IMAGENET

crowdsourcing, benchmarking & other cool things

Fei-Fei Li

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

Computer Science Dept.
Psychology Dept.
Stanford University

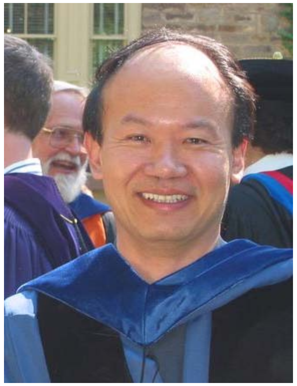
IMAGENET is team work!

WordNet friends

Christiane Fellbaum
Princeton U.

Dan Osherson
Princeton U.

co-PI

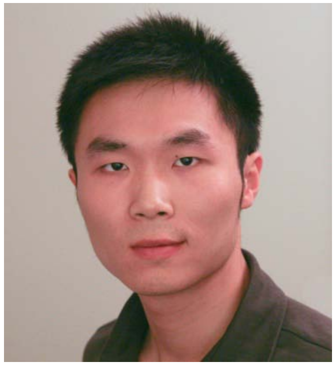


Kai Li
Princeton U.

Research collaborator; ImageNet Challenge boss

Alex Berg
Columbia U.

Graduate students



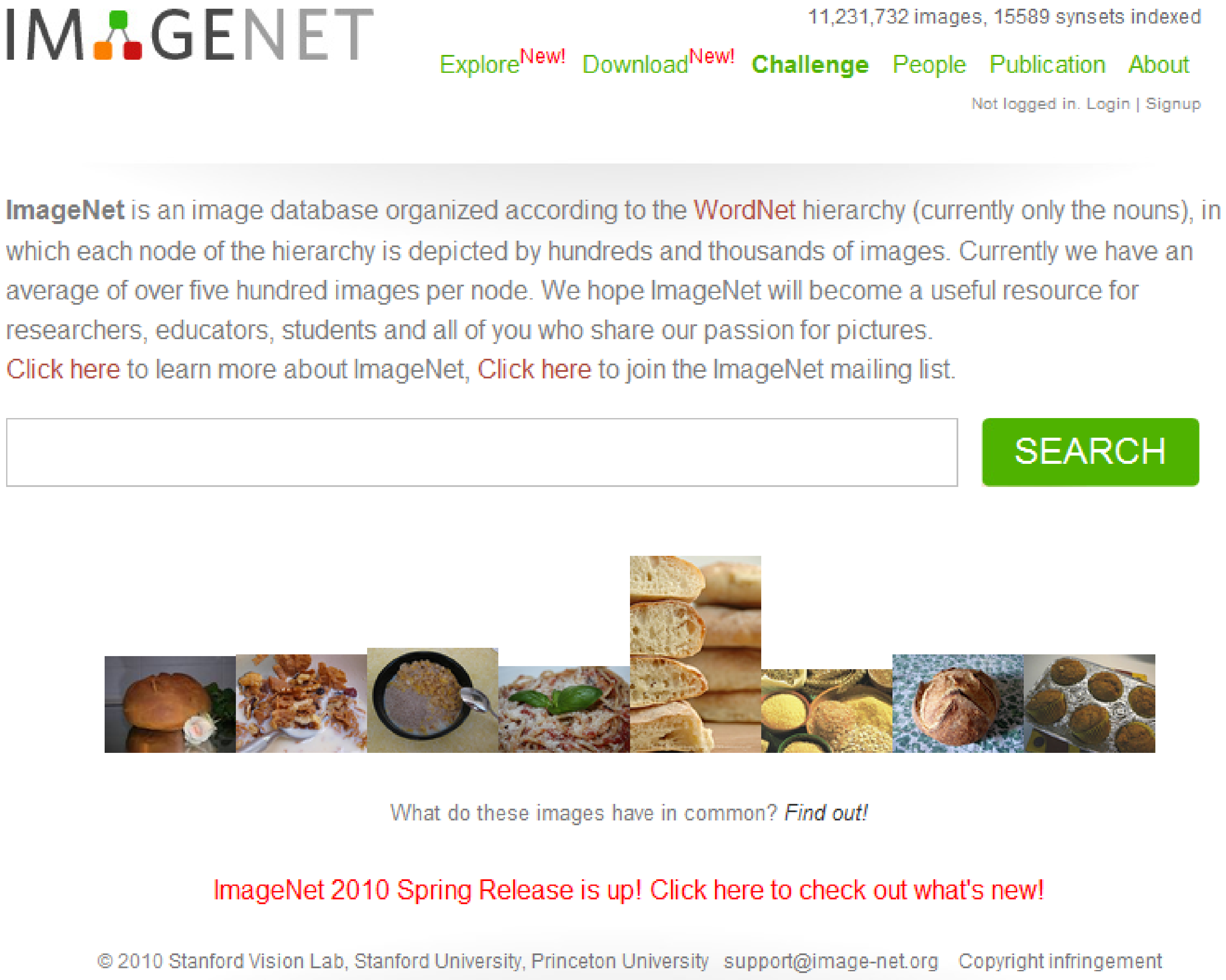
Jia Deng
Princeton/Stanford

Hao Su
Stanford U.

Other contributors

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

http://www.image-net.org



11,231,732 images, 15589 synsets indexed

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 the ImageNet mailing list.

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

ImageNet 2010 Spring Release is up! Click here to check out what's new!

© 2010 Stanford Vision Lab, Stanford University, Princeton University support@image-net.org Copyright infringement

outline

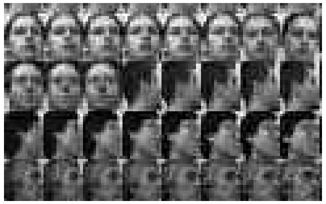
- Goal of ImageNet:
  - A dataset
  - A knowledge ontology
- Construction of ImageNet
  - 2-step process
  - Crowdsourcing: Amazon Mechanical Turk (AMT)
  - Properties of ImageNet
- Benchmarking: what does classifying 10k+ image categories tell us?
  - Computation matters
  - Size matters
  - Density matters
  - Hierarchy matters
- Human vision: Rosch revisited and quantified
  - Quantifying basic-, subordinate- and superordinate-level concepts
- In the horizon: ImageNet Spring 2010 release
  - The upcoming ImageNet Challenge (in partnership with PASCAL VOC)
  - Visualizing ImageNet
  - Etc.

outline

- Goal of ImageNet:
  - A dataset
  - A knowledge ontology
- 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
- Human vision: Rosch revisited and quantified
  - Quantifying basic-, subordinate- and superordinate-level concepts
- In the horizon: ImageNet Spring 2010 release
  - The upcoming ImageNet Challenge (in partnership with PASCAL VOC)
  - Visualizing ImageNet
  - Etc.

Datasets and computer vision

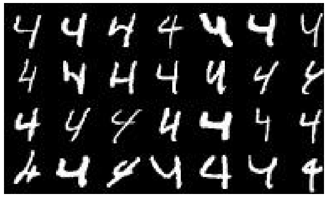
UIUC Cars (2004)
S. Agarwal, A. Awan, D. Roth



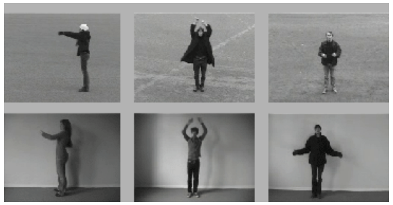
CMU/VASC Faces (1998)
H. Rowley, S. Baluja, T. Kanade

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

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



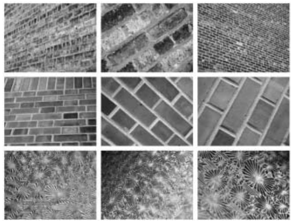
MNIST digits (1998-10)
Y LeCun & C. Cortes



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

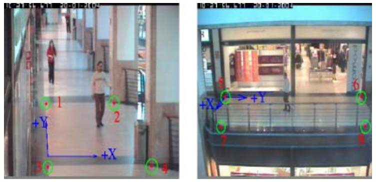
Sign Language (2008)
P. Buehler, M. Everingham, A. Zisserman

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

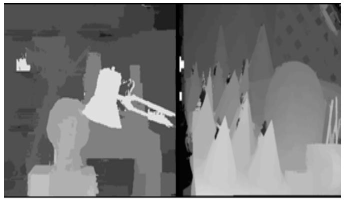


3D Textures (2005)
S. Lazebnik, C. Schmid, J. Ponce

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



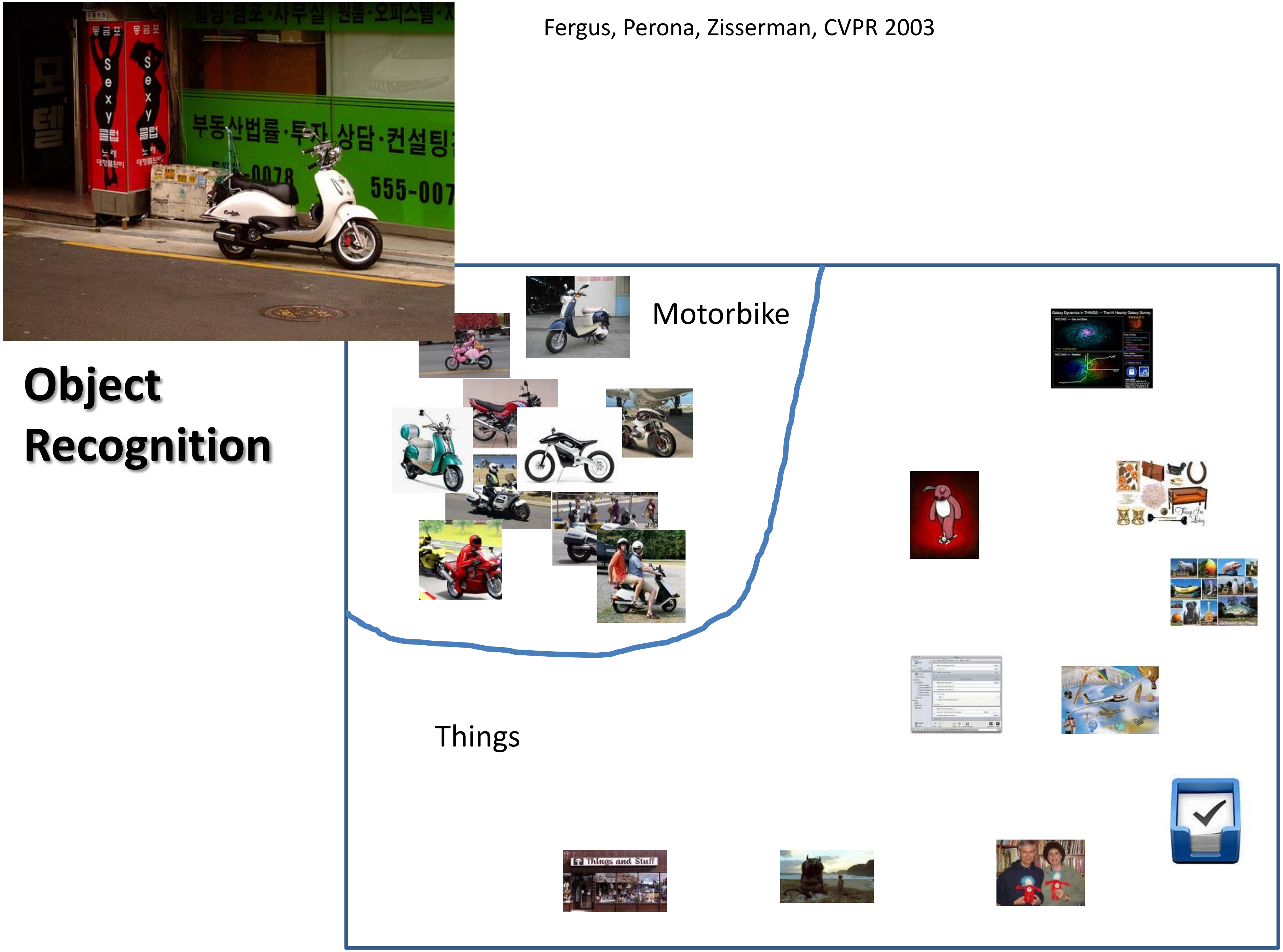
CAVIAR Tracking (2005)
R. Fisher, J. Santos-Victor J. Crowley



Middlebury Stereo (2002)
D. Scharstein R. Szeliski

Fergus, Perona, Zisserman, CVPR 2003

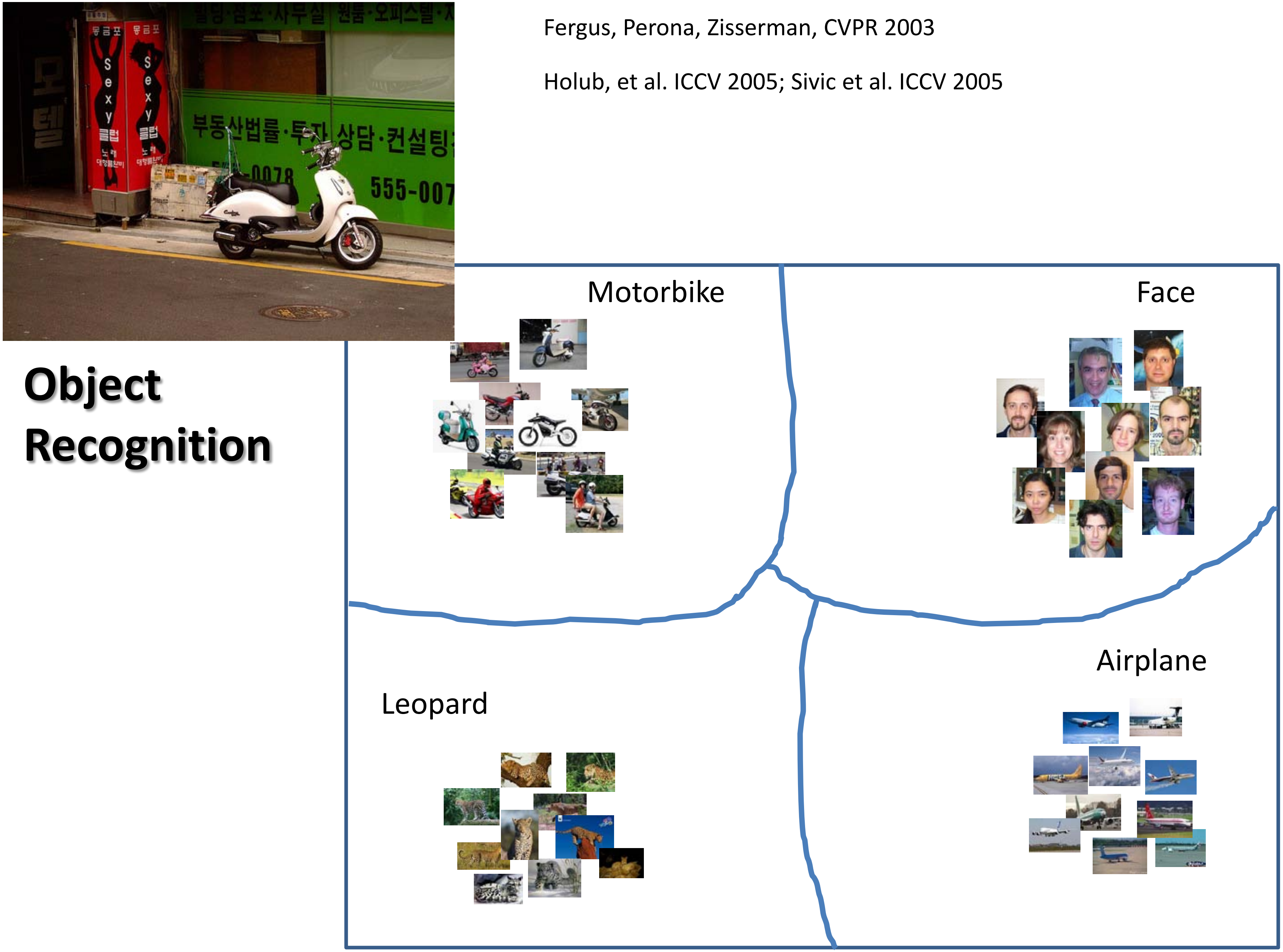
Object Recognition



Object Recognition

Fergus, Perona, Zisserman, CVPR 2003

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



Motorbike

Face

Leopard

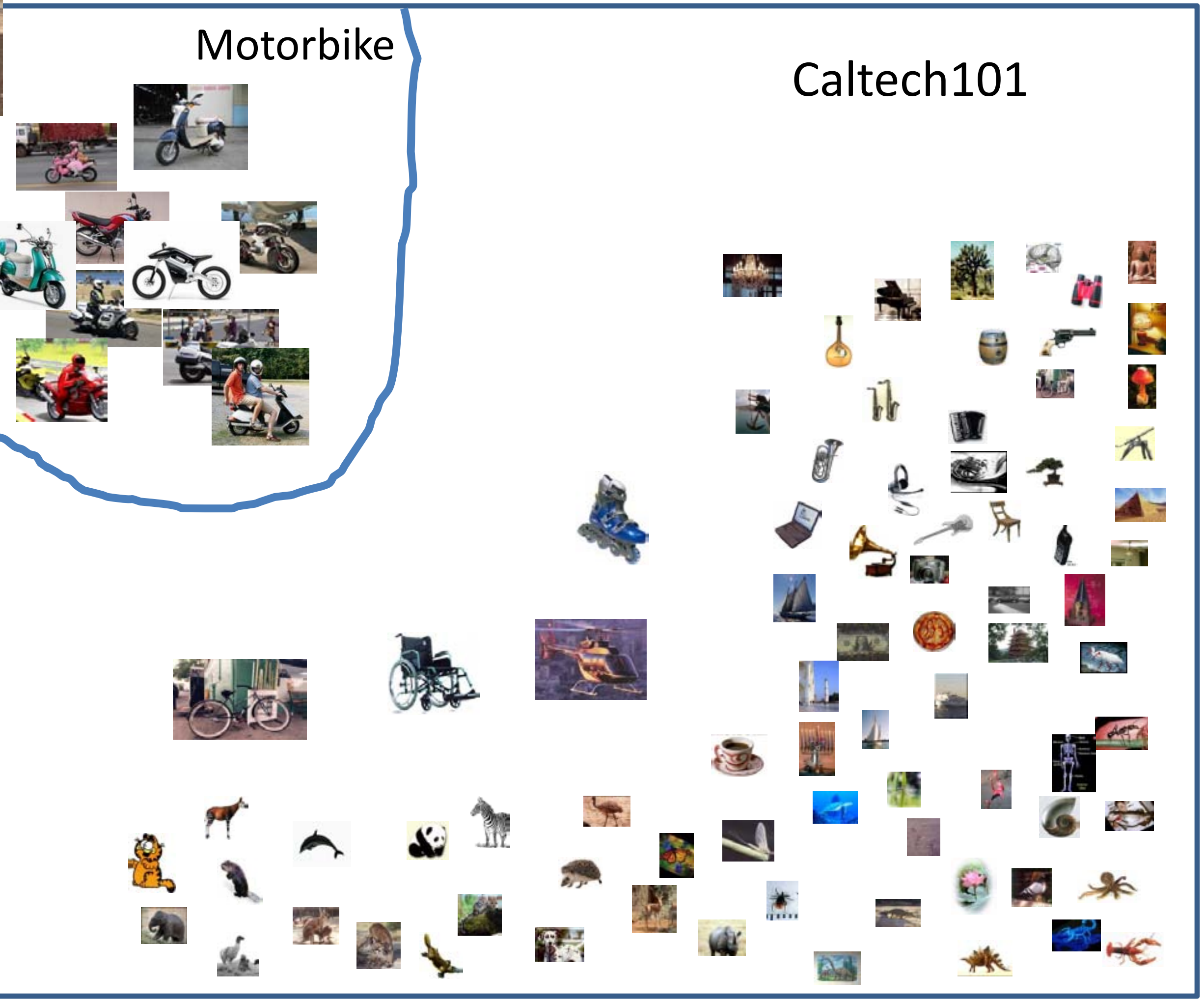
Airplane

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]

