# Visual recognition at the Large-Scale

Fei-Fei Li

(publish under **L. Fei-Fei**)

Computer Science Dept.

Psychology Dept.

Stanford University


~10,000 to 30,000

# http://www.image-net.org

IMAGENET

9,956,478 images, 14841 synsets indexed

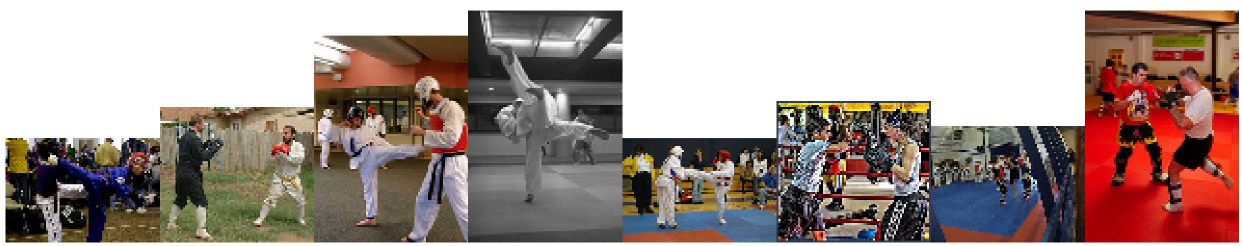
Explore Download **Challenge** New! People Publication Sponsors About

Not logged in. Login | Signup

**ImageNet** is an image database organized according to the WordNet hierarchy (currently only the nouns), in which each node of the hierarchy is depicted by hundreds and thousands of images. Currently we have an average of over five hundred images per node. We hope ImageNet will become a useful resource for researchers, educators, students and all of you who share our passion for pictures.

Click Here to learn more about ImageNet, Click Here to join ImageNet mailing list.

SEARCH

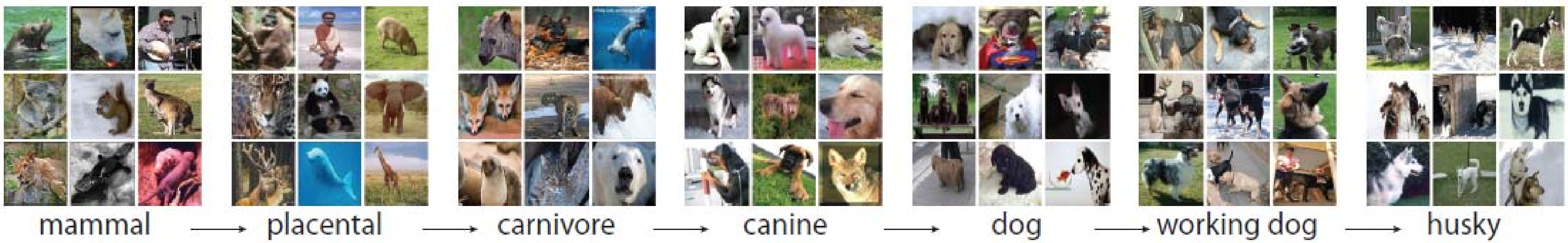


What do these images have in common? *Find out!*

Update Notice: ImageNet 2010 Spring Version will be released in April, 2010

# IMAGENET is a knowledge ontology

- Taxonomy




- S: (n) Eskimo dog, **husky** (breed of heavy-coated Arctic sled dog)
  - *direct hypernym* / ***inherited hypernym*** / *sister term*
    - S: (n) working dog (any of several breeds of usually large powerful dogs bred to work as draft animals and guard and guide dogs)
      - S: (n) dog, domestic dog, Canis familiaris (a member of the genus Canis (probably descended from the common wolf) that has been domesticated by man since prehistoric times; occurs in many breeds) *"the dog barked all night"*
        - S: (n) canine, canid (any of various fissiped mammals with nonretractile claws and typically long muzzles)
          - S: (n) carnivore (a terrestrial or aquatic flesh-eating mammal) *"terrestrial carnivores have four or five clawed digits on each limb"*
            - S: (n) placental, placental mammal, eutherian, eutherian mammal (mammals having a placenta; all mammals except monotremes and marsupials)
              - S: (n) mammal, mammalian (any warm-blooded vertebrate having the skin more or less covered with hair; young are born alive except for the small subclass of monotremes and nourished with milk)
                - S: (n) vertebrate, craniate (animals having a bony or cartilaginous skeleton with a segmented spinal column and a large brain enclosed in a skull or cranium)
                  - S: (n) chordate (any animal of the phylum Chordata having a notochord or spinal column)
                    - S: (n) animal, animate being, beast, brute, creature, fauna (a living organism characterized by voluntary movement)
                      - S: (n) organism, being (a living thing that has (or can develop) the ability to act or function independently)
                        - S: (n) living thing, animate thing (a living (or once living) entity)
                          - S: (n) whole, unit (an assemblage of parts that is regarded as a single entity) *"how big is that part compared to the whole?"; "the team is a unit"*
                            - S: (n) object, physical object (a tangible and visible entity; an entity that can cast a shadow) *"it was full of rackets, balls and other objects"*
                              - S: (n) physical entity (an entity that has physical existence)
                                - S: (n) entity (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))

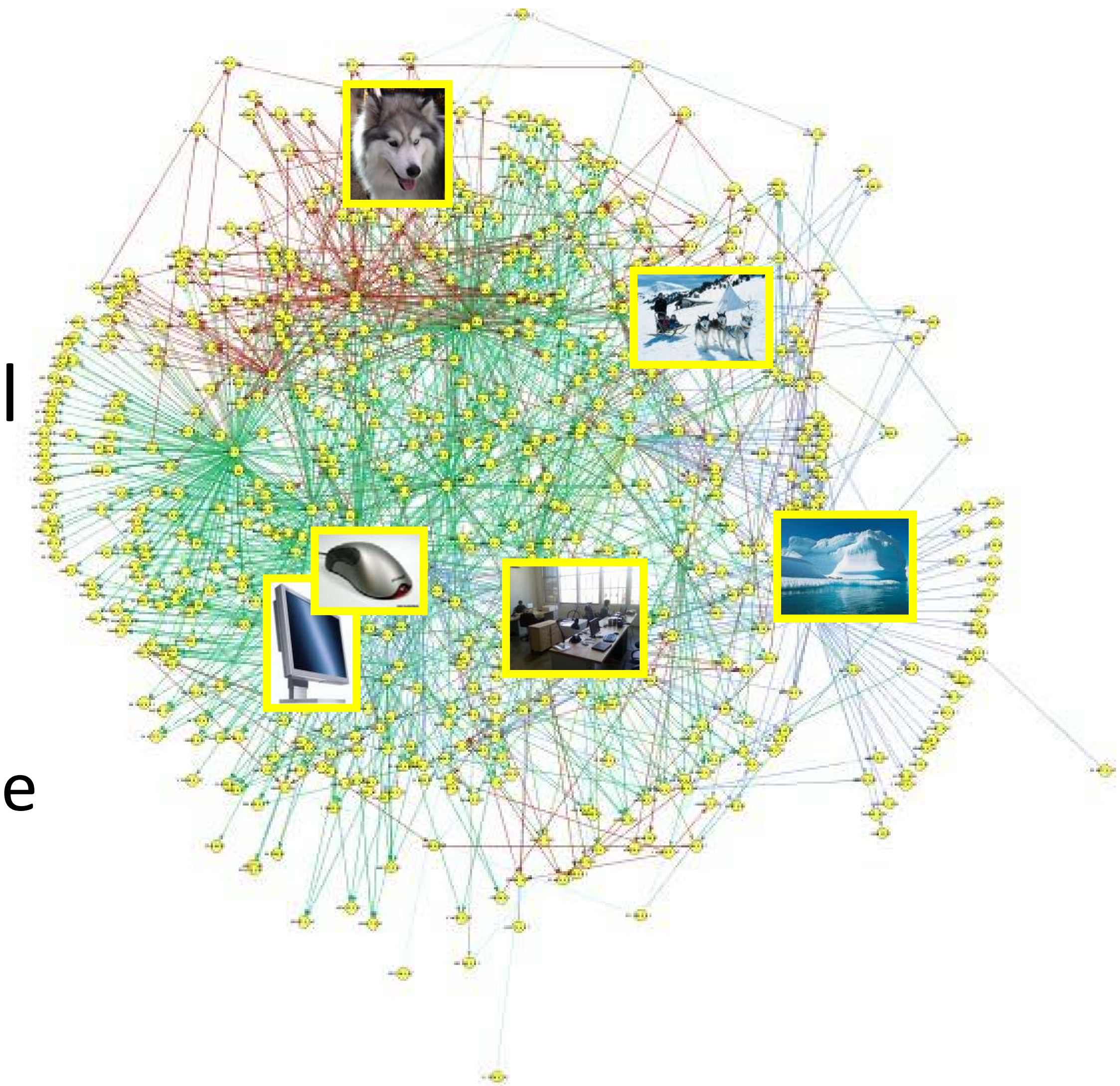
# IMAGENET is a knowledge ontology

- Taxonomy
- Partonomy

- S: (n) **car**, auto, automobile, machine, motorcar (a motor vehicle with four wheels; usually propelled by an internal combustion engine) *"he needs a car to get to work"*
  - *direct hyponym* / *full hyponym*
  - ***part meronym***
    - S: (n) accelerator, accelerator pedal, gas pedal, gas, throttle, gun (a pedal that controls the throttle valve) *"he stepped on the gas"*
    - S: (n) air bag (a safety restraint in an automobile; the bag inflates on collision and prevents the driver or passenger from being thrown forward)
    - S: (n) auto accessory (an accessory for an automobile)
    - S: (n) automobile engine (the engine that propels an automobile)
    - S: (n) automobile horn, car horn, motor horn, horn, hooter (a device on an automobile for making a warning noise)
    - S: (n) buffer, fender (a cushion-like device that reduces shock due to an impact)
    - S: (n) bumper (a mechanical device consisting of bars at either end of a vehicle to absorb shock and prevent serious damage)
    - S: (n) car door (the door of a car)
    - S: (n) car mirror (a mirror that the driver of a car can use)
    - S: (n) car seat (a seat in a car)
    - S: (n) car window (a window in a car)
    - S: (n) fender, wing (a barrier that surrounds the wheels of a vehicle to block splashing water or mud) *"in Britain they call a fender a wing"*
    - S: (n) first gear, first, low gear, low (the lowest forward gear ratio in the gear box of a motor vehicle; used to start a car moving)
    - S: (n) floorboard (the floor of an automobile)
    - S: (n) gasoline engine, petrol engine (an internal-combustion engine that burns gasoline; most automobiles are driven by gasoline engines)
    - S: (n) glove compartment (compartment on the dashboard of a car)
    - S: (n) grille, radiator grille (grating that admits cooling air to car's radiator)
    - S: (n) high gear, high (a forward gear with a gear ratio that gives the greatest vehicle velocity for a given engine speed)
    - S: (n) hood, bonnet, cowl, cowling (protective covering consisting of a metal part that covers the engine) *"there are powerful engines under the hoods of new cowling in order to repair the plane's engine"*
    - S: (n) luggage compartment, automobile trunk, trunk (compartment in an automobile that carries luggage or shopping or tools) *"he put his golf bag in the trunk*
    - S: (n) rear window (car window that allows vision out of the back of the car)
    - S: (n) reverse, reverse gear (the gears by which the motion of a machine can be reversed)
    - S: (n) roof (protective covering on top of a motor vehicle)
    - S: (n) running board (a narrow footboard serving as a step beneath the doors of some old cars)
    - S: (n) stabilizer bar, anti-sway bar (a rigid metal bar between the front suspensions and between the rear suspensions of cars and trucks; serves to stabilize the cl
    - S: (n) sunroof, sunshine-roof (an automobile roof having a sliding or raisable panel) *"'sunshine-roof' is a British term for 'sunroof'"*
    - S: (n) tail fin, tailfin, fin (one of a pair of decorations projecting above the rear fenders of an automobile)
    - S: (n) third gear, third (the third from the lowest forward ratio gear in the gear box of a motor vehicle) *"you shouldn't try to start in third gear"*
    - S: (n) window (a transparent opening in a vehicle that allow vision out of the sides or back; usually is capable of being opened)

# IMAGENET is a knowledge ontology

- Taxonomy
- Partonomy
- The "social network" of visual concepts
  - Prior knowledge
  - Context
  - Hidden knowledge and structure among visual concepts



# outline

- Construction of ImageNet
  - 2-step process
  - Crowdsourcing: Amazon Mechanical Turk (AMT)
  - Properties of ImageNet
- Benchmarking: what does classifying 10k+ image categories tell us?
  - Computation matters
  - Size matters
  - Density matters
  - Hierarchy matters
- A “semanticvisual” hierarchy for personal albums
  - Building it from Flickr images and user tags
  - Using the hierarchy for visual recognition tasks

# outline

- Construction of ImageNet
  - 2-step process
  - Crowdsourcing: Amazon Mechanical Turk (AMT)
  - Properties of ImageNet
- Benchmarking: what does classifying 10k+ image categories tell us?
  - Computation matters
  - Size matters
  - Density matters
  - Hierarchy matters
- A “semanticvisual” hierarchy for personal albums
  - Building it from Flickr images and user tags
  - Using the hierarchy for visual recognition tasks

# Constructing IMAGENET

Step 1:
Collect candidate images
via the Internet

→

Step 2:
Clean up the candidate
Images by humans

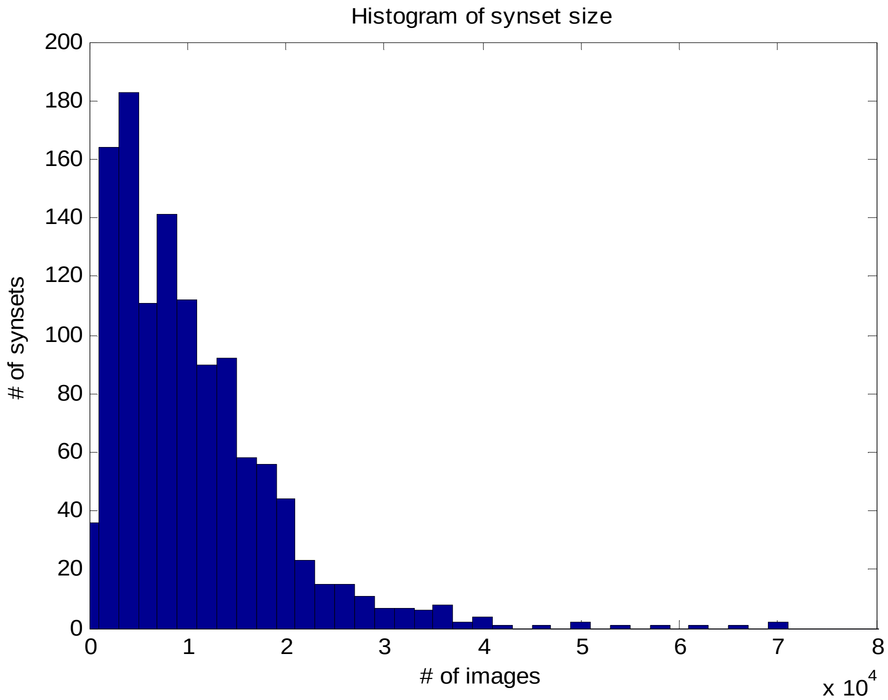
# Step 1: Collect Candidate Images from the Internet

