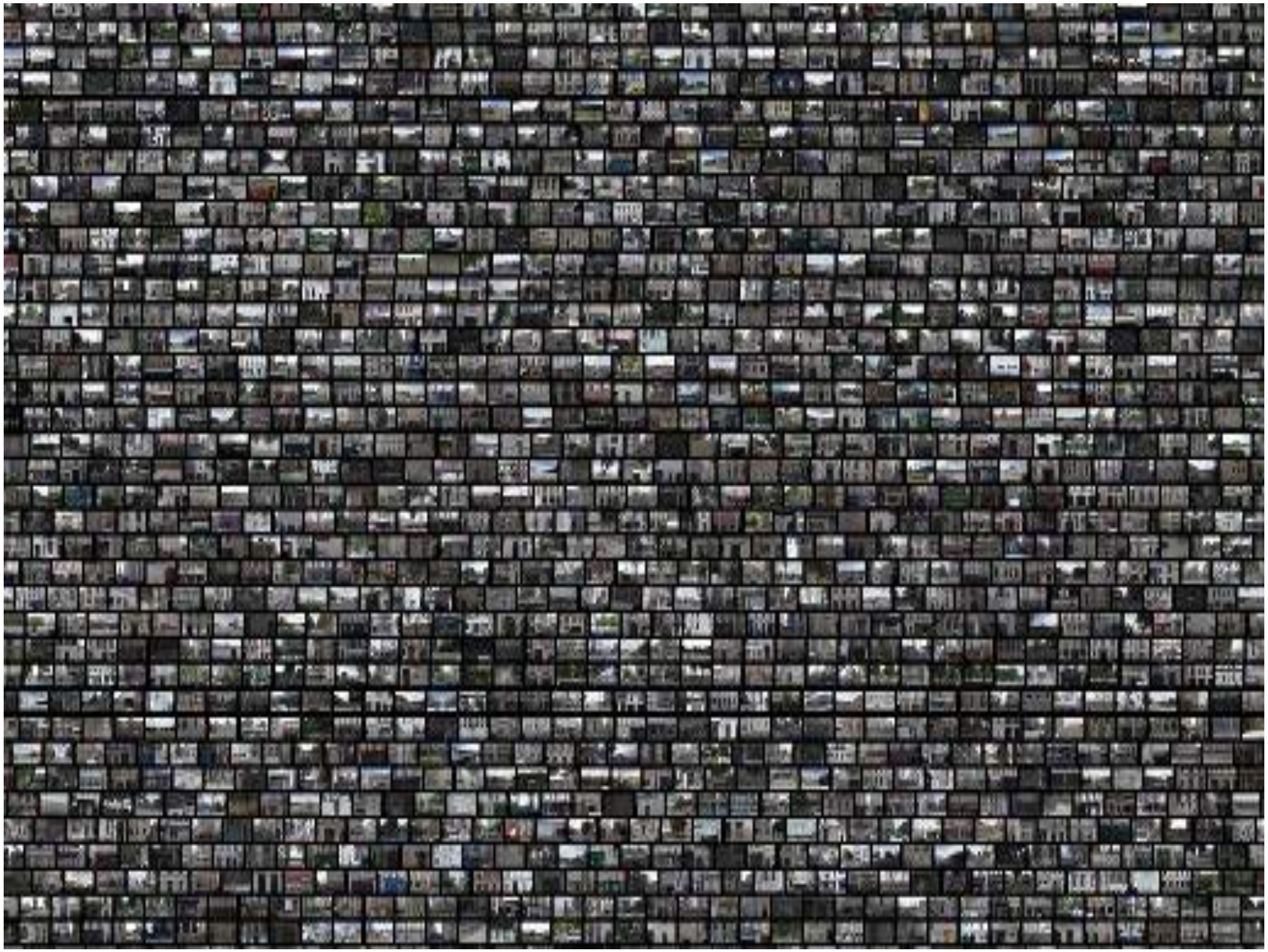
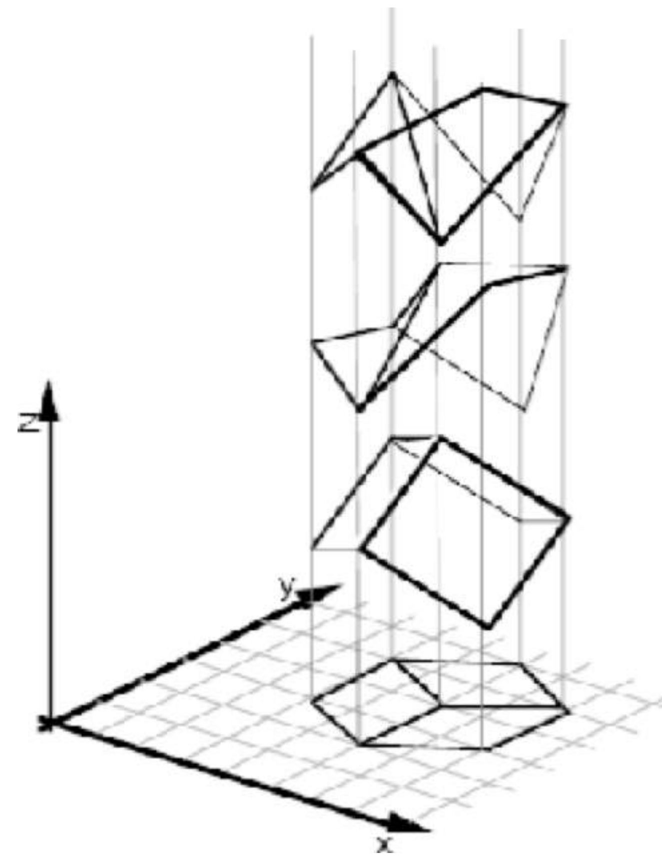
The Promise and Perils of Big Data



Some Slides from A. Gupta, A. Efros and A. Torralba

Why do we need data?

Most vision problems are ambiguous



So, how do we solve these problems?

- Magic of data !
- Use data to learn better likelihoods: how things look like.
- Use data to learn priors of what is more likely than others.

But how much data do we need?

The extremes of learning

Extrapolation problem
Generalization
Transfer learning

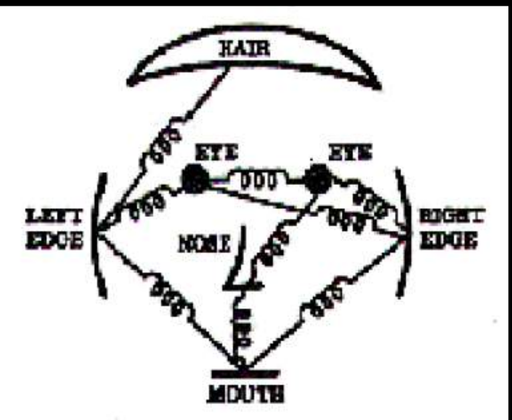
Interpolation problem
Correspondence
Finding the differences

Current Datasets

$1 \quad 10 \quad 10^2 \quad 10^3 \quad 10^4 \quad 10^{6\text{-}7} \quad 10^{10} \quad \infty$

Number of training samples

Datasets before 2012



So how much data does humans use?

What's the Capacity of Visual Long Term Memory?

What we know...

Standing (1973)

10,000 images

83% Recognition

... people can remember thousands of images

High Fidelity Visual Memory is possible (Hollingworth 2004)

What we don't know...

... what people are remembering for each item?

According to Standing

"Basically, my recollection is that we just separated the pictures into distinct thematic categories: e.g. cars, animals, single-person, 2-people, plants, etc.) Only a few slides were selected which fell into each category, and they were visually distinct."

Dogs
Playing Cards

"Gist" Only

Sparse Details

Highly Detailed

Slide by Aude Oliva

how much data does computer vision researchers use?

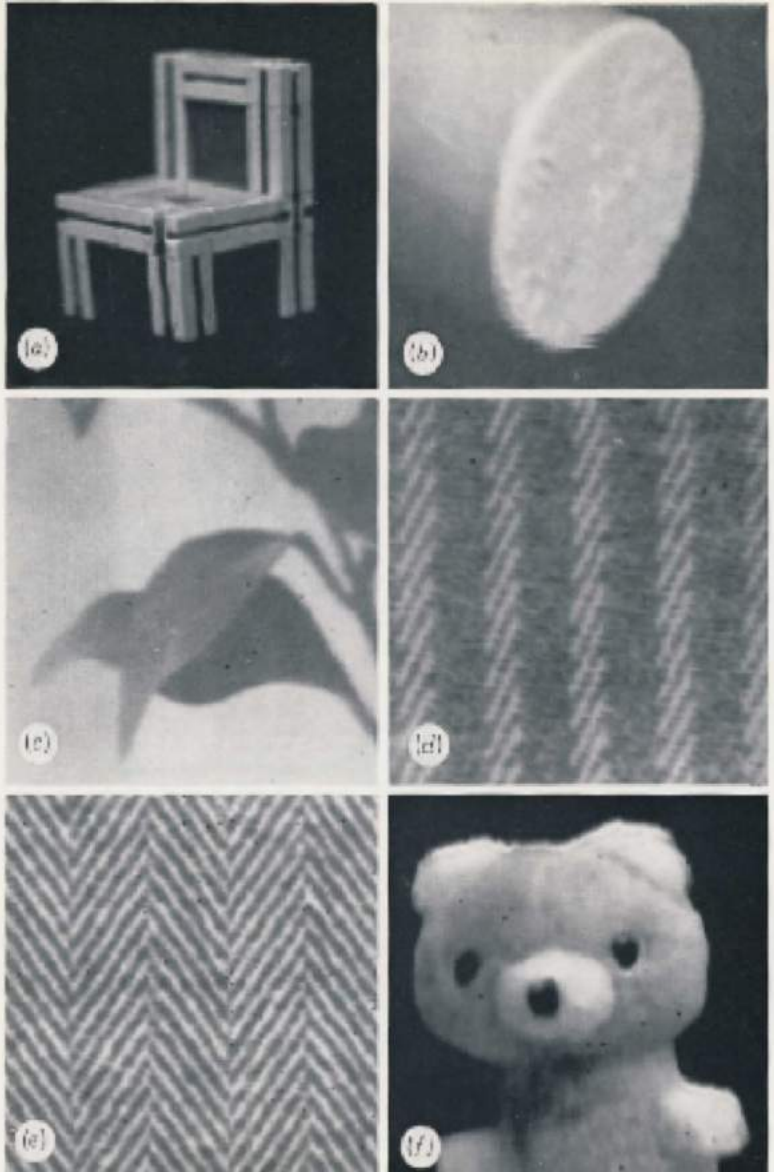
10^0
images

1972

10^1 images

1

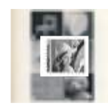
images



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 P1500hPhotowriter 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

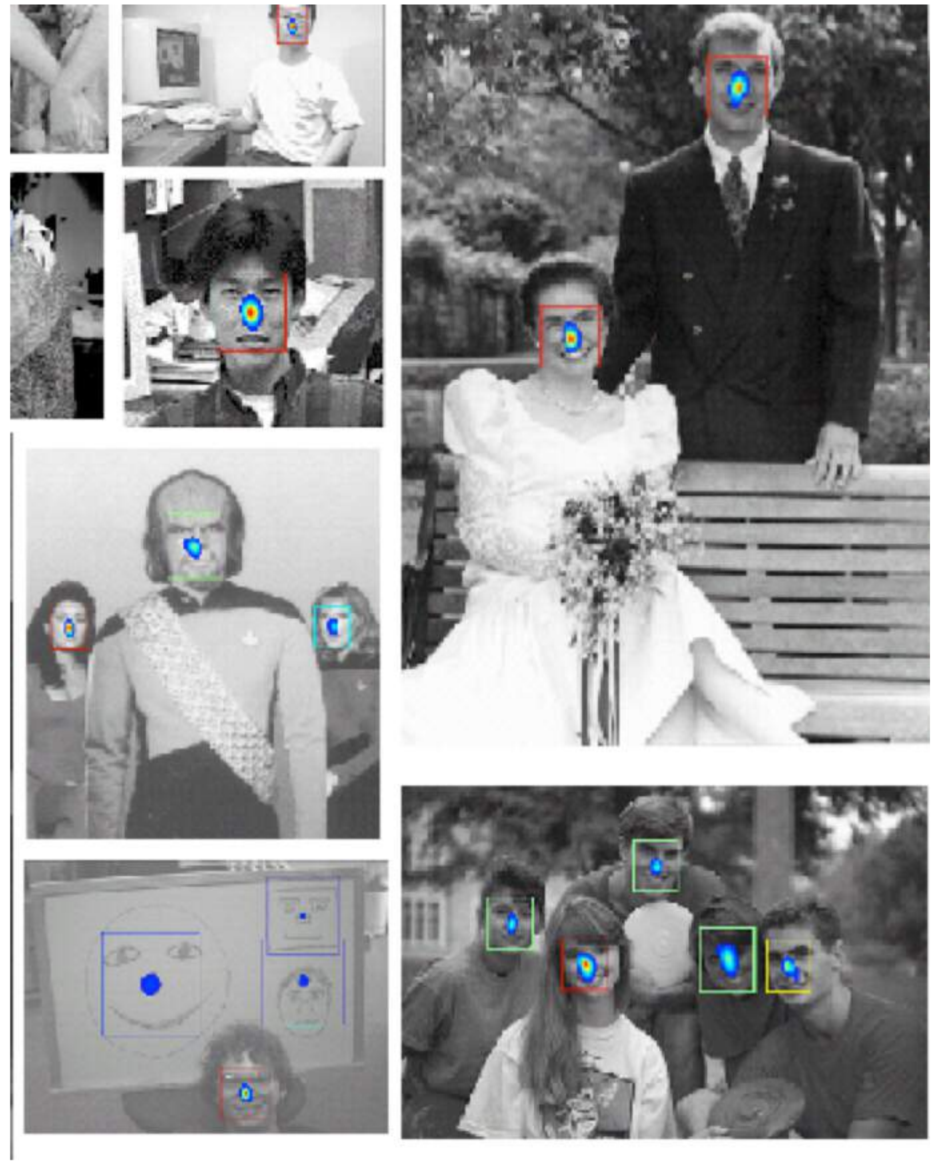
$10^{2\text{-}4}$ images



The faces and cars scale

$10^{2\text{-}4}$ images

In 1996 DARPA released 14000 images, from over 1000 individuals.