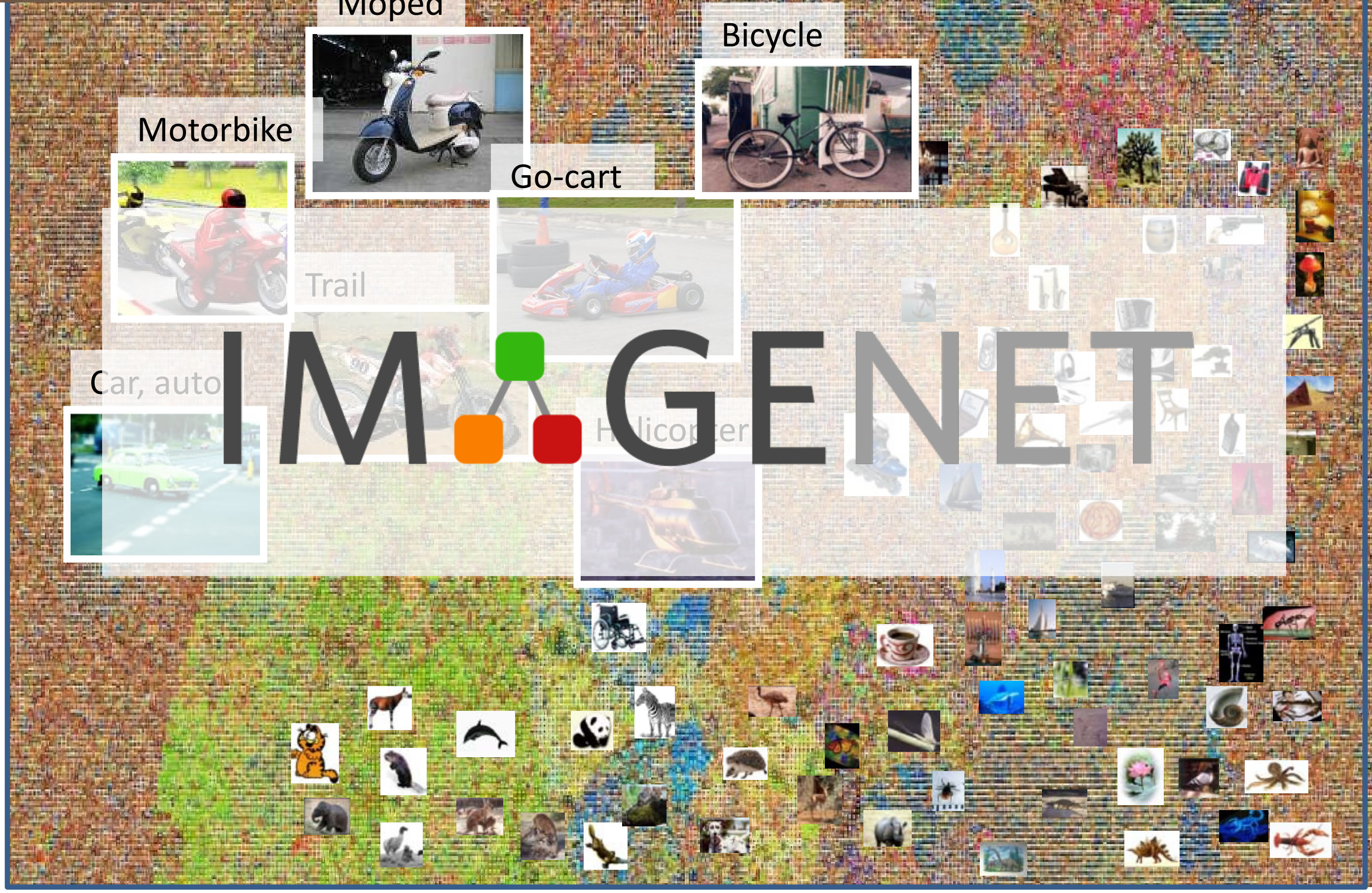
TinyImage

Torralba et al. 2007

Lotus Hill

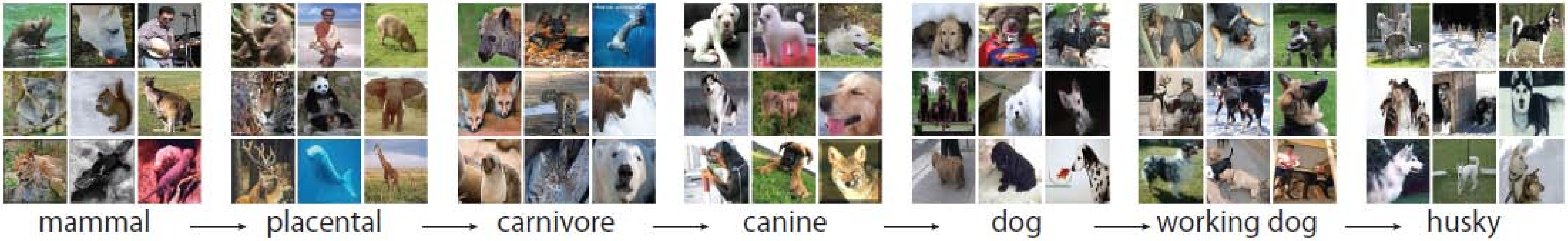
[Yao et al, 2007]

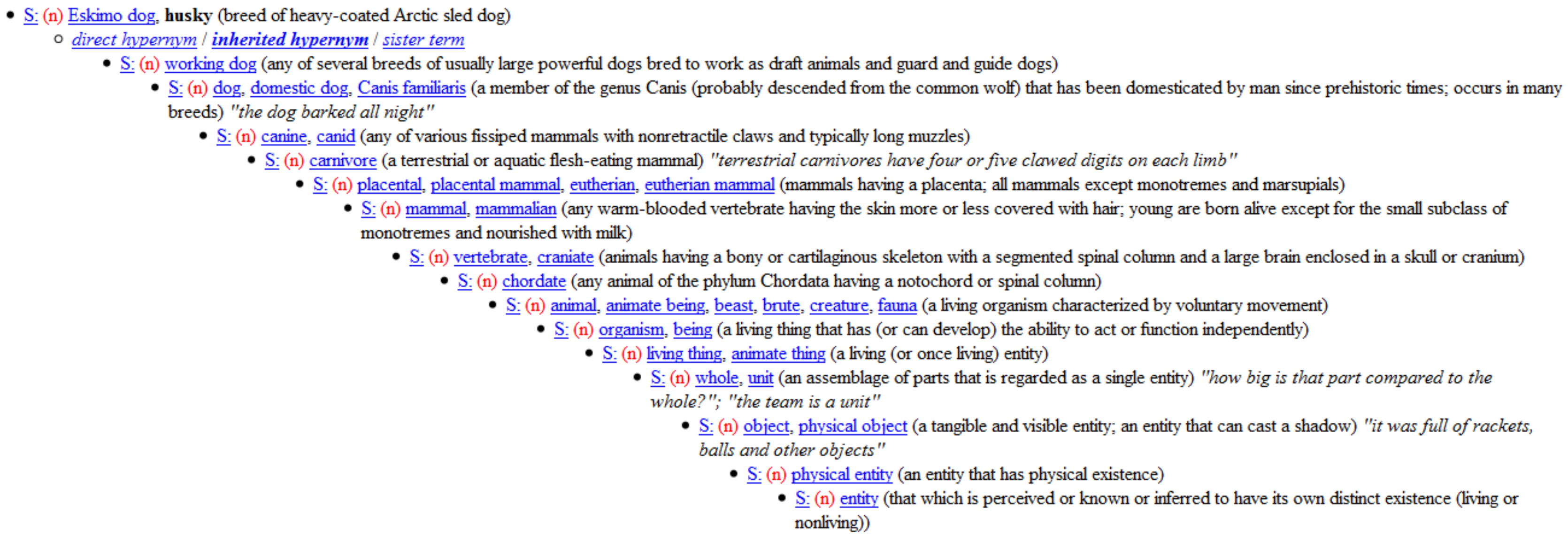
Background image courtesy: Antonio Torralba



IMAGENET is a knowledge ontology

- Taxonomy

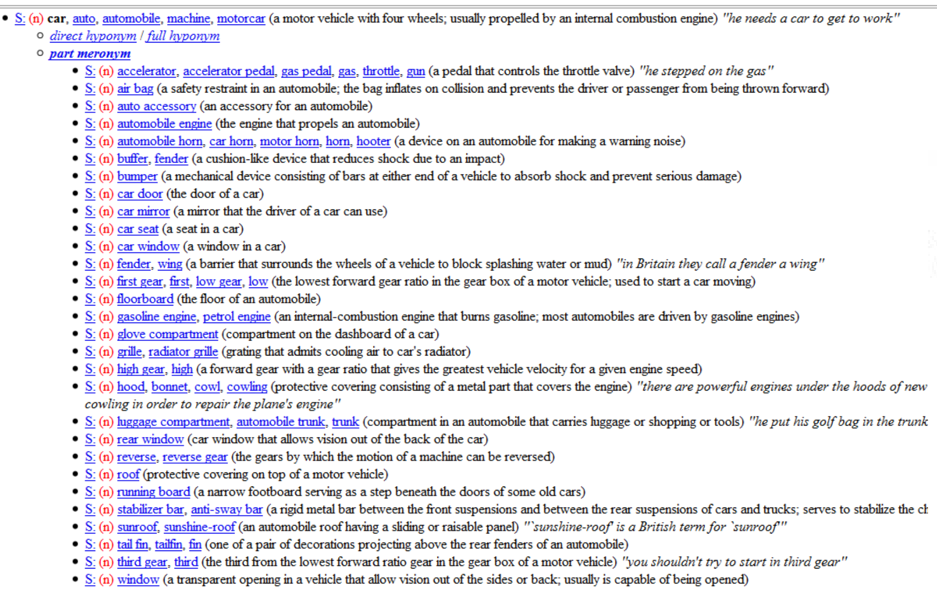

mammal → placental → carnivore → canine → dog → working dog → husky



- 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

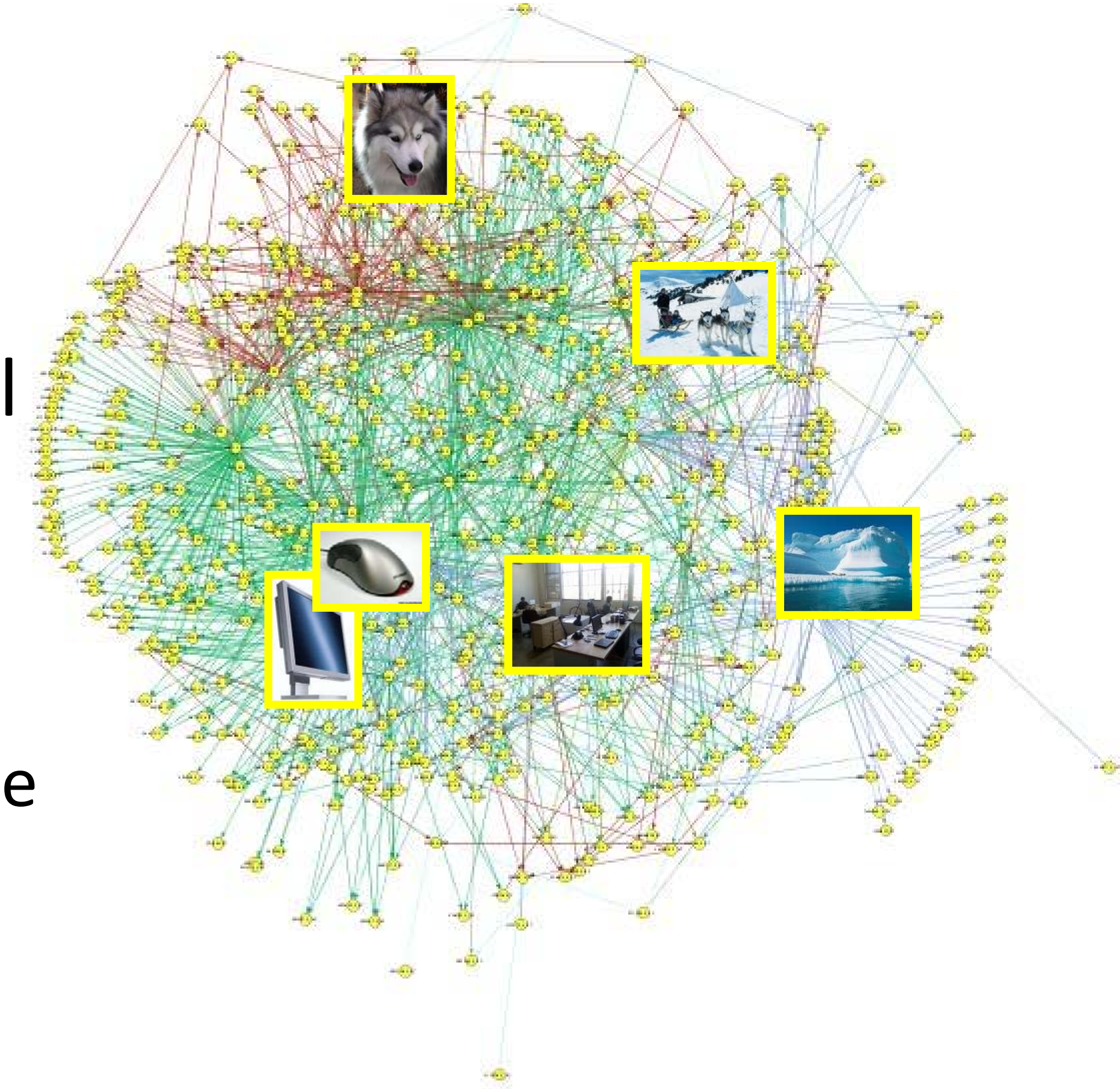
- Taxonomy
- Partonomy



- 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 ch
    - 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

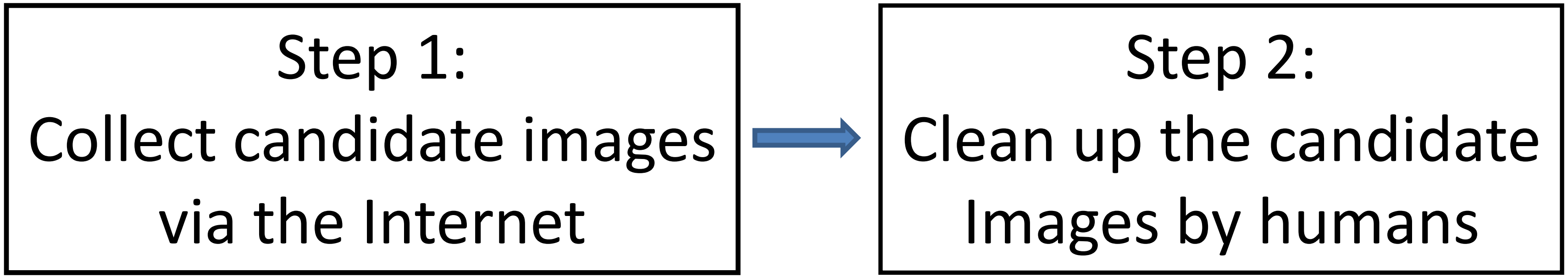
- Taxonomy
- Partonomy
- The “social network” of visual concepts
  - Prior knowledge
  - Context
  - Hidden knowledge and structure among visual concepts



outline

- Goal of ImageNet:
  - A dataset
  - A knowledge ontology
- 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
- Human vision: Rosch revisited and quantified
  - Quantifying basic-, subordinate- and superordinate-level concepts
- In the horizon: ImageNet Spring 2010 release
  - The upcoming ImageNet Challenge (in partnership with PASCAL VOC)
  - Visualizing ImageNet
  - Etc.

Constructing IMAGENET



Step 1:
Collect candidate images
via the Internet

Step 2:
Clean up the candidate
Images by humans

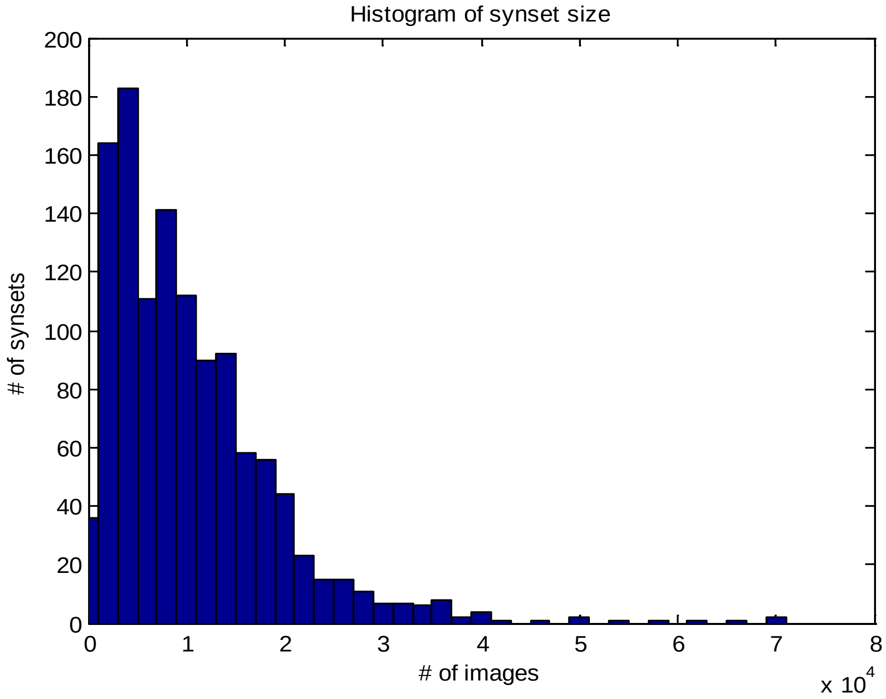
Step 1: Collect Candidate Images from the Internet

- 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 YAHOO! Live Search picsearch™ flickr™ Google™ Image Search
- Parallel downloading

Step 1: Collect Candidate Images from the Internet

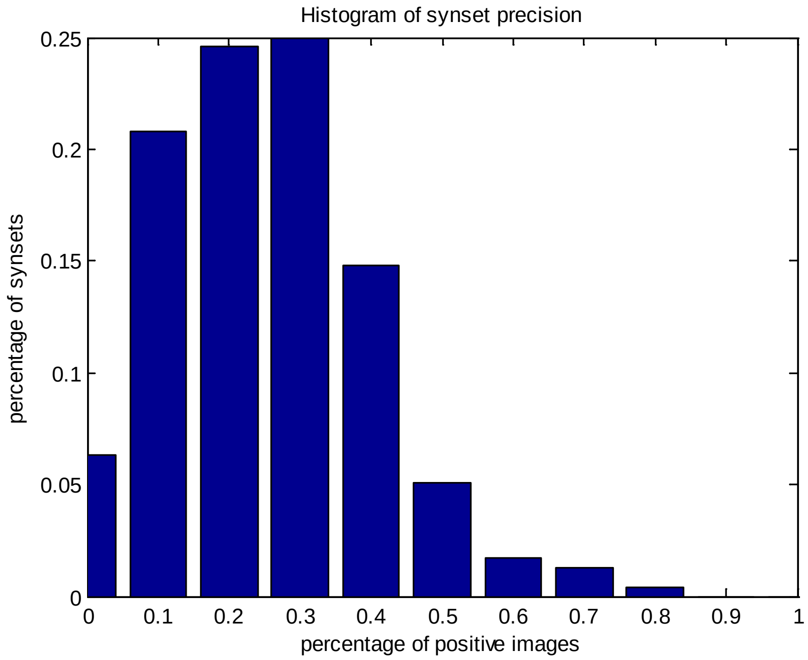
- "Mammal" subtree (1180 synsets)
  - Average # of images per synset: 10.5K



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

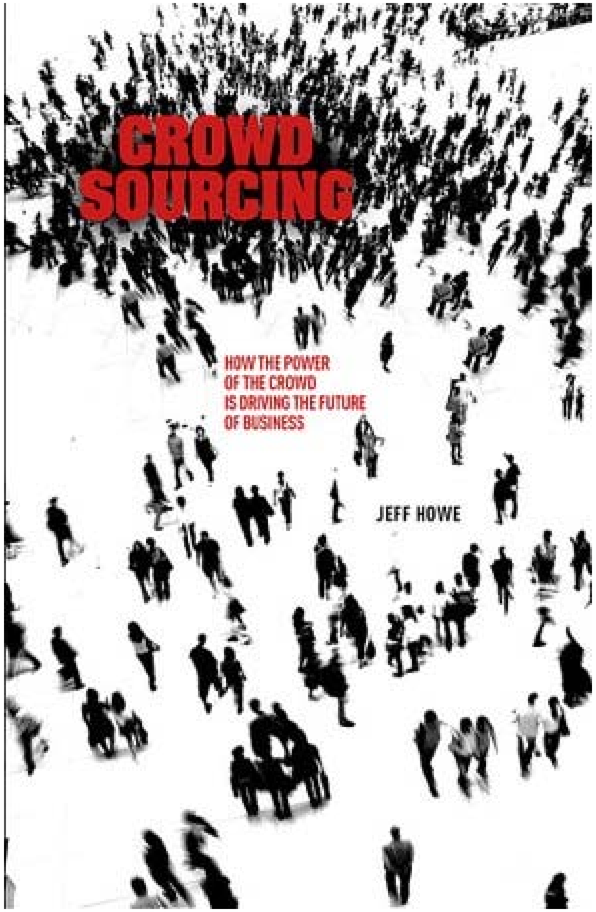
- # of synsets: 40,000 (subject to: imageability analysis)
- # of candidate images to label per synset: 10,000
- # of people needed to verify: 2-5
- Speed of human labeling: 2 images/sec (one fixation: ~200msec)

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

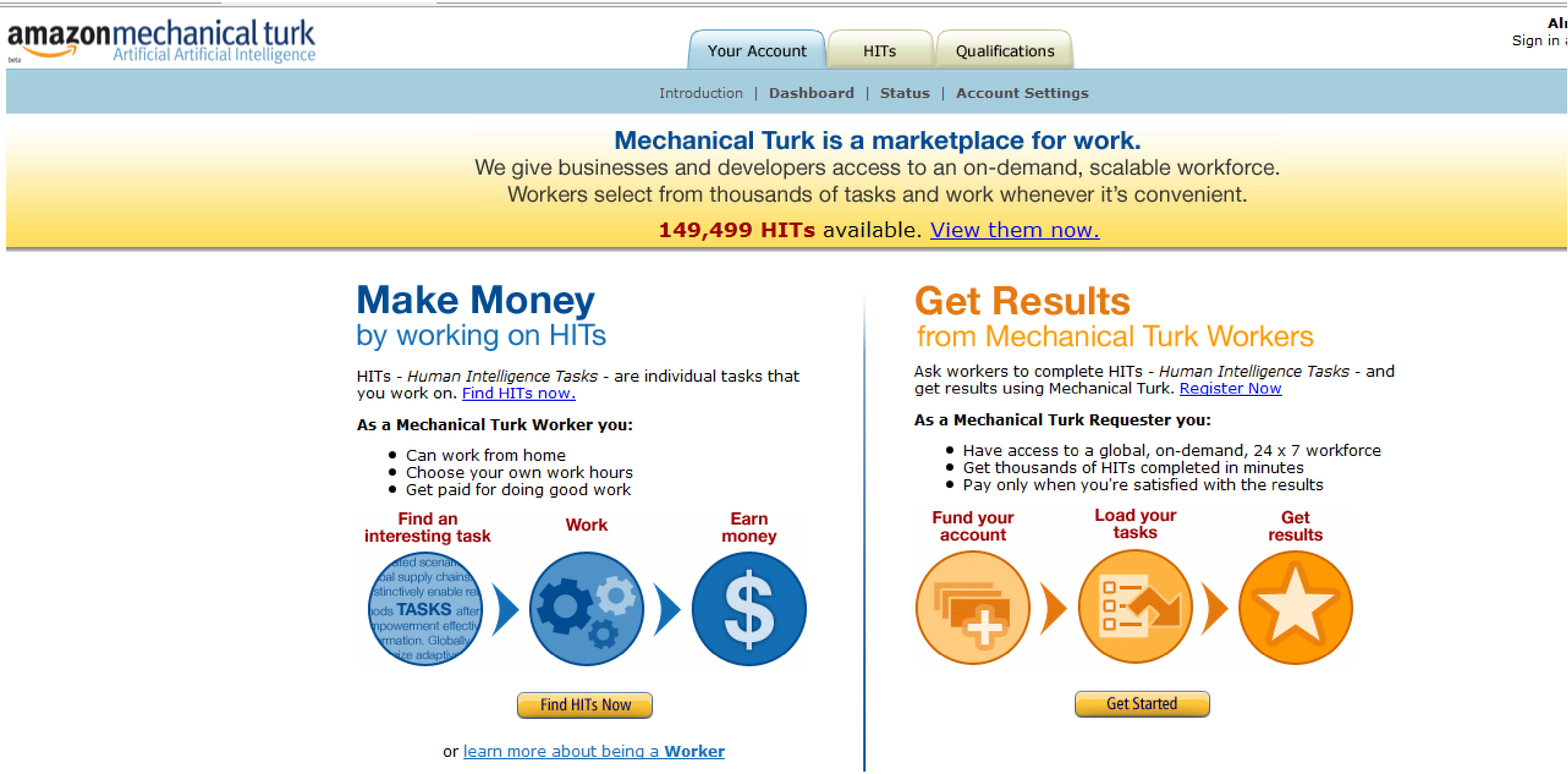
Moral of the story:

no graduate students would want to do this project!