- Query expansion
  - Synonyms: *German shepherd, German police dog, German shepherd dog, Alsatian*
  - Appending words from ancestors: *sheepdog, dog*
- Multiple languages
  - Italian, Dutch, Spanish, Chinese

    *e.g.* ovejero alemán, pastore tedesco,德国牧羊犬
- More engines
- Parallel downloading

# Step 1: Collect Candidate Images from the Internet

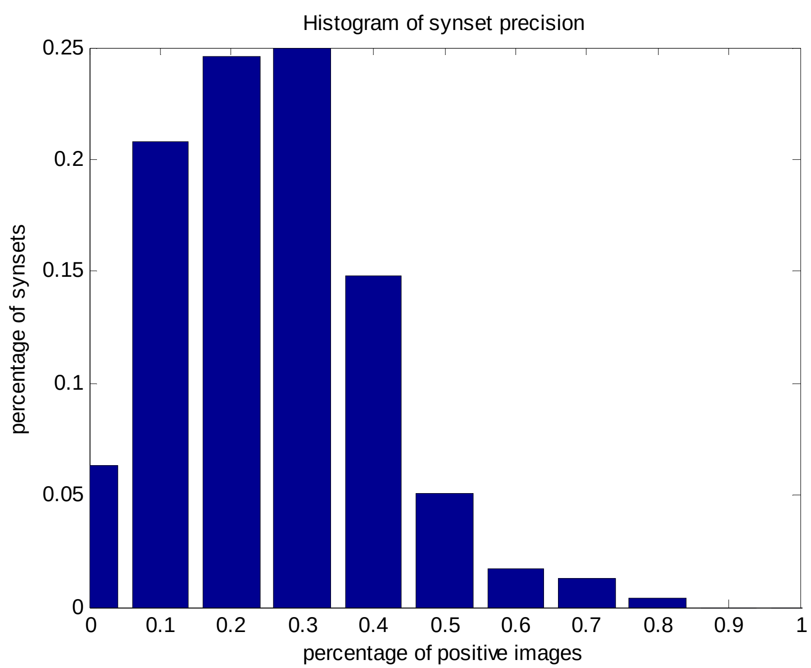
- “Mammal” subtree ( 1180 synsets )
  – Average # of images per synset: 10.5K




| Most populated | Least populated |
|---|---|
| Humankind (118.5k) | Algeripithecus minutus (90) |
| Kitty, kitty-cat ( 69k) | Striped muishond (107) |
| Cattle, cows ( 65k) | Mylodonitid (127) |
| Pooch, doggie ( 62k) | Greater pichiciego (128) |
| Cougar, puma ( 57k) | Damaraland mole rat (188) |
| Frog, toad ( 53k ) | Western pipistrel (196) |
| Hack, jade, nag (50k) | Muishond (215) |

# Step 1: Collect Candidate Images from the Internet

- "Mammal" subtree (1180 synsets )
  - Average accuracy per synset: 26%




| Most accurate | Least accurate |
|---|---|
| Bottlenose dolpin (80%) | Fanaloka (1%) |
| Meerkat (74%) | Pallid bat (3%) |
| Burmese cat (74%) | Vaquita (3%) |
| Humpback whale (69%) | Fisher cat (3%) |
| African elephant (63%) | Walrus (4%) |
| Squirrel (60%) | Grison (4%) |
| Domestic cat (59%) | Pika, Mouse hare (4%) |

# Step 2: verifying the images by humans

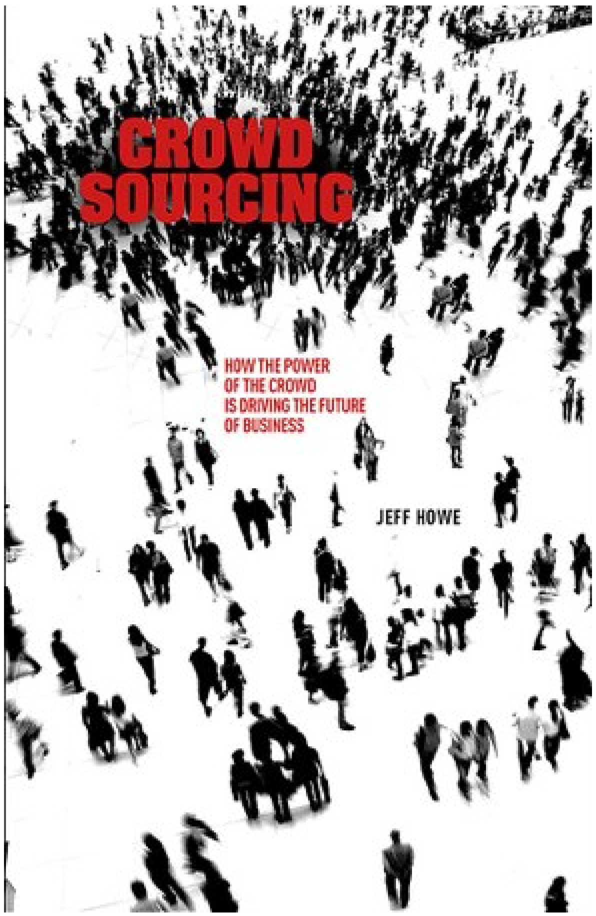
- # of synsets: 40,000 (subject to: imageability analysis)
- # of candidate images to label per synset: 10,000
- # of people needed to verify: 2-5
- Speed of human labeling: 2 images/sec (one fixation: ~200msec)

$$\boxed{40{,}000 \times 10{,}000 \times 3 / 2 = 600{,}000{,}000\,\text{sec} \approx 19\,\text{years}}$$

Moral of the story:

no graduate students would want to do this project!

# In summer 2008, we discovered crowdsourcing




amazon mechanical turk
Artificial Artificial Intelligence

Al
Sign in a

Your Account | HITs | Qualifications

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.

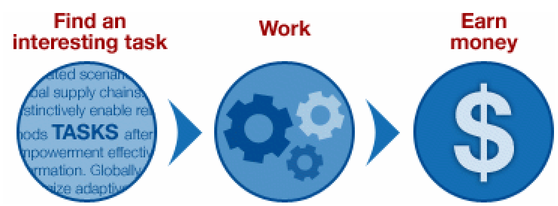
**149,499 HITs** available. View them now.

# 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 HITs Now

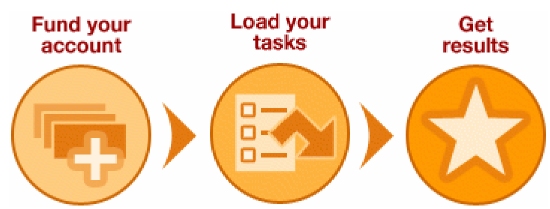
or learn more about being a **Worker**

# 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




Get Started

# Step 2: verifying the images by humans

- # of synsets: 40,000 (subject to: imageability analysis)
- # of candidate images to label per synset: 10,000
- # of people needed to verify: 2-5
- Speed of human labeling: 2 images/sec (one fixation: ~200msec)
- Massive parallelism (N ~ 10^2-3)

$$40{,}000 \times 10{,}000 \times 3 / 2 = 600{,}000{,}000 \sec \approx \frac{19 \text{years}}{N}$$

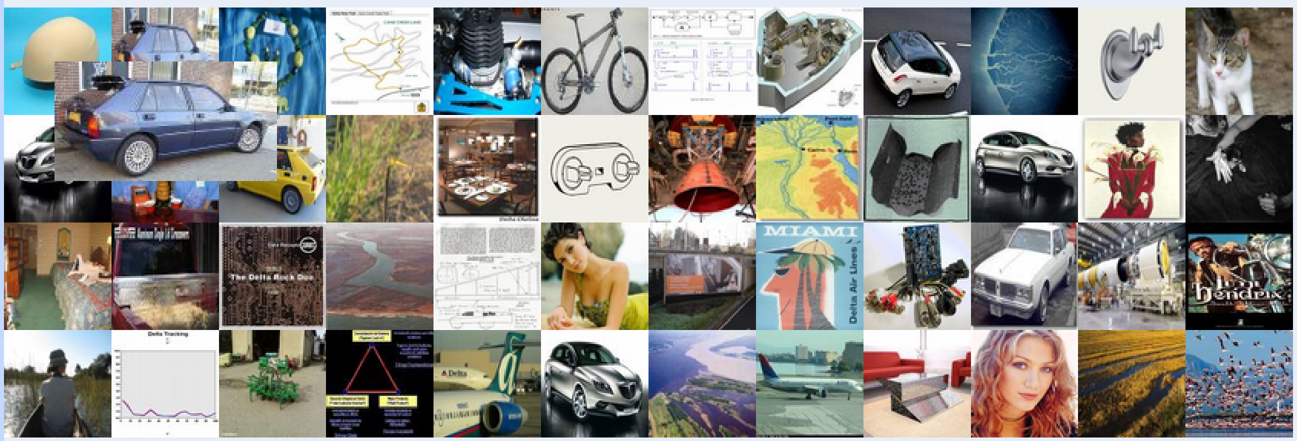
# IMAGENET Basic User Interface

Click on the good images.

Main | Instructions | Unsure? Look up in Wikipedia | Google | [ *Additional input* ] No good photos? Have expertise? comments? Click here!

**First time workers please click here for instructions.**

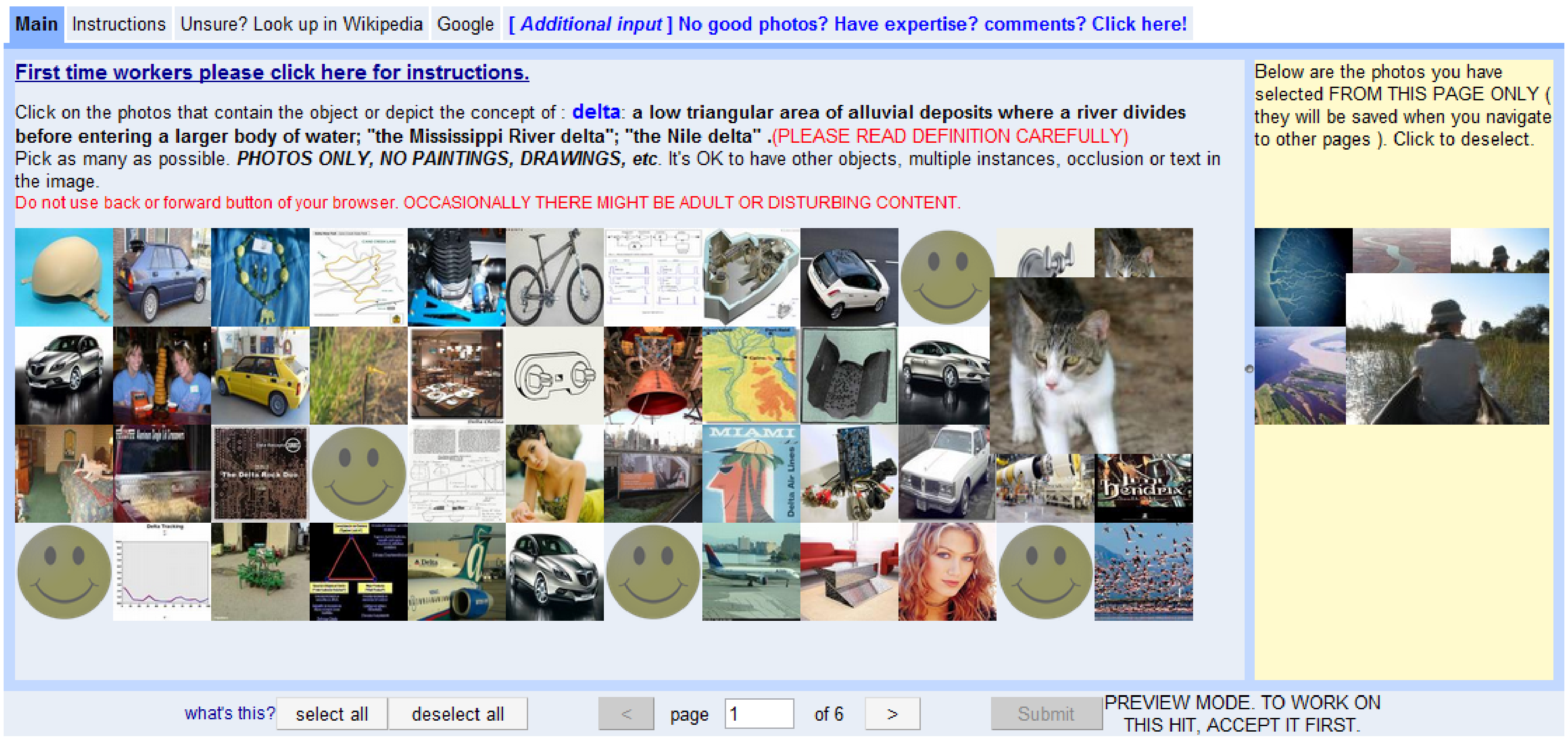
Click on the photos that contain the object or depict the concept of : **delta**: **a low triangular area of alluvial deposits where a river divides before entering a larger body of water; "the Mississippi River delta"; "the Nile delta"** .(PLEASE READ DEFINITION CAREFULLY)
Pick as many as possible. ***PHOTOS ONLY, NO PAINTINGS, DRAWINGS, etc***. It's OK to have other objects, multiple instances, occlusion or text in the image.
Do not use back or forward button of your browser. OCCASIONALLY THERE MIGHT BE ADULT OR DISTURBING CONTENT.




