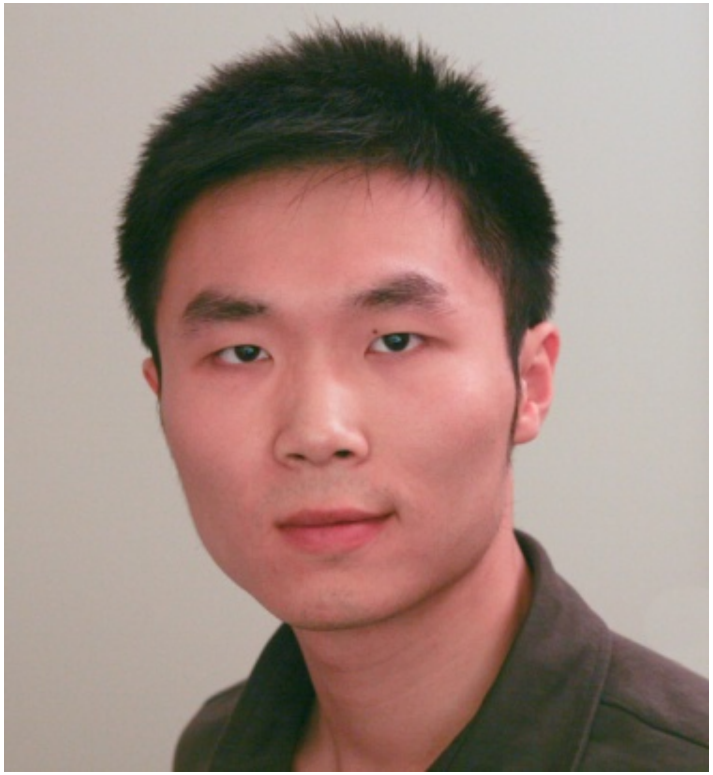
Large-Scale Visual Recognition Powered by Big Data and Big Crowd

Fei-Fei Li
Stanford University



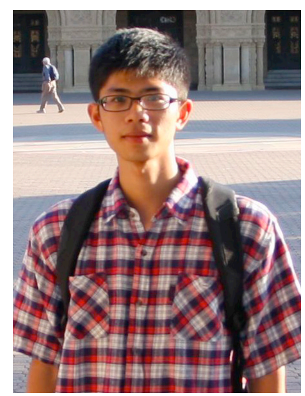
Dr. Jia Deng
Stanford U. -> U. Michigan

Prof. Kai Li
Princeton U.

Prof. Alex Berg
Stony Brook U.

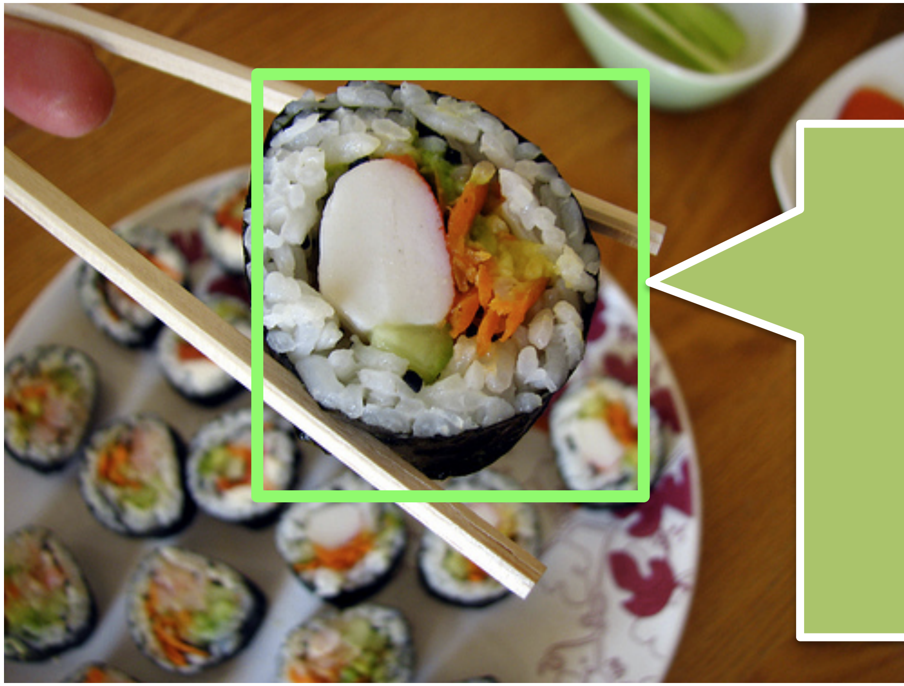
Sanjeev Satheesh
Stanford U.

Jonathan Krause
Stanford U.



Zhiheng Huang
Stanford U.

Olga Russakovsky
Stanford U.



California Roll

Ingredients: Rice, Seaweed, Crab, Cucumber, Avocado

Calories: 40

Fat: 7g

Carb: 40g

Protein: 5g

Gluten Free

Amanita phalloides

http://en.wikipedia.org/wiki/Amanita_phalloides

TOXIC. DO NOT EAT

Mountain Lion

DO NOT RUN

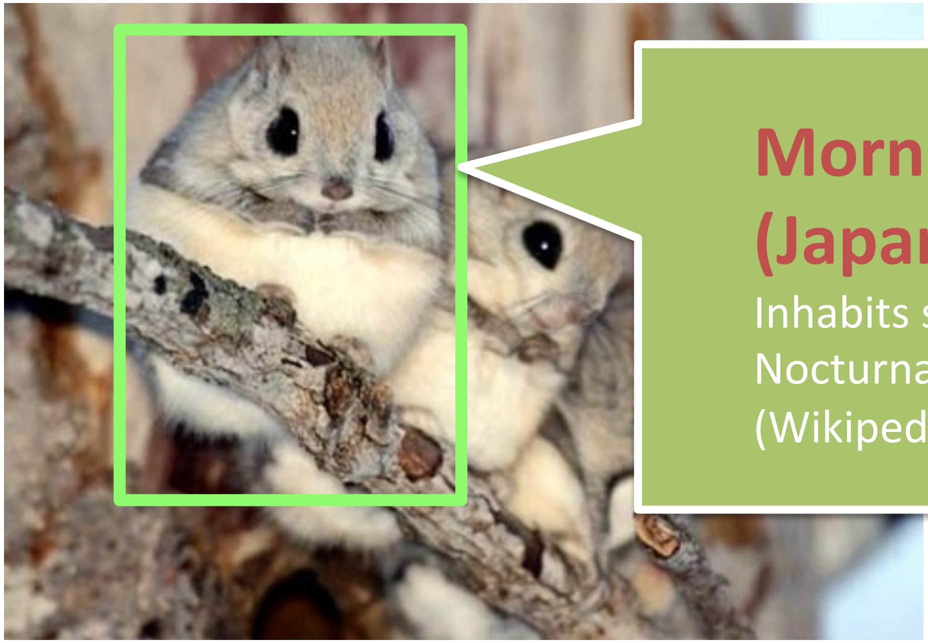
Raise arms to appear larger.

Show your teeth

IKEA POANG Chair

ON SALE

$29.00 at ikea.com



Mornonga
(Japanese flying squirrel)

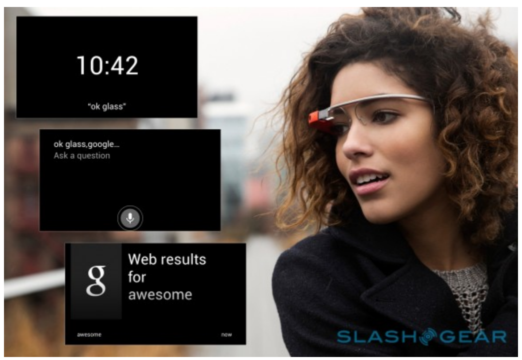
Inhabits sub-alpine forests in Japan.
Nocturnal. Eats seeds, fruit, tree leaves
(Wikipedia)

I wish my computer to recognize EVERYTHING

Surveillance

Robotics

Assistive tools



Wearable devices

Recognition Engine

Driverless cars

Smart photo album

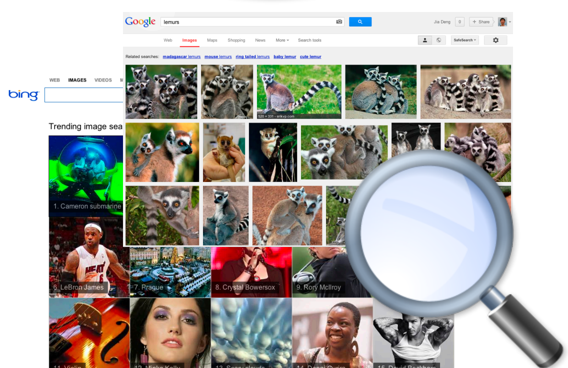


Image search

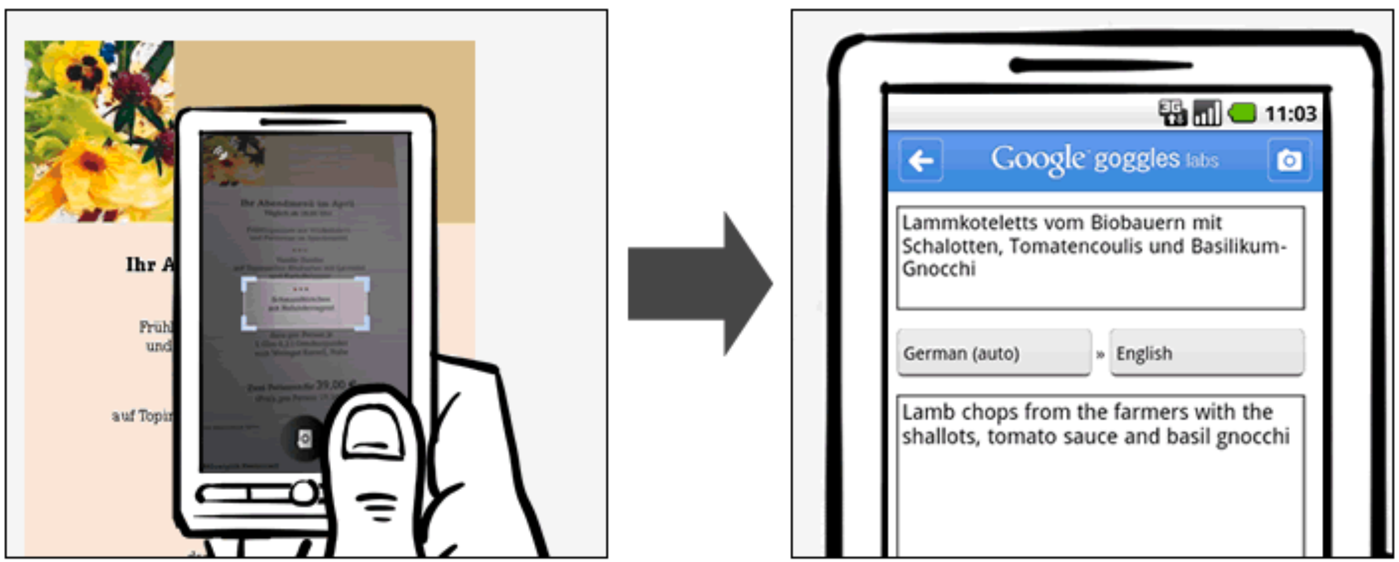
Mining social media

What can computers already recognize?


AUTO
10M
IN
39
Nikon
The Nikon 560. Detects up to 12 faces.

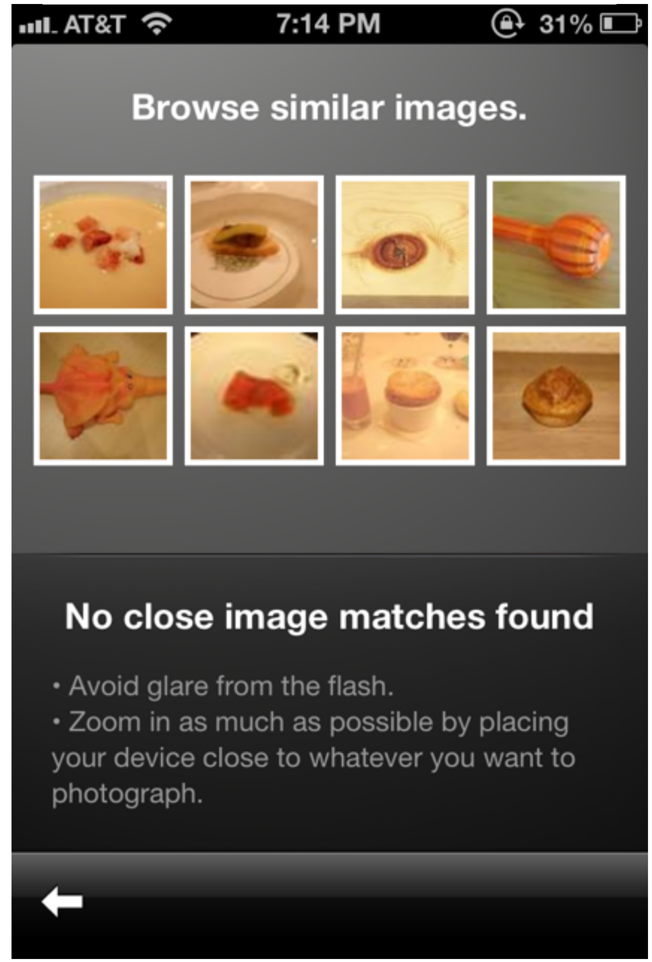
Google Goggles

Use pictures to search the web.



