CSE 590V: Computer vision seminar

Fall 2011

Course webpage: http://www.cs.washington.edu/education/courses/cse590v/11au/

Logistics

- Time: Tuesdays from 1:30pm-2:30pm
- Location: CSE 403
- Organizers: Neeraj Kumar and Bryan Russell
- Class mailing list: cse590v @ cs washington edu (subscribe at course webpage)

Course webpage: http://www.cs.washington.edu/education/courses/cse590v/11au/

http://www.cs.washington.edu/education/courses/cse590v/11au/

Google

CSE 590V: Computer vision seminar

Fall 2011

Late stroll by Leonid Afremov

Course description

CSE 590V is a seminar/reading group focused on recent work in computer vision. We will cover papers from recent and upcoming conferences related to computer vision (CVPR, ICCV, ECCV, NIPS, SIGGRAPH). The seminar is open to everyone. We especially encourage first year graduate students who may be considering research in computer vision or related areas to participate.

Logistics

Time: Tuesdays from 1:30pm-2:30pm

Location: CSE 403

Course description

- This is a seminar on recent work in computer vision
- We will cover papers from recent conferences related to computer vision: CVPR, ICCV, ECCV, NIPS, SIGGRAPH
- We have organized the papers into topics
- Each week, we will discuss the papers for a topic

Potential list of topics covered in class

- Datasets and active learning (covered today)
- Attributes (covered next time)
- Poselets
- Person detection
- Scene understanding
- Large scale recognition
- Learning
- Events and actions
- Language
- Cross-domain/multi-modal learning & matching
- Crowds & videos/social networks
- Shading and lighting
- Multi-view geometry
- RGB-D perception
- Cognitive science & saliency
- Misc/cool papers

Course expectations (everybody)

- Read the assigned paper(s) beforehand
- Come ready to discuss the papers
- Make a list of 3 items to discuss, for example:
  - Question
  - Extension
  - Critique

Course expectations (students)

- Give a presentation on one of the topics
- E-mail Neeraj or Bryan top 3 preferred topic choices by this Friday
- We will assign topics by next week
- Friday before you give your presentation, meet with Neeraj and Bryan to discuss upcoming presentation
- We will award a prize to the best presentation

Volunteer(s) for next week?

- Topic: attributes
  - Automatic Attribute Discovery and Characterization. Tamara Berg, Alexander Berg, Jonathan Shih. ECCV 2010.
  - Relative Attributes. Devi Parikh, Kristen Grauman. ICCV 2011.
  - Attribute Learning in Large-scale Datasets. O. Russakovsky and L. Fei-Fei. Workshop on Parts and Attributes, assoc. with ECCV 2010.
  - Interactively Building a Discriminative Vocabulary of Nameable Attributes. Devi Parikh, Kristen Grauman. CVPR 2011.

Summary to do list

- Today: sign up for course mailing list
- Friday: e-mail Neeraj and Bryan top 3 preferred topics
- Next Tuesday: read assigned attribute papers (we will let you know which ones we will focus on)

Datasets for object recognition and scene understanding

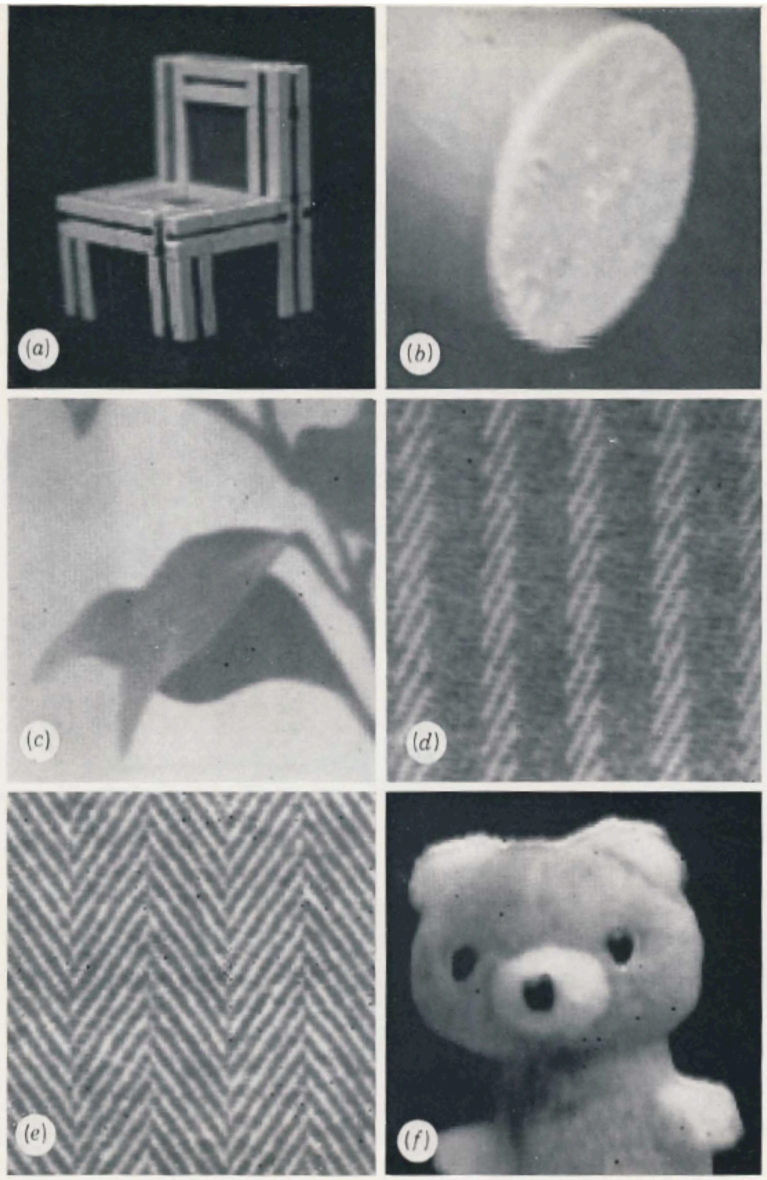
Bryan Russell

In the beginning, things weren't always so easy…

1972

Slide credit: A. Torralba

Slide credit: A. Torralba



URE 4. This figure provides a high quality reproduction of the six images discussed in the text. (*a*) and (*b*) were taken with a considerably modified Information International Incorporated Vidissector, and the rest were taken with a Telemation TMC-2100 vidicon camera attached to a Spatial Data Systems digitizer (Camera Eye 108). The full dynamic range from black to white is represented by 256 grey-levels. The images reproduced here were created by an Optronics P150ohPhotowriter from intensity arrays that measured 128 elements square. This size of intensity array corresponds to viewing a 1 in square at 5 ft with the human retina. The image of the period at the end of this sentence probably covers more than 40 retinal receptors. The reader should view the images from a distance of about 5 ft when assessing the performance of the programs.

Marr, 1976

Slide credit: A. Torralba

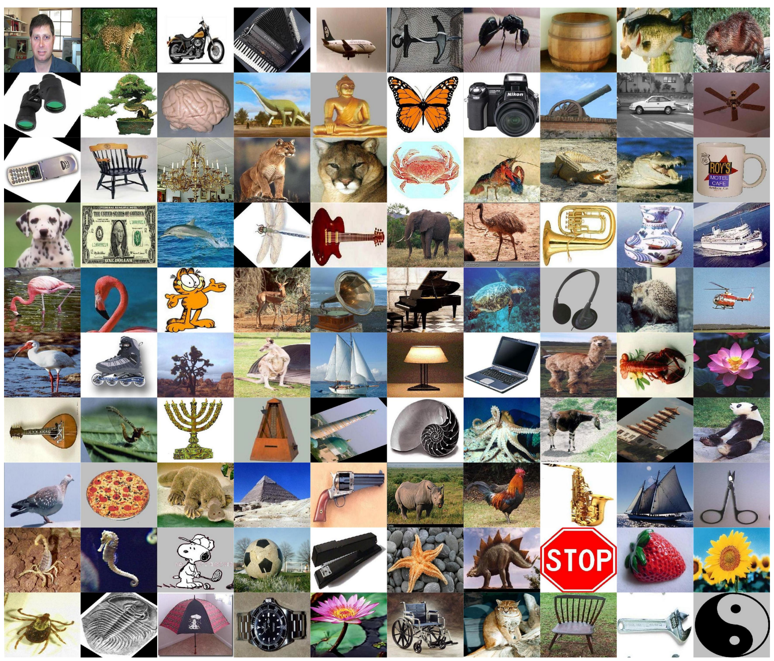
URE 4. This figure provides a high quality reproduction of the six images discussed in the text. (*a*) and (*b*) were taken with a considerably modified Information International Incorporated Vidissector, and the rest were taken with a Telemation TMC-2100 vidicon camera attached to a Spatial Data Systems digitizer (Camera Eye 108). The full dynamic range from black to white is represented by 256 grey-levels. The images reproduced here were created by an Optronics P150ohPhotowriter from intensity arrays that measured 128 elements square. This size of intensity array corresponds to viewing a 1 in square at 5 ft with the human retina. The image of the period at the end of this sentence probably covers more than 40 retinal receptors. The reader should view the images from a distance of about 5 ft when assessing the performance of the programs.

Marr, 1976

The rise of the modern dataset…

Caltech 101 and 256

101 object classes



Fei-Fei, Fergus, Perona, 2004

9,146 images

256 object classes

Griffin, Holub, Perona, 2007

30,607 images

Slide credit: A. Torralba

Hao Wooi Lim's blog

Where my thoughts are stored in byte-addressable little-endian format memory.

FRIDAY, AUGUST 21, 2009

Table of results for Caltech 101

This is a table documenting some of the best results some paper obtained in Caltech-101 dataset.

Results shown here are all trained using 30 samples from each category.

1. Group-Sensitive Multiple Kernel Learning for Object Categorization (ICCV 2009)
   Cited 17 times. **84.3%**
   Additional Info: GS-MKL
2. LP-Beta + Geometric blur + PHOW gray/color + Self-Similarity
   82.1% +- 0.3%
3. Learning Subcategory Relevances for Category Recognition (CVPR 2008)
   Cited 19 times. **81.9%**
   Poster: Link (PDF)
4. Object Recognition as Ranking Holistic Figure-Ground Hypotheses (CVPR 2010)
   Cited 8 times. **81.9%**
   Additional Info: Regression with Post-Processing.
5. Image Classification using Random Forests and Ferns (2007)
   Cited 130 times. **81.3%**
   Additional Info: Bosch Multi-way SVM
6. In Defense of Nearest-Neighbor Based Image Classification (CVPR 2008)
   Cited 139 times. **79.23%**
   Additional Info: NBNN (5 descriptors)
7. Visual Geometric Group (VGG)'s implementation of Multiple Kernel Image Classifier trained on dense SIFT, self-similarity, and geometric blur features
   78.20% +- 0.4%
   Additional Info: Result of 77.8% is obtained by combining dense SIFT, self-similarity, and geometric blur features with the multiple kernel learning