In summer 2008, we discovered crowdsourcing


CROWD SOURCING
HOW THE POWER OF THE CROWD IS DRIVING THE FUTURE OF BUSINESS
JEFF HOWE

amazon mechanical turk
beta
Artificial Artificial Intelligence



amazonmechanical turk
Artificial Artificial Intelligence
beta
Your Account
HITs
Qualifications
Al
Sign in a
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

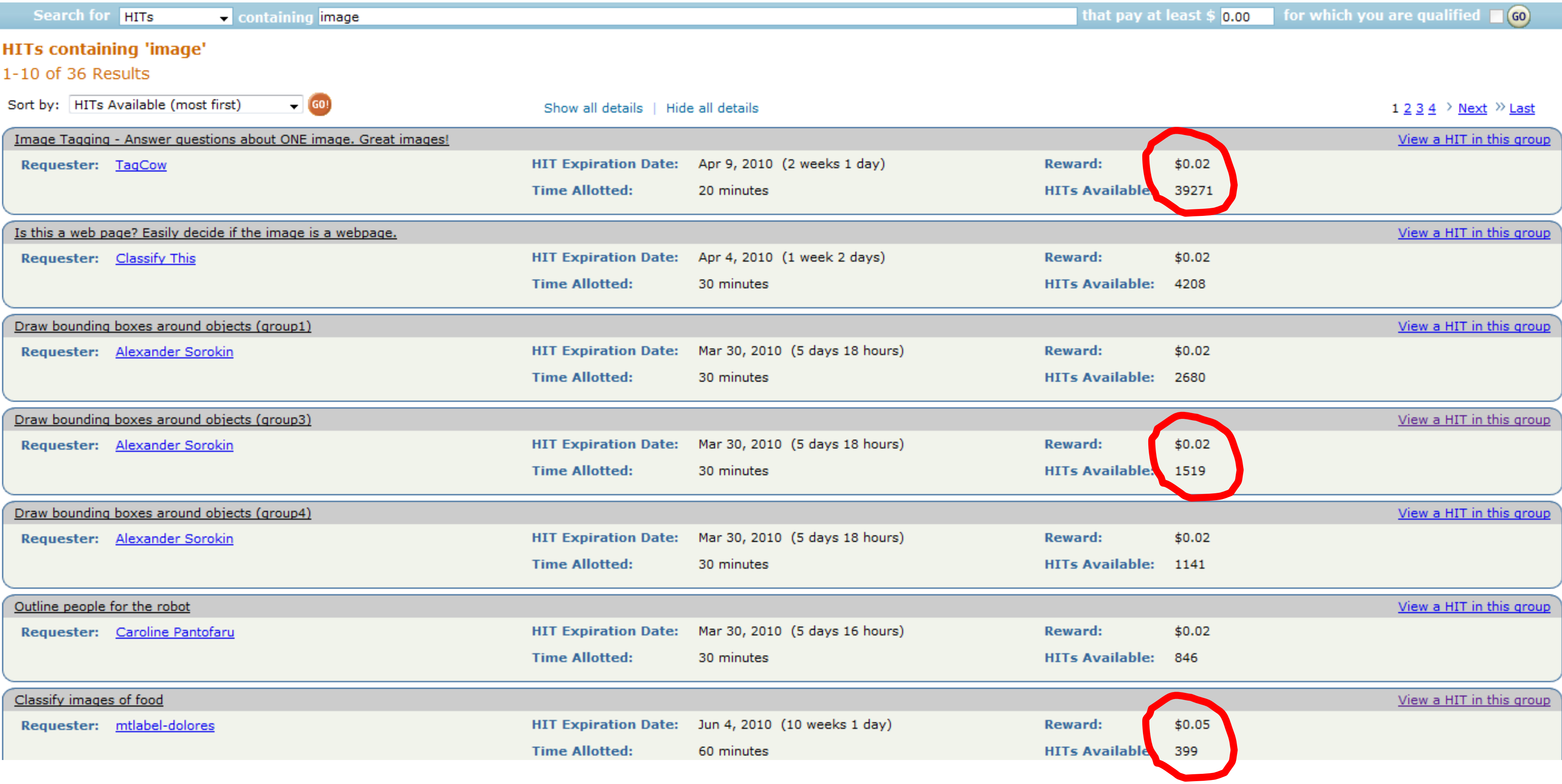
or learn more about being a **Worker**

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



Search for HITs containing image that pay at least $ 0.00 for which you are qualified □ GO

HITs containing 'image'

1-10 of 36 Results

Sort by: HITs Available (most first) GO!

Show all details | Hide all details

1 2 3 4 › Next » Last

Image Tagging - Answer questions about ONE image. Great images! View a HIT in this group

Requester: TagCow
HIT Expiration Date: Apr 9, 2010 (2 weeks 1 day)
Reward: $0.02
Time Allotted: 20 minutes
HITs Available: 39271

Is this a web page? Easily decide if the image is a webpage. View a HIT in this group

Requester: Classify This
HIT Expiration Date: Apr 4, 2010 (1 week 2 days)
Reward: $0.02
Time Allotted: 30 minutes
HITs Available: 4208

Draw bounding boxes around objects (group1) View a HIT in this group

Requester: Alexander Sorokin
HIT Expiration Date: Mar 30, 2010 (5 days 18 hours)
Reward: $0.02
Time Allotted: 30 minutes
HITs Available: 2680

Draw bounding boxes around objects (group3) View a HIT in this group

Requester: Alexander Sorokin
HIT Expiration Date: Mar 30, 2010 (5 days 18 hours)
Reward: $0.02
Time Allotted: 30 minutes
HITs Available: 1519

Draw bounding boxes around objects (group4) View a HIT in this group

Requester: Alexander Sorokin
HIT Expiration Date: Mar 30, 2010 (5 days 18 hours)
Reward: $0.02
Time Allotted: 30 minutes
HITs Available: 1141

Outline people for the robot View a HIT in this group

Requester: Caroline Pantofaru
HIT Expiration Date: Mar 30, 2010 (5 days 16 hours)
Reward: $0.02
Time Allotted: 30 minutes
HITs Available: 846

Classify images of food View a HIT in this group

Requester: mtlabel-dolores
HIT Expiration Date: Jun 4, 2010 (10 weeks 1 day)
Reward: $0.05
Time Allotted: 60 minutes
HITs Available: 399

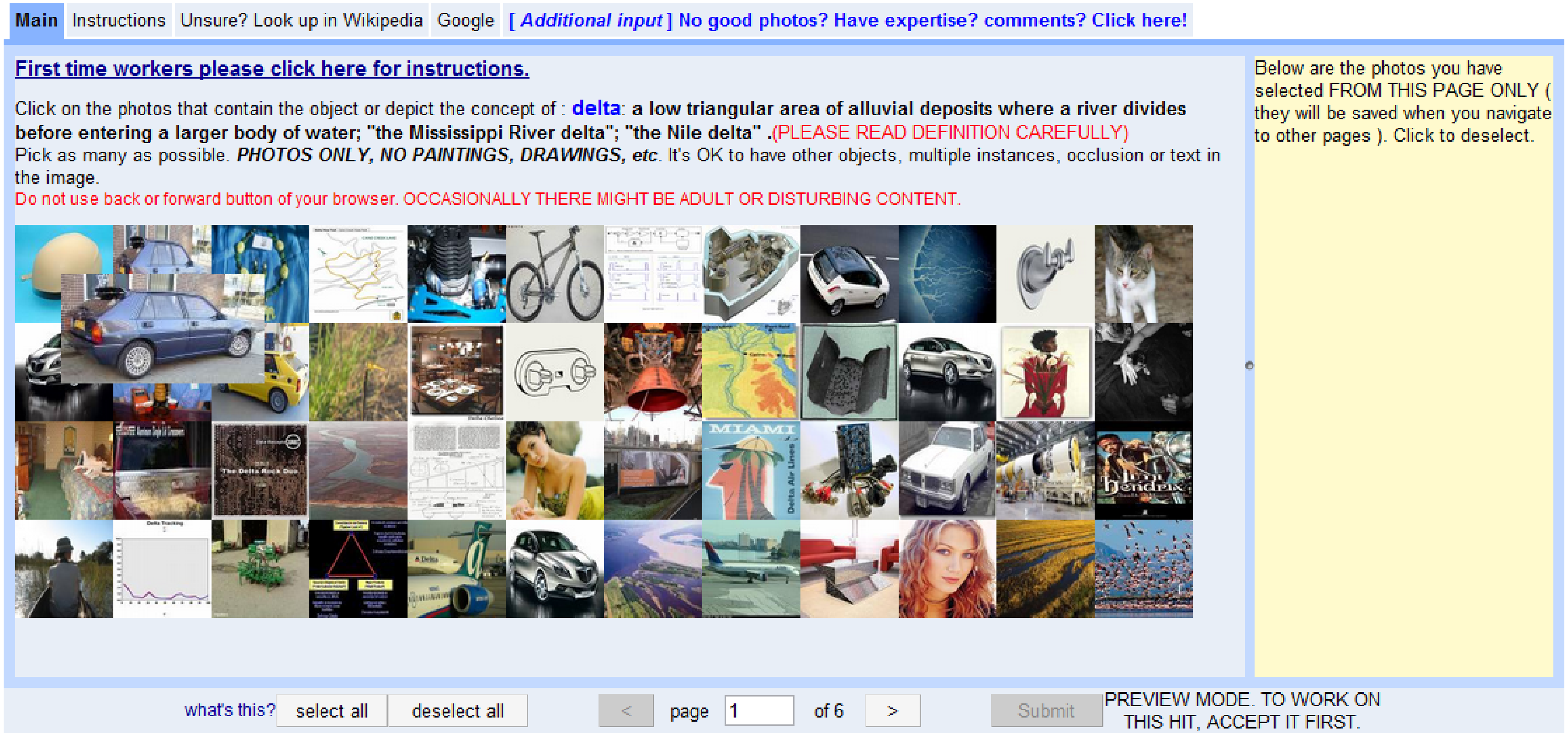
Step 2: verifying the images by humans

- # 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)

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

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!

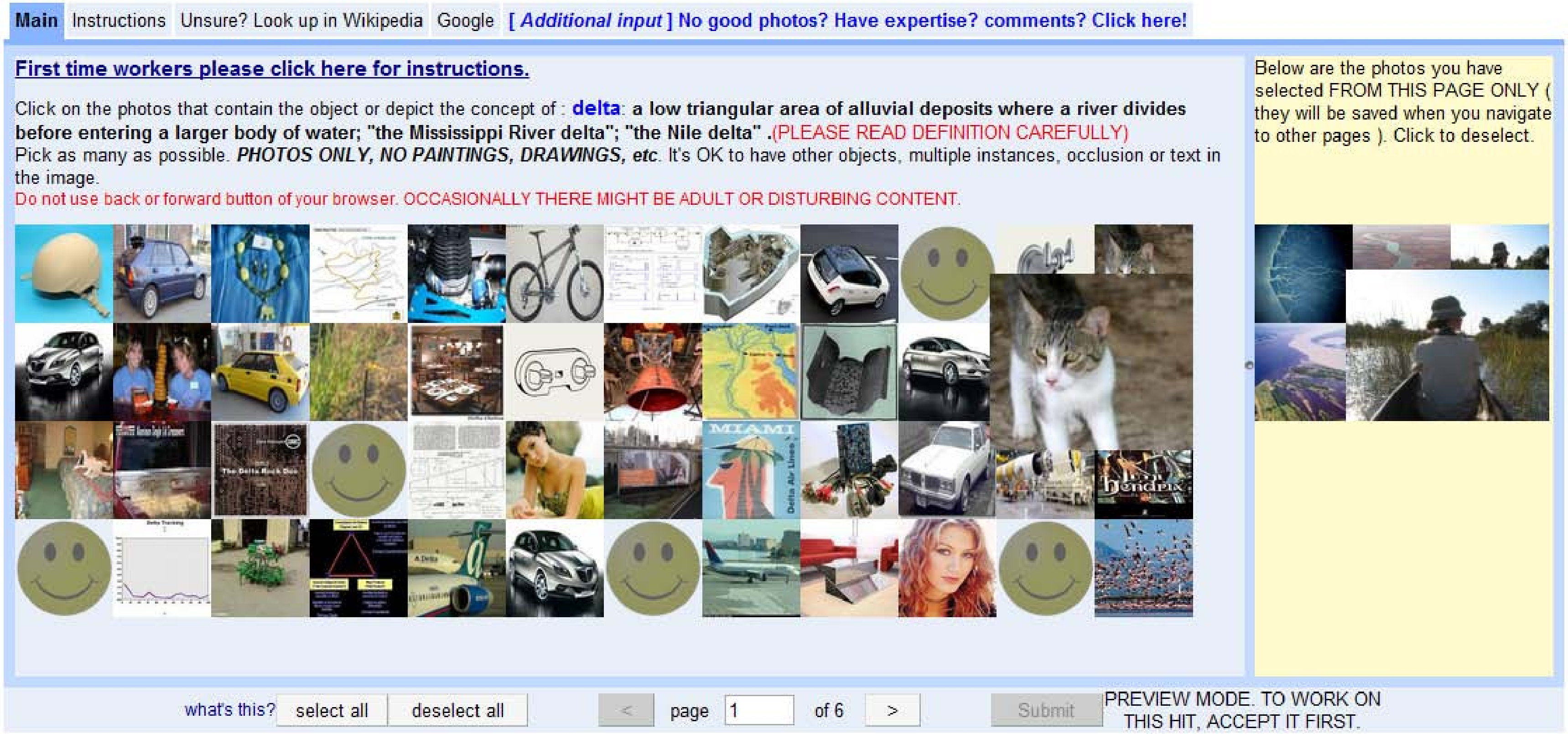
First time workers please click here for instructions.

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.

PREVIEW MODE. TO WORK ON THIS HIT, ACCEPT IT FIRST.

IMAGENET Basic User Interface



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.

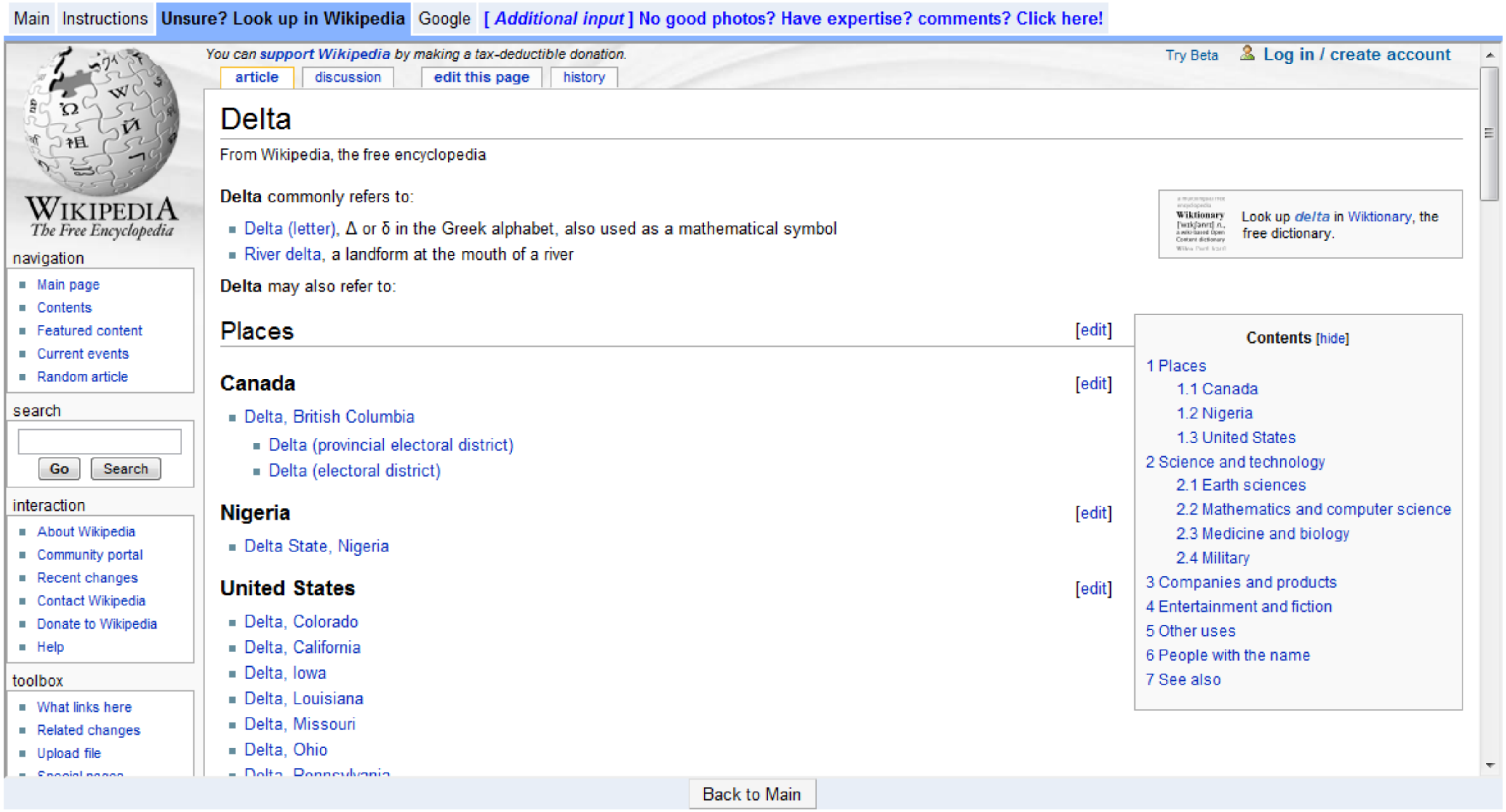
Click on the photos that contain the object or depict the concept of : **delta**: **a low triangular area of alluvial deposits where a river divides before entering a larger body of water; "the Mississippi River delta"; "the Nile delta" .**(PLEASE READ DEFINITION CAREFULLY)
Pick as many as possible. ***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.

PREVIEW MODE. TO WORK ON THIS HIT, ACCEPT IT FIRST.

Enhancement 1

- Provide wiki and google links



You can ***support Wikipedia*** *by making a tax-deductible donation.*

Delta

From Wikipedia, the free encyclopedia

Look up *delta* in Wiktionary, the free dictionary.