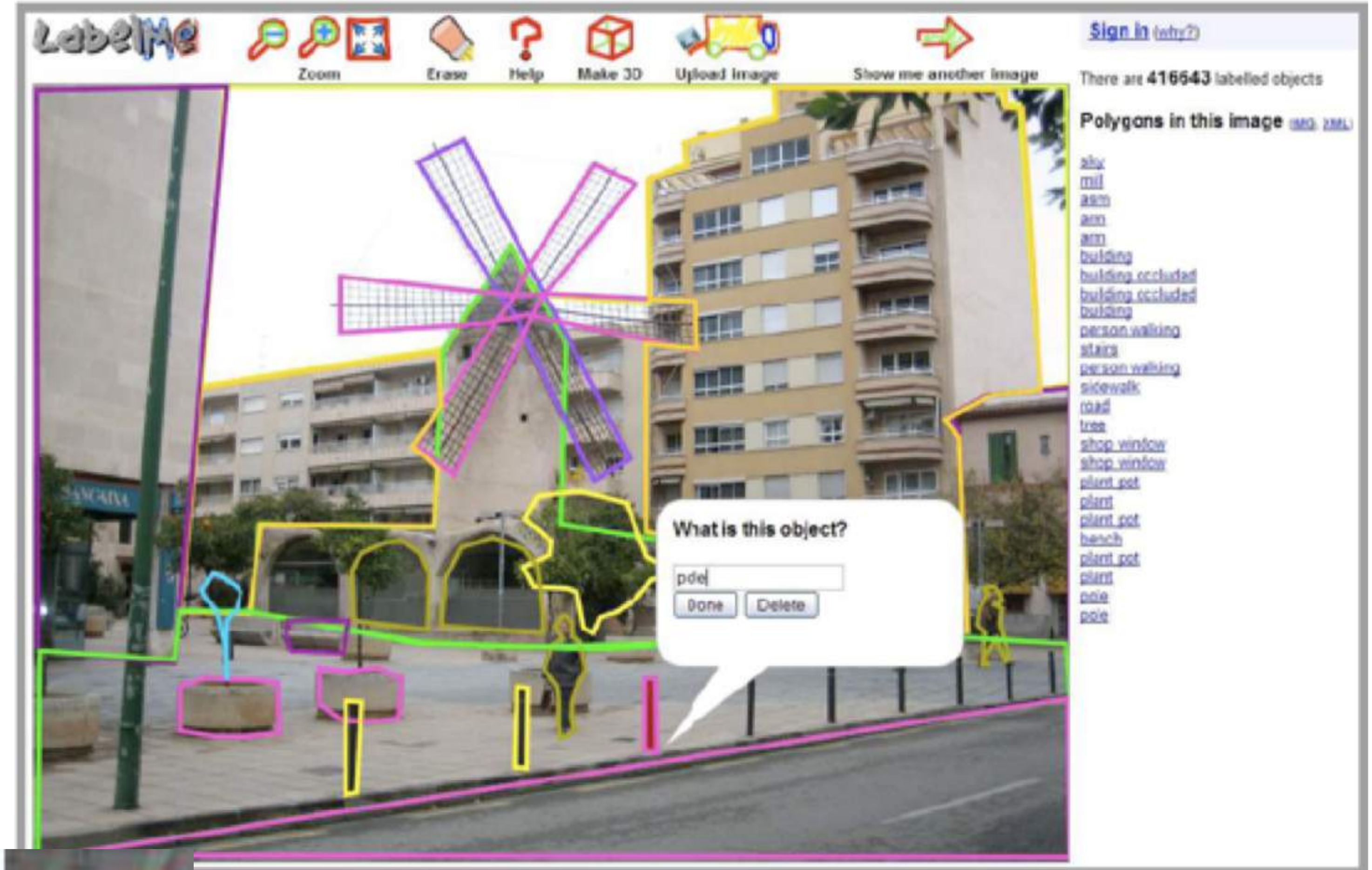


$10^{2\text{-}4}$
images

10^5
images

LabelMe

10^5
images



Tool went online July 1st, 2005
530,000 object annotations collected

Labelme.csail.mit.edu

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

2004-2006 Caltech 101 and 256







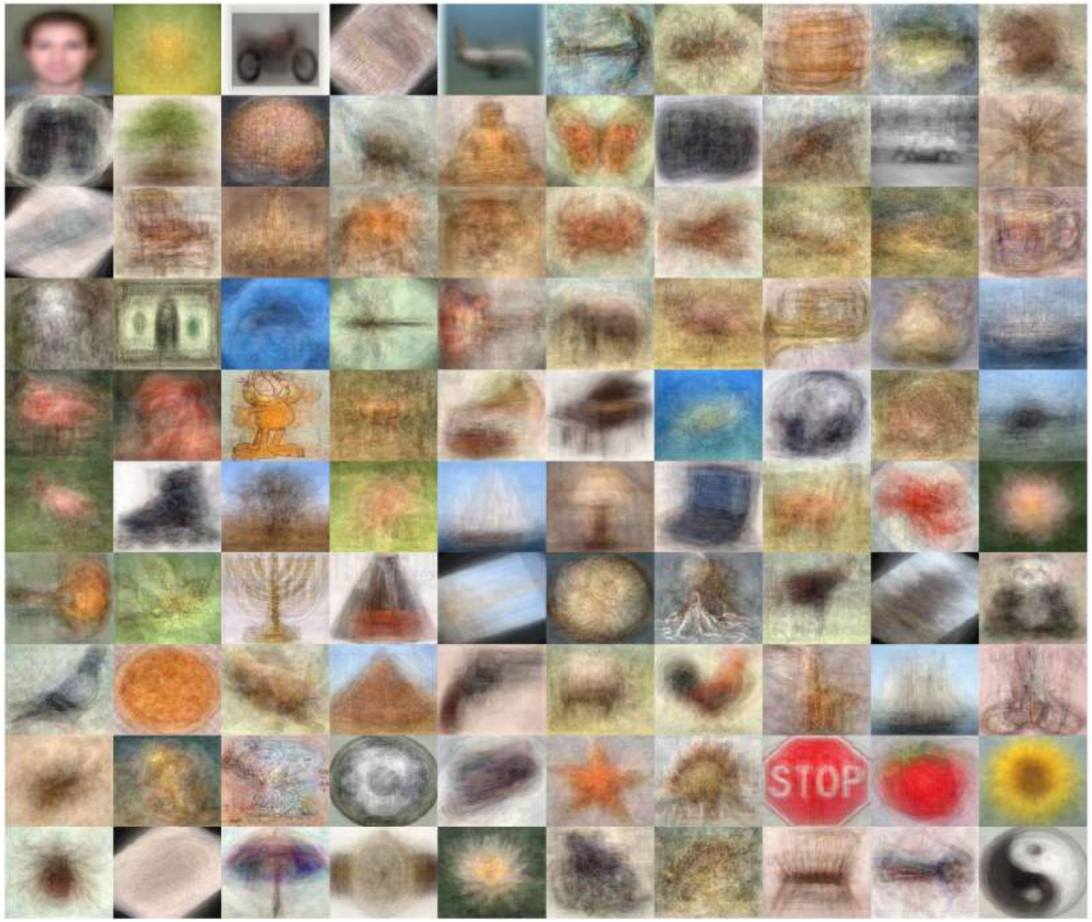







2006 80 Million Tiny Images

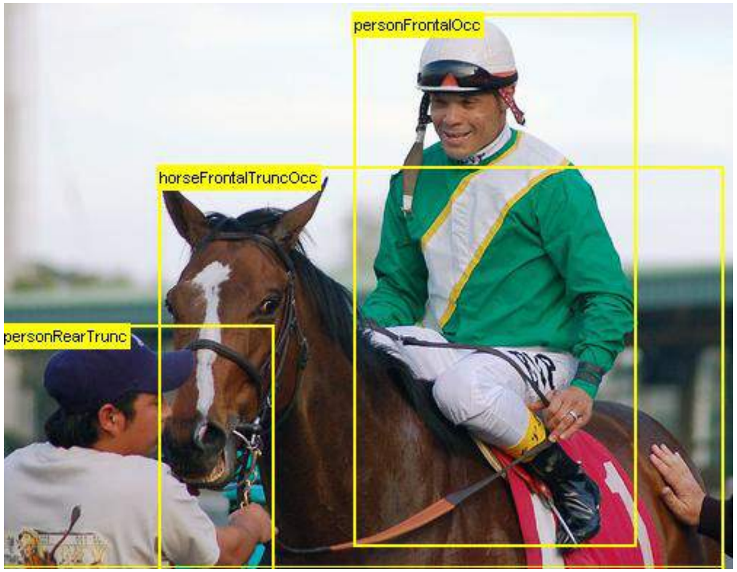
15-30 training images, up to ~70% accuracy.



Antonio Torralba

2007 PASCAL VOC

20 classes

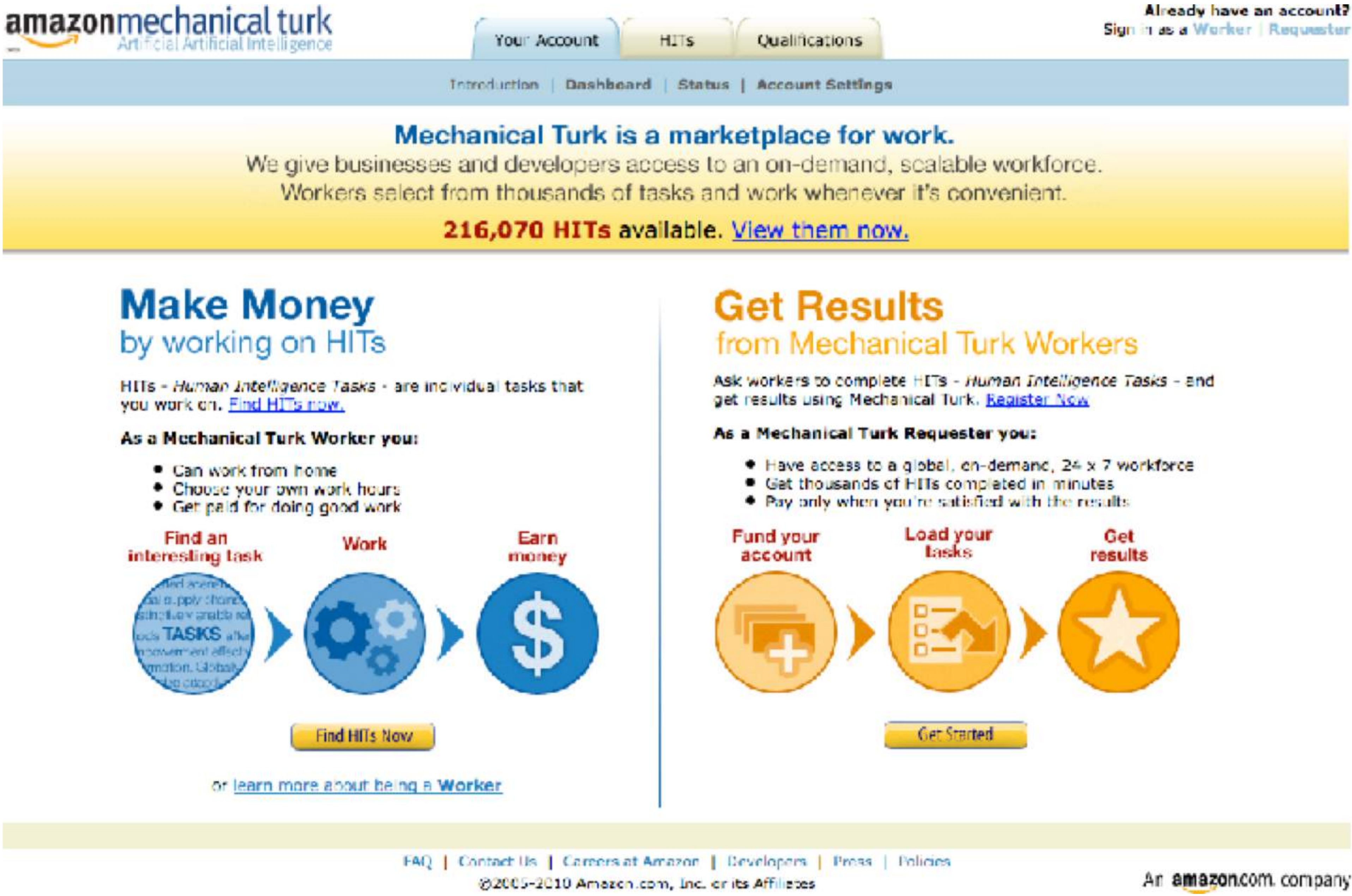


The PASCAL Visual Object Classes (VOC) Challenge, Everingham, Van Gool, Williams, Winn and Zisserman, *IJCV*, 2010

2009 Caltech Pedestrian

1 class, lots of instances.

Pedestrian Detection: An Evaluation of the State of the Art,
Dollár, Wojek, Schiele and Perona, *PAMI*, 2012



amazonmechanical turk
Artificial Artificial Intelligence

Your Account | HITs | Qualifications

Already have an account?
Sign in as a Worker | Requester

Introduction | Dashboard | Status | Account Settings

Mechanical Turk is a marketplace for work.
We give businesses and developers access to an on-demand, scalable workforce.
Workers select from thousands of tasks and work whenever it's convenient.

216,070 HITs available. View them now.

Make Money
by working on HITs

HITs - *Human Intelligence Tasks* - are individual tasks that you work on. Find HITs now.

As a Mechanical Turk Worker you:

- Can work from home
- Choose your own work hours
- Get paid for doing good work

Find an interesting task → Work → Earn money

Find HITs Now

or learn more about being a **Worker**

Get Results
from Mechanical Turk Workers

Ask workers to complete HITs - *Human Intelligence Tasks* - and get results using Mechanical Turk. Register Now

As a Mechanical Turk Requester you:

- Have access to a global, on-demand, 24 x 7 workforce
- Get thousands of HITs completed in minutes
- Pay only when you're satisfied with the results

Fund your account → Load your tasks → Get results

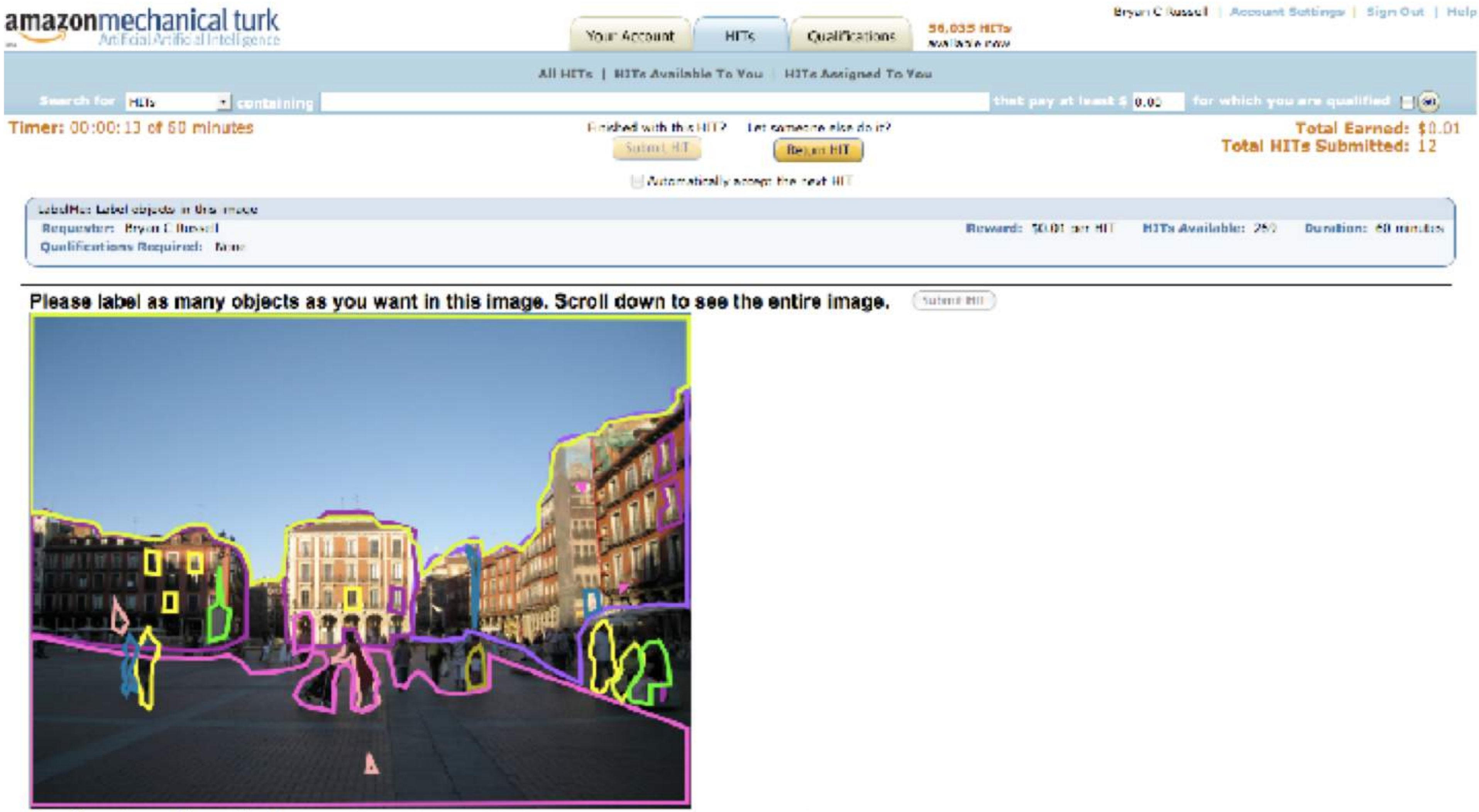
Get Started

FAQ | Contact Us | Careers at Amazon | Developers | Press | Policies

An amazon.com. company

Labeling for money

amazon mechanical turk
Artificial Artificial Intelligence



Alexander Sorokin, David Forsyth, "Utility data annotation with Amazon Mechanical Turk", First IEEE Workshop on Internet Vision at CVPR 08.

2009 ImageNet

22K categories, 14M images

ImageNet: A Large-Scale Hierarchical Image Database,
Deng, Dong, Socher, Li, Li and Fei-Fei, *CVPR,* 2009

2010 SUN

908 scene categories

Beer garden

SUN Database: Large-scale Scene Recognition from Abbey to Zoo
Xiao, Hays, Ehinger, Oliva, and Torralba, *CVPR*, 2010.