Below are the photos you have selected FROM THIS PAGE ONLY ( they will be saved when you navigate to other pages ). Click to deselect.

what's this? | select all | deselect all | < | page 1 of 6 | > | Submit | PREVIEW MODE. TO WORK ON THIS HIT, ACCEPT IT FIRST.

# IMAGENET Basic User Interface

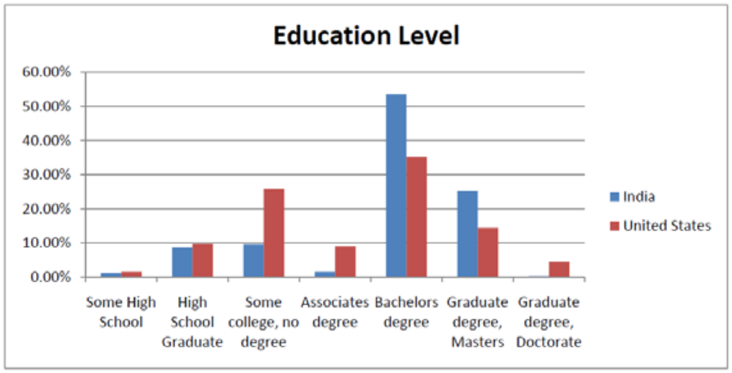


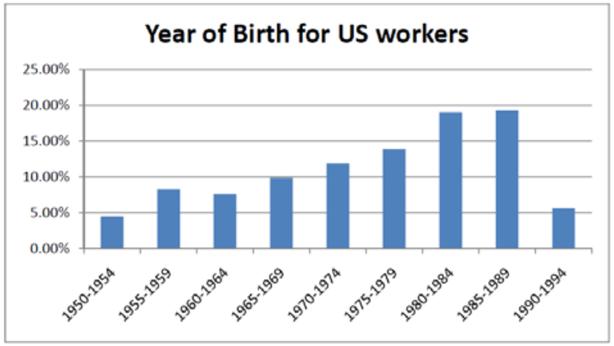
# So are we exploiting chained prisoners?

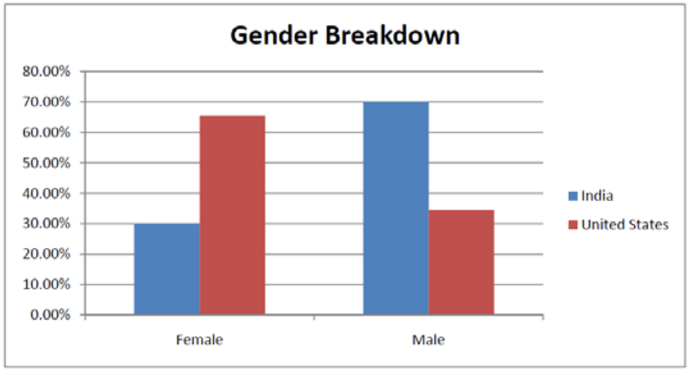
amazon mechanical turk
beta
Artificial Artificial Intelligence

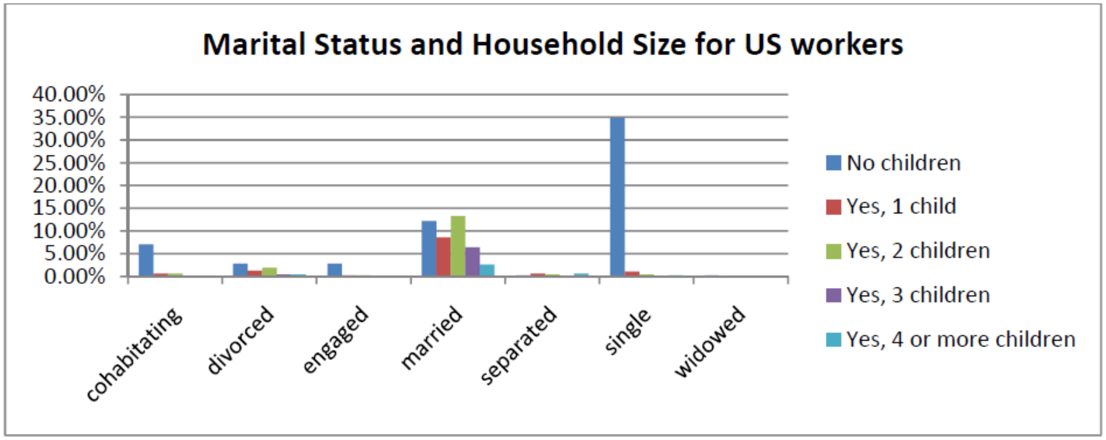
# Demography of AMT workers

| | |
|---|---|
| United States | 46.80% |
| **India** | **34.00%** |
| Miscellaneous | 19.20% |



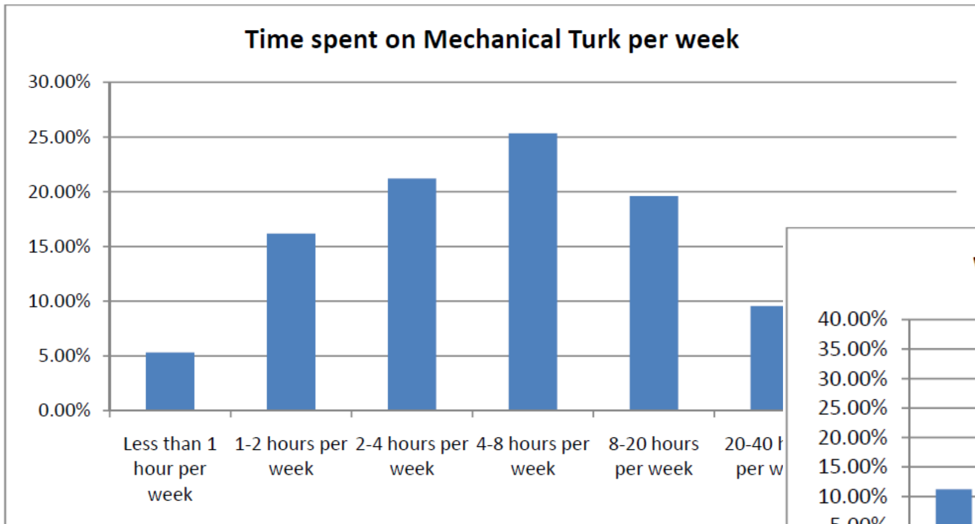



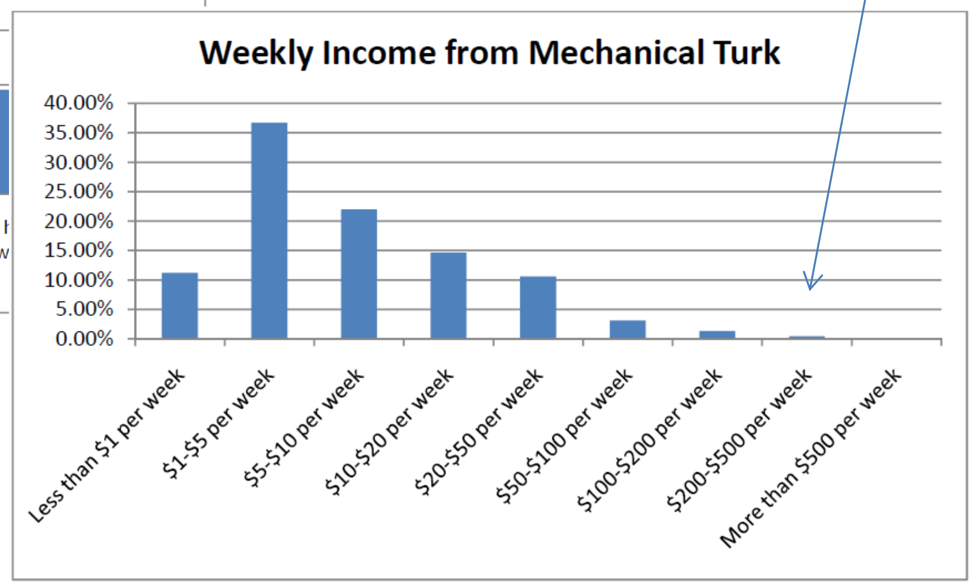




**Panos Ipeirotis, NYU, Feb, 2010**

# Demography of AMT workers

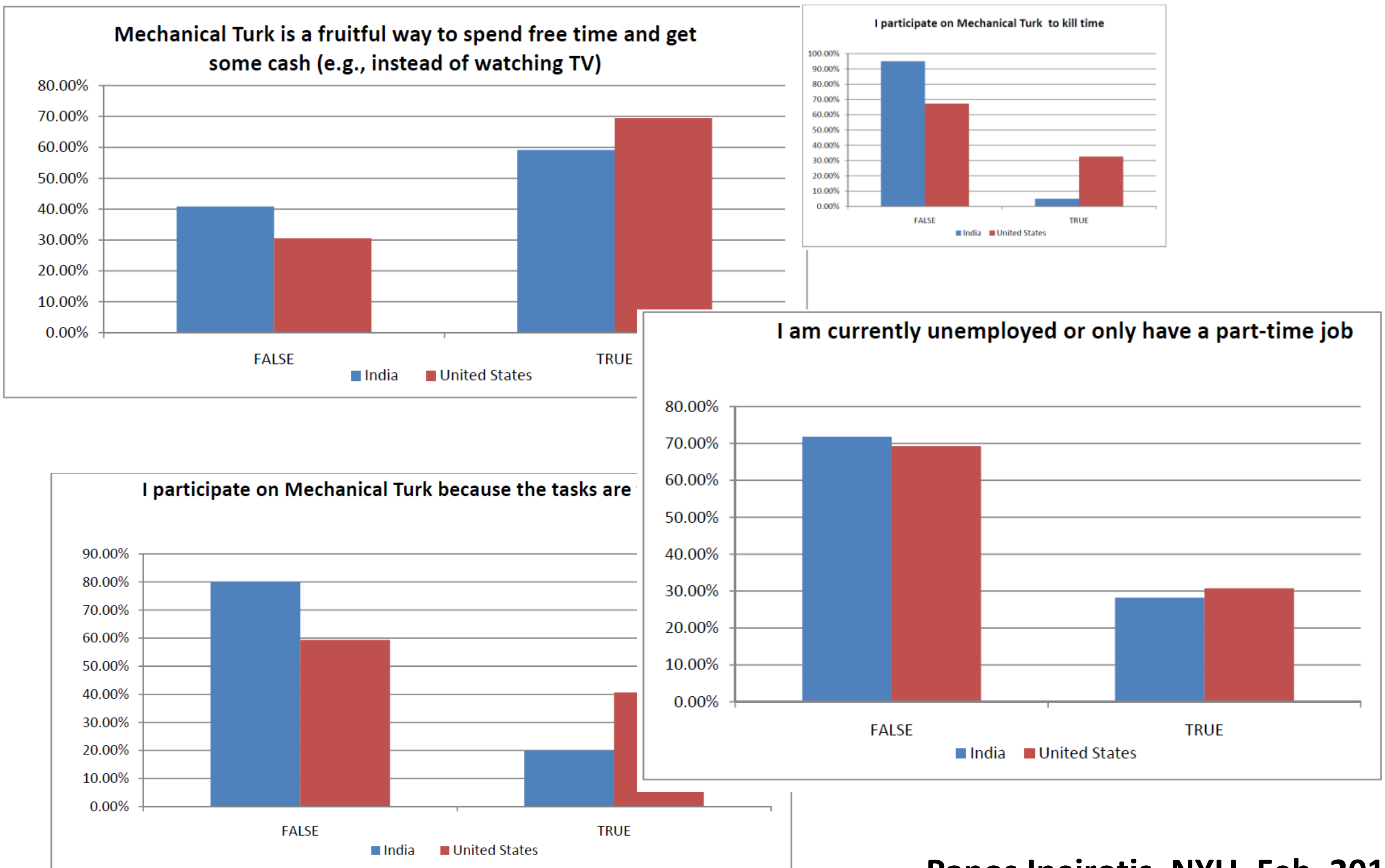



Typical Stanford
Graduate student's income



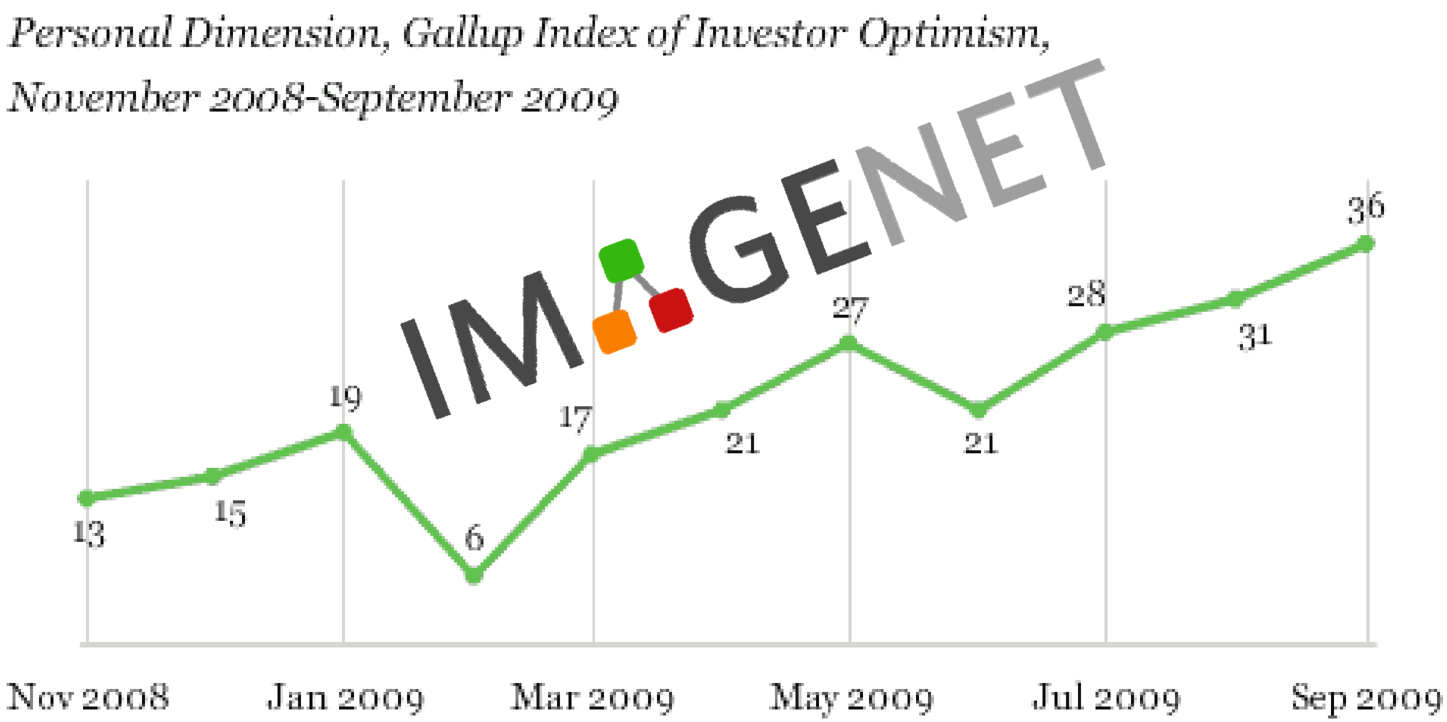

**Panos Ipeirotis, NYU, Feb, 2010**

# Demography of AMT workers




**Panos Ipeirotis, NYU, Feb, 2010**

# U.S. economy 2008 - 2009



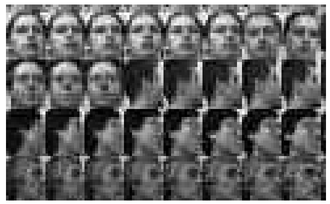

IMAGENET hired more than 25,000 AMT workers in this period of time!!