Google Goggles

Use pictures to search the web.



PASCAL VOC [Everingham et al. 2006-2012]

Airplane	**Dining table**
Bird	**Dog**
Boat	**Horse**
Bike	**Motorbike**
Bottle	**Person**
Bus	**Potted plant**
Car	**Sheep**
Cat	**Sofa**
Chair	**Train**
Cow	**TV monitor**

No Coffee Mugs!

What about Gas Pumps!

Many Things Do Not Work Yet…

How many things are there?

WordNet

10K+
[Biederman '87]

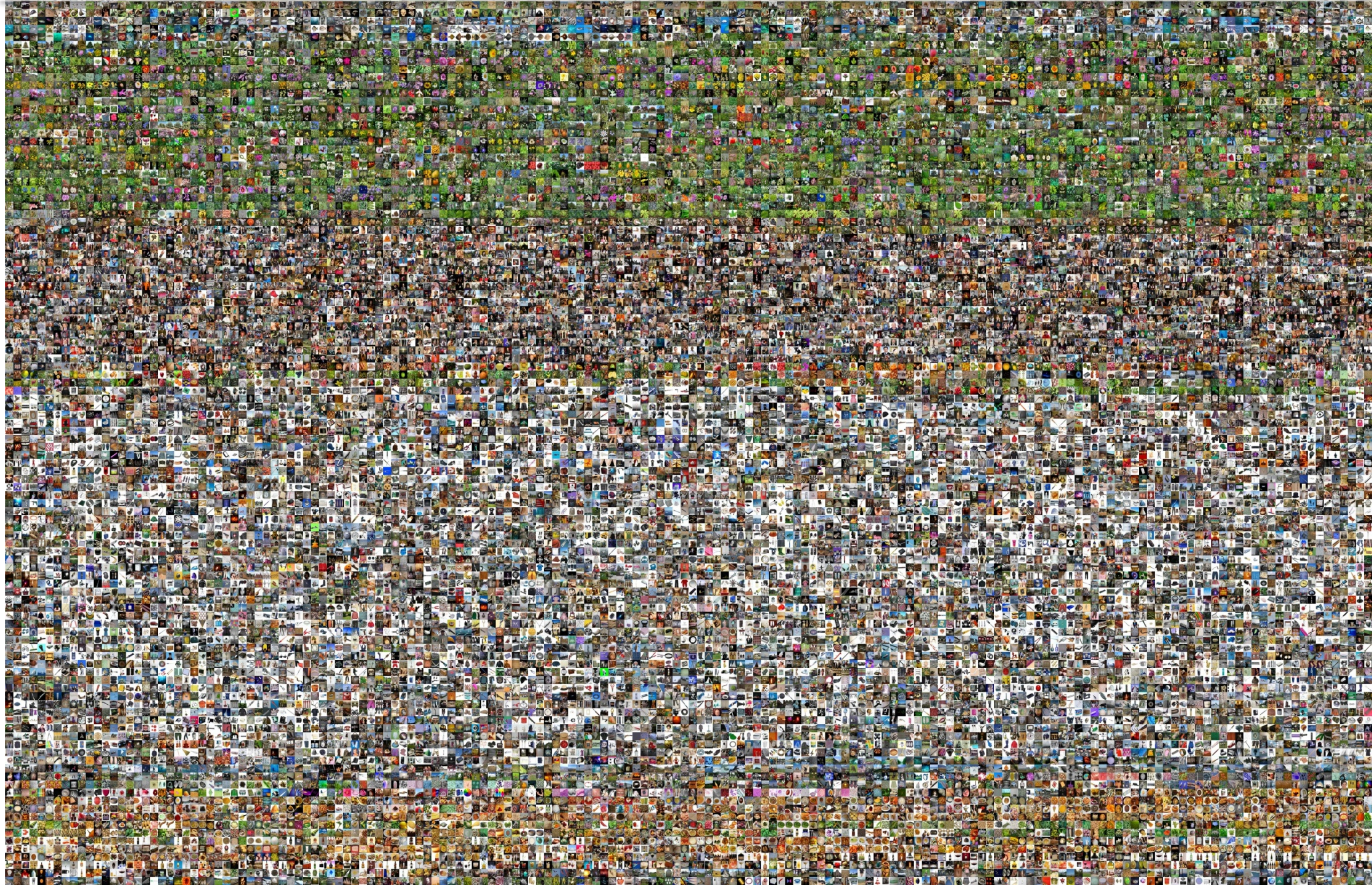
60K+
product
categories

80K+
English nouns
[Miller '95; Fellbaum '98]

3.5M+
unique tags
[Sigurbjörnsson & Zwol '08]

4.1M+
articles

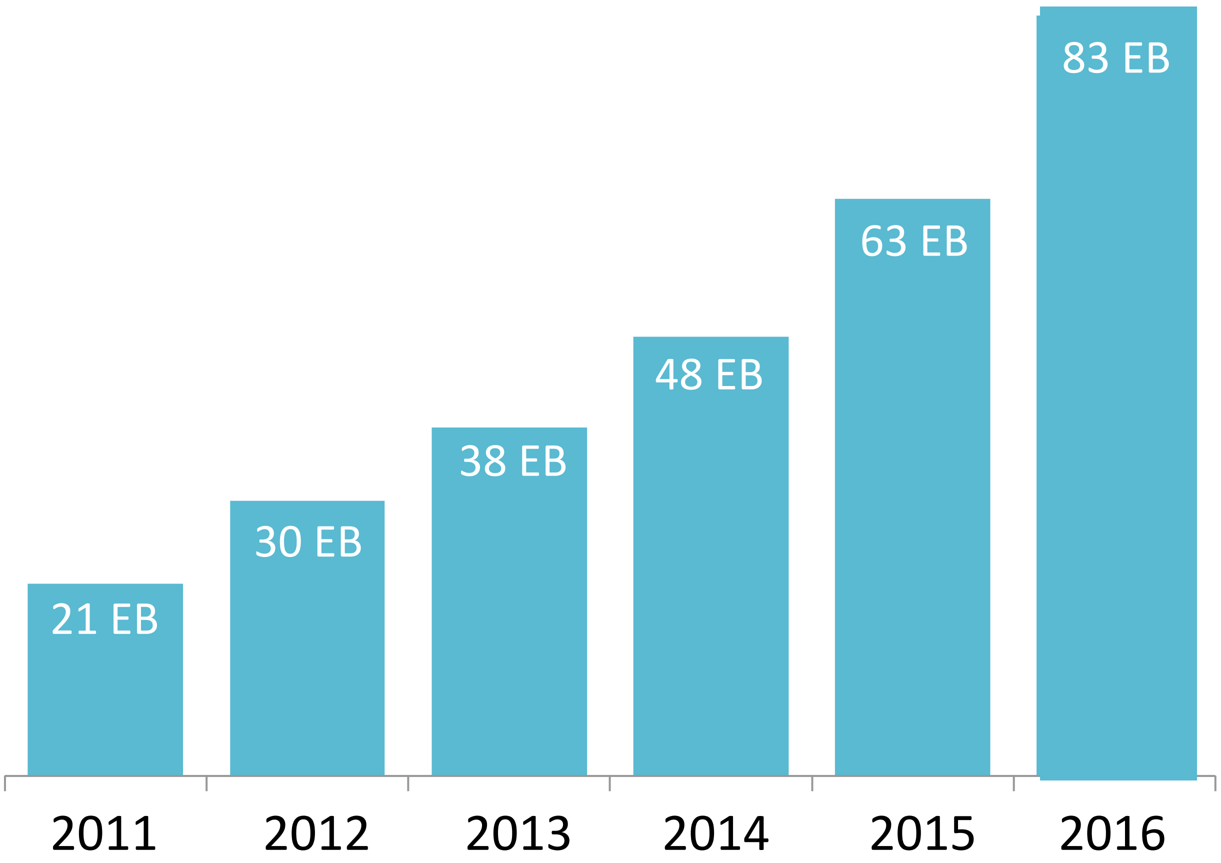
From 20 classes to Millions?




4 September 2008 | www.nature.com/nature | £10
THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE
nature
THE BITER BIT
Viral infections for viruses
TROPICAL CYCLONES
The strong get stronger
BLACK HOLE PHYSICS
A new window on the Galactic Centre
BIG DATA
NATUREJOBS
Minnesota musings
SCIENCE IN THE PETABYTE ERA
nature 455, 1-136 4 September 2008
www.nature.com/nature

Big Data from the Internet

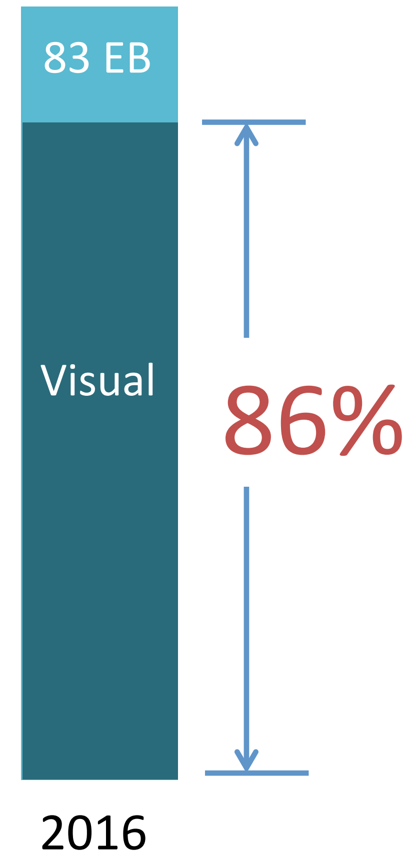
Global Consumer Internet Traffic Per Month



Source: Cisco

72 hours of videos / min

300 million images / day



Source: Cisco

Big Data from the Internet

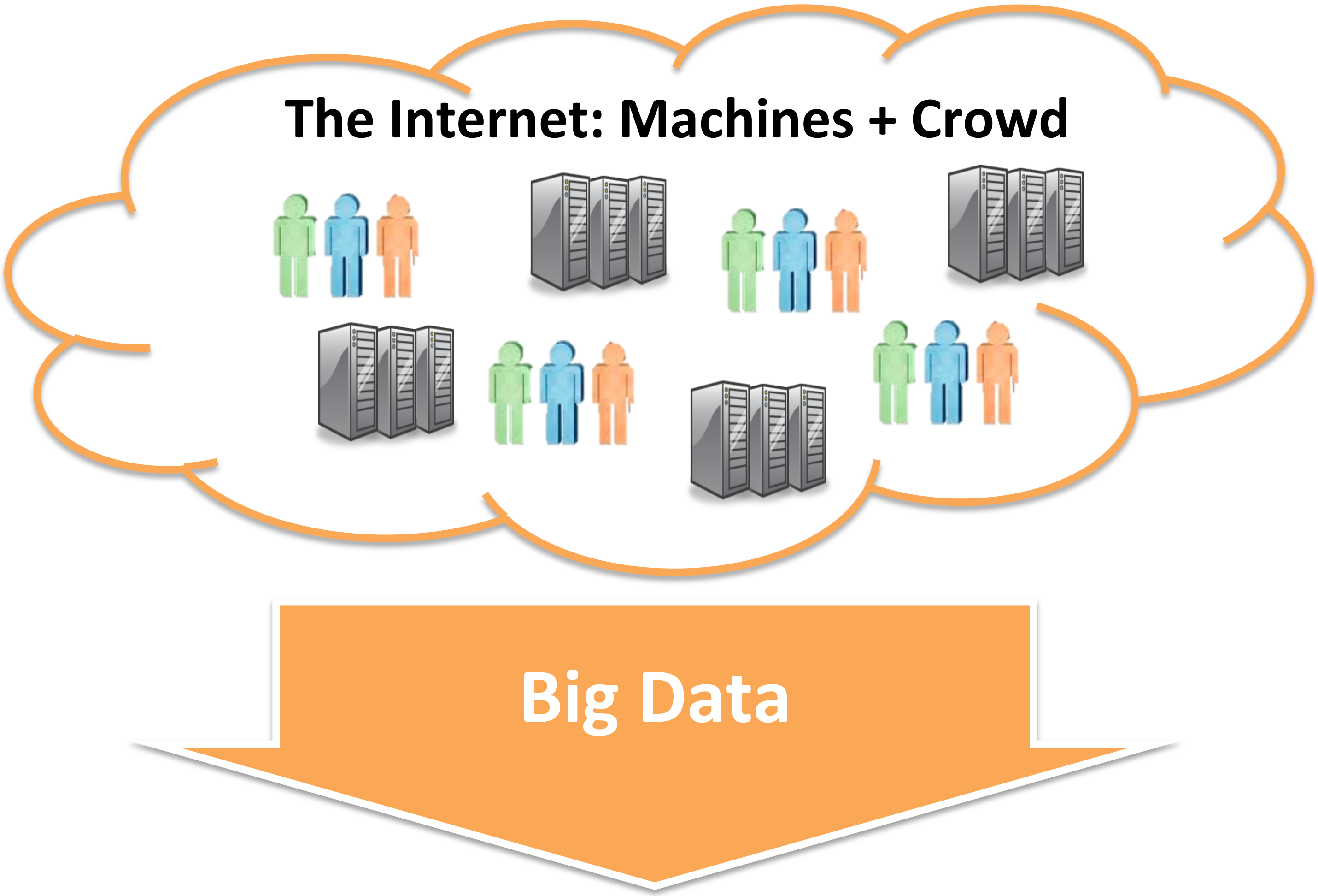
→ The Internet can teach EVERYTHING

Google pitbullfrog

Evolution Gone Wild

Future plants and animals

http://www.worth1000.com/contests/12705/contest

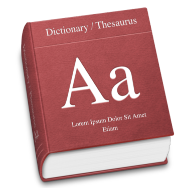


Teach machines to recognize EVERYTHING

PASCAL VOC

20

[Everingham et al.'06-'12]



10K+

[Biederman '87]

Goal: Build a recognition engine on 10K classes

And Never Make A Mistake!