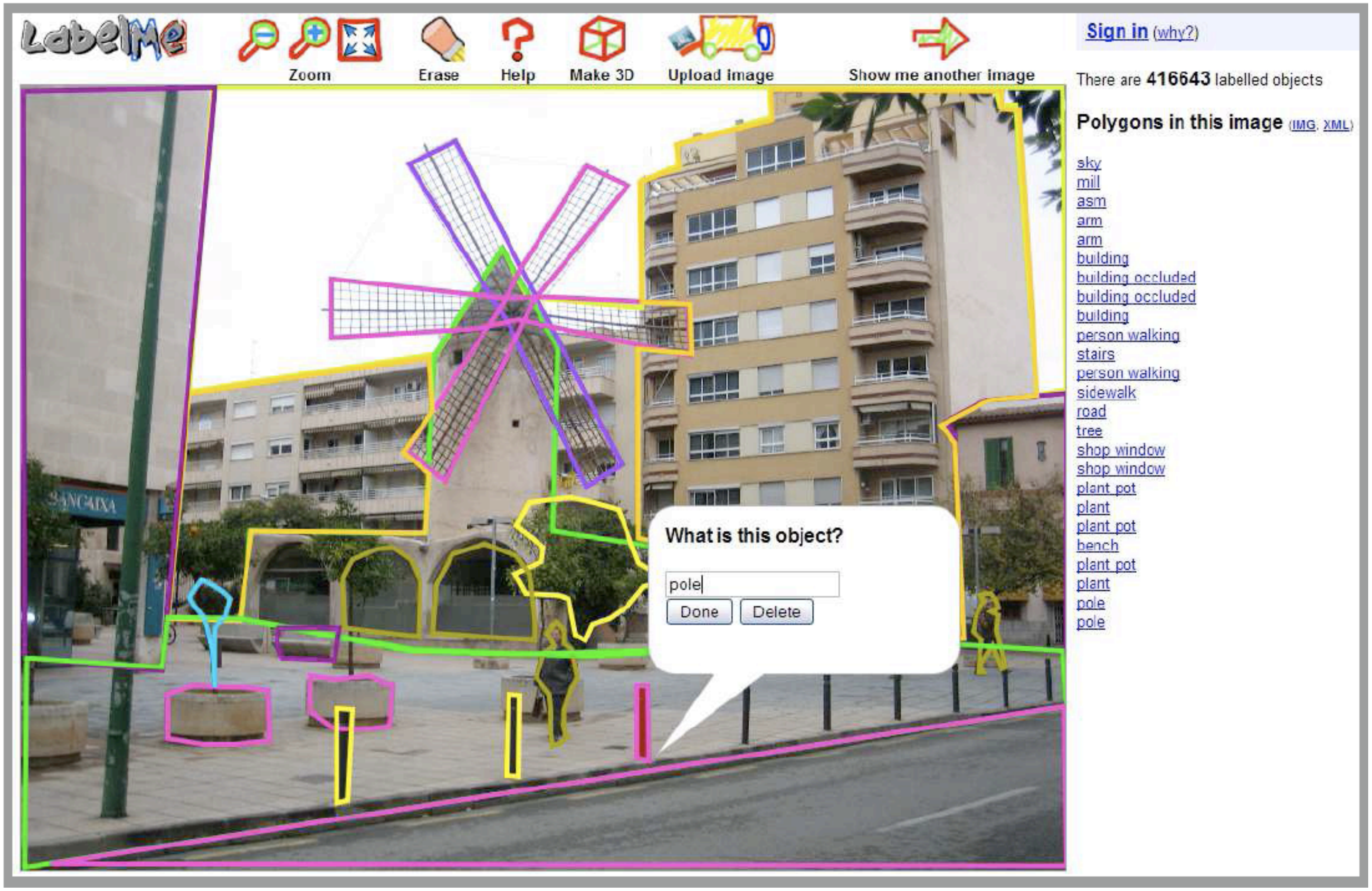
http://zybler.blogspot.com/2009/08/table-of-results-for-famous-public.html

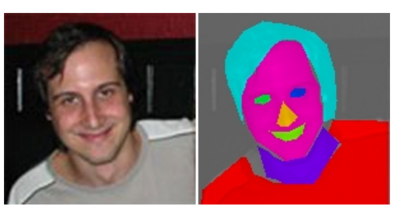
MSRC

591 images, 23 object classes
Pixel-wise segmentation

J. Winn, A. Criminisi, and T. Minka, 2005

LabelMe





Tool went online July 1st, 2005
825,597 object annotations collected
199,250 images available for labeling

labelme.csail.mit.edu

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

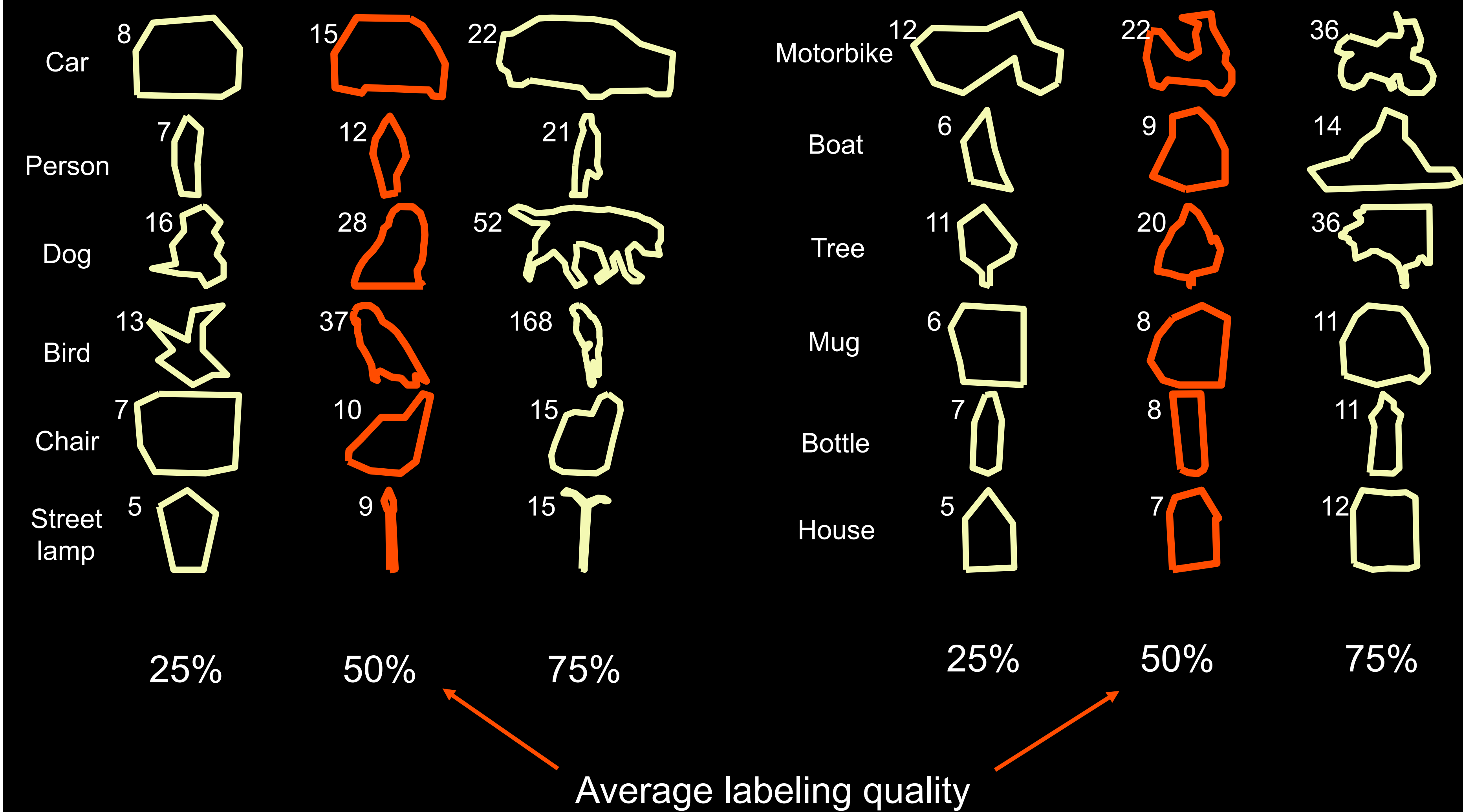
LabelMe

Labeling tool Help Downloads New! About Matlab Toolbox Upload Images Stats

Search

Your query (**street**) matches **13238** images

Quality of the labeling



Extreme labeling

The other extreme of extreme labeling

… things do not always look good…

Testing

Most common labels:

test

adksdsa

woiieiie

…

Sophisticated testing

Most common labels:

Star

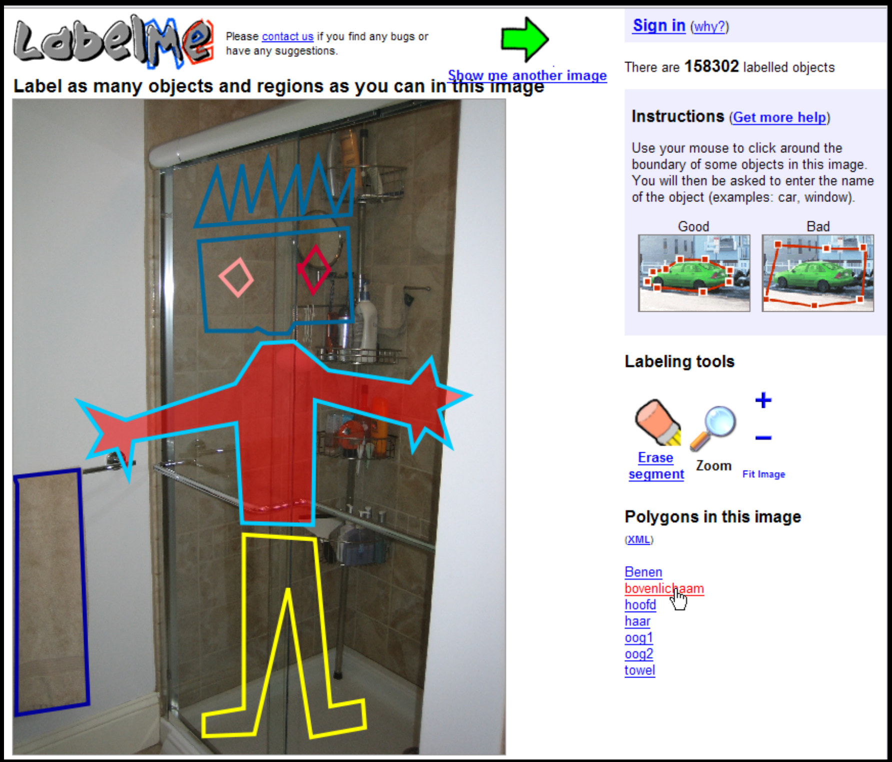
Square

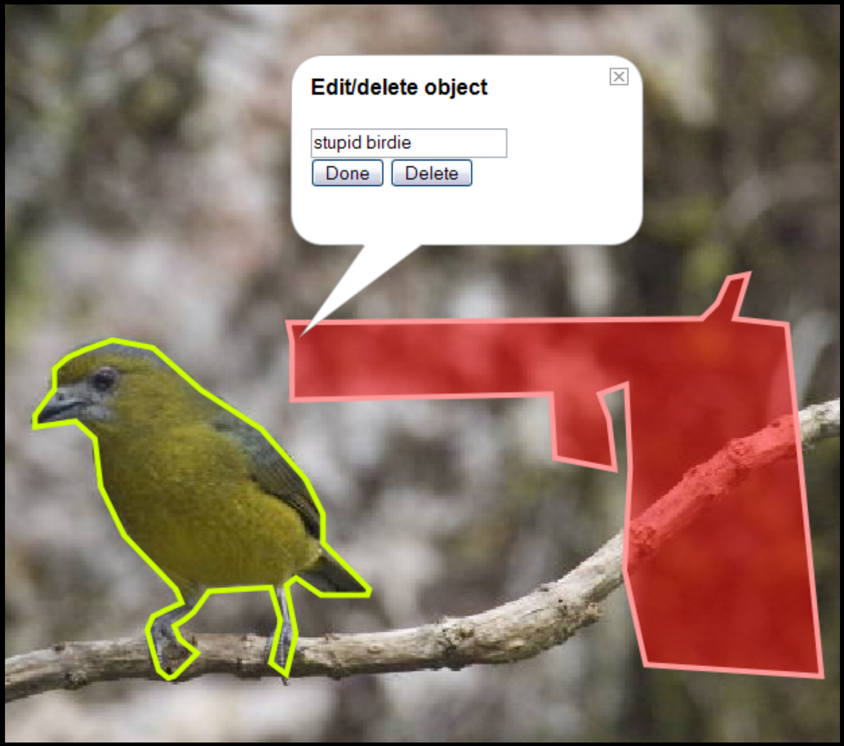
Nothing

…

Creative testing

Do not try this at home





Most common labels:

Stupid birdie

Tourist checking hottie

Man’s derriere

…

Visual Object Classes Challenge 2011 (VOC2011)

PASCAL2
Pattern Analysis, Statistical Modelling and Computational Learning

[click on an image to see the annotation]

2011 version - 20 object classes:
Person: person
Animal: bird, cat, cow, dog, horse, sheep
Vehicle: aeroplane, bicycle, boat, bus, car, motorbike, train
Indoor: bottle, chair, dining table, potted plant, sofa, tv/monitor

The train/val data has 11,530 images containing
27,450 ROI annotated objects and 5,034 segmentations

- Three main competitions: classification, detection, and segmentation
- Three "taster" competitions: person layout, action classification, and ImageNet large scale recognition

M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, A. Zisserman

Classification Results: VOC2010 data

Competition "comp1" (train on VOC2010 data)

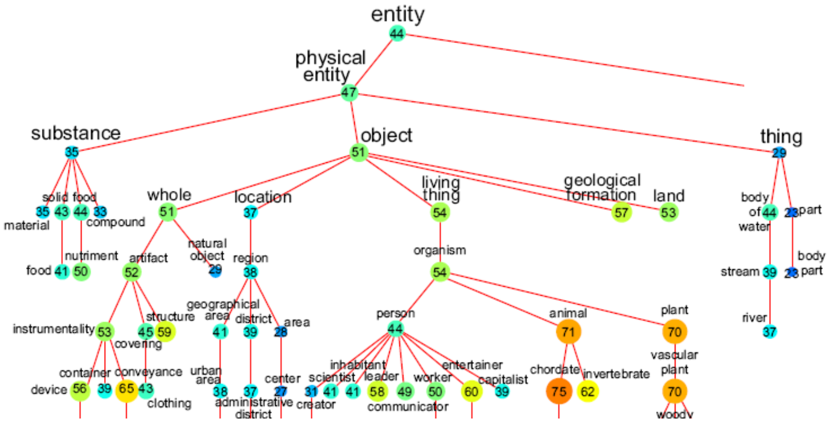
Average Precision (AP %)

	aero plane	bicycle	bird	boat	bottle	bus	car	cat	chair	cow	dining table	dog	horse	motor bike	person	potted plant	sheep	sofa	train	tv/ monitor
BONN_FGT_SEGM	88.0	61.6	53.1	63.3	34.8	77.5	72.3	71.1	41.1	56.0	39.6	64.3	68.9	75.4	87.5	32.5	59.3	40.8	78.7	61.4
BUPT_LPBETA_MULTFEAT	82.1	38.6	39.5	46.5	15.5	55.0	46.4	46.5	39.9	21.3	31.2	37.6	45.8	41.4	75.5	15.6	41.7	25.0	62.5	44.3
BUPT_SPM_SC_HOG	79.6	47.0	42.9	52.3	21.3	66.6	50.1	58.7	44.3	21.8	32.7	46.0	49.7	51.7	72.4	13.2	44.1	28.1	61.5	48.8
BUPT_SVM_MULTFEAT	81.1	45.3	47.3	46.3	20.1	42.3	36.4	49.1	37.5	20.6	38.5	43.8	44.9	54.4	68.6	18.0	48.2	26.0	57.7	40.3
BUT_FU_SVM_SIFT	89.7	63.9	64.5	68.3	36.8	77.9	68.5	72.0	57.2	47.2	56.7	63.5	66.8	74.2	85.0	32.8	54.3	49.1	82.6	66.8
CVC_FLAT	89.4	57.6	63.0	68.5	32.0	76.7	64.7	66.9	51.5	48.4	50.0	54.8	63.1	69.9	83.5	33.6	54.8	46.1	82.2	65.9
CVC_PLUS	91.0	61.8	66.7	71.1	37.7	78.9	67.8	72.2	55.8	51.0	55.8	59.4	65.3	73.0	84.0	39.9	56.9	48.5	83.9	68.1
CVC_PLUSDET	91.7	70.0	66.8	71.3	49.0	81.4	77.5	71.2	60.0	52.6	55.7	61.0	70.9	76.7	88.4	43.2	59.7	53.8	84.7	71.3
HIT_PROTOLEARN_2	60.7	22.1	22.7	29.0	15.0	34.9	27.8	31.6	31.9	14.1	17.4	28.9	24.0	20.6	55.8	9.2	22.0	16.8	30.9	24.6
LIG_MSVM_FUSE_CONCEPT	74.4	43.0	37.5	50.4	22.0	60.7	47.1	46.8	47.5	22.2	35.0	42.1	42.9	48.4	73.8	15.6	31.8	28.9	63.8	46.6
LIP6UPMC_KSVM_BASELINE	78.4	54.1	49.9	61.1	24.6	68.3	58.0	59.9	50.7	35.7	42.5	55.0	60.8	63.1	71.1	25.9	51.5	39.9	74.1	59.6
LIP6UPMC_MKL_L1	78.5	55.9	54.6	62.5	25.0	69.3	59.5	60.0	51.3	37.9	46.7	54.0	60.5	64.0	72.8	32.8	52.6	38.5	72.7	61.1
LIP6UPMC_RANKING	78.8	51.3	46.1	58.2	19.5	68.6	55.6	59.4	46.8	30.7	36.0	49.3	52.3	60.0	76.3	17.8	49.1	35.3	66.3	56.6
LIRIS_MKL_TRAINVAL	87.5	57.0	61.7	68.2	29.9	76.6	61.9	67.5	56.9	35.1	50.6	55.1	62.2	69.3	83.6	35.9	52.9	42.7	79.8	66.3
NEC_V1_HOGLBP_NONLIN_SVM	93.3	71.7	69.9	76.9	42.0	85.3	77.4	79.3	60.0	55.8	60.6	71.1	75.7	77.7	86.8	33.5	61.5	55.8	87.5	69.9
NEC_V1_HOGLBP_NONLIN_SVMDET	93.3	72.9	69.9	77.2	47.9	85.6	79.7	79.4	61.7	56.6	61.1	71.1	76.7	79.3	86.8	38.1	63.9	55.8	87.5	72.9
NII_SVMSIFT	69.3	40.3	27.3	44.1	19.5	54.1	23.9	44.4	42.9	20.3	31.1	37.5	36.6	40.5	68.8	9.3	24.6	20.2	55.6	43.9
NLPR_VSTAR_CLS_DICTLEARN	90.3	77.0	65.3	75.0	53.7	85.9	80.4	74.6	62.9	66.2	54.1	66.8	76.1	81.7	89.9	41.6	66.3	57.0	85.0	74.3
NTHU_LINSPARSE_2	77.9	44.0	37.4	48.5	19.0	63.6	49.0	51.0	45.5	27.6	32.1	41.7	46.9	49.7	68.5	13.2	40.3	30.1	61.7	46.3
NUDT_SVM_LDP_SIFT_PMK_SPMK	86.1	59.3	60.2	68.7	28.7	74.8	63.5	68.0	52.5	41.4	47.1	57.5	60.9	68.2	81.5	29.4	52.1	44.5	79.1	4.7
NUDT_SVM_WHGO_SIFT_CENTRIST_LLM	83.5	54.2	55.2	66.8	28.5	72.1	65.4	64.2	51.9	36.1	49.3	55.6	58.0	66.5	82.1	25.3	48.1	41.7	78.4	59.5
NUSPSL_EXCLASSIFIER	91.3	77.0	70.0	75.6	50.7	83.2	77.1	75.4	62.5	62.6	62.7	64.6	77.9	81.8	91.1	44.8	64.2	53.2	86.3	77.1
NUSPSL_KERNELREGFUSING	93.0	79.0	71.6	77.8	54.3	85.2	78.6	78.8	64.5	64.0	62.7	69.6	82.0	84.4	91.6	48.6	64.9	59.6	89.4	76.4
NUSPSL_MFDETSVM	91.9	77.1	69.5	74.7	52.5	84.3	77.3	76.2	63.0	63.5	62.9	65.0	79.5	83.2	91.2	45.5	65.4	55.0	87.0	77.2
RITSU_CBVR_WKF	85.6	57.2	54.9	64.5	29.2	71.2	57.1	63.2	53.9	37.6	49.6	54.7	58.7	67.9	80.1	29.2	52.1	43.5	76.4	60.9
SURREY_MK_KDA	90.6	66.1	67.2	70.6	36.0	79.7	69.8	73.4	58.4	50.7	60.1	65.2	69.8	76.9	87.0	42.5	59.6	49.9	85.2	71.3
TIT_SIFT_GMM_MKL	87.2	56.6	59.6	66.0	32.6	72.7	63.1	64.8	54.6	41.2	49.3	58.8	59.1	68.2	82.9	31.2	49.2	43.2	75.0	63.4
UC3M_GENDISC	85.5	51.6	55.4	64.8	25.9	74.4	60.6	66.0	51.0	45.9	43.9	55.0	59.0	65.2	80.3	24.0	51.4	47.0	76.4	58.6
UVA_BW_NEWCOLOURSIFT	91.5	71.0	67.3	69.9	43.9	80.6	75.3	73.4	59.3	57.8	60.8	64.0	70.6	80.0	88.6	50.8	65.6	56.1	83.0	76.2
UVA_BW_NEWCOLOURSIFT_SRKDA	90.6	66.9	63.4	70.2	49.4	81.8	76.7	70.9	60.0	57.1	60.5	64.5	67.4	79.1	90.2	53.3	63.5	58.0	81.9	74.4
WLU_SPM_EMDIST	75.8	48.9	36.8	44.3	21.2	65.8	52.1	52.1	45.4	28.2	35.0	45.3	47.8	54.2	71.0	14.7	39.8	32.7	62.2	48.0
XRCE_IFV	87.1	59.6	59.9	69.7	31.3	76.4	62.9	64.3	52.5	42.4	55.1	59.7	64.3	70.4	83.9	32.6	53.3	50.4	80.0	67.6