# outline

- Construction of ImageNet
  - 2-step process
  - Crowdsourcing: Amazon Mechanical Turk (AMT)
  - Properties of ImageNet
- Benchmarking: what does classifying 10k+ image categories tell us?
  - Computation matters
  - Size matters
  - Density matters
  - Hierarchy matters
- A “semanticvisual” hierarchy for personal albums
  - Building it from Flickr images and user tags
  - Using the hierarchy for visual recognition tasks

# Datasets and computer vision

**UIUC Cars (2004)**
S. Agarwal, A. Awan, D. Roth

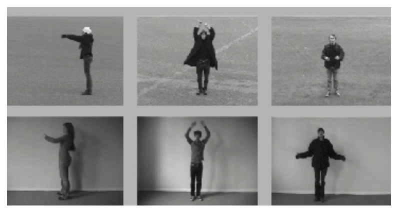


**CMU/VASC Faces (1998)**
H. Rowley, S. Baluja, T. Kanade

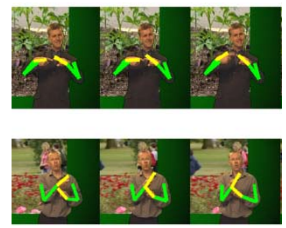
**FERET Faces (1998)**
P. Phillips, H. Wechsler, J. Huang, P. Raus

**COIL Objects (1996)**
S. Nene, S. Nayar, H. Murase

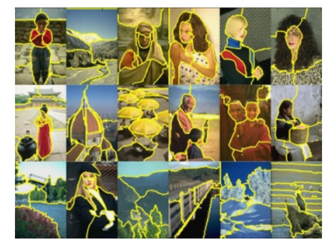
**MNIST digits (**1998-10)
Y LeCun & C. Cortes



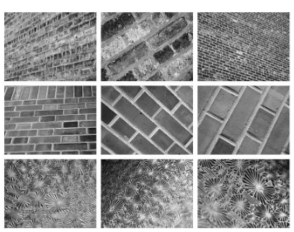
**KTH human action** (2004)
I. Leptev & B. Caputo



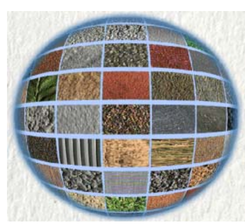
**Sign Language** (2008)
P. Buehler, M. Everingham, A. Zisserman



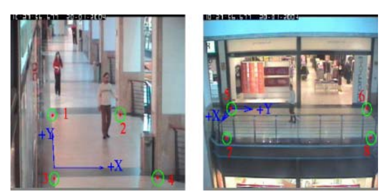
**Segmentation** (2001)
D. Martin, C. Fowlkes, D. Tal, J. Malik.



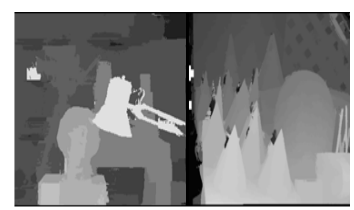
**3D Textures** (2005)
S. Lazebnik, C. Schmid, J. Ponce



**CuRRET Textures** (1999)
K. Dana B. Van Ginneken S. Nayar J. Koenderink



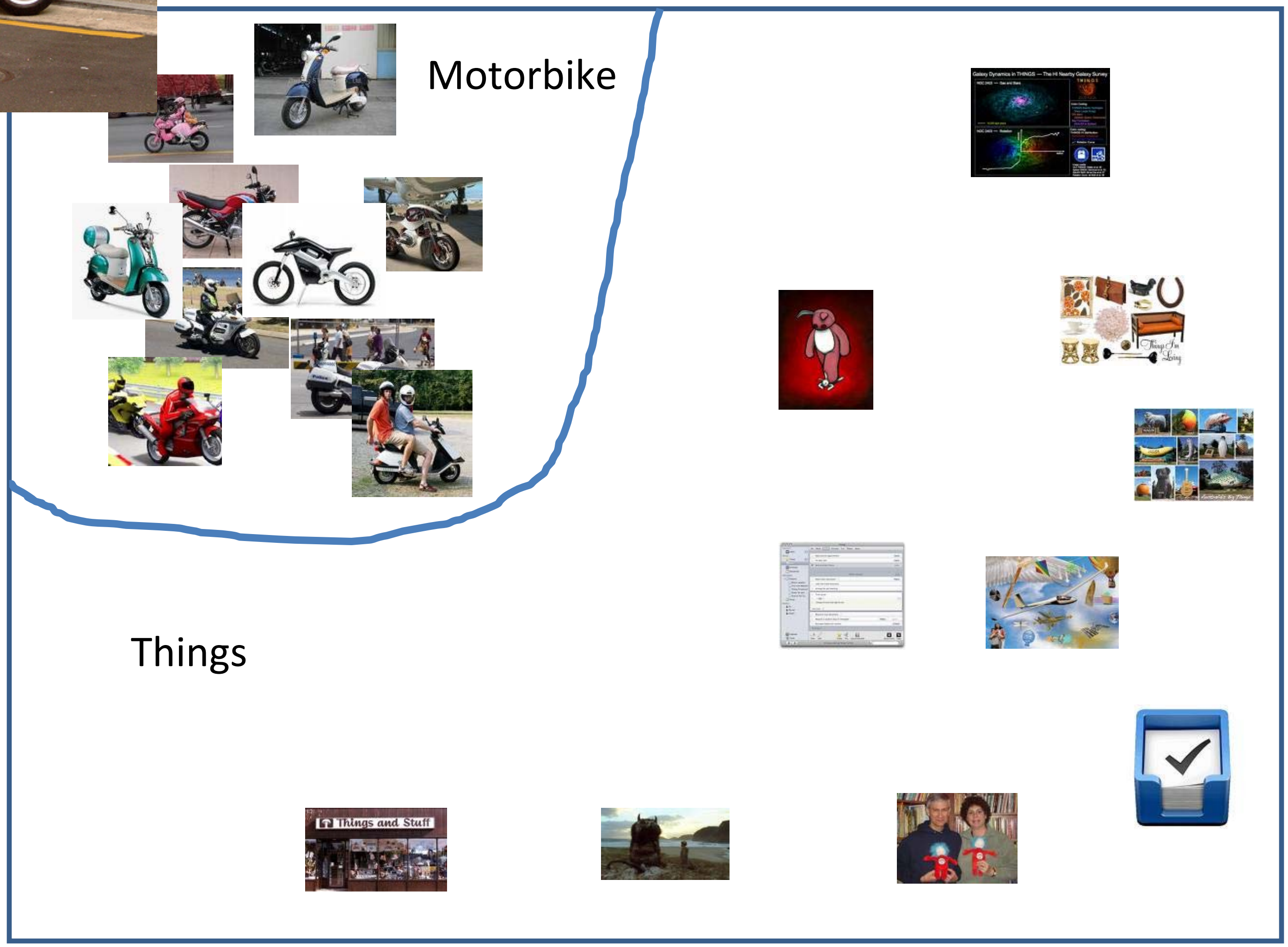
**CAVIAR Tracking** (2005)
R. Fisher, J. Santos-Victor J. Crowley



