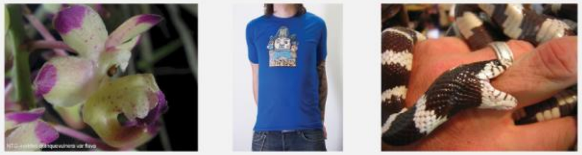
3



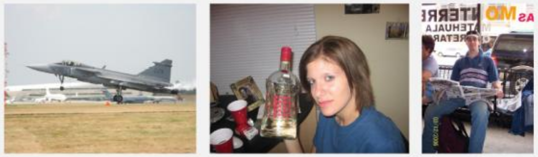
4



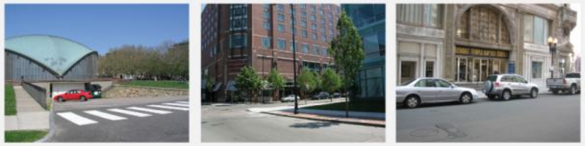
5



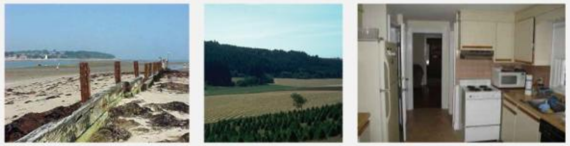
6



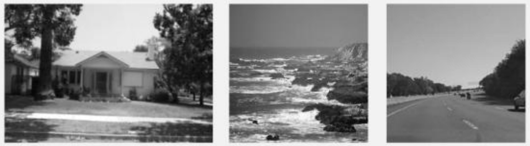
7



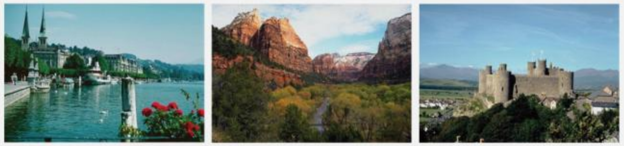
8



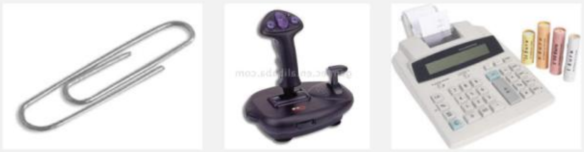
9



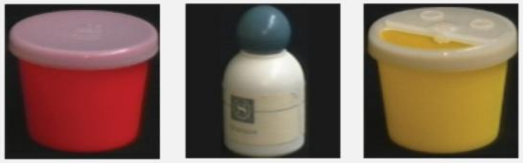
10



11



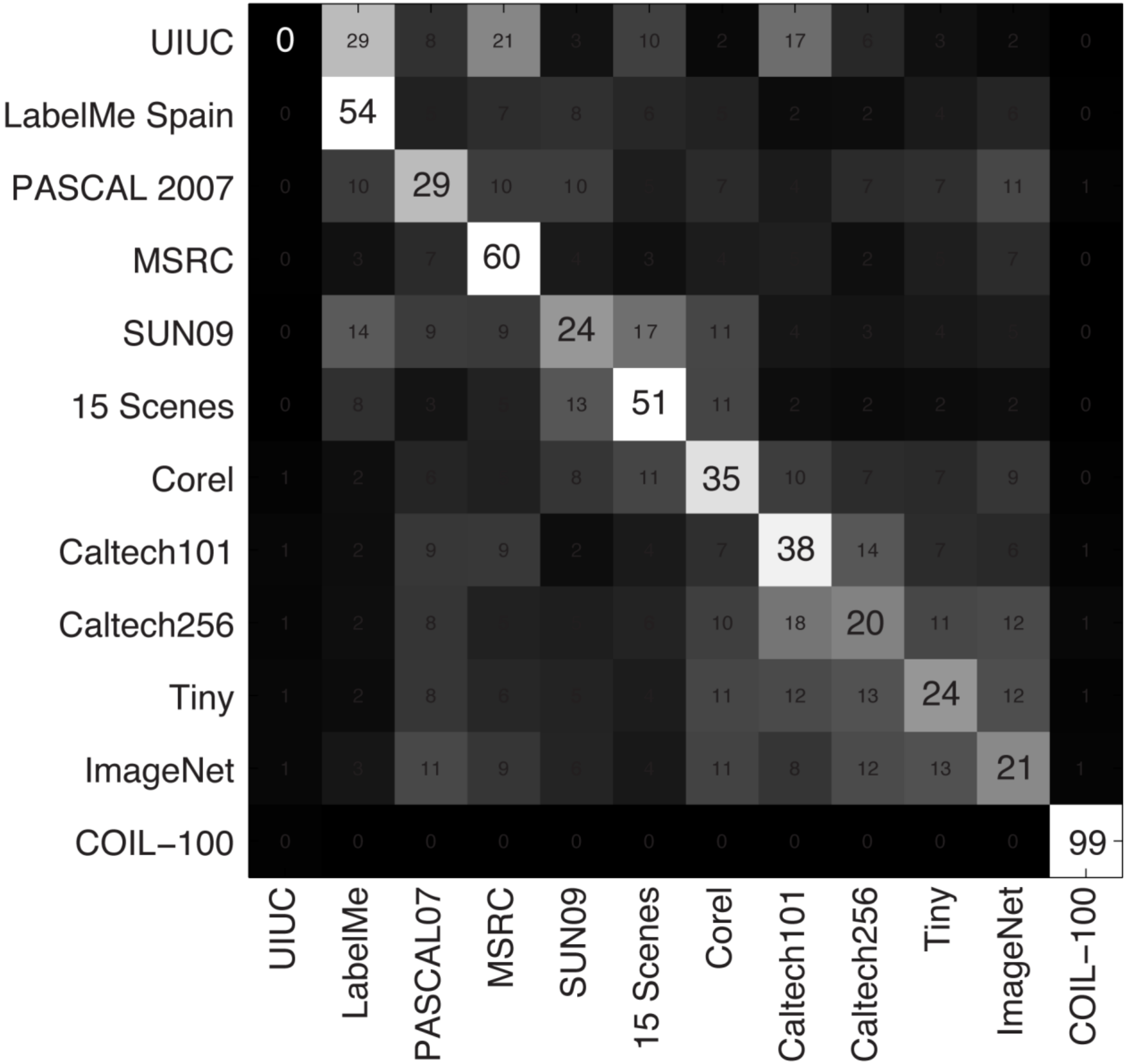
12



___ Caltech 101
___ Caltech 256
___ MSRC
___ UIUC cars
___ Tiny Images
___ Corel
___ PASCAL 2007
___ LabelMe
___ COIL-100
___ ImageNet
___ 15 Scenes
___ SUN'09

SVM plays *“Name that dataset!”*

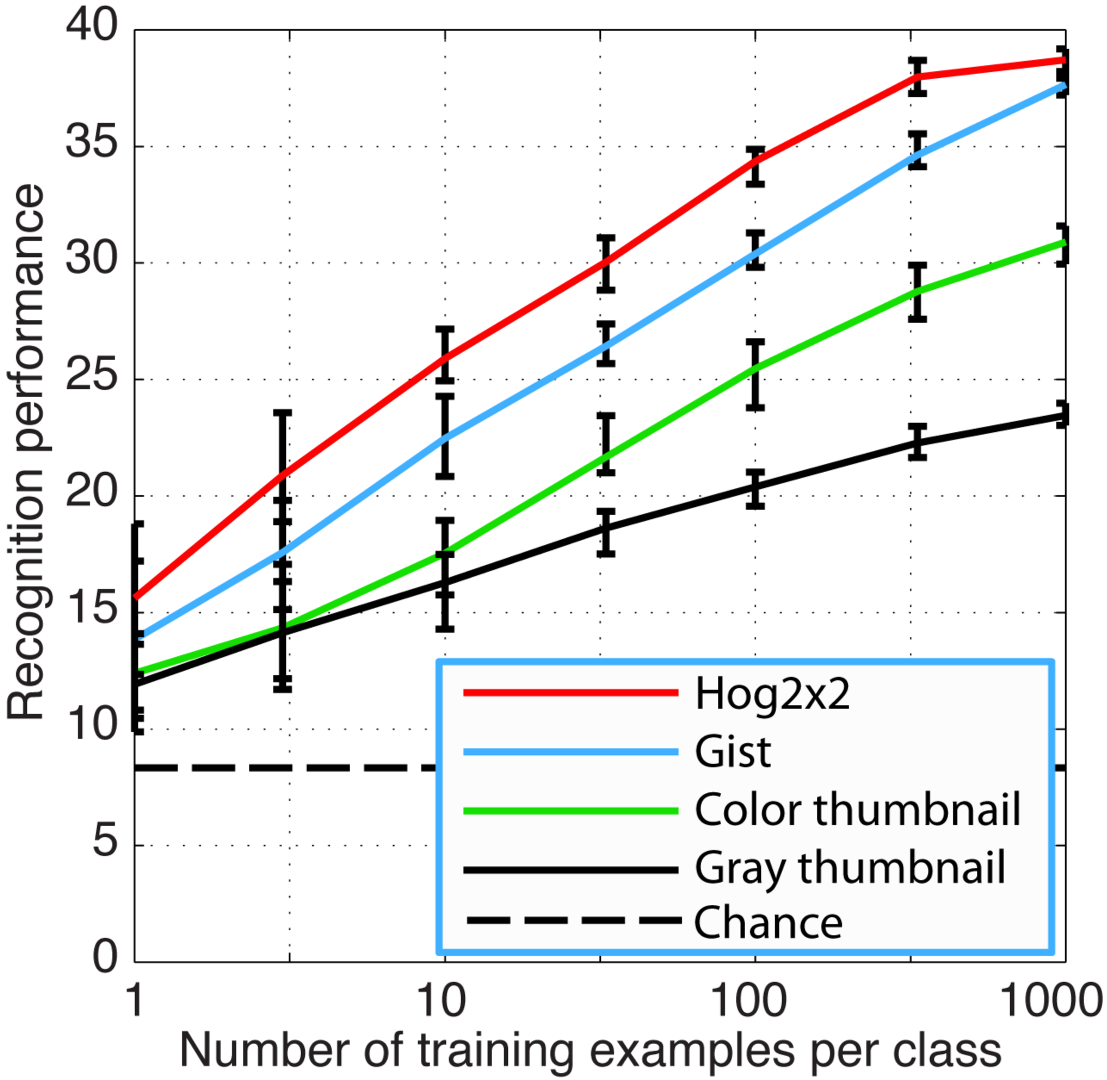
SVM plays *“Name that dataset!”*



	UIUC	LabelMe	PASCAL07	MSRC	SUN09	15 Scenes	Corel	Caltech101	Caltech256	Tiny	ImageNet	COIL–100
UIUC	0	29	8	21	3	10	2	17	6	3	2	0
LabelMe Spain	0	54	5	7	8	6	5	2	2	4	6	0
PASCAL 2007	0	10	29	10	10	5	7	4	7	7	11	1
MSRC	0	3	7	60	4	3	4	5	2	5	7	0
SUN09	0	14	9	9	24	17	11	4	3	4	5	0
15 Scenes	0	8	3	5	13	51	11	2	2	2	2	0
Corel	1	2	6	5	8	11	35	10	7	7	9	0
Caltech101	1	2	9	9	2	4	7	38	14	7	6	1
Caltech256	1	2	8	5	5	6	10	18	20	11	12	1
Tiny	1	2	8	6	5	4	11	12	13	24	12	1
ImageNet	1	3	11	9	6	4	11	8	12	13	21	1
COIL–100	0	0	0	0	0	0	0	0	0	0	0	99

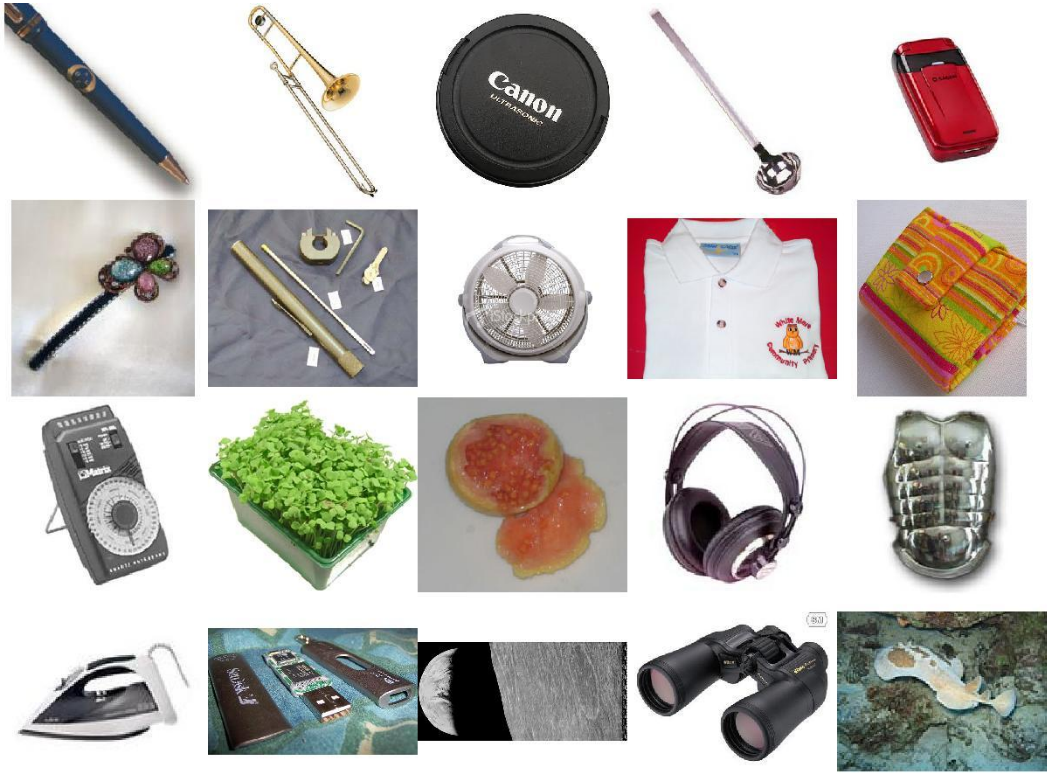
- 12 1-vs-all classifiers
- Standard full-image features
- 39% performance (chance is 8%)

SVM plays *“Name that dataset!”*



Dataset look-alikes

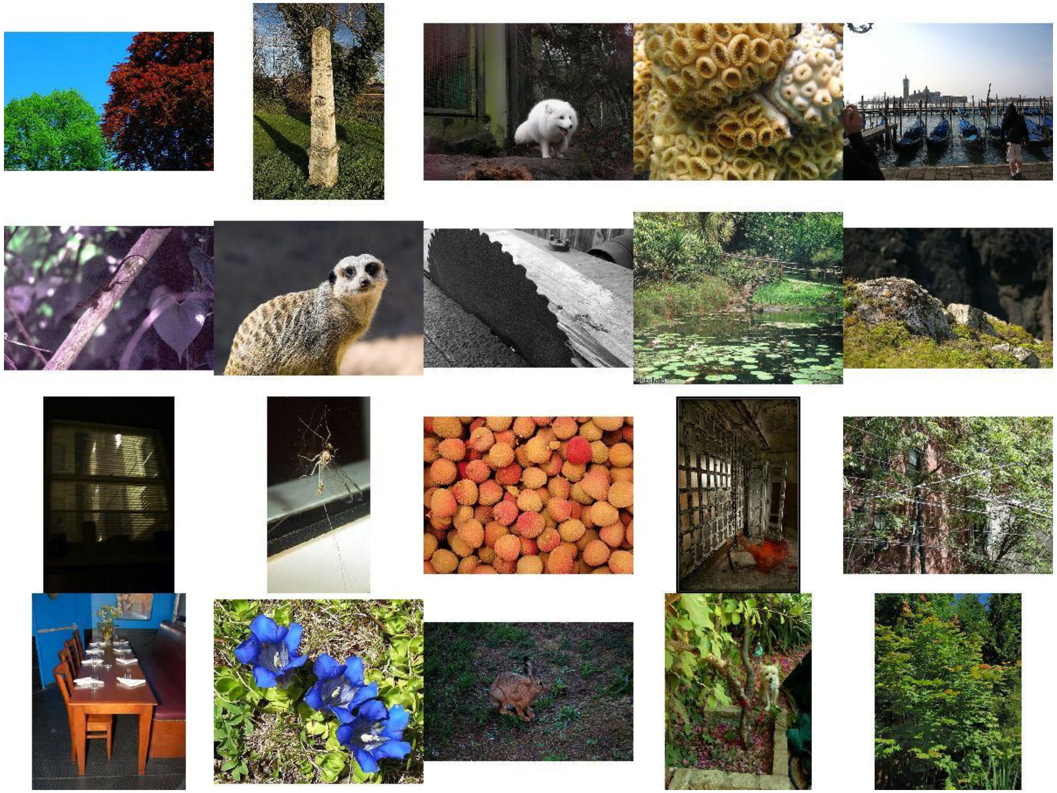
ImageNet pretending to be …



… Caltech 256

Dataset look-alikes

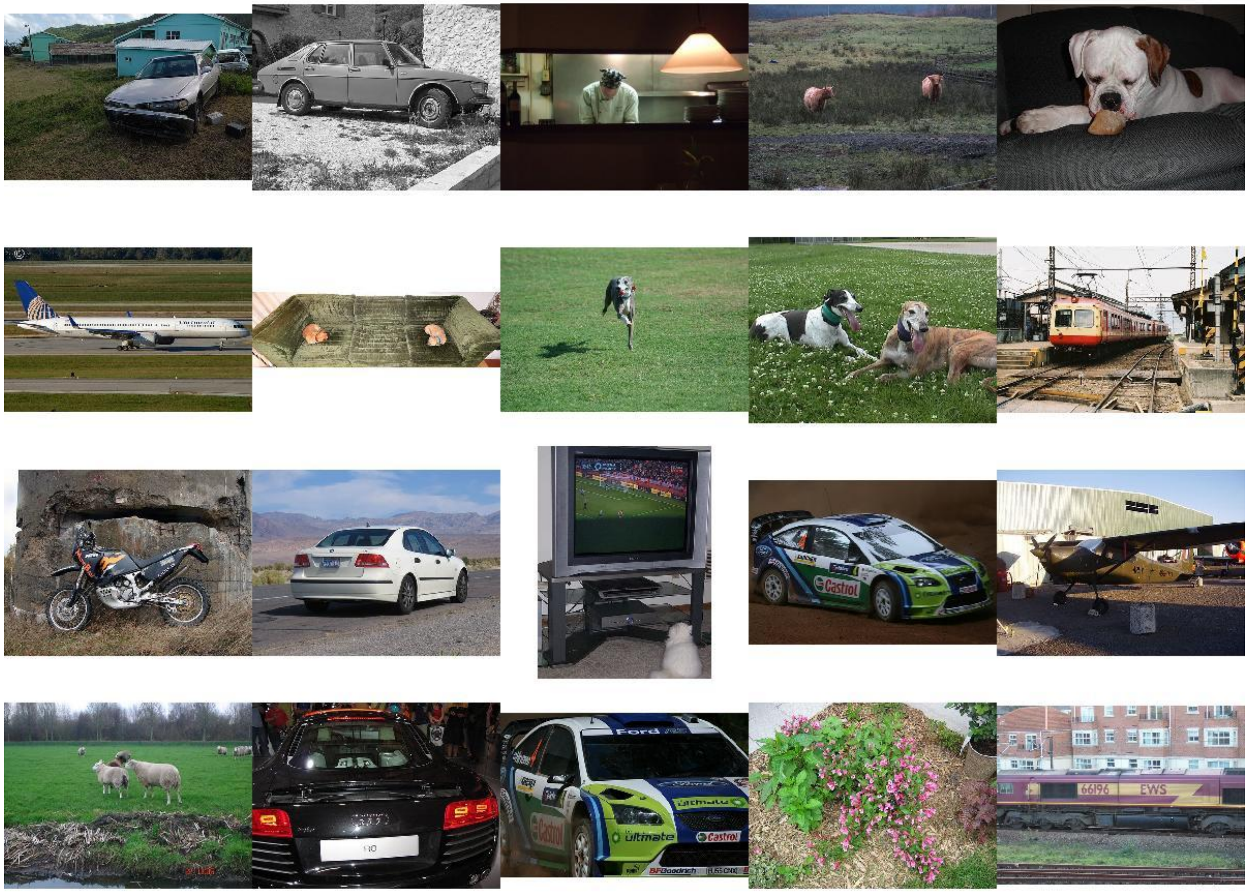
ImageNet pretending to be …



… COREL

Dataset look-alikes

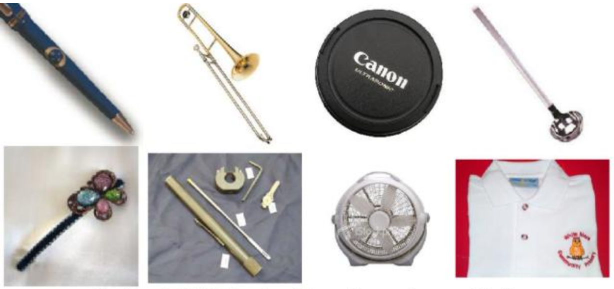
PASCAL VOC pretending to be …



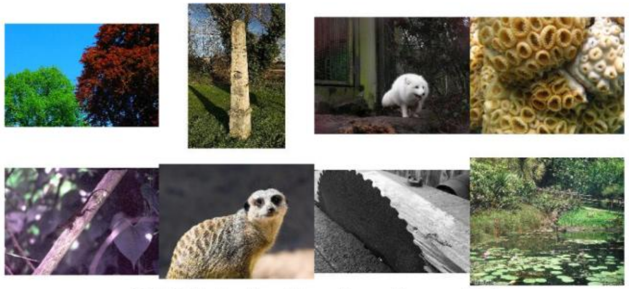
… MSRC

Dataset look-alikes

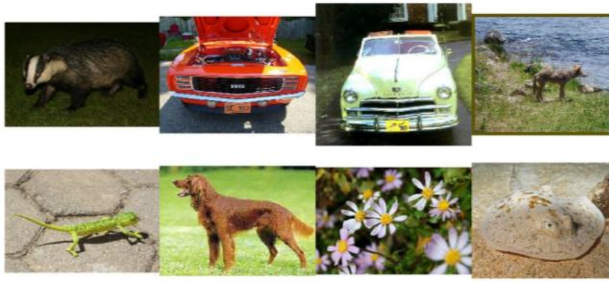
ImageNet pretending to be:



Caltech 256 look-alikes from ImageNet

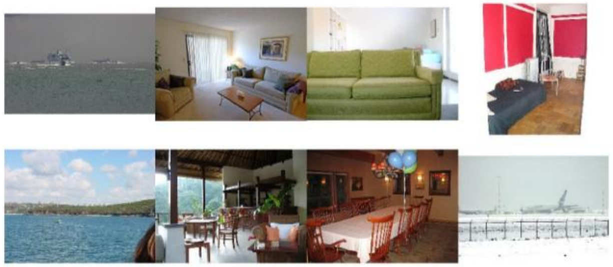


COREL look-alikes from ImageNet

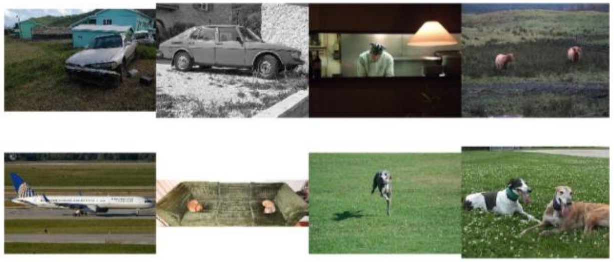


MSRC look-alikes from ImageNet

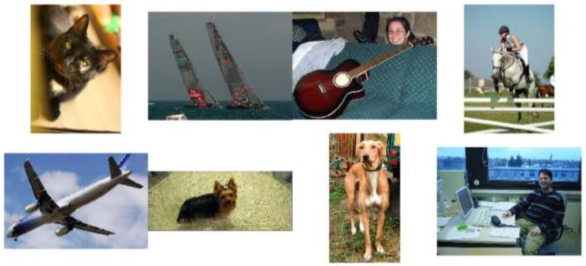
PASCAL VOC pretending to be:



15 scenes look-a-likes from PASCAL 2007



MSRC look-alikes from PASCAL 2007



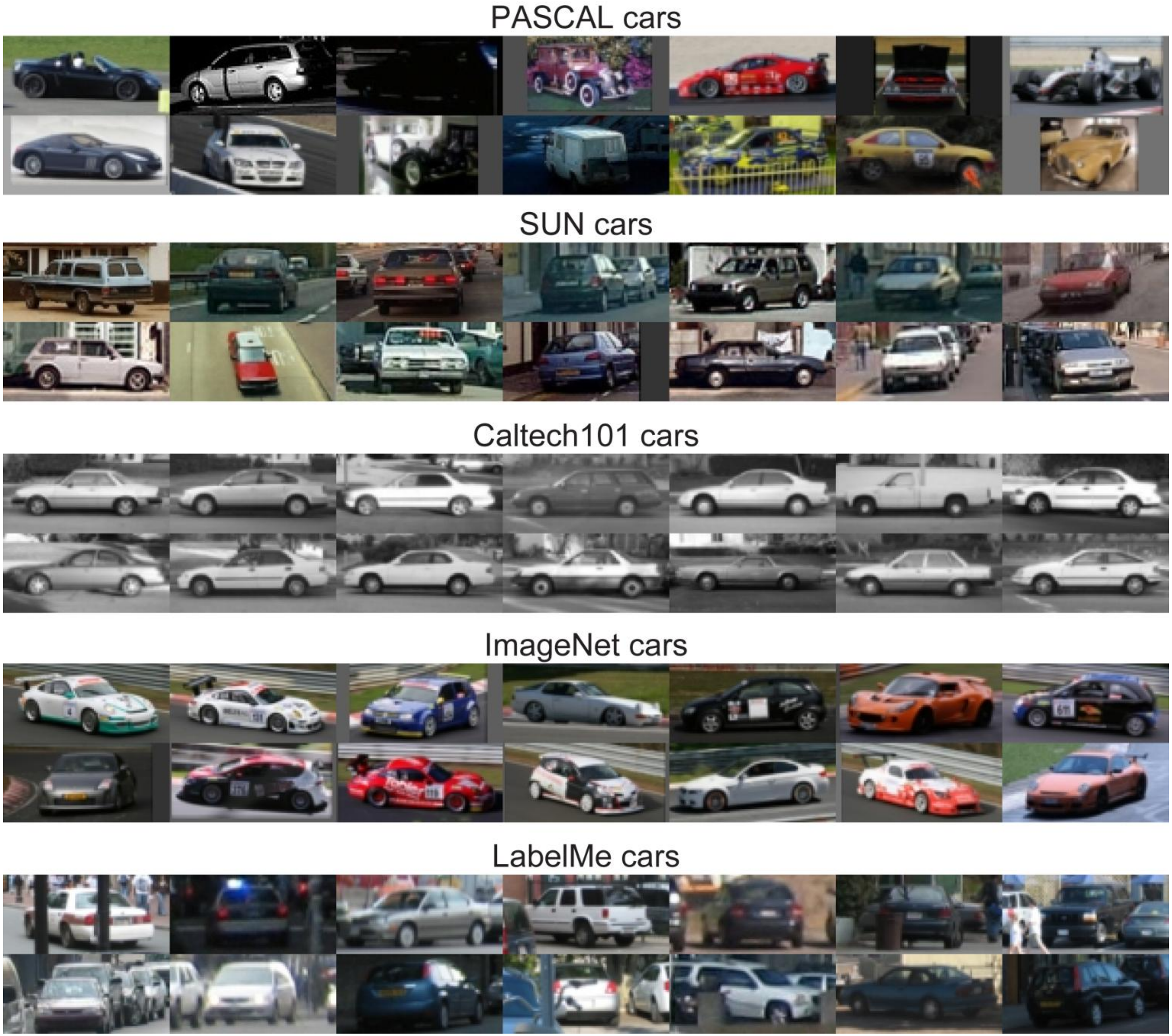
Caltech 101 look-alikes from PASCAL 2007

Datasets have different goals…

- Some are object-centric (e.g. Caltech, ImageNet)
- Otherwise are scene-centric (e.g. LabelMe, SUN’09)

- What about playing *“name that dataset”* on bounding boxes?