EVA

Engine for Visual Annotation

The EVA system, powered by **ImageNet**, can annotate images with guaranteed accuracies. It currently recognizes over **10,000** visual categories. See the project page to find out more.

Paste a URL | Upload an image

ANNOTATE

Agenda

How to build a large-scale recognition engine using big data

STEP 1: ?

STEP 2: ?

STEP 3: ?

Agenda

How to build a large-scale recognition engine using big data

STEP 1: Build a Large Knowledge Base

STEP 2: ?

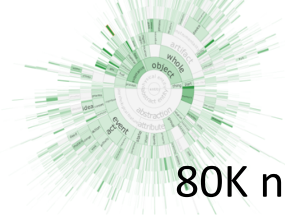
STEP 3: ?

Get a list of everything

↓

Crawl the web

WordNet



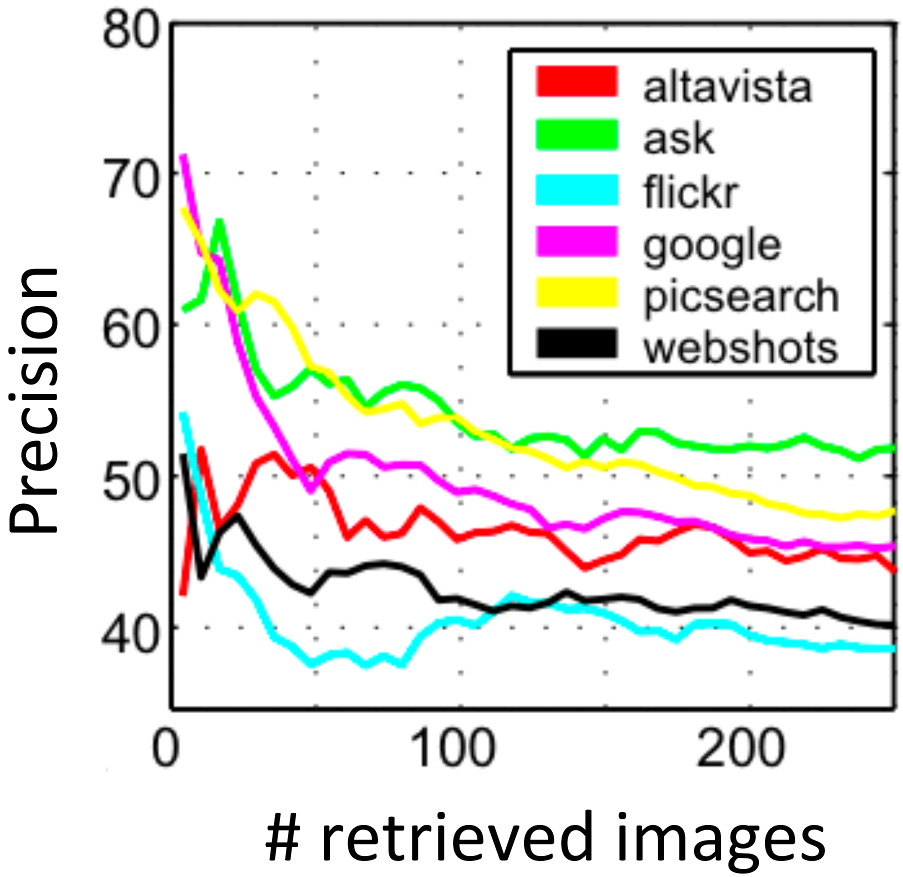
80K nouns

- Expert constructed
- Rich structure
  - Taxonomy, Partonomy
- Widely used

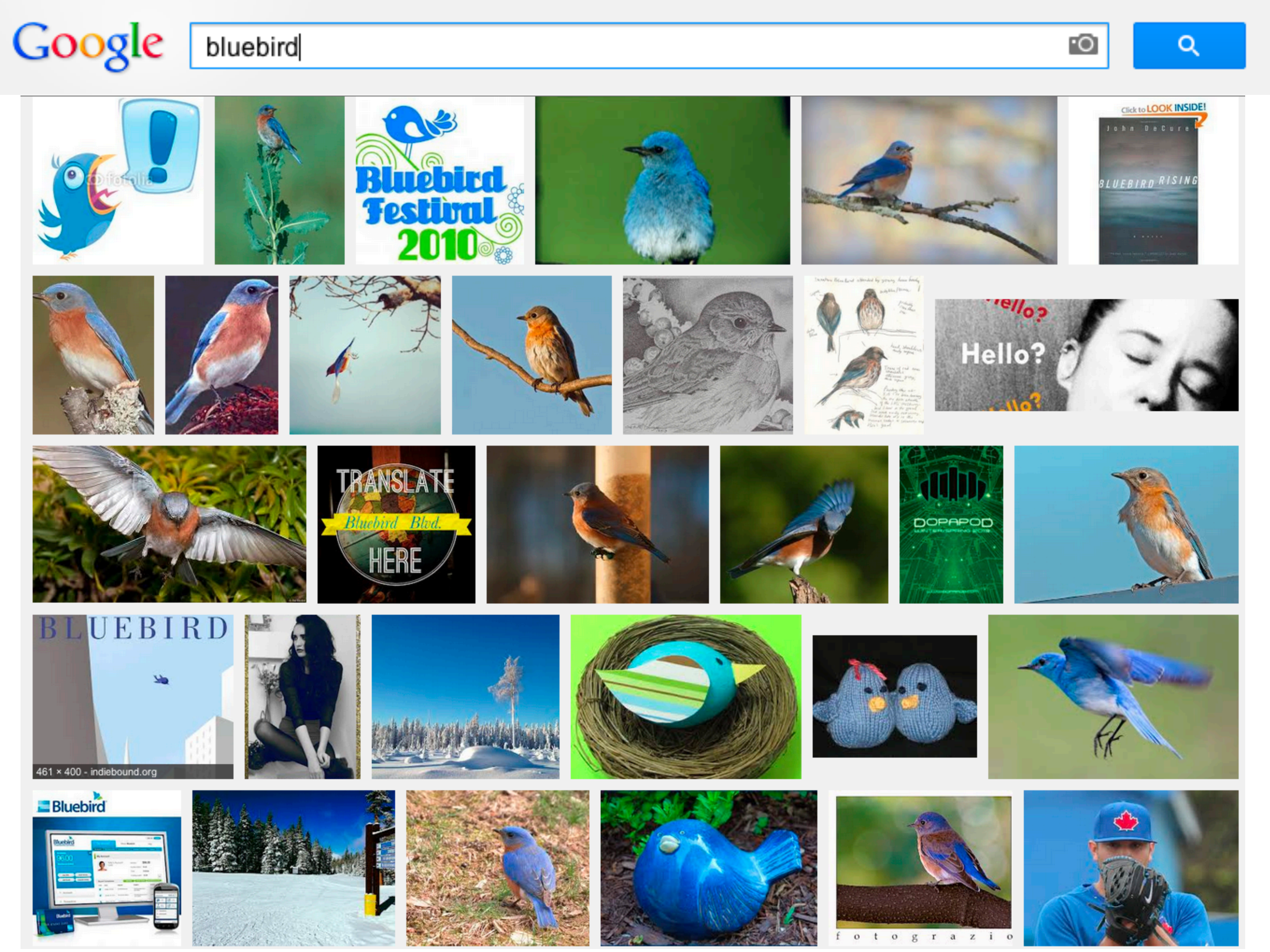
Google flickr

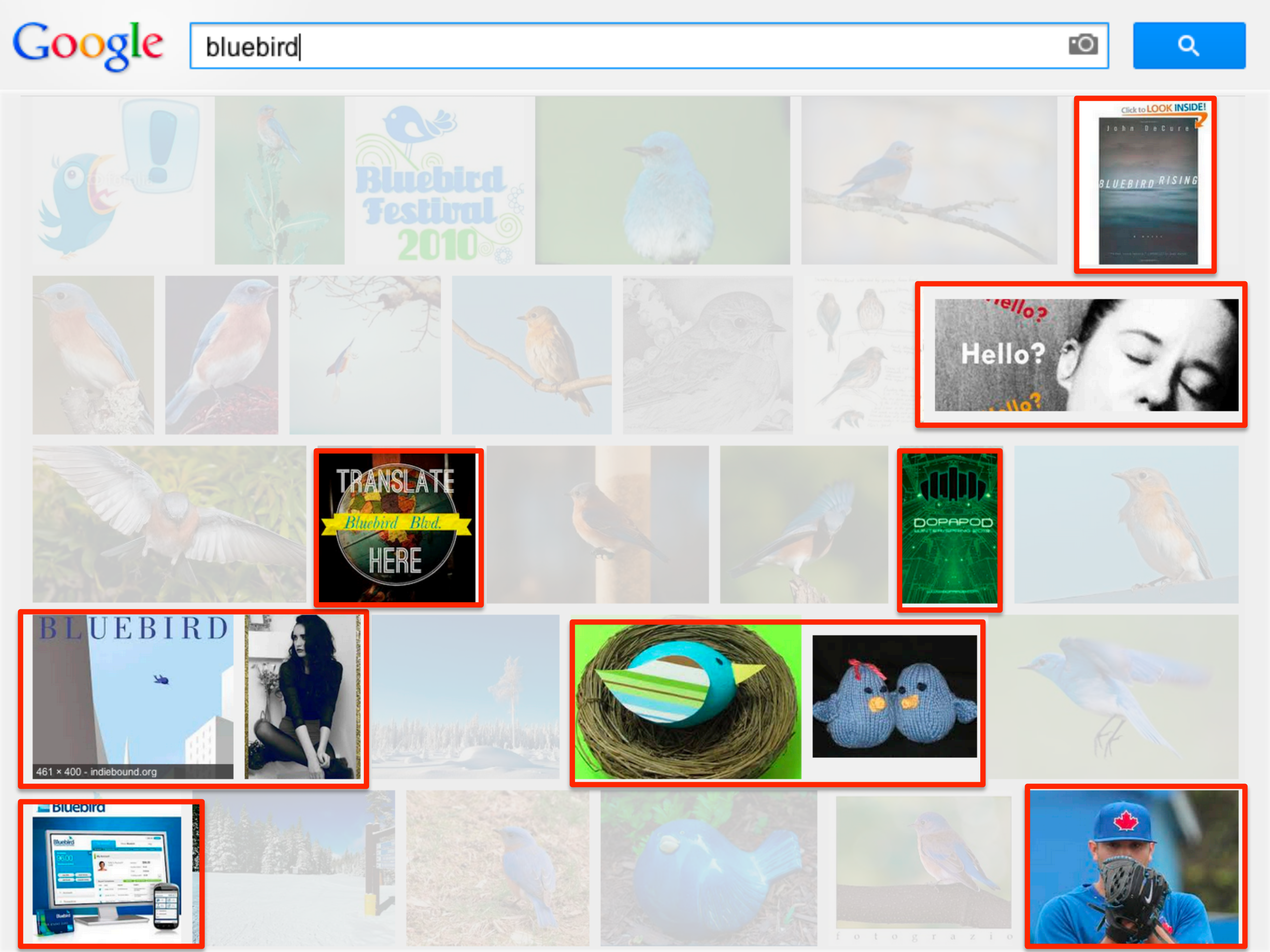
[Torralba, Fergus, Freeman ’08]
[Yao, Yang, Zhu ’07]
[Everingham et al ’06]
[Russell et al ’05]
[Griffin & Perona ’03]
[Fei-Fei, Fergus, Perona ’03]

What you get may not be what you want.