Slide credit: A. Torralba

80.000.000 tiny images

75.000 non-abstract nouns from WordNet



7 Online image search engines

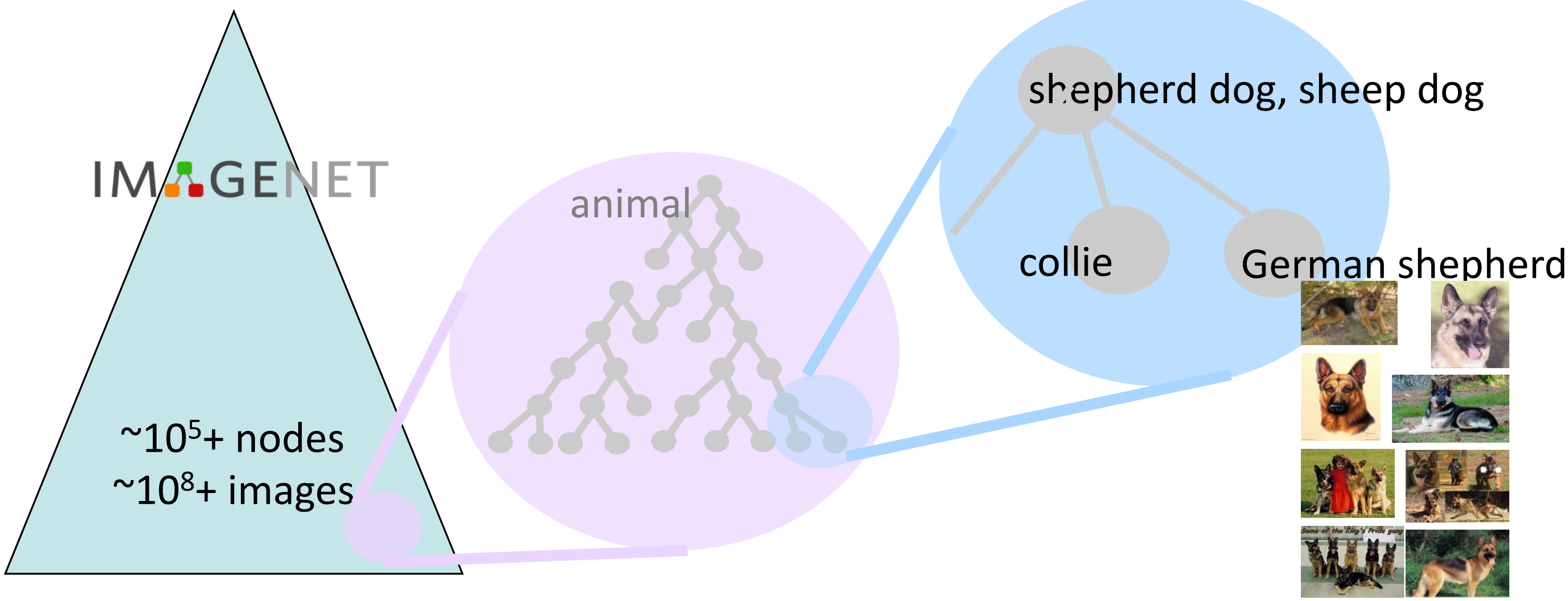
And after 1 year downloading images

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

Slide credit: A. Torralba

IMAGENET

- An ontology of images based on WordNet
- ImageNet currently has
  - 13,000+ categories of visual concepts
  - 10 million human-cleaned images (~700im/categ)
  - 1/3+ is released online @ **www.image-net.org**



Deng, Dong, Socher, Li & Fei-Fei, CVPR 2009

- Collected all the terms from WordNet that described scenes, places, and environments
    - Any concrete noun which could reasonably complete the phrase “I am in a place”, or “let’s go to the place”

- 899 scene categories
- 130,519 images
- 397 scene categories with at least 100 images
- 63,726 labeled objects

J. Xiao, J. Hays, K. Ehinger, A. Oliva, and A. Torralba, CVPR 2010

Collecting datasets (towards $10^{6\text{-}7}$ examples)

- ESP game (CMU)

Luis Von Ahn and Laura Dabbish 2004

- LabelMe (MIT)

Russell, Torralba, Freeman, 2005

- StreetScenes (CBCL-MIT)

Bileschi, Poggio, 2006

- WhatWhere (Caltech)

Perona et al, 2007

- PASCAL challenges (2006-2011)

M. Everingham et al.

- Lotus Hill Institute

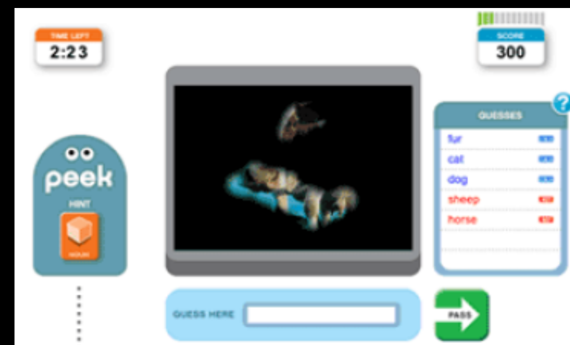
Song-Chun Zhu et al 2007

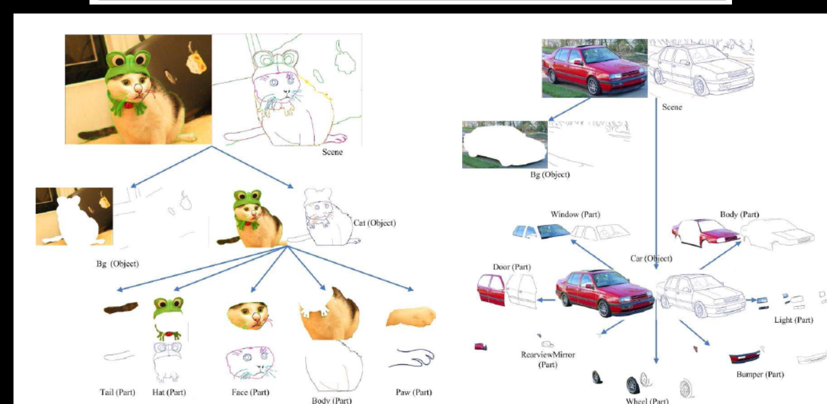
- ImageNet (Stanford)

J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li and L. Fei-Fei

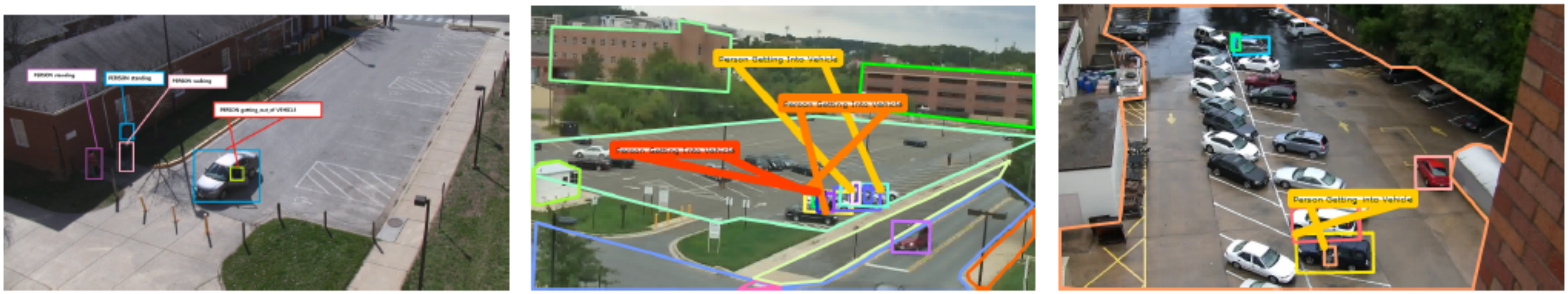
- Tiny images

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

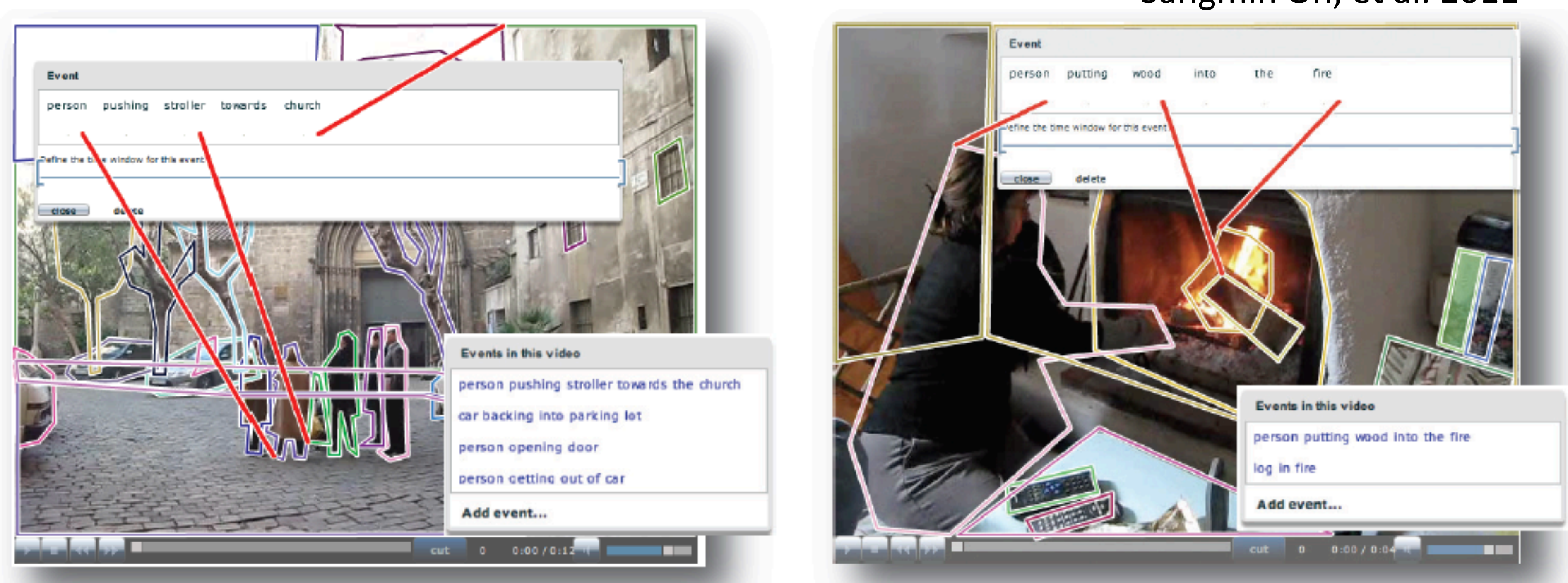




Video: event and action recognition



Sangmin Oh, et al. 2011



J. Yuen, et al. 2009

Video: event and action recognition

	KTH	Weizmann	HOHA 1	TRECVID	VIRAT This Work
# of Event Types	6	10	8	10	23
Avg. # of samples per class	100	9	~85	3~1670	10~1500
Max. Resolution (w x h)	160 x 120	180 x 144	~540 x 240	720 x 576	1920 x 1080
Human Height in Pixels	80~100	60~70	100~1200	20~200	20~180
Human to video height ratio	65~85%	42~50%	50~500%	4~36%	2~20%
# Scenes	N/A	N/A	Many	5	17
Viewpoint Type	Side	Side	Varying	5 / Varying	Varying
Natural Background Clutter	No	No	Yes	Yes	Yes
Incidental Objects/Activities	No	No	Yes, Varying	Yes	Yes
End-to-end Activities	No	No	Yes, Varying	Yes	Yes
Tight Bounding boxes	Cropped	Cropped	No	No	Yes
Multiple annotations on movers	No	No	No	No	Yes
Camera Motion	No	No	Varying	No	Varying
				100 hours	29 hours

Sangmin Oh, et al. 2011

All the following slides are from A. Torralba and A. Ef

Unbiased Look at Dataset Bias

Alyosha Efros (CMU)

Antonio Torralba (MIT)

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

Excesses of the “Data Revolution”

- Are we getting too obsessed with evaluation?
  - The dictatorship of the PR curve over the pixels…
  - Hard to jump out of algorithmic local minima
  - Too much value for “winning” a challenge
  - Easy to overfit over time
- There are all behavioral problems
  - Can be fixed with proper “Best Practices”

Are datasets measuring the right thing?