Similar results



Performance: 61%
(chance: 20%)

Cross-Dataset Generalization

MSRC

Classifier trained on MSRC cars

Cross-dataset Performance

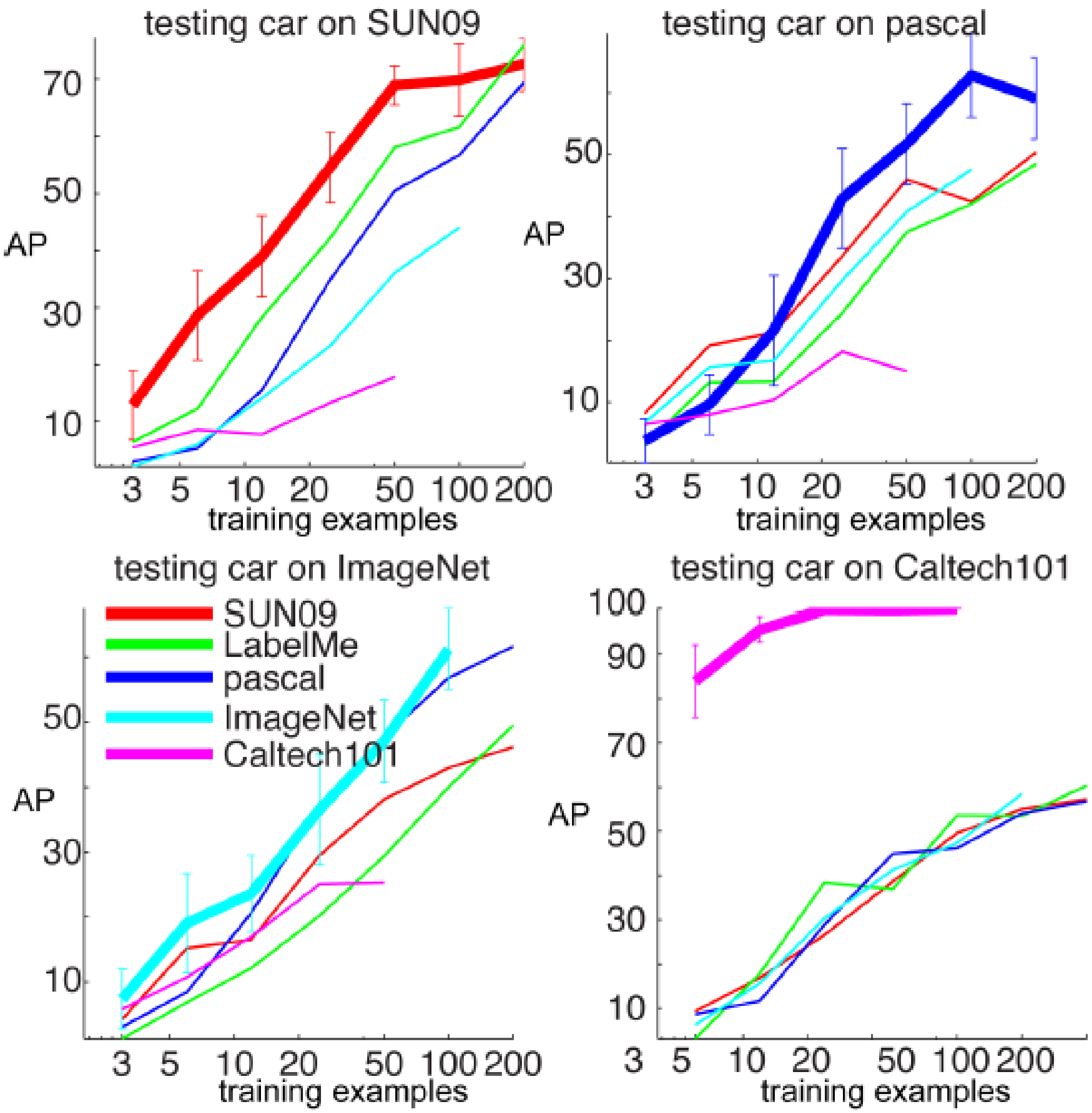
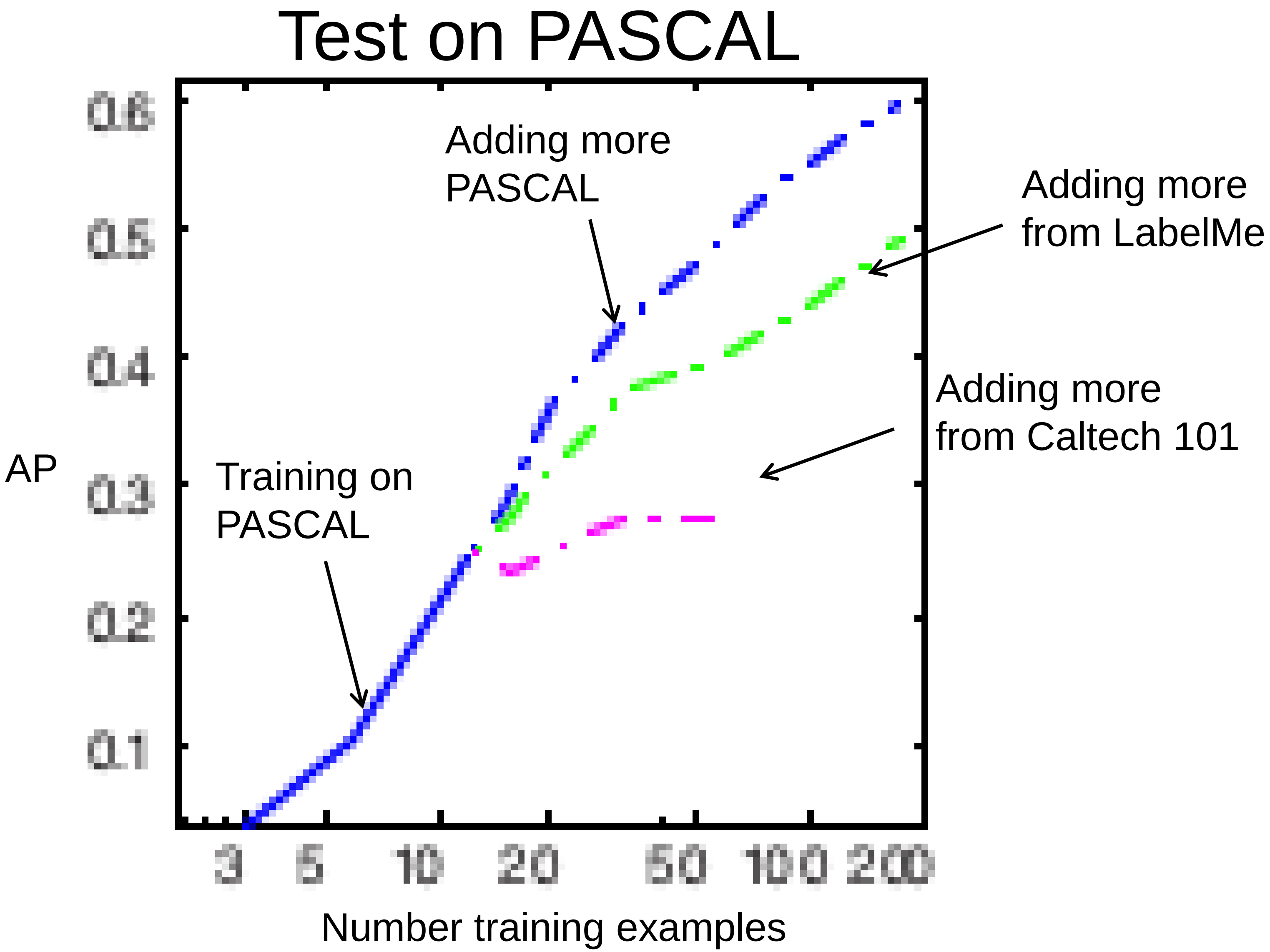


Figure 6. Cross-dataset generalization for "car" detection as function of training data

Mixing datasets



Test on PASCAL

Adding more PASCAL

Adding more from LabelMe

Adding more from Caltech 101

Training on PASCAL

Dataset Value

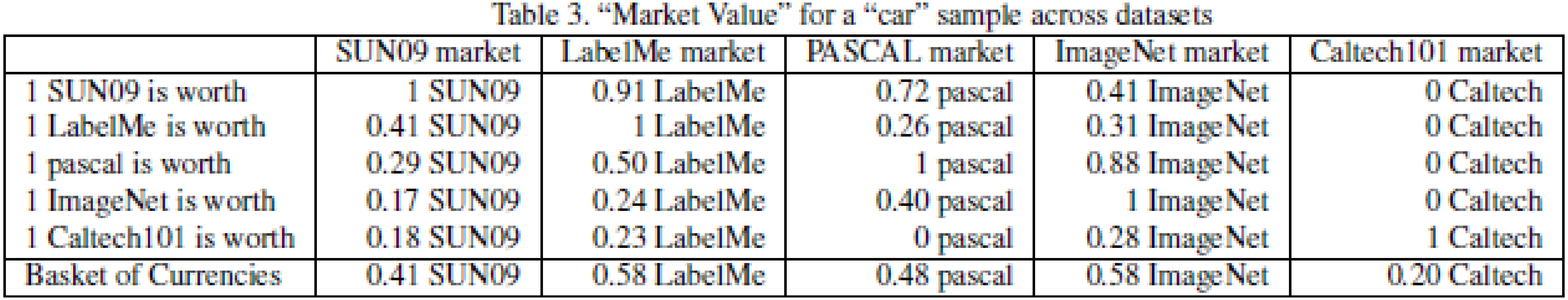


Table 3. "Market Value" for a "car" sample across datasets

	SUN09 market	LabelMe market	PASCAL market	ImageNet market	Caltech101 market
1 SUN09 is worth	1 SUN09	0.91 LabelMe	0.72 pascal	0.41 ImageNet	0 Caltech
1 LabelMe is worth	0.41 SUN09	1 LabelMe	0.26 pascal	0.31 ImageNet	0 Caltech
1 pascal is worth	0.29 SUN09	0.50 LabelMe	1 pascal	0.88 ImageNet	0 Caltech
1 ImageNet is worth	0.17 SUN09	0.24 LabelMe	0.40 pascal	1 ImageNet	0 Caltech
1 Caltech101 is worth	0.18 SUN09	0.23 LabelMe	0 pascal	0.28 ImageNet	1 Caltech
Basket of Currencies	0.41 SUN09	0.58 LabelMe	0.48 pascal	0.58 ImageNet	0.20 Caltech

Overall…

- Caltech, MSRC – bad
- PASCAL, ImageNet – better

We are getting better. The new datasets are better than the old ones.

A green pasture for research: “Understanding and Living with dataset bias”

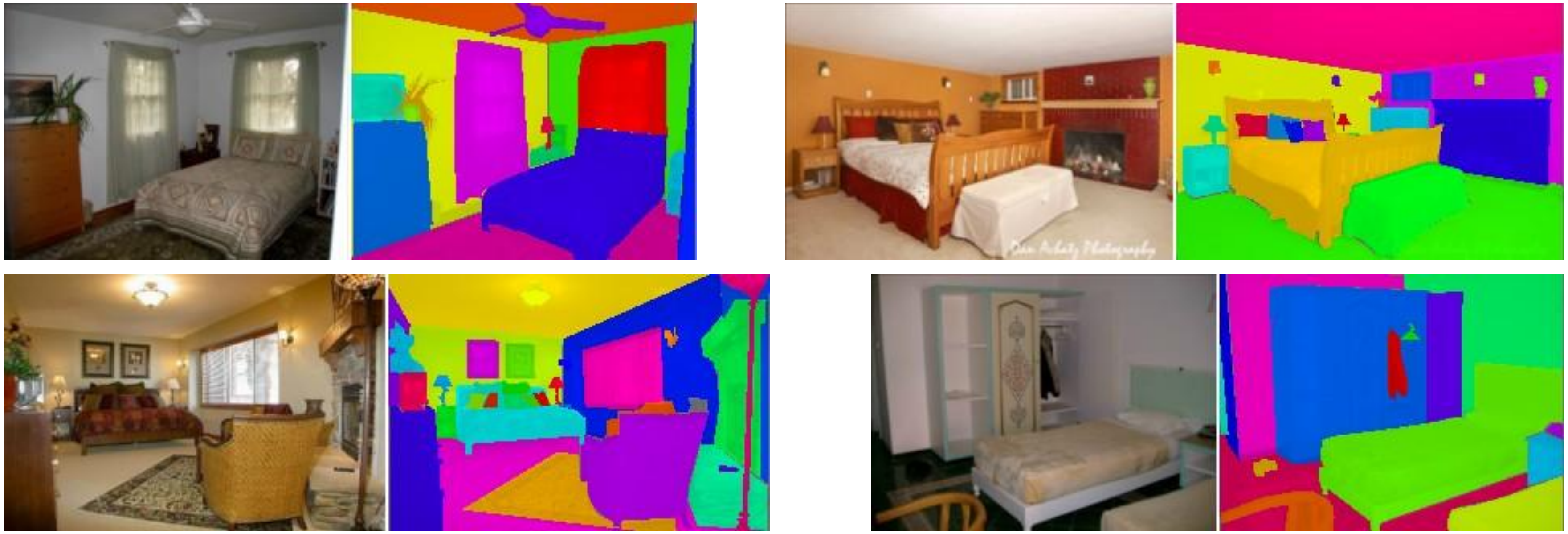
Where does the bias come from?

How do we live with it?

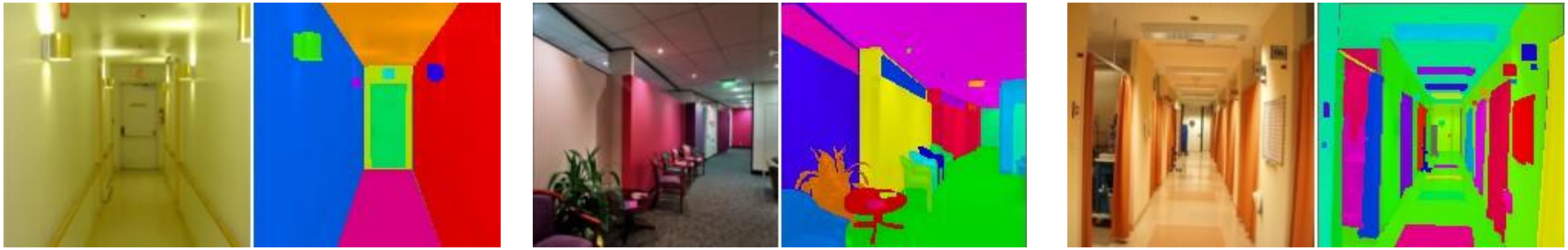
Where do this bias comes from?

Photographer bias

SUN database bedrooms



SUN database corridors

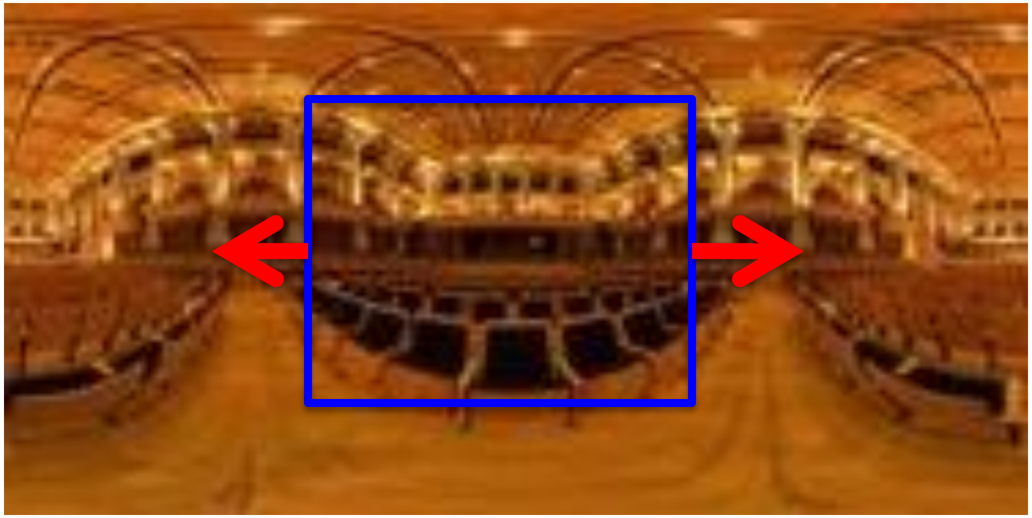


Viewpoint Annotation for Truth

Adjust the view of the panoramic image on the right so that it matches the view shown on the left.

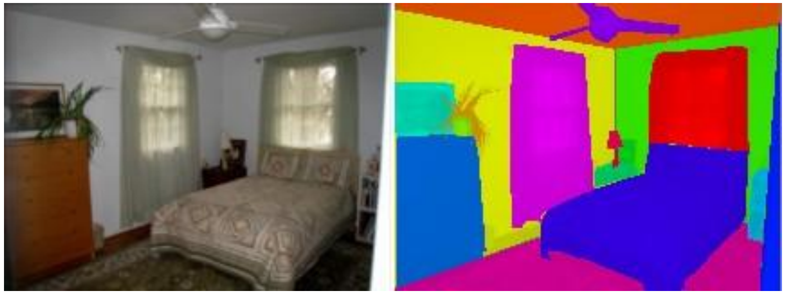
Target View:

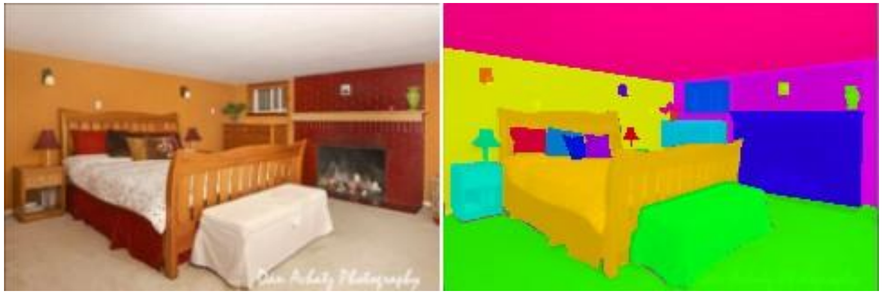
Panorama: Adjust the view to match the target view.

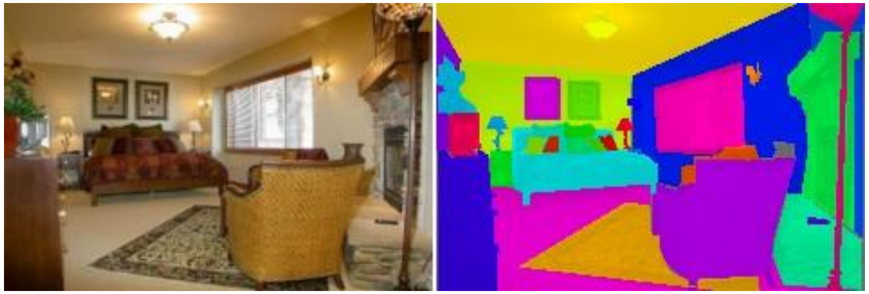


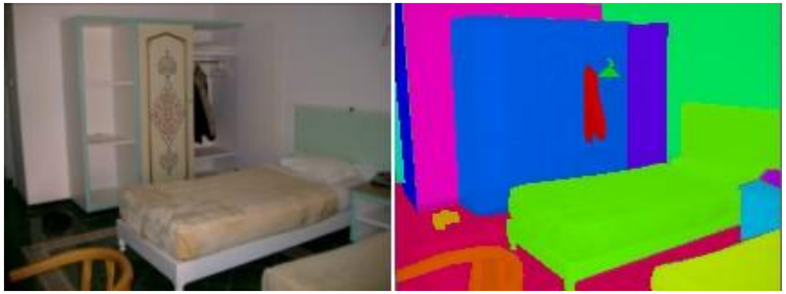
Amazon Mechanical Turks $0.01 Task.