Delta commonly refers to:

- Delta (letter), Δ or δ in the Greek alphabet, also used as a mathematical symbol
- River delta, a landform at the mouth of a river

Delta may also refer to:

Places [edit]

Canada [edit]

- Delta, British Columbia
  - Delta (provincial electoral district)
  - Delta (electoral district)

Nigeria [edit]

- Delta State, Nigeria

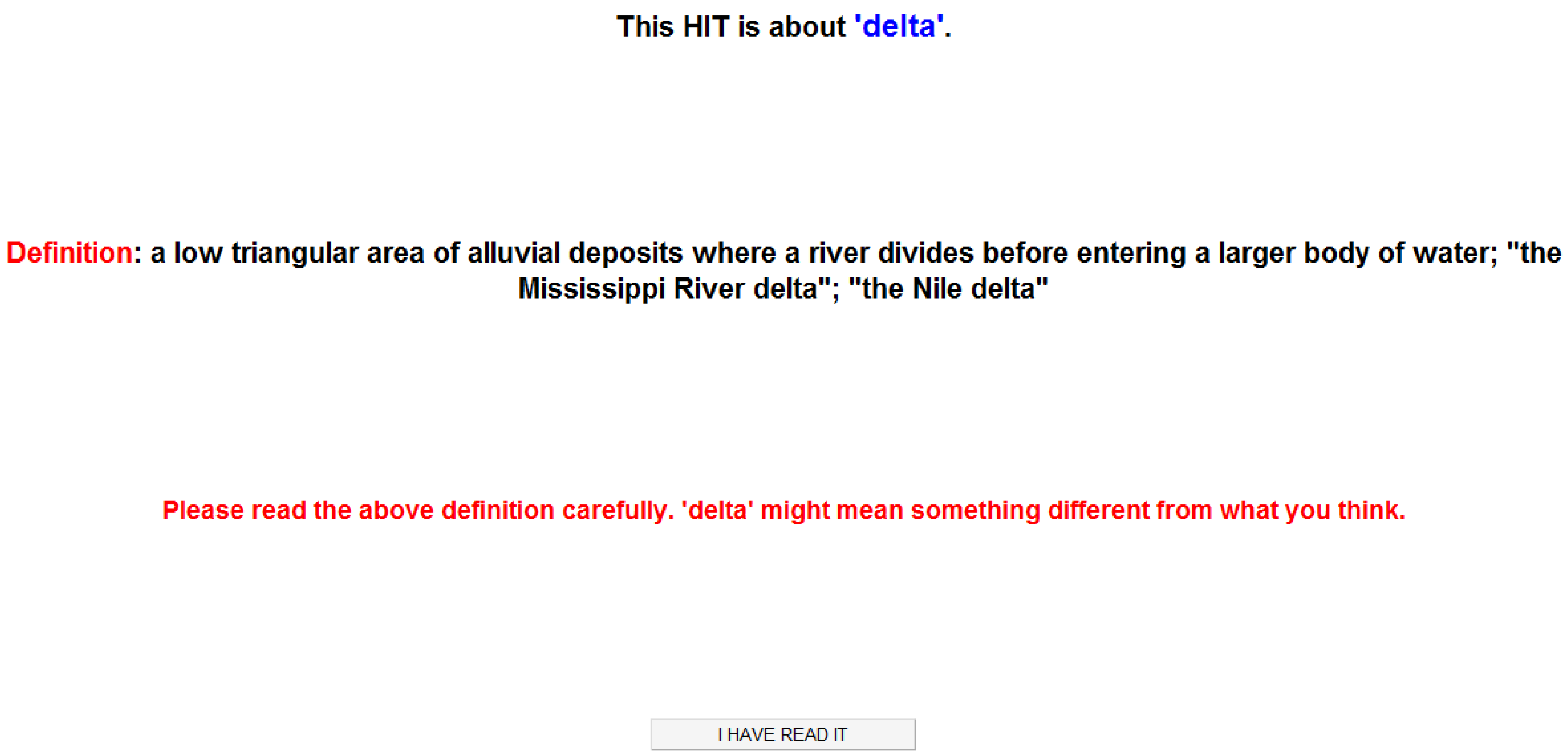
United States [edit]

- Delta, Colorado
- Delta, California
- Delta, Iowa
- Delta, Louisiana
- Delta, Missouri
- Delta, Ohio
- Delta, Pennsylvania

Enhancement 2

- Make sure workers read the definition.
  - Words are ambiguous. E.g.
    - Box: *any one of several designated areas on a ball field where the batter or catcher or coaches are positioned*
    - Keyboard: *holder consisting of an arrangement of hooks on which keys or locks can be hung*
  - These synsets are hard to get right
  - Some workers do not read or understand the definition.

Definition quiz

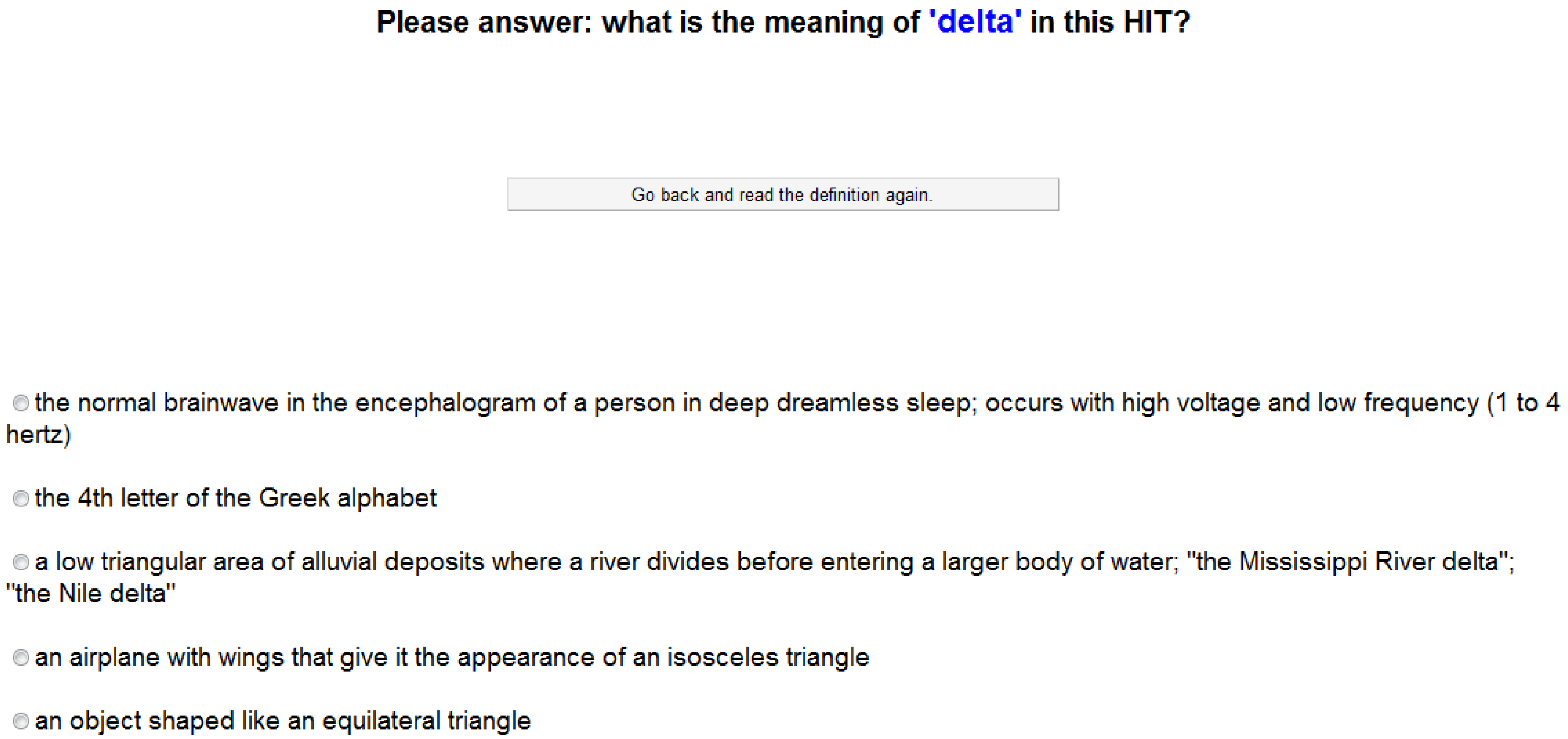


This HIT is about 'delta'.

Definition: 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 the above definition carefully. 'delta' might mean something different from what you think.

Definition quiz



Please answer: what is the meaning of 'delta' in this HIT?

○ the normal brainwave in the encephalogram of a person in deep dreamless sleep; occurs with high voltage and low frequency (1 to 4 hertz)

○ the 4th letter of the Greek alphabet

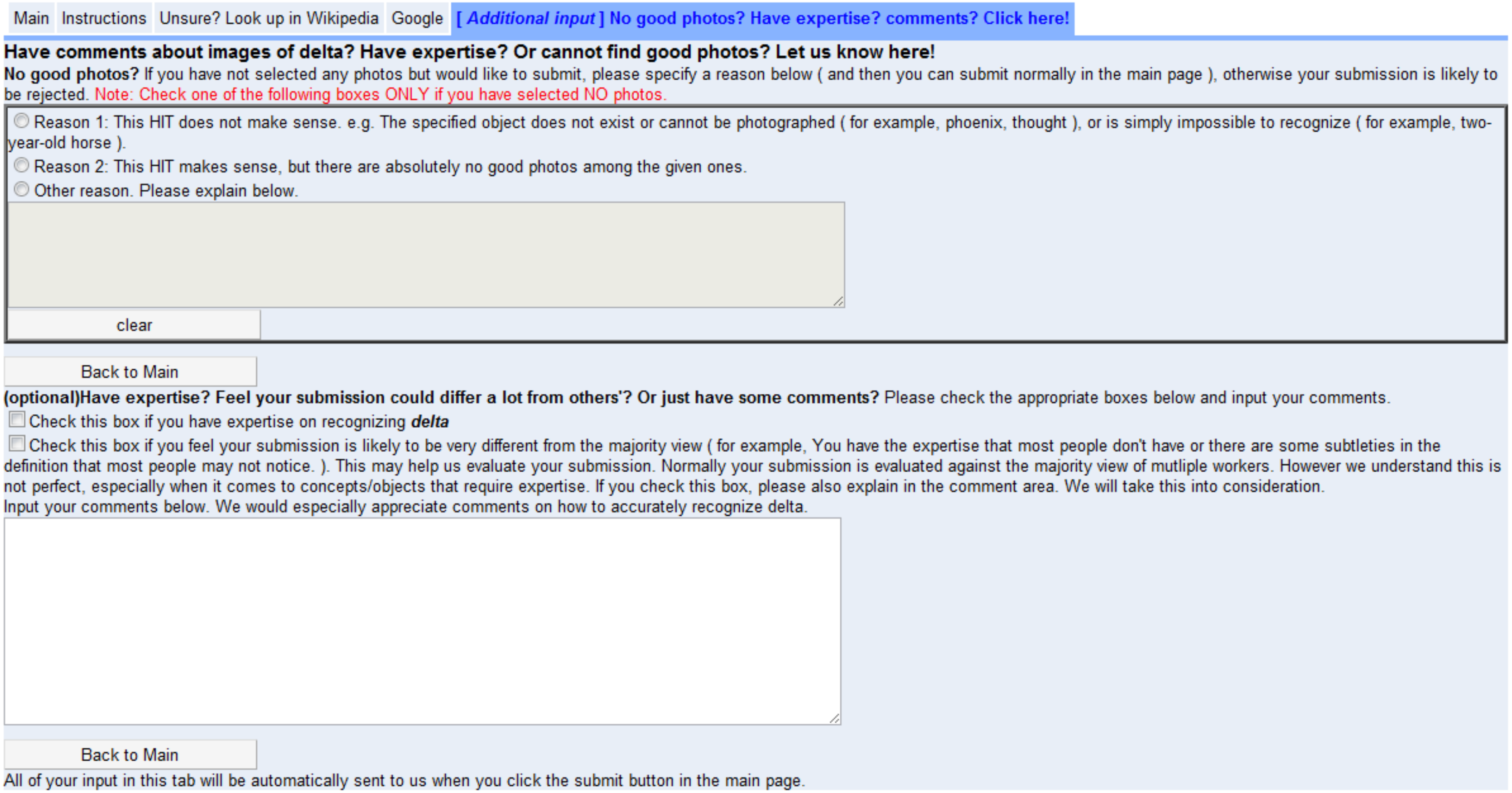
○ a low triangular area of alluvial deposits where a river divides before entering a larger body of water; "the Mississippi River delta"; "the Nile delta"

○ an airplane with wings that give it the appearance of an isosceles triangle

○ an object shaped like an equilateral triangle

Enhancement 3

- Allow more feedback. E.g. “unimagable synsets” expert opinion


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

Have comments about images of delta? Have expertise? Or cannot find good photos? Let us know here!

No good photos? If you have not selected any photos but would like to submit, please specify a reason below (and then you can submit normally in the main page), otherwise your submission is likely to be rejected. Note: Check one of the following boxes ONLY if you have selected NO photos.

○ Reason 1: This HIT does not make sense. e.g. The specified object does not exist or cannot be photographed (for example, phoenix, thought), or is simply impossible to recognize (for example, two-year-old horse).

○ Reason 2: This HIT makes sense, but there are absolutely no good photos among the given ones.

○ Other reason. Please explain below.

(optional)Have expertise? Feel your submission could differ a lot from others'? Or just have some comments? Please check the appropriate boxes below and input your comments.

□ Check this box if you have expertise on recognizing ***delta***

□ Check this box if you feel your submission is likely to be very different from the majority view (for example, You have the expertise that most people don't have or there are some subtleties in the definition that most people may not notice.). This may help us evaluate your submission. Normally your submission is evaluated against the majority view of mutliple workers. However we understand this is not perfect, especially when it comes to concepts/objects that require expertise. If you check this box, please also explain in the comment area. We will take this into consideration.

Input your comments below. We would especially appreciate comments on how to accurately recognize delta.

All of your input in this tab will be automatically sent to us when you click the submit button in the main page.

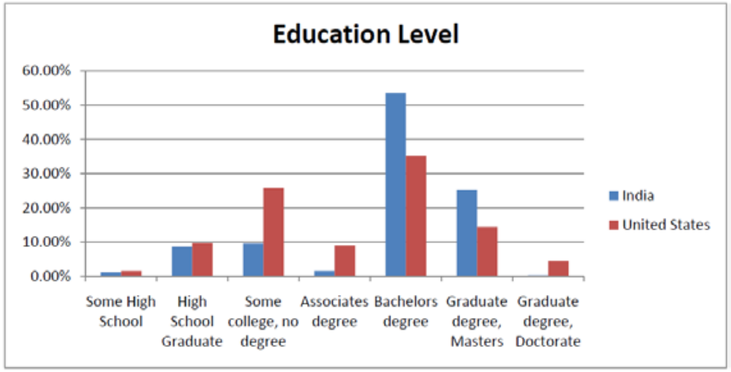
IMAGENET is built by crowdsourcing

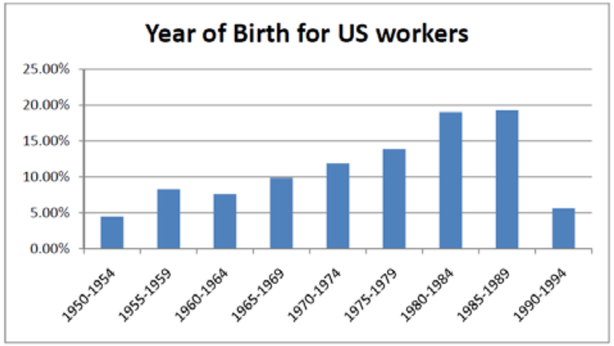
- July 2008: 0 images
- Dec 2008: 3 million images, 6000+ synsets
- April 2010: 11 million images, 15,000+ synsets

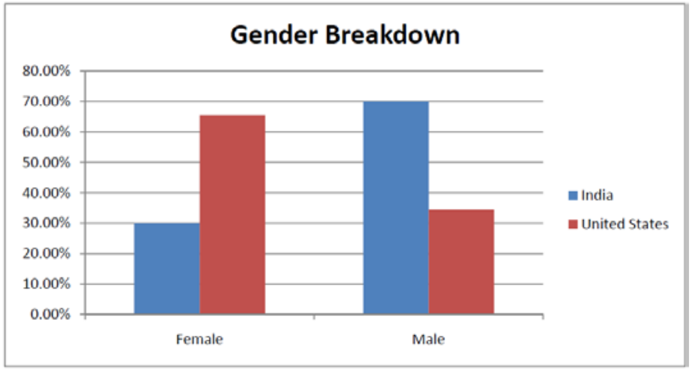
So are we exploiting chained prisoners?