- In Machine Learning:

  Dataset is The World

- In Recognition

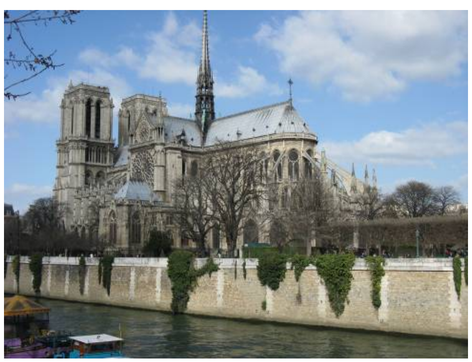
  Dataset is a *representation* of The World

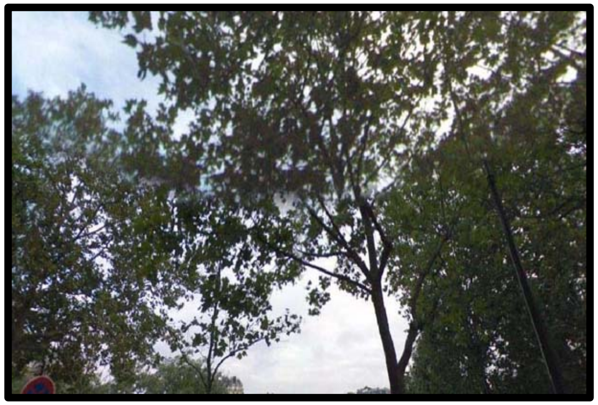
- ML solution: domain transfer
- Vision question: Do datasets provide a good representation?

Visual Data is Inherently Biased

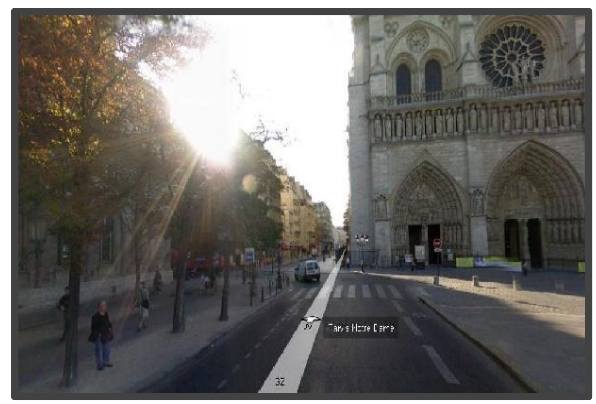
- Internet is a tremendous repository of visual data (Flickr, YouTube, Picassa, etc)
- But it’s <u>not</u> random samples of visual world

Flickr Paris





Google StreetView Paris



Knopp, Sivic, Pajdla, ECCV 2010

Sampled Alyosha's Paris









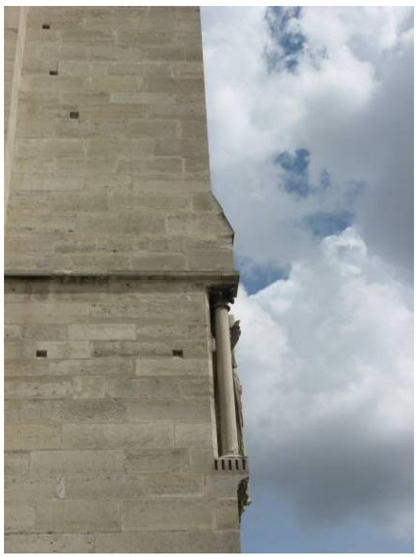



Sampling Bias

- People like to take pictures on vacation

Photographer Bias

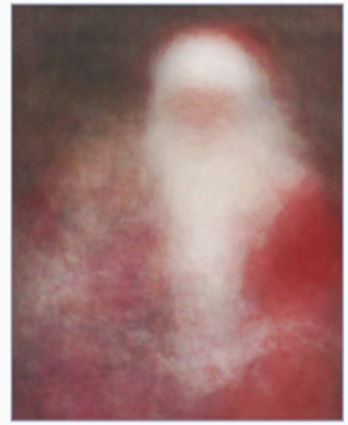
- People want their pictures to be recognizable and/or interesting



vs.