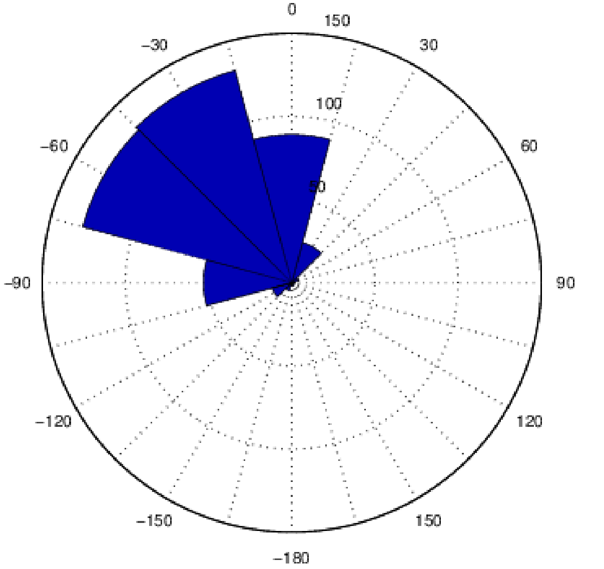
Pictures of bedrooms

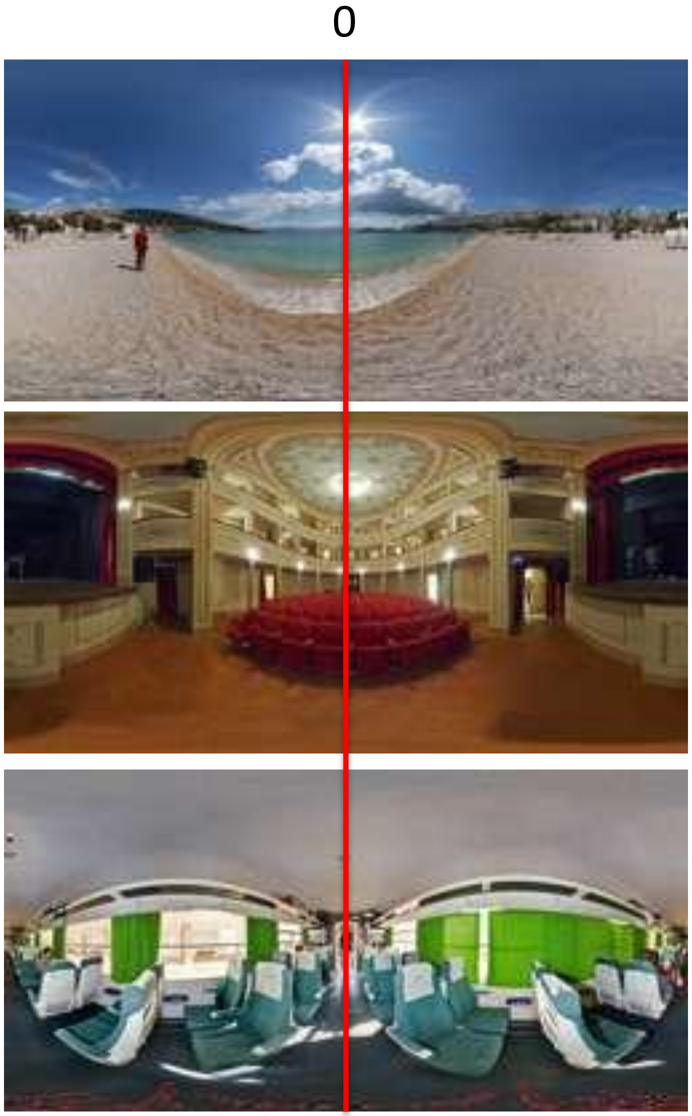




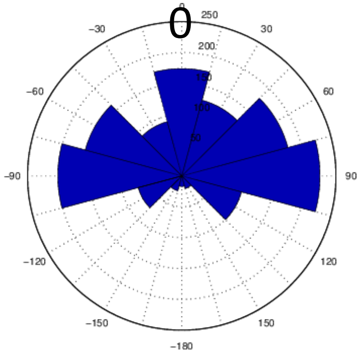




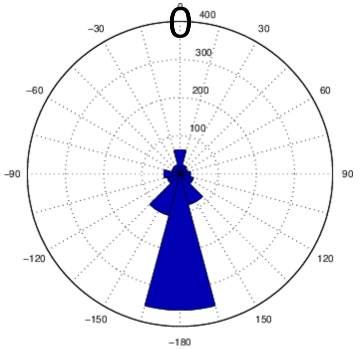




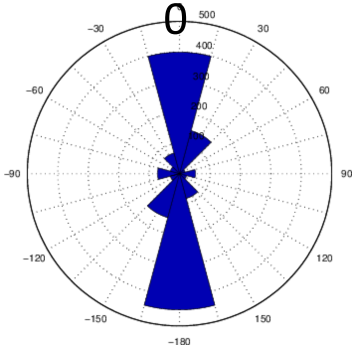
Beach

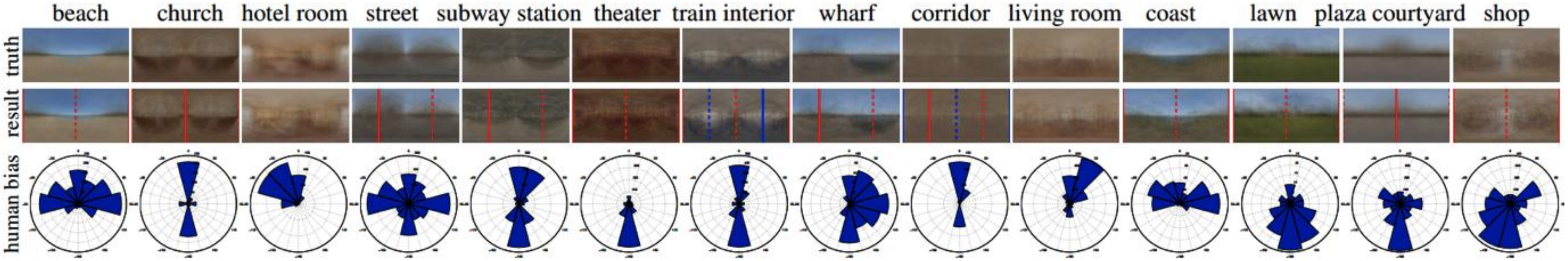


Theater

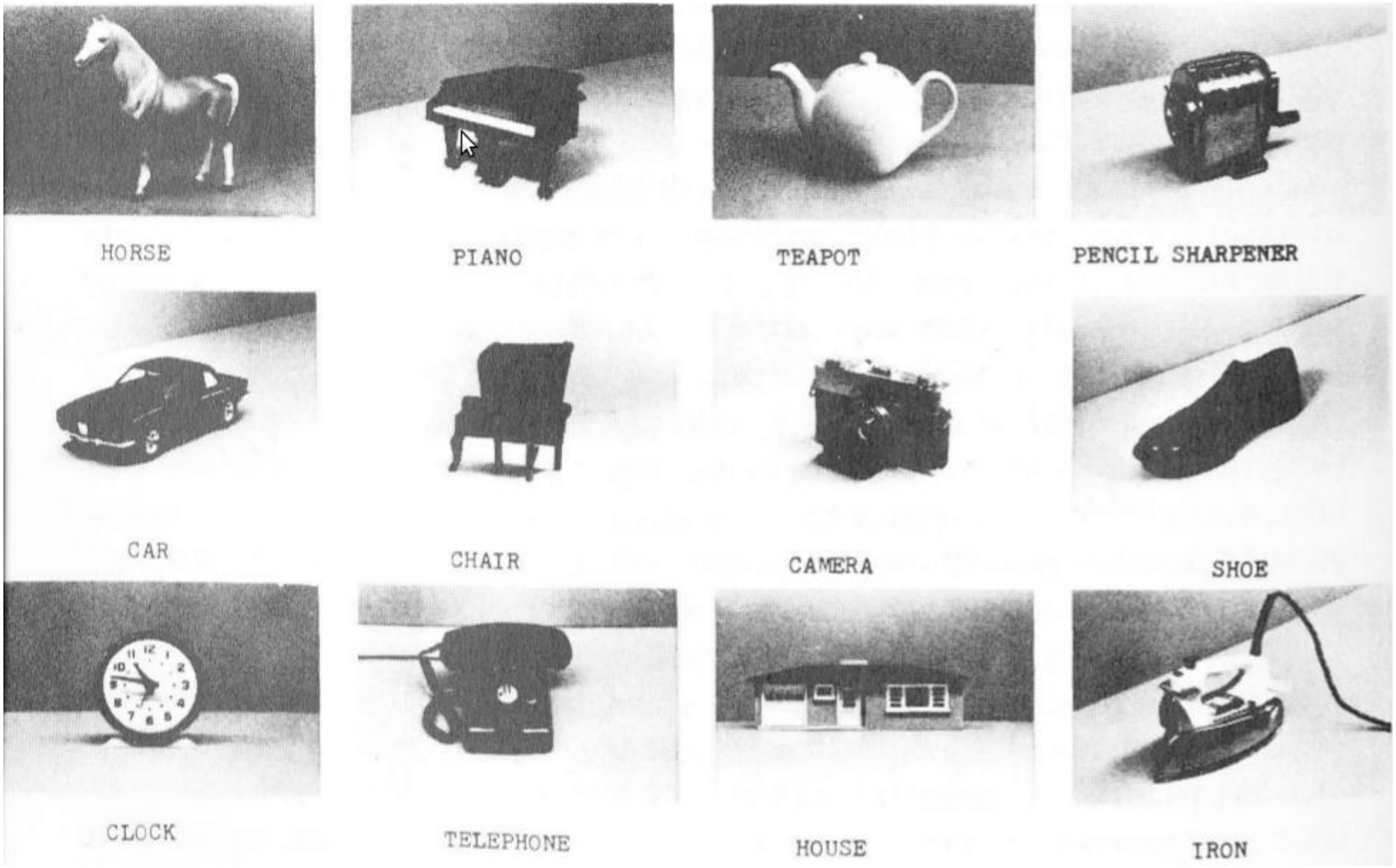


Inside train



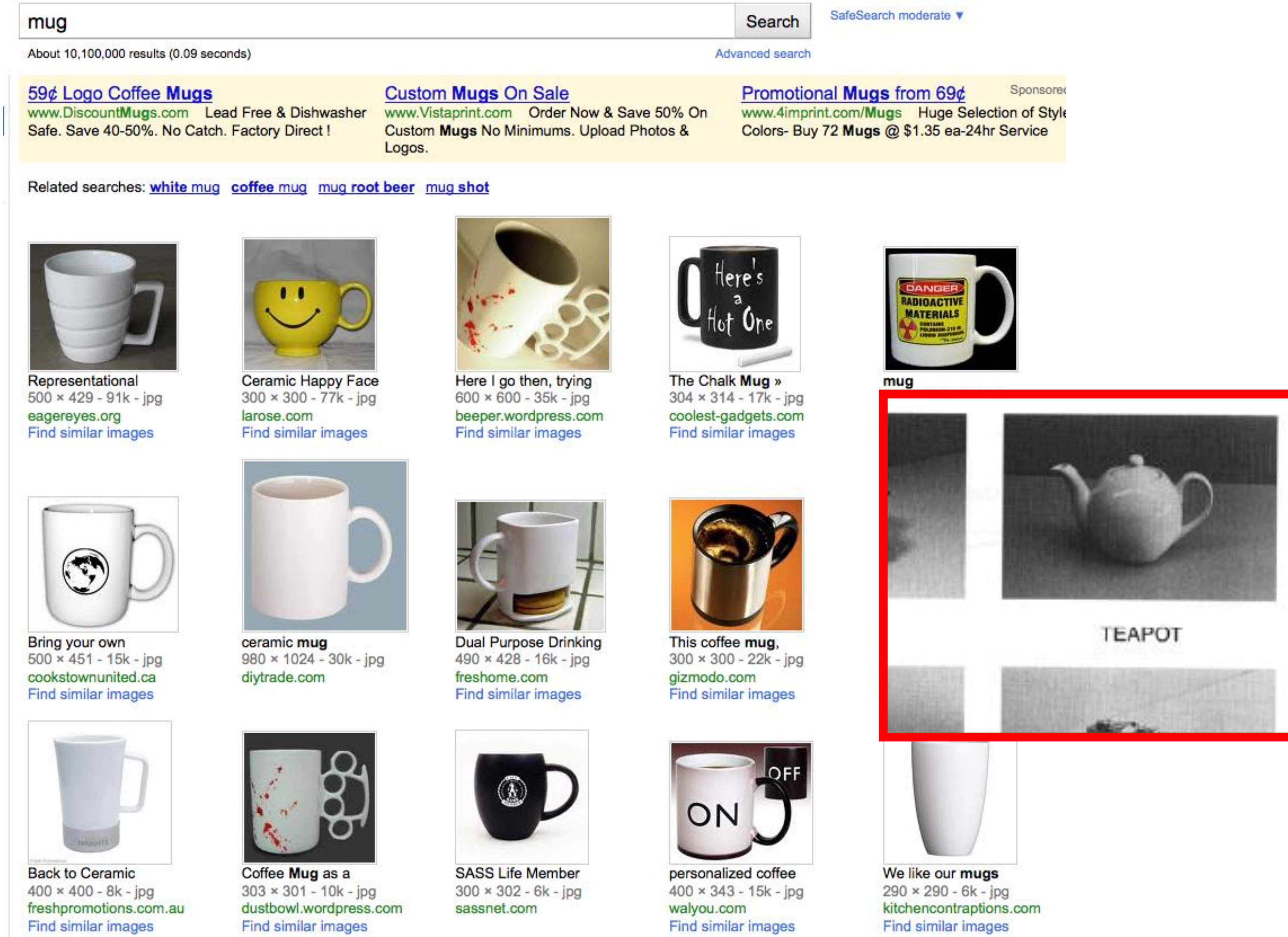


Canonical view of objects



S. Palmer, E. Rosch, and P. Chase. Canonical perspective and the perception of objects. Attention and Performance IX, 1981.

Some bias comes from the way the data is collected



mug

Search

SafeSearch moderate ▼

About 10,100,000 results (0.09 seconds)

Advanced search

59¢ Logo Coffee **Mugs**
www.Discount**Mugs**.com Lead Free & Dishwasher Safe. Save 40-50%. No Catch. Factory Direct !

Custom **Mugs** On Sale
www.Vistaprint.com Order Now & Save 50% On Custom **Mugs** No Minimums. Upload Photos & Logos.

Promotional **Mugs** from 69¢
www.4imprint.com/**Mugs** Huge Selection of Style Colors- Buy 72 **Mugs** @ $1.35 ea-24hr Service

Sponsored

Related searches: **white** mug **coffee** mug mug **root beer** mug **shot**

Representational
500 × 429 - 91k - jpg
eagereyes.org
Find similar images

Ceramic Happy Face
300 × 300 - 77k - jpg
larose.com
Find similar images

Here I go then, trying
600 × 600 - 35k - jpg
beeper.wordpress.com
Find similar images

The Chalk **Mug** »
304 × 314 - 17k - jpg
coolest-gadgets.com
Find similar images

mug

Bring your own
500 × 451 - 15k - jpg
cookstownunited.ca
Find similar images

ceramic **mug**
980 × 1024 - 30k - jpg
diytrade.com

Dual Purpose Drinking
490 × 428 - 16k - jpg
freshome.com
Find similar images

This coffee **mug**,
300 × 300 - 22k - jpg
gizmodo.com
Find similar images

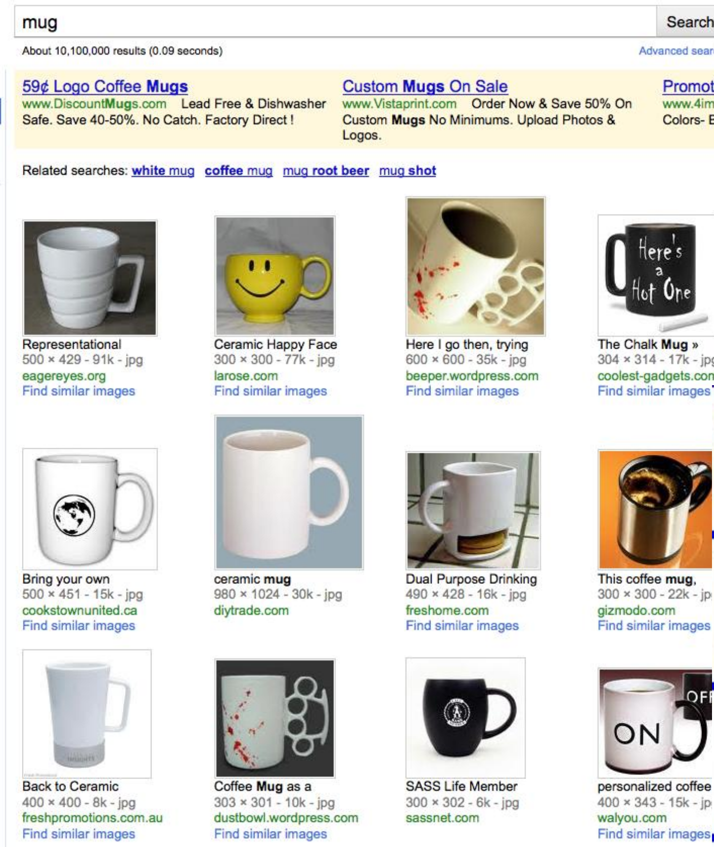
Back to Ceramic
400 × 400 - 8k - jpg
freshpromotions.com.au
Find similar images

Coffee **Mug** as a
303 × 301 - 10k - jpg
dustbowl.wordpress.com
Find similar images

SASS Life Member
300 × 302 - 6k - jpg
sassnet.com

personalized coffee
400 × 343 - 15k - jpg
walyou.com
Find similar images

We like our **mugs**
290 × 290 - 6k - jpg
kitchencontraptions.com
Find similar images



mug Search

SafeSearch moderate ▾

About 10,100,000 results (0.09 seconds)

Advanced search

Sponsored

[59¢ Logo Coffee Mugs](www.DiscountMugs.com)
www.DiscountMugs.com Lead Free & Dishwasher Safe. Save 40-50%. No Catch. Factory Direct !

Custom Mugs On Sale
www.Vistaprint.com Order Now & Save 50% On Custom **Mugs** No Minimums. Upload Photos & Logos.

Promotional Mugs from 69¢
www.4imprint.com/Mugs Huge Selection of Style Colors- Buy 72 **Mugs** @ $1.35 ea-24hr Service

Related searches: **white** mug **coffee** mug mug **root beer** mug **shot**

Representational
500 × 429 - 91k - jpg
eagereyes.org
Find similar images

Ceramic Happy Face
300 × 300 - 77k - jpg
larose.com
Find similar images

Here I go then, trying
600 × 600 - 35k - jpg
beeper.wordpress.com
Find similar images

The Chalk **Mug** »
304 × 314 - 17k - jpg
coolest-gadgets.com
Find similar images

mug
300 × 279 - 54k - jpg
reynosawatch.org

Bring your own
500 × 451 - 15k - jpg
cookstownunited.ca
Find similar images

ceramic **mug**
980 × 1024 - 30k - jpg
diytrade.com

Dual Purpose Drinking
490 × 428 - 16k - jpg
freshome.com
Find similar images

This coffee **mug**,
300 × 300 - 22k - jp
gizmodo.com
Find similar images

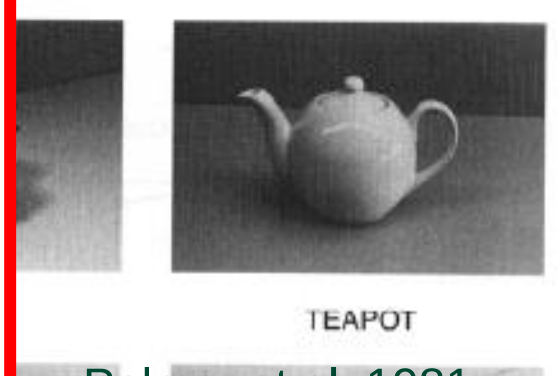
Back to Ceramic
400 × 400 - 8k - jpg
freshpromotions.com.au
Find similar images

Coffee **Mug** as a
303 × 301 - 10k - jpg
dustbowl.wordpress.com
Find similar images

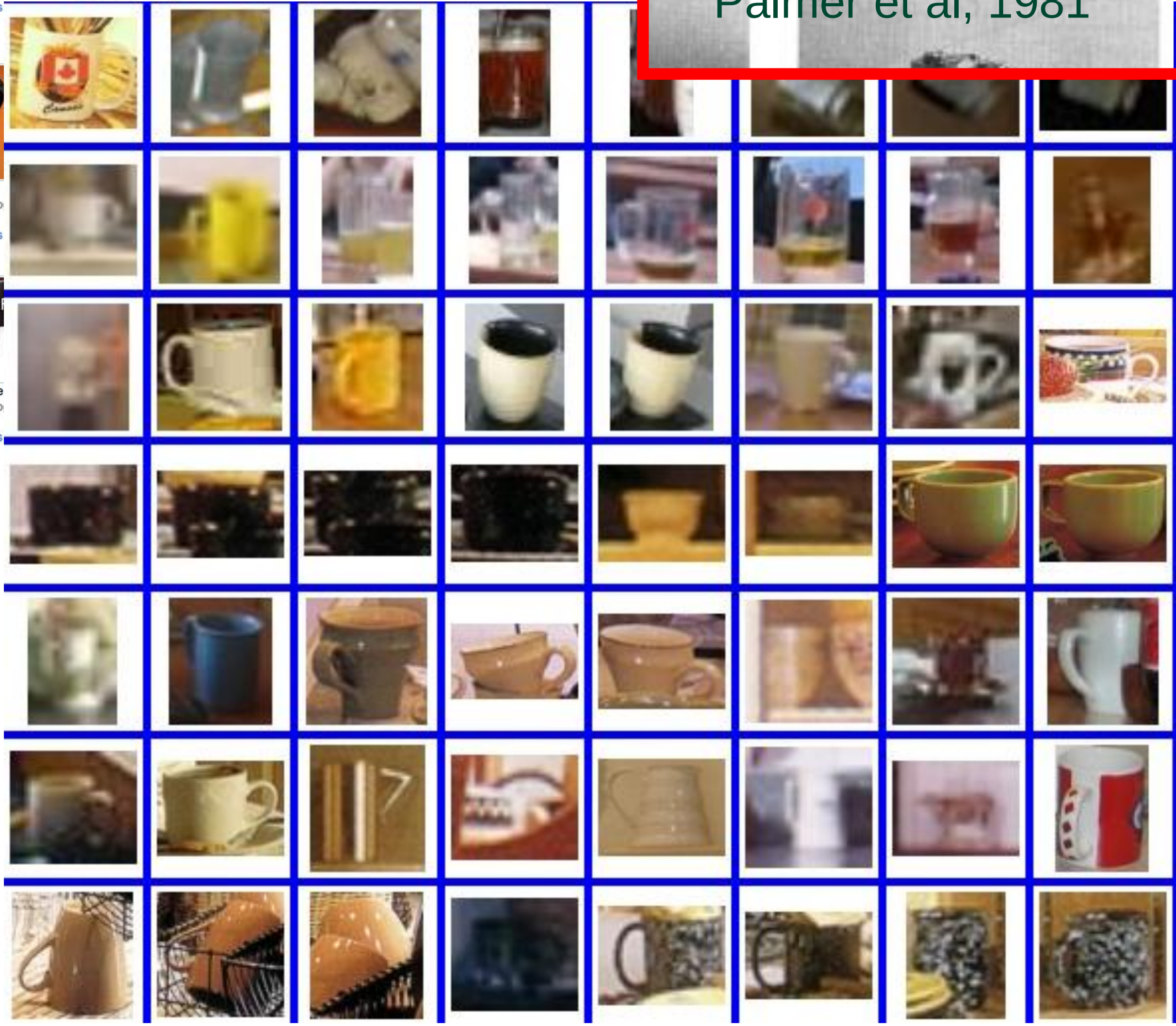
SASS Life Member
300 × 302 - 6k - jpg
sassnet.com

personalized coffee
400 × 343 - 15k - jp
walyou.com
Find similar images

Google mugs



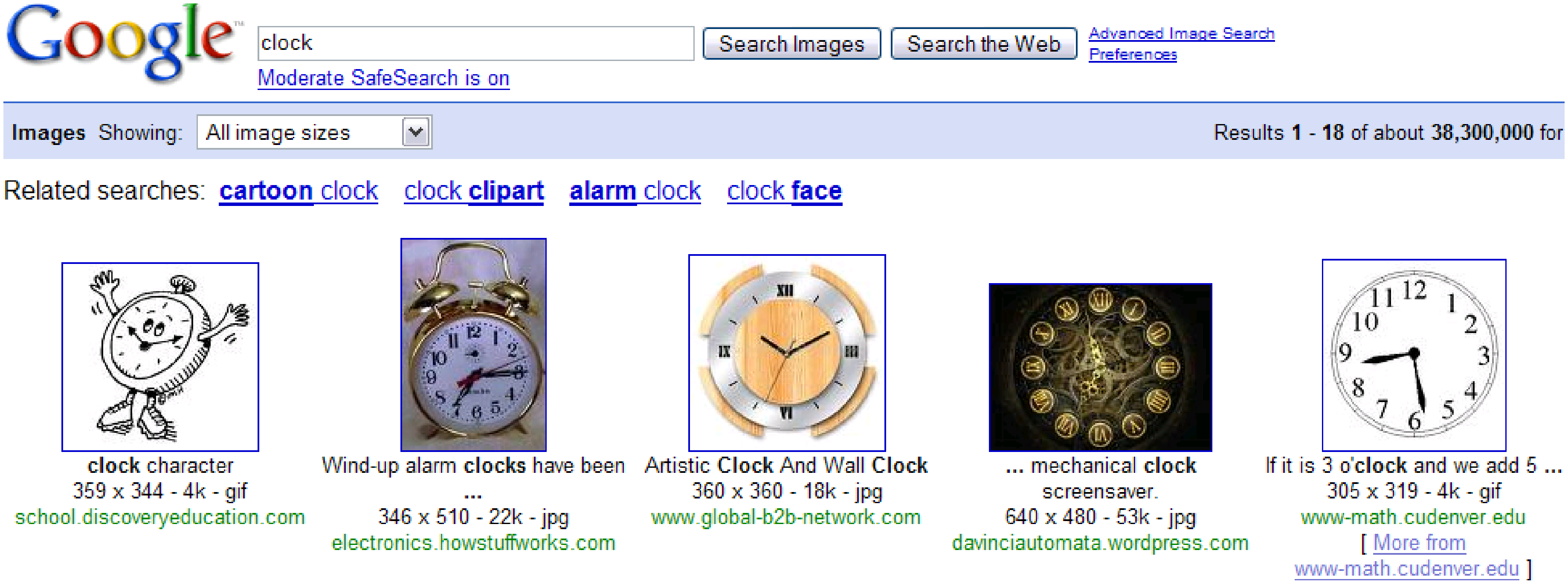
Palmer et al, 1981



Mugs from LabelMe

Palmer et al, 1981

CLOCK



clock

Search Images | Search the Web | Advanced Image Search | Preferences

Moderate SafeSearch is on

Images Showing: All image sizes

Results **1** - **18** of about **38,300,000** for

Related searches: **cartoon** clock clock **clipart** **alarm** clock clock **face**

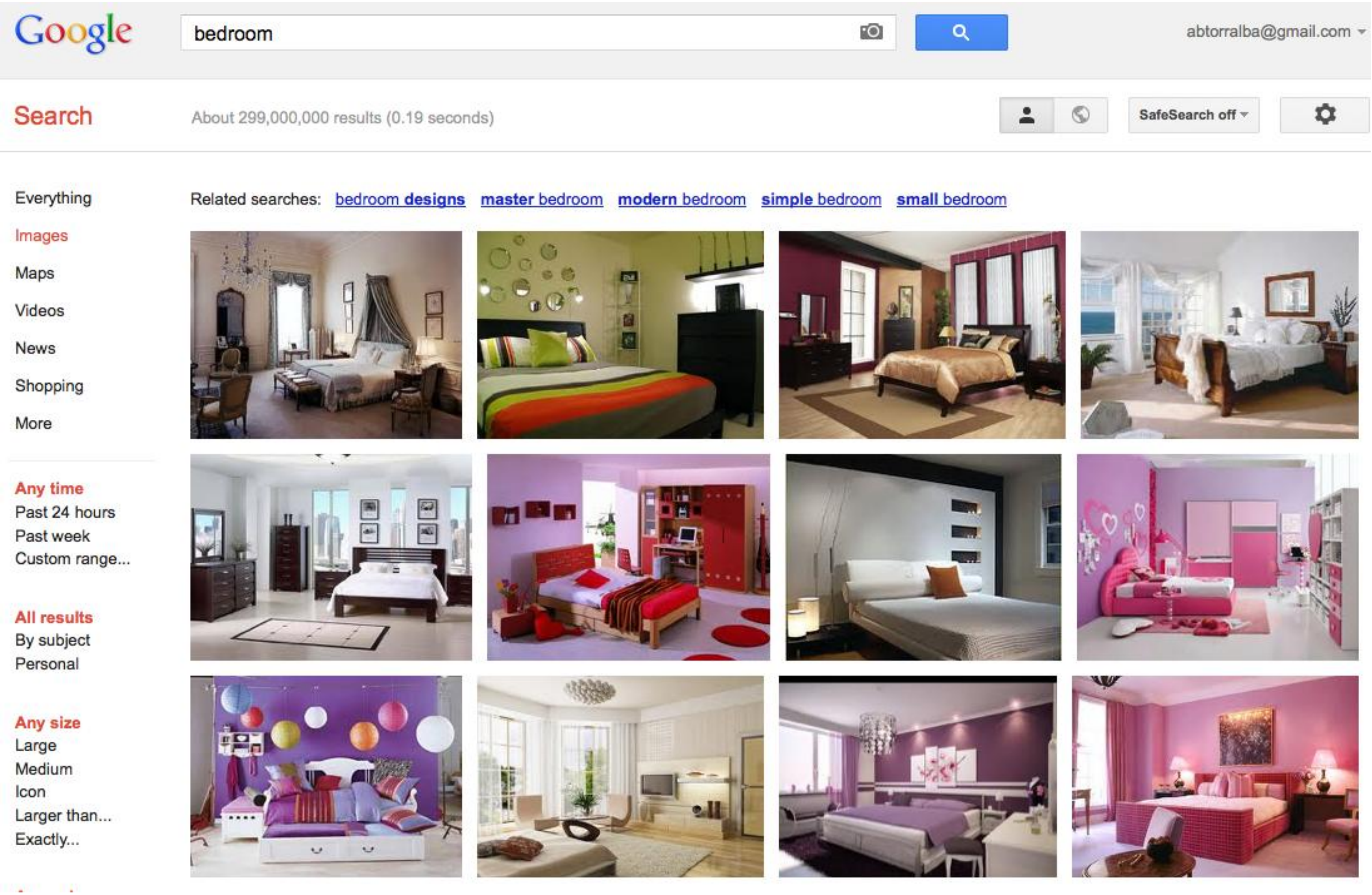
clock character
359 x 344 - 4k - gif
school.discoveryeducation.com

Wind-up alarm **clocks** have been ...
346 x 510 - 22k - jpg
electronics.howstuffworks.com

Artistic **Clock** And Wall **Clock**
360 x 360 - 18k - jpg
www.global-b2b-network.com

... mechanical **clock** screensaver.
640 x 480 - 53k - jpg
davinciautomata.wordpress.com