**Middlebury Stereo** (2002)
D. Scharstein R. Szeliski

Fergus, Perona, Zisserman, CVPR 2003

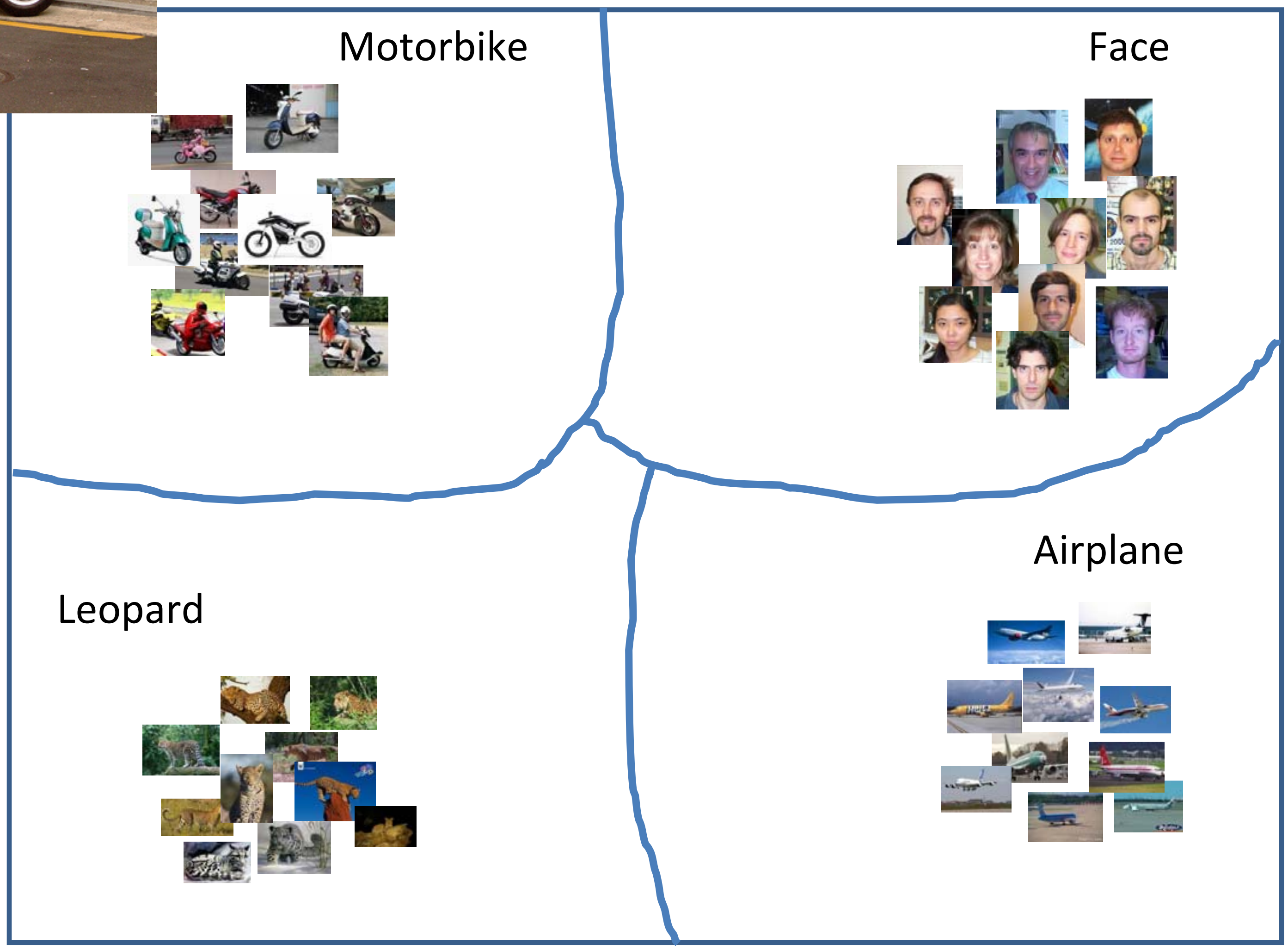
# Object Recognition




# Object Recognition

Fergus, Perona, Zisserman, CVPR 2003

Holub, et al. ICCV 2005; Sivic et al. ICCV 2005

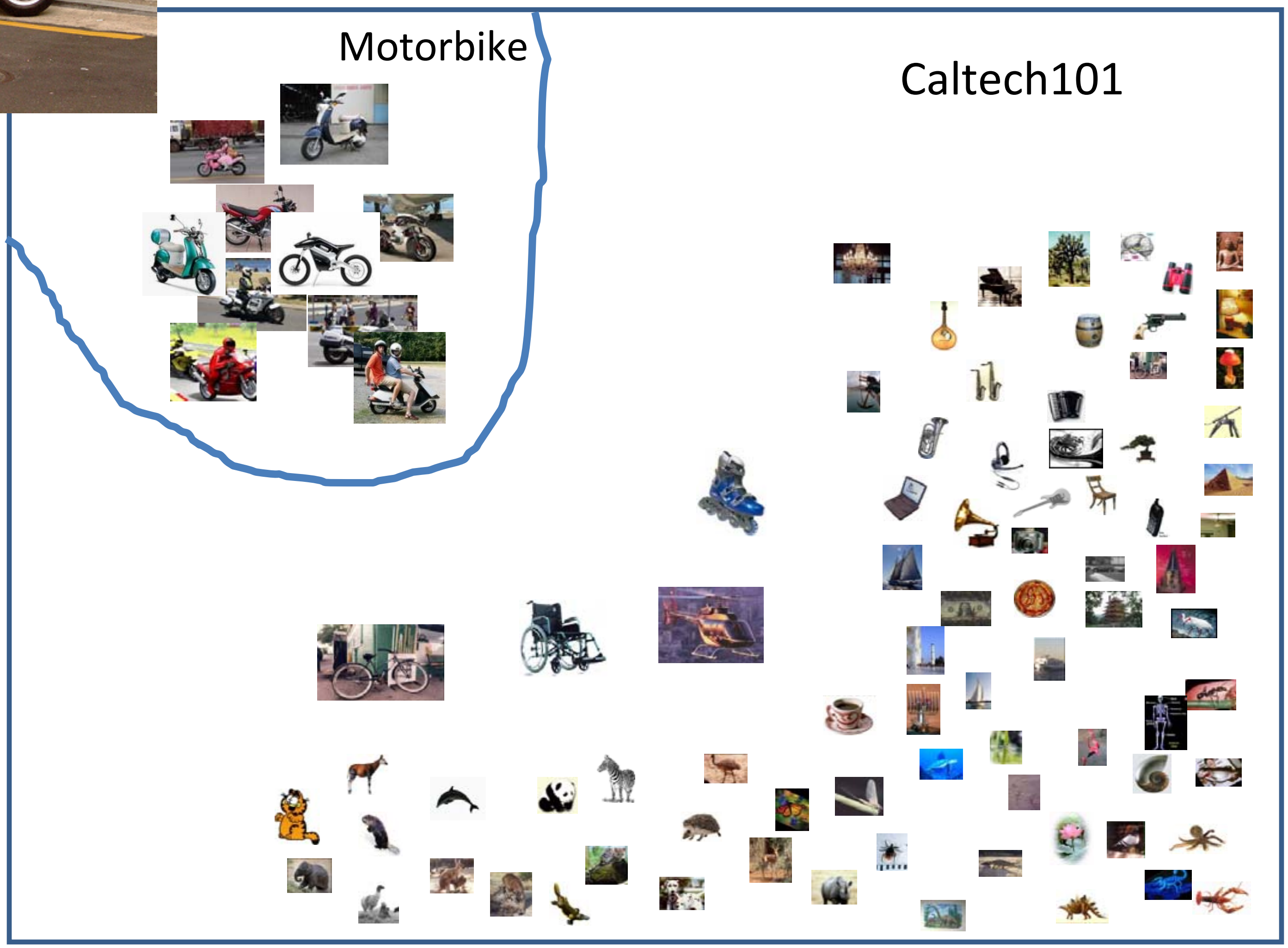



Fergus, Perona, Zisserman, CVPR 2003

Holub, et al. ICCV 2005; Sivic et al. ICCV 2005

Fei-Fei et al. CVPR 2004; Grauman et al. ICCV 2005; Lazebnik et al. CVPR 2006
Zhang & Malik, 2006; Varma & Sizzerman 2008; Wang et al. 2006; [....]

# Object Recognition




PASCAL

[Everingham et al, 2009]

MSRC

[Shotton et al. 2006]


Fergus, Perona, Zisserman, CVPR 2003

Holub, et al. ICCV 2005; Sivic et al. ICCV 2005

Fei-Fei et al. CVPR 2004; Grauman et al. ICCV 2005; Lazebnik et al. CVPR 2006
Zhang & Malik, 2006; Varma & Sizzerman 2008; Wang et al. 2006; [….]

Biederman 1987

# Object Recognition

ESP

[Ahn et al, 2006]

LabelMe

[ Russell et al, 2005]

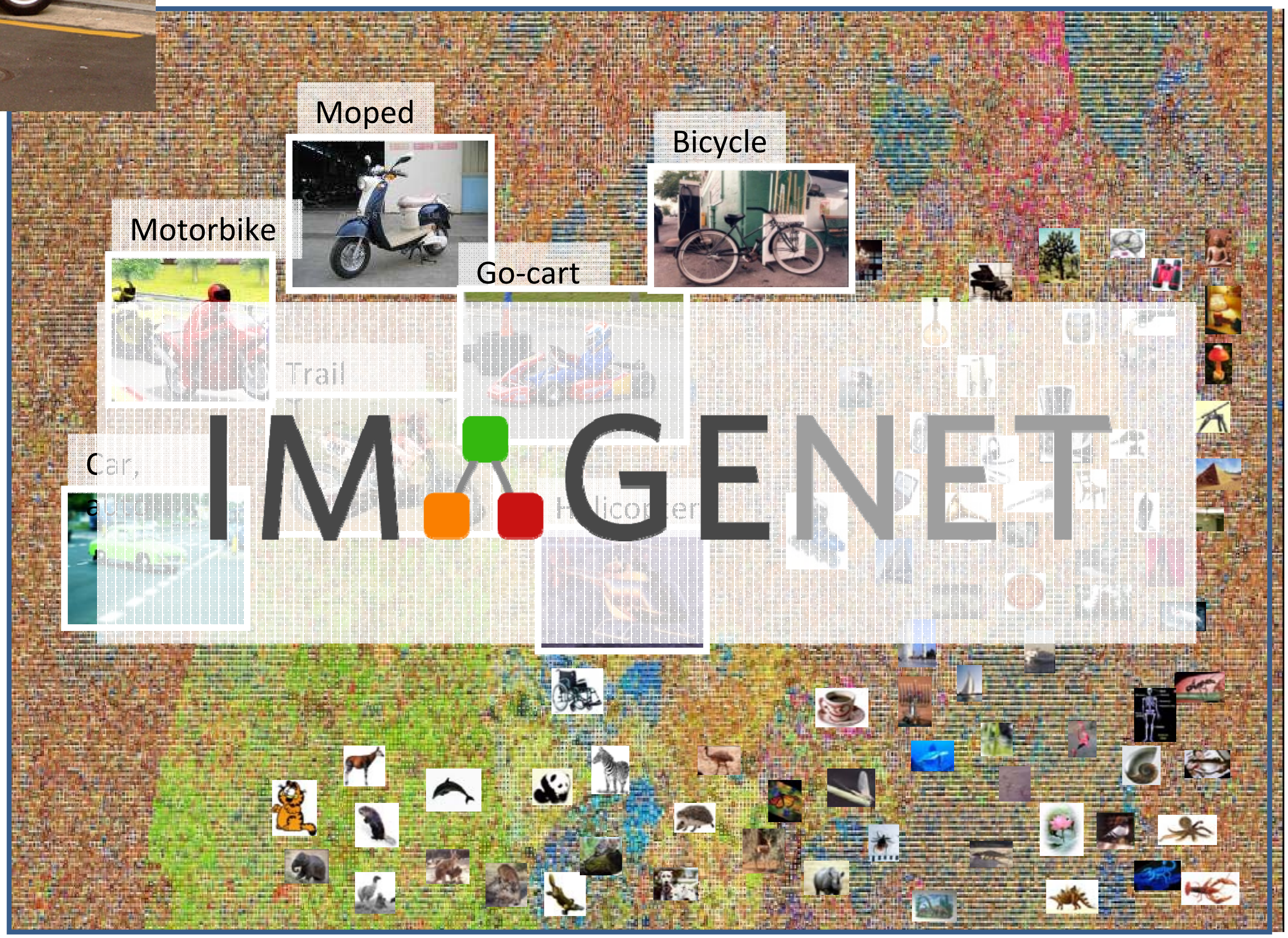
TinyImage

Torralba et al. 2007

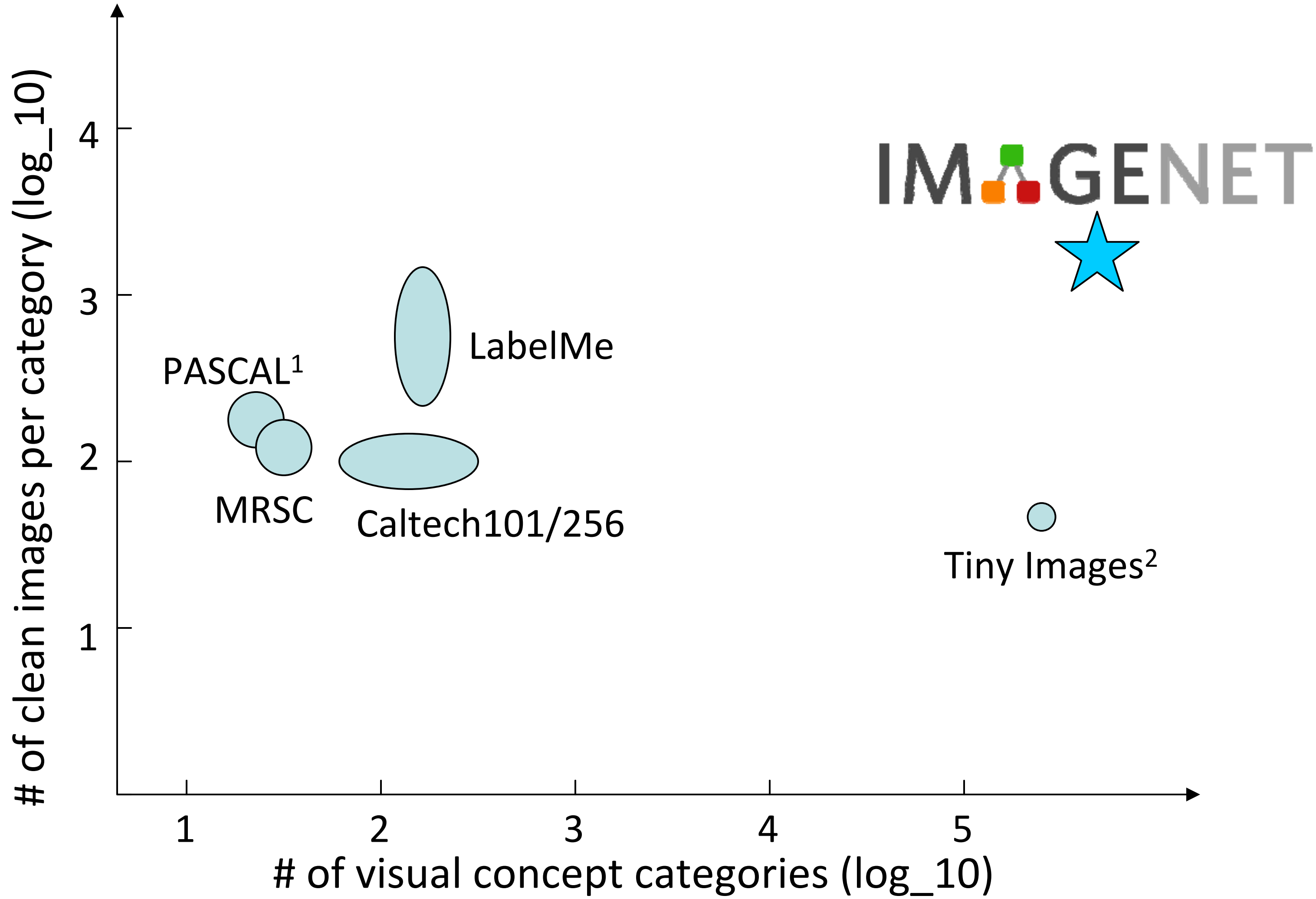
Lotus Hill

[ Yao et al, 2007]

Background image courtesy: Antonio Torralba



# Comparison among free datasets




1. Excluding the Caltech101 datasets from PASCAL
2. No image in this dataset is human annotated. The # of clean images per category is a rough estimation

# Basic evaluation setup

- IMAGENET
  - 10,000 categories
  - 9 million images
  - 50%-50% train test split
- Multi-class classification in 1-vs-all framework
  - GIST+NN: filter banks; nearest neighbor (Oliva & Torralba, 2001)
  - BOW+NN: SIFT, 1000 codewords, BOW; nearest neighbor
  - BOW+SVM: SIFT, 1000 codewords, BOW; linear SVM
  - SPM+SVM: SIFT, 1000 codewords, Spatial Pyramid; intersection kernel SVM (Lazebnik et al. 2006)

Deng, Berg, Li, & Fei-Fei, *submitted*

# Computation issues first

- BOW+SVM
  - Train one 1-vs-all with LIBLINEAR → 1 CPU hour
  - 10,000 categories → 1 CPU year
- SPM + SVM
  - Maji & Berg 2009, LIBLINEAR with piece-wise linear encoding
  - Memory bottleneck. Modification required.
  - 10,000 categories → 6 CPU year
- Parallelized on a cluster
  - Weeks for a single run of experiments

Deng, Berg, Li, & Fei-Fei, *submitted*

# Size matters

- 6.5% for 10K categories
- Better than we expected (instead of dropping at the rate of 10x; it's roughly at about 2x)
- An ordering switch between SVM and NN methods when the # of categories becomes large

Some unpublished results omitted.

Deng, Berg, Li, & Fei-Fei, *submitted*

# Size matters

- 6.5% for 10K categories
- Better than we expected (instead of dropping at the rate of 10x; it's roughly at about 2x)
- An ordering switch between SVM and NN methods when the # of categories becomes large
- When dataset size varies, conclusion we can draw about different categories varies

Some unpublished results omitted.