[Torralba, Fergus, Freeman ’08]


Google
bluebird
fotolia
Bluebird Festival 2010
Click to LOOK INSIDE!
John DeCure
BLUEBIRD RISING
Hello?
TRANSLATE
Bluebird Blvd.
HERE
DOPAPOD
BLUEBIRD
461 × 400 - indiebound.org
Bluebird
f o t o g r a z i o


Google
bluebird
Bluebird Festival 2010
Click to LOOK INSIDE!
John DeCure
BLUEBIRD RISING
Hello?
TRANSLATE
Bluebird Blvd.
HERE
DOPAPOD
BLUEBIRD
461 × 400 - indiebound.org
Bluebird

Get a list of everything

↓

Crawl the web

↓

Clean up

80K nouns

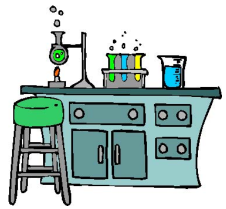
- Expert constructed
- Rich structure
  - Taxonomy, Partonomy
- Senses disambiguated
- Widely used

Google flickr

[Torralba, Fergus, Freeman ’08]
[Yao, Yang, Zhu ’07]
[Everingham et al ’06]
[Russell et al ’05]
[Griffin & Perona ’03]
[Fei-Fei, Fergus, Perona ’03]