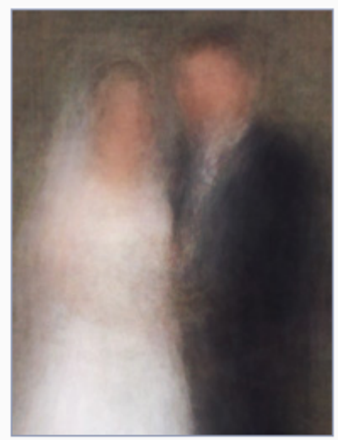
Social Bias

Little Leaguer



Kids with Santa

The Graduate



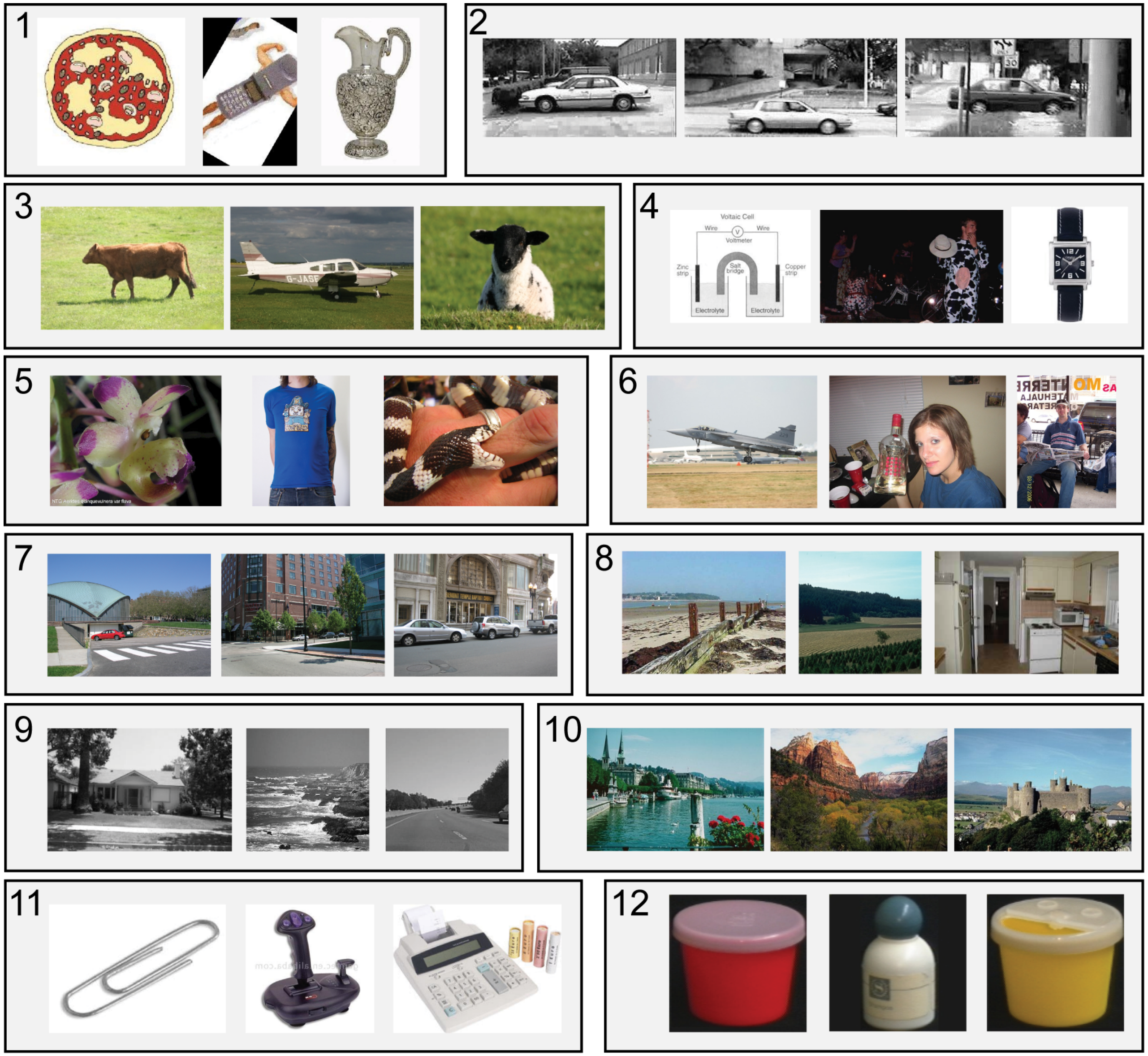
Newlyweds

“100 Special Moments” by Jason Salavon

Our Question

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

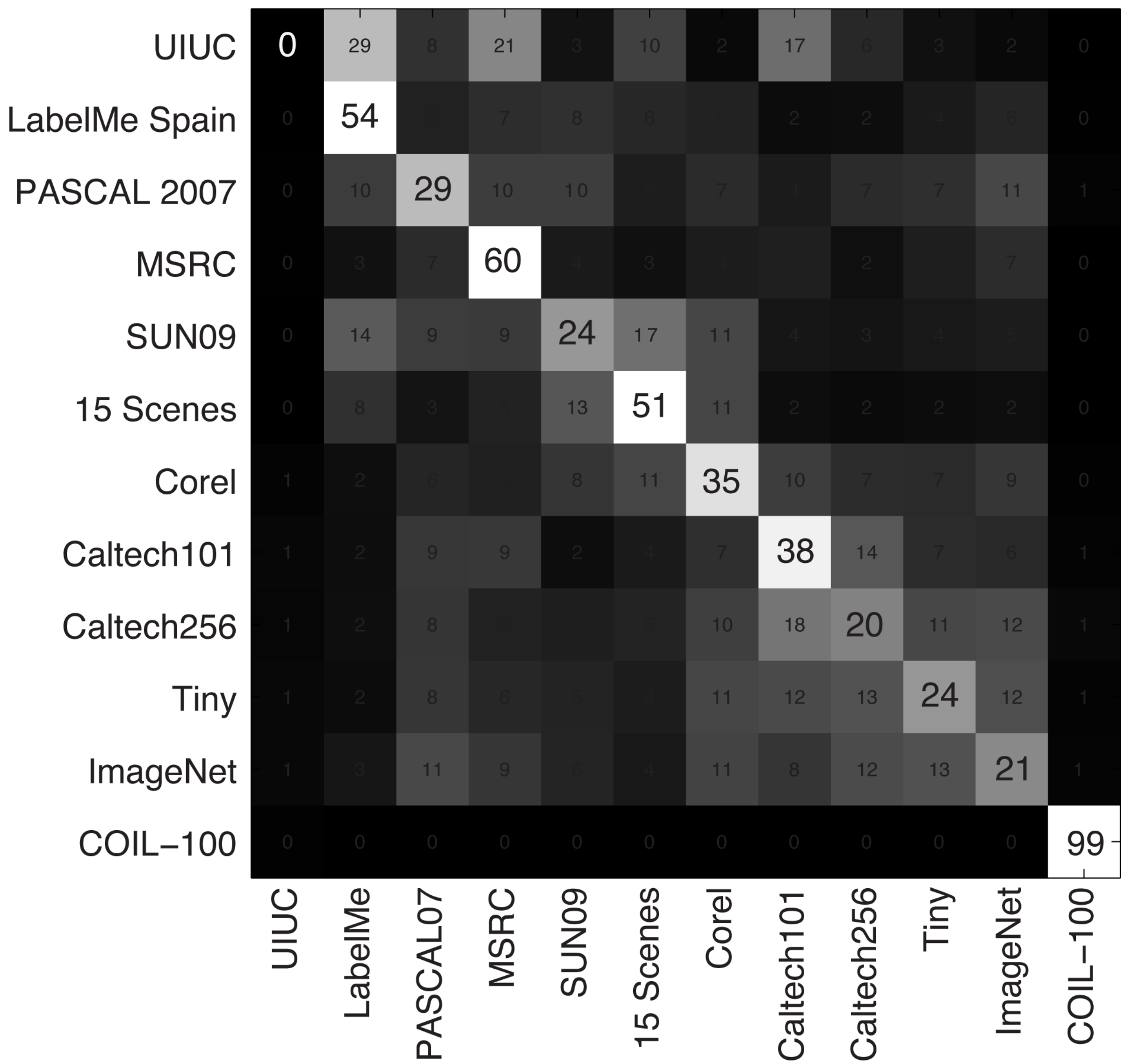
“*Name That Dataset!*” game



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

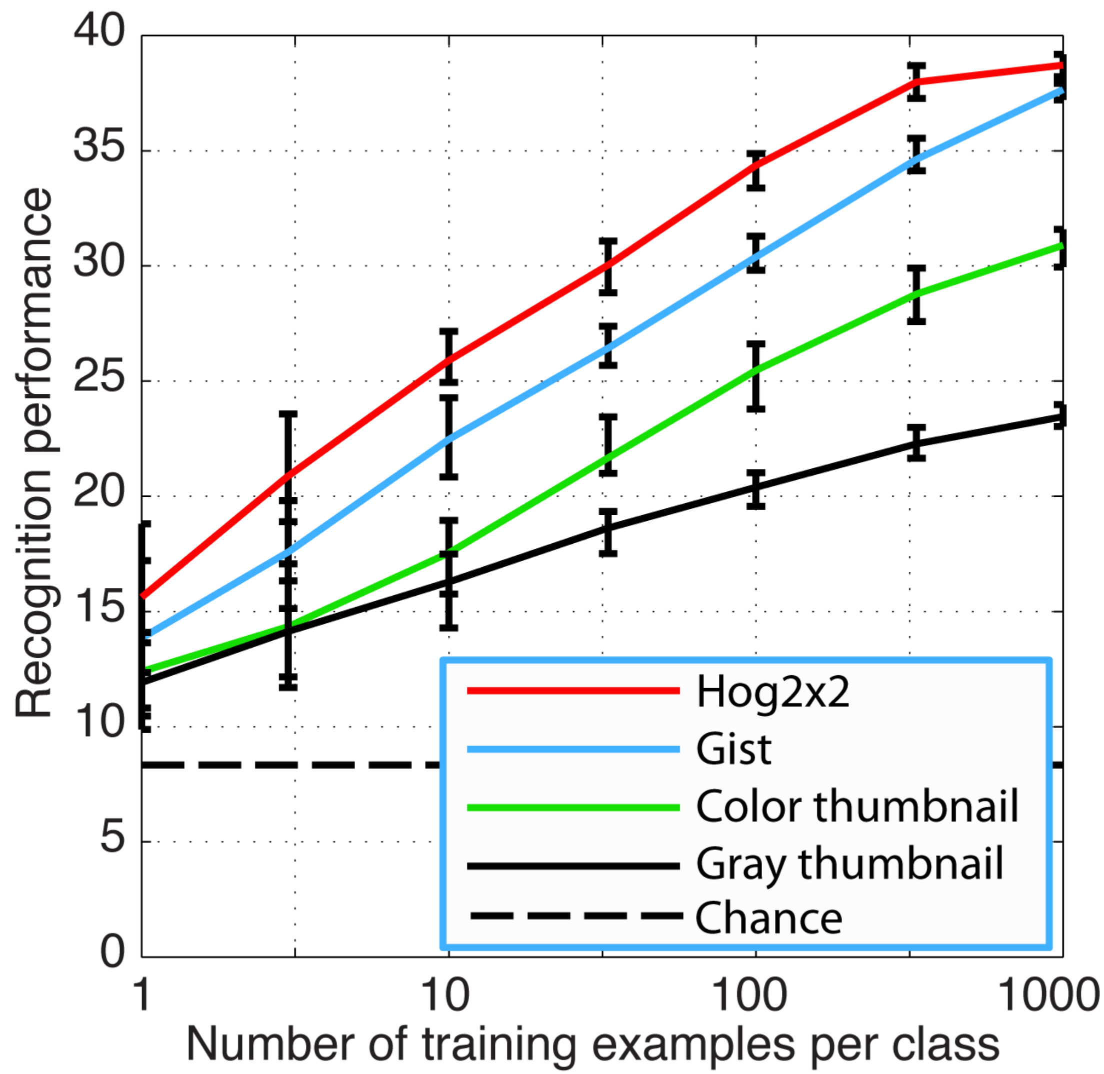
SVM plays *"Name that dataset!"*

SVM plays *“Name that dataset!”*



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

SVM plays *"Name that dataset!"*



Dataset look-alikes

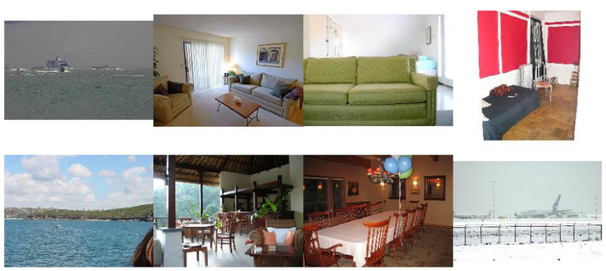
ImageNet pretending to be:

Caltech 256 look-alikes from ImageNet

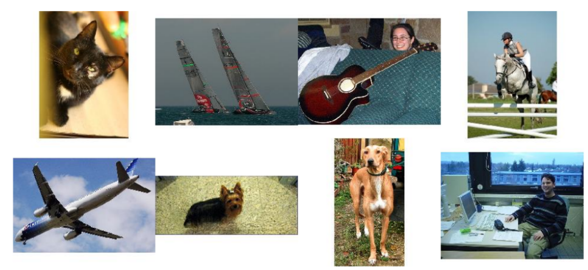

COREL look-alikes from ImageNet


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

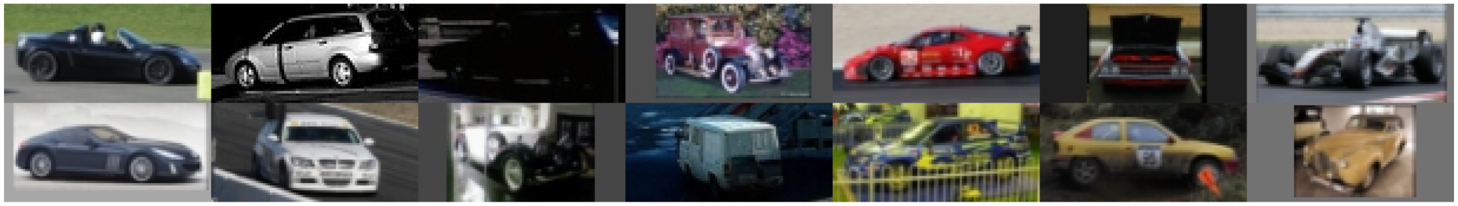
Datasets have different goals…

- 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

PASCAL cars



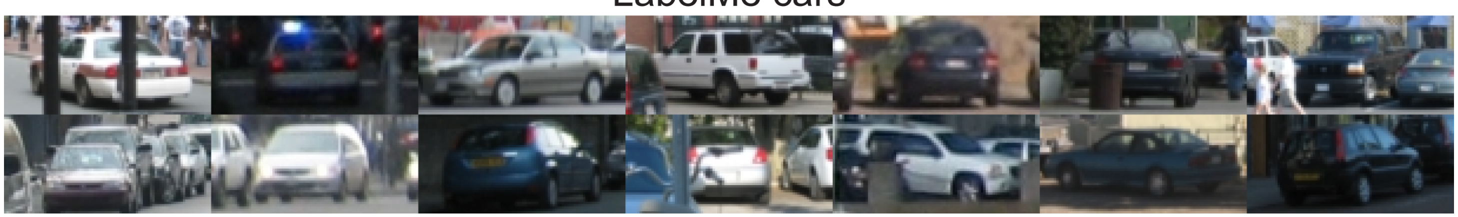
SUN cars

Caltech101 cars

Performance: 61%
(chance: 20%)

ImageNet cars

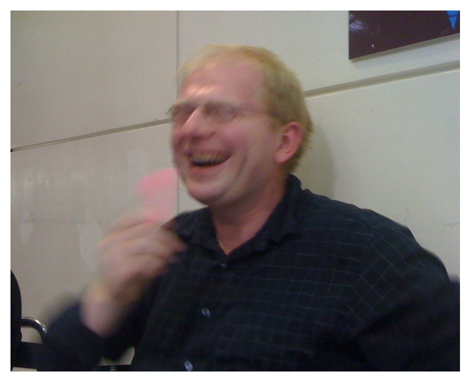
LabelMe cars



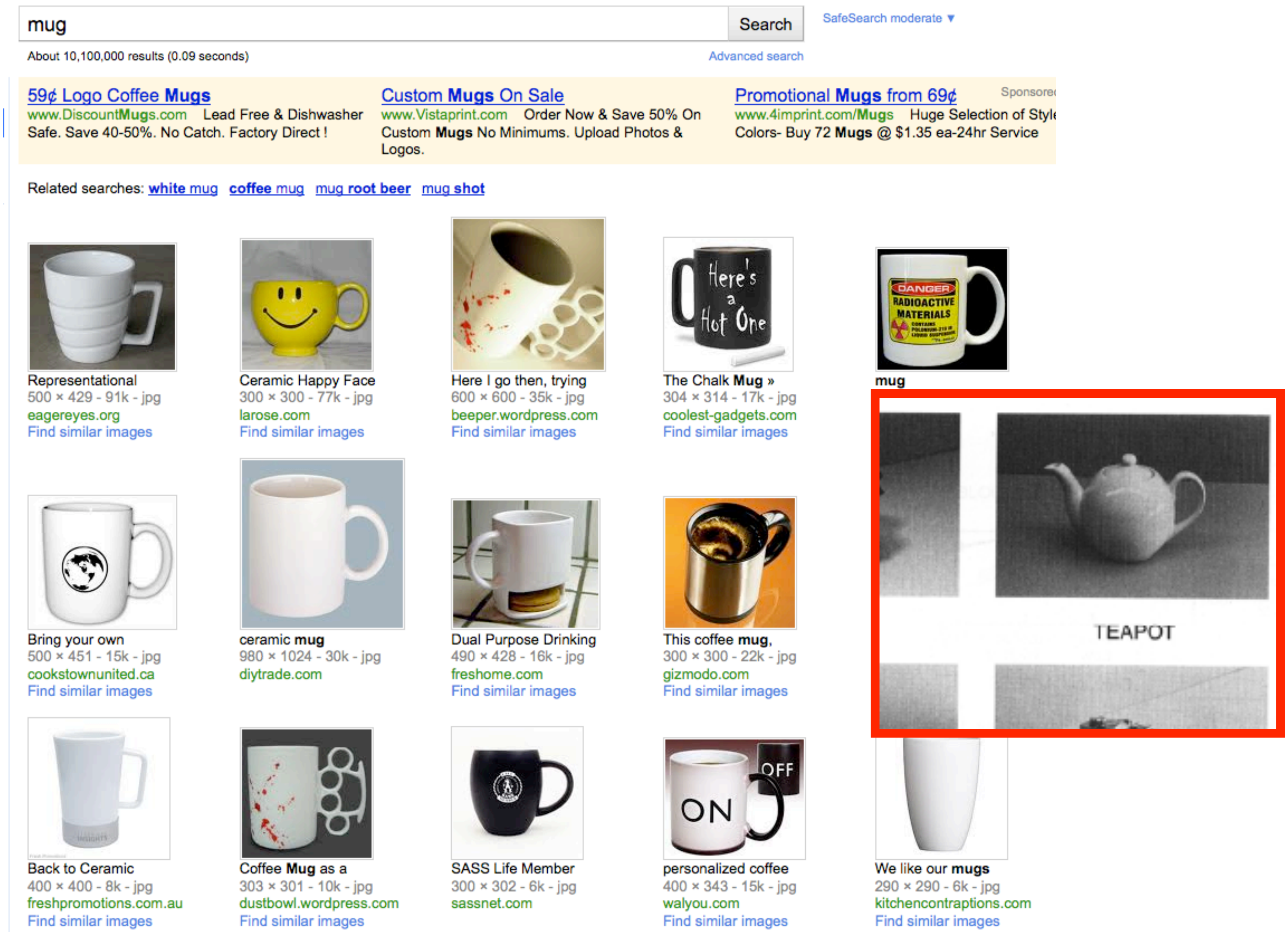
Where do this bias comes from?