Deng, Berg, Li, & Fei-Fei, *submitted*

# Size matters

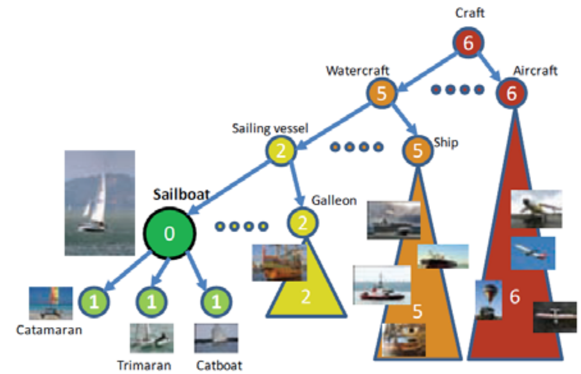
- 6.5% for 10K categories
- Better than we expected (instead of dropping at the rate of 10x; it's roughly at about 2x)
- An ordering switch between SVM and NN methods when the # of categories becomes large
- When dataset size varies, conclusion we can draw about different categories varies
- Purely semantic organization of concepts (by WordNet) exhibits meaningful visual structure (ordered by DFS)

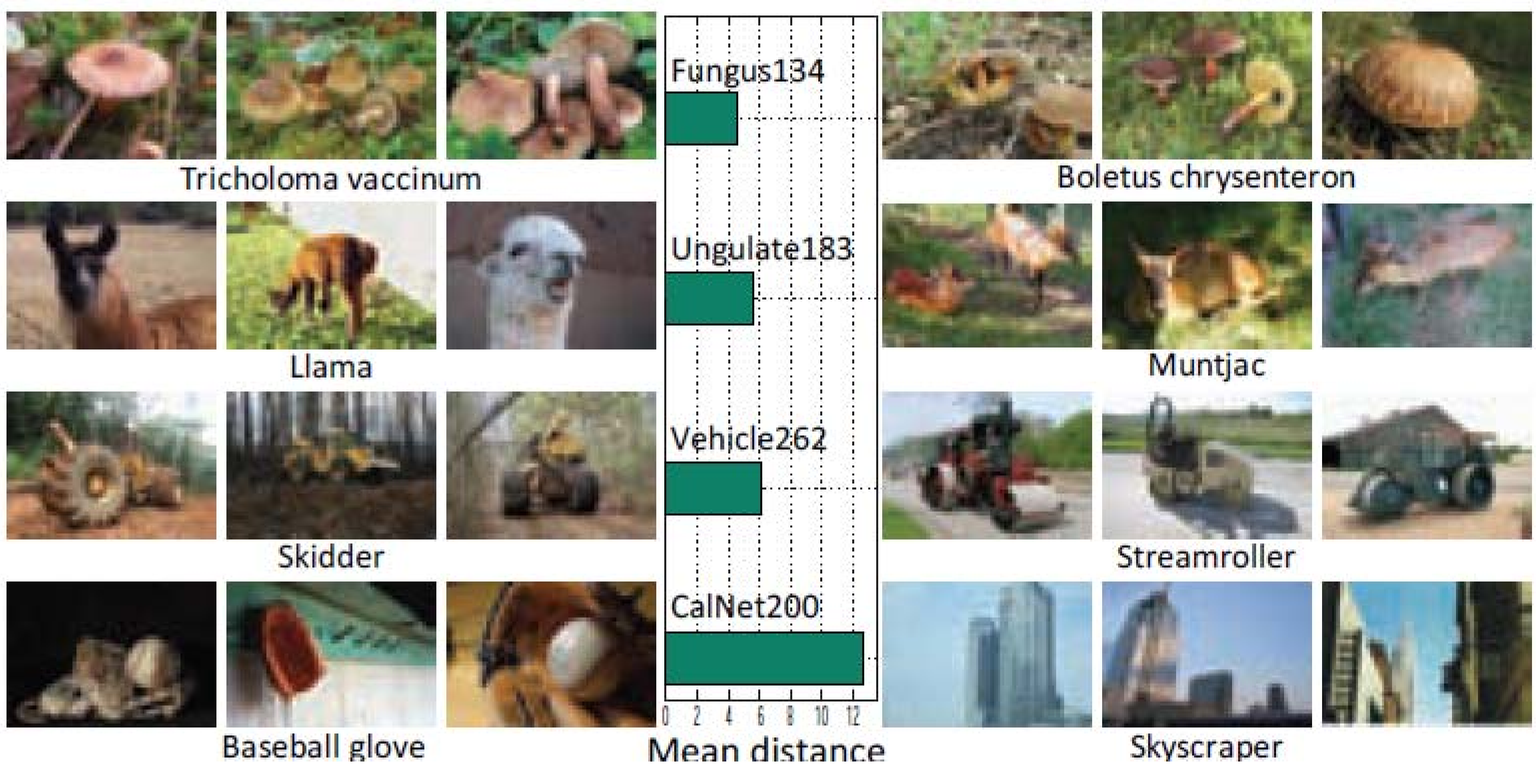
Some unpublished results omitted.

Deng, Berg, Li, & Fei-Fei, *submitted*

# Density matters

- Datasets have very different "density" or "sparcity"




Deng, Berg, Li, & Fei-Fei, *submitted*

# Density matters

- Datasets have very different “density” or “sparcity”
- there is a significant difference in difficulty between different datasets, independent of feature and classier choice.

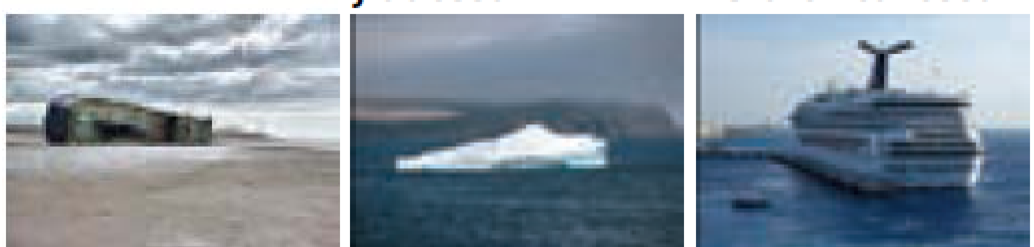
Some unpublished results omitted.

Deng, Berg, Li, & Fei-Fei, *submitted*

# Hierarchy matters

- Classifying a “dog” as “cat” is probably not as bad as classifying it as “microwave”
- A simple way to incorporate classification cost

***Query*** | ***Prediction flat cost*** | ***Prediction hierarchical cost***



Shipwreck | Iceberg (17) | Cruise ship (4)

Deng, Berg, Li, & Fei-Fei, *submitted*

# Hierarchy matters

- Classifying a "dog" as "cat" is probably not as bad as classifying it as "microwave"
- A simple way to incorporate hierarchical classification cost



Deng, Berg, Li, & Fei-Fei, *submitted*

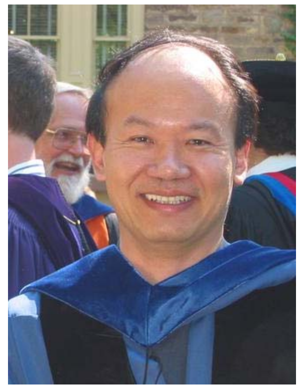
# IMAGENET is team work!

## WordNet friends

Christiane Fellbaum
Princeton U.

Dan Osherson
Princeton U.

## co-PI

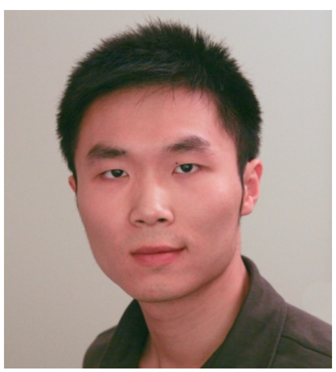


Kai Li
Princeton U.

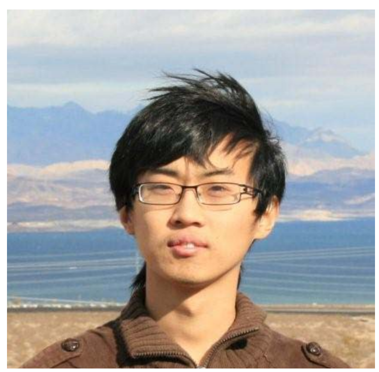
## Research collaborator; ImageNet Challenge boss

Alex Berg
Columbia U.

## Graduate students



Jia Deng
Princeton/Stanford



Hao Su
Stanford U.

Other contributors

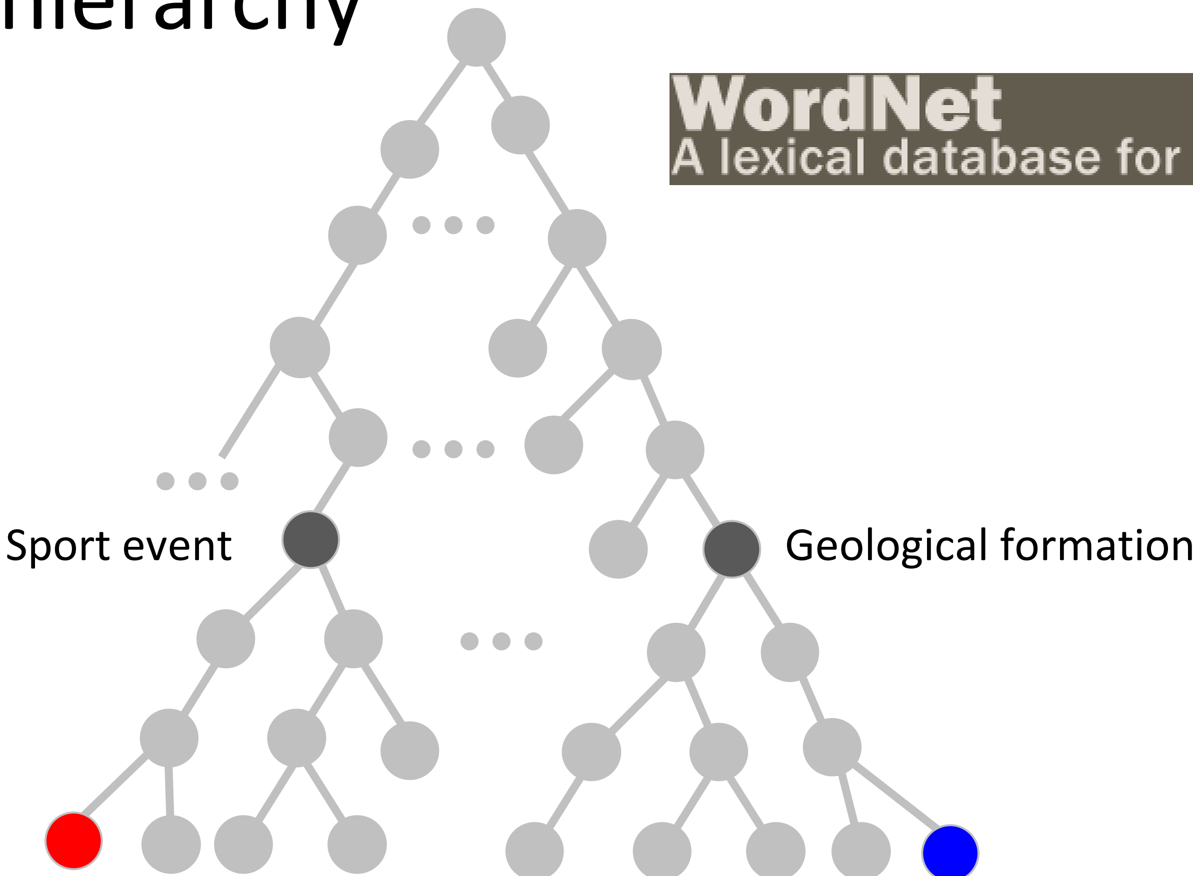
- Princeton graduate students
  - Wei Dong
  - Zhe Wang
- Stanford graduate students
  - John Le
  - Pao Siangliulue
- AMT partner
  - Dolores Lab

# outline

- Construction of ImageNet
  - 2-step process
  - Crowdsourcing: Amazon Mechanical Turk (AMT)
  - Properties of ImageNet
- Benchmarking: what does classifying 10k+ image categories tell us?
  - Computation matters
  - Size matters
  - Density matters
  - Hierarchy matters
- A “semanticvisual” hierarchy for personal albums
  - Building it from Flickr images and user tags
  - Using the hierarchy for visual recognition tasks