<u>Graduate Students</u>

Good at complex tasks



Good quality

Very few of them

High cost

Estimate: 20 Years, $2M+

Graduate Students

Good at complex tasks

Good quality

Very few of them

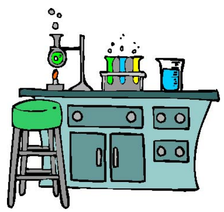
High cost

The Crowd

amazonmechanical turk
beta
Artificial Artificial Intelligence

Graduate Students

Good at complex tasks



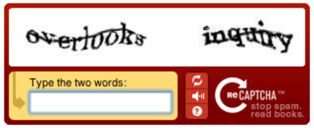
Good quality

Very few of them

High cost

The Crowd

Good at simple tasks

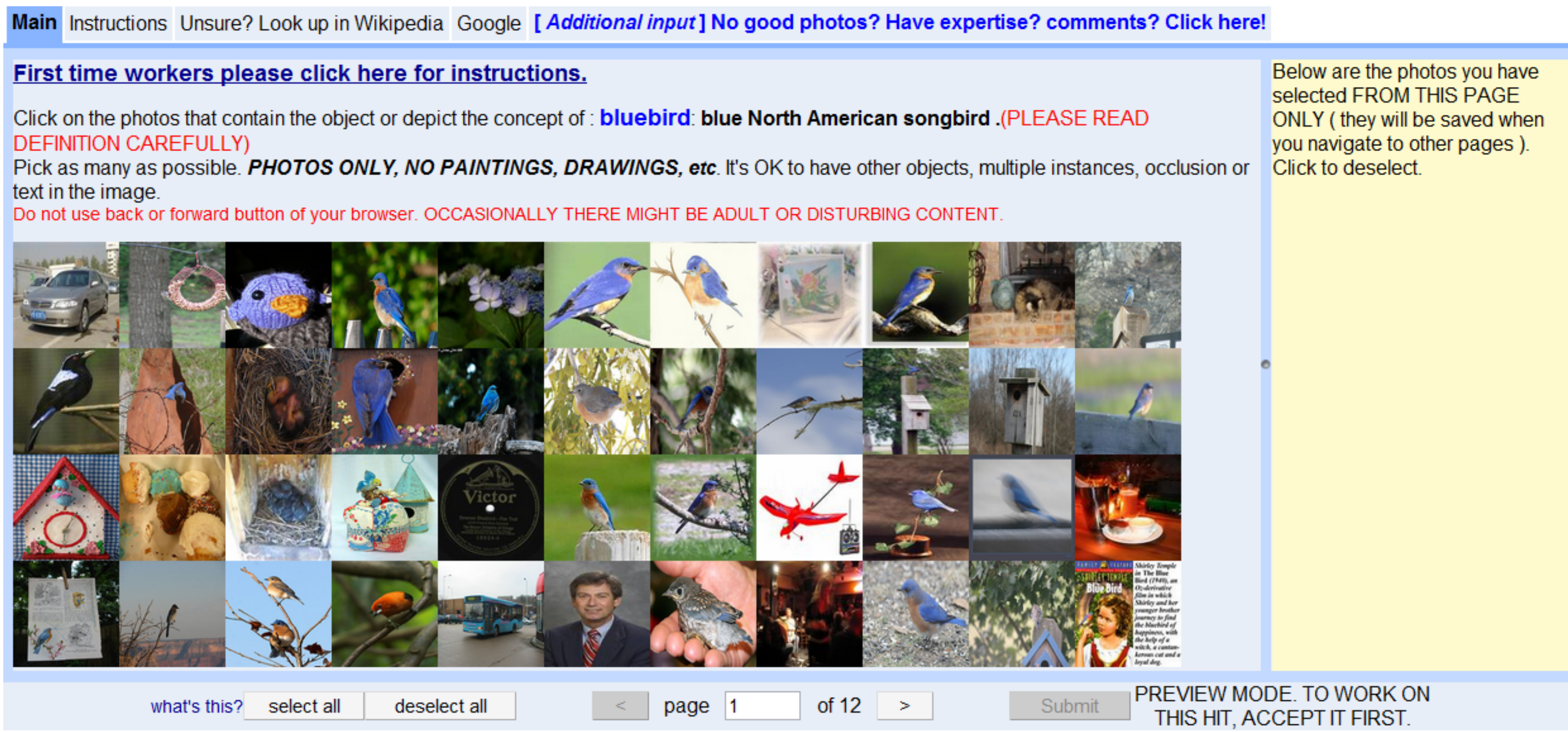


Mixed quality

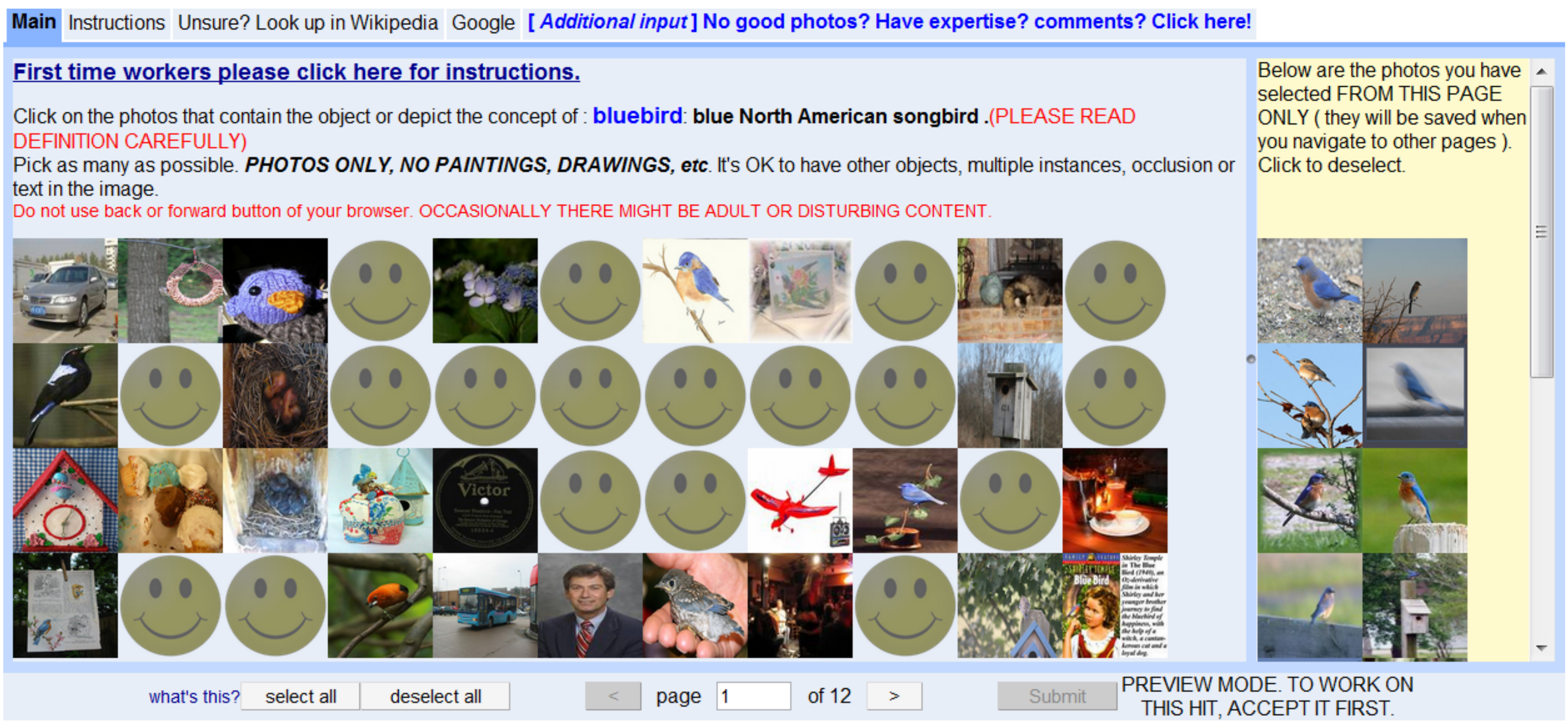
Many of them

Low cost

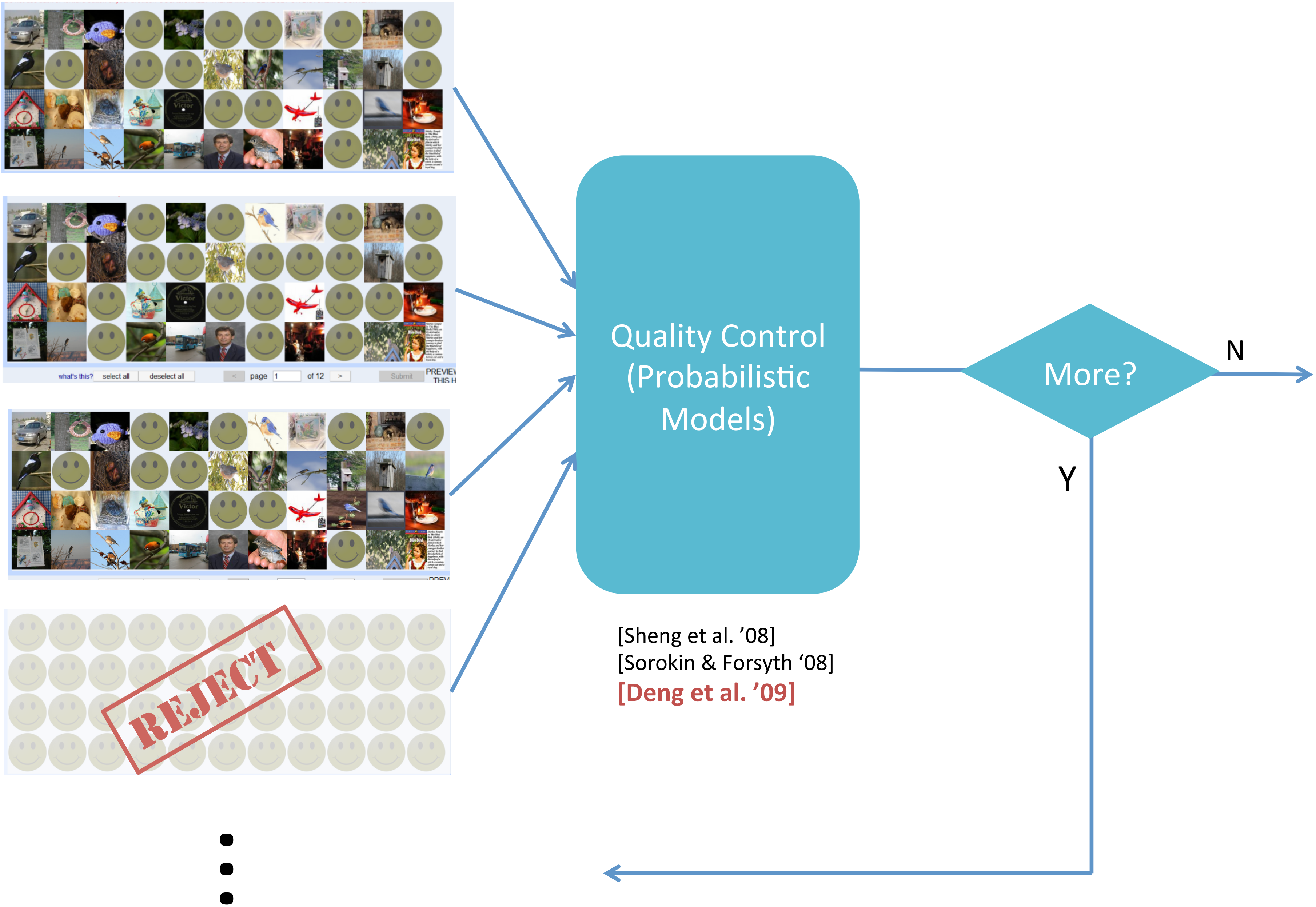
Cleaning up by AMT workers

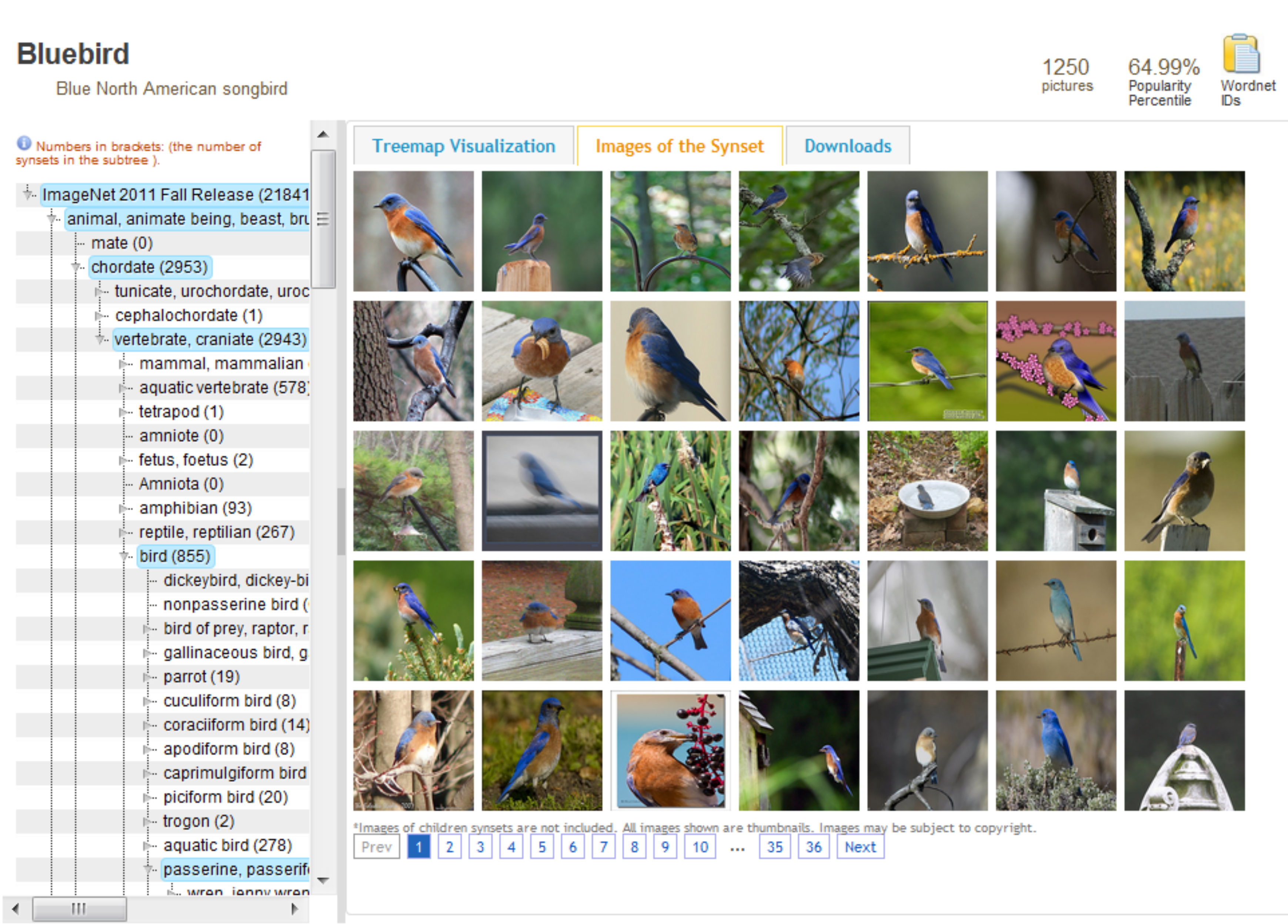


Cleaning up by AMT workers



Dealing with Mixed Quality





Bluebird
Blue North American songbird
1250 pictures
64.99% Popularity Percentile
Wordnet IDs
Numbers in brackets: (the number of synsets in the subtree).
ImageNet 2011 Fall Release (21841
animal, animate being, beast, bru
mate (0)
chordate (2953)
tunicate, urochordate, uroc
cephalochordate (1)
vertebrate, craniate (2943)
mammal, mammalian
aquatic vertebrate (578
tetrapod (1)
amniote (0)
fetus, foetus (2)
Amniota (0)
amphibian (93)
reptile, reptilian (267)
bird (855)
dickeybird, dickey-bi
nonpasserine bird (
bird of prey, raptor, r
gallinaceous bird, g
parrot (19)
cuculiform bird (8)
coraciiform bird (14)
apodiform bird (8)
caprimulgiform bird
piciform bird (20)
trogon (2)
aquatic bird (278)
passerine, passerifi
Treemap Visualization
Images of the Synset
Downloads
*Images of children synsets are not included. All images shown are thumbnails. Images may be subject to copyright.
Prev 1 2 3 4 5 6 7 8 9 10 ... 35 36 Next

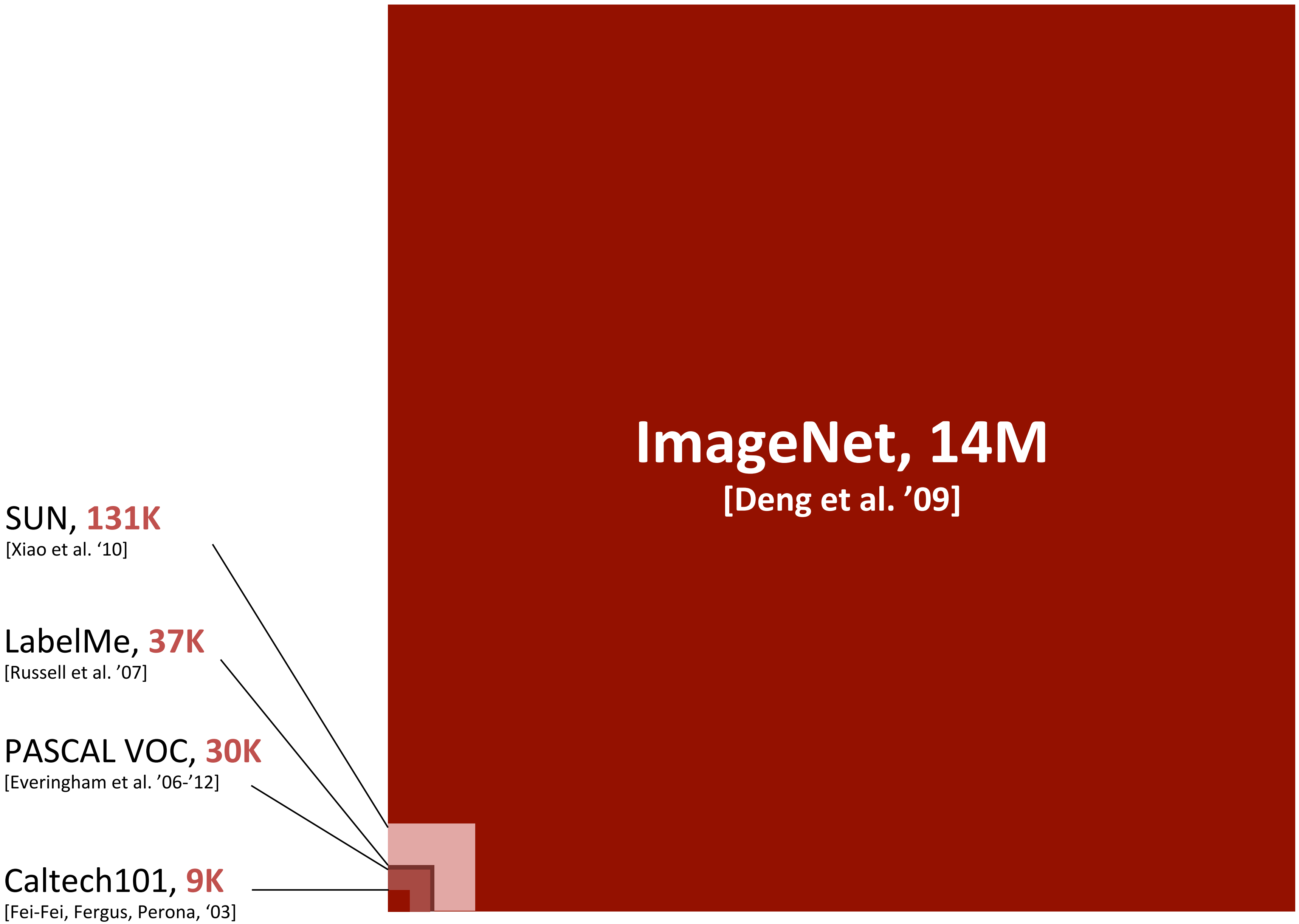
IMAGENET [Deng et al. 2009]

www.image-net.org

22K categories and **14M** images

- Animals
  - Bird
  - Fish
  - Mammal
  - Invertebrate
- Plants
  - Tree
  - Flower
- Food
- Materials
- Structures
- Artifact
  - Tools
  - Appliances
  - Structures
- Person
- Scenes
  - Indoor
  - Geological Formations
- Sport Activities