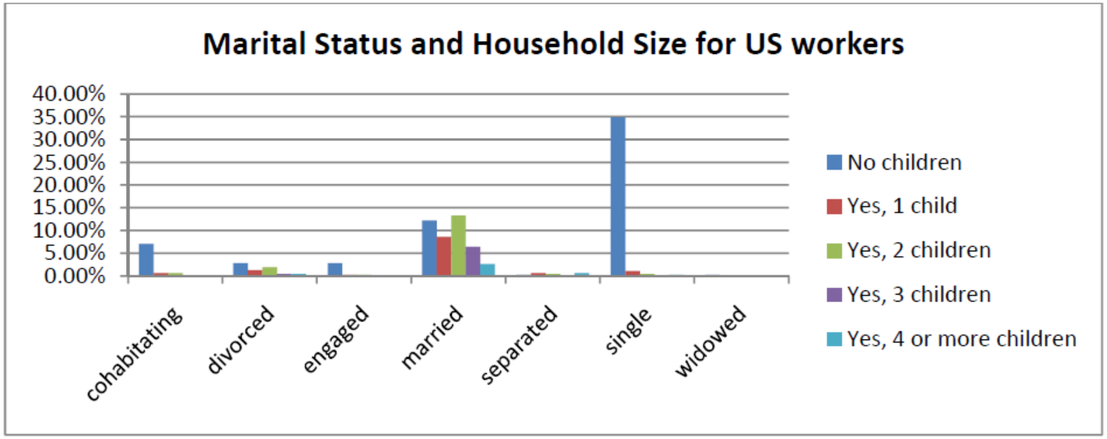
Demography of AMT workers

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





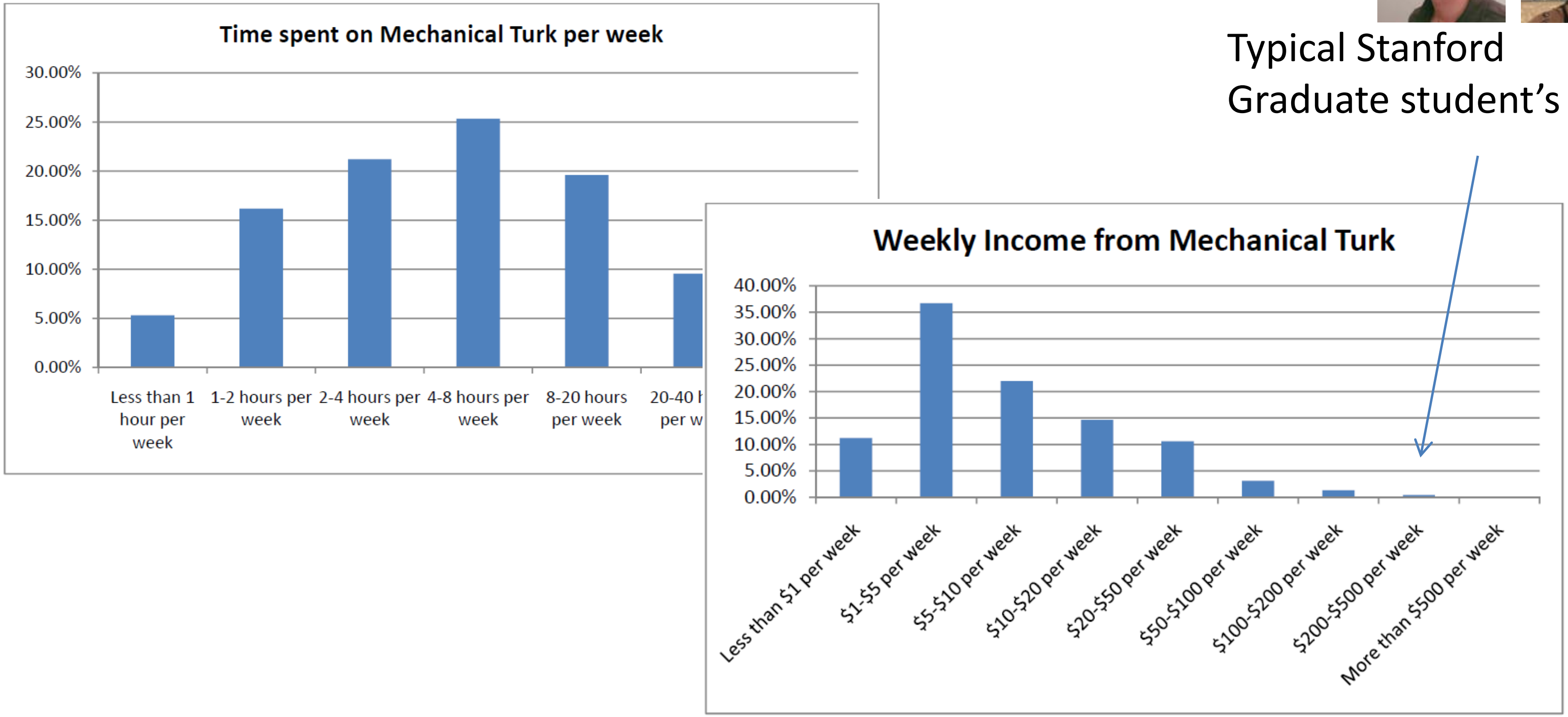




Panos Ipeirotis, NYU, Feb, 2010

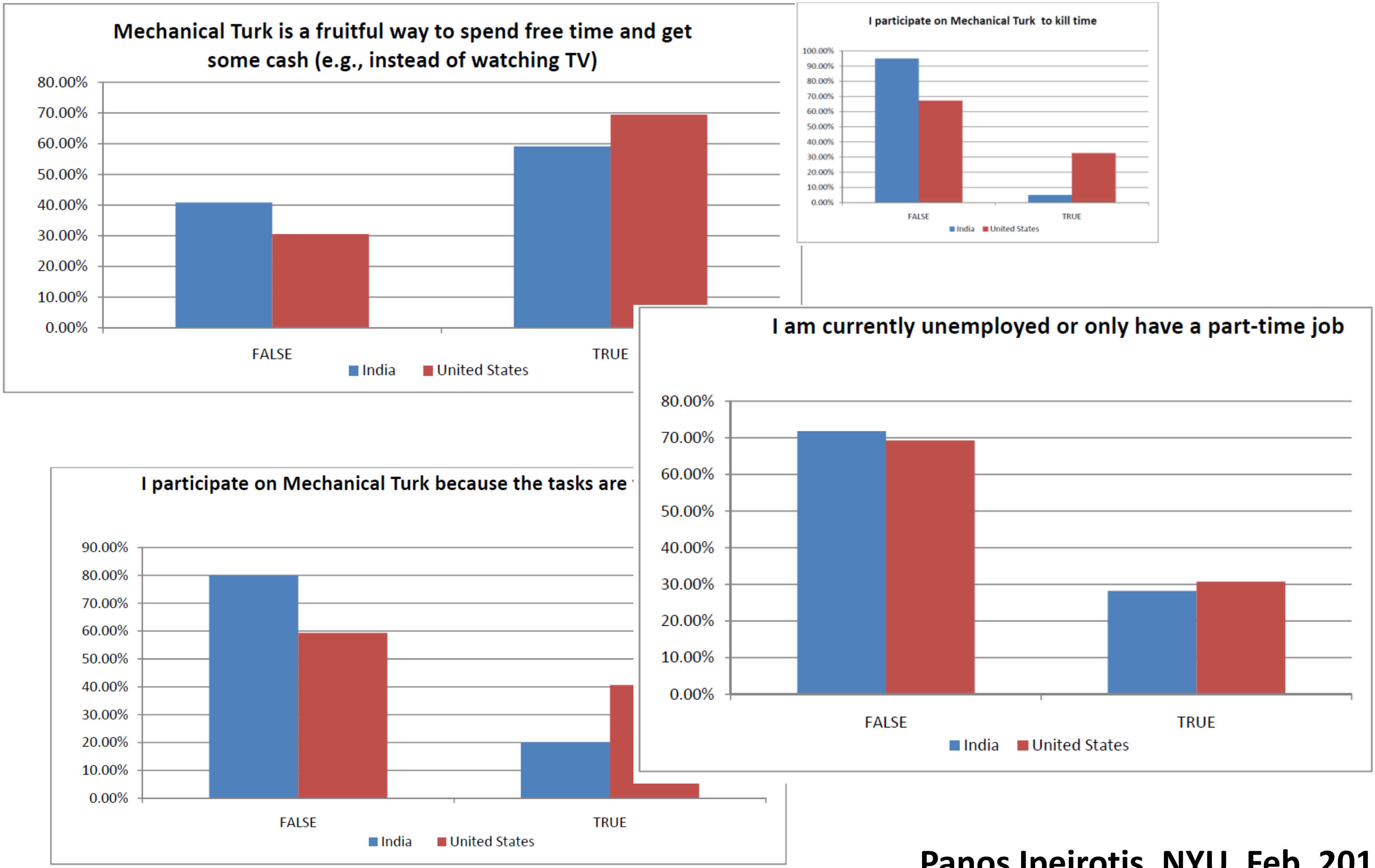
Demography of AMT workers

Typical Stanford
Graduate student's income



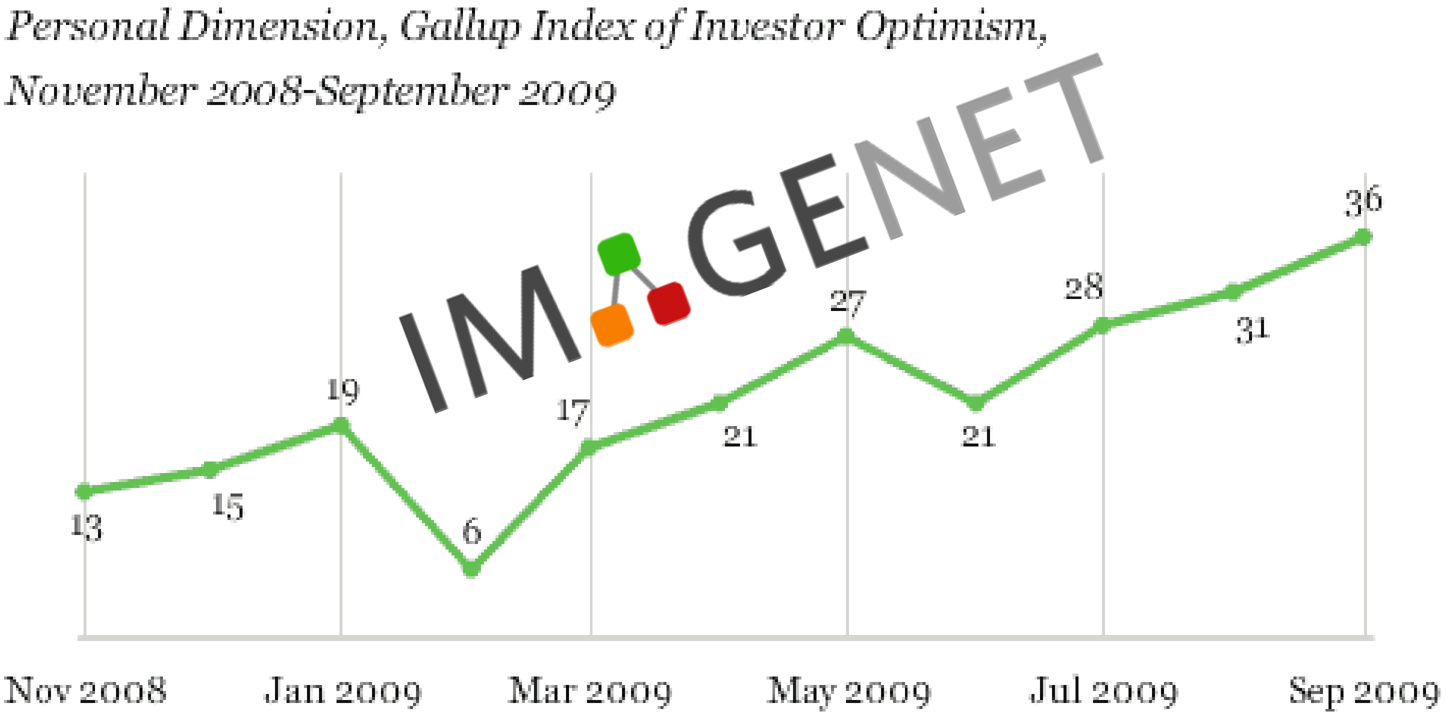
Panos Ipeirotis, NYU, Feb, 2010

Demography of AMT workers



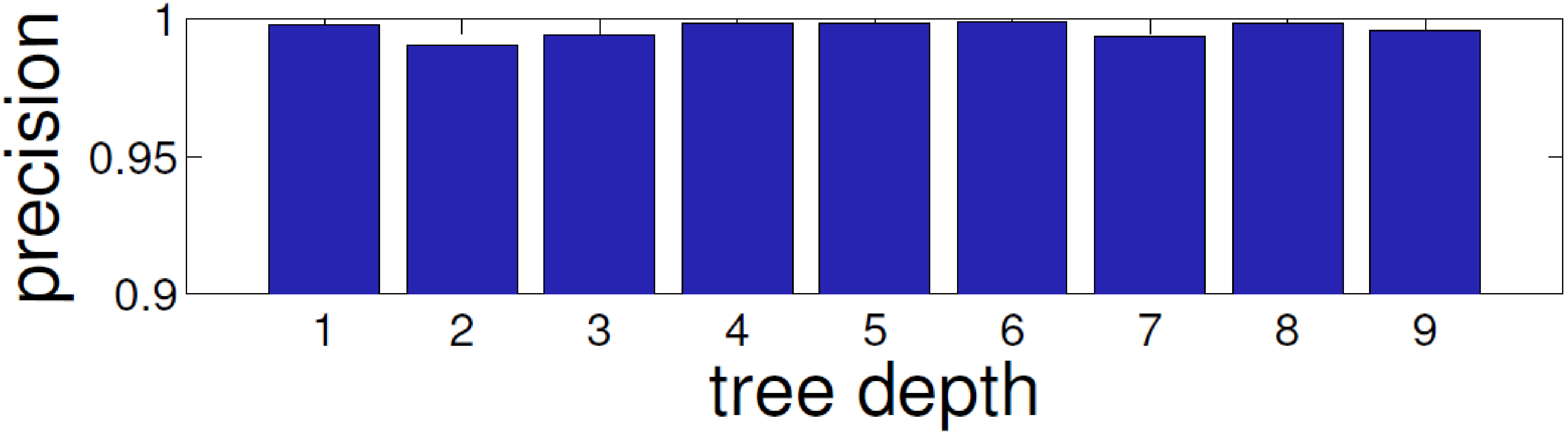
Panos Ipeirotis, NYU, Feb, 2010

U.S. economy 2008 - 2009



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

Accuracy



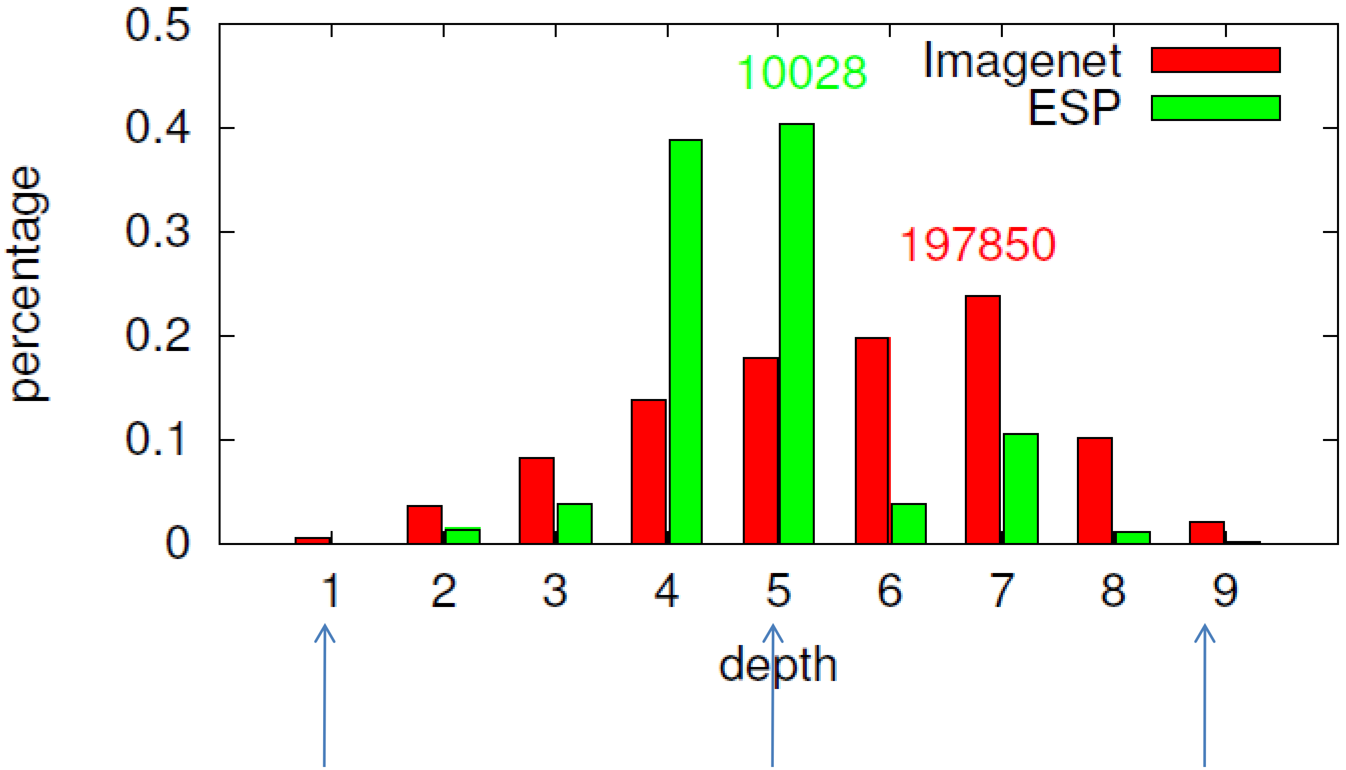
e.g. mammal

e.g. dog

e.g. German Shepherd

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

Diversity



e.g. mammal

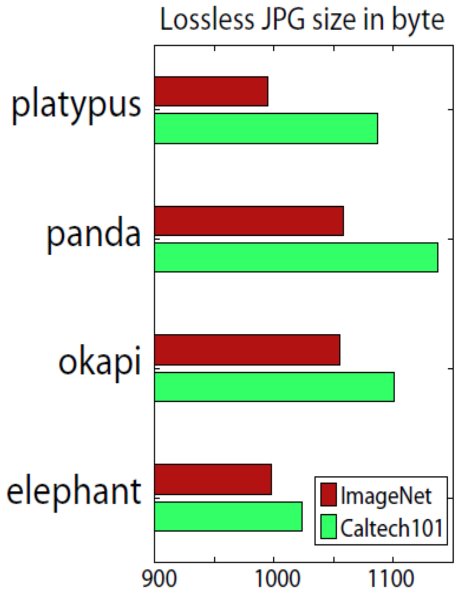
e.g. dog

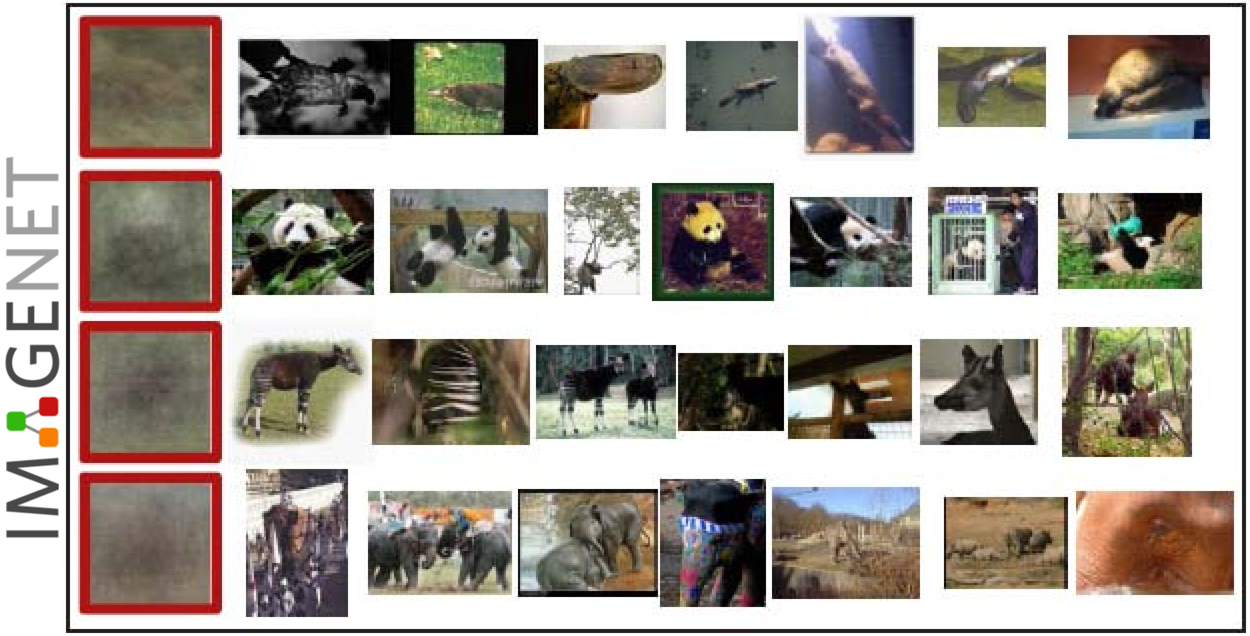
e.g. German Shepherd

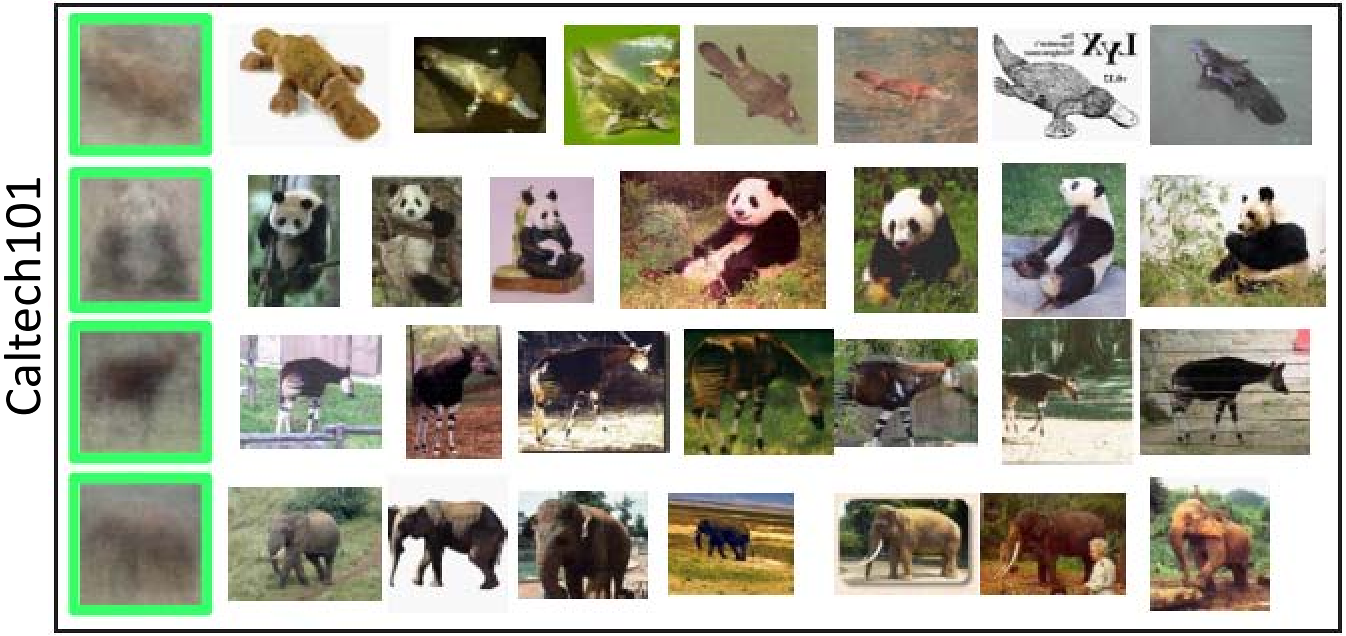
ESP: Ahn et al. 2006

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

Diversity

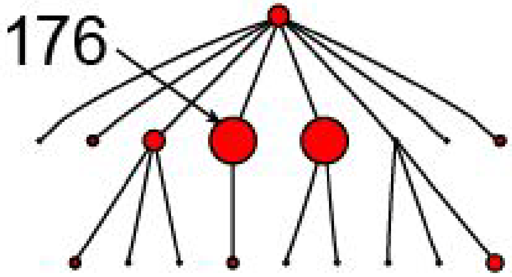

Lossless JPG size in byte
platypus
panda
okapi
elephant
ImageNet
Caltech101
900
1000
1100