2014 COCO

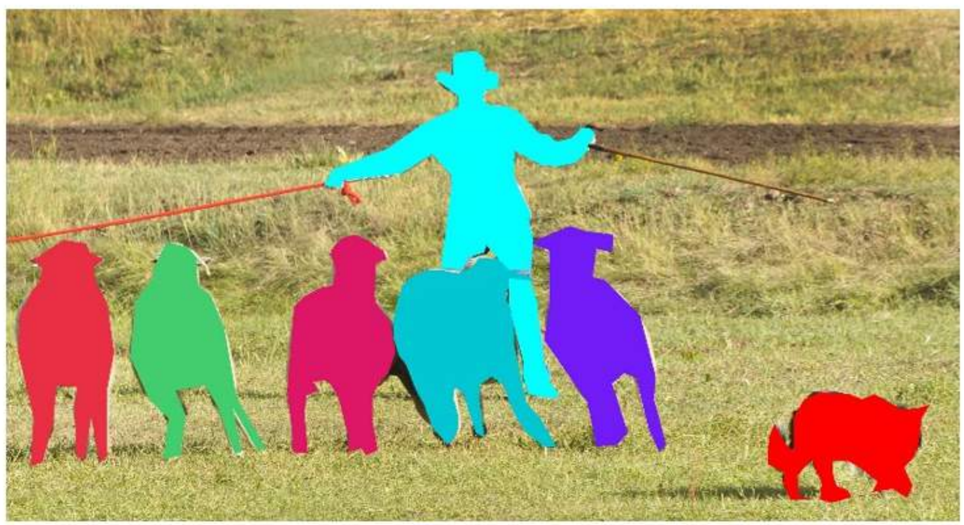
80 object categories with instance masks

Detection in Context

Instance segmentation

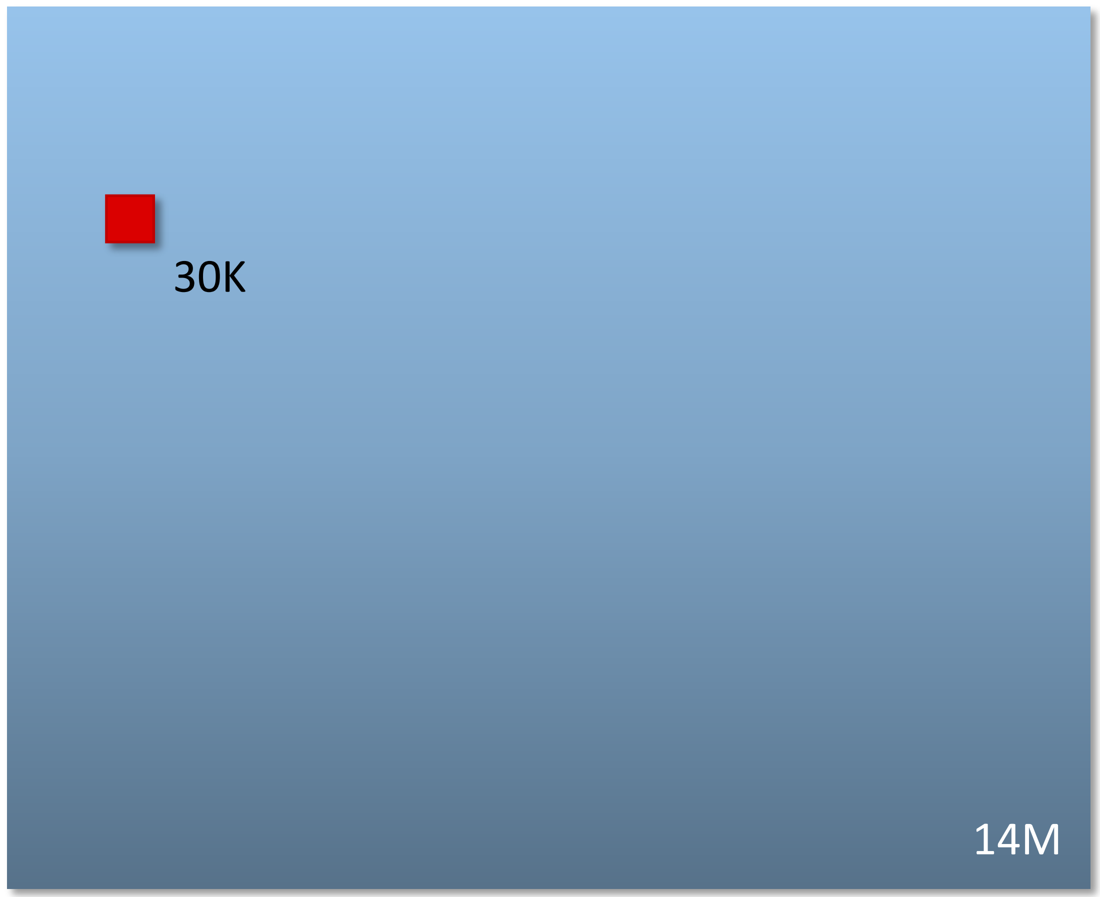
Non-iconic instances

Microsoft COCO
Common Objects in Context



Images

2009 2012



ImageNet

Datasets in perspective

Number of images on my hard drive: 10^5

Number of images seen during my first 10 years: 10^8
(3 images/second * 60 * 60 * 16 * 365 * 10 = 630720000)

Number of images seen by all humanity: 10^{20}
106,456,367,669 humans[1] * 100 years * 3 images/second * 60 * 60 * 16 * 365 =
1 from http://www.prb.org/Articles/2002/HowManyPeopleHaveEverLivedonEarth.aspx

Number of all 32x32 images: 10^{7373}
$256^{32*32*3} \sim 10^{7373}$

When do we need big data?

Two Kinds of Things in the World

Navier–Stokes momentum equation (*convective form*)

$$\rho \frac{D\mathbf{u}}{Dt} = -\nabla \bar{p} + \mu \nabla^2 \mathbf{u} + \tfrac{1}{3}\mu \nabla(\nabla \cdot \mathbf{u}) + \rho \mathbf{g}.$$

Nasdaq 10 Years
Daily Chart

Unreasonable Effectiveness of Data

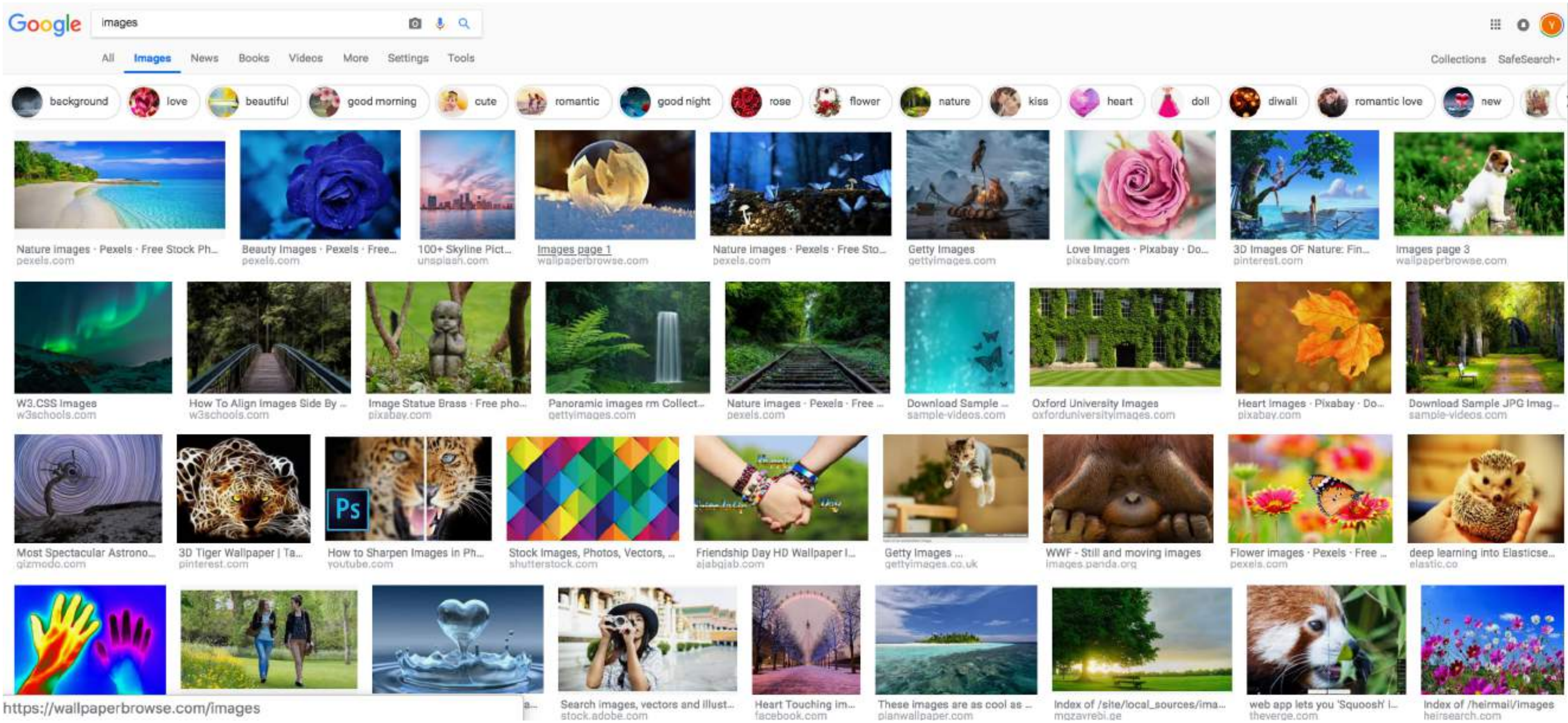
[Halevy, Norvig, Pereira 2009]

- Parts of our world can be explained by elegant mathematics
  - Physics, astronomy, chemistry, etc.

- But much cannot
  - economics, psychology, …
  - visual understanding?

- Enter **The Magic of Data**
  - Great advances in speech, vision & NLP

Lots of Data Available



Thanks to the Internet

Unreasonable Effectiveness of Data

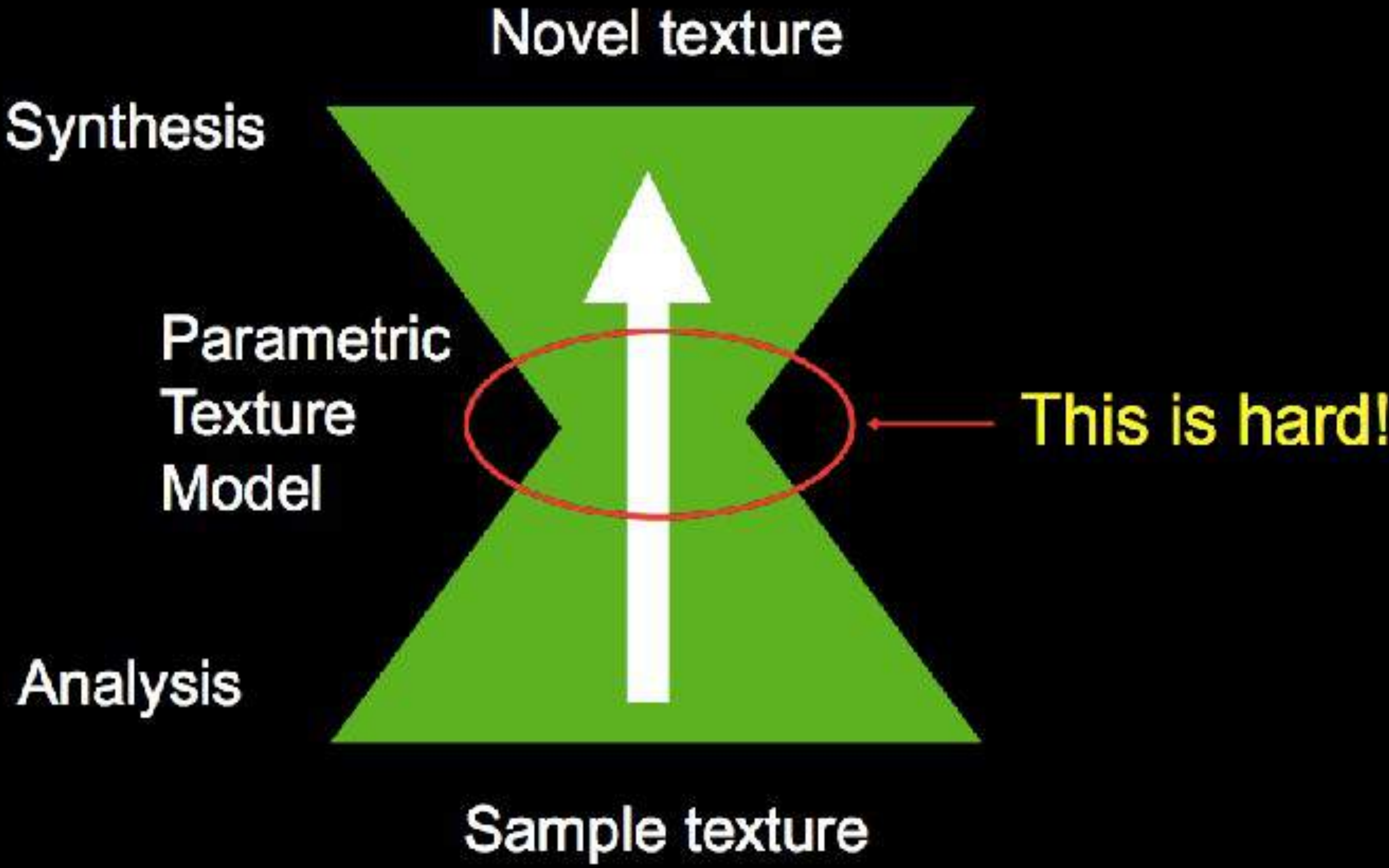
Simple Algorithms (Dumb) + Lot of Data are better than Complicated algorithms

A Quick Example:
Texture Generation

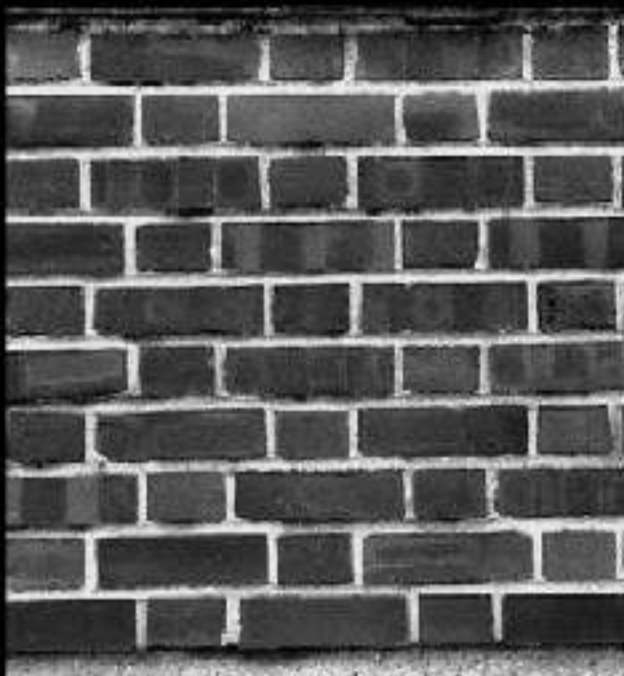
Texture Synthesis

Texture Synthesis

Classical Texture Synthesis



Throwing away too much too soon?