# Semantic hierarchy

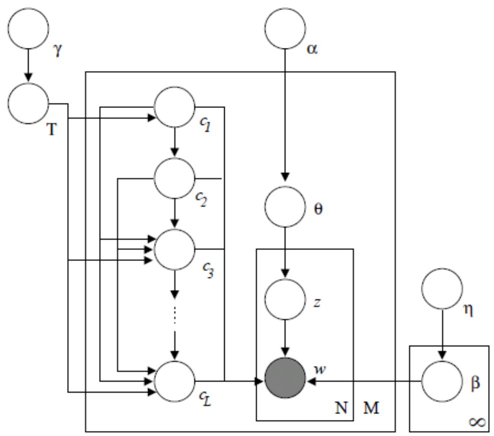
WordNet
A lexical database for English



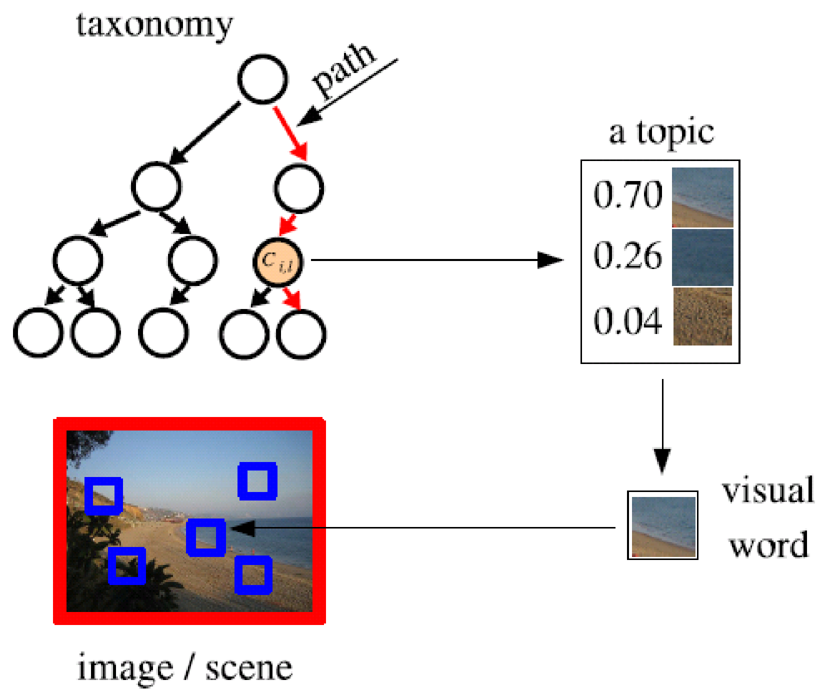

Snow boarding

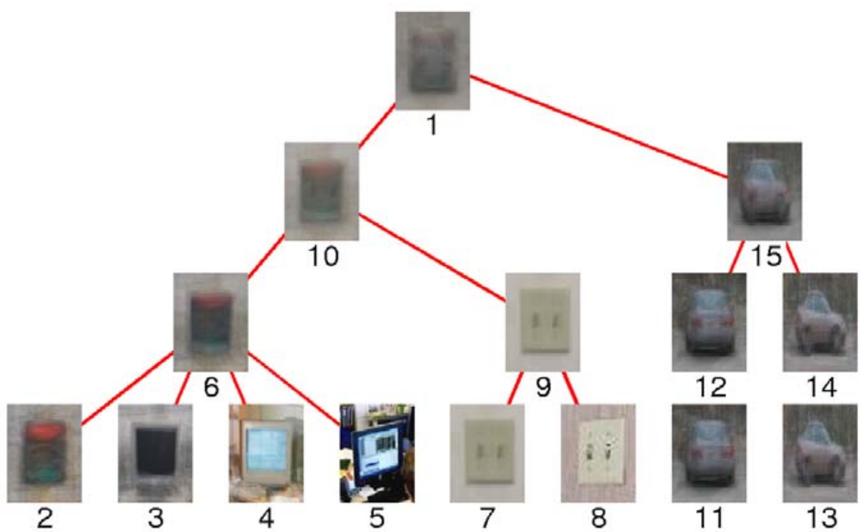
Snow mountain

# (purely) visual hierarchy

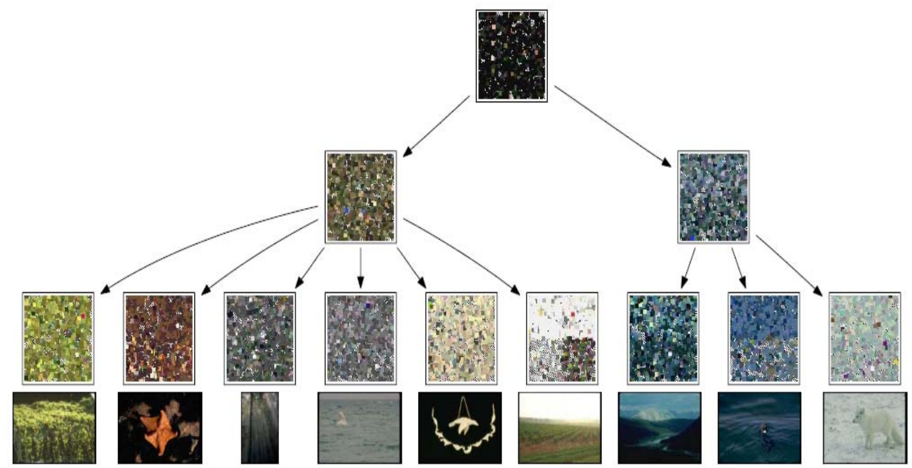



Nested-CRP, Blei et al. NIPS 2004



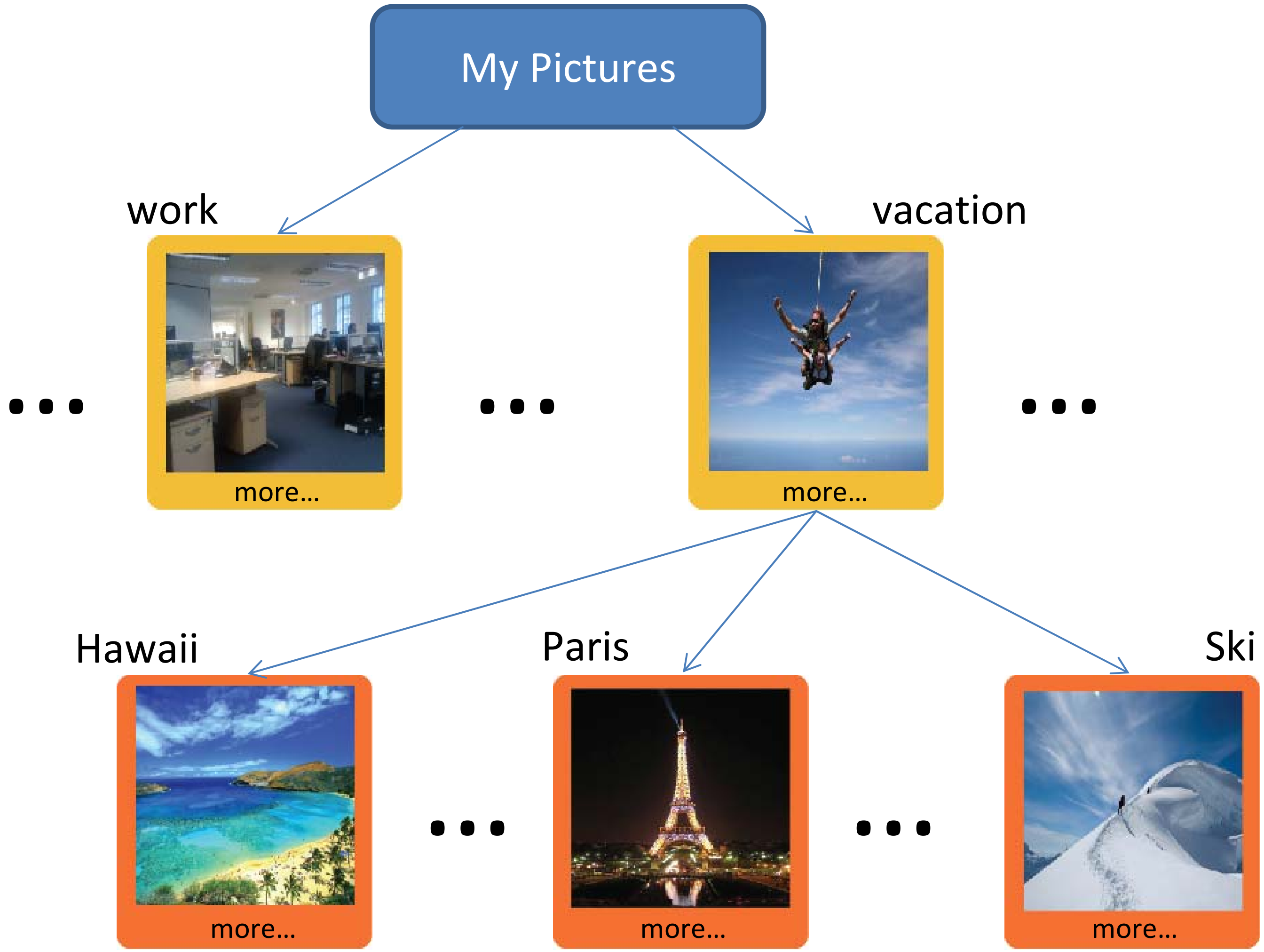




Sivic, Russell, Zisserman, Freeman, Efros, CVPR 2008



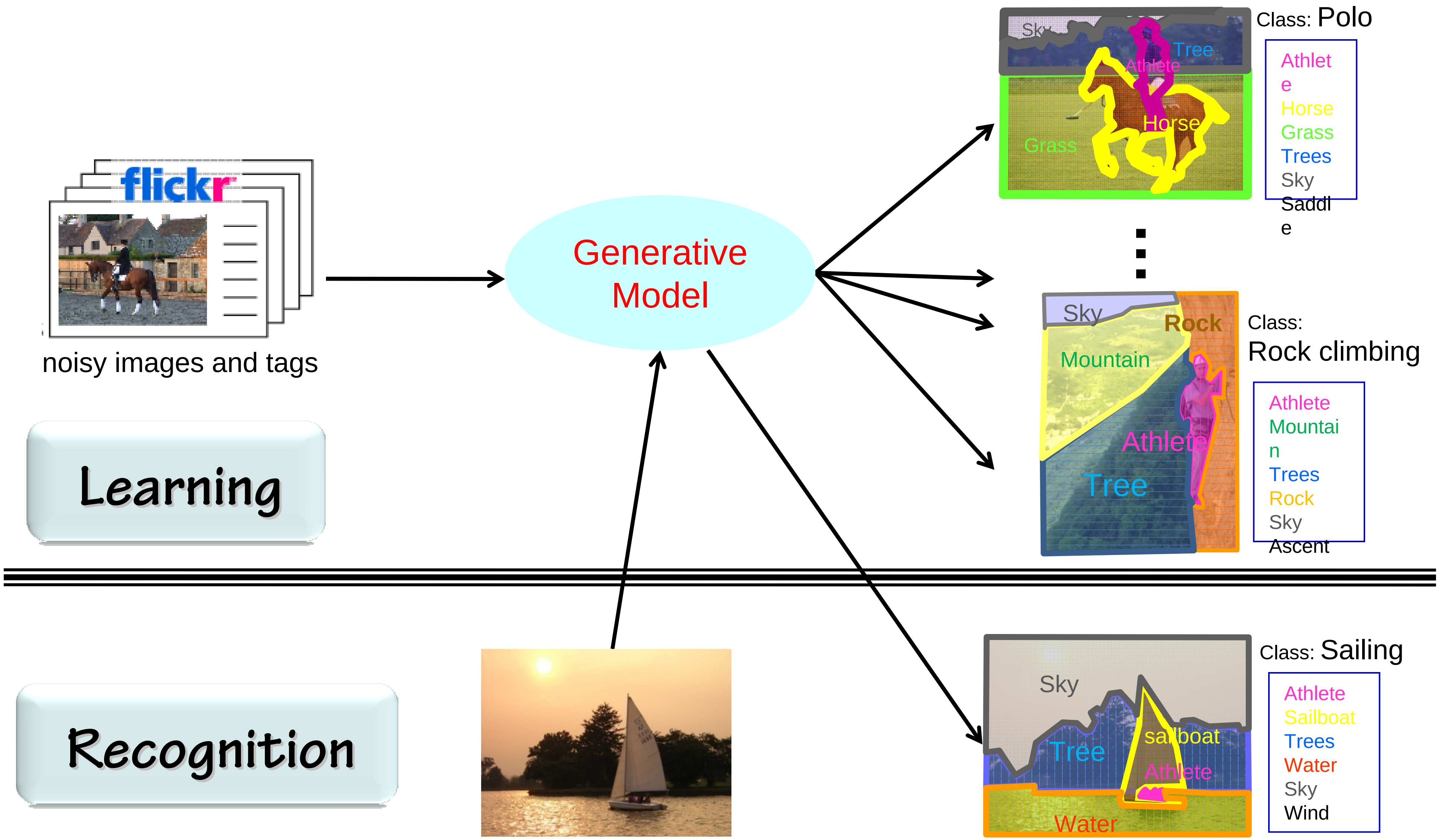
Bart, Porteous, Perona, Welling, CVPR 2008

# A “*semantivisual*” hierarchy of images



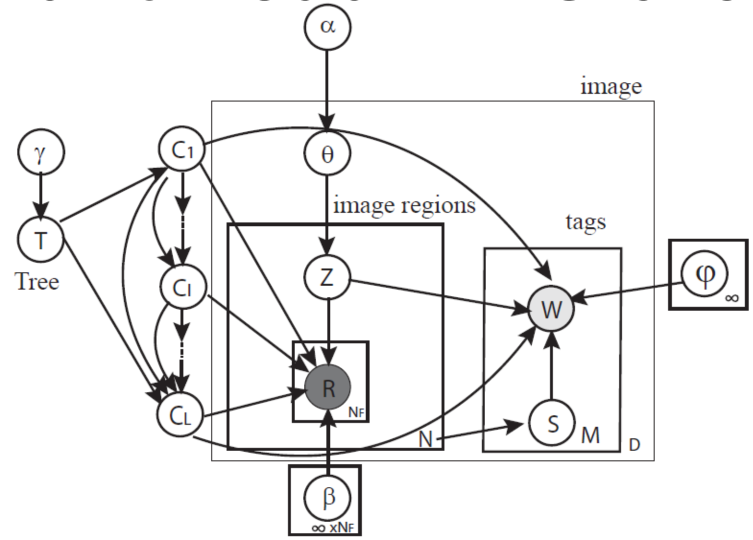

Li, Wang, Lim, Blei & Fei-Fei, *CVPR,* 2010

# “Towards total scene understanding”

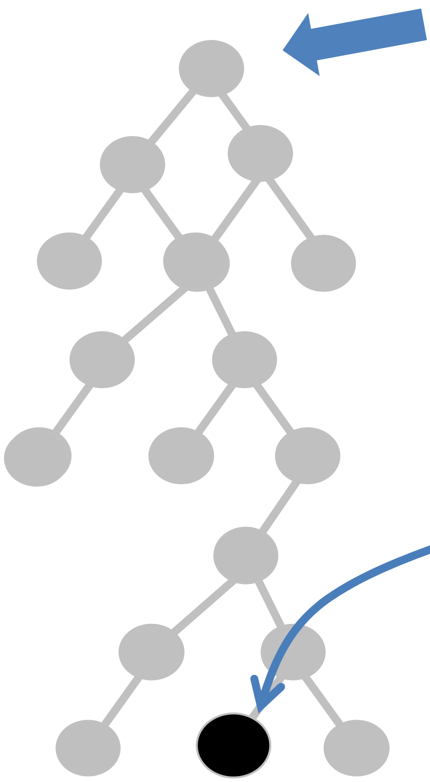


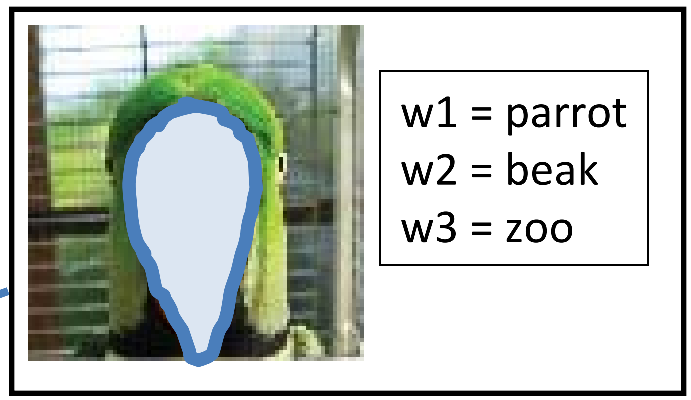

L.-J. Li, R. Socher and L. Fei-Fei, Towards Total Scene Understanding: Classification, annotation and segmentation in an Automatic Framework. *IEEE CVPR,* 2009. Oral.