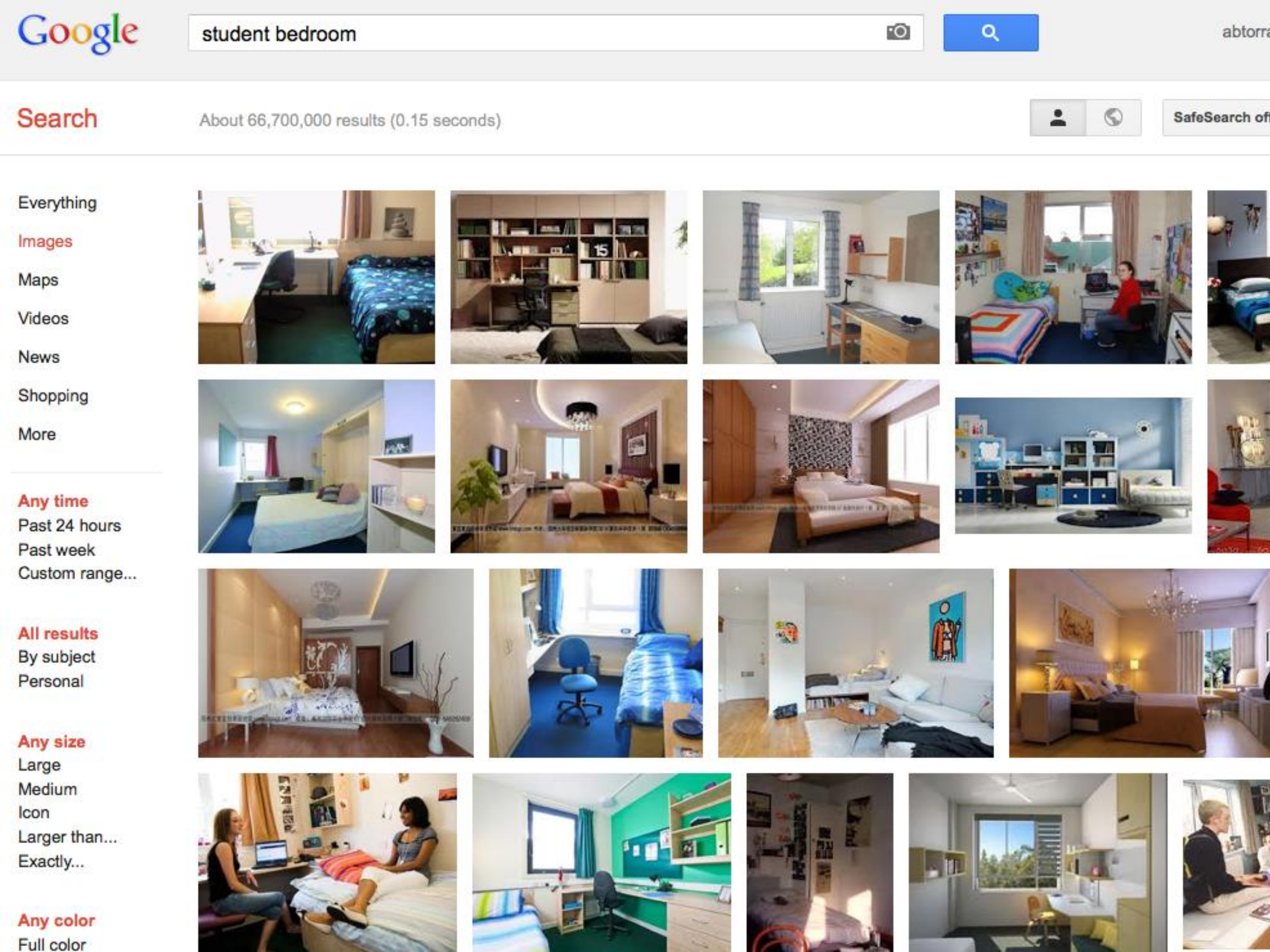
If it is 3 o'**clock** and we add 5 ...
305 x 319 - 4k - gif
www-math.cudenver.edu
[More from www-math.cudenver.edu]



bedroom

About 299,000,000 results (0.19 seconds)

Related searches: bedroom **designs** **master** bedroom **modern** bedroom **simple** bedroom **small** bedroom

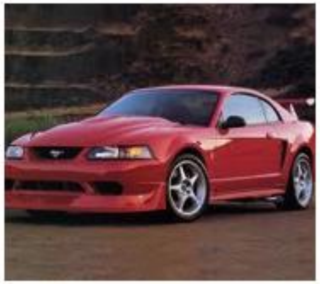


student bedroom

About 66,700,000 results (0.15 seconds)

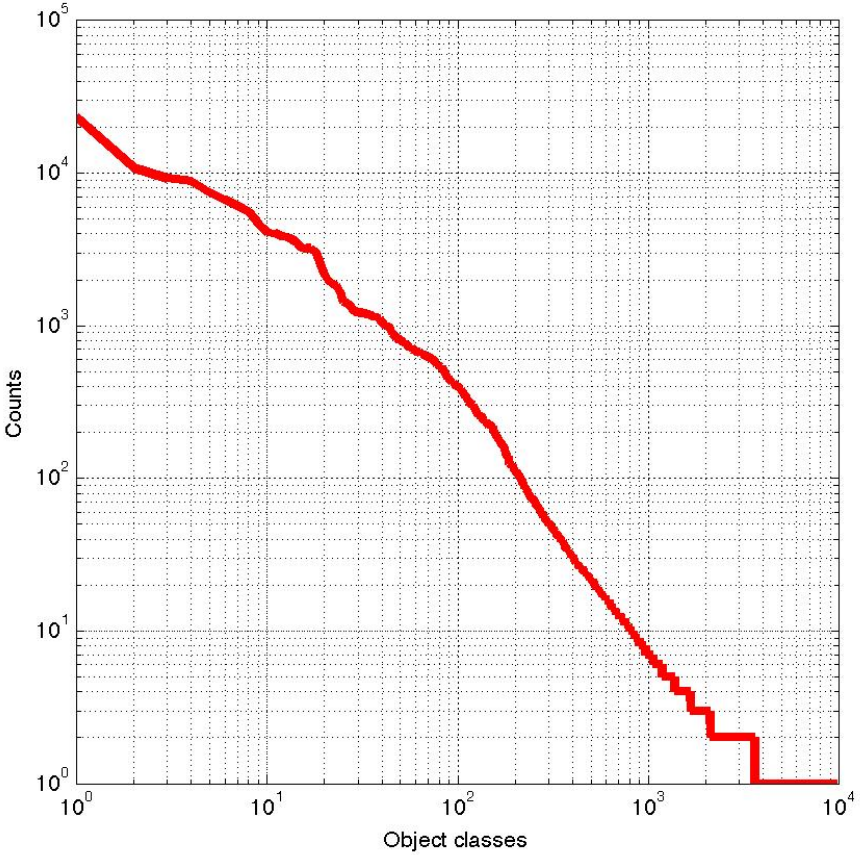
www.bigstock.com · 7067629

The world is biased

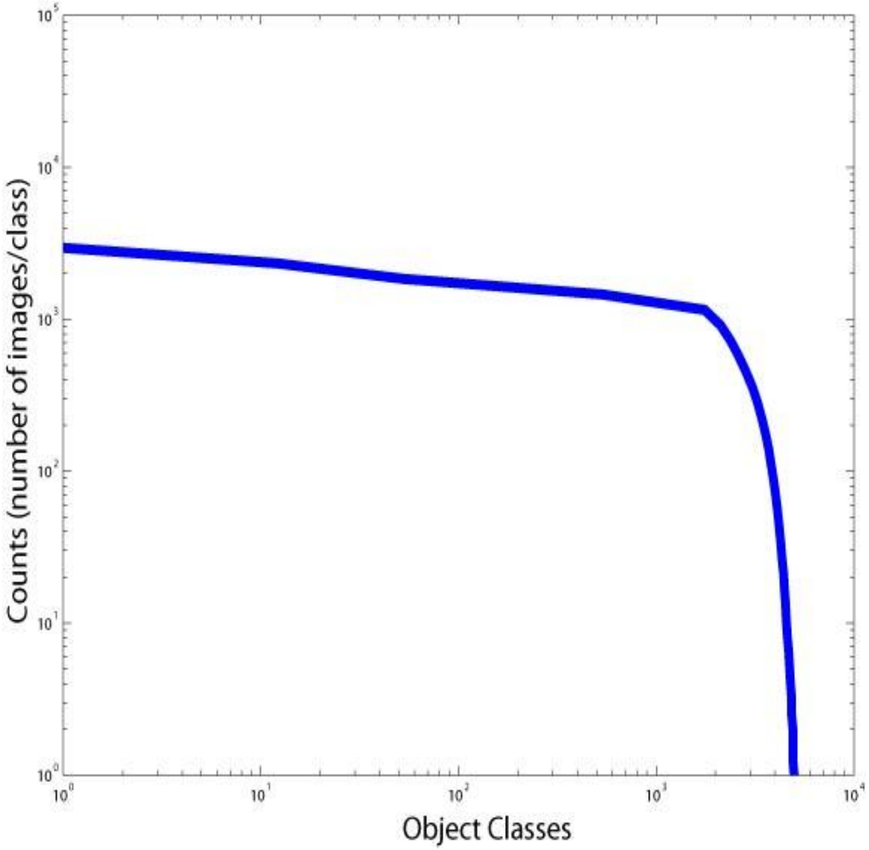


Hyundai NEOS

Distribution bias



LabelMe



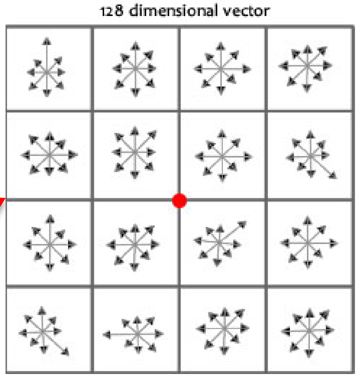
IMAGENET

Feature bias

Researchers meet

Delay
1 year

Descriptor



Delay
1 year

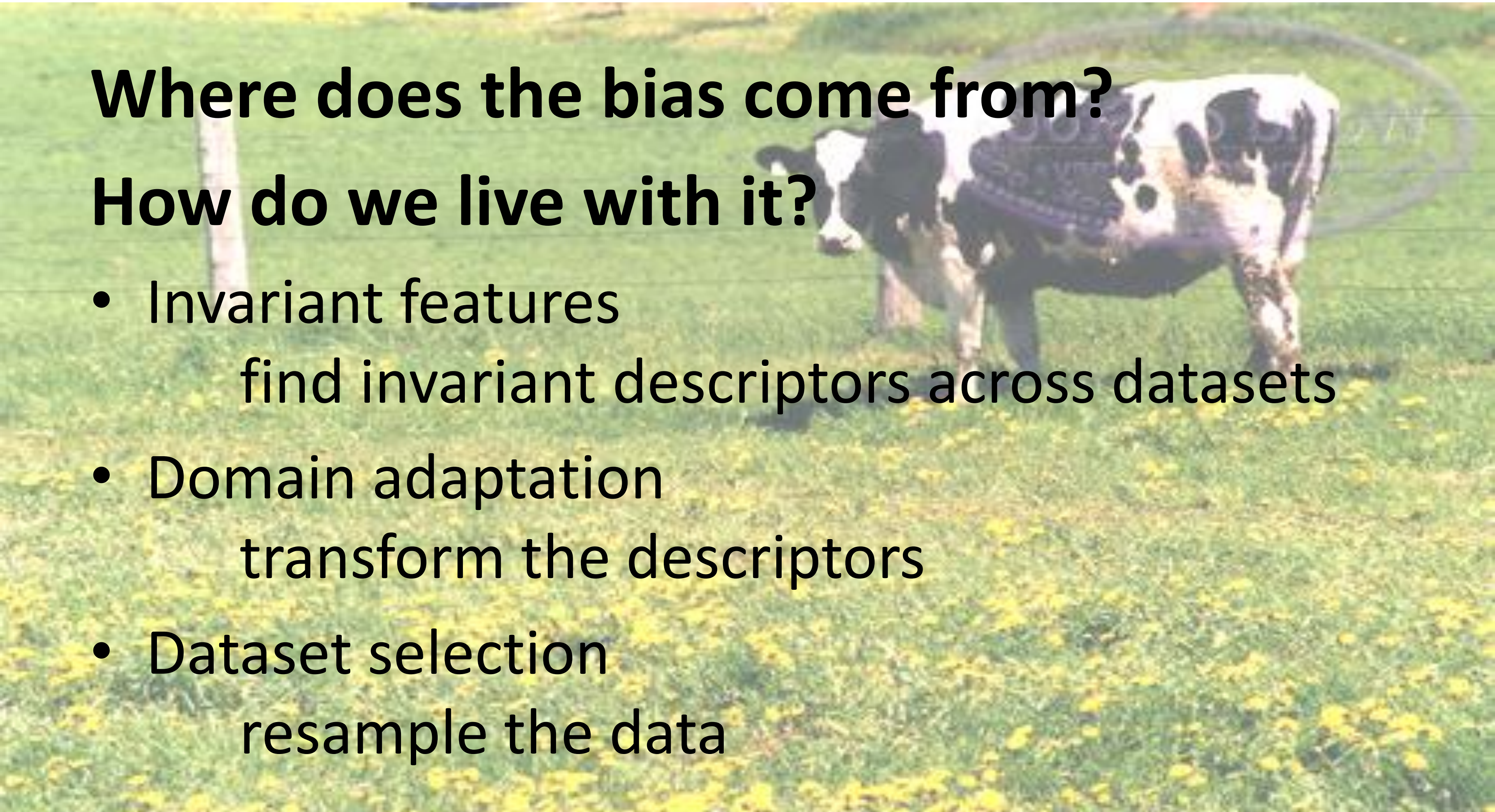
Researchers meet

Delay
1 year

Images /
Benchmarks

Delay
1 year

A green pasture for research: “Understanding and Living with dataset bias”



Where does the bias come from?

How do we live with it?

- Invariant features

  find invariant descriptors across datasets

- Domain adaptation

  transform the descriptors

- Dataset selection

  resample the data

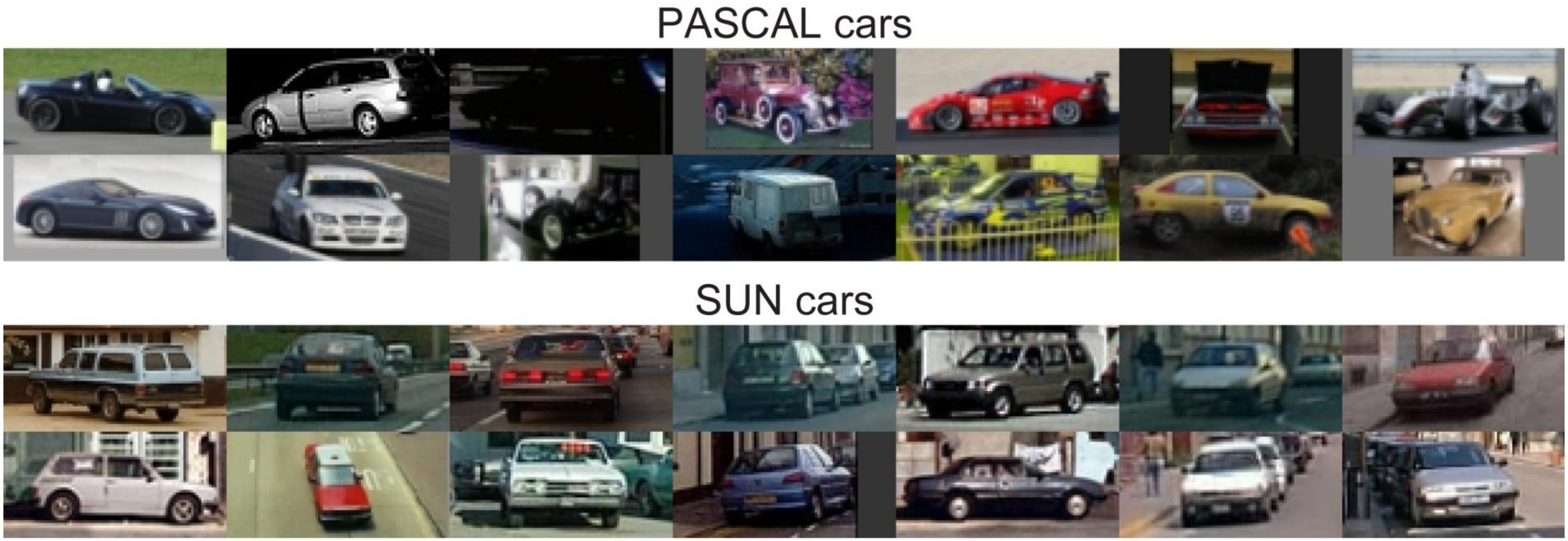
Duan, Tsang, Xu, Maybank. Domain transfer svm for video concept detection. CVPR. (2009)

Saenko, Kulis, Fritz, Darrell. Adapting Visual Category Models to New Domains. ECCV 2010

Gopalan, Li, and Chellappa. Domain Adaptation for Object Recognition: An Unsupervised Approach. ICCV 2011

Boqing Gong, Yuan Shi, Fei Sha. Geodesic Flow Kernel for Unsupervised Domain Adaptation. CVPR 2012.

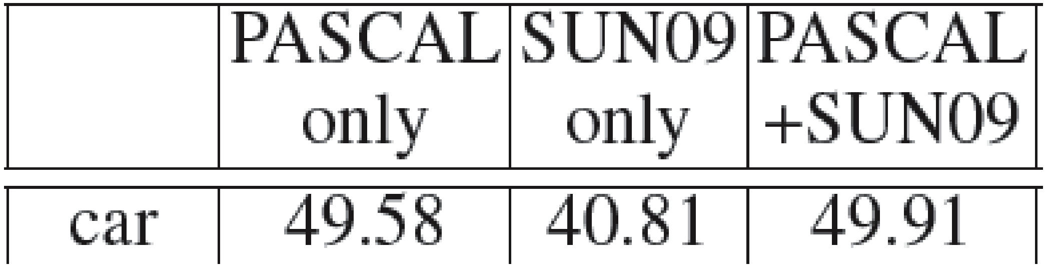
Mixing datasets



PASCAL cars

SUN cars

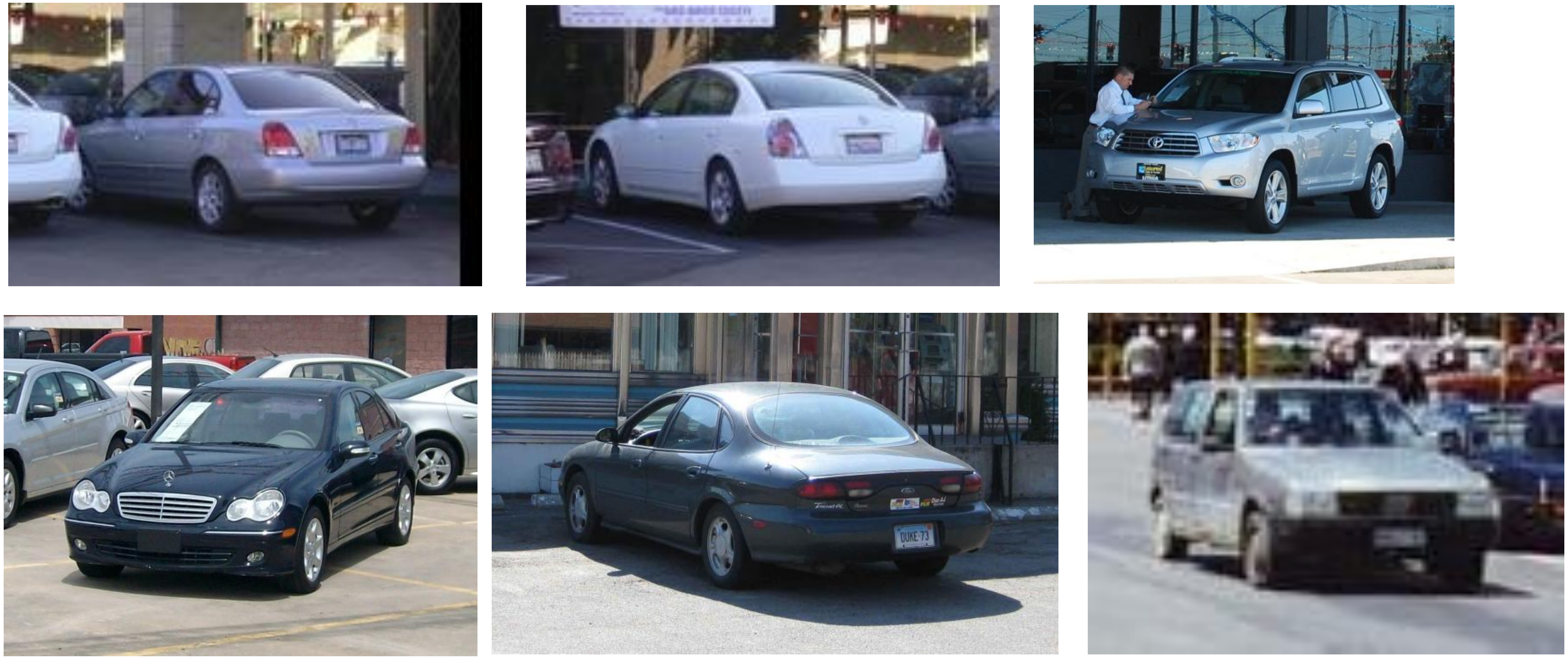
If we test on PASCAL and we train with:



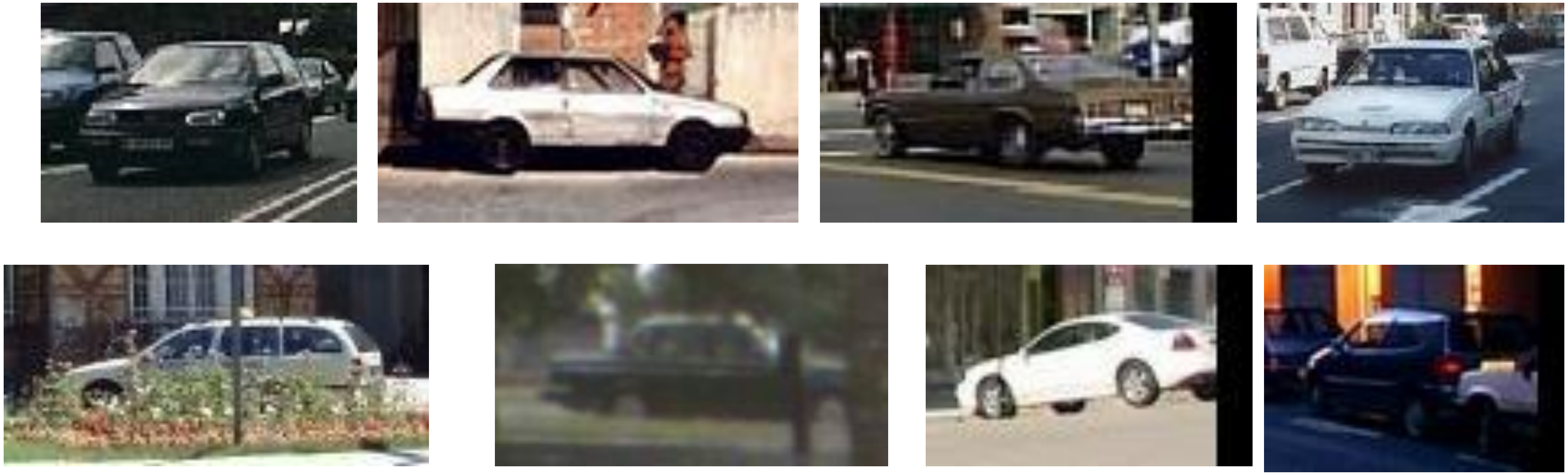
	PASCAL only	SUN09 only	PASCAL +SUN09
car	49.58	40.81	49.91

Transfer Learning by Borrowing Examples for Multiclass Object Detection
J. J. Lim, R. Salakhutdinov, A. Torralba. NIPS, 2011.

Car examples from SUN database

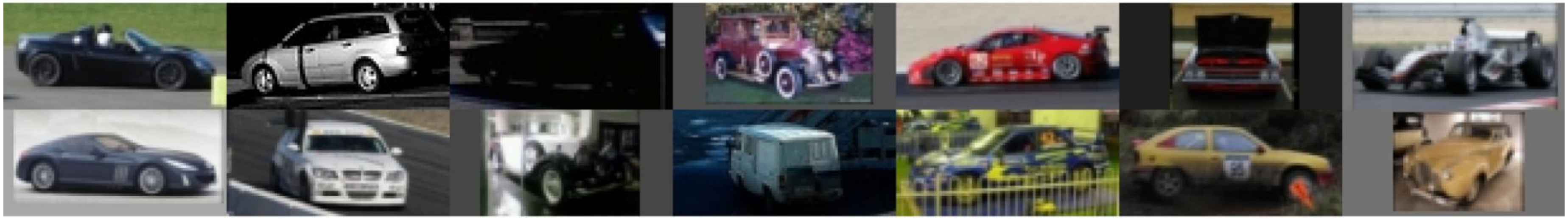


…

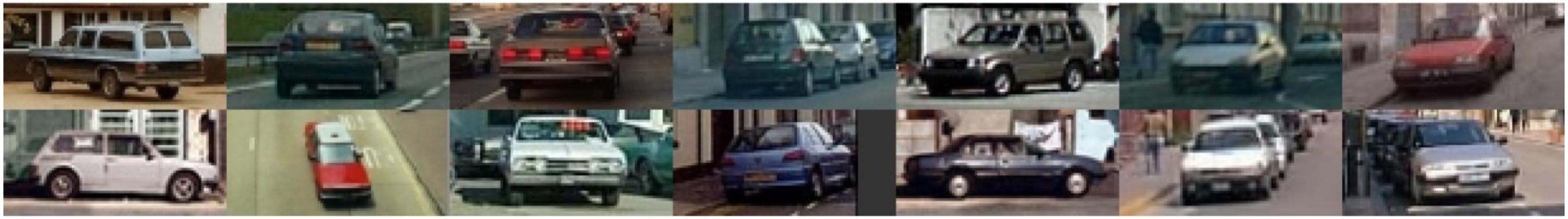


Mixing datasets

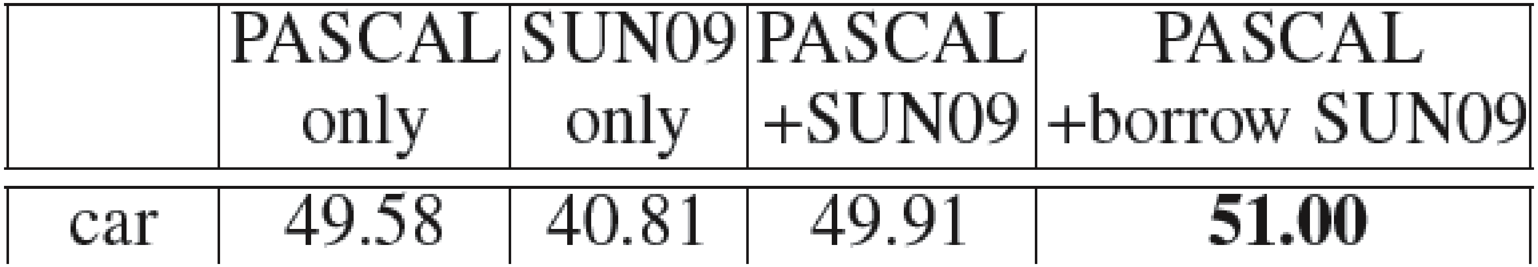
PASCAL cars



SUN cars



If we test on PASCAL and we train with:

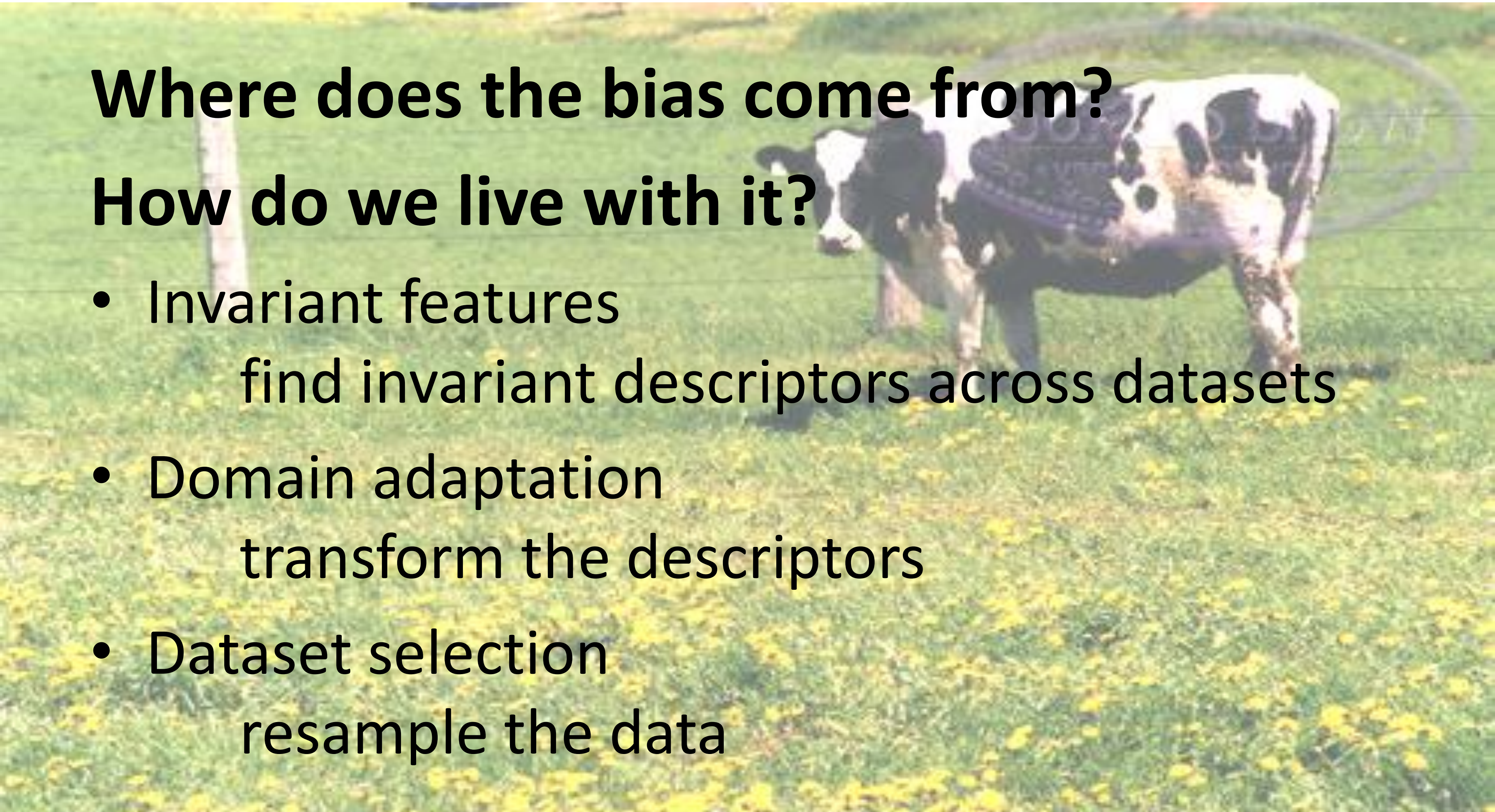


	PASCAL only	SUN09 only	PASCAL +SUN09	PASCAL +borrow SUN09
car	49.58	40.81	49.91	**51.00**

Less is more if we take the *good* data

Transfer Learning by Borrowing Examples for Multiclass Object Detection
J. J. Lim, R. Salakhutdinov, A. Torralba. NIPS, 2011.