IMAGENET

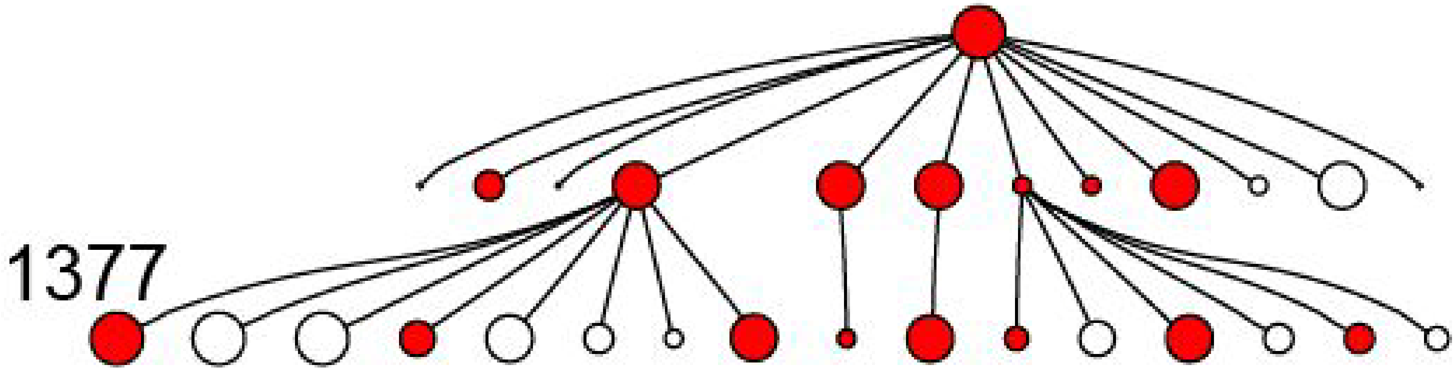

Caltech101

Semantic hierarchy

ESP Cattle Subtree

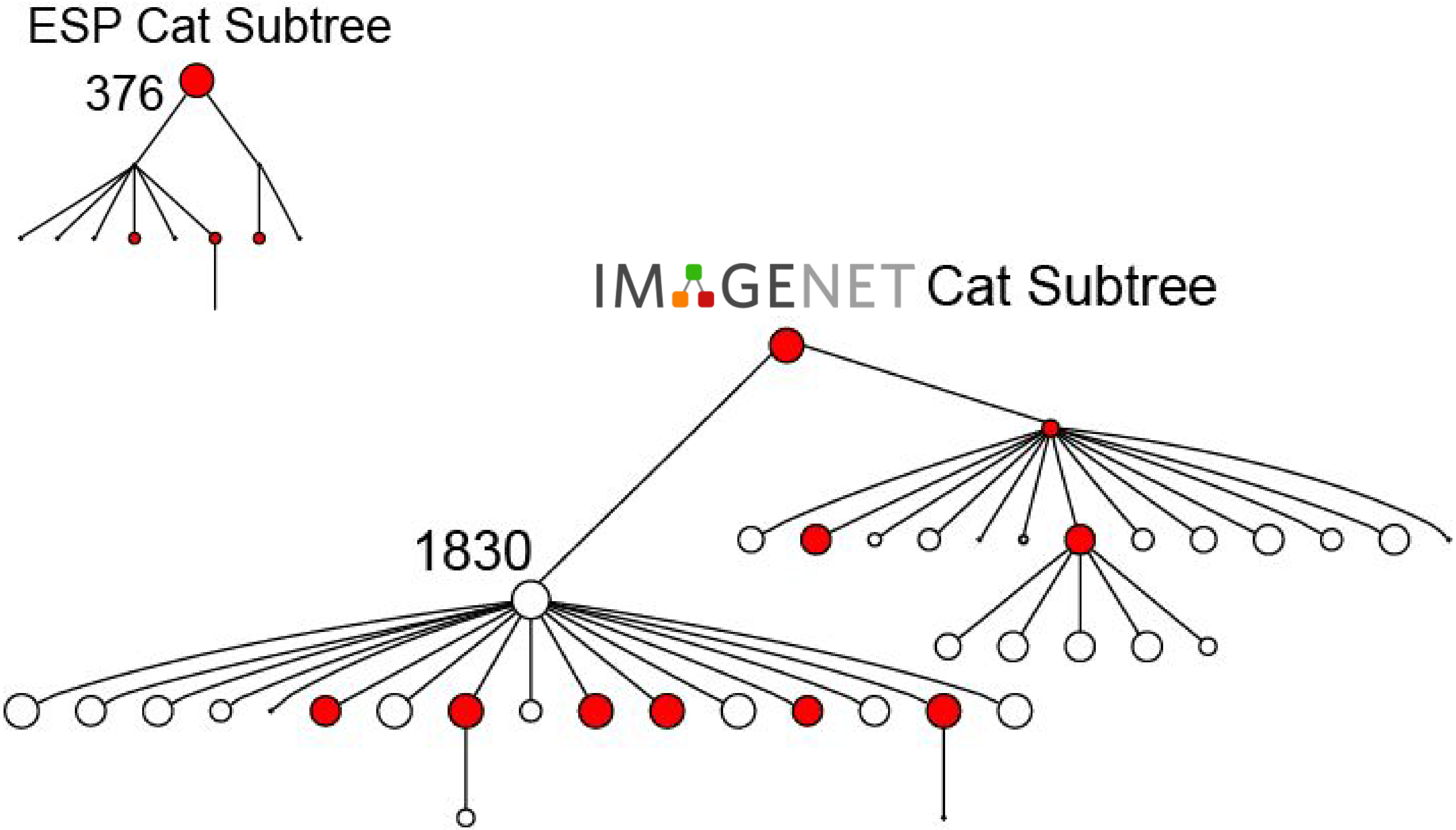


IMAGENET Cattle Subtree



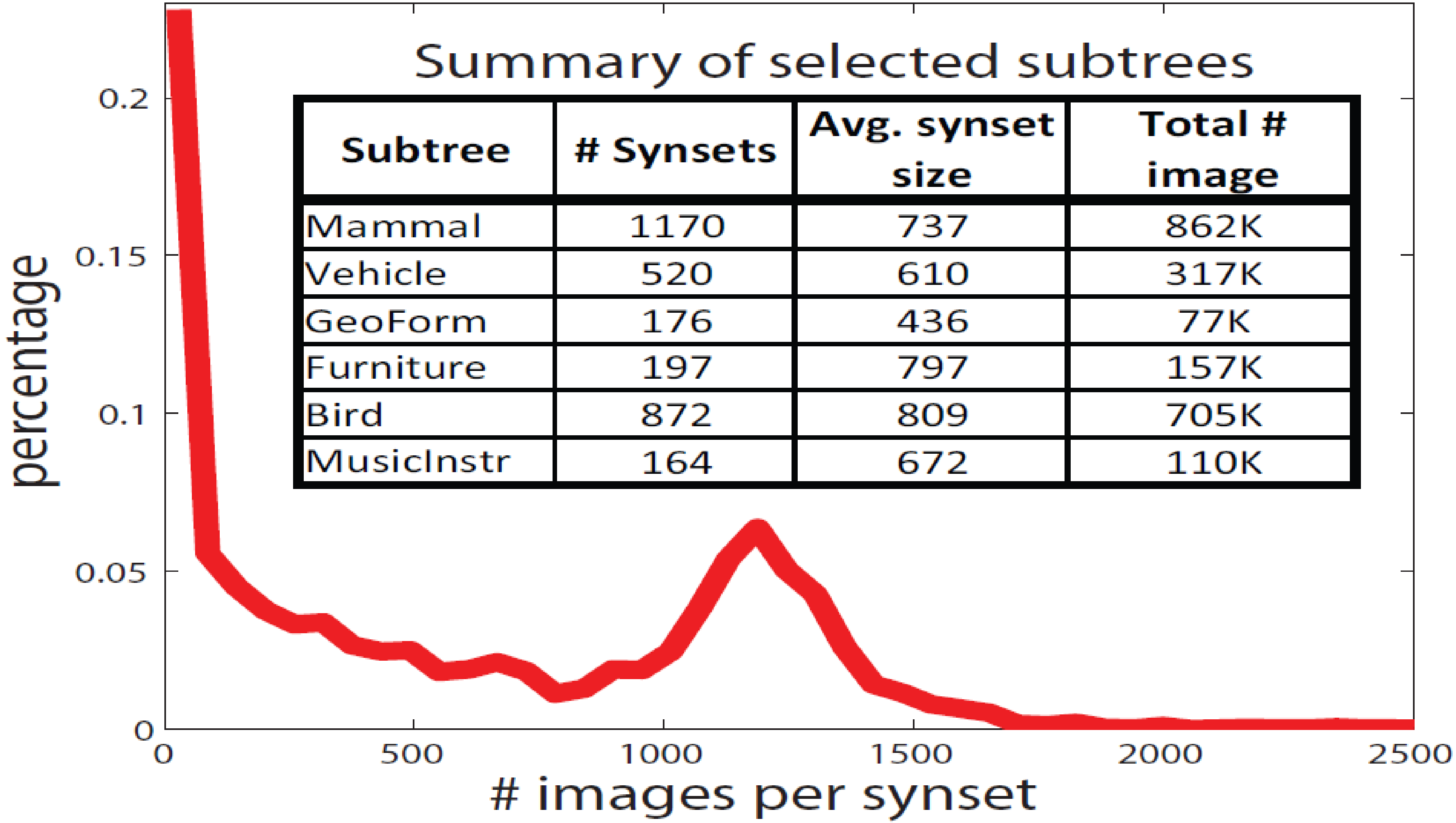
Deng, Dong, Socher, Li, Li, & Fei-Fei, *CVPR*, 2009

Semantic hierarchy



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

Scale

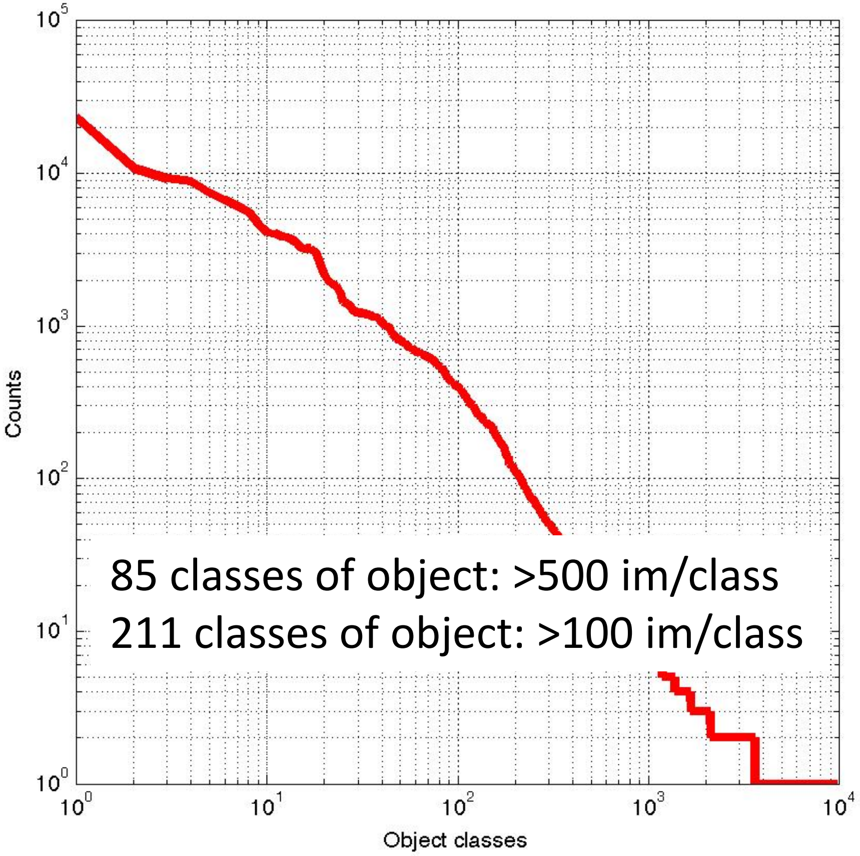


Summary of selected subtrees

Subtree	# Synsets	Avg. synset size	Total # image
Mammal	1170	737	862K
Vehicle	520	610	317K
GeoForm	176	436	77K
Furniture	197	797	157K
Bird	872	809	705K
MusicInstr	164	672	110K

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

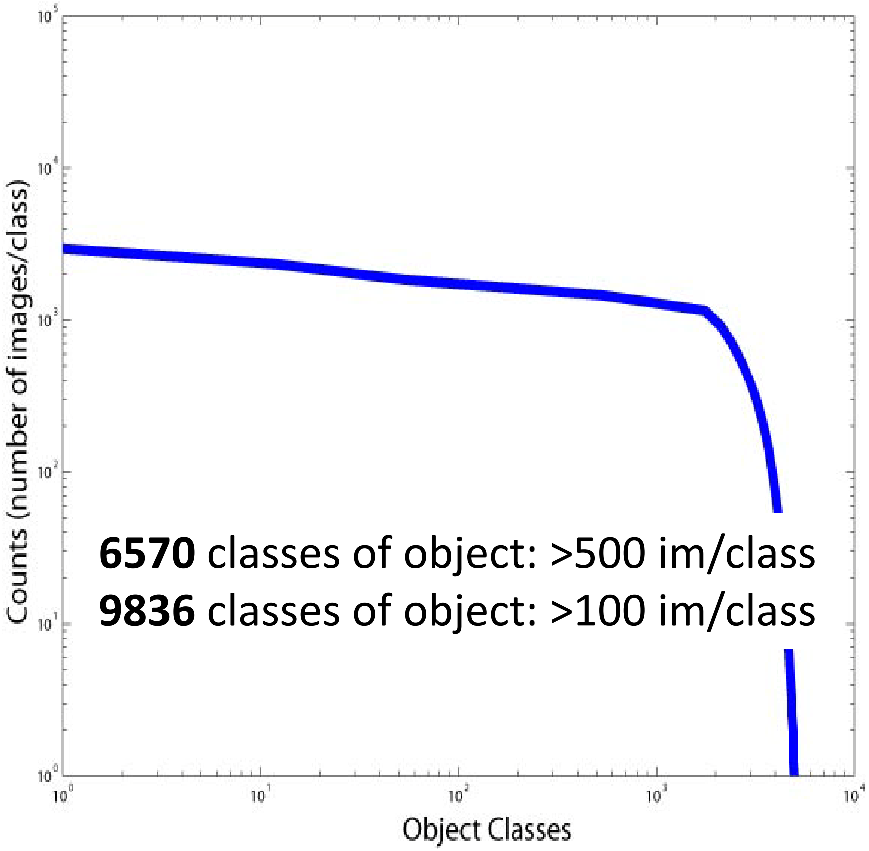
Scale



85 classes of object: >500 im/class
211 classes of object: >100 im/class

LabelMe

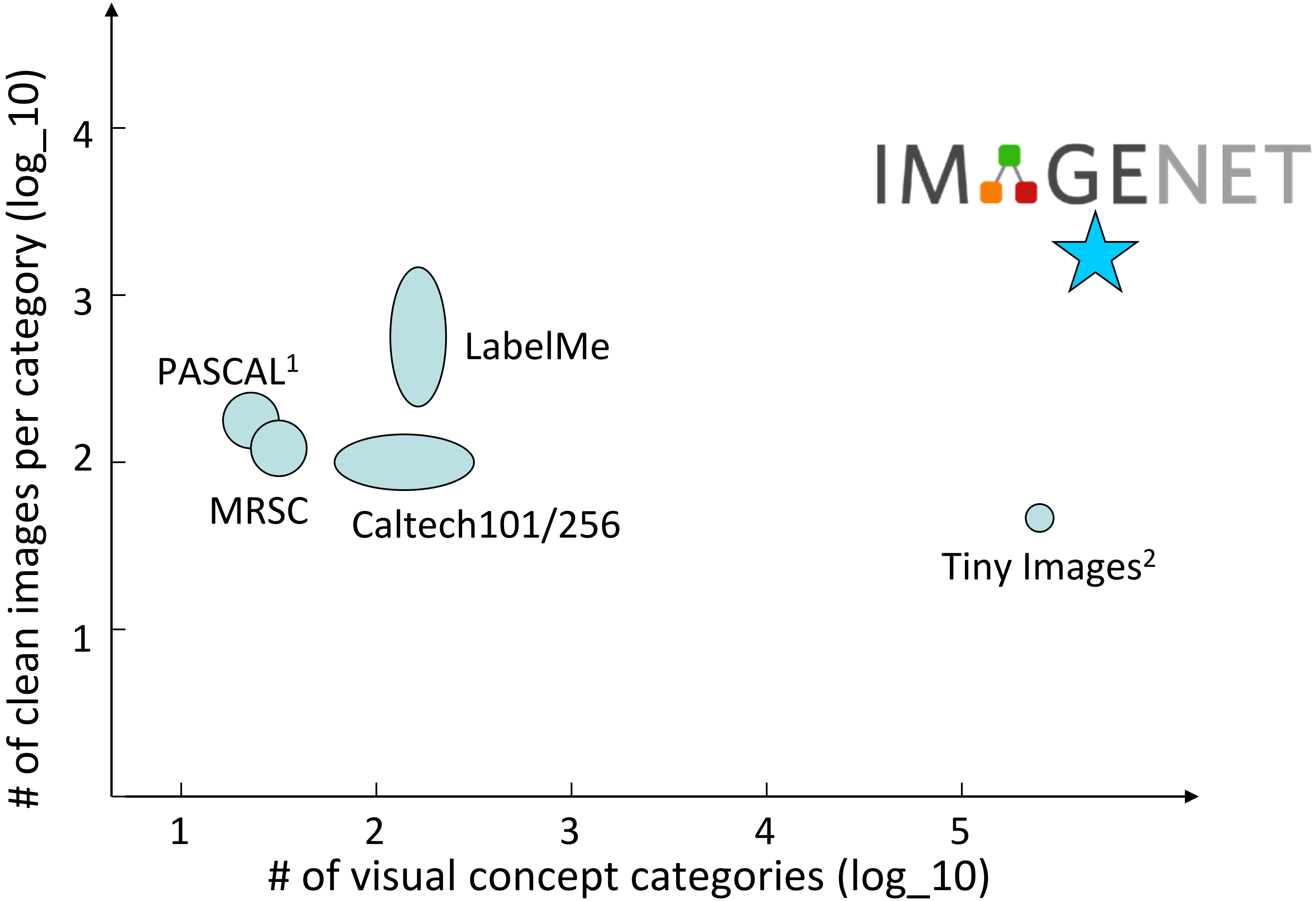
Russell et al. 2005;
statistics obtained in 2009



6570 classes of object: >500 im/class
9836 classes of object: >100 im/class

IMAGENET

Comparison among free datasets

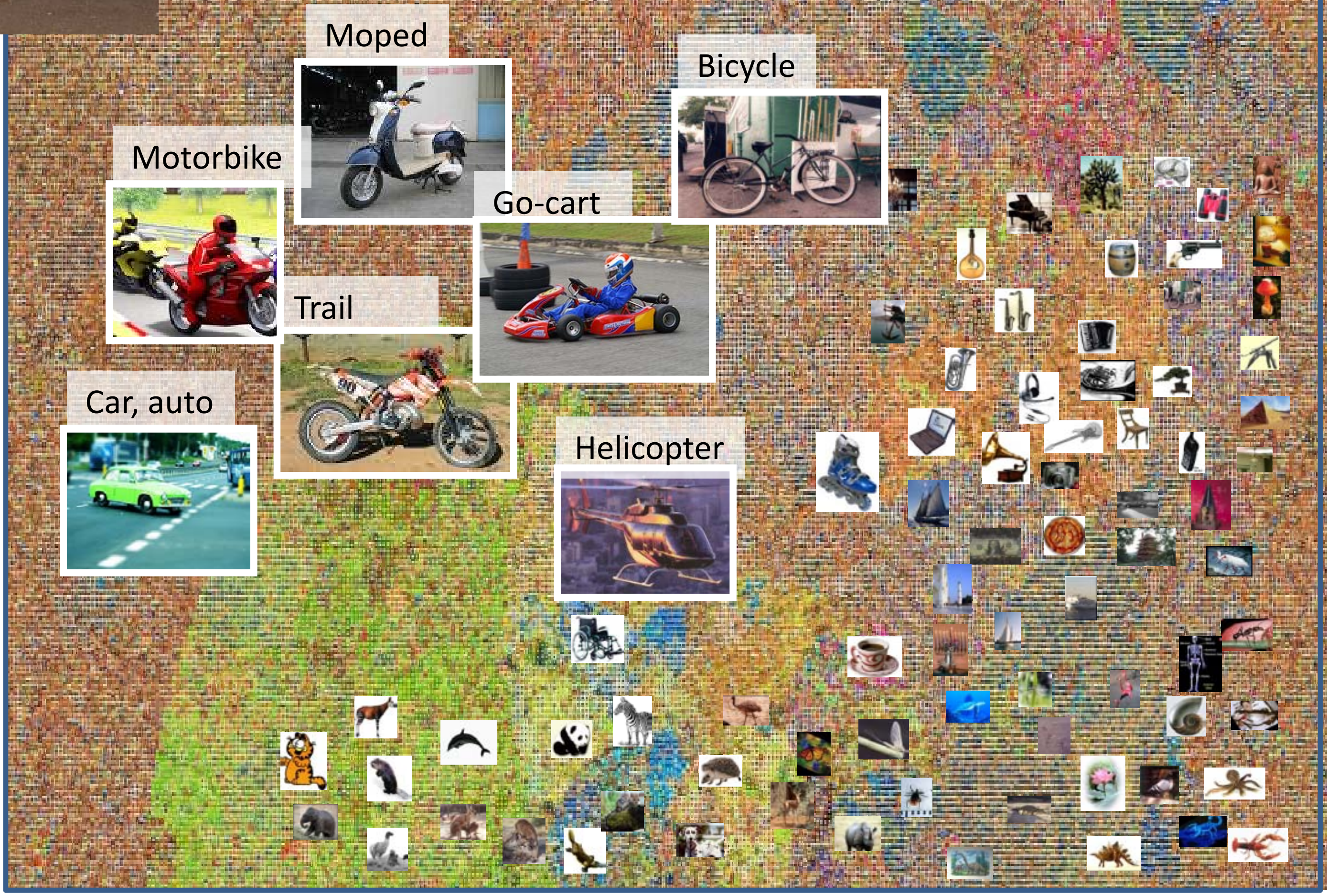


1. Excluding the Caltech101 datasets from PASCAL
2. No image in this dataset is human annotated. The # of clean images per category is a rough estimation

outline

- Goal of ImageNet:
  - A dataset
  - A knowledge ontology
- 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
- Human vision: Rosch revisited and quantified
  - Quantifying basic-, subordinate- and superordinate-level concepts
- In the horizon: ImageNet Spring 2010 release
  - The upcoming ImageNet Challenge (in partnership with PASCAL VOC)
  - Visualizing ImageNet
  - Etc.

What does classifying more than 10,000 image categories tell us?