input texture

synthesized texture

Non-parametric Approach

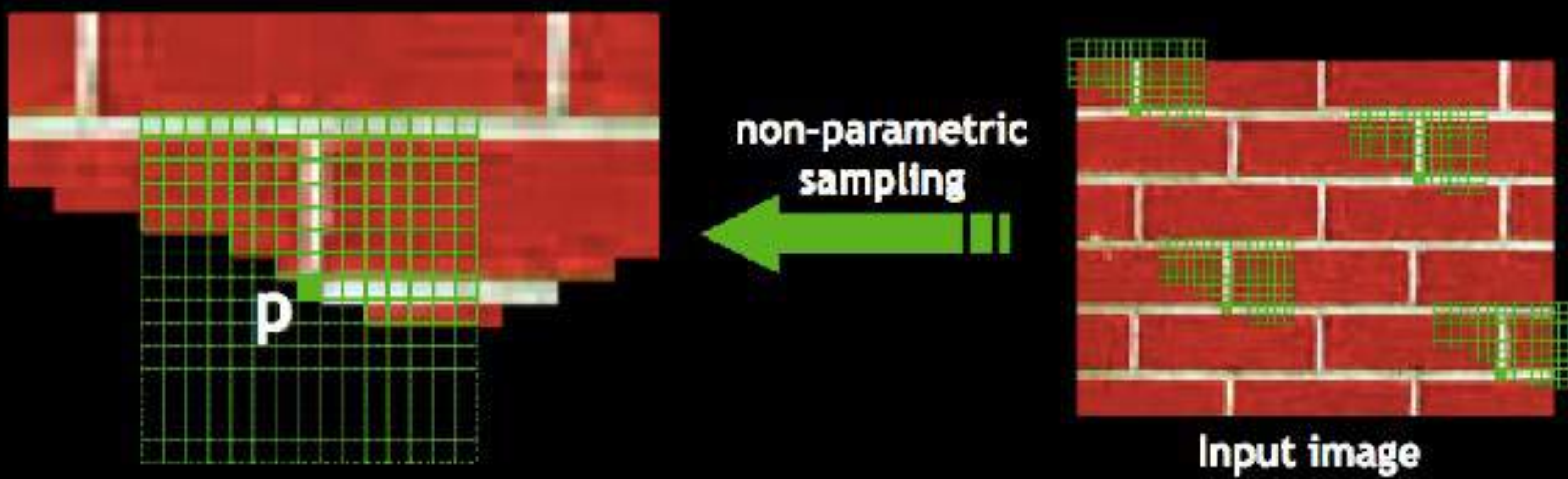
Novel texture

Synthesis

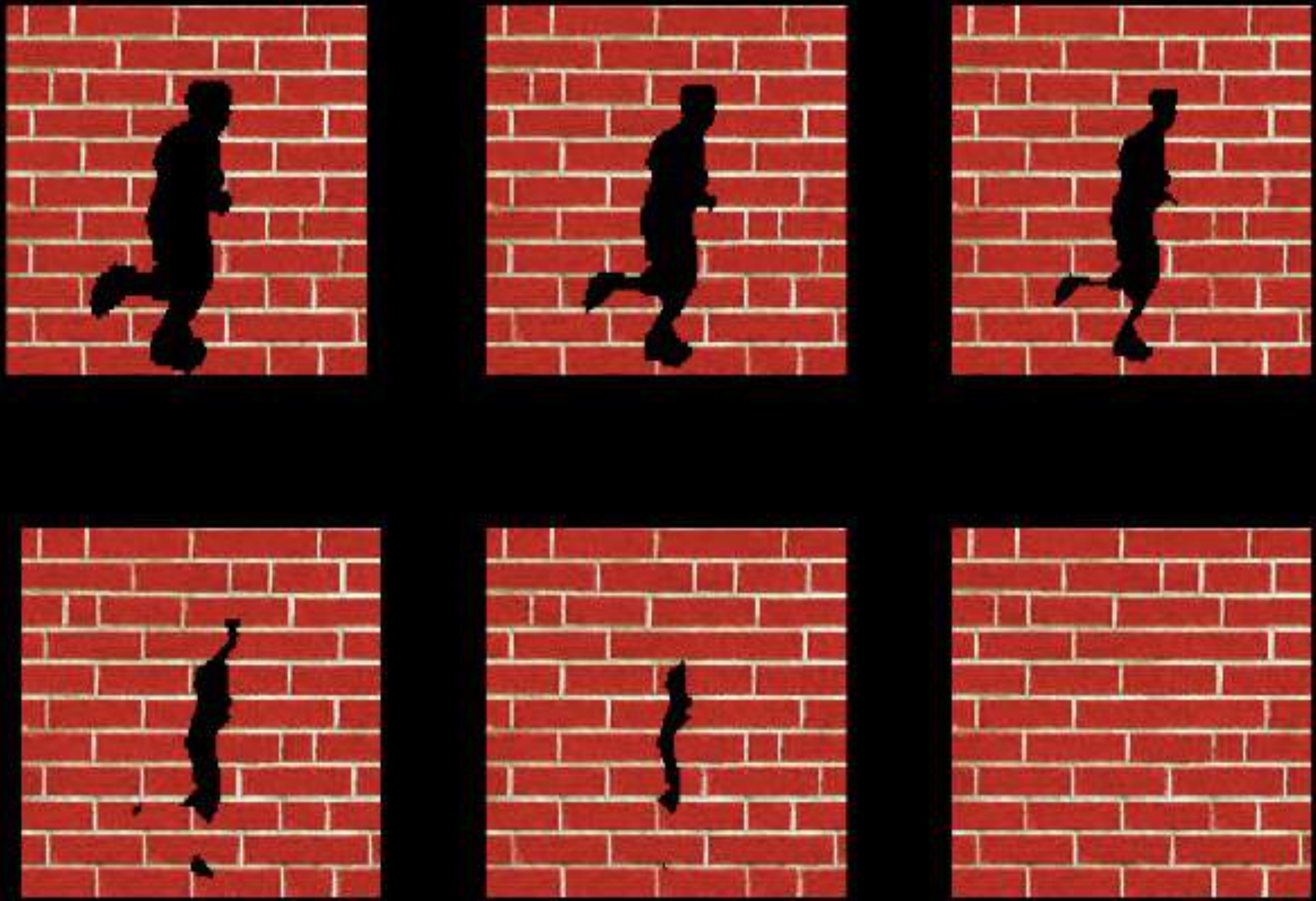
Analysis

Sample texture

[Efros & Leung, '99, Efros & Freeman '01]

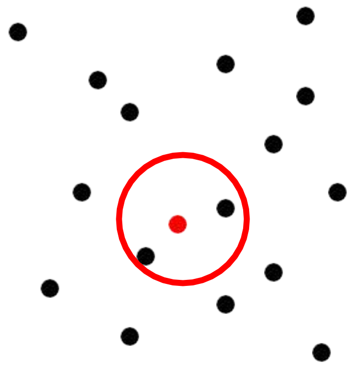


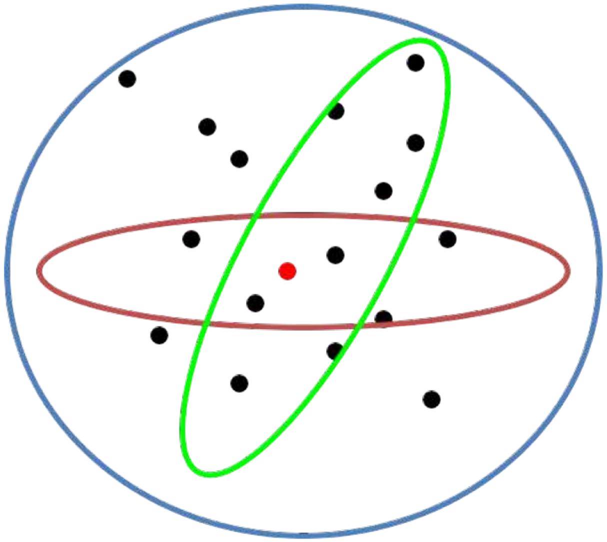
Texture Growing



So, how do we use big data?

Two ways to use Lots of Data





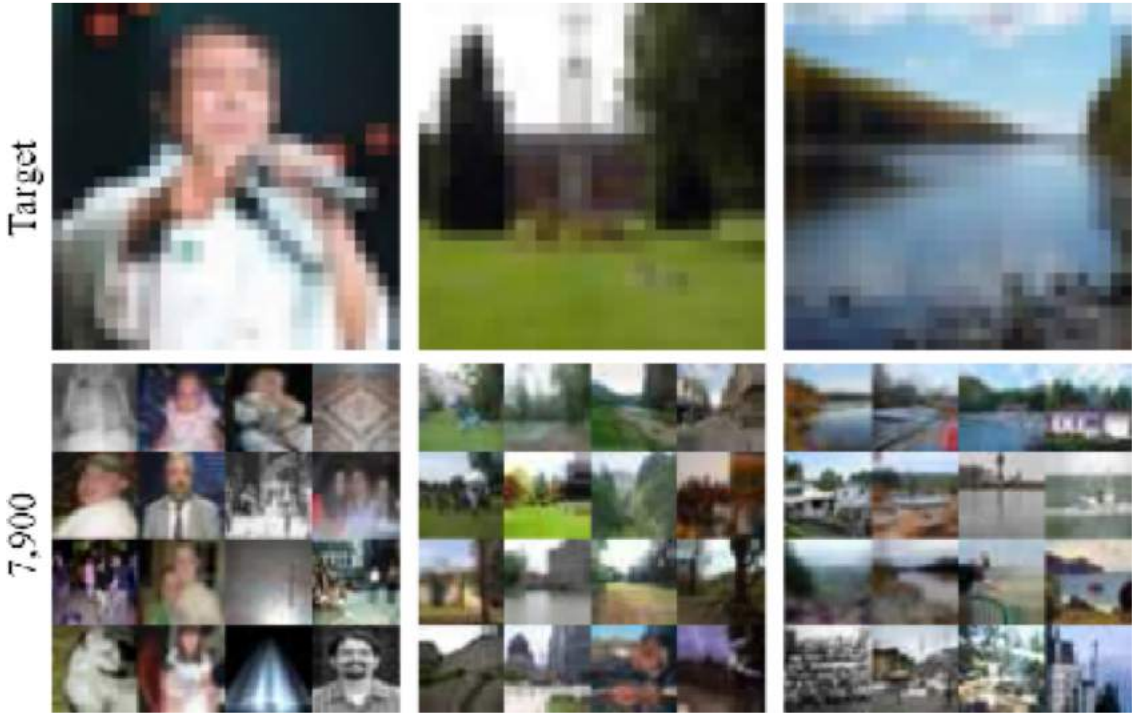
Brute Force Vision: Find that needle in the haystack and disregard the rest (a.k.a. kNN)

See what different subsets of data think of you (a.k.a. Parametric Models)

Lots

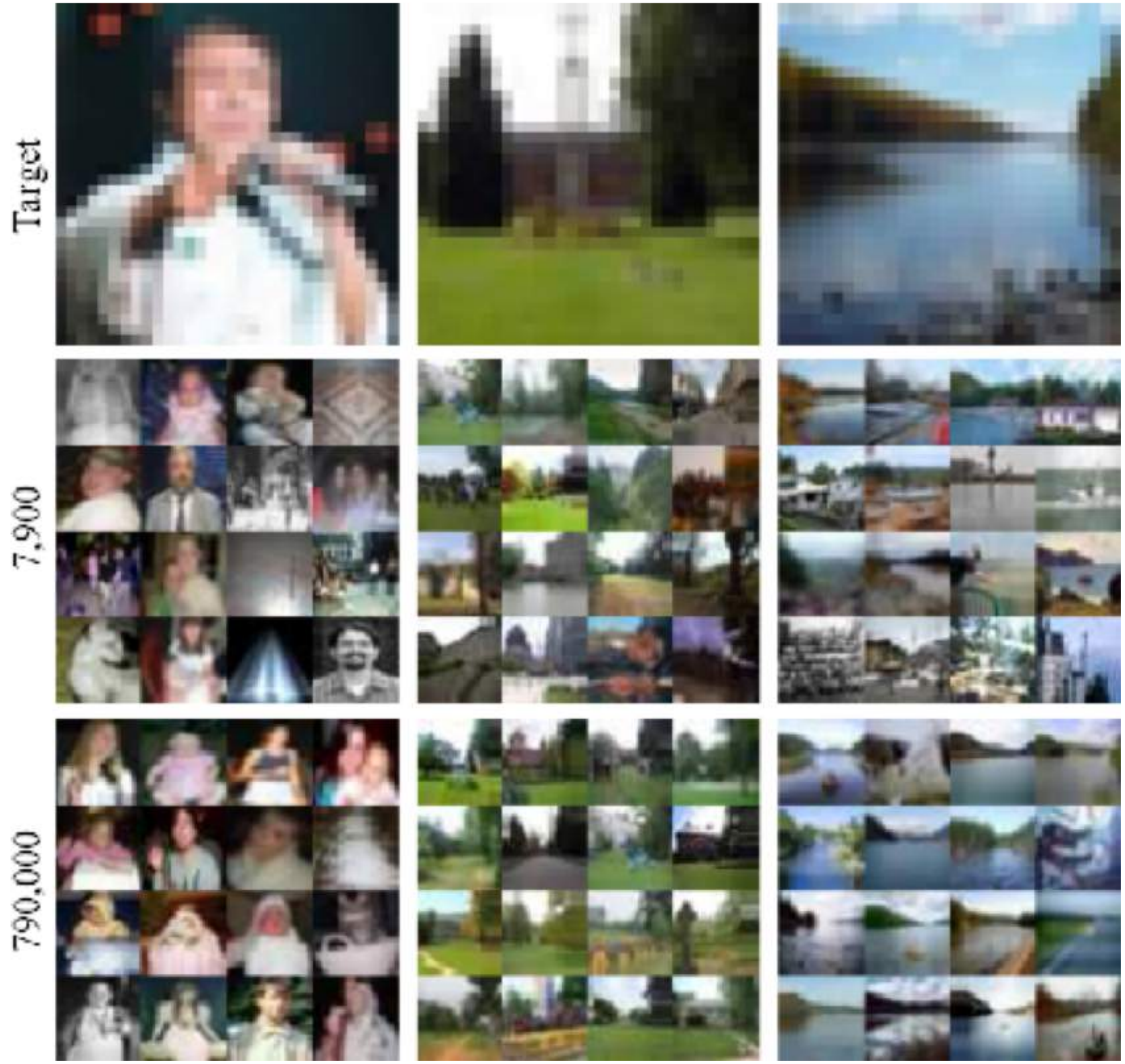
Of

Images



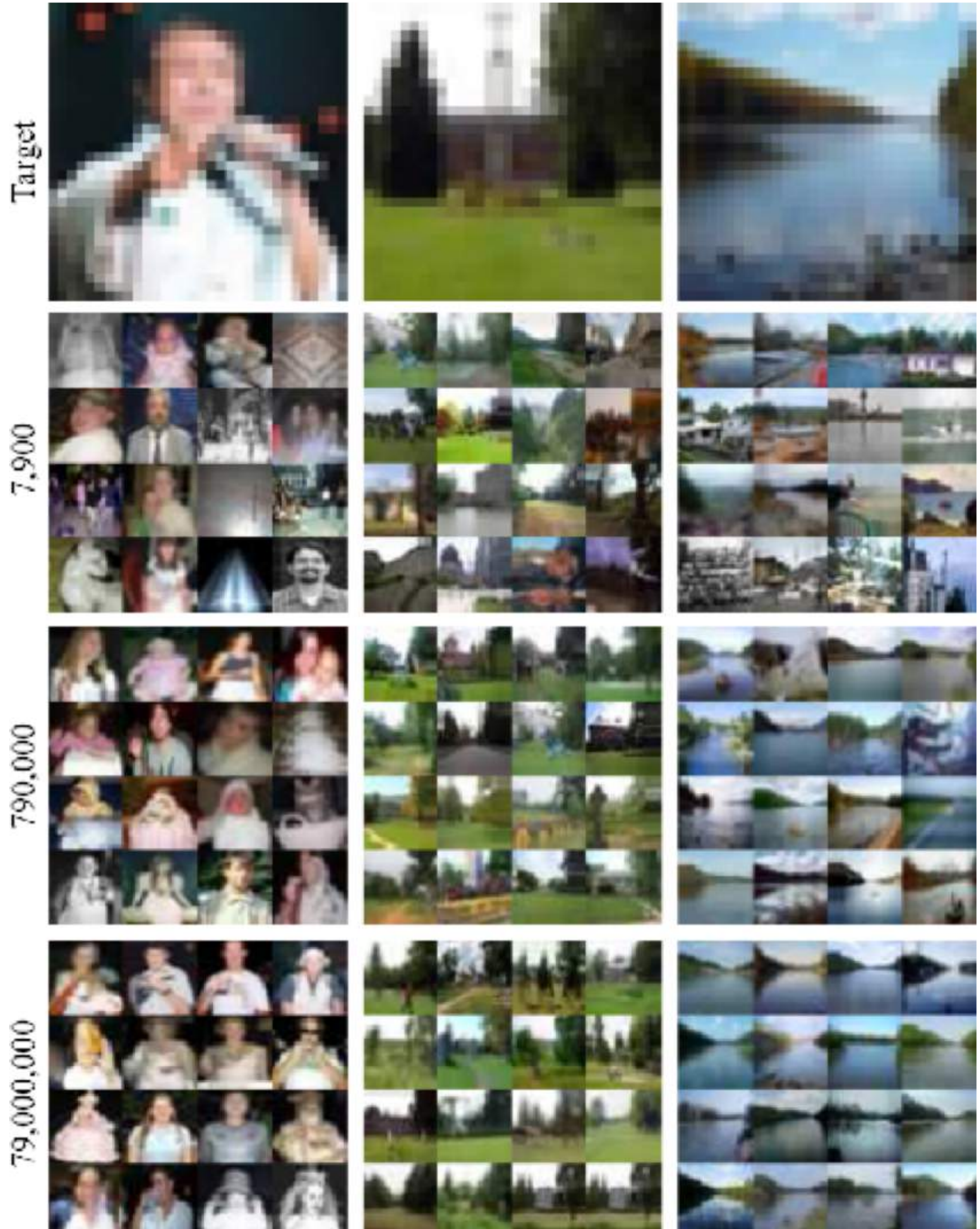
A. Torralba, R. Fergus, W.T.Freeman. PAMI 2008