A green pasture for research: “Understanding and Living with dataset bias”

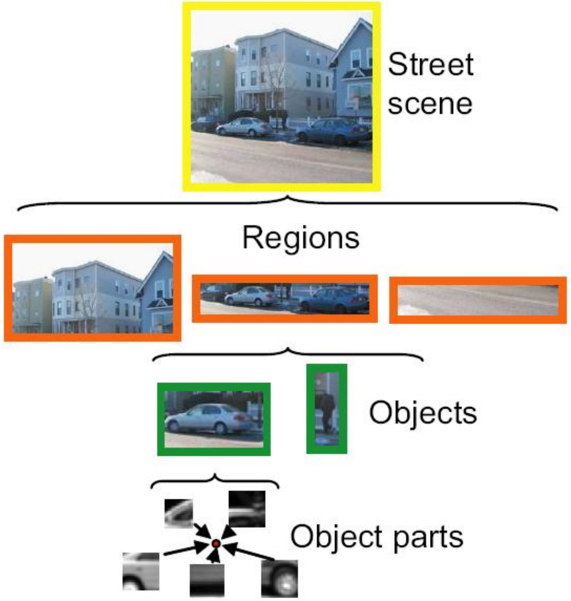


Where does the bias come from?

How do we live with it?

- Invariant features
  find invariant descriptors across datasets
- Domain adaptation
  transform the descriptors
- Dataset selection
  resample the data

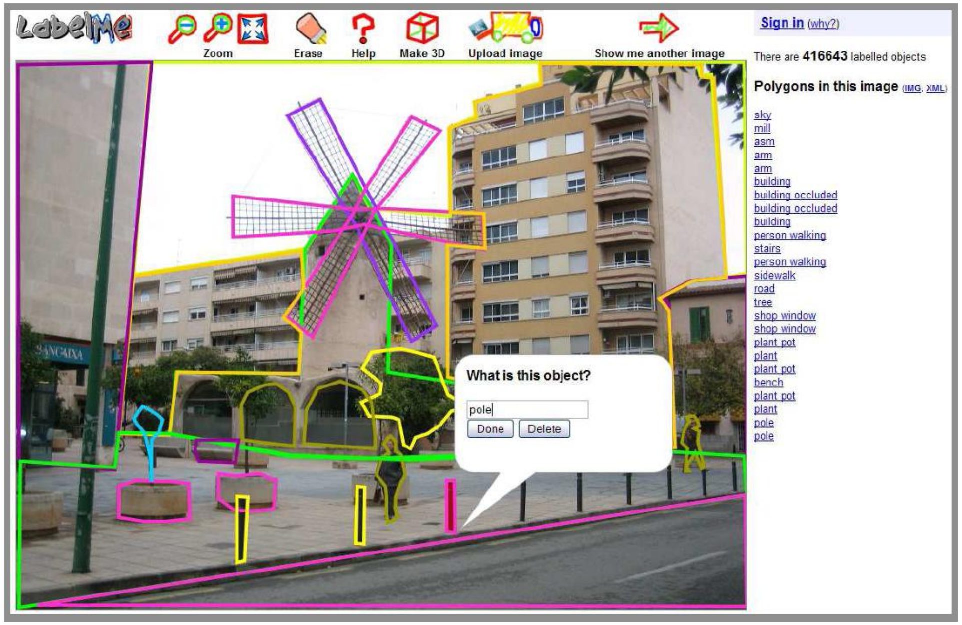
Discussion


Street scene

Regions

Objects

Object parts



Sign in (why?)

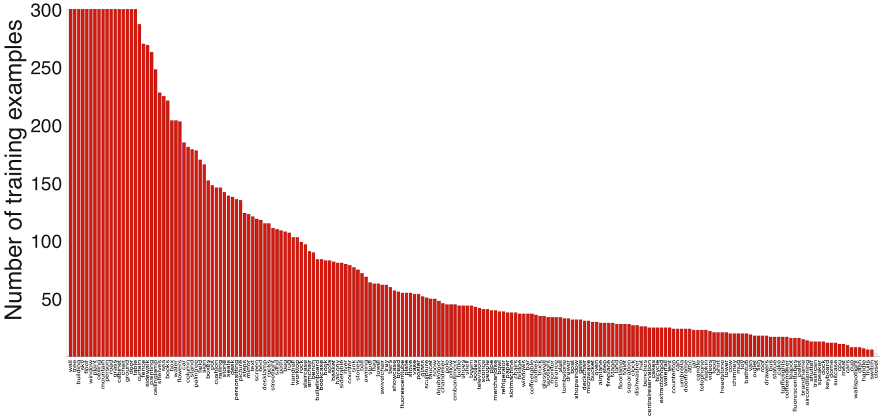
There are 416643 labelled objects

Polygons in this image (IMG, XML)

sky
mill
asm
arm
arm
building
building occluded
building occluded
building
person walking
stairs
person walking
sidewalk
road
tree
shop window
shop window
plant pot
plant
plant pot
bench
plant pot
plant
pole
pole

What is this object?

pole



This work is partially funded by NSF Career 0747120

Dataset bias

Power law: the two extremes of learning coexist

For lots of data: sift flow
-Reduce context and describe non-parametric context.

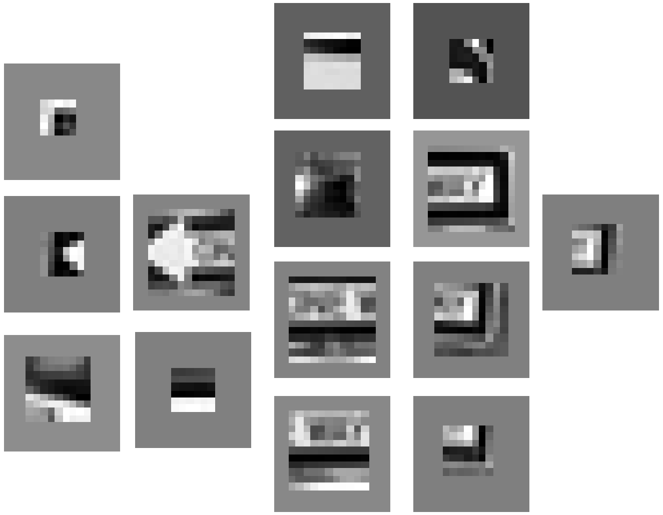
Out of context test to decide what is missing on a context model (slides from cifar)

Some symptoms of one-vs-all multiclass approaches

What is the best representation to detect a traffic sign?

Very regular object: template matching will do the job

Parts derived from training a binary classifier.



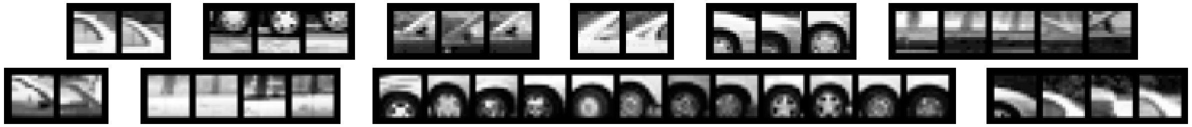
~100% detection rate with 0 false alarms

Some of these parts cannot be used for anything else than this object.

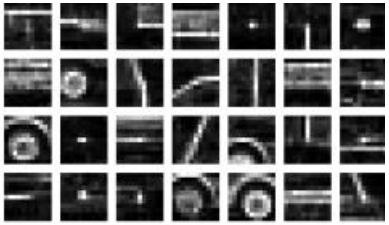
Some symptoms of one-vs-all multiclass approaches

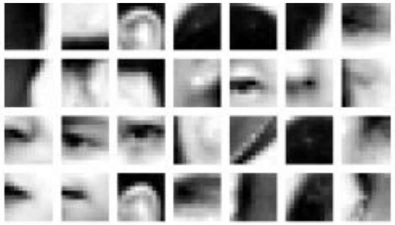
Part-based object representation (looking for meaningful parts):

- A. Agarwal and D. Roth



- M. Weber, M. Welling and P. Perona



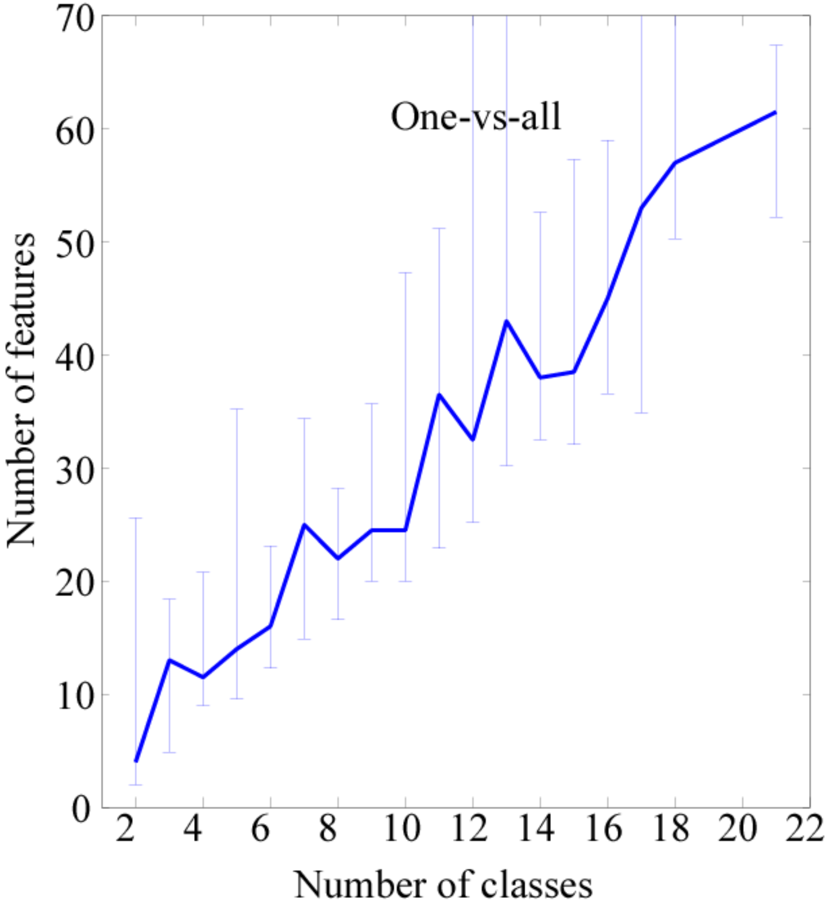


…

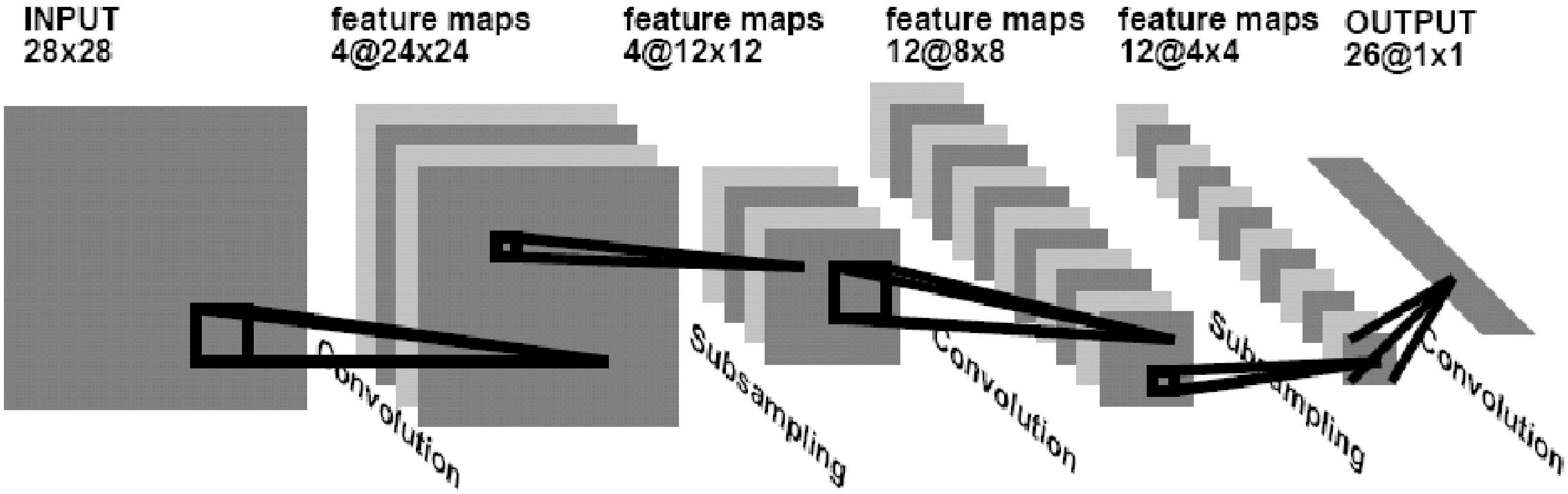
These studies try to recover parts that are meaningful. But is this the right thing to do? The derived parts may be too specific, and they are not likely to be useful in a general system.

Some symptoms of one-vs-all multiclass approaches

Computational cost grows linearly with Nclasses * Nviews * Nstyles …



Convolutional Neural Network



Le Cun et al, 98

Translation invariance is already built into the network

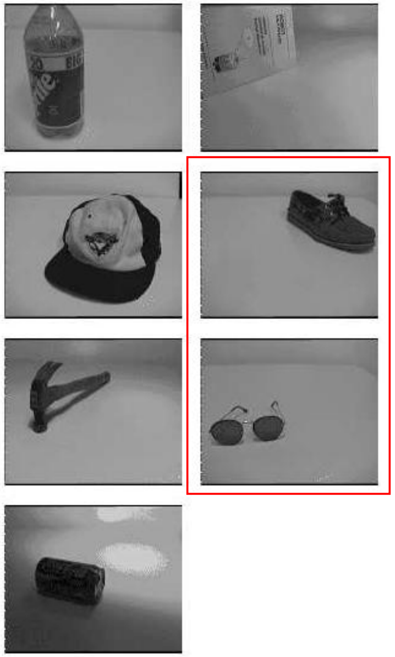
The output neurons share all the intermediate levels

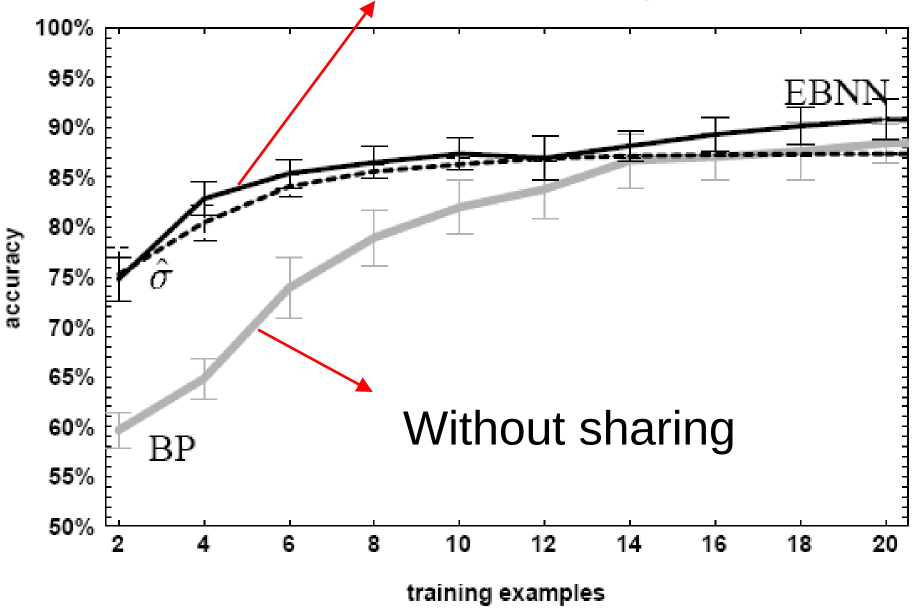
Sharing invariances

S. Thrun. Is Learning the n-th Thing Any Easier Than Learning The First? NIPS 1996

Knowledge is transferred between tasks via a learned model of the invariances of the domain: object recognition is invariant to rotation, translation, scaling, lighting, … These invariances are common to all object recognition tasks.

Toy world





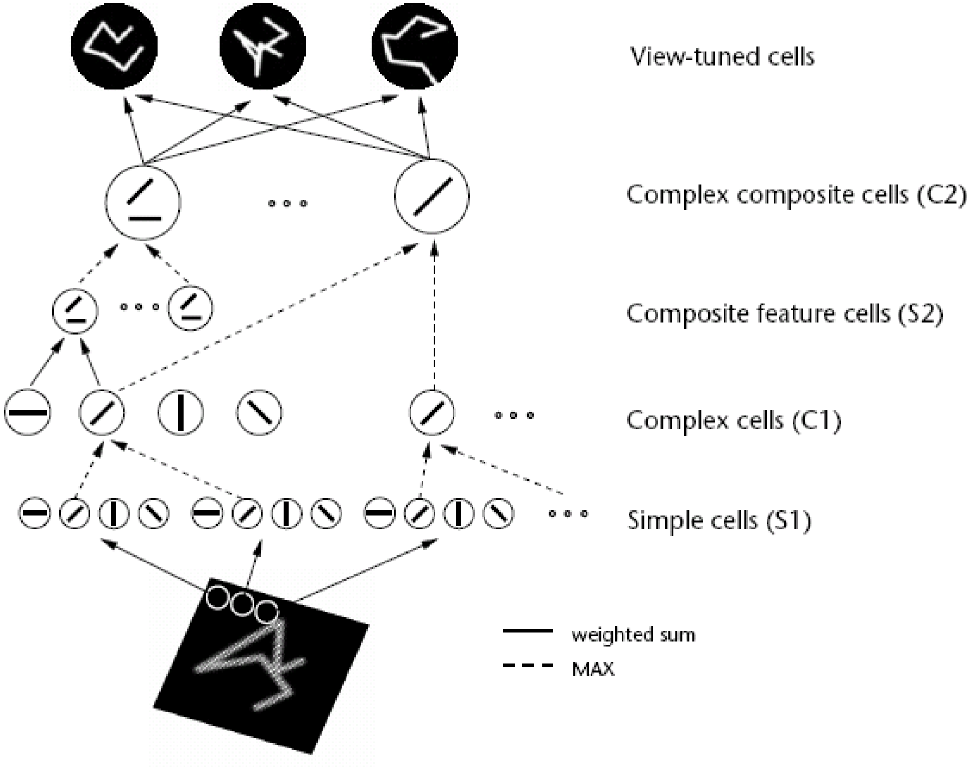
With sharing

Without sharing

Models of object recognition

I. Biederman, “Recognition-by-components: A theory of human image understanding,” *Psychological Review*, 1987.

M. Riesenhuber and T. Poggio, “Hierarchical models of object recognition in cortex,” *Nature Neuroscience* 1999.

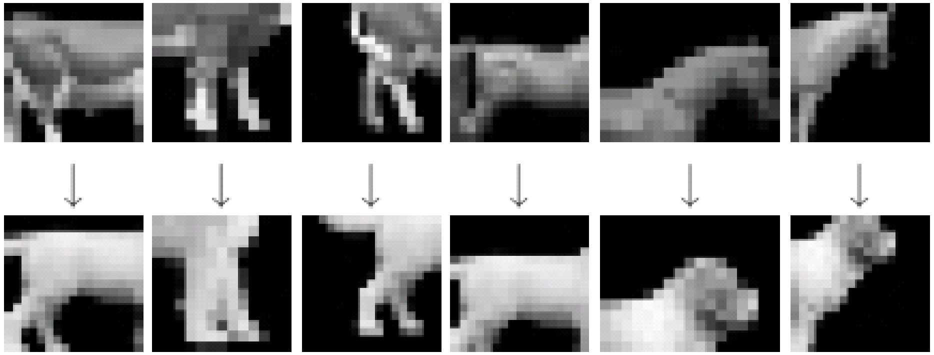


T. Serre, L. Wolf and T. Poggio. “Object recognition with features inspired by visual cortex”. CVPR 2005

Sharing patches

- Bart and Ullman, 2004

For a new class, use only features similar to features that where good for other classes:

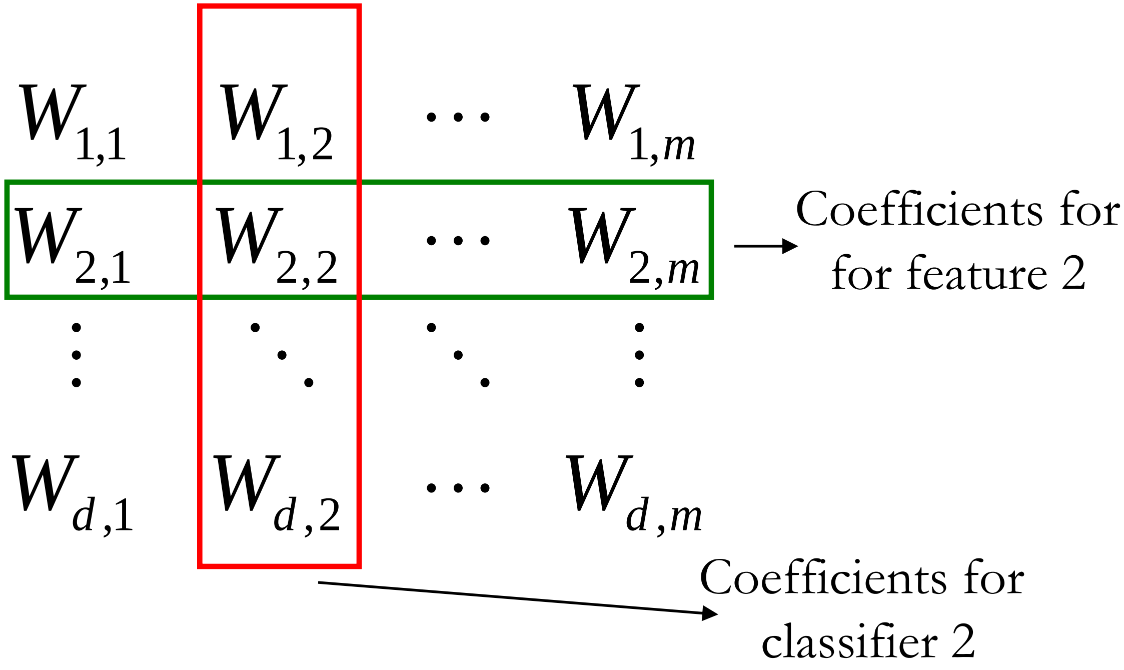


Proposed Dog features →

Figure 1. Feature adaptation. (a) Top row: features extracted from multiple images of cows (first three) and horses (last three), as described in section 3.1. Bottom row: features adapted to the dogs class by the proposed cross-generalization algorithm (section 3.2), using a single dog image.