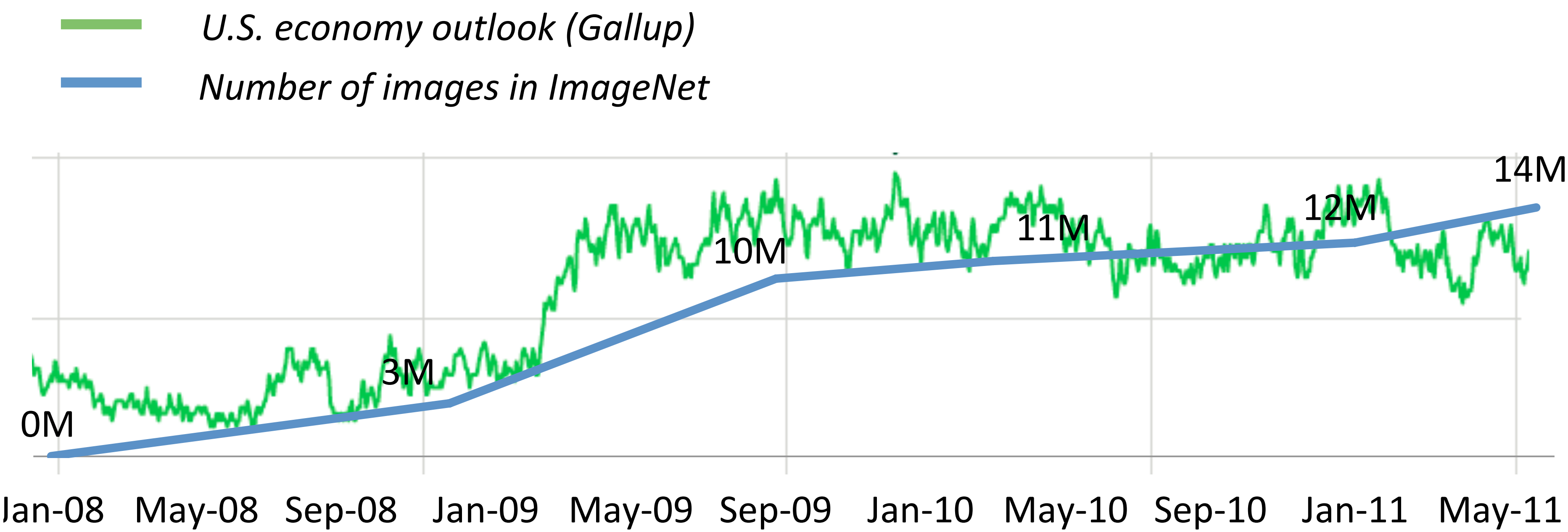
Number of Labeled Images



hired **50K+** AMT workers

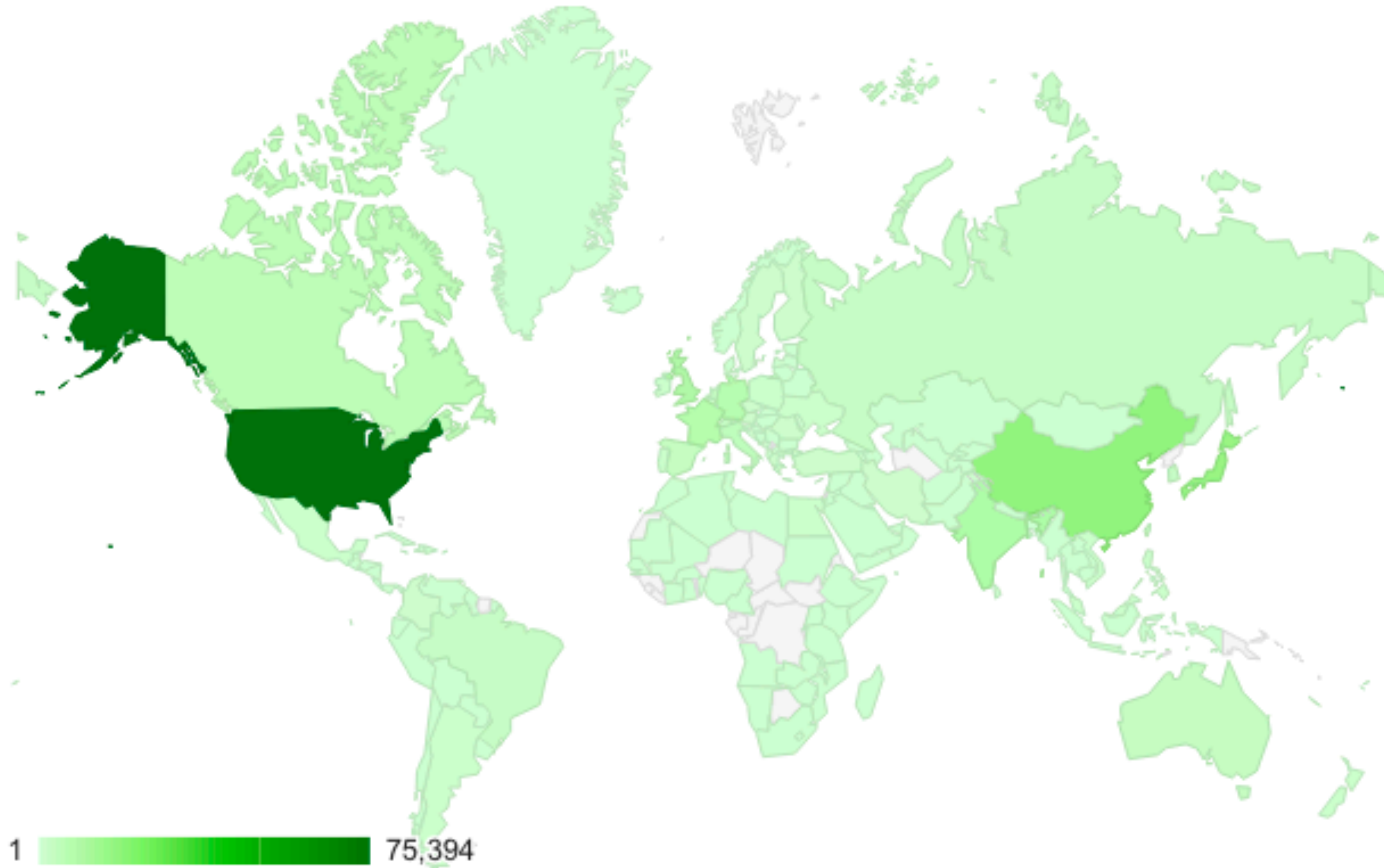
who looked at **160M+** images

and made **550M+** binary decisions



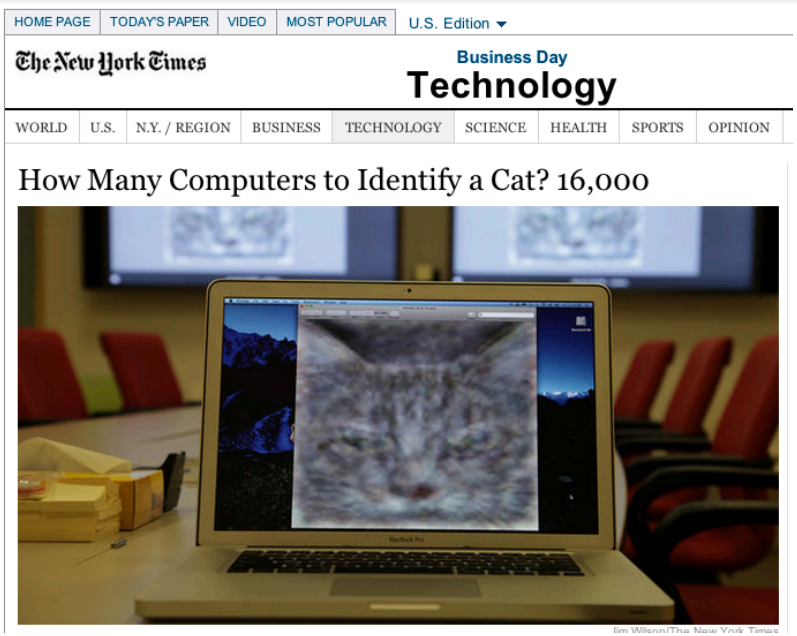
Research Impact of ImageNet

3000+ registered users, visits from 175 countries



ECCV 2012
Best paper Award

Kuettel, Guillaumin, Ferrari. **Segmentation Propagation in ImageNet.** ECCV 2012



HOME PAGE | TODAY'S PAPER | VIDEO | MOST POPULAR | U.S. Edition

The New York Times

Business Day

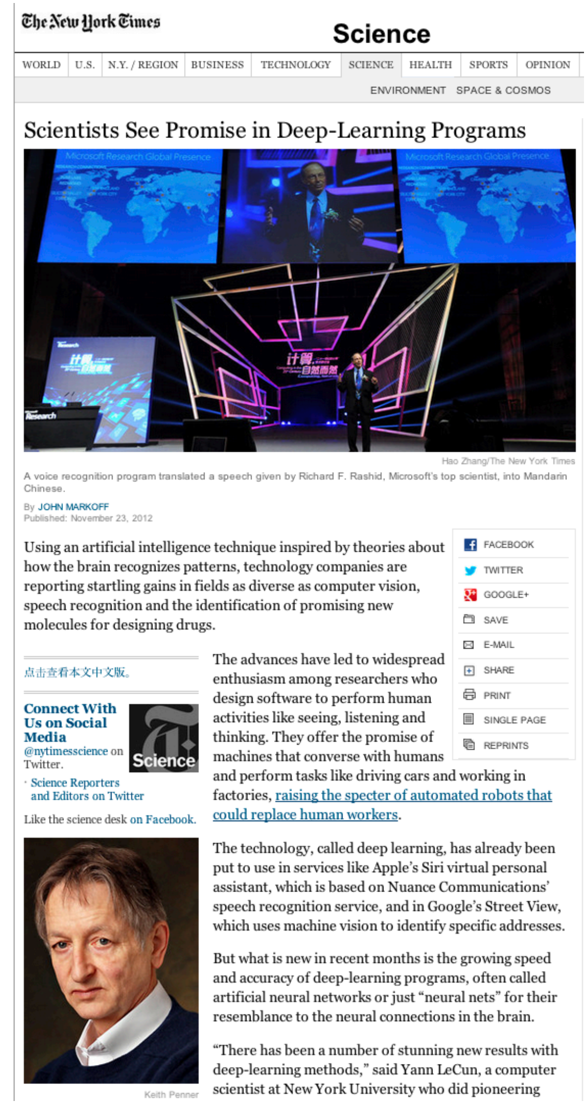
Technology

WORLD | U.S. | N.Y. / REGION | BUSINESS | TECHNOLOGY | SCIENCE | HEALTH | SPORTS | OPINION

How Many Computers to Identify a Cat? 16,000

Jim Wilson/The New York Times

Le et al. **Building high-level features using large scale unsupervised learning**. ICML 2012.



The New York Times

Science

WORLD | U.S. | N.Y. / REGION | BUSINESS | TECHNOLOGY | SCIENCE | HEALTH | SPORTS | OPINION

ENVIRONMENT SPACE & COSMOS

Scientists See Promise in Deep-Learning Programs

Hao Zhang/The New York Times

A voice recognition program translated a speech given by Richard F. Rashid, Microsoft's top scientist, into Mandarin Chinese.

By JOHN MARKOFF
Published: November 23, 2012

FACEBOOK
TWITTER
GOOGLE+
SAVE
E-MAIL
SHARE
PRINT
SINGLE PAGE
REPRINTS

Using an artificial intelligence technique inspired by theories about how the brain recognizes patterns, technology companies are reporting startling gains in fields as diverse as computer vision, speech recognition and the identification of promising new molecules for designing drugs.

点击查看本文中文版。

Connect With Us on Social Media
@nytimesscience on Twitter.
· Science Reporters and Editors on Twitter
Like the science desk on Facebook.

The advances have led to widespread enthusiasm among researchers who design software to perform human activities like seeing, listening and thinking. They offer the promise of machines that converse with humans and perform tasks like driving cars and working in factories, raising the specter of automated robots that could replace human workers.

The technology, called deep learning, has already been put to use in services like Apple's Siri virtual personal assistant, which is based on Nuance Communications' speech recognition service, and in Google's Street View, which uses machine vision to identify specific addresses.

But what is new in recent months is the growing speed and accuracy of deep-learning programs, often called artificial neural networks or just "neural nets" for their resemblance to the neural connections in the brain.

"There has been a number of stunning new results with deep-learning methods," said Yann LeCun, a computer scientist at New York University who did pioneering

Keith Penner

Krizhevsky, Sutskever, Hinton. **ImageNet classification with deep convolutional neural networks.** NIPS 2012

Seeking a Better Way to Find Web Images

By JOHN MARKOFF
Published: November 19, 2012

STANFORD, Calif. — You may think you can find almost anything on the Internet.

But even as images and video rapidly come to dominate the Web, search engines can ordinarily find a given image only if the text entered by a searcher matches the text with which it was labeled. And the labels can be unreliable, unhelpful ("fuzzy" instead of "rabbit") or simply nonexistent.

To eliminate those limits, scientists will need to create a new generation of visual search technologies — or else, as the Stanford computer scientist Fei-Fei Li recently put it, the Web will be in danger of "going dark."

Now, along with computer scientists from Princeton, Dr. Li, 36, has built the world's largest visual database in an effort to mimic the human vision system. With more than 14 million labeled objects, from obsidian to orangutans to ocelots, the database has because a vital resource for computer vision researchers.

The labels were created by humans. But now machines can learn from the vast database to recognize similar, unlabeled objects, making possible a striking increase in recognition accuracy.

This summer, for example, two Google computer scientists, Andrew Y. Ng and Jeff Dean, tested the new system, known as ImageNet, on a huge collection of labeled photos.