Lots Of Images



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

Lots Of Images



kNN matching is great…

- **because we live in a (mostly) boring world!**

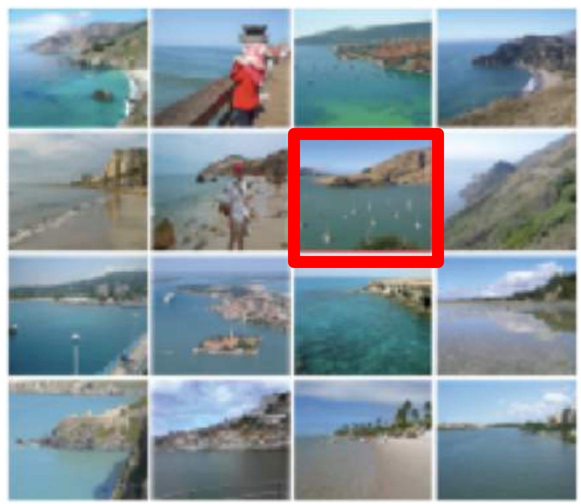
 

Image Completion

Original Image | Input | Criminisi et al. | MS *Smart Erase*

Instead, generate proposals using millions of images

Input



16 nearest neighbors
(gist+color matching)

output

Hays, Efros, 2007

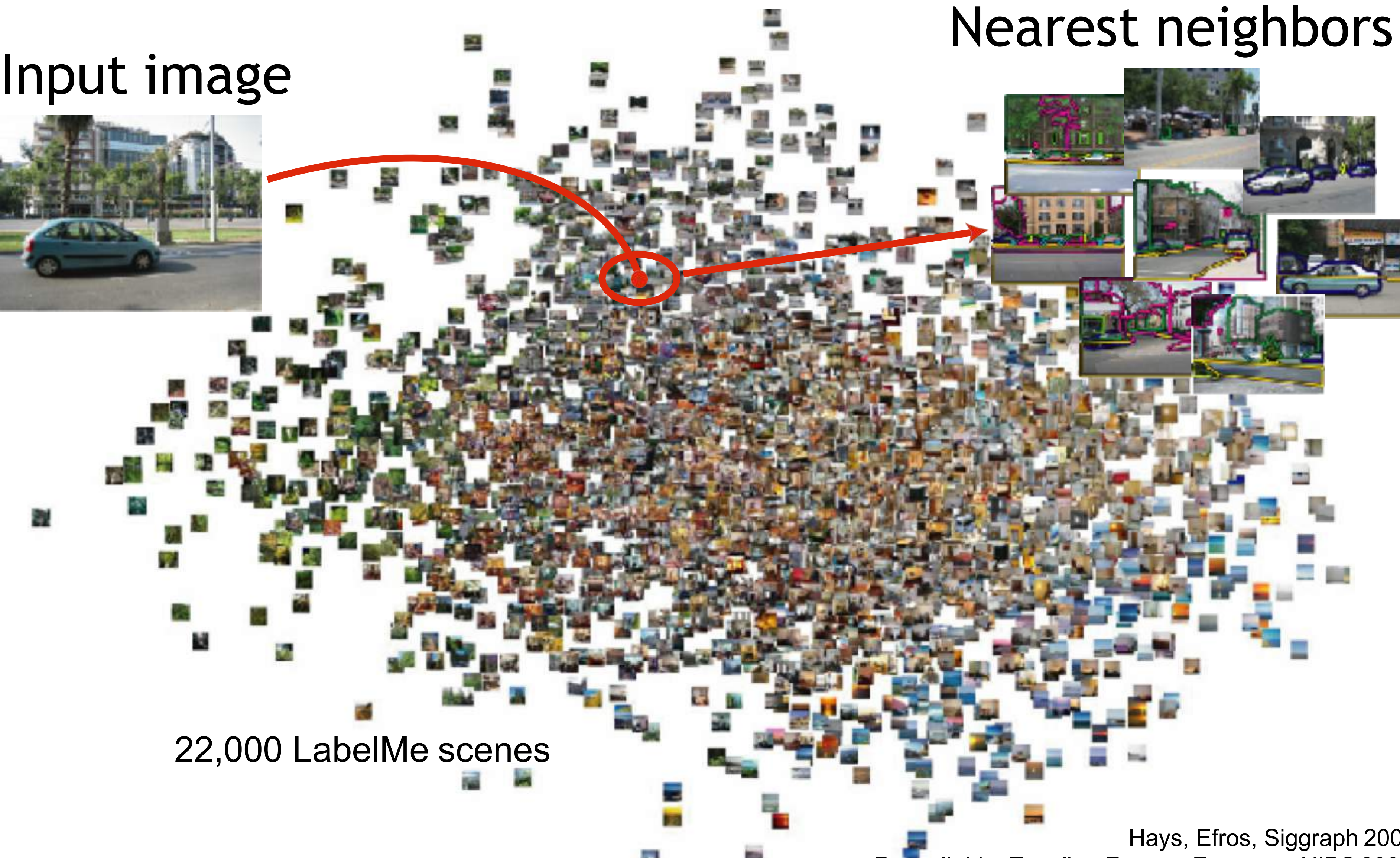
Automatic Colorization Result

Grayscale input High resolution

Colorization of input using average

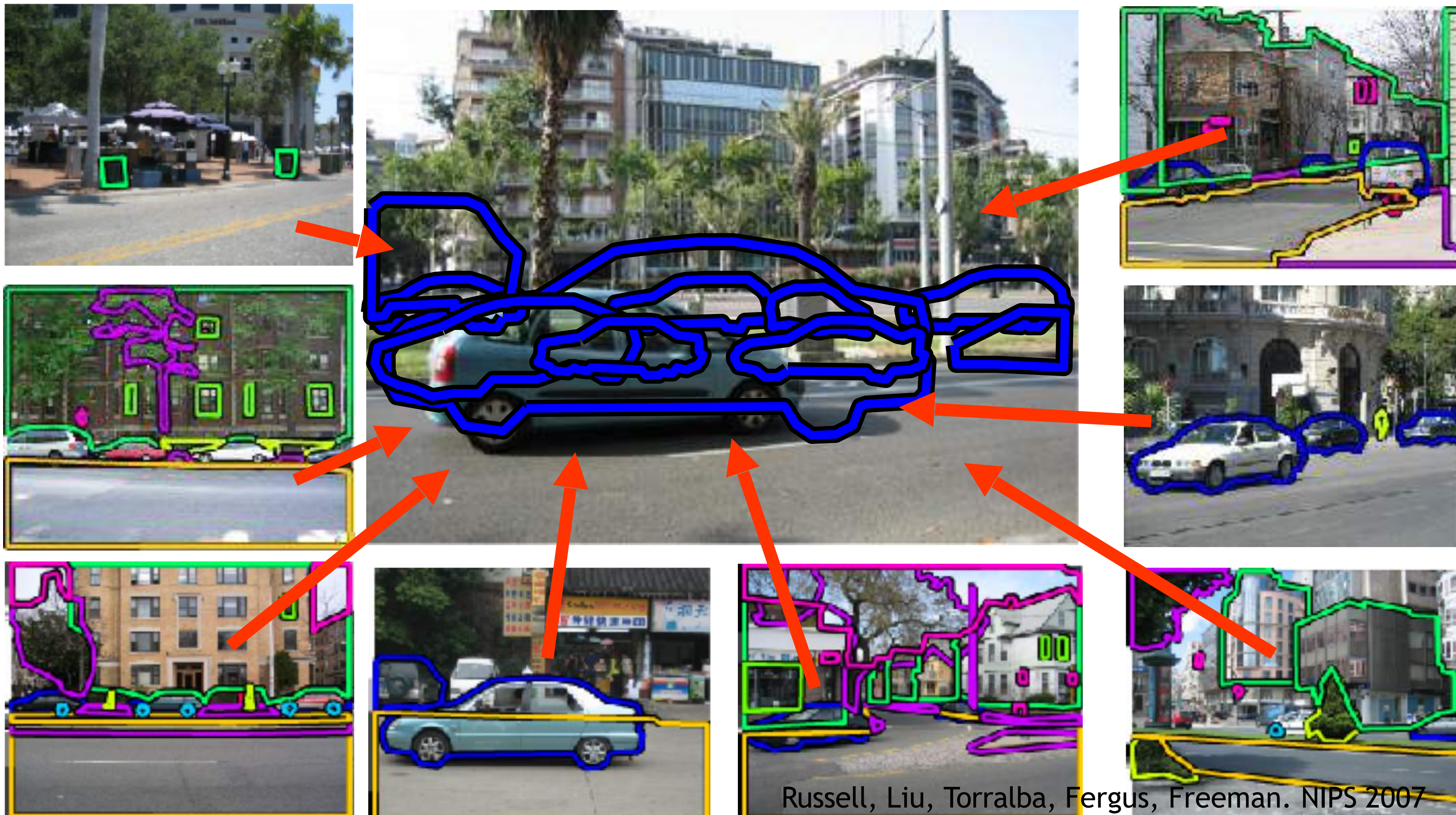
A. Torralba, R. Fergus, W.T.Freeman. 2008

With a good image similarity and a lot of data…



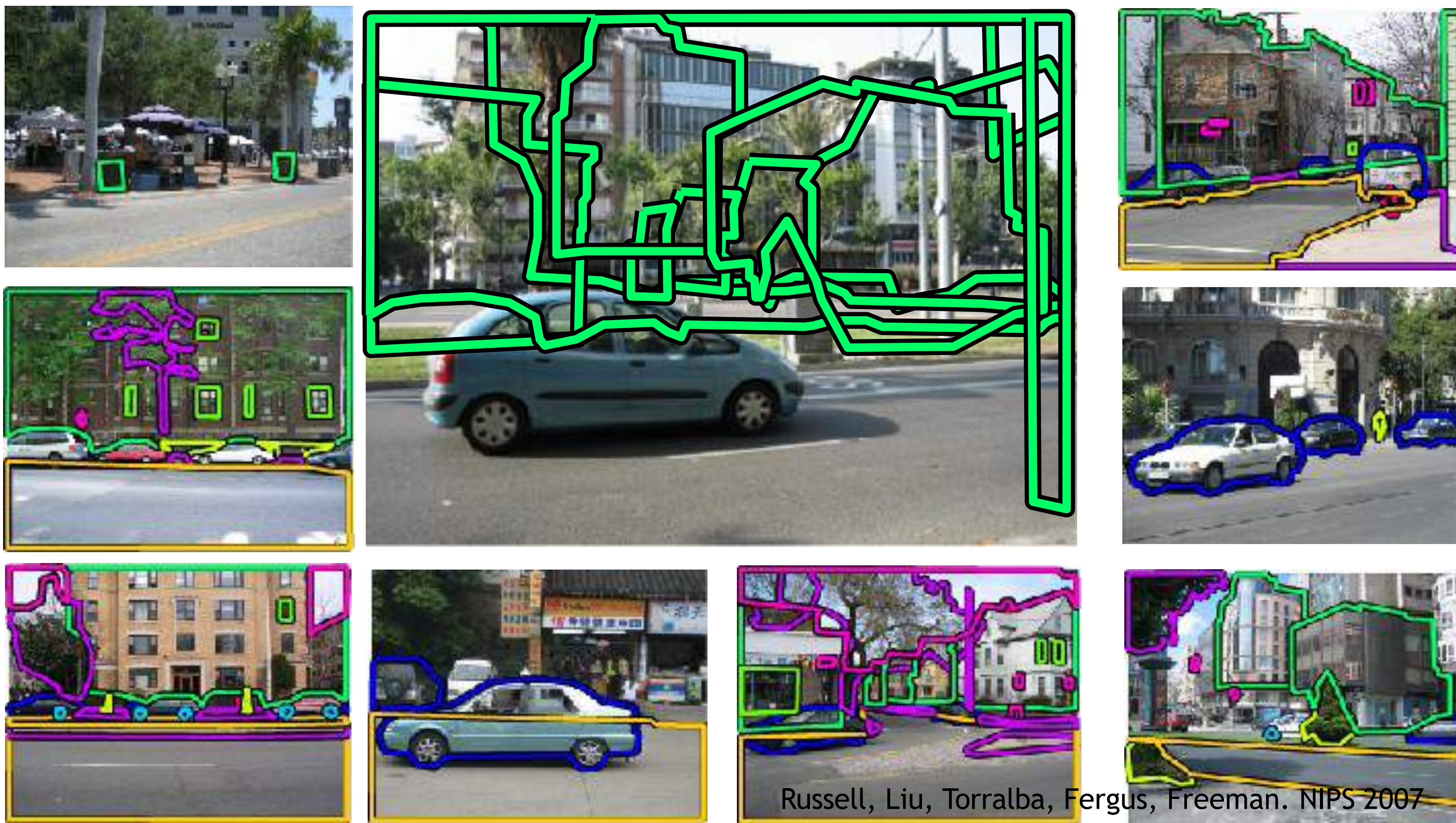
Hays, Efros, Siggraph 2006
Russell, Liu, Torralba, Fergus, Freeman. NIPS 2007

With a good image similarity and a lot of data…



Russell, Liu, Torralba, Fergus, Freeman. NIPS 2007

With a good image similarity and a lot of data…

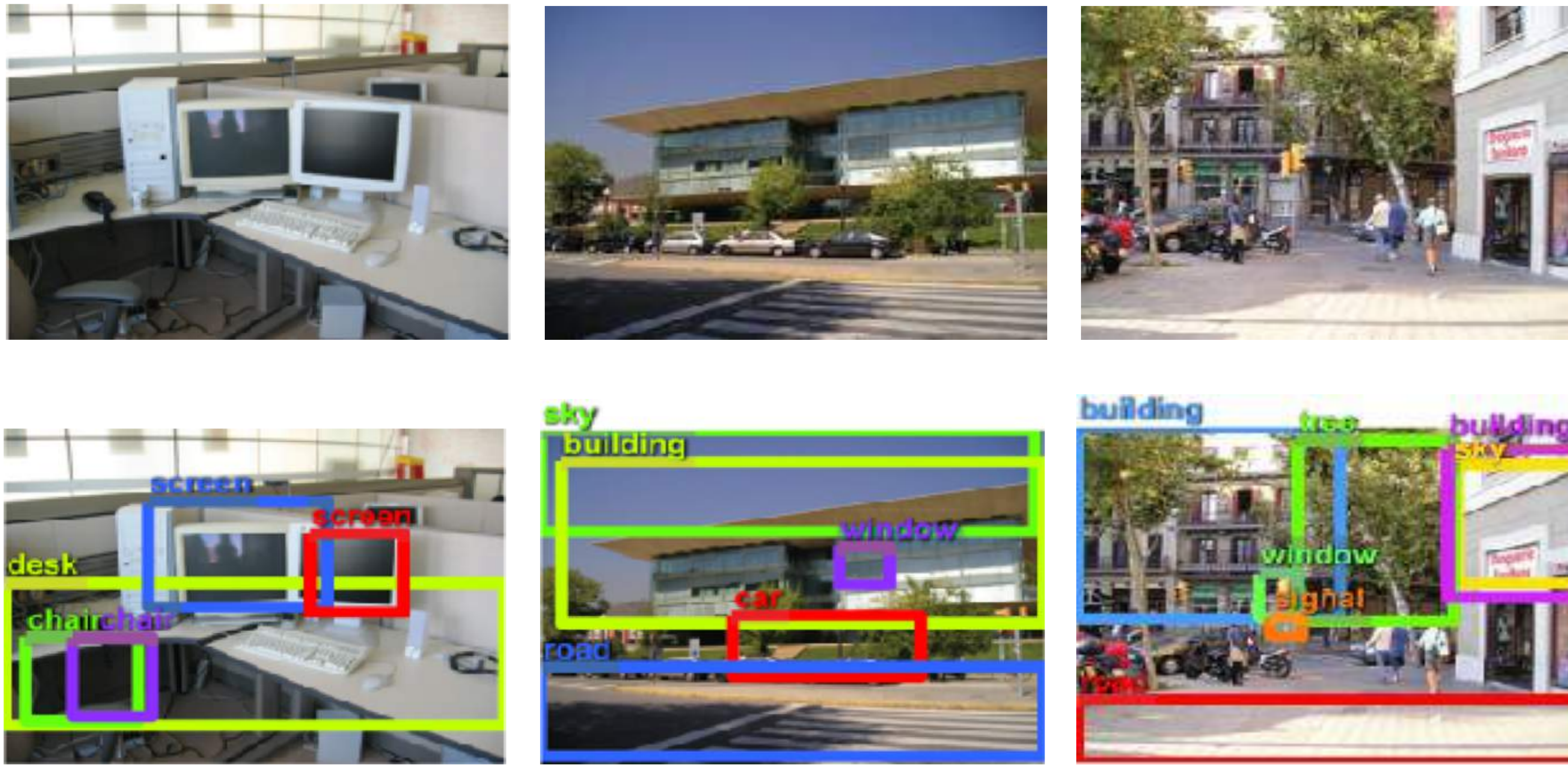


Russell, Liu, Torralba, Fergus, Freeman. NIPS 2007

With a good image similarity and a lot of data…

Russell, Liu, Torralba, Fergus, Freeman. NIPS 2007

Outputs



Russell, Liu, Torralba, Fergus, Freeman. NIPS 2007

While many scenes are boring…

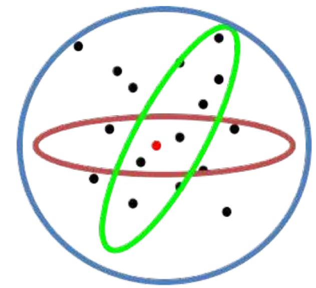
Slide by Antonio Torralba

Some scenes are unique

Slide by Antonio Torralba

Generalization to rare scenes

- Moving away from kNN methodology…
- Use data to make connections



→ Ideal Feature Extractor → Ideal Decision Function →

window, top-left
clock, top-middle
shelf, left
…
drawing, middle
hat, bottom right