Transfer Learning for Image Classification with Sparse Prototype Representations

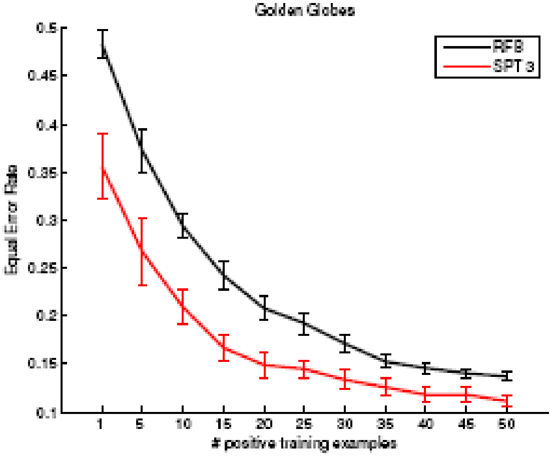
A. Quattoni, M. Collins, T. Darrell, CVPR 2008



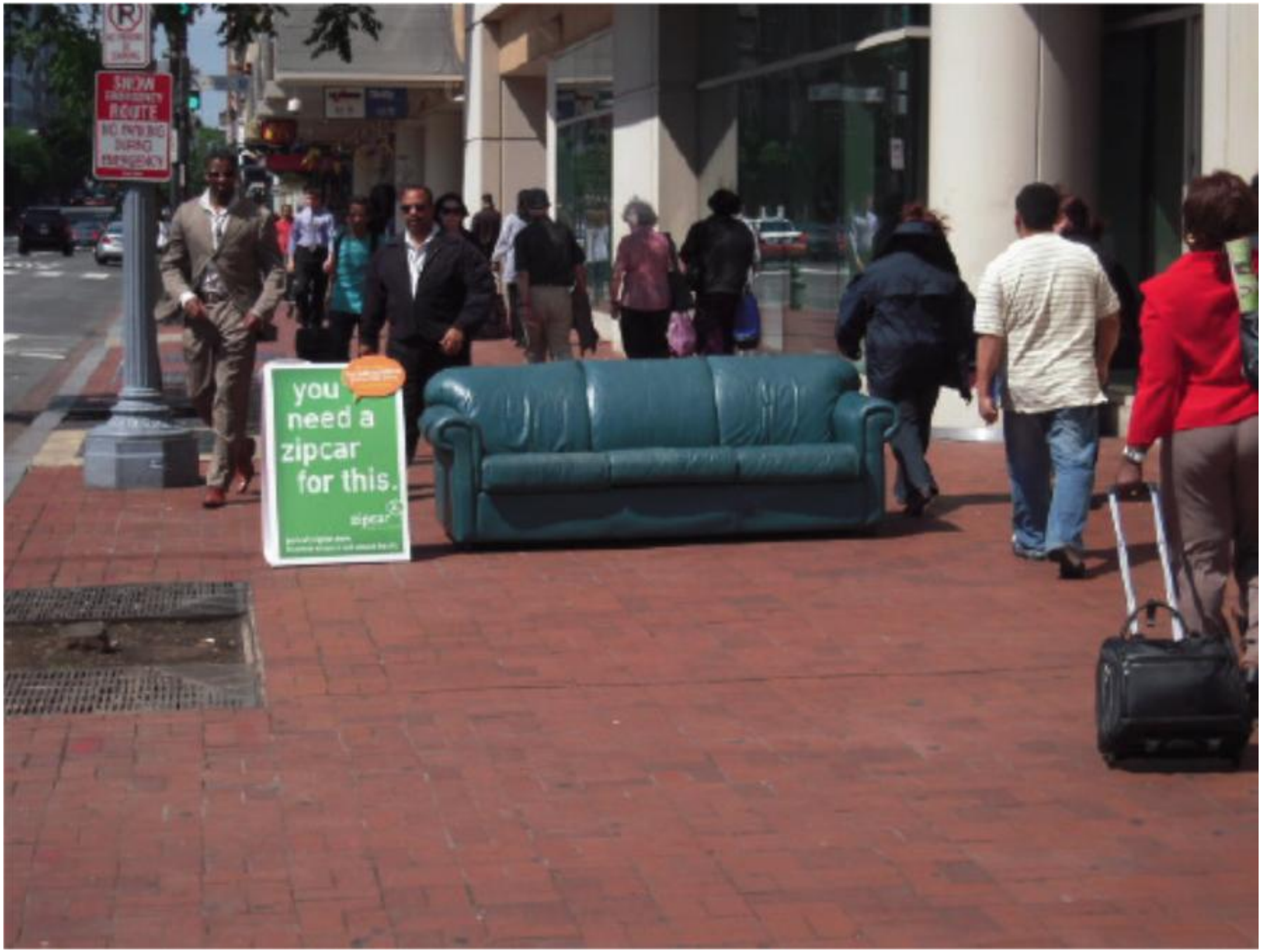
$$\begin{matrix} W_{1,1} & W_{1,2} & \cdots & W_{1,m} \\ W_{2,1} & W_{2,2} & \cdots & W_{2,m} \\ \vdots & \ddots & \ddots & \vdots \\ W_{d,1} & W_{d,2} & \cdots & W_{d,m} \end{matrix}$$

Coefficients for for feature 2

Coefficients for classifier 2

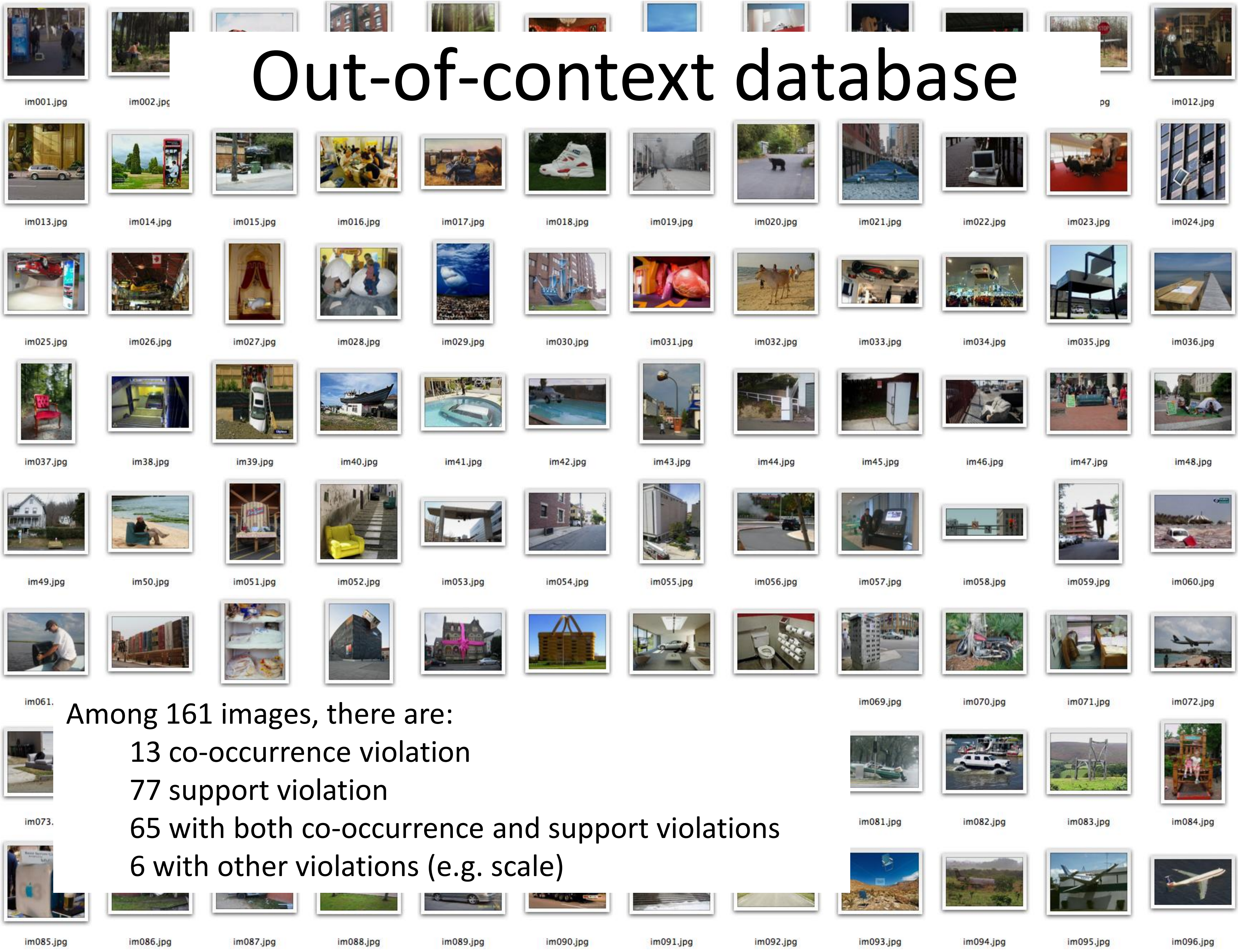


$$\min_{\mathbf{W}} \sum_{k=1}^{m} \frac{1}{|D_k|} \sum_{(x,y)\in D_k} l\,(f_k(x), y) + C \sum_{i=1}^{d} \max_k (|W_{ik}|)$$


SNOW
EMERGENCY
ROUTE
NO PARKING
DURING
EMERGENCY
you need a zipcar for this.
zipcar

showcase[0.82]
curtain[0.81]
curtain[0.86]
curtain[0.90]
person[0.99]
person[0.96]
person[0.91]
platform[0.83]
gate[0.80]
clothes[0.82]
pillow[0.96]
armchair[0.51]
floor[0.90]
grass[0.86]
car[0.93]
stand[0.84]
chair[0.82]
car[0.83]
car[0.92]
car[0.93]

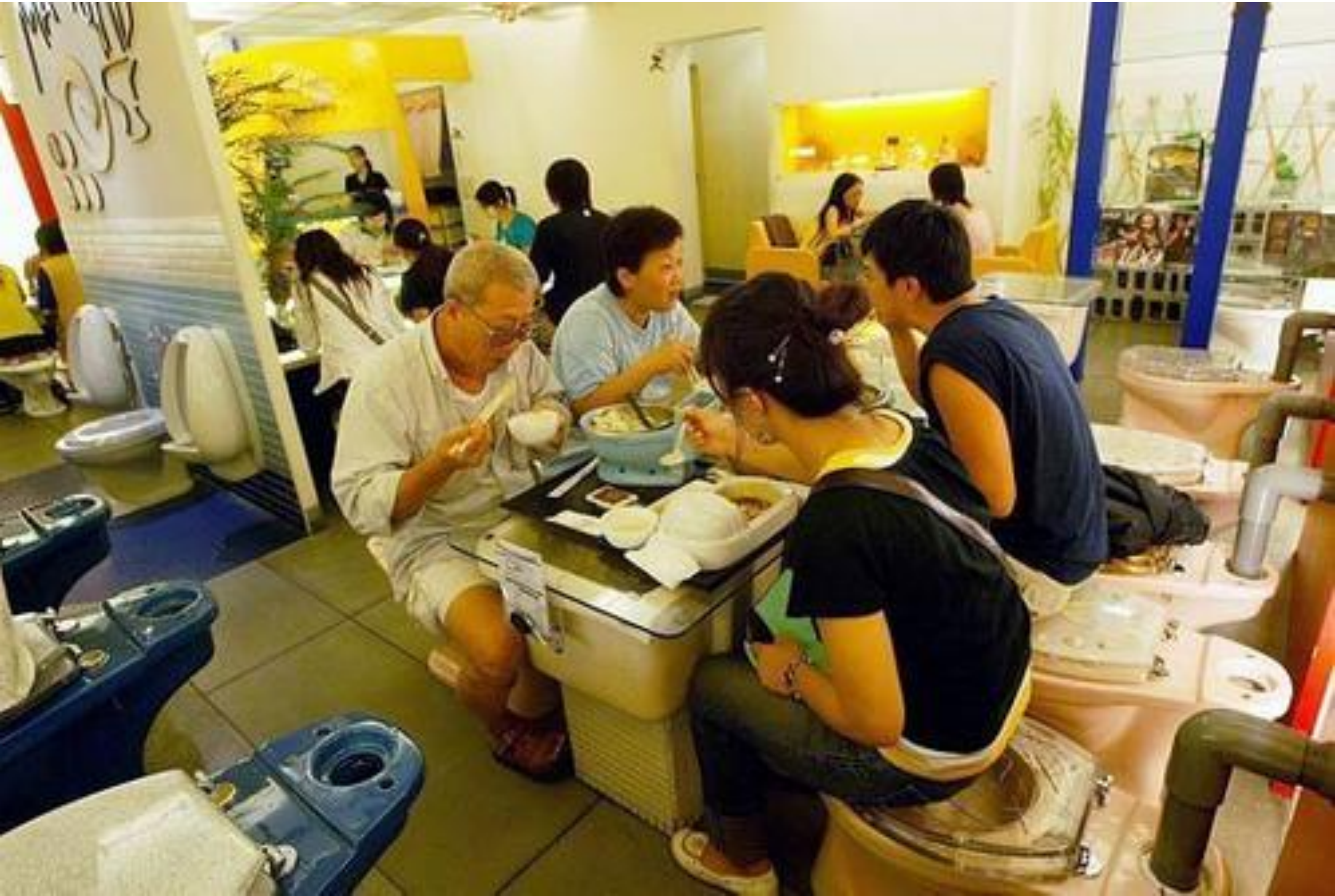
Out-of-context database



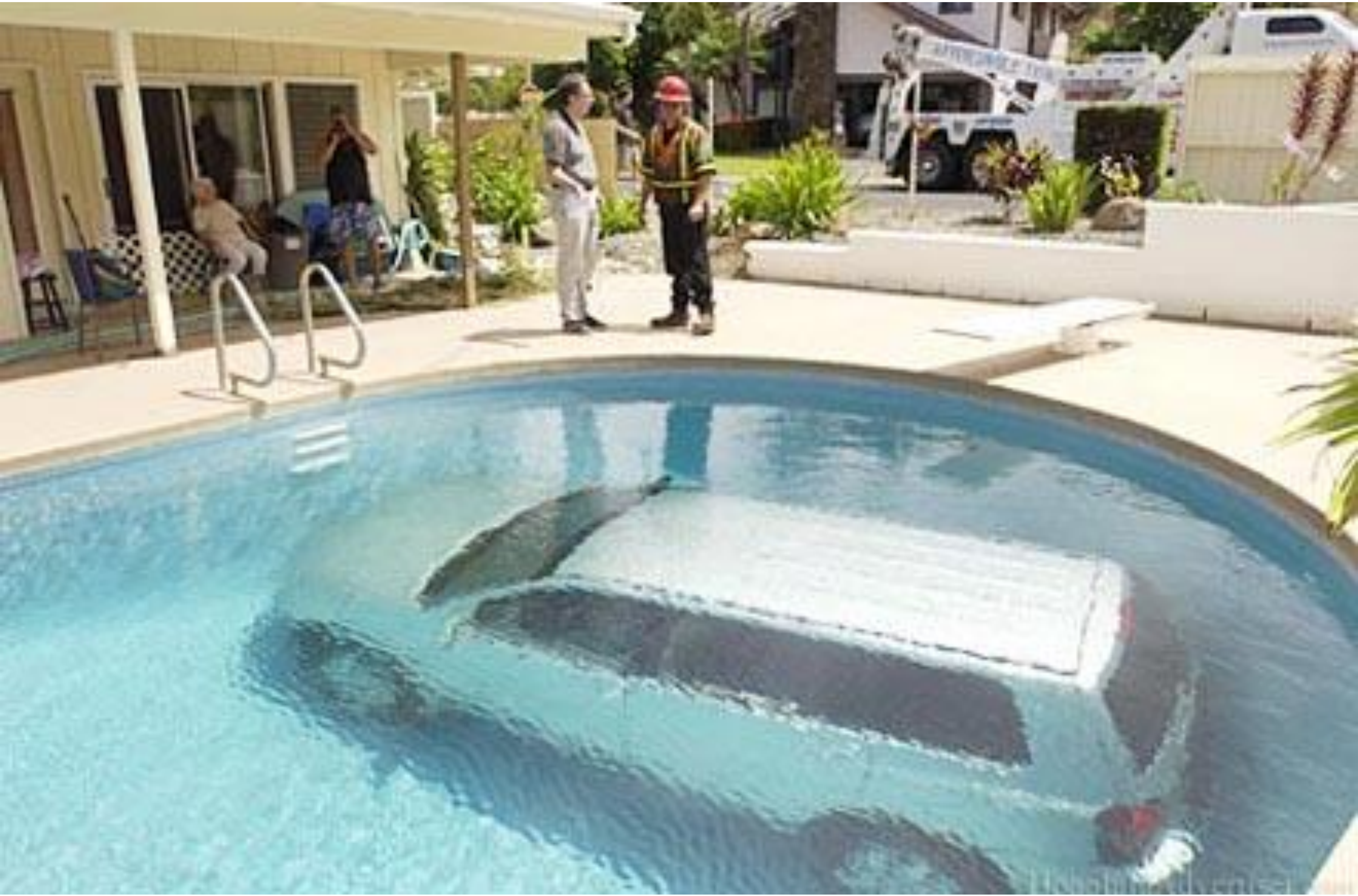
Among 161 images, there are:

- 13 co-occurrence violation
- 77 support violation
- 65 with both co-occurrence and support violations
- 6 with other violations (e.g. scale)

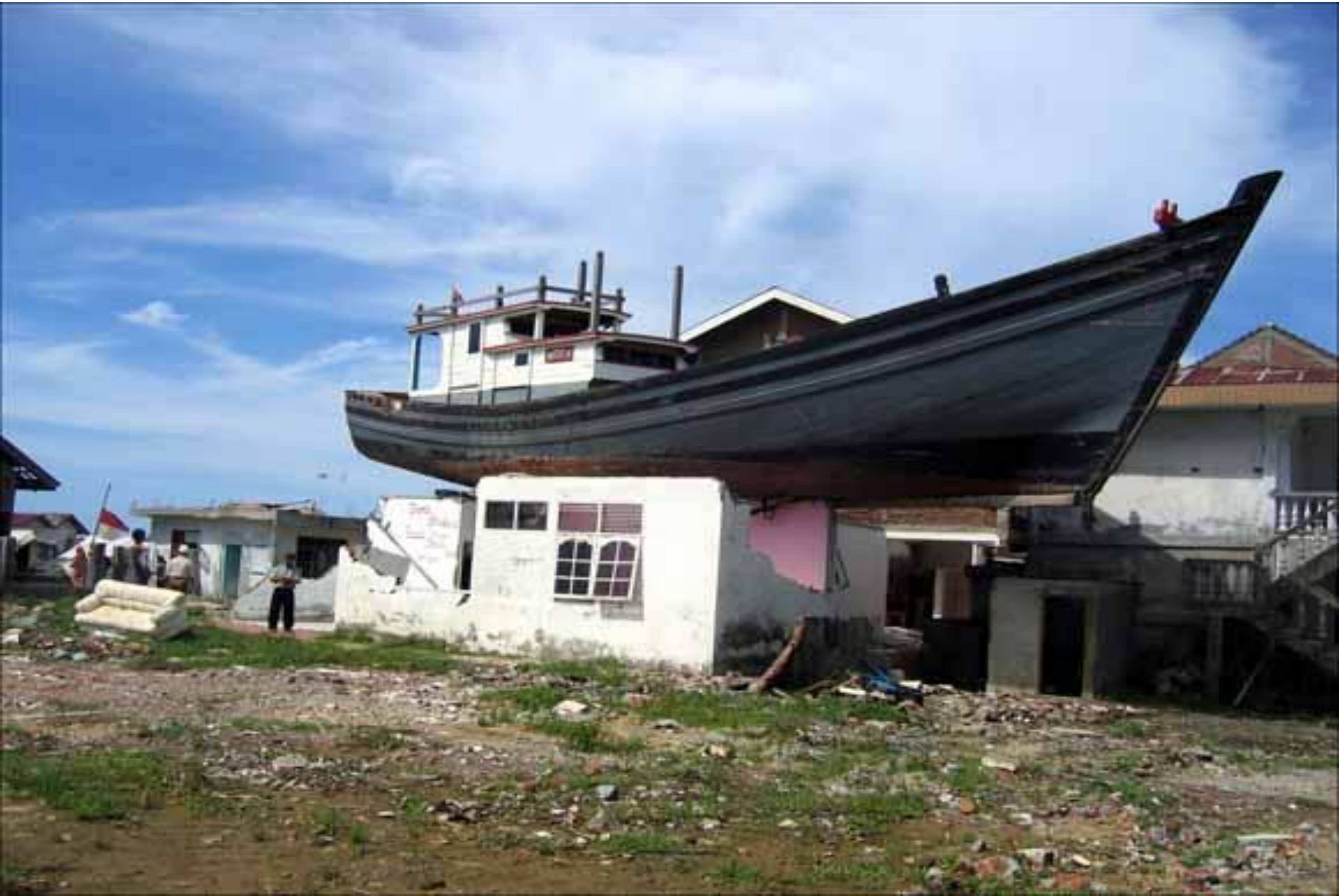
Out of context objects in the real world



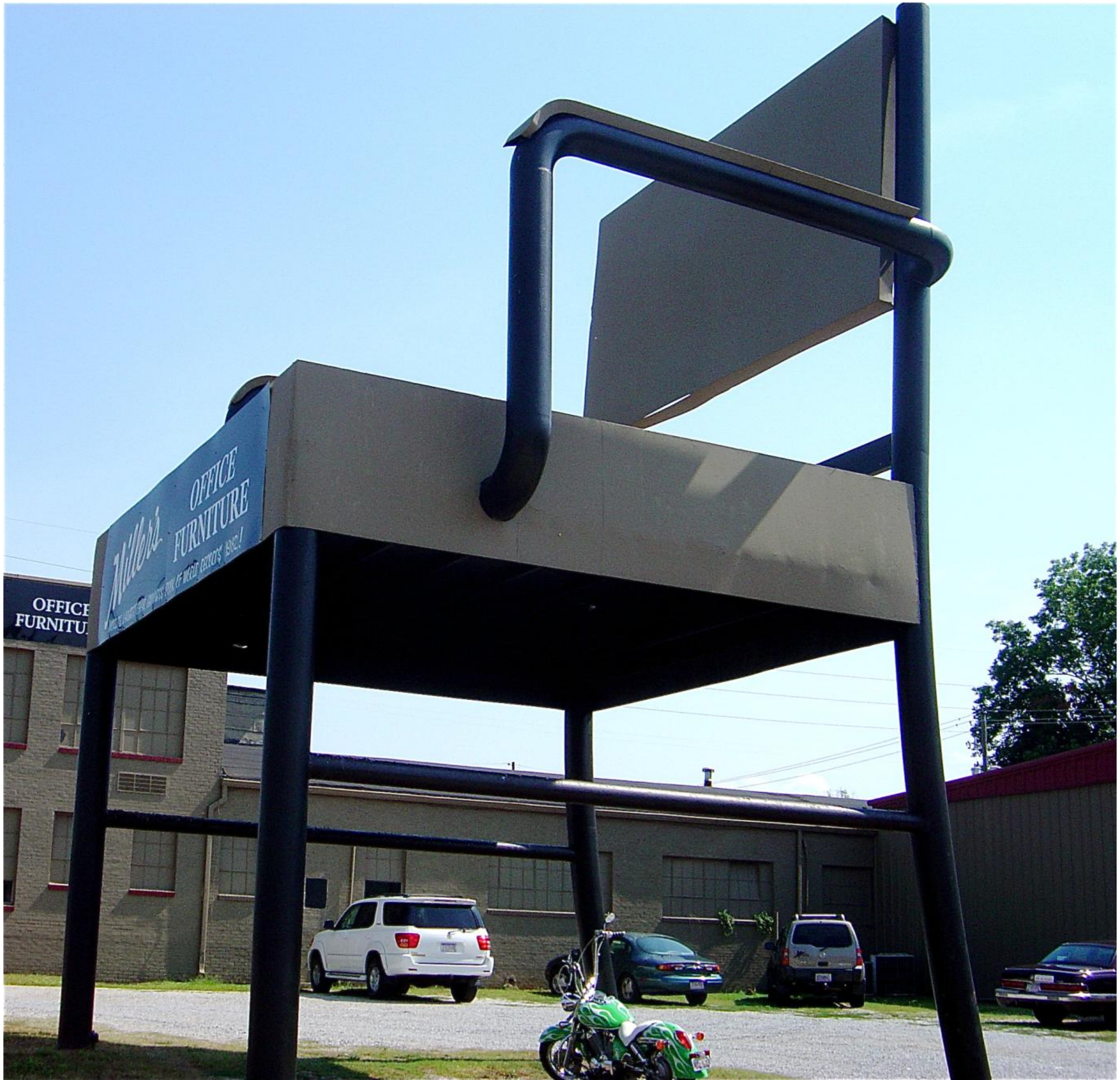
Out of context objects in the real world



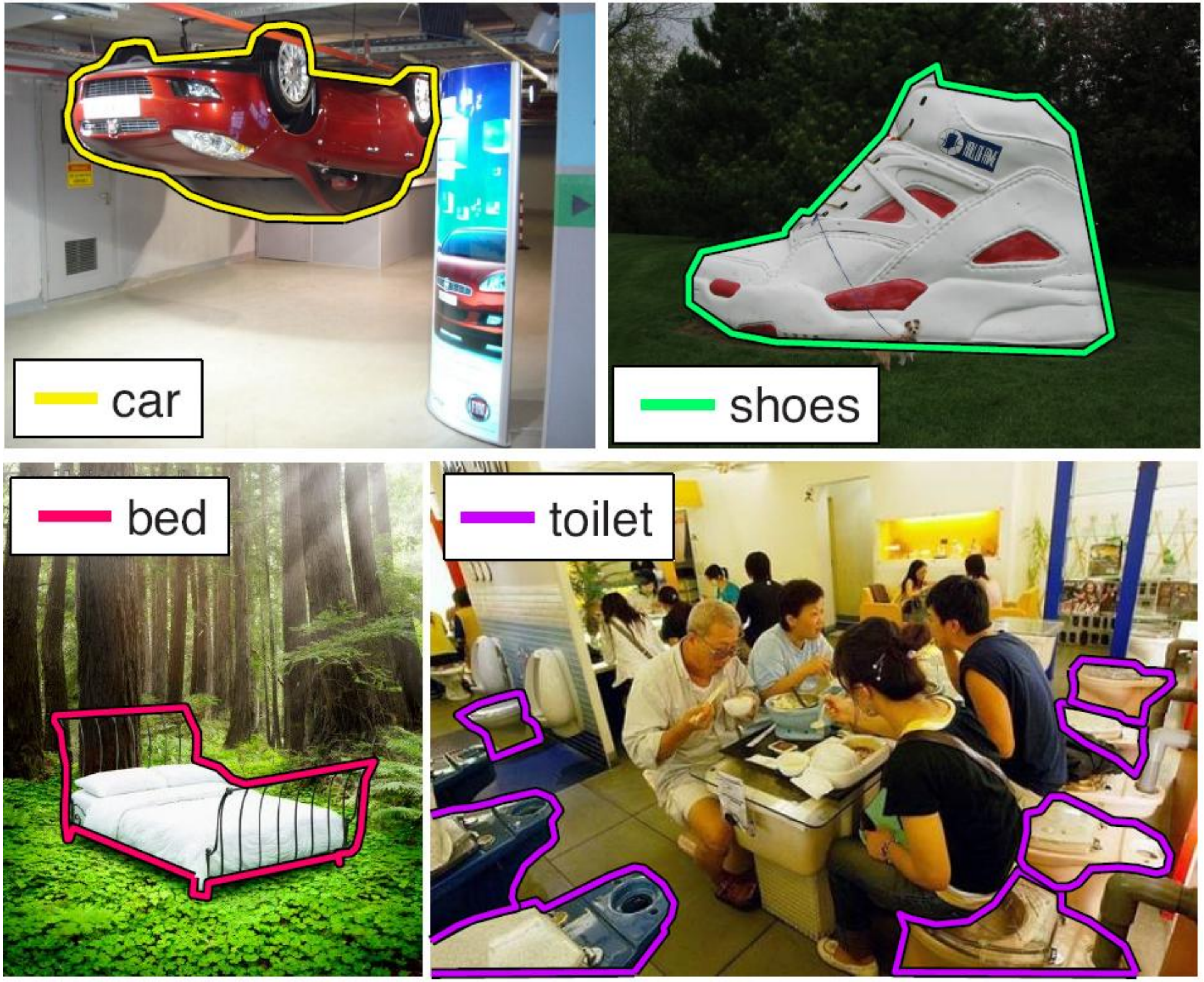
Out of context objects in the real world