Agenda

How to build a large-scale recognition engine using big data

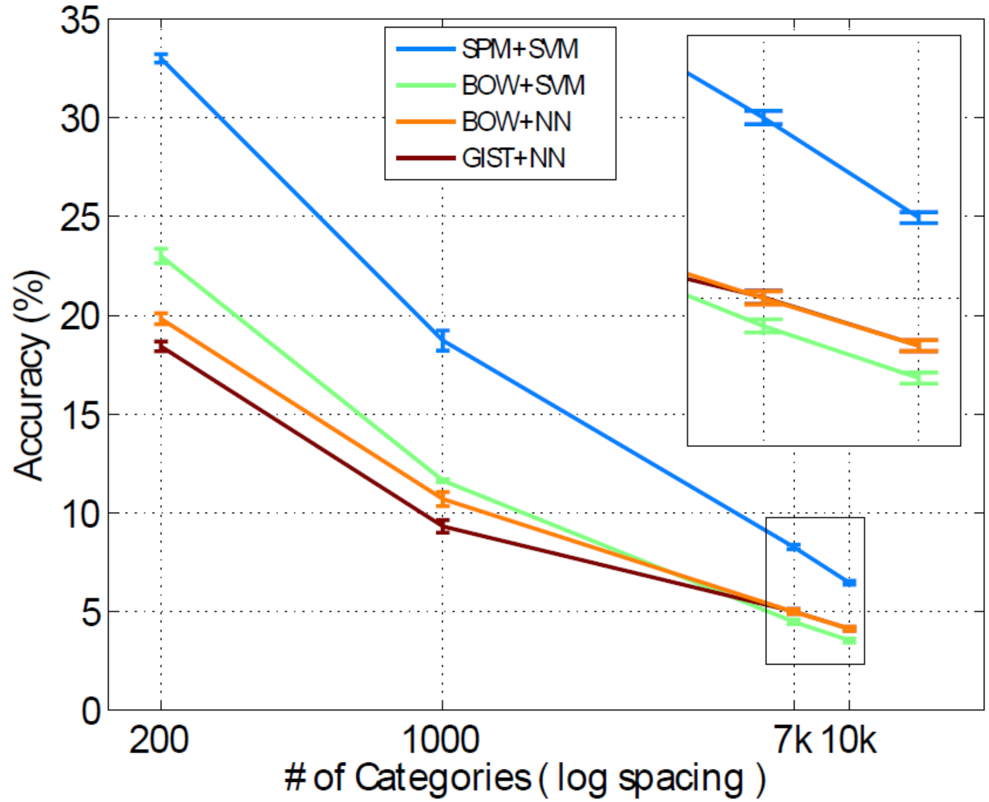
STEP 1: **Build a Large Knowledge Base (ImageNet)**

STEP 2: **?**

STEP 3: **?**

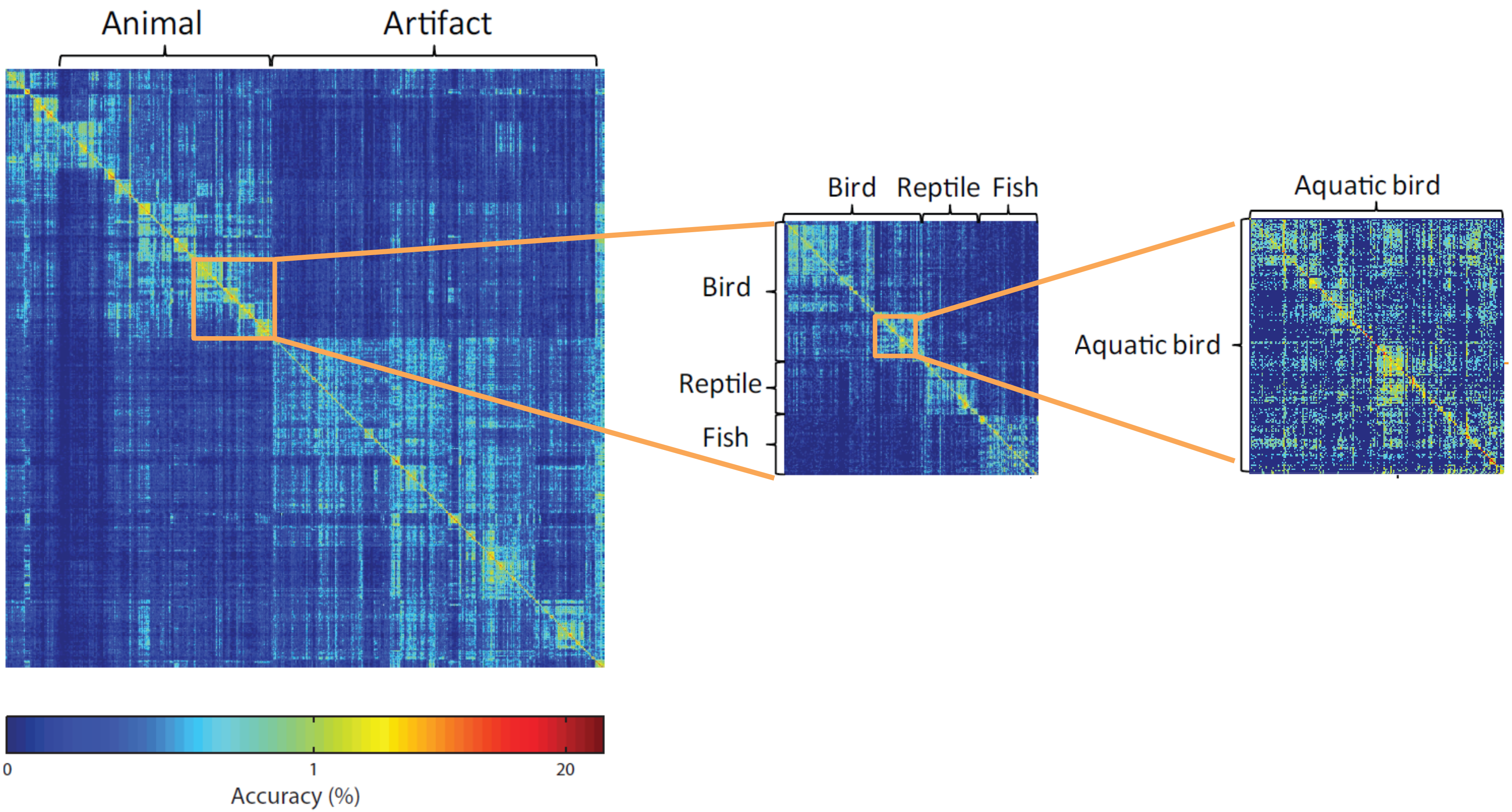
Learn to Classify 10K Classes

- 9 Million images
- 4 methods
  - SPM+SVM [Lazebnik et al. ’06]
  - BOW+SVM [Csurka et al. ’04]
  - BOW+NN
  - GIST+NN [Oliva et al. ’01]
- 6.4% for 10K categories



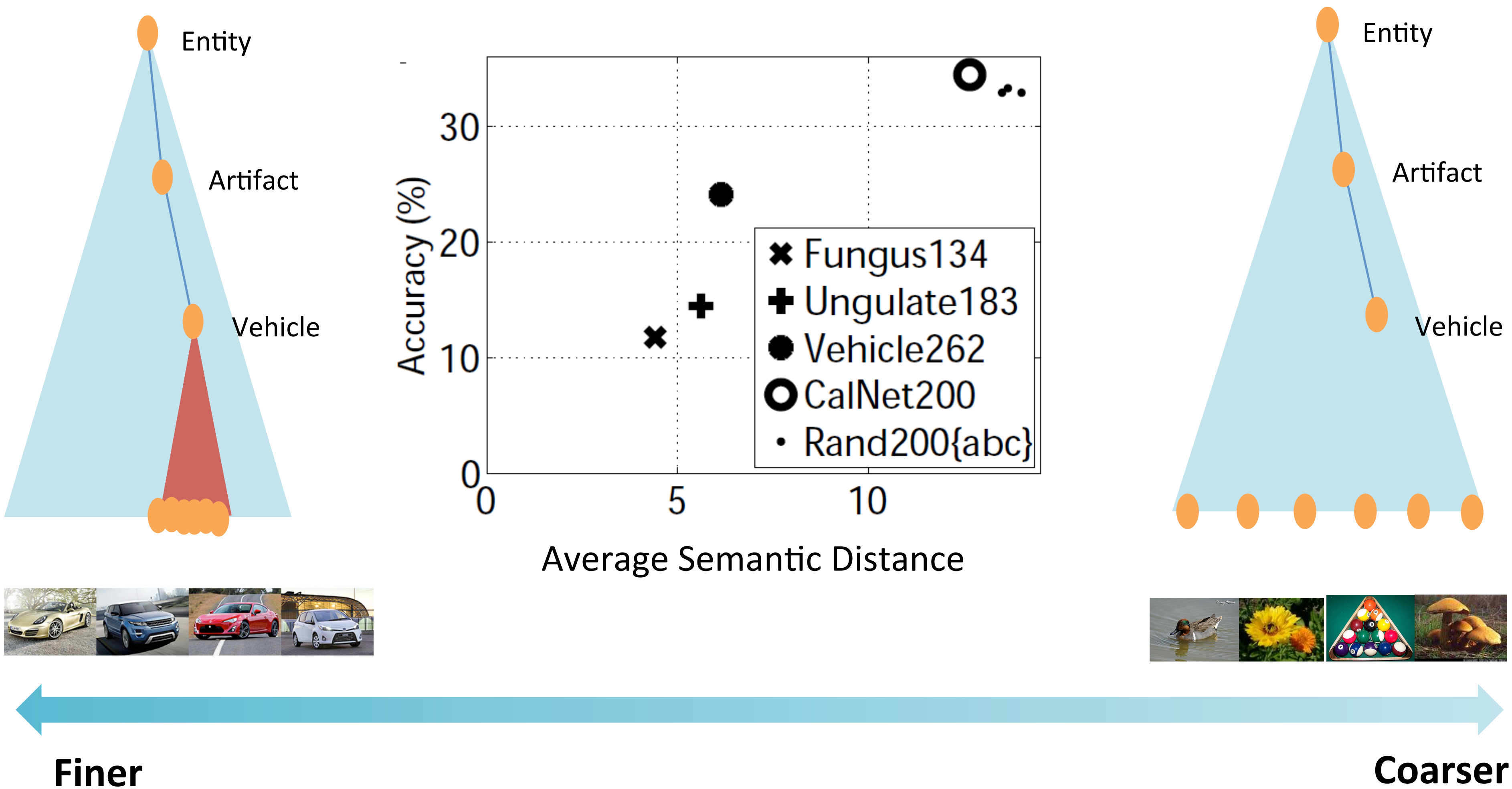
Deng, Berg, Li, & Fei-Fei, *ECCV2010*

Learn to Classify 10K Classes



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

Fine-grained categories are a lot harder



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

Agenda

How to build a large-scale recognition engine using big data

STEP 1: **Build a Large Knowledge Base (ImageNet)**

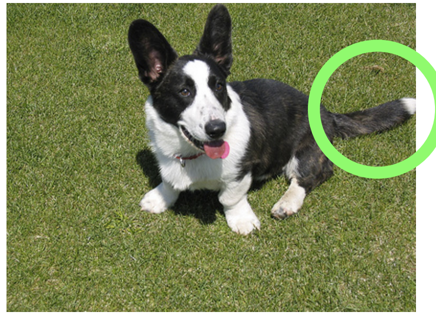
STEP 2: **Fine-Grained Recognition**

STEP 3: **?**

Why is Fine-Grained Recognition Difficult?

What breed is this dog?

Why is Fine-Grained Recognition Difficult?

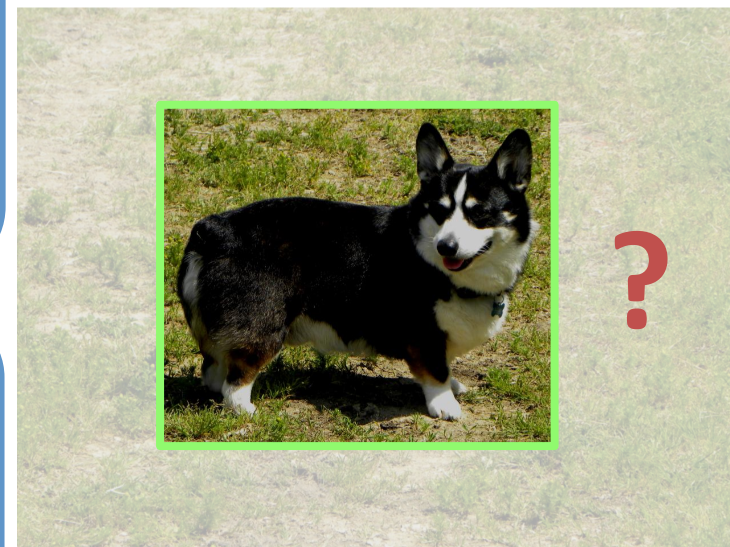



...

Cardigan Welsh Corgi


...