Parametric Models

Ideal Feature Extractor → Ideal Decision Function →

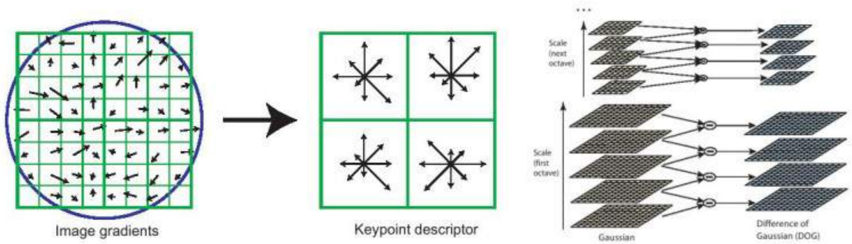
window, top-left
clock, top-middle
shelf, left
...

drawing, middle
hat, bottom right

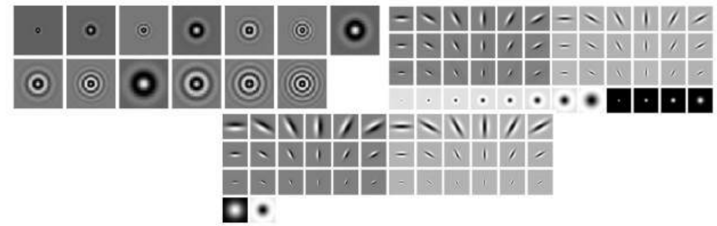
Q: What types of features shall we consider?

Q: What types of decision functions shall we consider?

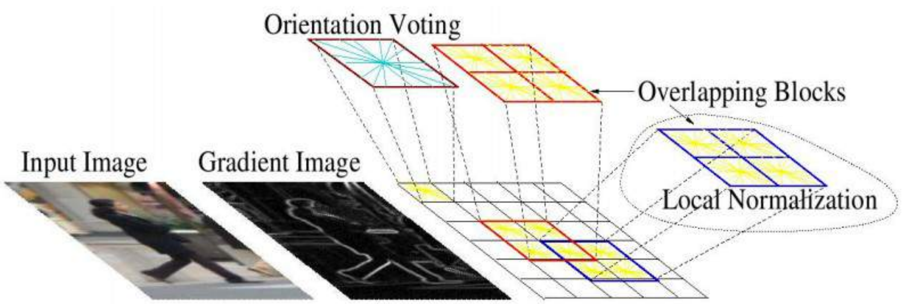
Computer Vision Features



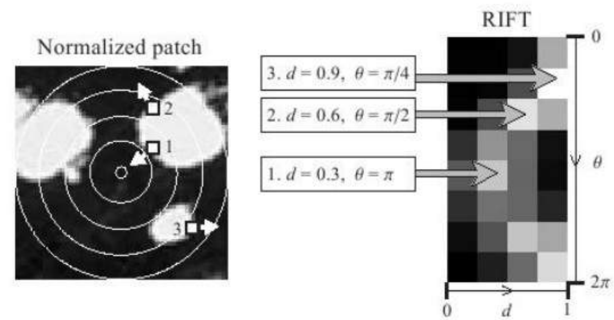
SIFT



Textons

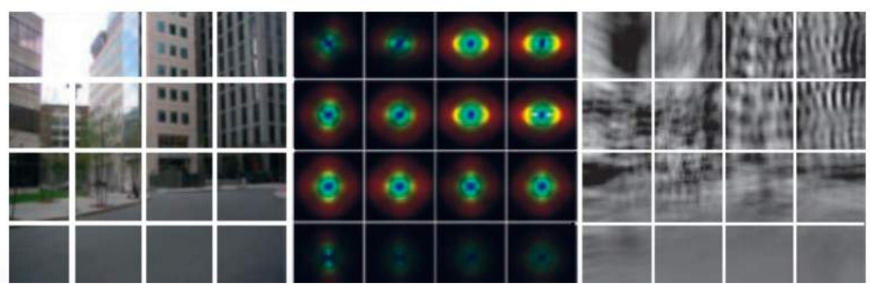


HoG



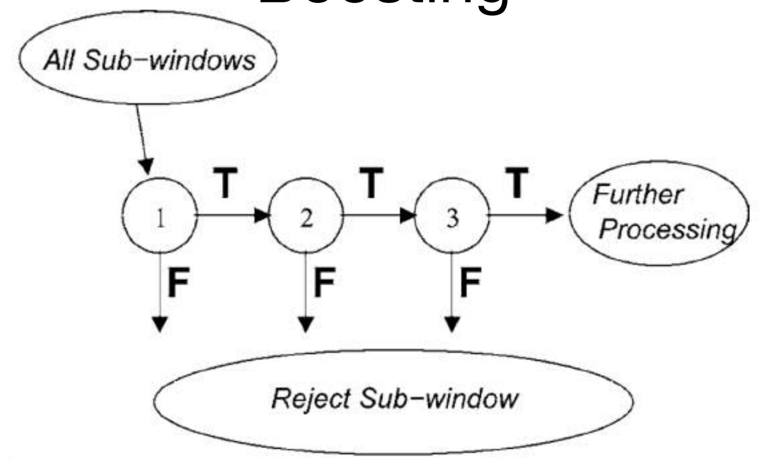
RIFT

GIST

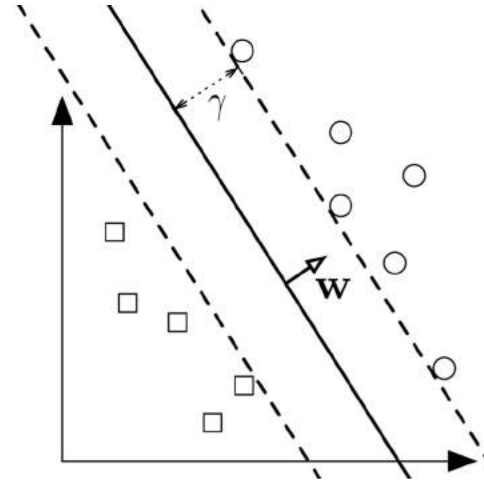


Computer Vision Decision Functions

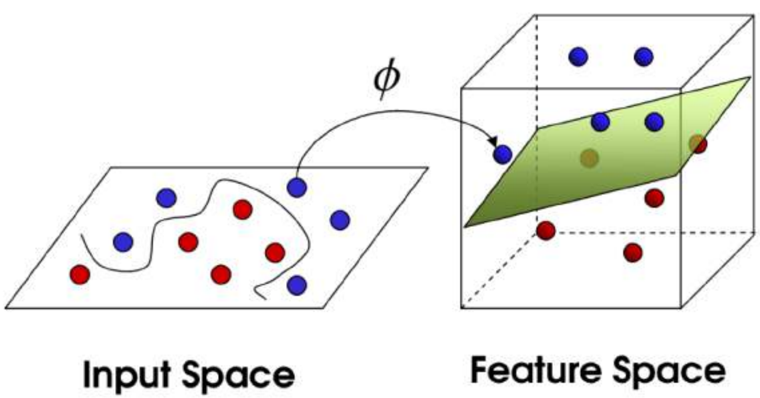
Boosting



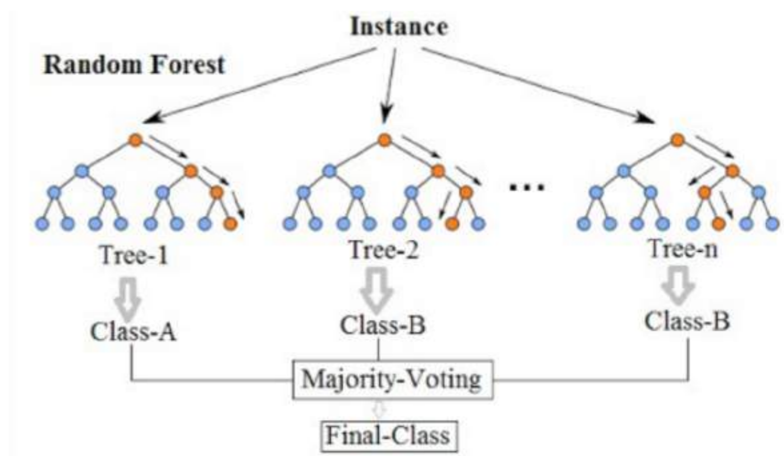
Linear SVM



Kernel Machines



Random Forest



Can we learn everything in one shot?
- representation learning

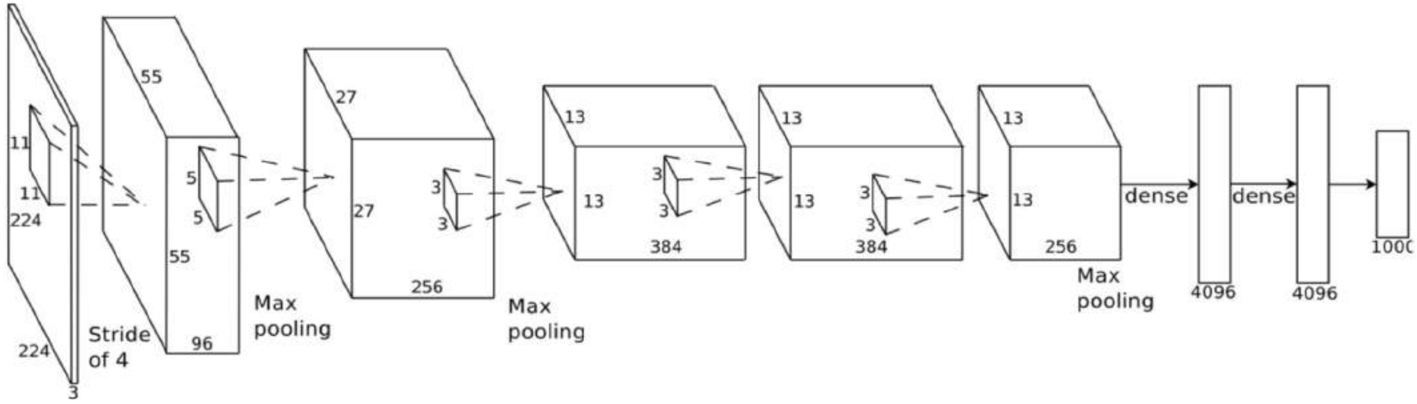


Image Completion using Deep Models

Iizuka, Simo-Serra, Ishikawa, SIGGRAPH 2017

Image Colorization using Deep Models

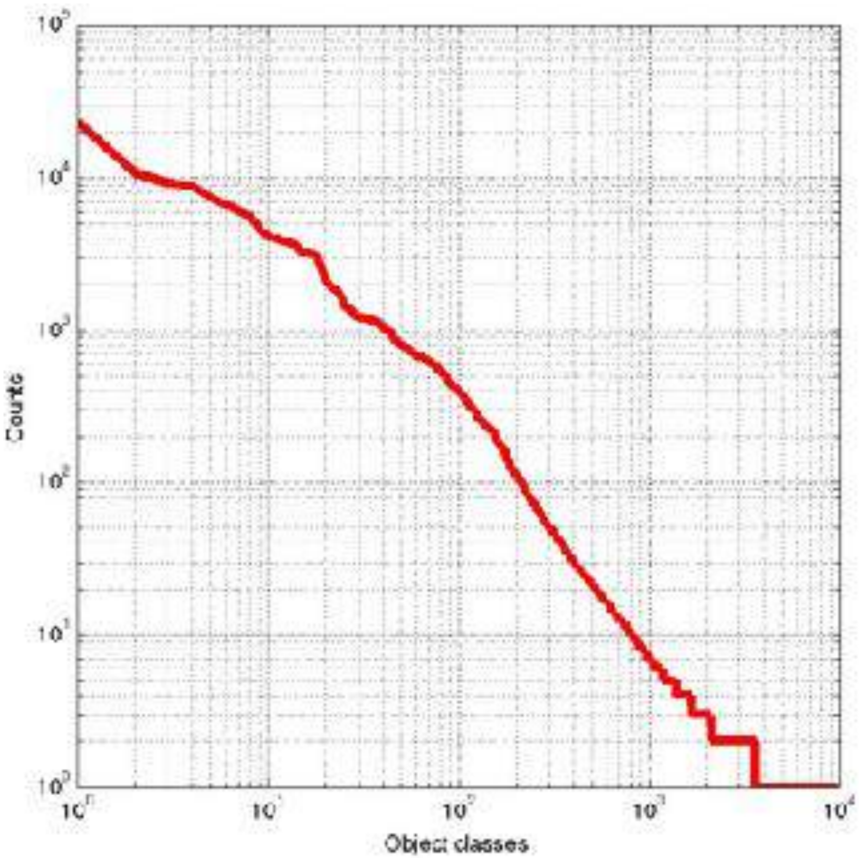
Zhang, Isola, Efros, ECCV 2016

Object Detection and Segmentation

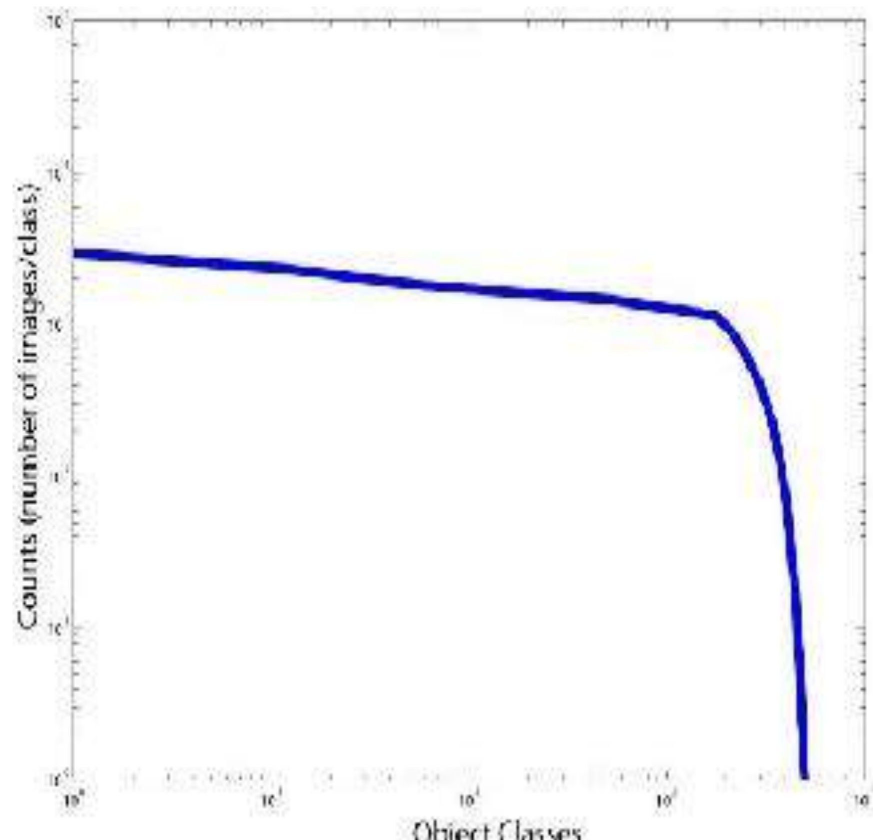
“Mask RCNN”, He, Gkioxari, Dollár, Girshick, *ICCV,* 2017