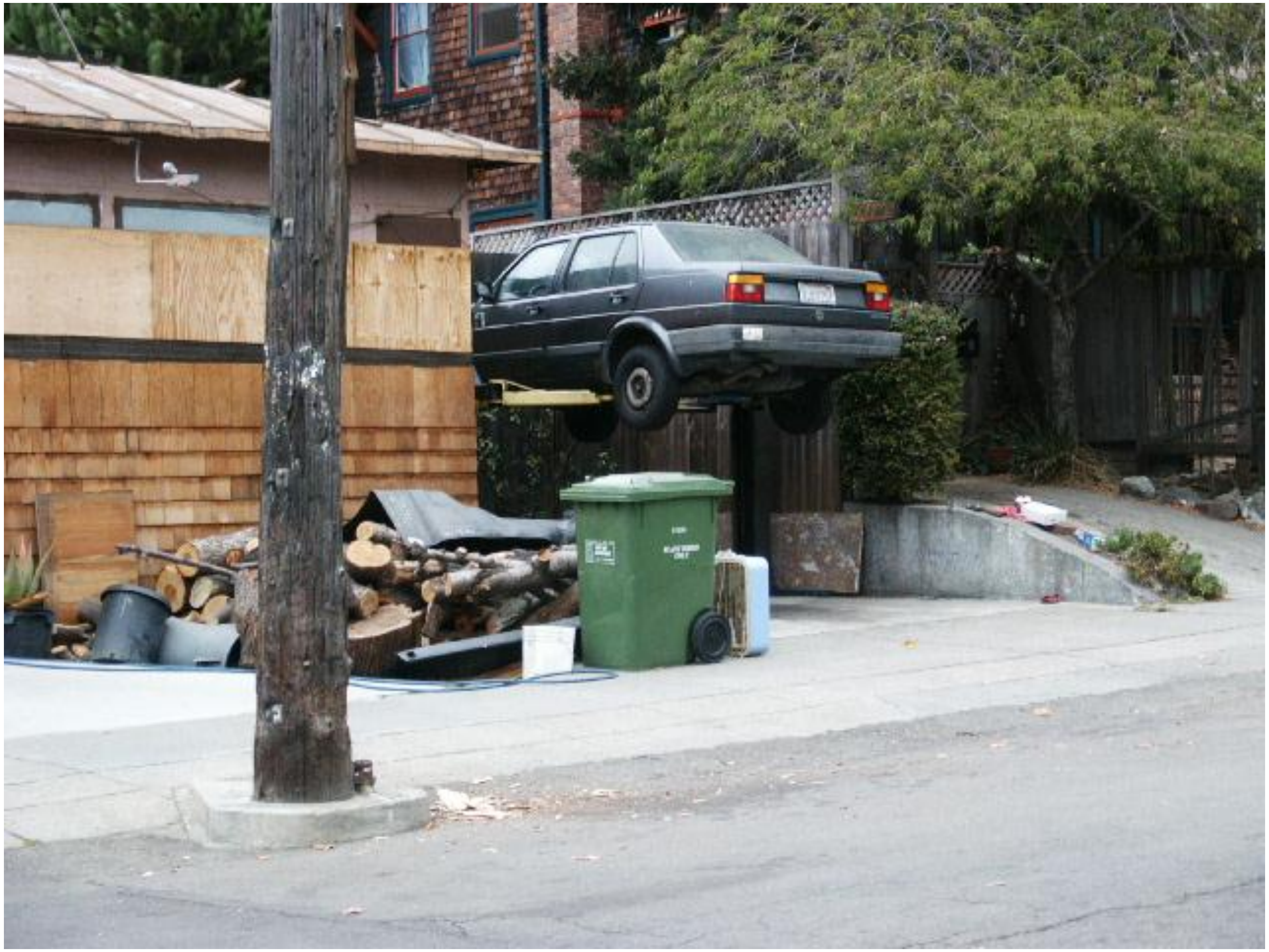
Out of context objects in the real world



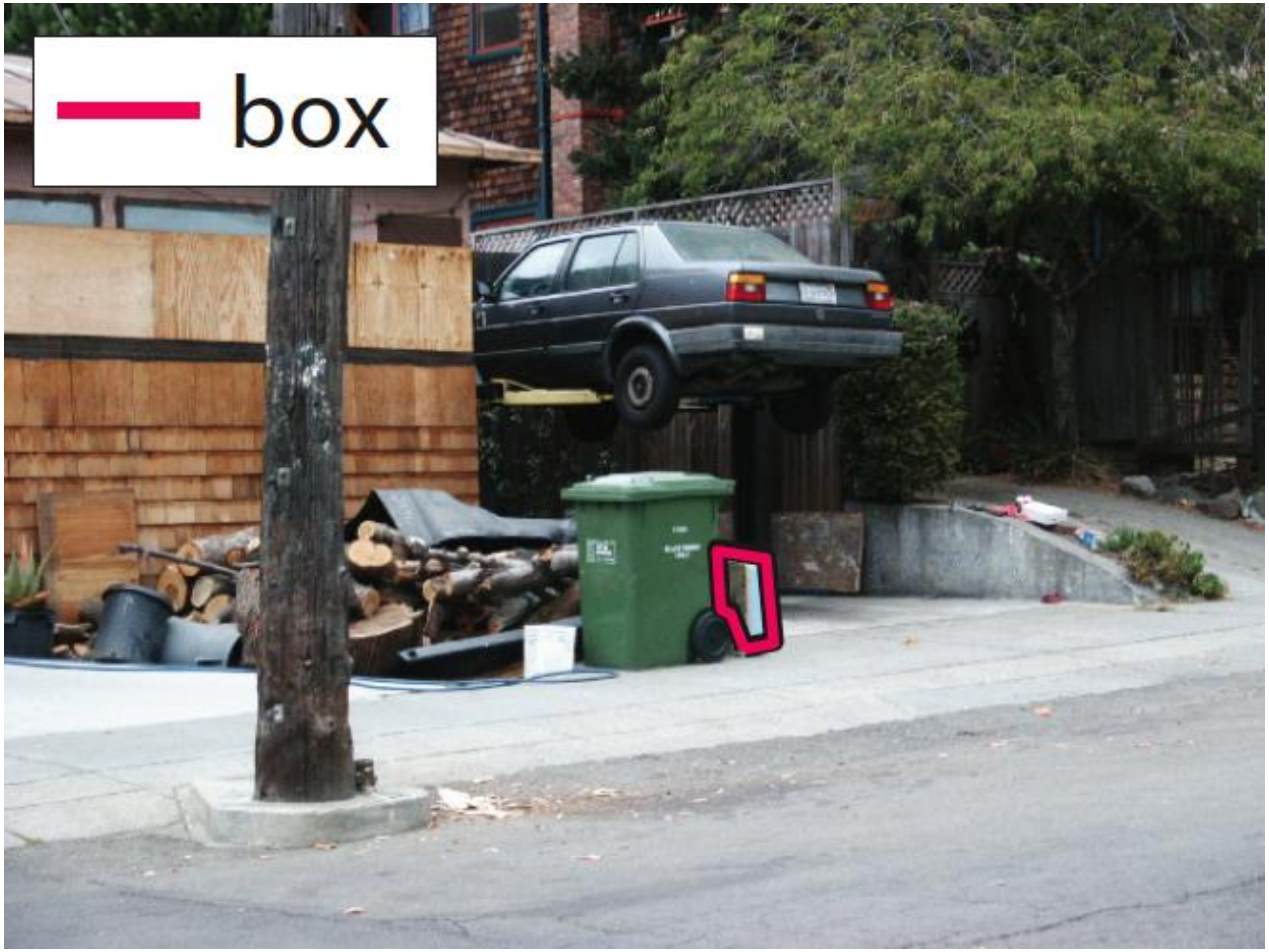
Detecting out of context objects



Detecting out of context objects



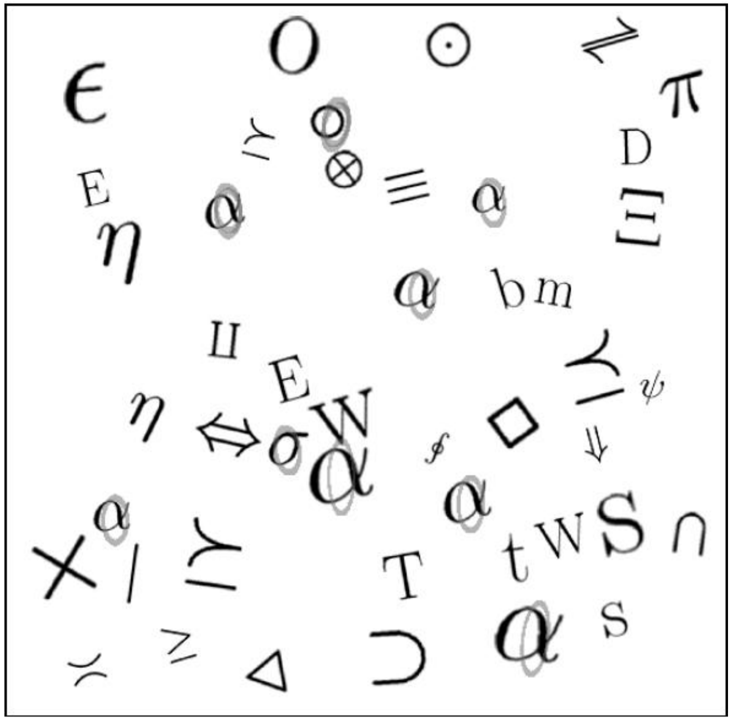
Detecting out of context objects



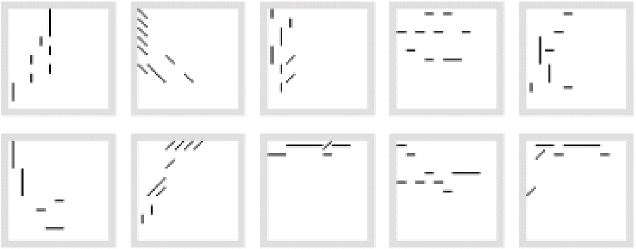
Reusable Parts

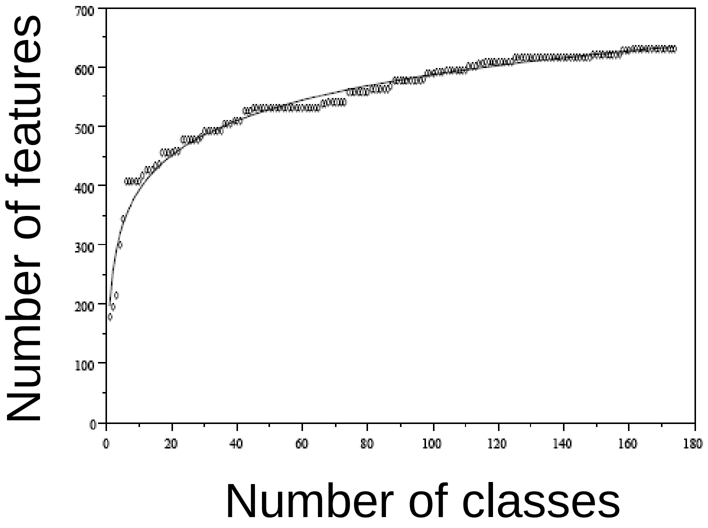
Krempp, Geman, & Amit “Sequential Learning of Reusable Parts for Object Detection”. TR 2002

Goal: Look for a vocabulary of edges that reduces the number of features.

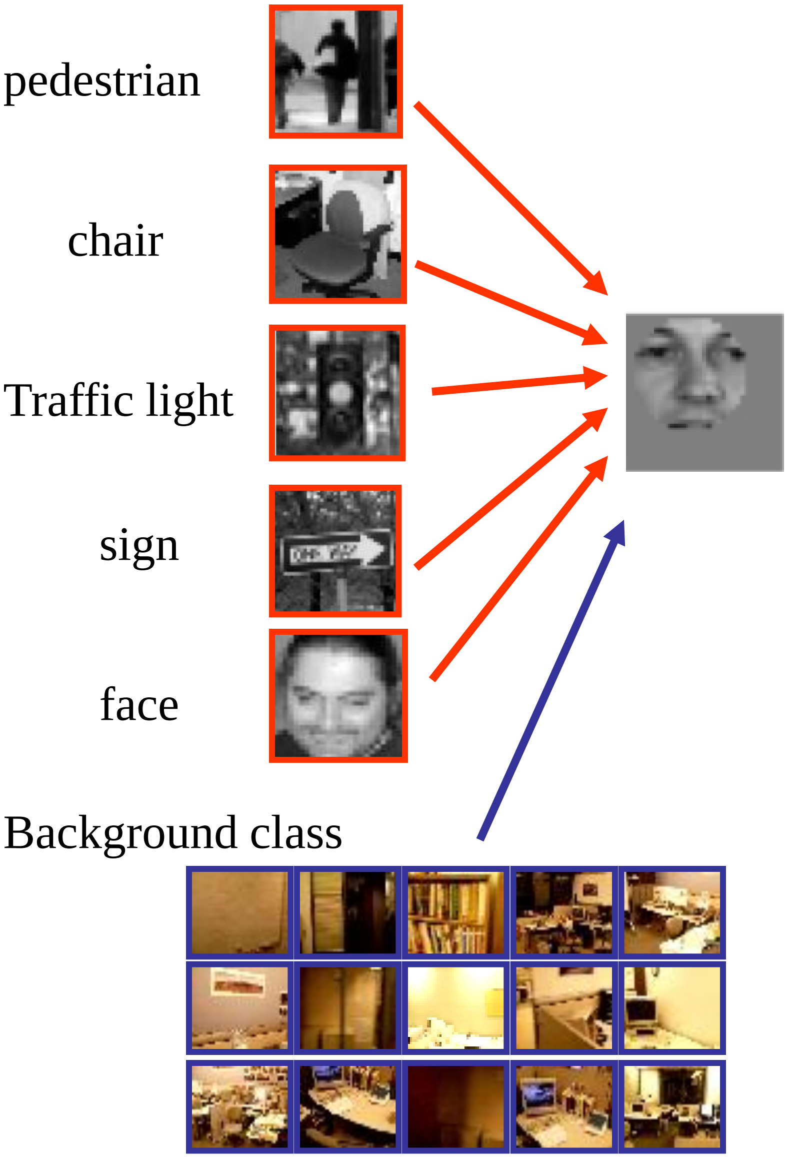


Examples of reused parts





Specific feature



pedestrian

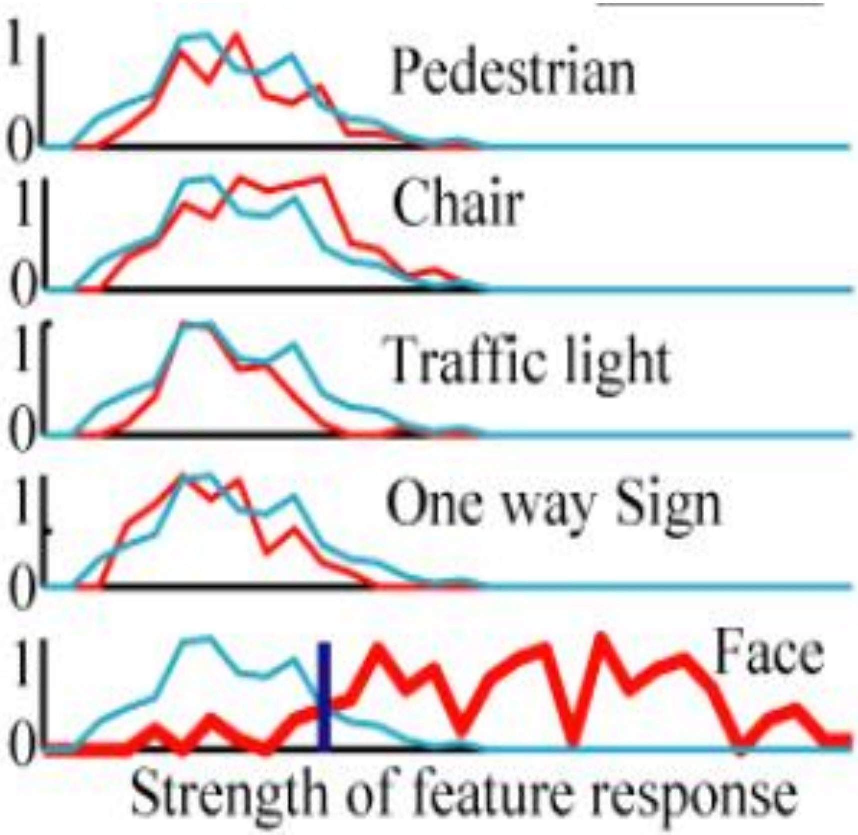
chair

Traffic light

sign

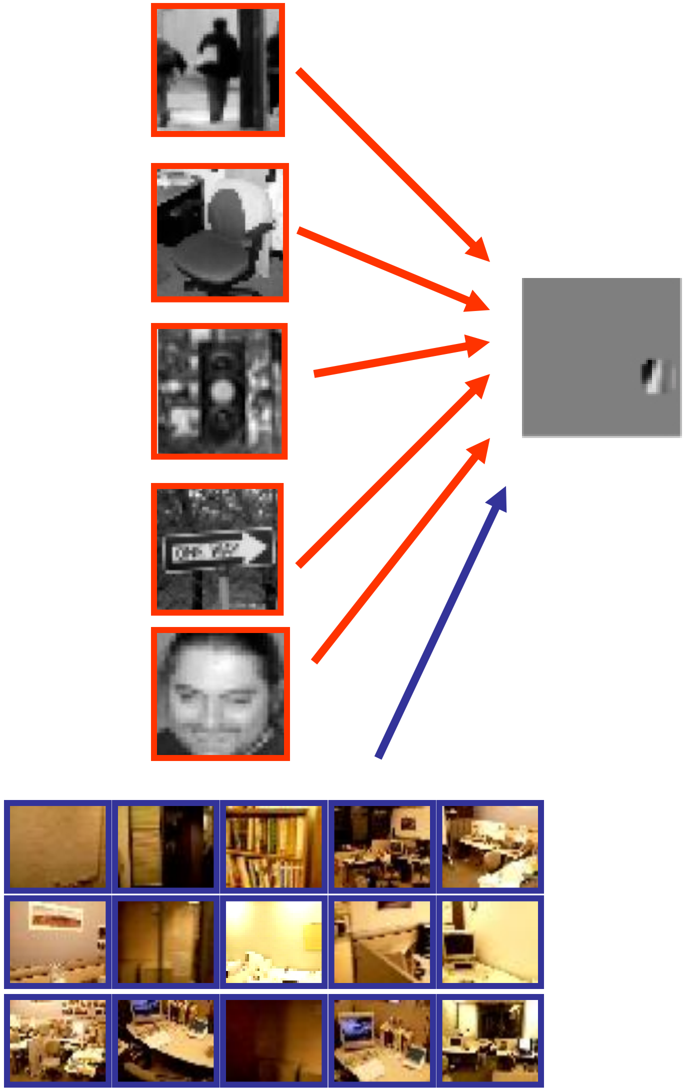
face

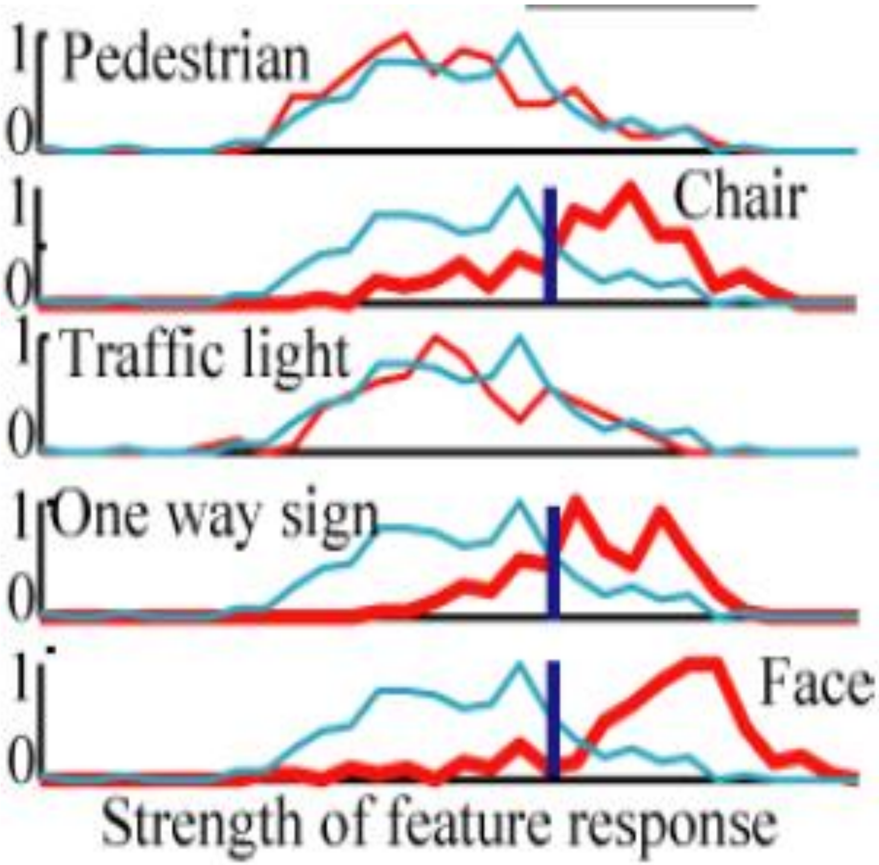
Background class



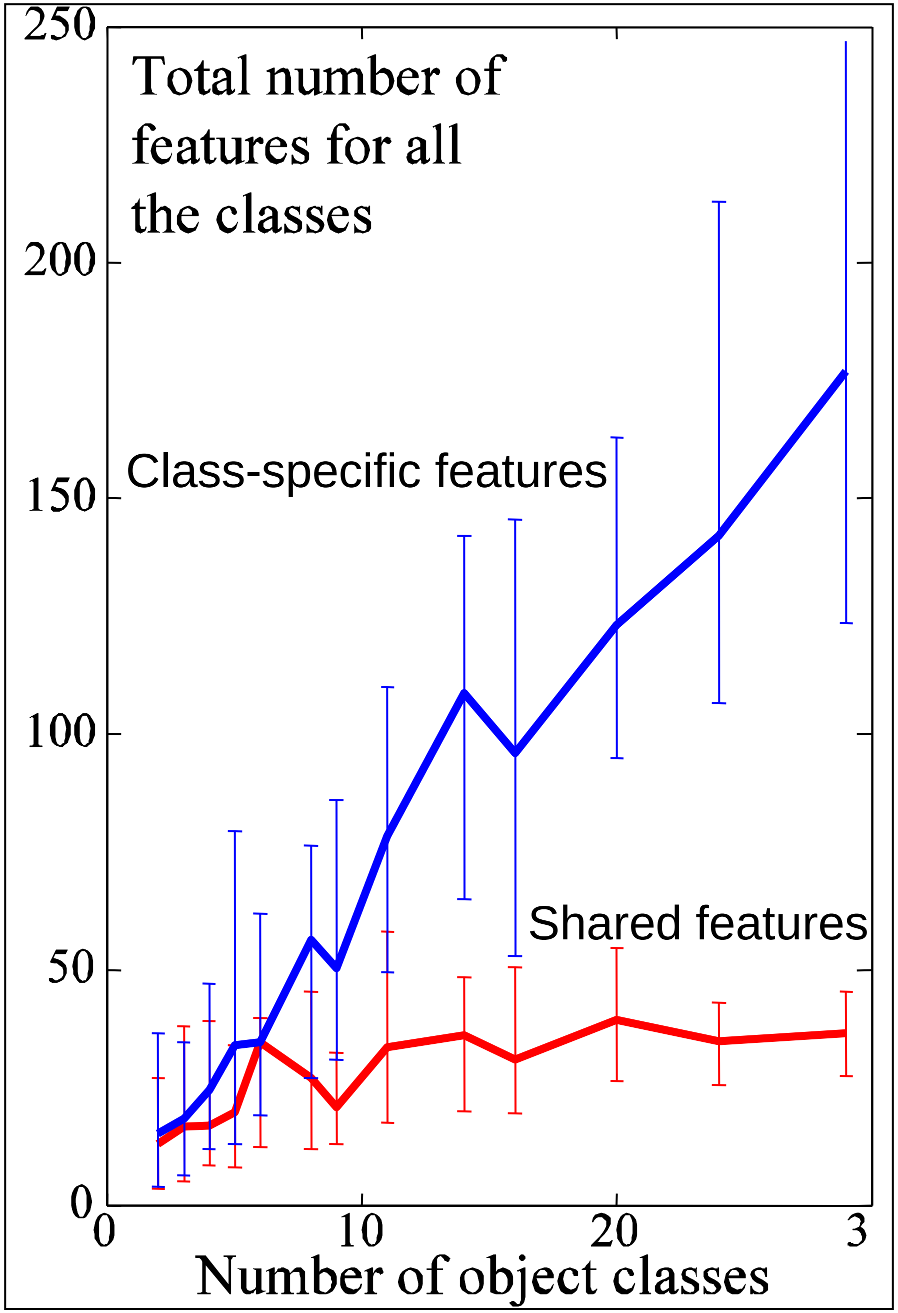
Non-shared feature: this feature is too specific to faces.

Shared feature





shared feature



50 training samples/class
29 object classes
2000 entries in the dictionary

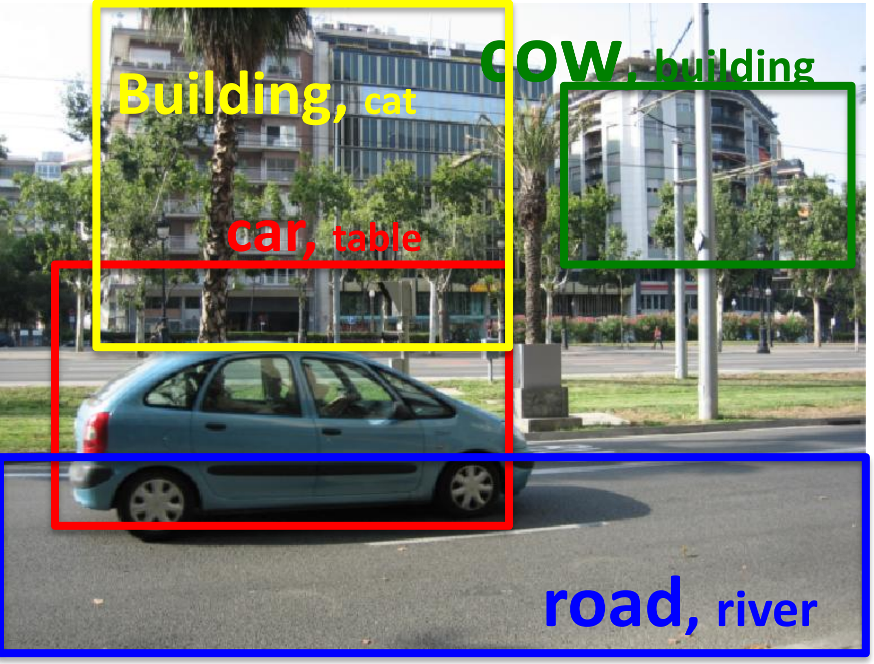
Results averaged on 20 runs
Error bars = 80% interval

Torralba, Murphy, Freeman. CVPR 2004. PAMI 2007

Goal: to assign labels c_k to each candidate so that they are in contextual agreement.