Background image courtesy: Antonio Torralba

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, *ECCV2010*

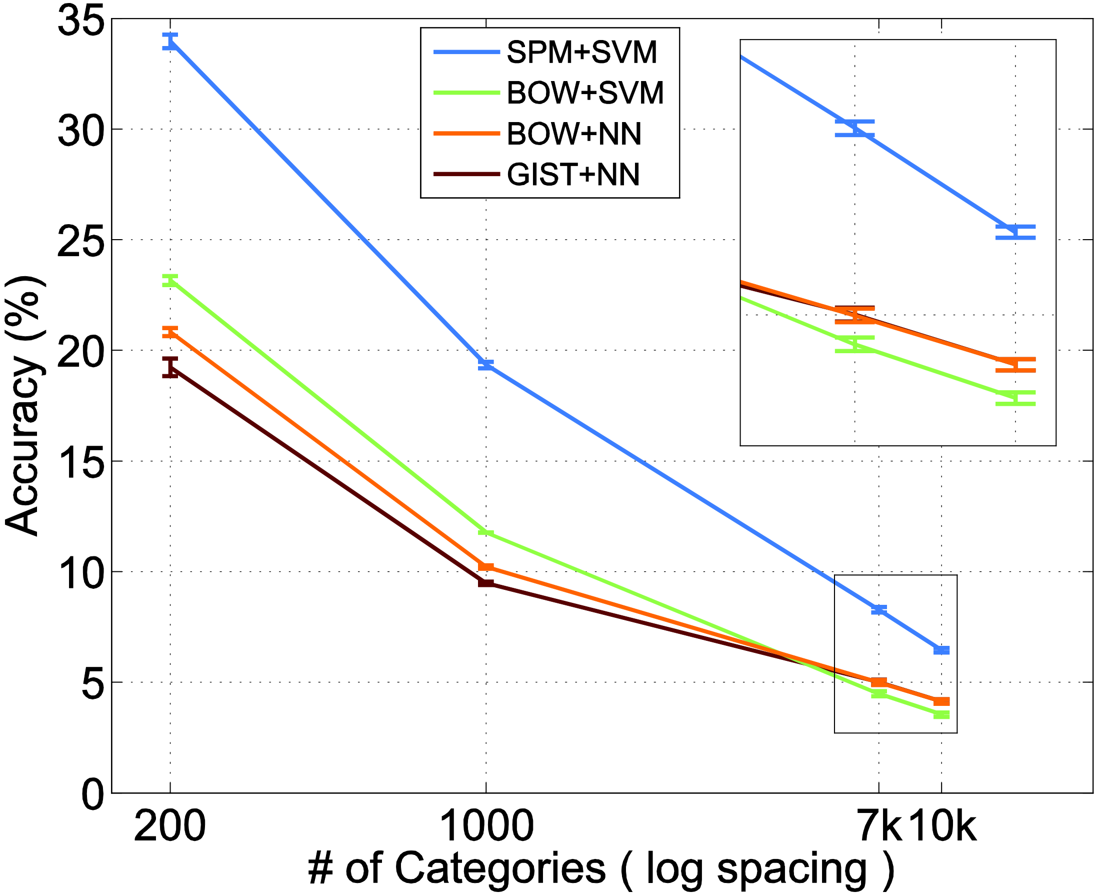
Computation issues first

- 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, *ECCV2010*

Size matters

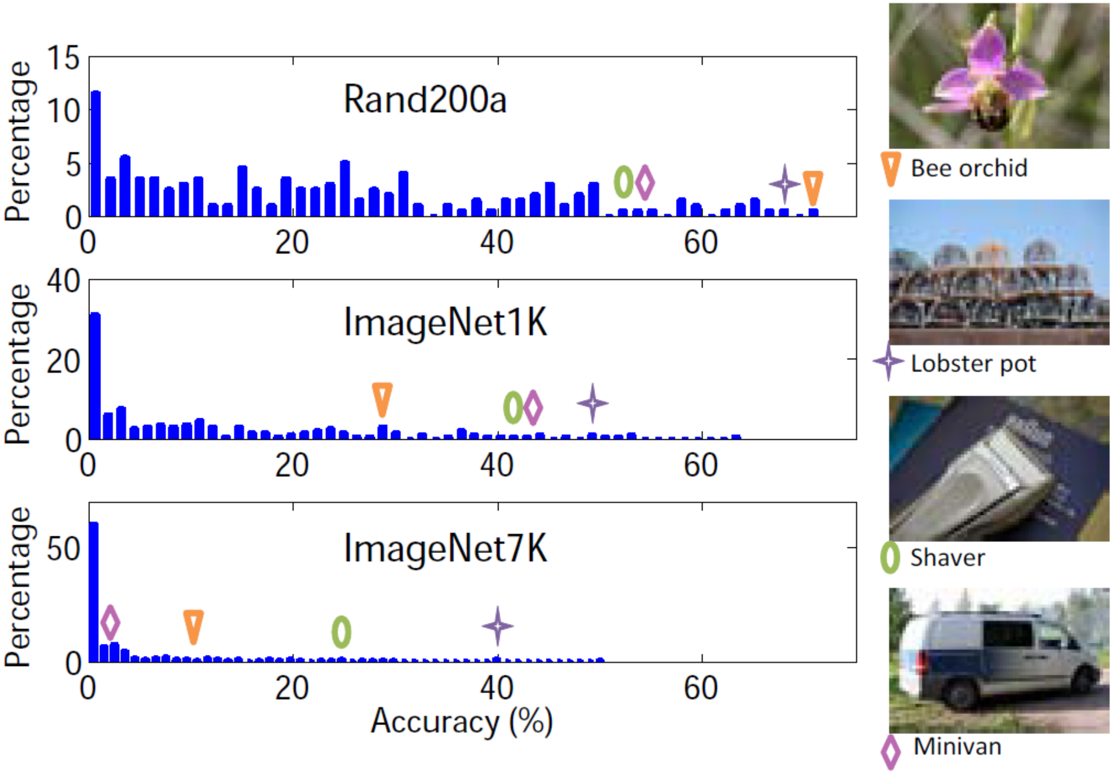
- 6.4% for 10K categories
- Better than we expected (instead of dropping at the rate of 10x; it's roughly at about 2x)
- An ordering switch between SVM and NN methods when the # of categories becomes large



Deng, Berg, Li, & Fei-Fei, *ECCV2010*

Size matters

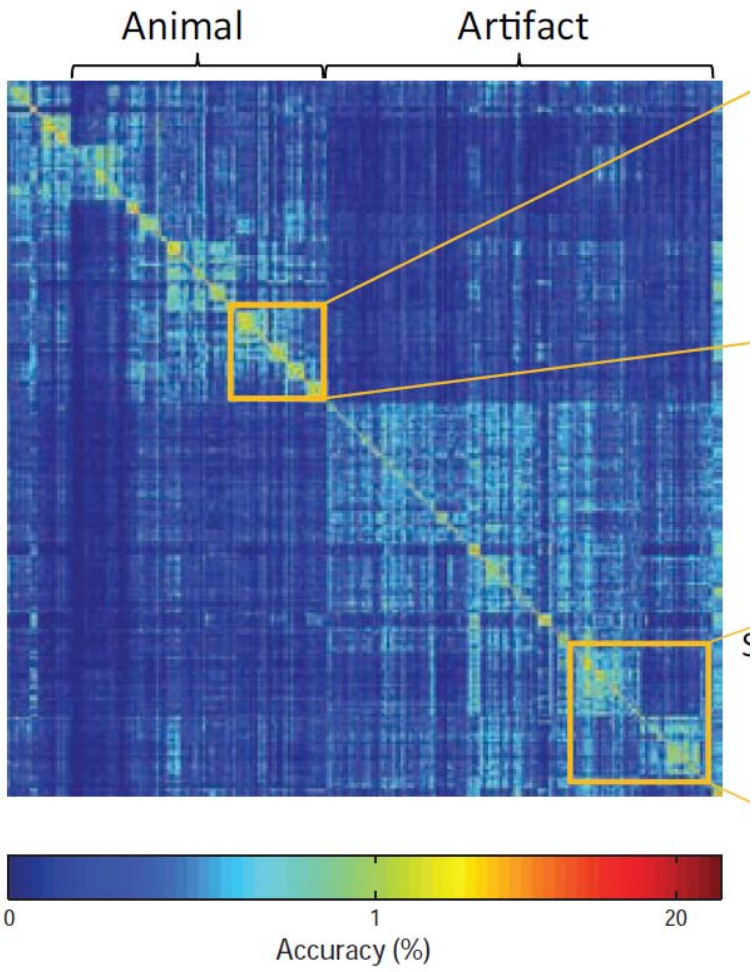
- 6.4% for 10K categories
- Better than we expected (instead of dropping at the rate of 10x; it's roughly at about 2x)
- An ordering switch between SVM and NN methods when the # of categories becomes large
- When dataset size varies, conclusion we can draw about different categories varies



Deng, Berg, Li, & Fei-Fei, *ECCV2010*

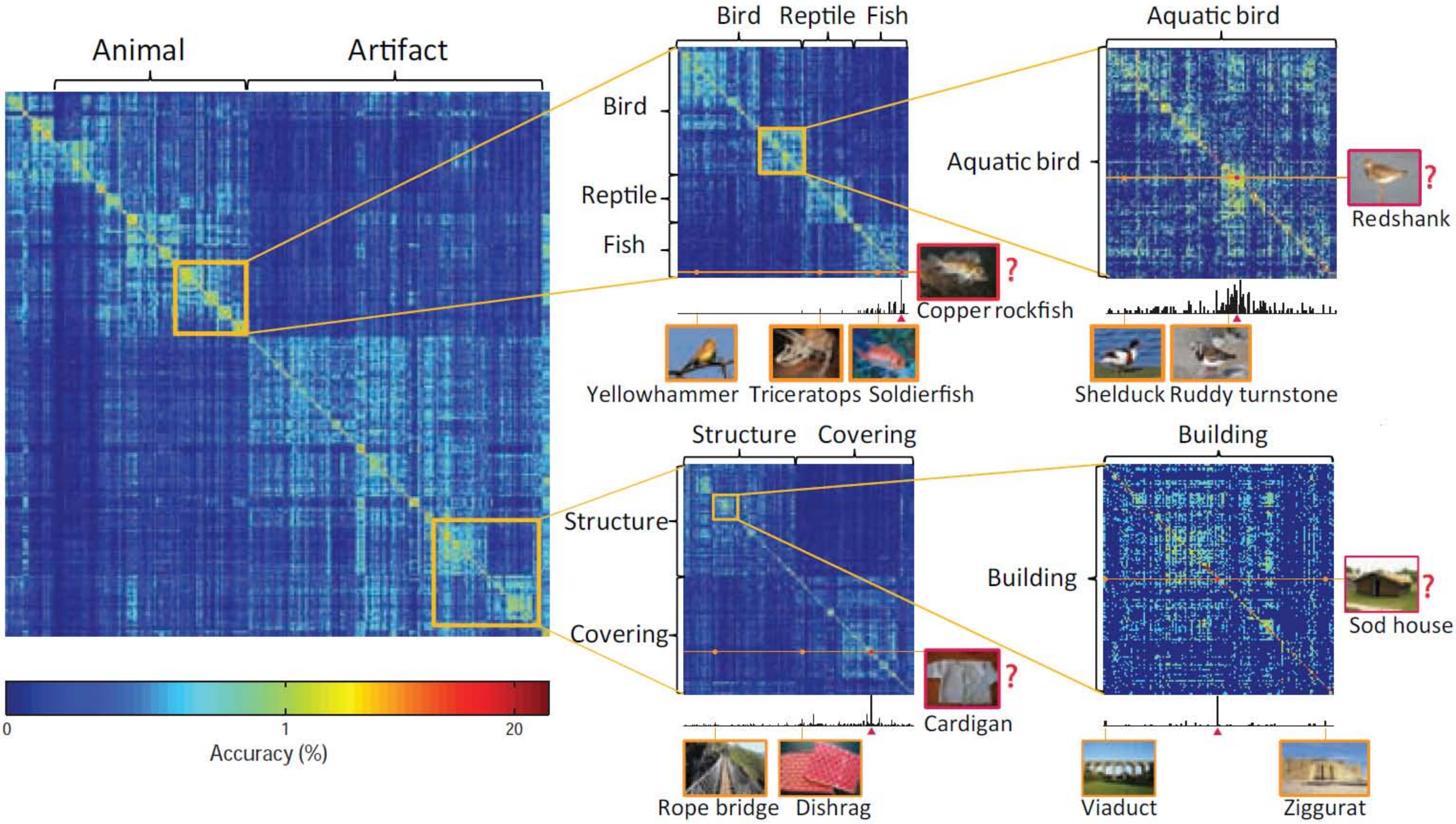
Size matters

- 6.4% 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)



Deng, Berg, Li, & Fei-Fei, *ECCV2010*

Size matters

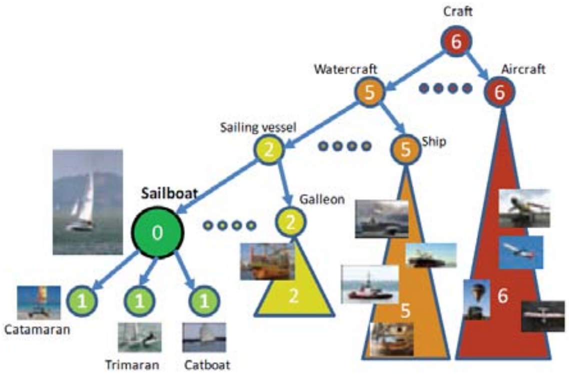


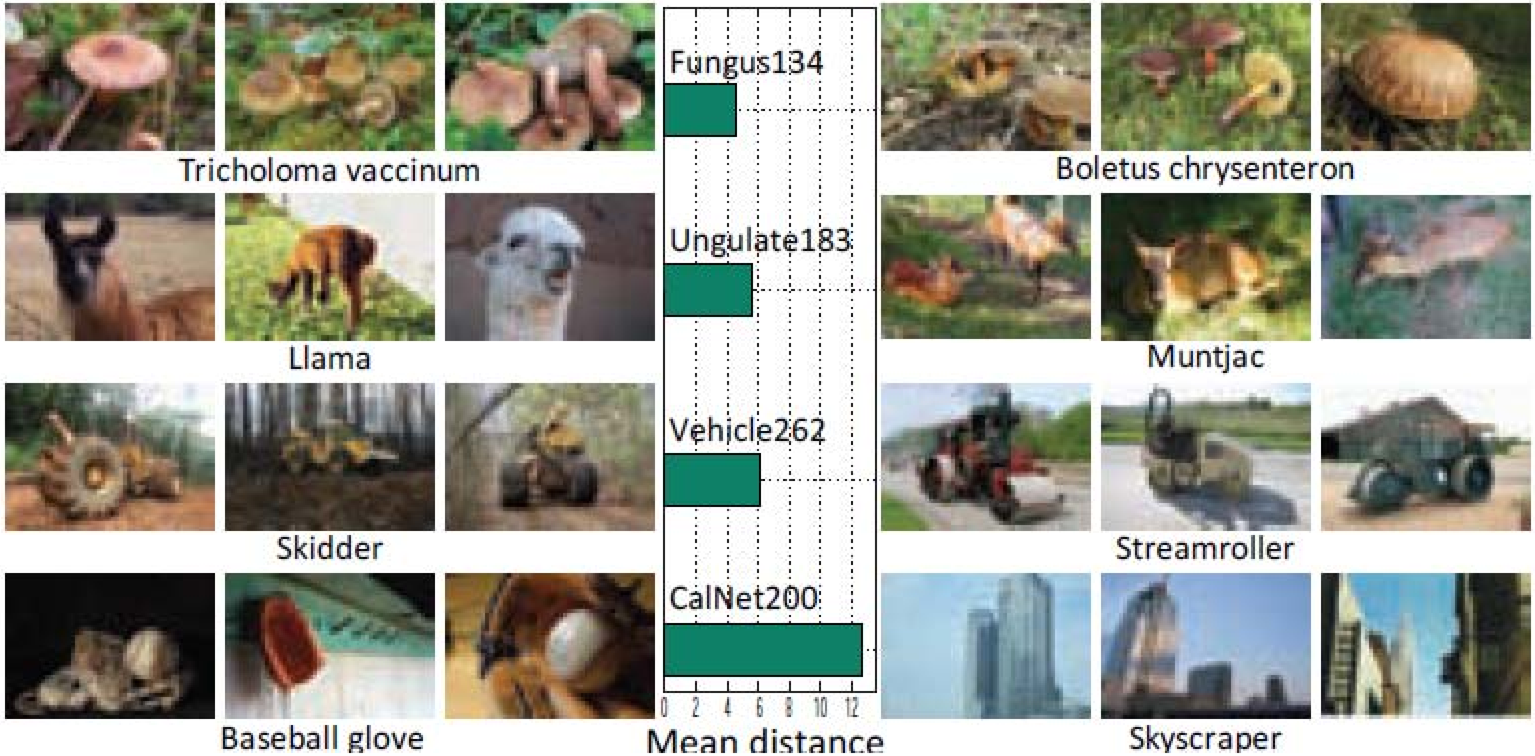
exhibits meaningful visual structure (ordered by DFS)

Deng, Berg, Li, & Fei-Fei, *ECCV2010*

Density matters

- Datasets have very different “density” or “sparsity”

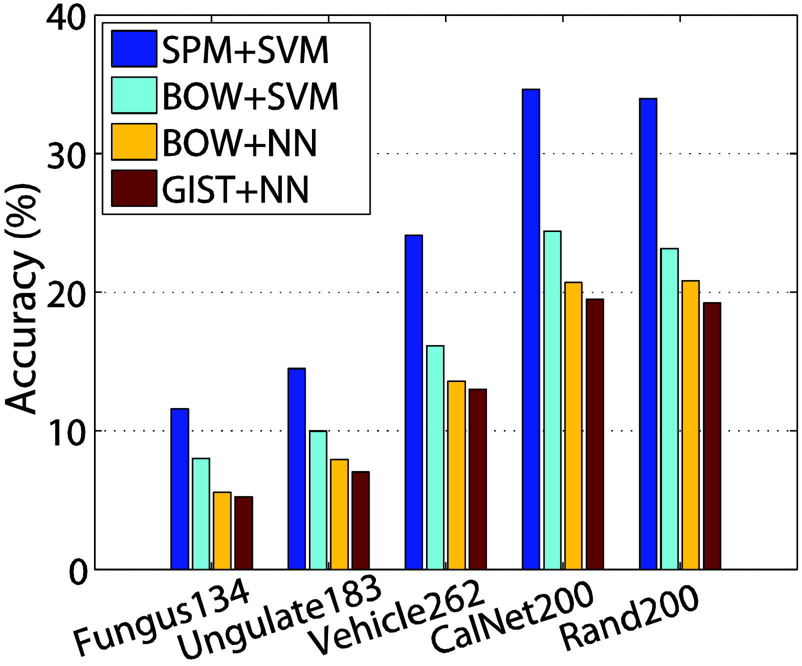


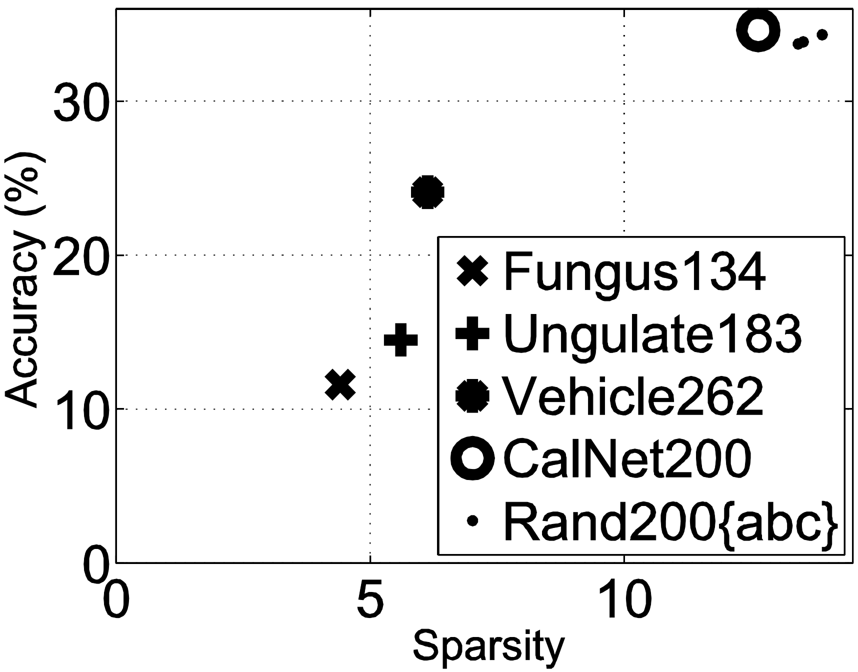


Deng, Berg, Li, & Fei-Fei, *ECCV2010*

Density matters

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





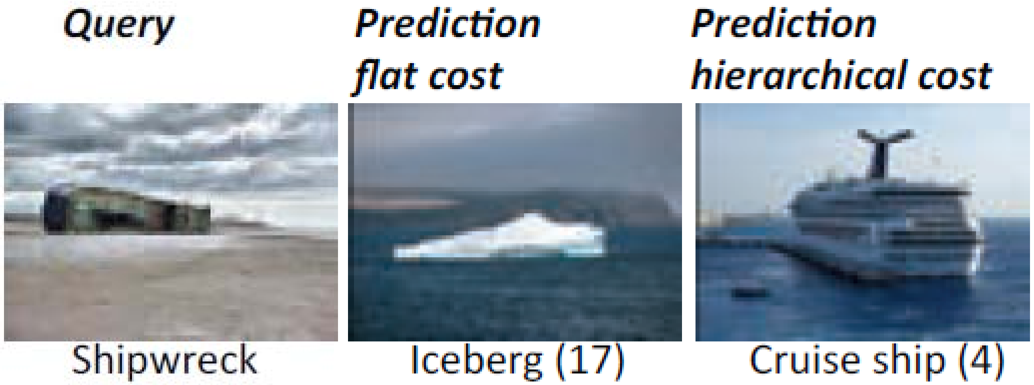
Dense ⟵ ⟶ Sparse   Dense ⟵ ⟶ Sparse

Deng, Berg, Li, & Fei-Fei, *ECCV2010*

Hierarchy matters

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

$$C_{i,j} = \begin{cases} 0 & i=j\text{, or } i \text{ is a descendent of } j \\ h(i,j) & h \text{ is the height of the lowest common ancestor in WordNet} \end{cases}$$

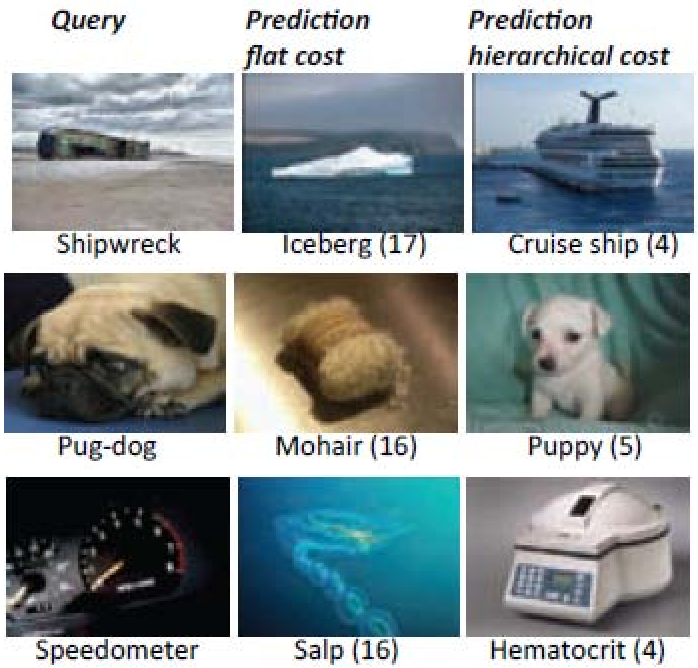


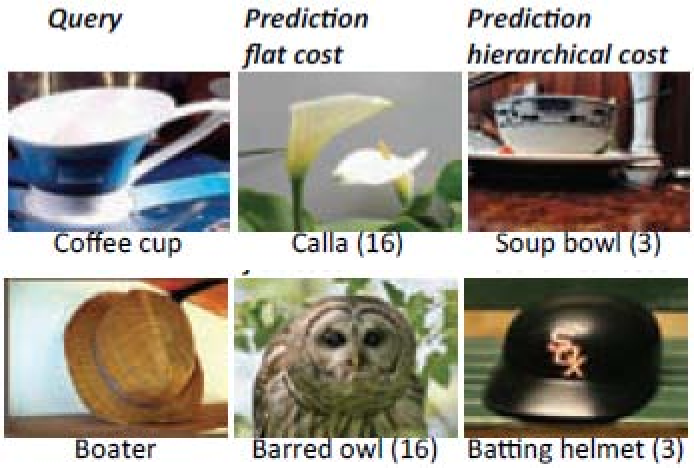
Deng, Berg, Li, & Fei-Fei, *ECCV2010*

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

$$C_{i,j} = \begin{cases} 0 & \textit{i=j}\text{, or } \textit{i} \text{ is a descendent of } \textit{j} \\ h(i,j) & \textit{h is the height of the lowest common ancestor in WordNet} \end{cases}$$