The Bias of Data

Object distributions

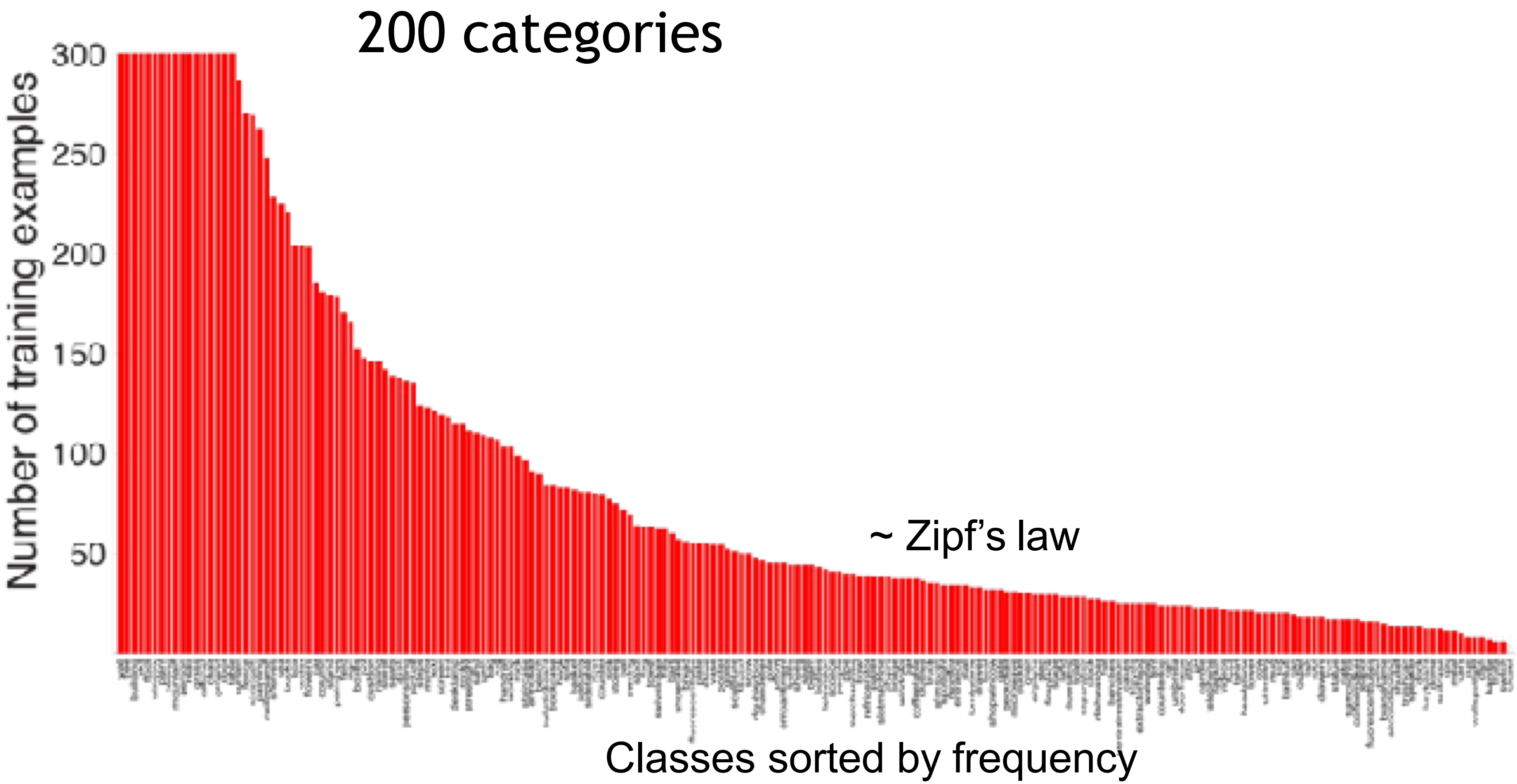


LabelMe



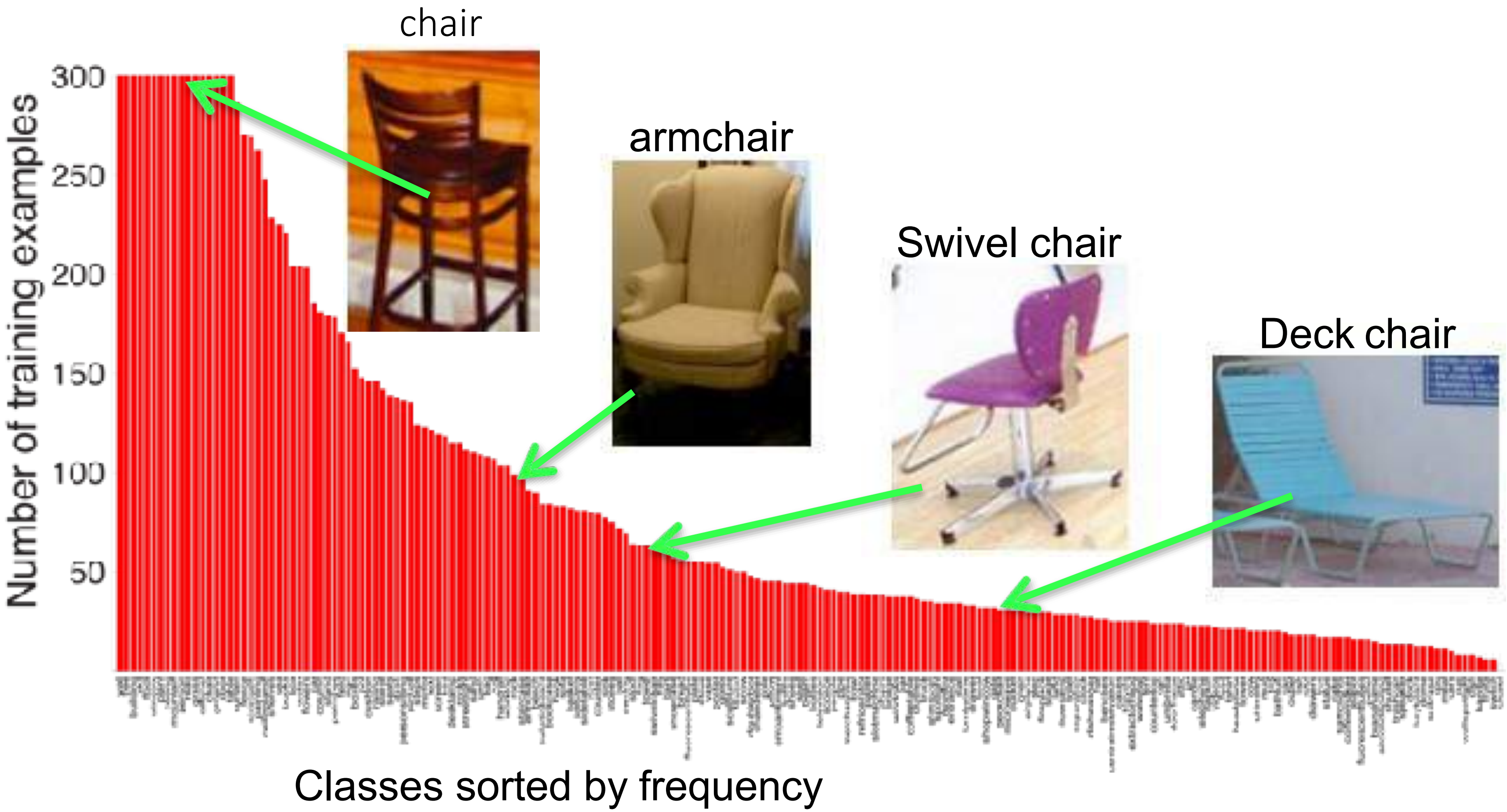
IMAGENET

SUN database



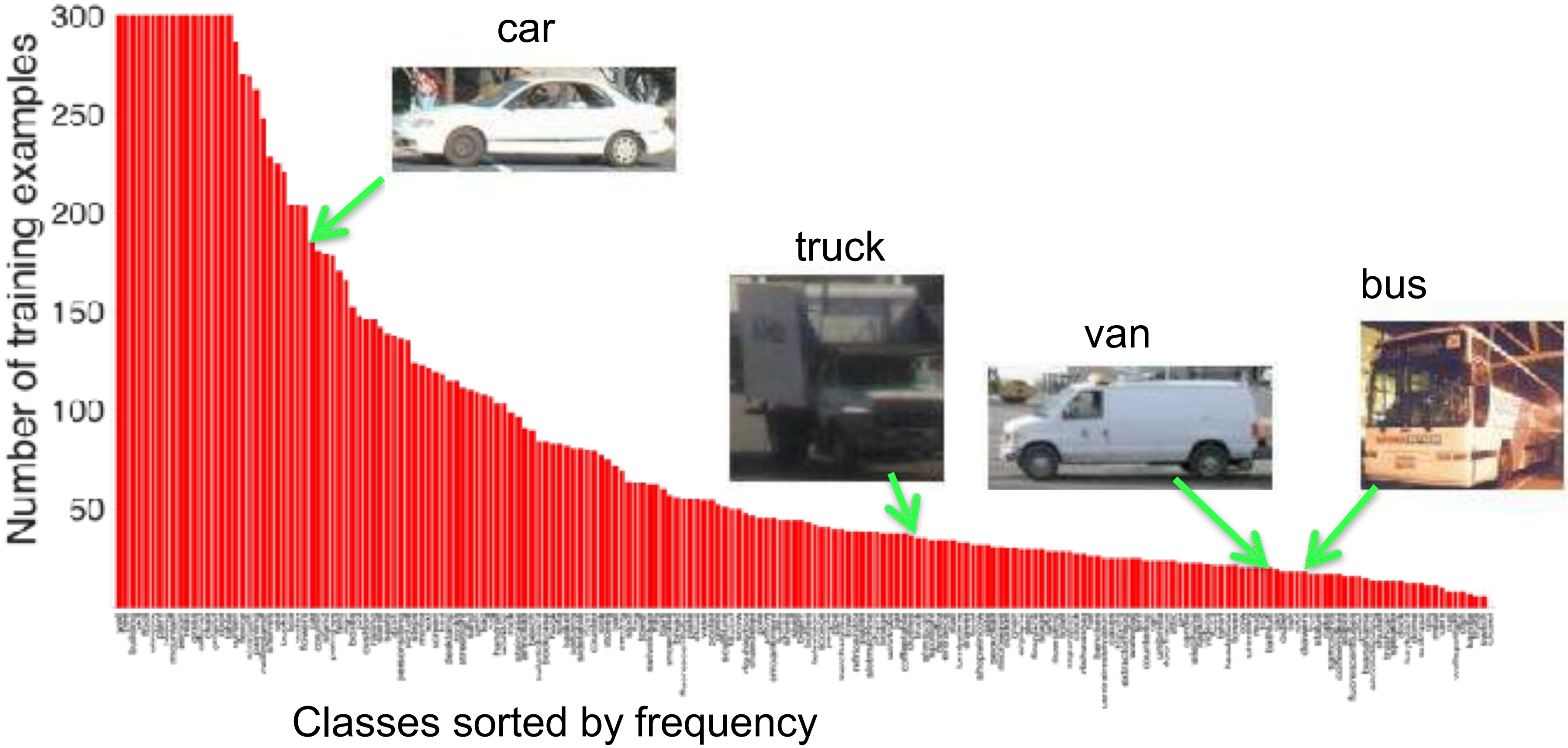
The first 9 objects account for 50% of all training examples
17 classes with more than 300 examples
109 classes with less than 50 examples

Rare objects are similar to frequent objects



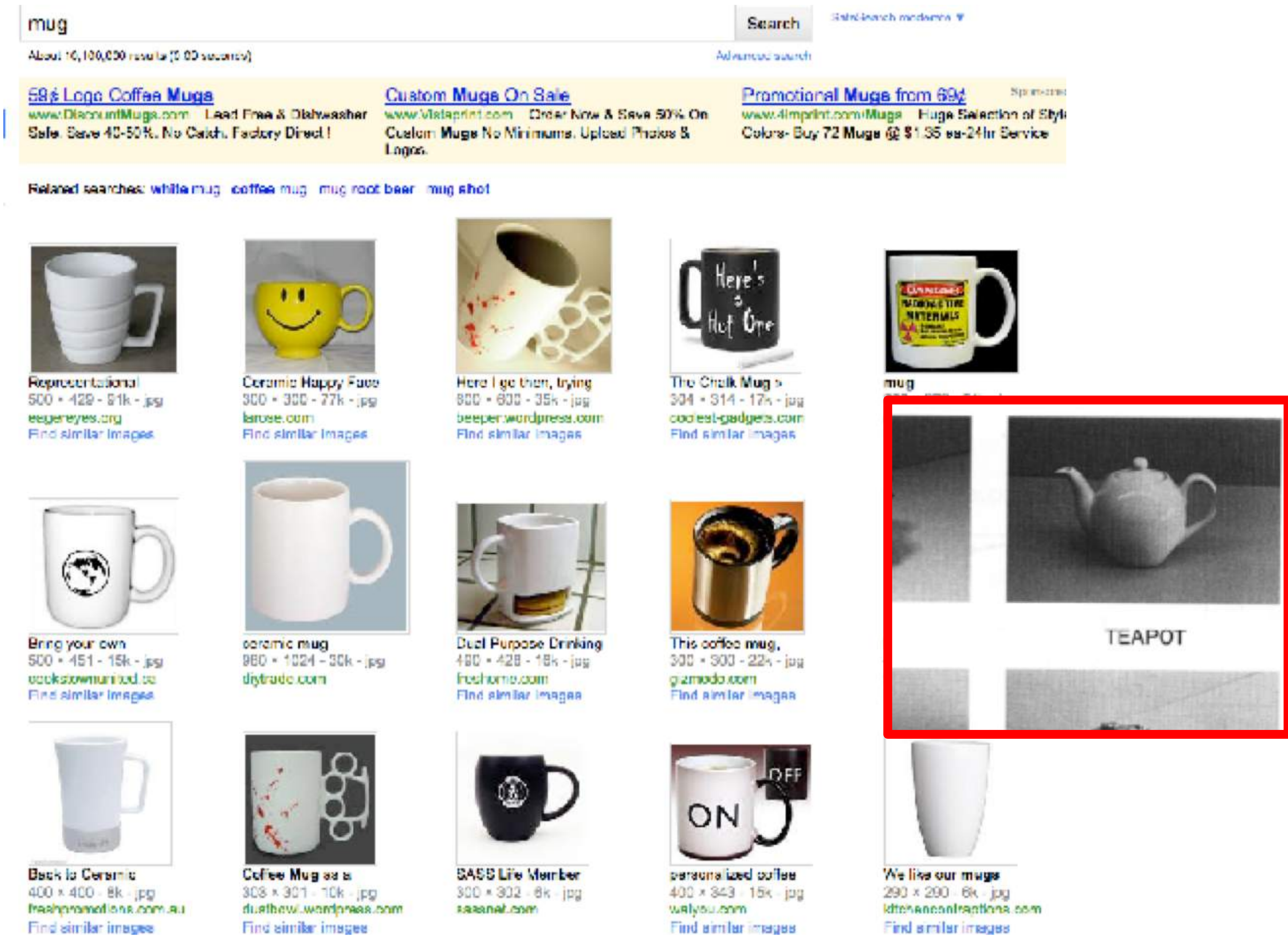
Salakhutdinov, Torralba, and Tenenbaum, CVPR, 2011