M possible object labels
N regions

Label: c_k = [1…M] with k = [1…N]
Scores: s_k = vector length M



We want to optimize the joint probability of all the labels:

$$p(c_1 = m_1, \ldots, c_N = m_N \mid s_1, \ldots, s_N)$$

Solution 3: Approximated model of dependencies:

$$p(c_1=m_1,\ldots, c_N=m_N | s_1,\ldots, s_N) =$$

$$= \frac{\prod_{i=1\ldots N} p(s_i | c_i=m_i)\, p(c_1=m_1,\ldots,c_N=m_N)}{Z(s_1,\ldots s_N)}$$

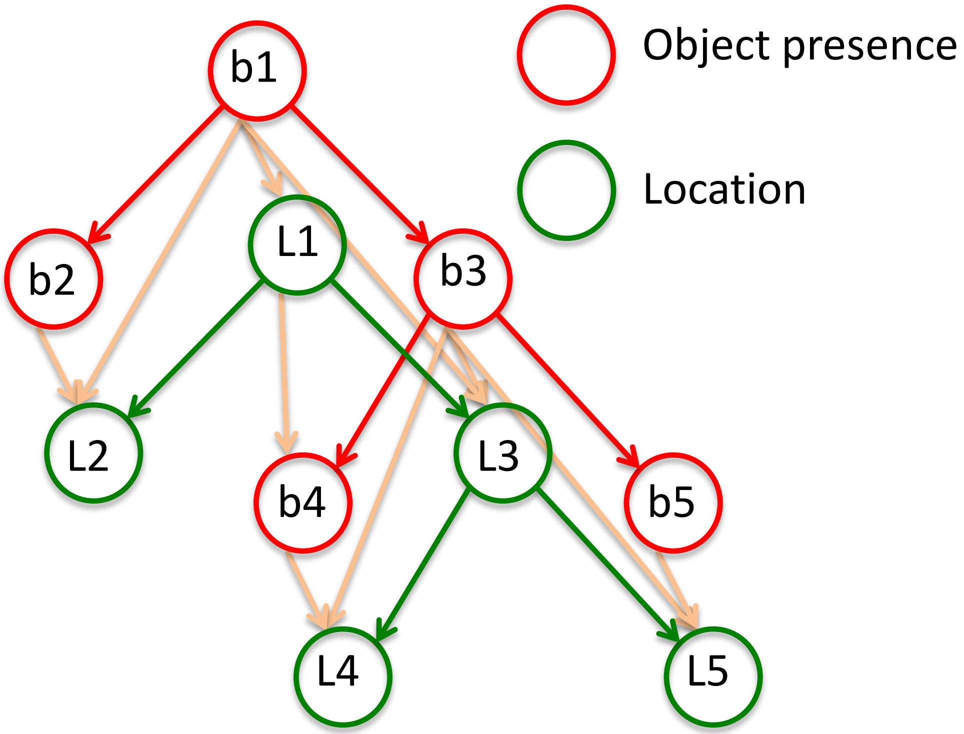
$$p(c_1=m_1,\ldots,c_N=m_N) = \exp(\sum_{i,j=1\ldots N} \Phi(c_i=m_i, c_j=m_j))$$

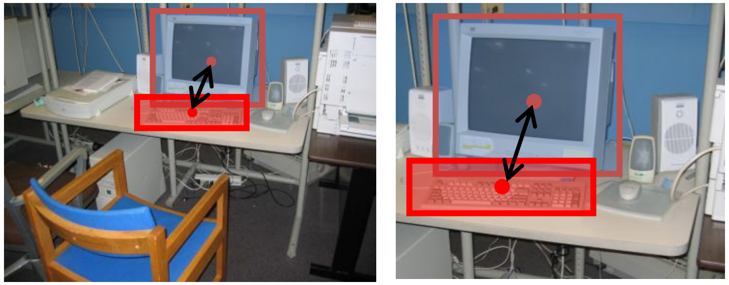
$\Phi(c_i=m_i, c_j=m_j)$ = co-ocurrence matrix on training set (count how many times two objects appear together).

Problem: learning $p(c_1=m_1,\ldots,c_N=m_N)$ will be easier, but recognition may still be slow.

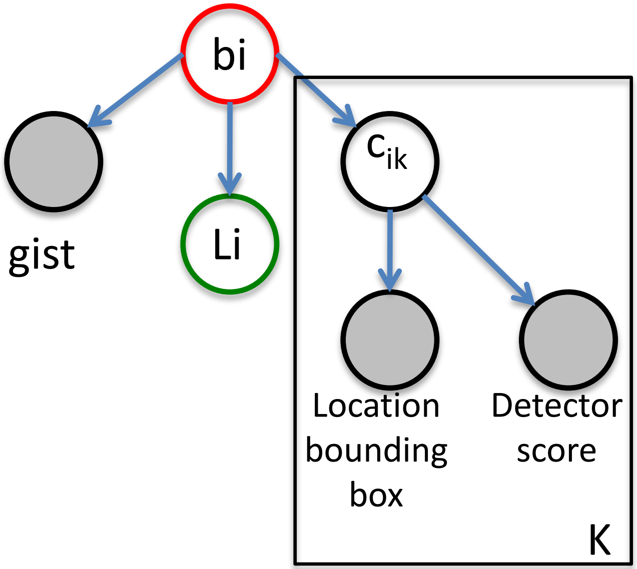
Tree structured context model

Prior model



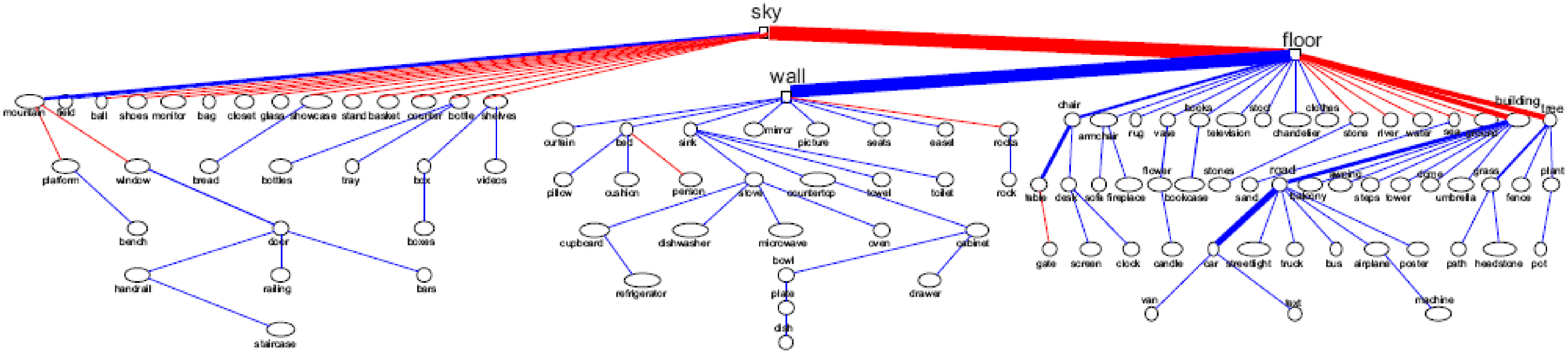


Observation model



Learning: Chow-Liu algorithm

Tree learned from SUN 09

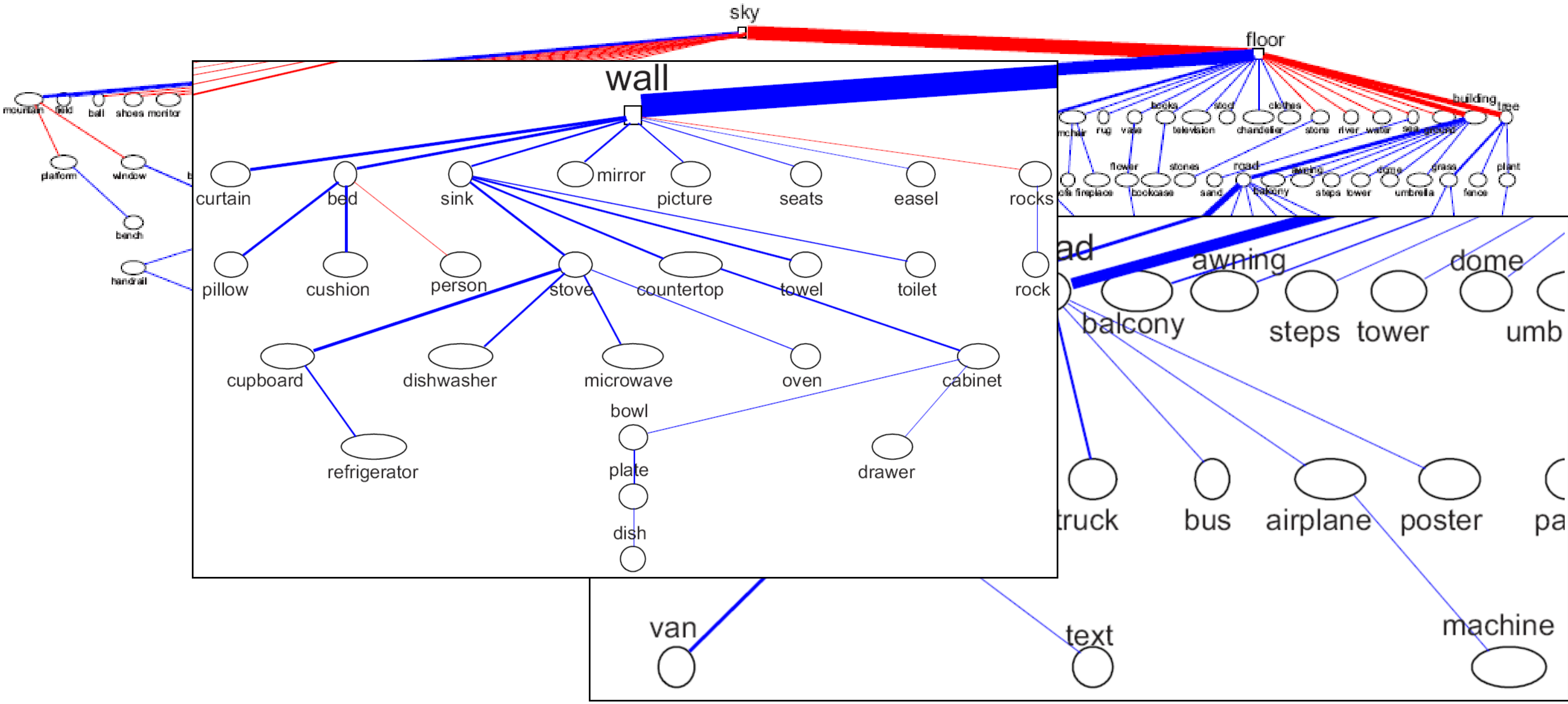


107 object categories

4317 training images

25/106 edges and 7/top-53 edges (≈13%) negative

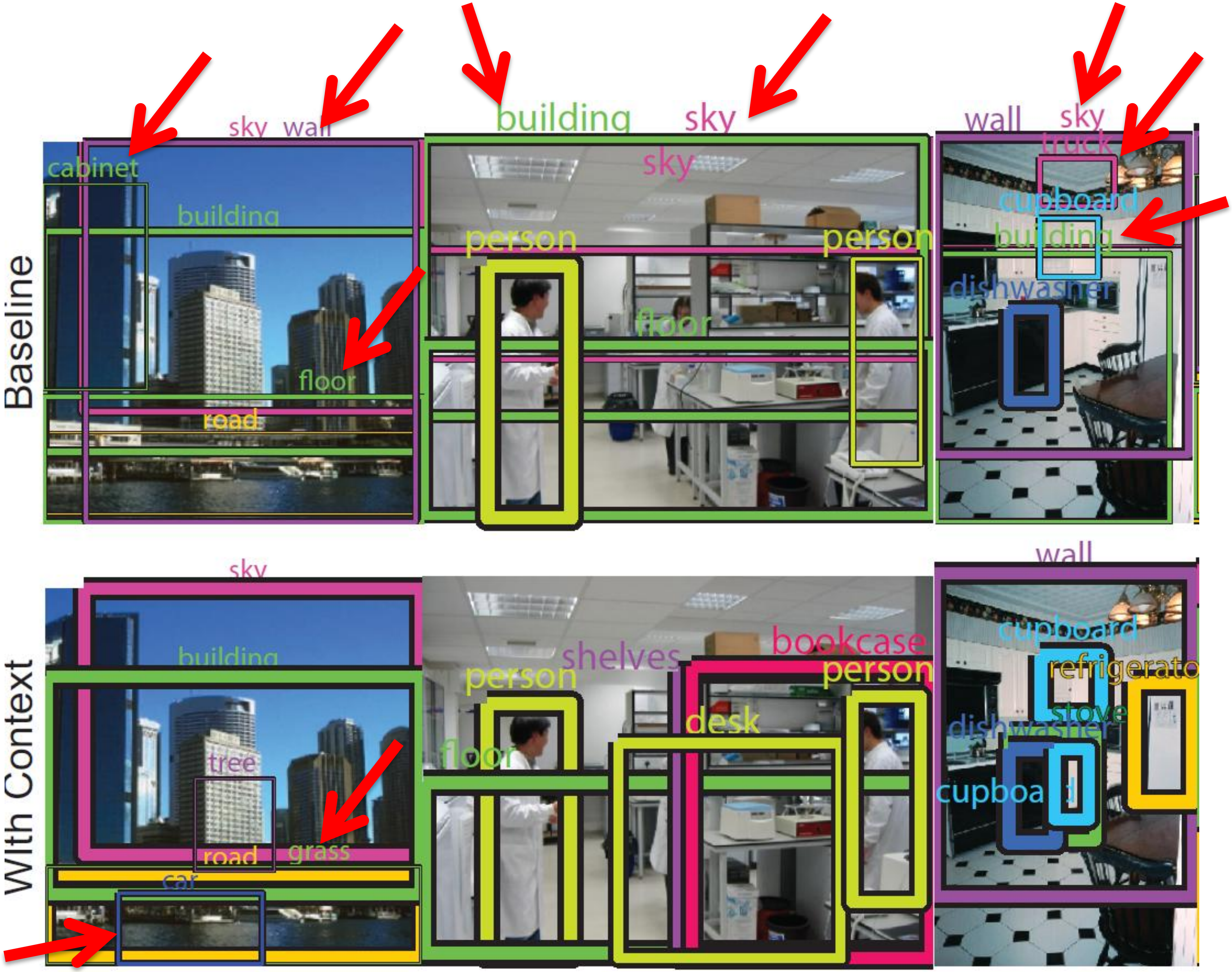
Learning object dependencies



107 object categories

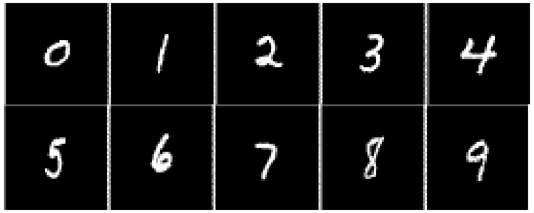
4317 training images

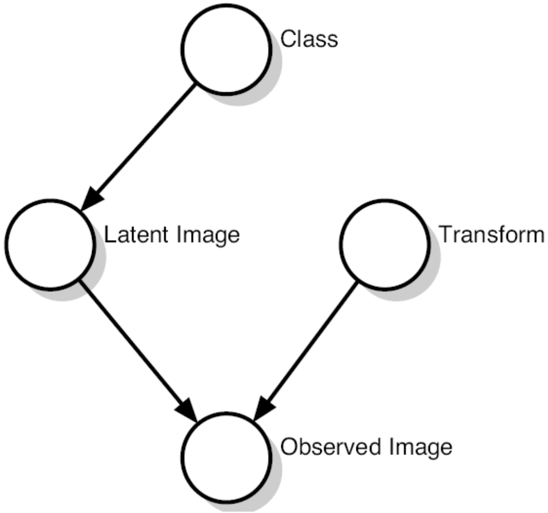
25/106 edges and 7/top-53 edges (≈13%) negative

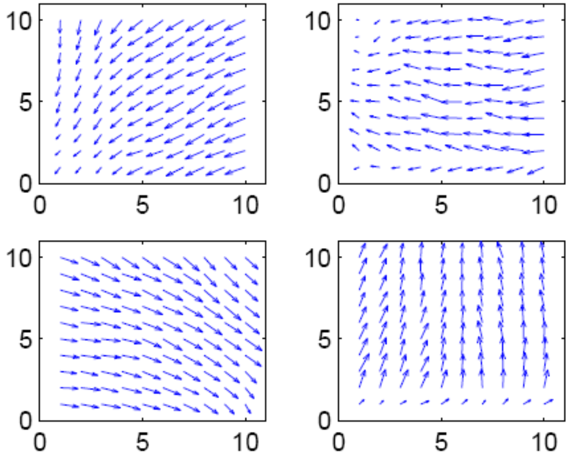

Baseline
sky
wall
cabinet
building
floor
road
building
sky
sky
person
person
floor
wall
sky
truck
cupboard
building
dishwasher
With Context
sky
building
tree
road
grass
car
shelves
person
bookcase
person
desk
floor
wall
cupboard
refrigerato
stove
dishwasher
cupboard

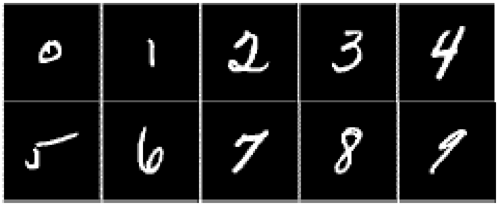
Sharing transformations

Miller, E., Matsakis, N., and Viola, P. (2000). Learning from one example through shared densities on transforms. In IEEE Computer Vision and Pattern Recognition.

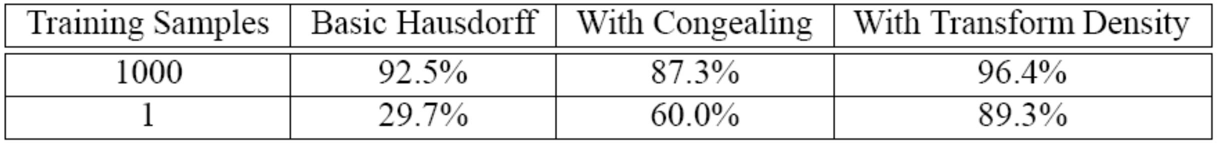








Transformations are shared
and can be learnt from other tasks.

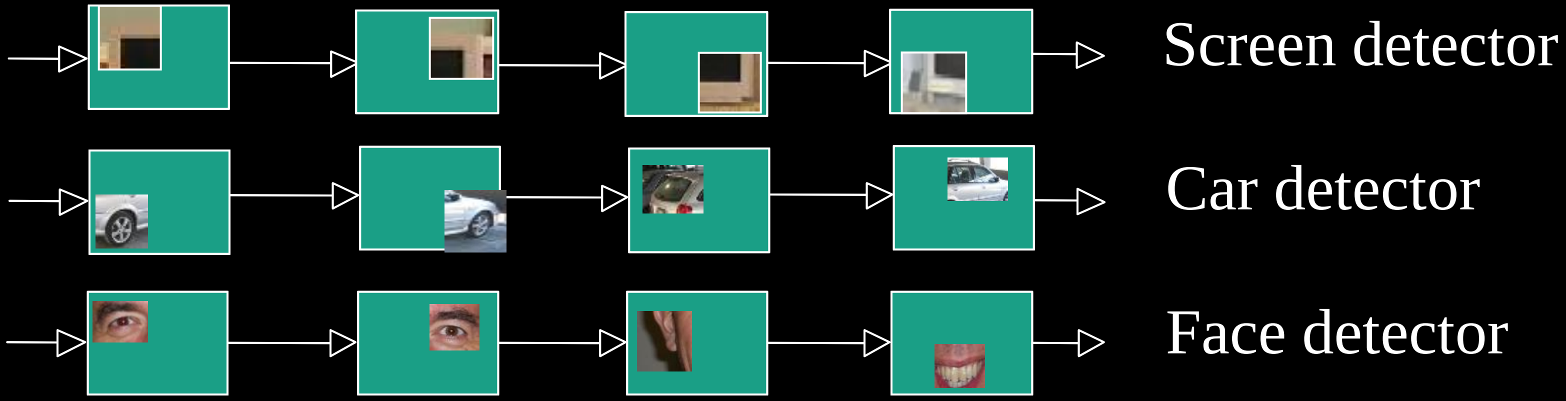


Training Samples	Basic Hausdorff	With Congealing	With Transform Density
1000	92.5%	87.3%	96.4%
1	29.7%	60.0%	89.3%

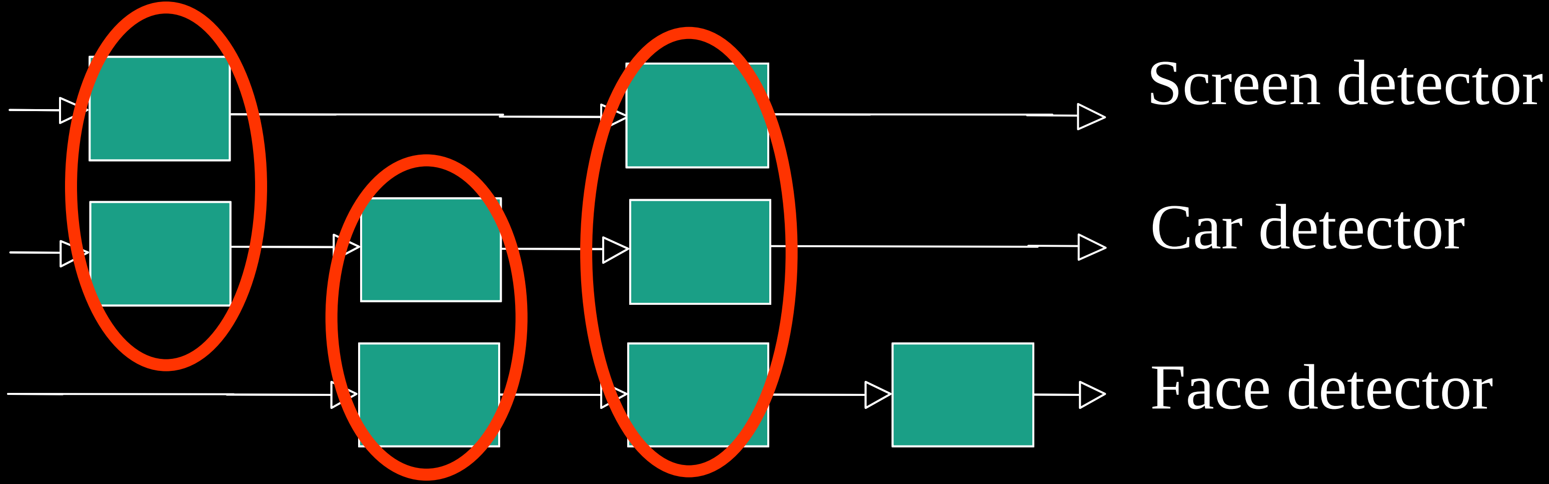
Additive models and boosting

Torralba, Murphy, Freeman. CVPR 2004. PAMI 2007

- Independent binary classifiers:

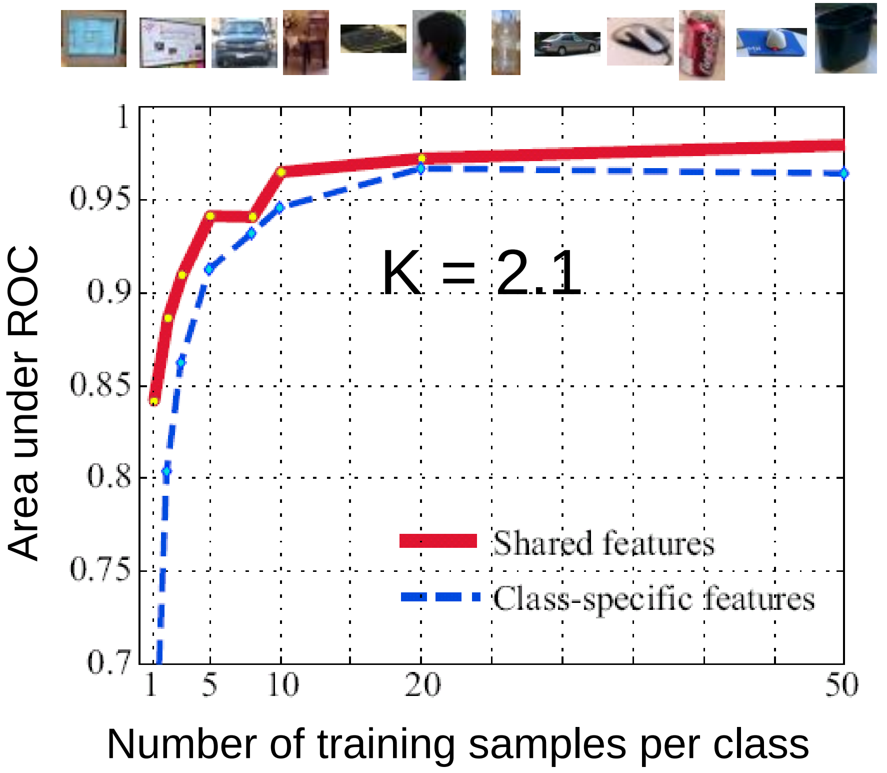


- Binary classifiers that share features:

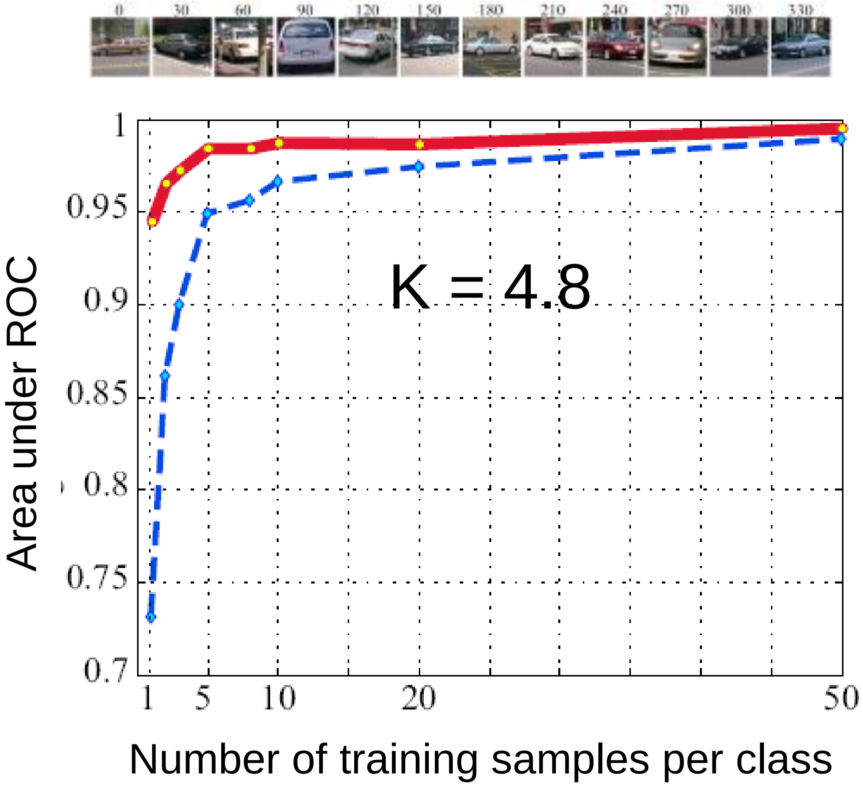


Generalization as a function of object similarities

12 unrelated object classes



12 viewpoints



Torralba, Murphy, Freeman. CVPR 2004. PAMI 2007