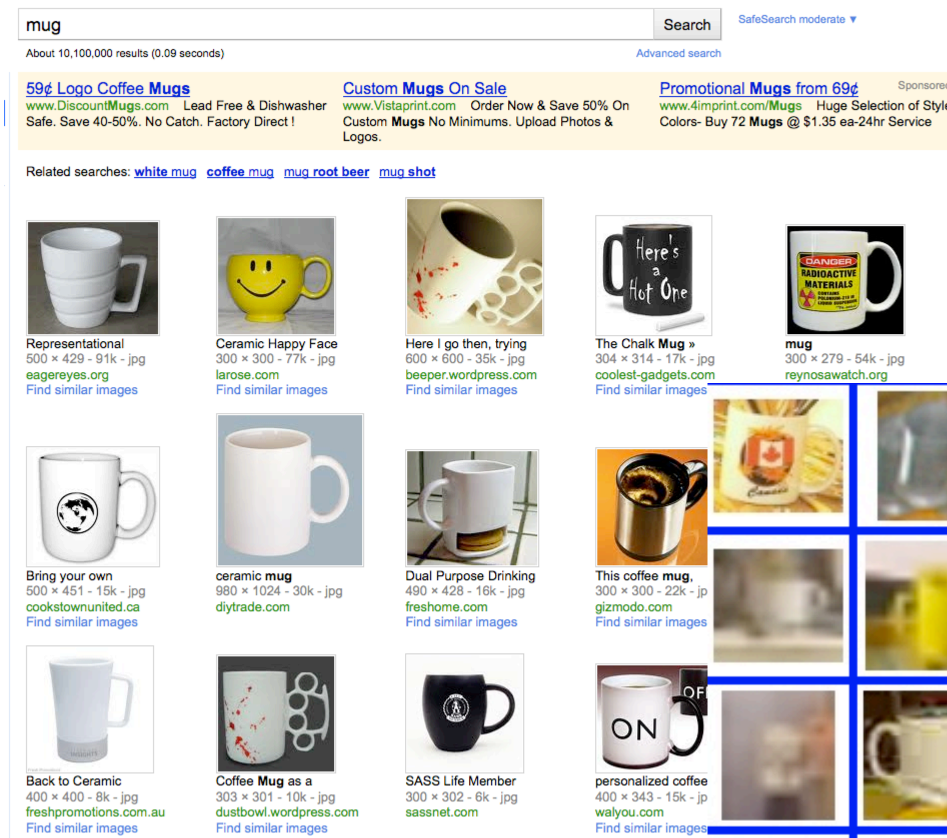
Some bias is in the world

Some bias is in the world



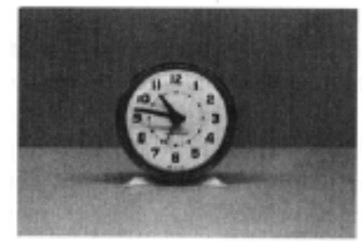
Some bias comes from the way the data is collected



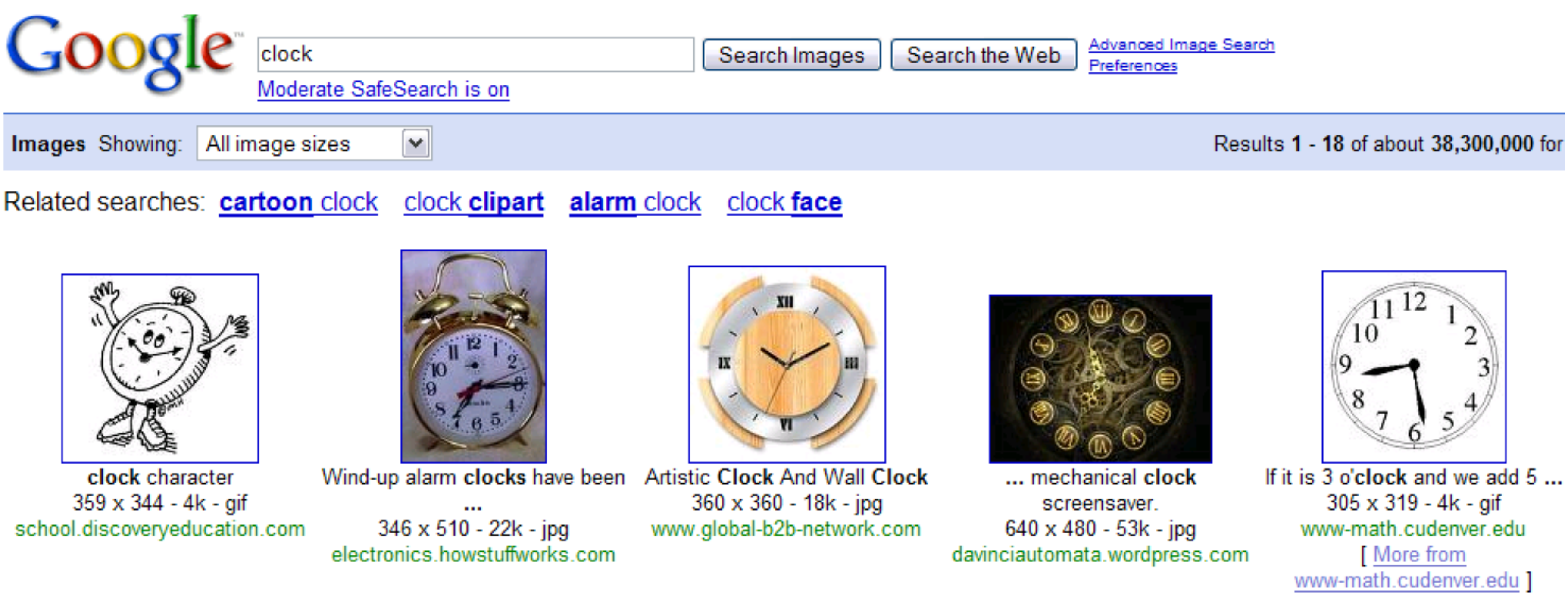


Google mugs

Mugs from LabelMe



CLOCK


Google
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]

Measuring Dataset Bias

Cross-Dataset Generalization

MSRC

Classifier trained on MSRC cars

Cross-dataset Performance

Table 1. Cross-dataset generalization. Object detection and classification performance (AP) for "car" and "person" when training on one dataset (rows) and testing on another (columns), i.e. each row is: training on one dataset and testing on all the others. "Self" refers to training and testing on the same dataset (same as diagonal), and "Mean Others" refers to averaging performance on all except self.

task	Test on: / Train on:	SUN09	LabelMe	PASCAL	ImageNet	Caltech101	MSRC	Self	Mean others	Percent drop
"car" classification	SUN09	**28.2**	29.5	16.3	14.6	16.9	21.9	28.2	19.8	**30%**
	LabelMe	14.7	**34.0**	16.7	22.9	43.6	24.5	34.0	24.5	**28%**
	PASCAL	10.1	25.5	**35.2**	43.9	44.2	39.4	35.2	32.6	**7%**
	ImageNet	11.4	29.6	36.0	**57.4**	52.3	42.7	57.4	34.4	**40%**
	Caltech101	7.5	31.1	19.5	33.1	**96.9**	42.1	96.9	26.7	**73%**
	MSRC	9.3	27.0	24.9	32.6	40.3	**68.4**	68.4	26.8	**61%**
	Mean others	10.6	28.5	22.7	29.4	39.4	34.1	53.4	27.5	48%
"car" detection	SUN09	**69.8**	50.7	42.2	42.6	54.7	69.4	69.8	51.9	**26%**
	LabelMe	61.8	**67.6**	40.8	38.5	53.4	67.0	67.6	52.3	**23%**
	PASCAL	55.8	55.2	**62.1**	56.8	54.2	74.8	62.1	59.4	**4%**
	ImageNet	43.9	31.8	46.9	**60.7**	59.3	67.8	60.7	49.9	**18%**
	Caltech101	20.2	18.8	11.0	31.4	**100**	29.3	100	22.2	**78%**
	MSRC	28.6	17.1	32.3	21.5	67.7	**74.3**	74.3	33.4	**55%**
	Mean others	42.0	34.7	34.6	38.2	57.9	61.7	72.4	44.8	**48%**
"person" classification	SUN09	**16.1**	11.8	14.0	7.9	6.8	23.5	16.1	12.8	**20%**
	LabelMe	11.0	**26.6**	7.5	6.3	8.4	24.3	26.6	11.5	**57%**
	PASCAL	11.9	11.1	**20.7**	13.6	48.3	50.5	20.7	27.1	**-31%**
	ImageNet	8.9	11.1	11.8	**20.7**	76.7	61.0	20.7	33.9	**-63%**
	Caltech101	7.6	11.8	17.3	22.5	**99.6**	65.8	99.6	25.0	**75%**
	MSRC	9.4	15.5	15.3	15.3	93.4	**78.4**	78.4	29.8	**62%**
	Mean others	9.8	12.3	13.2	13.1	46.7	45.0	43.7	23.4	**47%**
"person" detection	SUN09	**69.6**	56.8	37.9	45.7	52.1	72.7	69.6	53.0	**24%**
	LabelMe	58.9	**66.6**	38.4	43.1	57.9	68.9	66.6	53.4	**20%**
	PASCAL	56.0	55.6	**56.3**	55.6	56.8	74.8	56.3	59.8	**-6%**
	ImageNet	48.8	39.0	40.1	**59.6**	53.2	70.7	59.6	50.4	**15%**
	Caltech101	24.6	18.1	12.4	26.6	**100**	31.6	100	22.7	**77%**
	MSRC	33.8	18.2	30.9	20.8	69.5	**74.7**	74.7	34.6	**54%**
	Mean others	44.4	37.5	31.9	38.4	57.9	63.7	71.1	45.6	**36%**