# A “*semantivisual*” hierarchy of images




R: Region Appearance
W: Words
N: Node in the tree
T: Tree



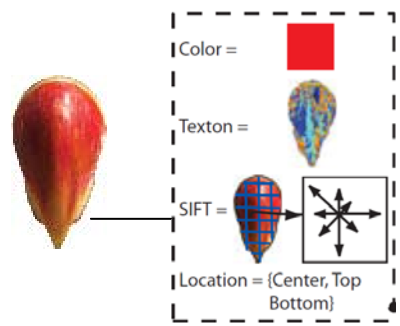



p(N_ik|R _i, W_k) ∝

p(R_i|R_rest,T)

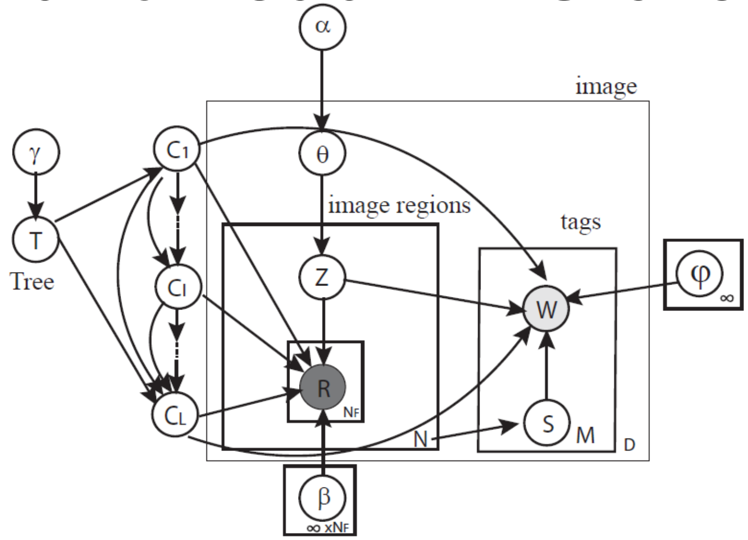
p(W_k|W_rest,T,R_i)

p(N_ik|T,
N_ik_rest)

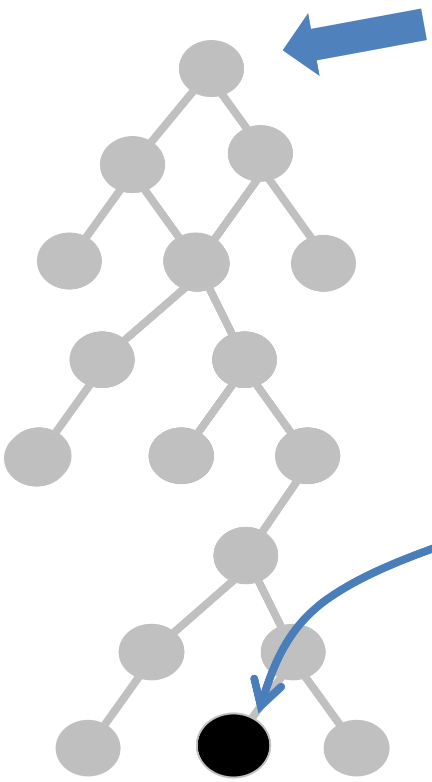



Li, Wang, Lim, Blei & Fei-Fei, *CVPR,* 2010

# A “*semantivisual*” hierarchy of images




R: Region Appearance
W: Words
N: Node in the tree
T: Tree



p(N_ik|R _i, W_k) $\propto$

p(R_i|R_rest,T)

p(W_k|W_rest,T,R_i)

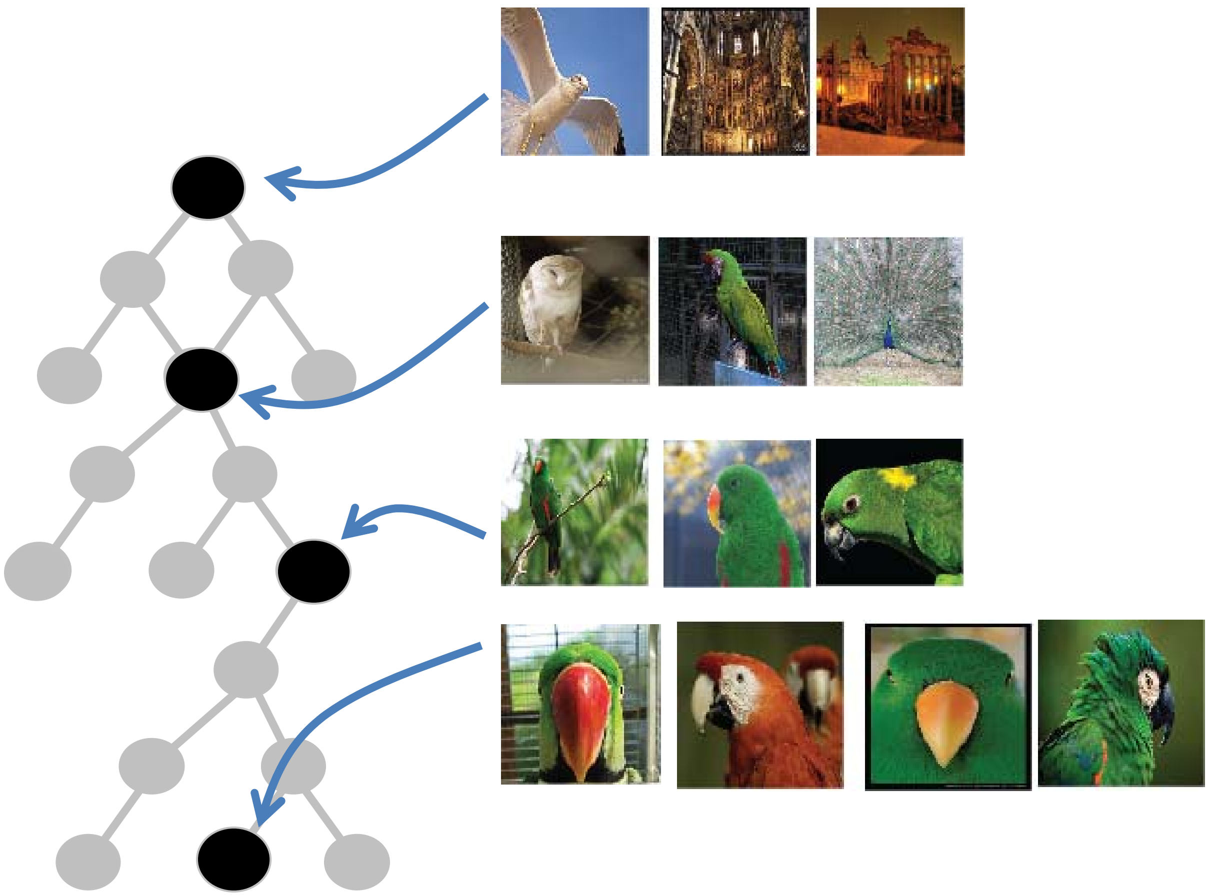
p(N_ik|T, N_ik_rest)

N* = argmax (N_{R})

N_{R}: No. Regions assigned to the node

Li, Wang, Lim, Blei & Fei-Fei, *CVPR,* 2010

# A "*semantivisual*" hierarchy of images



Li, Wang, Lim, Blei & Fei-Fei, *CVPR,* 2010


40 tags, 4000 images

animal, bride, building, cake, child, christmas, church, city, clouds, dessert, dinner, flower, spring, friends, fruit, green, high-school, calcio, italy, europe, london, love, nature, landscape, macro, paris, party, present, sea, sun, sky, seagull, soccer, reflection, sushi, vacation, trip, water, silhouette, and wife.

Li, Wang, Lim, Blei & Fei-Fei, *CVPR,* 2010

flickr from YAHOO!

4000 images


Li, Wang, Lim, Blei & Fei-Fei, *CVPR,* 2010

flickr from YAHOO!

4000 images


Li, Wang, Lim, Blei & Fei-Fei, *CVPR,* 2010

flickr from YAHOO!

4000 images


Li, Wang, Lim, Blei & Fei-Fei, *CVPR,* 2010

flickr from YAHOO!

4000 images


Li, Wang, Lim, Blei & Fei-Fei, *CVPR*, 2010

# Evaluating and using the hierarchy

Evaluate the quality of image concept clustering by path

- ☐ Garden
- ☐ Tree
- ☐ Water
- ☐ Rock
- ☐ Leaf
- ☐ Architecture

| | |
|---|---|
| *semantivisual* hierarchy | 92 % |
| nCRP | 70 % |

amazonmechanical turk
Artificial Artificial Intelligence

Evaluate the quality of hierarchy given a path of the tree

Photo

- ☐ Zoo -> Parrot -> Beak
- ☐ Zoo -> Beak -> Parrot
- ☐ Parrot -> Zoo -> Beak
- ☐ Parrot -> Beak -> Zoo
- ☐ Beak -> Zoo -> Parrot
- ☐ Beak -> Parrot -> Zoo

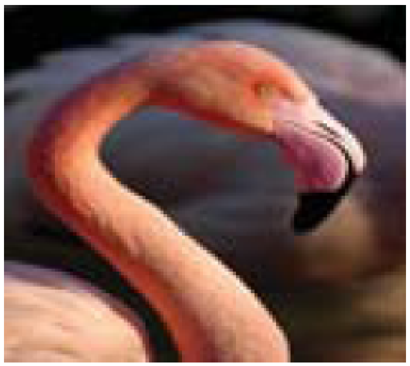
| | |
|---|---|
| *semantivisual* hierarchy | 59 % |
| nCRP | 50 % |
| Flickr | 45 % |

amazonmechanical turk
Artificial Artificial Intelligence

Li, Wang, Lim, Blei & Fei-Fei, *CVPR,* 2010

# Evaluating and using the hierarchy

- Hierarchical annotation



| Our Result | Flickr |
|---|---|
| Photo | Photo |
| ↓ | ↓ |
| Zoo | Bird |
| ↓ | ↓ |
| Flamingo | Animal |
| ↓ | |
| Head | |

| Our Result | Flickr |
|---|---|
| Photo | Photo |
| ↓ | ↓ |
| Football | Football |
| ↓ | ↓ |
| Stadium | London |
| ↓ | |
| Human | |

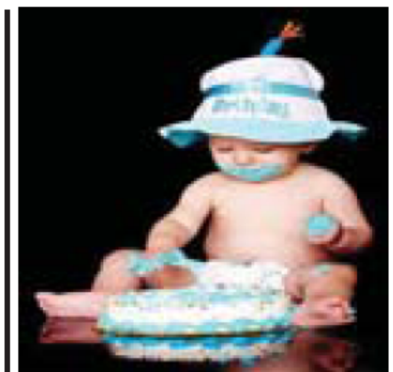
| Our Result | Flickr |
|---|---|
| Photo | Photo |
| ↓ | ↓ |
| Event | Wedding |
| ↓ | ↓ |
| Wedding | Bride |
| ↓ | |
| Dress | |

| method | accuracy |
|---|---|
| Our Model | 46% |
| nCRP | 16% |

Li, Wang, Lim, Blei & Fei-Fei, *CVPR,* 2010

# Evaluating and using the hierarchy

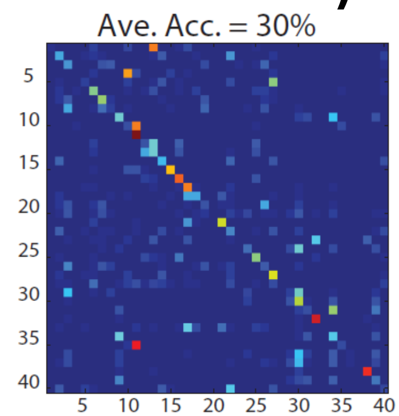
- Hierarchical annotation
- Image labeling (annotation)

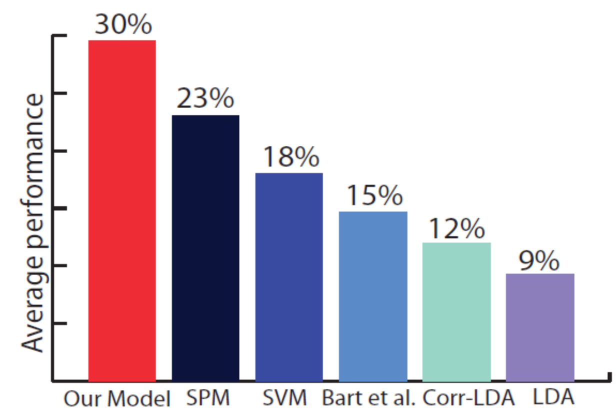
| |  |  |  | |
|---|---|---|---|---|
| Alipr | building photo landscape sky people | card people female fashion cloth | people ocean water landscape snow | 38% |
| Corr-LDA | cake dress garden architecture flower | photo birthday bird architecture portrait | light cloud photo city human | 44% |
| Ours | photo wedding gown bride flower | photo birthday kid cake human | photo cloud sky architecture building | 74% |

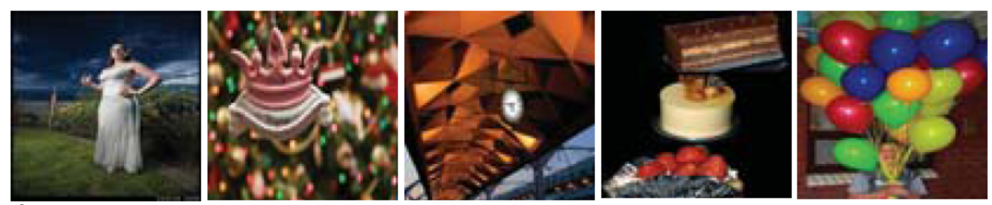
Li, Wang, Lim, Blei & Fei-Fei, *CVPR,* 2010

# Evaluating and using the hierarchy

- Hierarchical annotation
- Image labeling (annotation)
- Image classifica



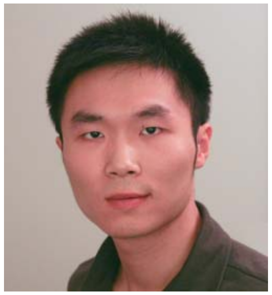



| | | | | | |
|---|---|---|---|---|---|
| Ours | Bride | Christmas | City | Dessert | Party |
| SPM | Building | Flower | Sushi | Child | Friends |
| SVM | London | Party | Sushi | Italy | Christmas |
| Bart et al. | Soccer | City | Present | Sun | Fruit |
| Corr-LDA | Friends | High School | Italia | Dinner | Cake |
| LDA | Italia | High School | Child | Party | Cake |

Li, Wang, Lim, Blei & Fei-Fei, *CVPR,* 2010

# Thank you!



Jia Deng
$4^{th}$ year PhD
Princeton;

"ImageNet"

Hao Su
$1^{st}$ year PhD
Stanford;

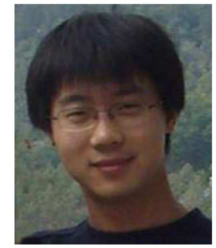
"ImageNet"

Li-Jia Li
$4^{th}$ year PhD
Stanford;

"Total scene understanding";
"Semantivisual hierarchy"

Chris
Baldassano

Juan Carlos
Niebles



Bangpeng
Yao