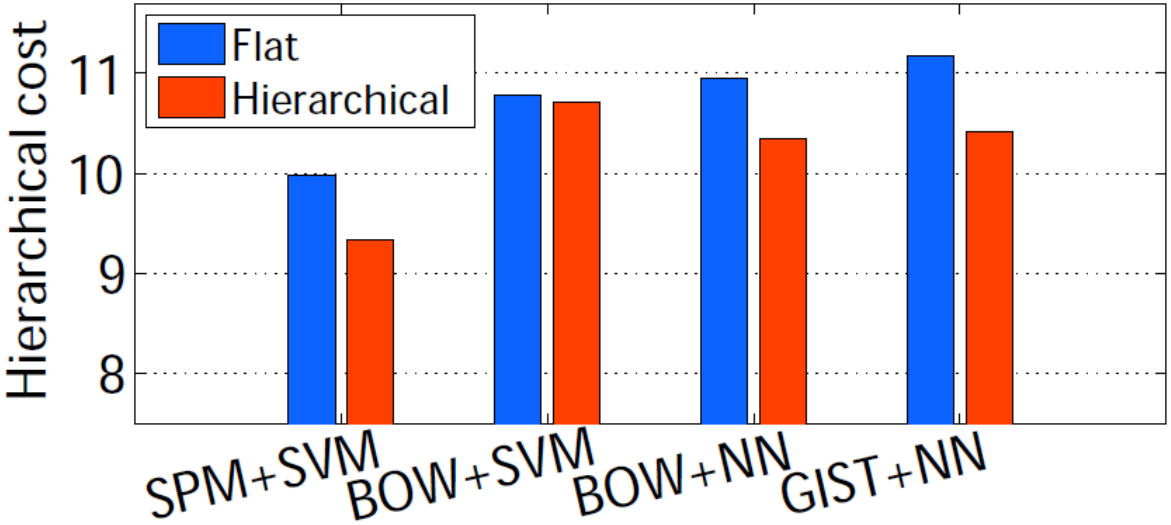


Deng, Berg, Li, & Fei-Fei, *ECCV2010*

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

$$C_{i,j} = \begin{cases} 0 & \textit{i=j}\text{, or } \textit{i} \text{ is a descendent of } \textit{j} \\ h(i,j) & \textit{h is the height of the lowest common ancestor in WordNet} \end{cases}$$

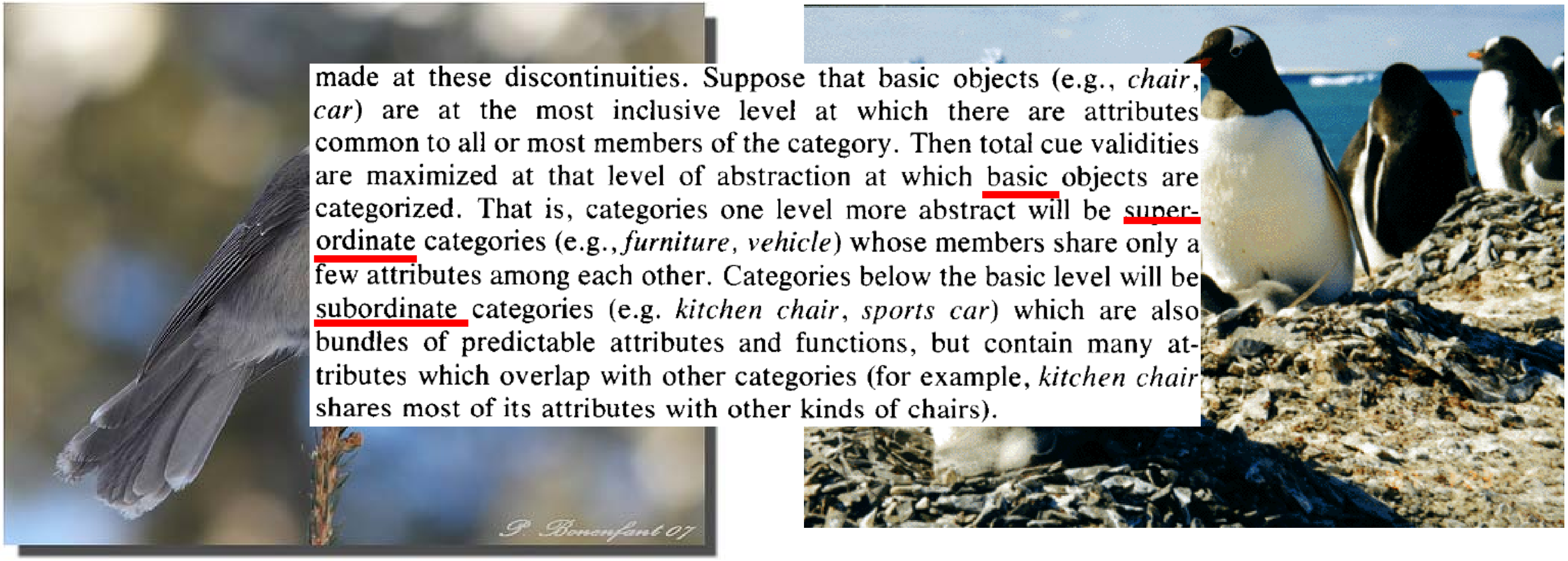


Deng, Berg, Li, & Fei-Fei, *ECCV2010*

outline

- Goal of ImageNet:
  - A dataset
  - A knowledge ontology
- 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
- Human vision: Rosch revisited and quantified
  - Quantifying basic-, subordinate- and superordinate-level concepts
- In the horizon: ImageNet Spring 2010 release
  - The upcoming ImageNet Challenge (in partnership with PASCAL VOC)
  - Visualizing ImageNet
  - Etc.

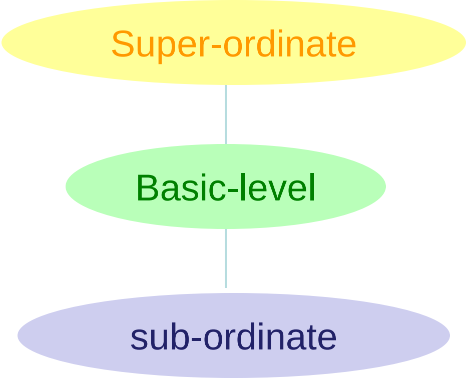
Eleanor Rosch re-visited and quantified


made at these discontinuities. Suppose that basic objects (e.g., *chair*, *car*) are at the most inclusive level at which there are attributes common to all or most members of the category. Then total cue validities are maximized at that level of abstraction at which basic objects are categorized. That is, categories one level more abstract will be superordinate categories (e.g., *furniture*, *vehicle*) whose members share only a few attributes among each other. Categories below the basic level will be subordinate categories (e.g. *kitchen chair*, *sports car*) which are also bundles of predictable attributes and functions, but contain many attributes which overlap with other categories (for example, *kitchen chair* shares most of its attributes with other kinds of chairs).

Vertebrate

Bird

Canadian gray jay


Super-ordinate

Basic-level

sub-ordinate

Vertebrate

Bird

Penguin

Rosch et al. 1976

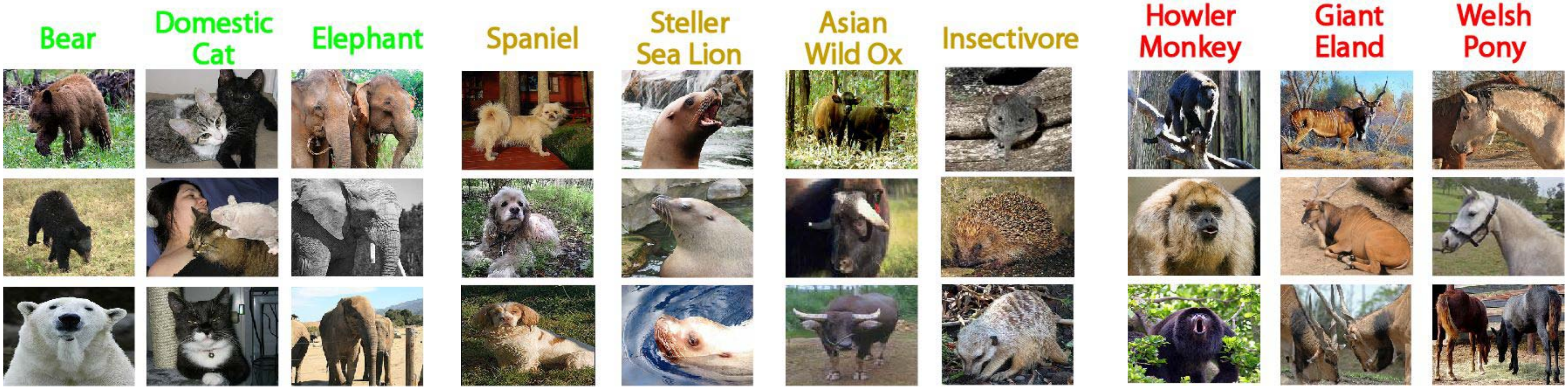
Eleanor Rosch re-visited and quantified

- What do we have? Multiple AMT workers vote on whether an image belongs to a synset
- Intuition. Divergence (d) of votes reflect discriminability of the image: the higher the d, the less discriminable the image.
- How do we measure? Information theoretic analysis (entropy)

$$d(\text{image}) = -(f\log(f) + (1\text{-}f)\log(1\text{-}f)) \qquad D(\text{synset}) = \text{average}(d)$$

* *where f is the normalized frequency of the 'yes' votes the image receives*

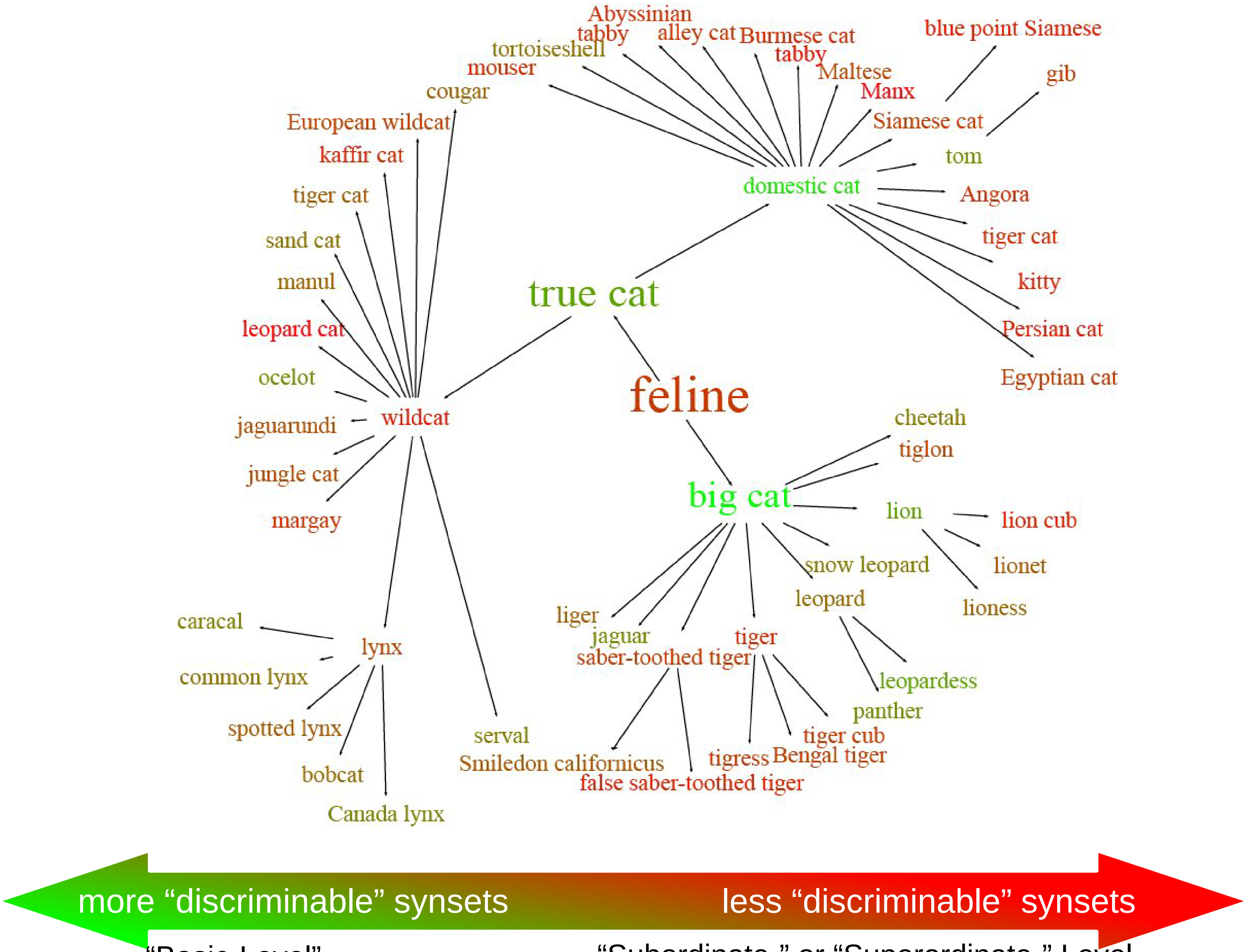
AMT worker / Image	1	2	3	4	5	6	7	8	9	10	d
	N	Y	N	Y	Y	Y	Y	Y	Y	Y	0.72
	N	N	N	N	N	N	N	N	N	N	0.00
	N	N	N	Y	Y	N	N	N	N	Y	0.88



more "discriminable" synsets

less "discriminable" synsets

Fei-Fei, Deng, Su, & Li, *VSS*, 2009



feline

true cat

domestic cat

Abyssinian

tabby

alley cat

Burmese cat

tabby

tortoiseshell

mouser

Maltese

Manx

blue point Siamese

Siamese cat

gib

tom

Angora

tiger cat

kitty

Persian cat

Egyptian cat

wildcat

cougar

European wildcat

kaffir cat

tiger cat

sand cat

manul

leopard cat

ocelot

jaguarundi

jungle cat

margay

lynx

caracal

common lynx

spotted lynx

bobcat

Canada lynx

serval

big cat

cheetah

tiglon

lion

lion cub

lionet

lioness

snow leopard

leopard

leopardess

panther

liger

jaguar

saber-toothed tiger

Smiledon californicus

false saber-toothed tiger

tiger

tigress

tiger cub

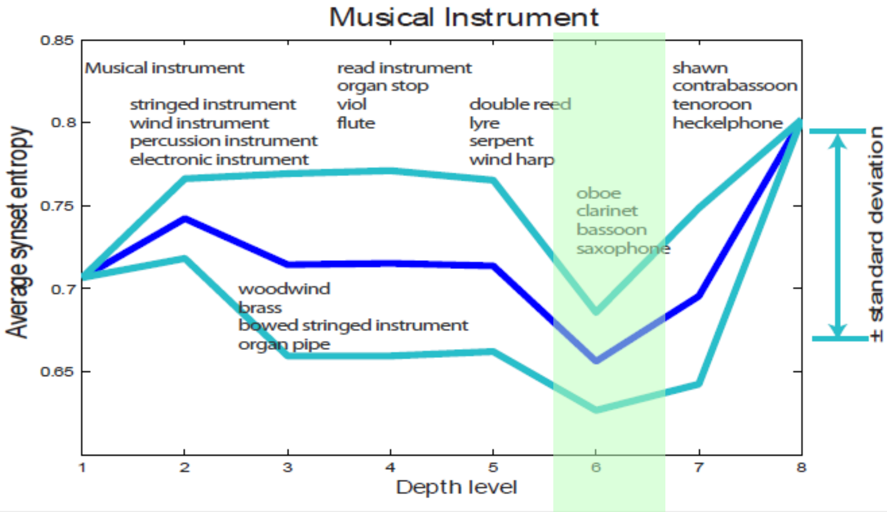
Bengal tiger

more “discriminable” synsets

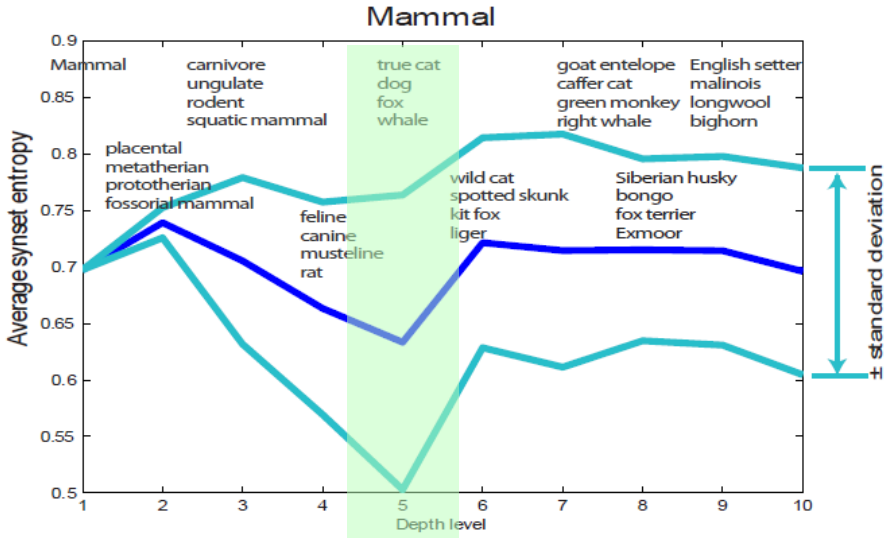
less “discriminable” synsets

“Basic-Level”

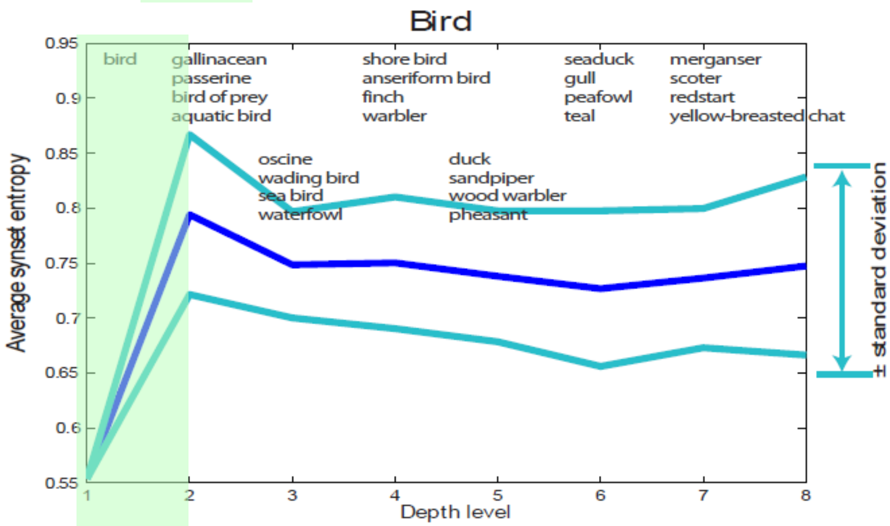
“Subordinate-” or “Superordinate-” Level



“Basic-Level”



“Basic-Level”



“Basic-Level”

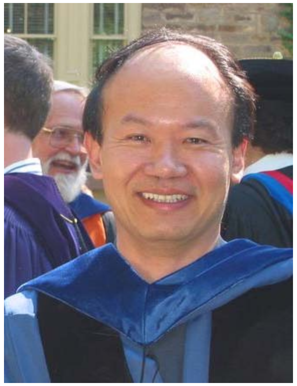
Fei-Fei, Deng, Su, & Li, *VSS*, 2009

Summary

- ImageNet is intended to serve as
  - A dataset
  - A knowledge ontology
- Construction of large-scale image dataset is a new research area
  - Crowdsourcing might be the future of many such tasks
- Benchmarking: what does classifying 10k+ image categories tell us?
  - Computation matters
  - Size matters
  - Density matters
  - Hierarchy matters
- Human vision: Rosch revisited and quantified
  - Quantifying basic-, subordinate- and superordinate-level concepts
- In the horizon: ImageNet Spring 2010 release
  - The upcoming ImageNet Challenge (in partnership with PASCAL VOC)

Thank you!

co-PI

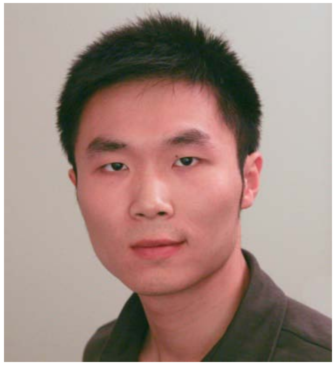


Kai Li
Princeton U.

Research collaborator;
ImageNet Challenge boss

Alex Berg
Columbia U.

Graduate students



Jia Deng
Princeton/Stanford U.

Hao Su
Stanford U.

Tomorrow 4pm:
Intelligence Seminar

Story Telling in Images:
modeling visual hierarchies
within and across images