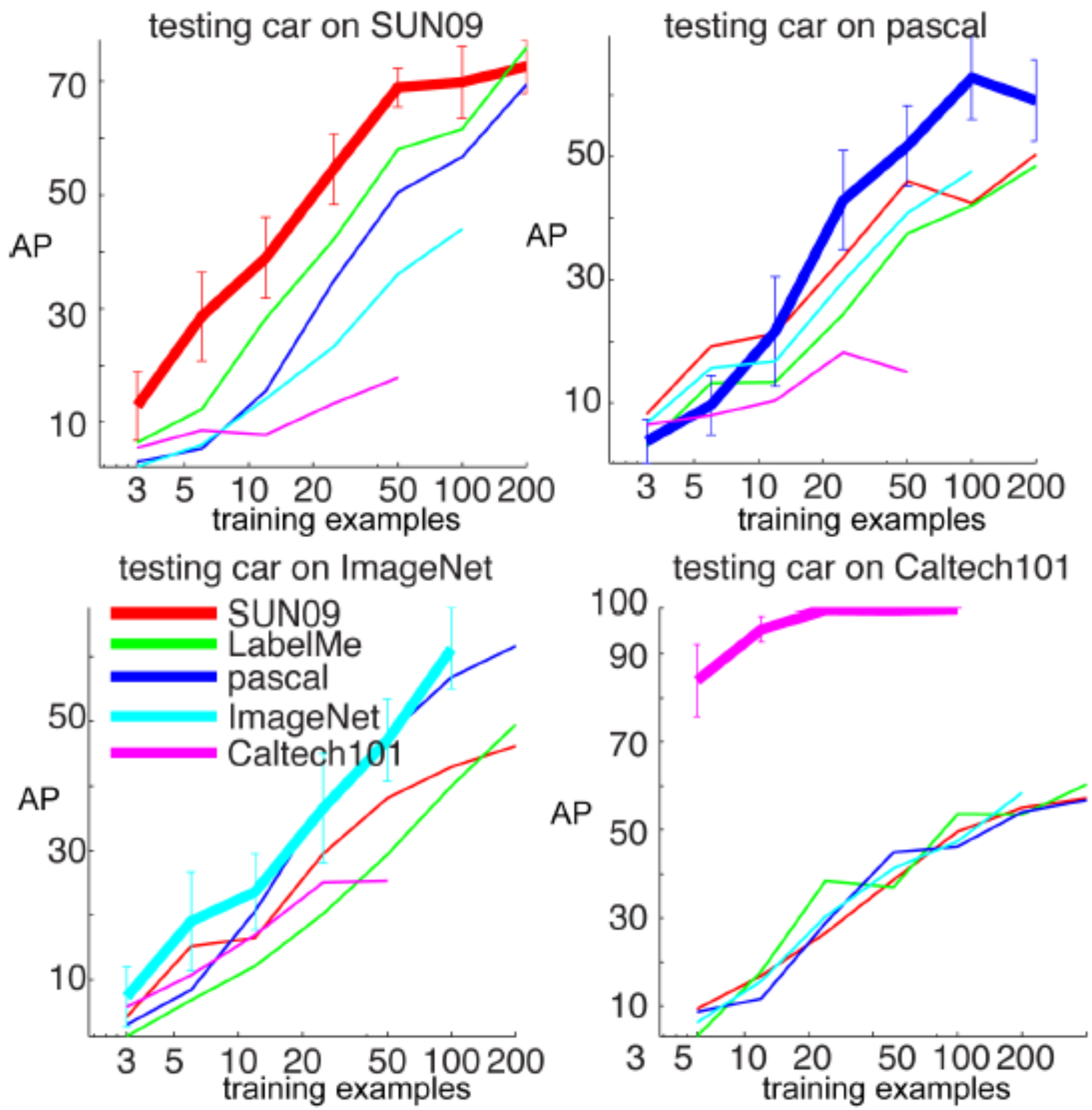


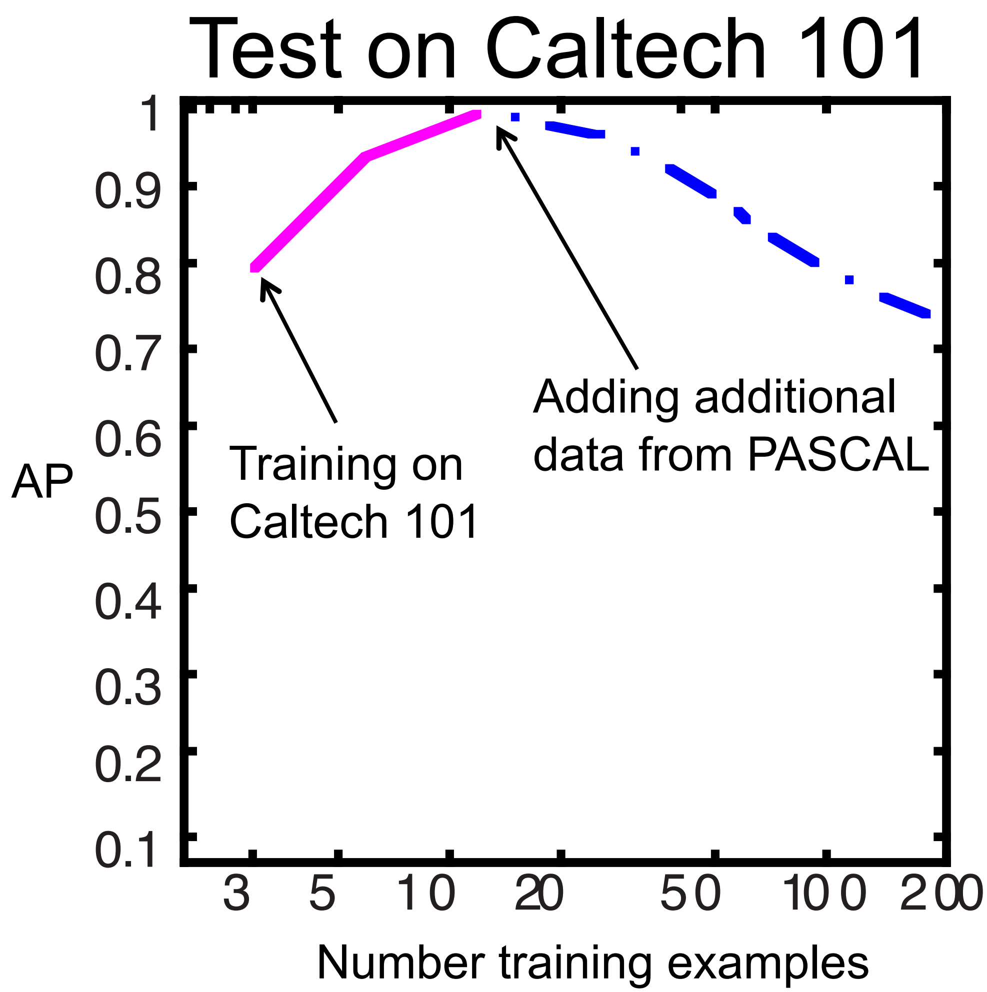
Figure 6. Cross-dataset generalization for "car" detection as function of training data

Dataset Value

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

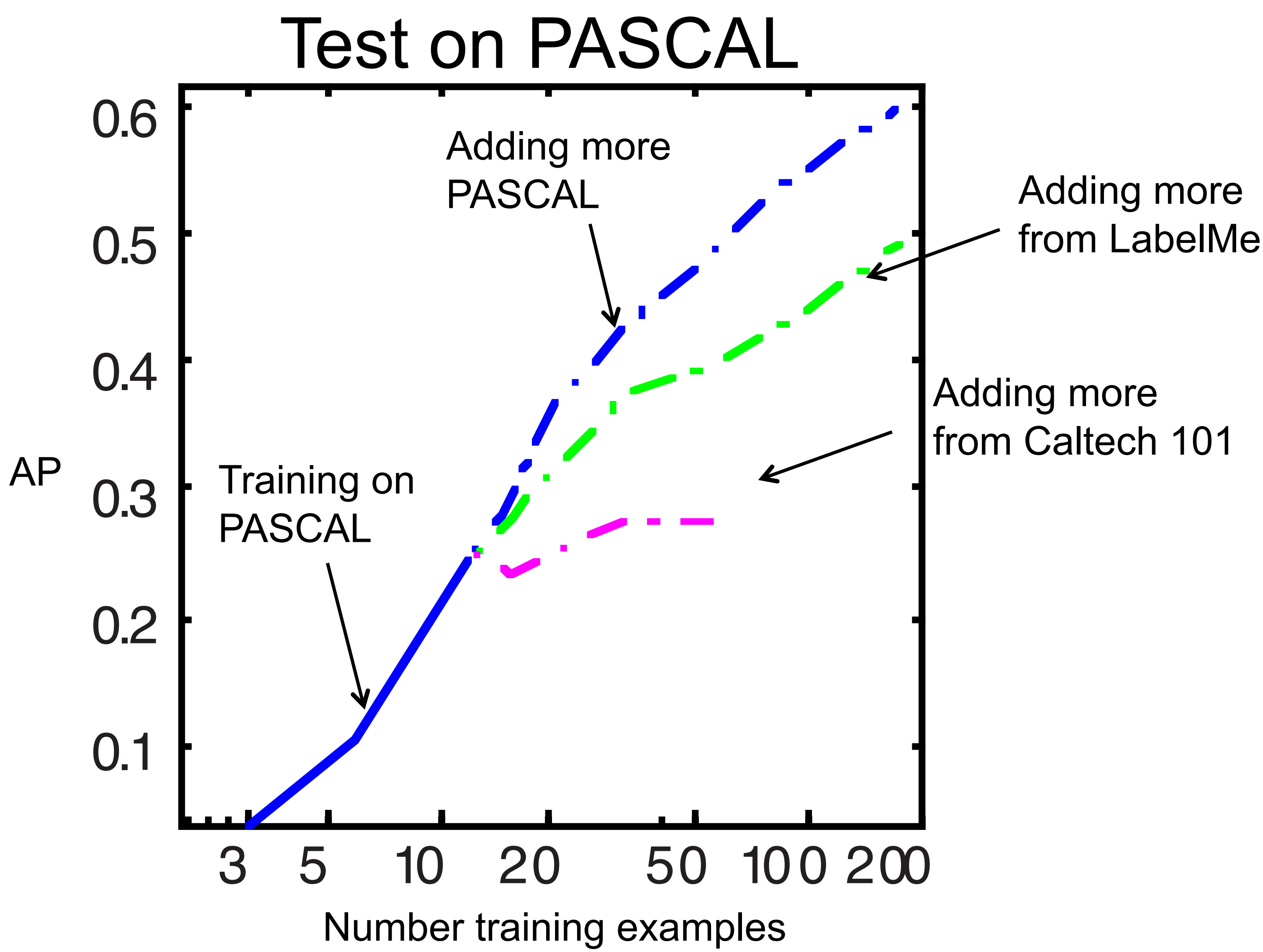
	SUN09 market	LabelMe market	PASCAL market	ImageNet market	Caltech101 market
1 SUN09 is worth	1 SUN09	0.91 LabelMe	0.72 pascal	0.41 ImageNet	0 Caltech
1 LabelMe is worth	0.41 SUN09	1 LabelMe	0.26 pascal	0.31 ImageNet	0 Caltech
1 pascal is worth	0.29 SUN09	0.50 LabelMe	1 pascal	0.88 ImageNet	0 Caltech
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

Mixing datasets



Task: car detection
Features: HOG

Mixing datasets



Negative Set Bias

Table 2. Measuring Negative Set Bias.

task	Positive Set: / Negative Set:	SUN09	LabelMe	PASCAL	ImageNet	Caltech101	MSRC	Mean
"car" detection	self	67.6	62.4	56.3	60.5	97.7	74.5	70.0
	all	53.8	51.3	47.1	65.2	97.7	70.0	64.1
	percent drop	20%	18%	16%	-8%	0%	6%	8%
"person" detection	self	67.4	68.6	53.8	60.4	100	76.7	71.1
	all	52.2	58.0	42.6	63.4	100	71.5	64.6
	percent drop	22%	15%	21%	-5%	0%	7%	9%

Not all the bias comes from the appearance of the objects we care about

Overall…

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

Causes for Pessimism

- Our best-performing techniques just don't work in the real world
  - E.g. try a person detector on Hollywood film
- The classifiers are inherently designed to overfit to type of data it's trained on.

  - we just don't have enough negative data to present this…

Causes for Optimism

- We are getting better. The new datasets are better than the old ones.
- Large dataset trend will alleviate this trend.

Summary

- Until now datasets are used to evaluate algorithms, but nobody has dared to evaluate them. Let's evaluate datasets.

Four Stages of Dataset Grief