Rare objects are similar to frequent objects

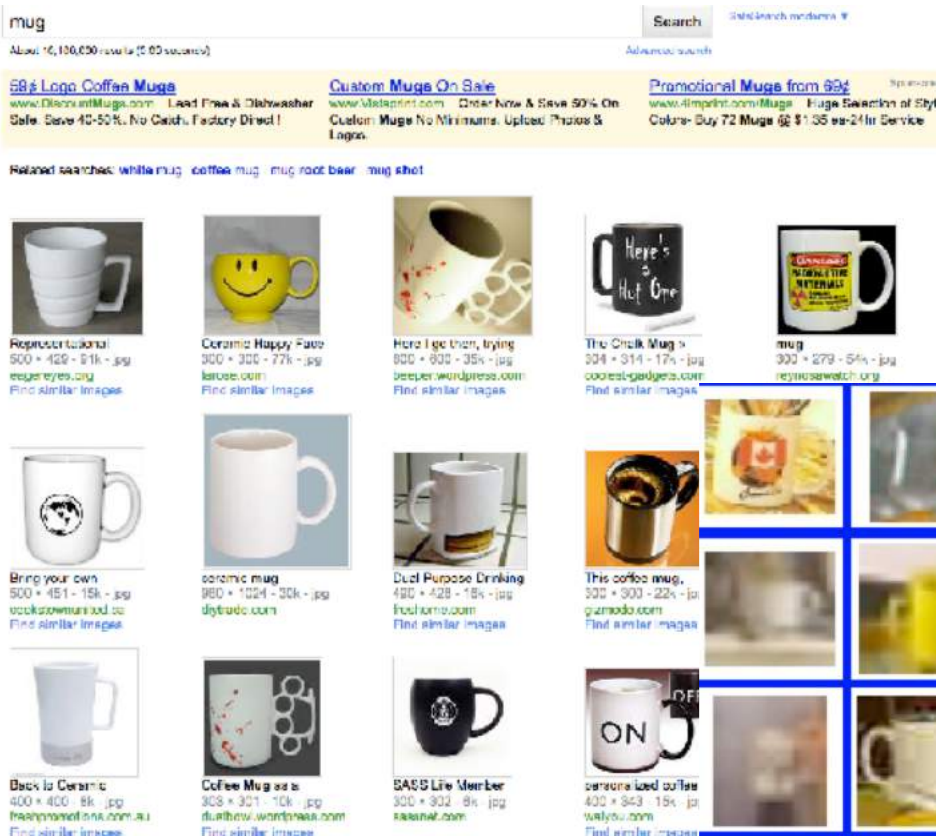


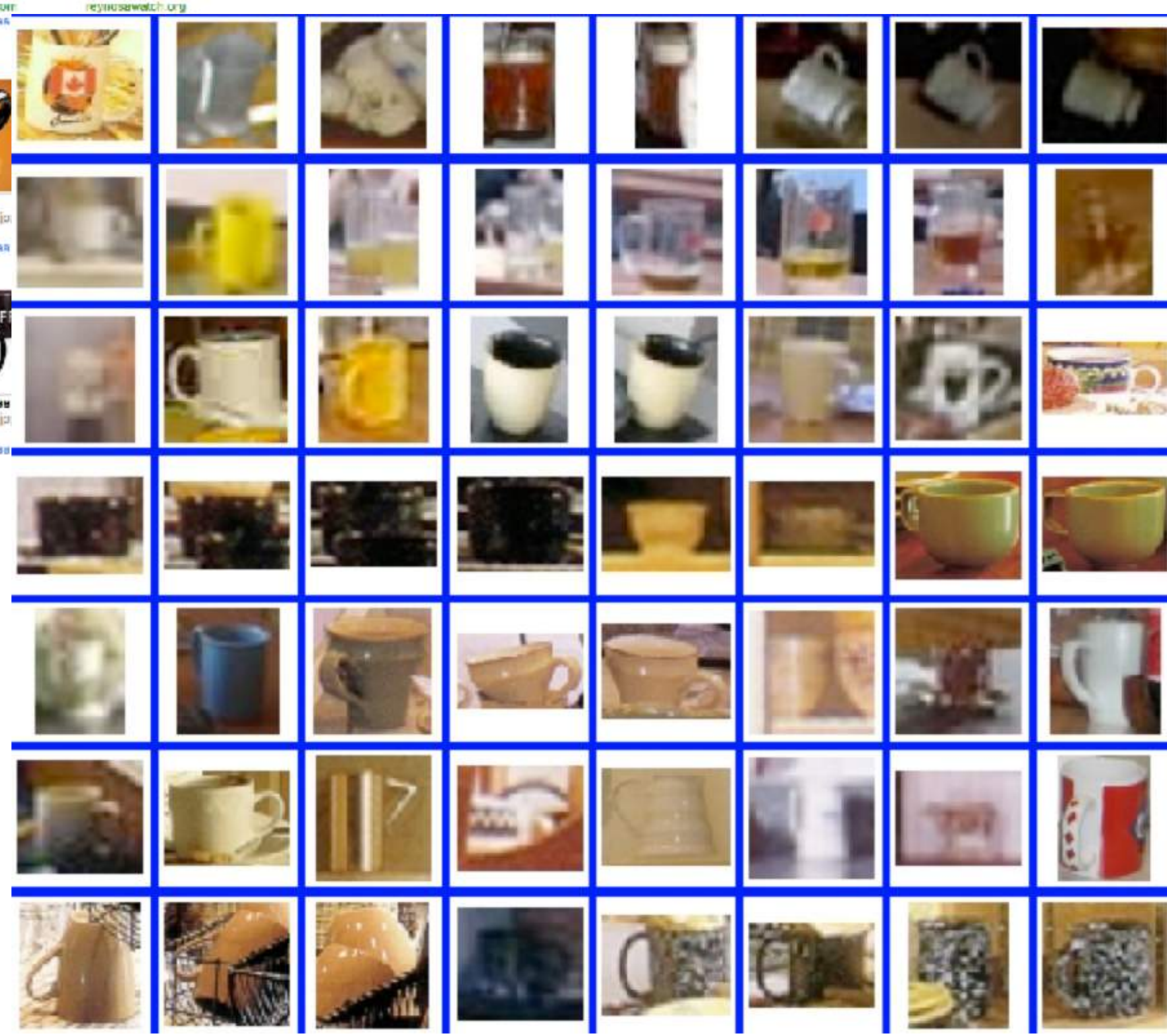
Salakhutdinov, Torralba, and Tenenbaum, CVPR, 2011

Some bias comes from the way the data is collected



Google mugs



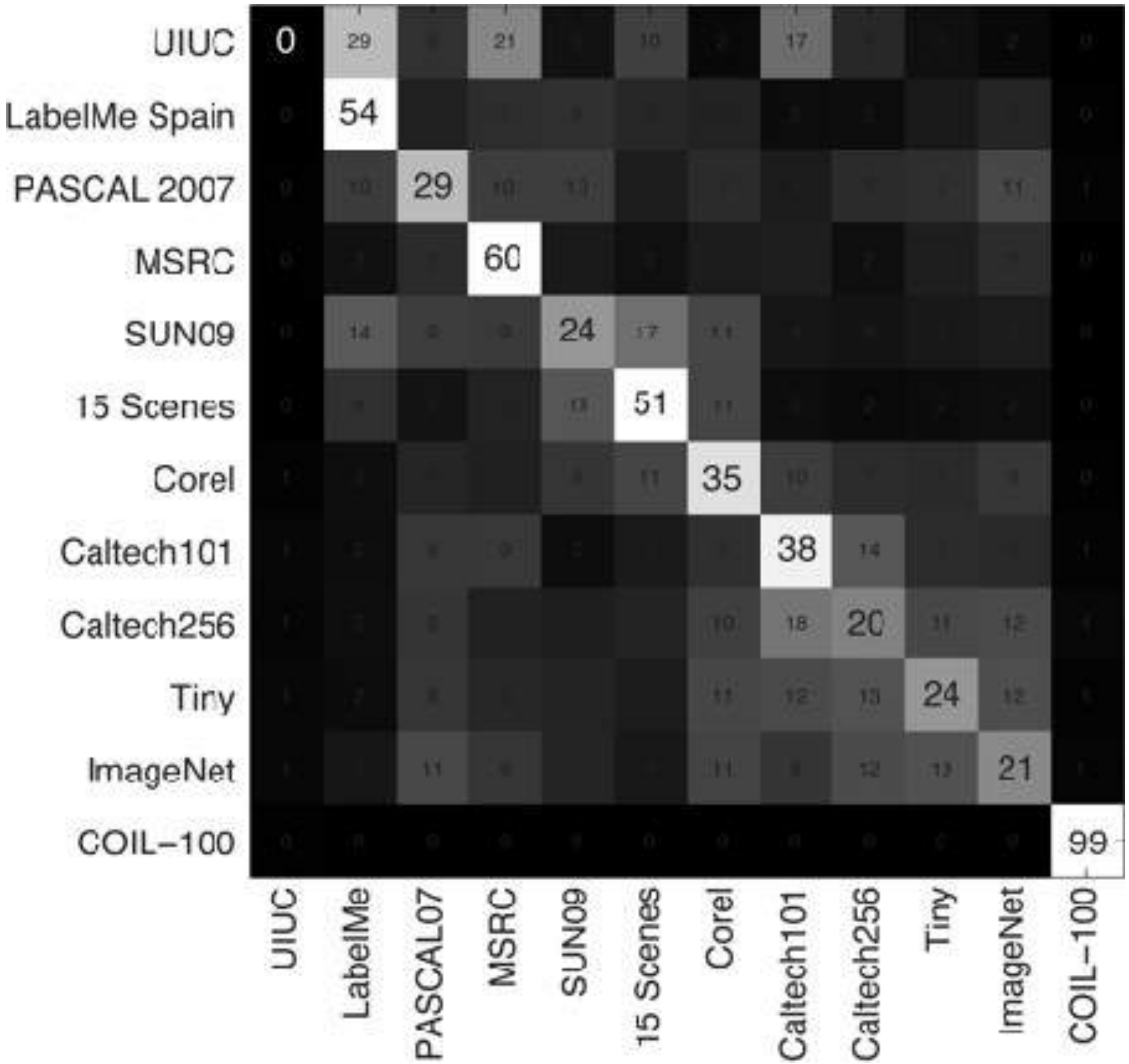


Mugs from LabelMe

"*Name That Dataset!*" game

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

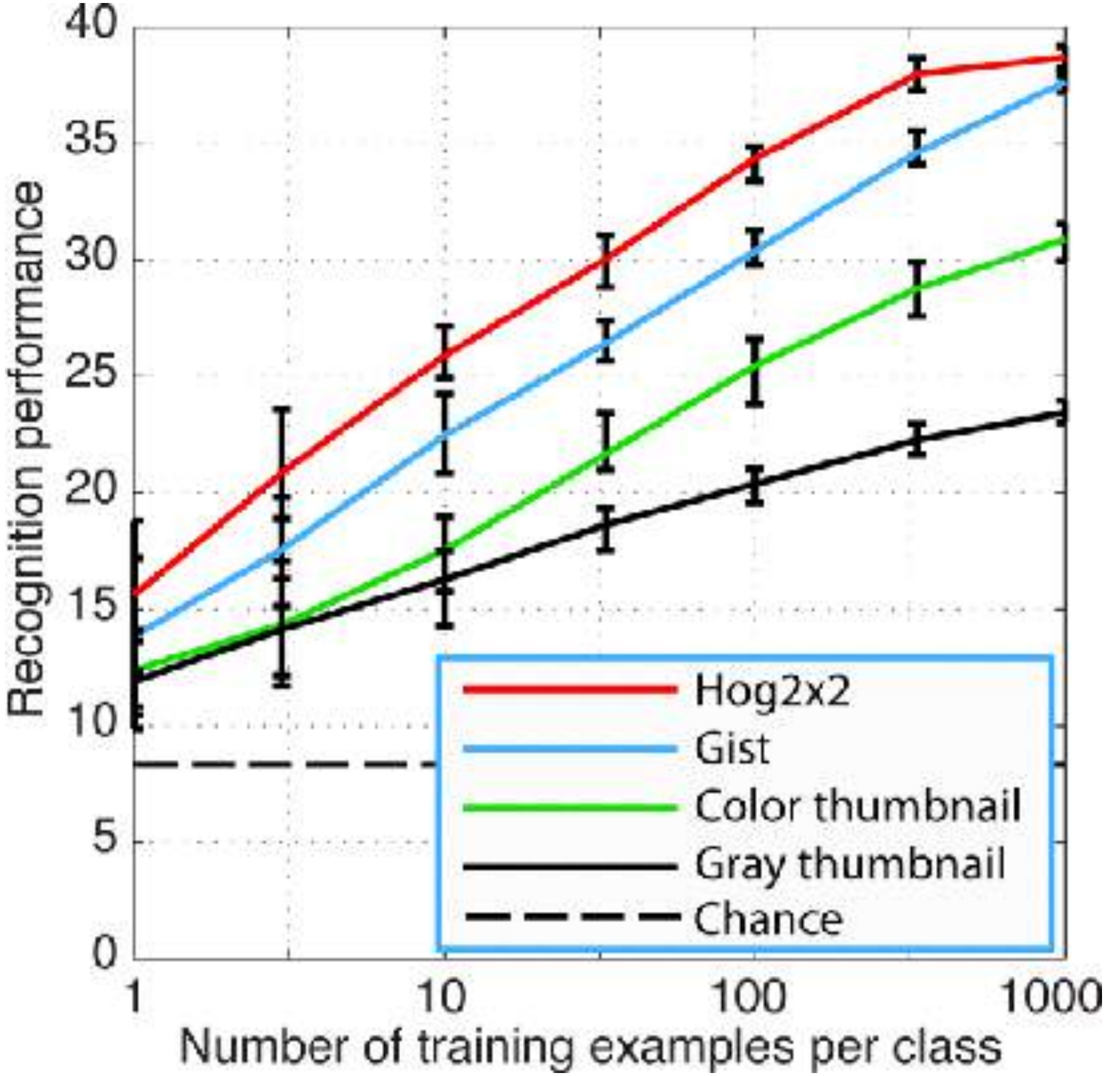
SVM plays *"Name that dataset!"*



- 12 1-vs-all classifiers
- Standard full-image features

- 39% performance (chance is 8%)

SVM plays "*Name that dataset!*"

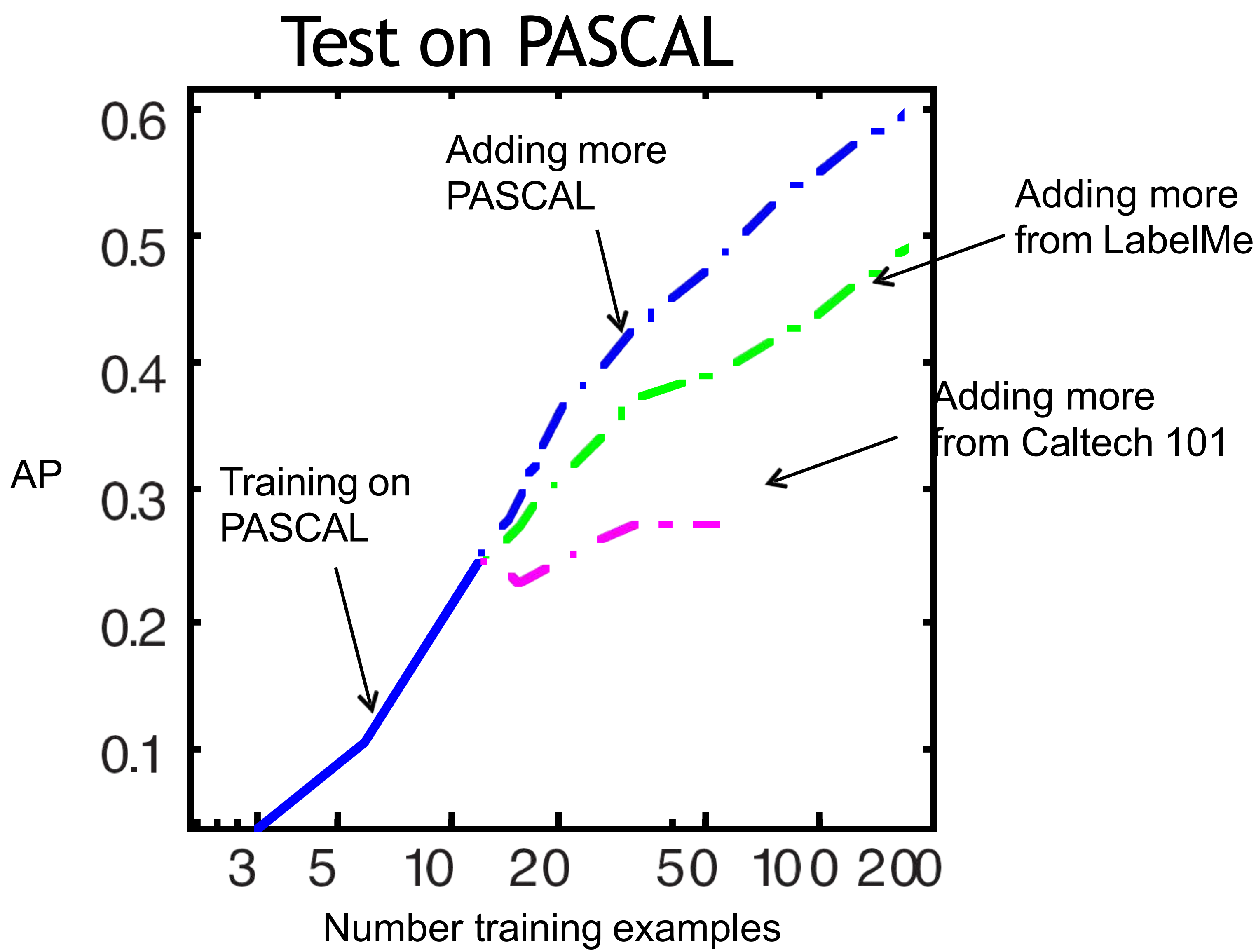


Datasets have different goals…

- 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?

Mixing datasets



- Why do we need data?
- How much data do we need?
- When do we need big data?
- How do we use big data?
- Why should we be careful of data?

The end