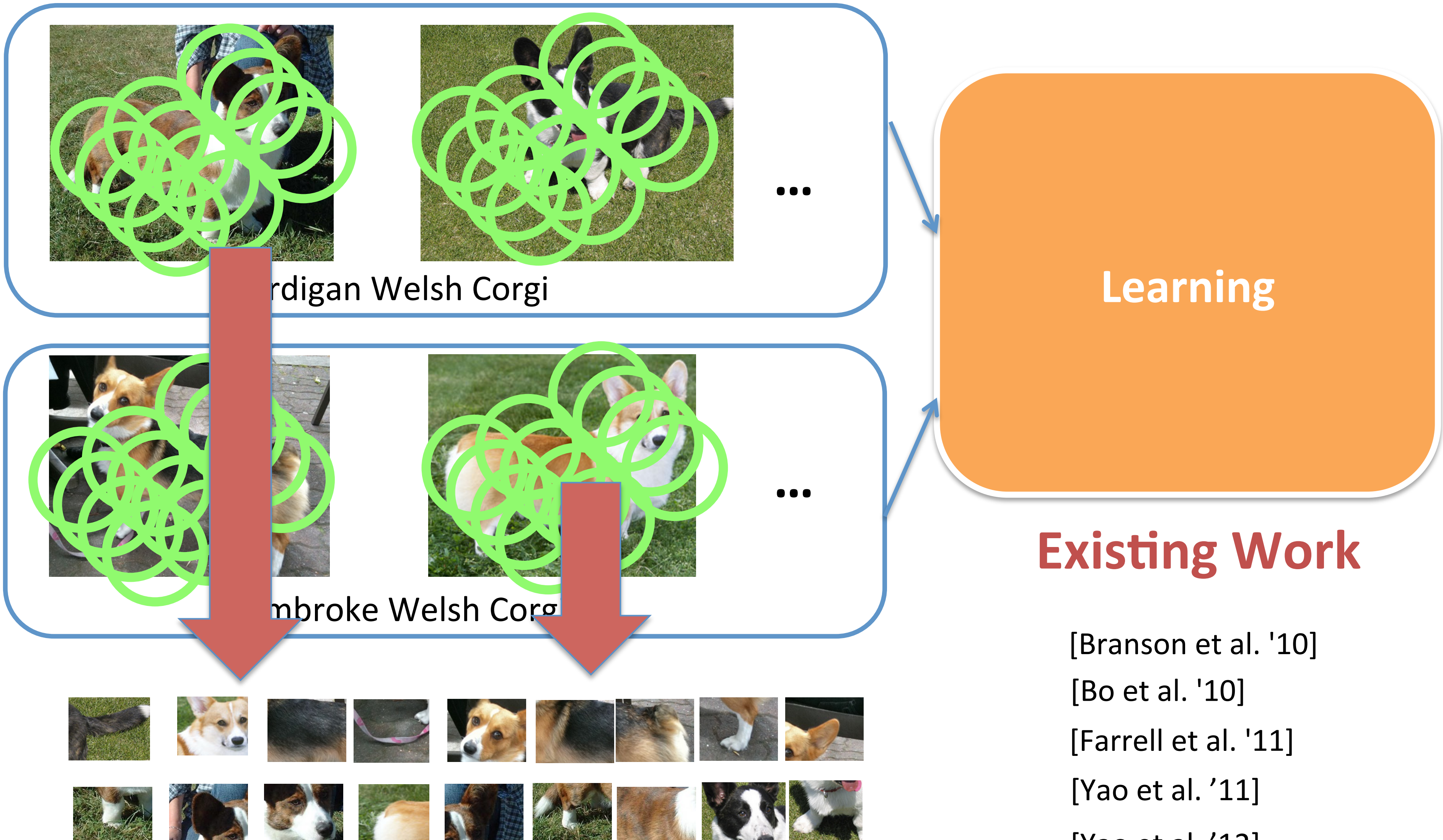
Pembroke Welsh Corgi



What breed is this dog?

Key: Find the right features.

Why is Fine-Grained Recognition Difficult?



Existing Work

[Branson et al. '10]

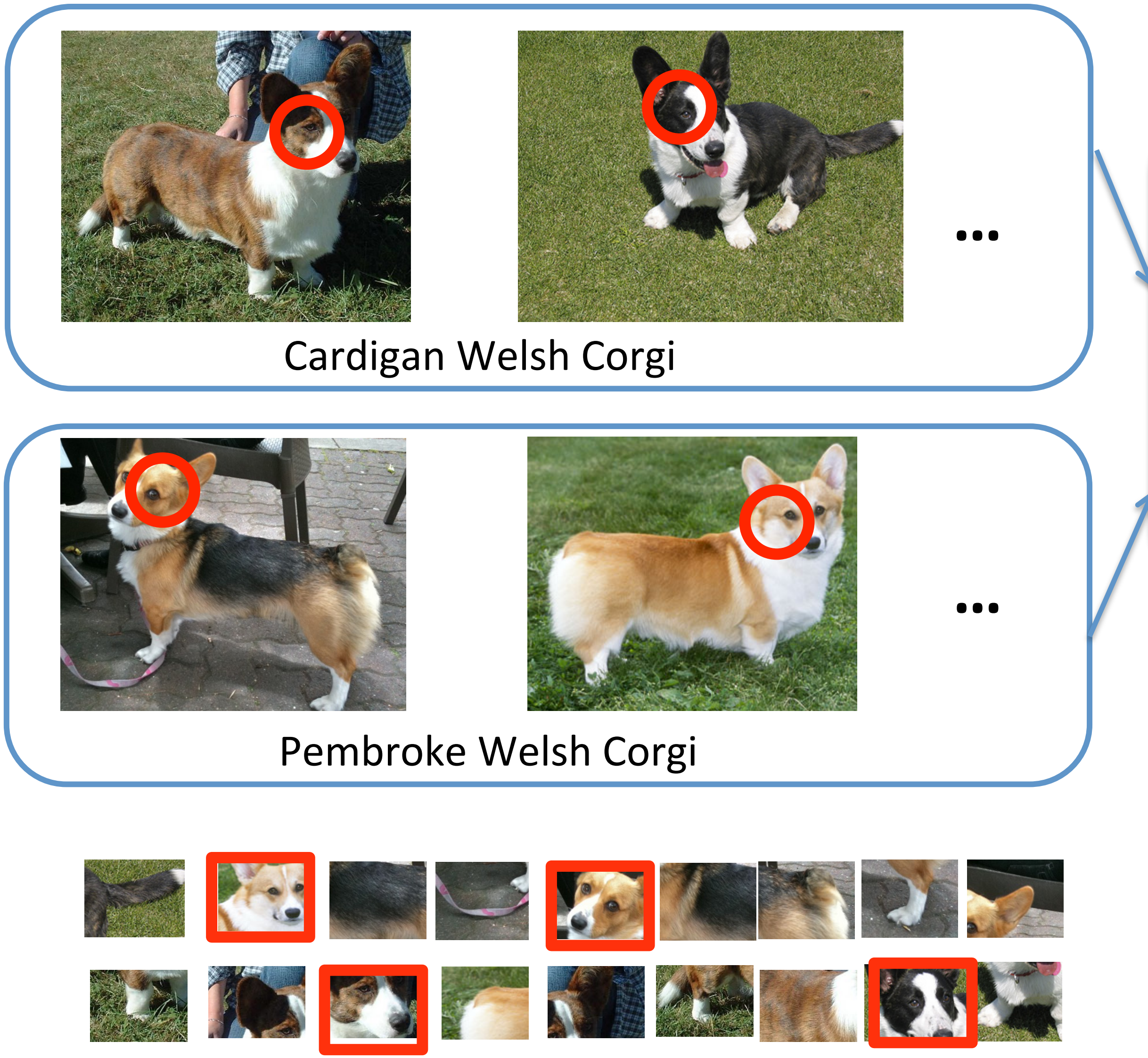
[Bo et al. '10]

[Farrell et al. '11]

[Yao et al. ’11]

[Yao et al. ’12]

Why is Fine-Grained Recognition Difficult?



Existing Work

[Branson et al. '10]

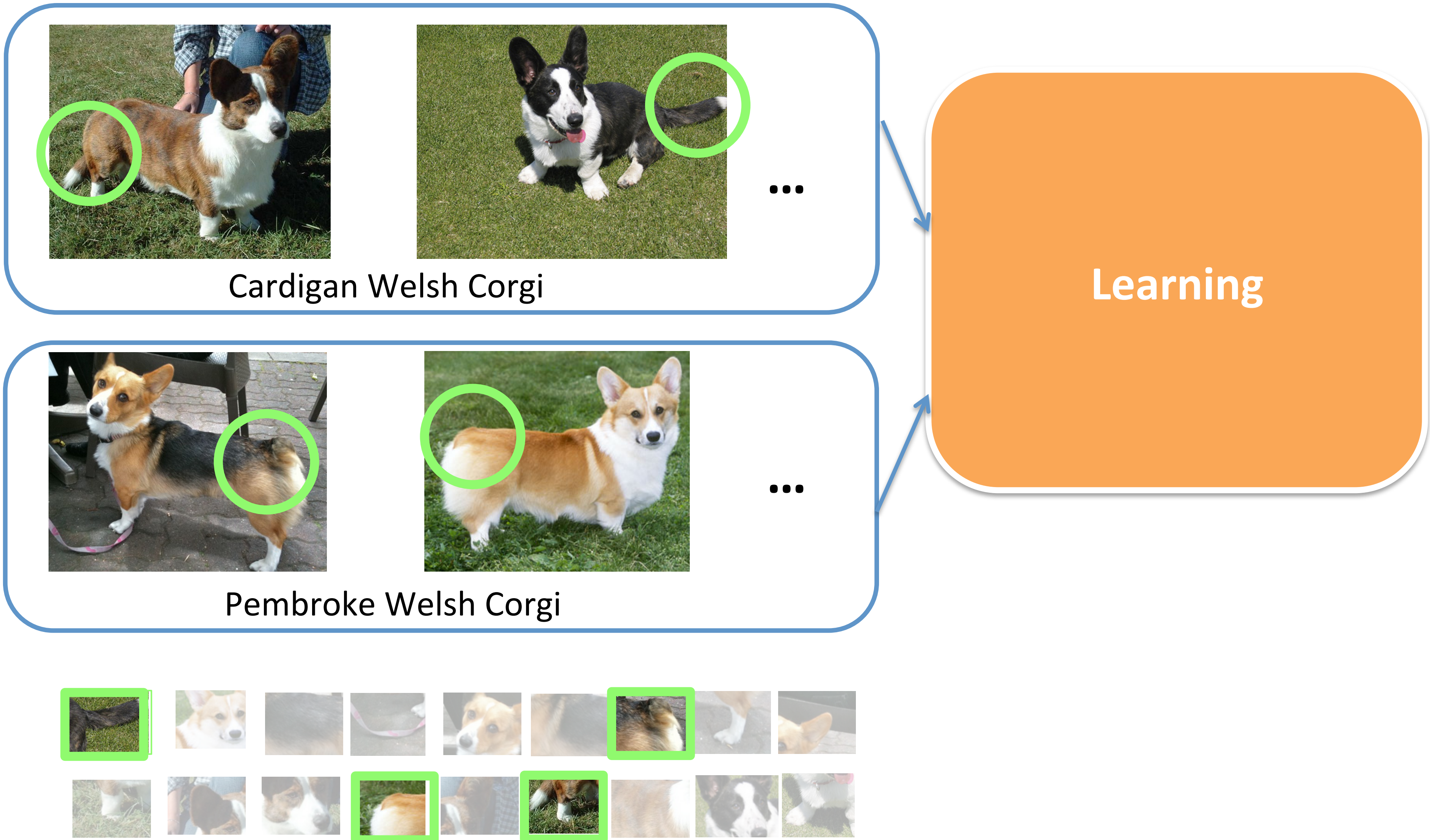
[Bo et al. '10]

[Farrell et al. '11]

[Yao et al. ’11]

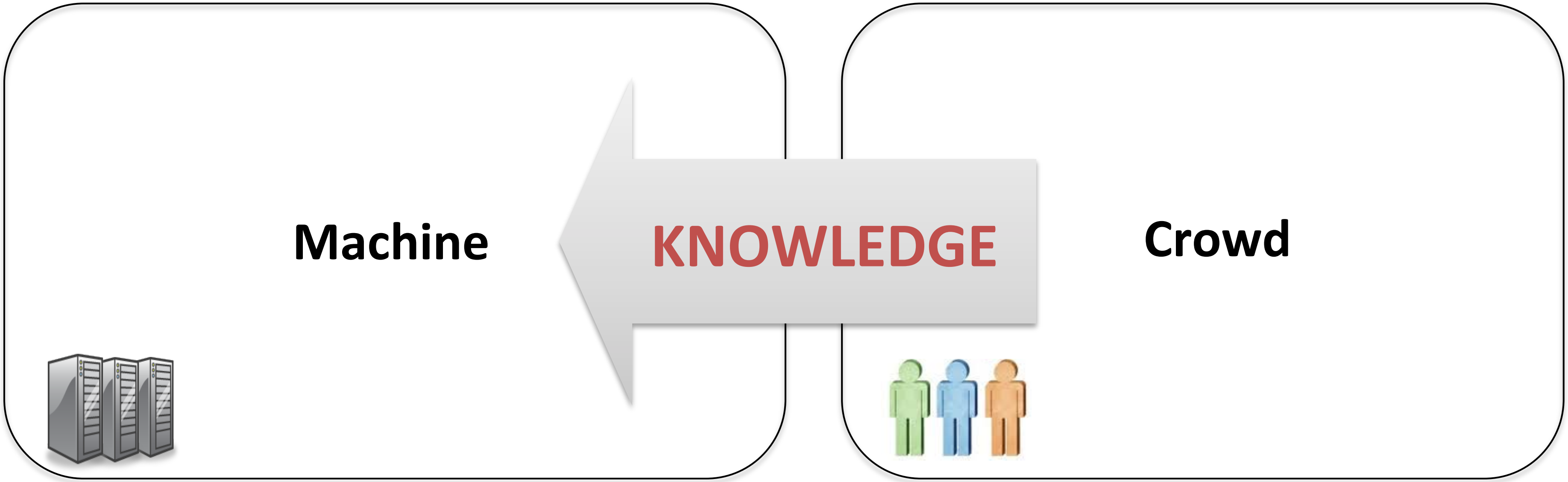
[Yao et al. ’12]

Why is Fine-Grained Recognition Difficult?

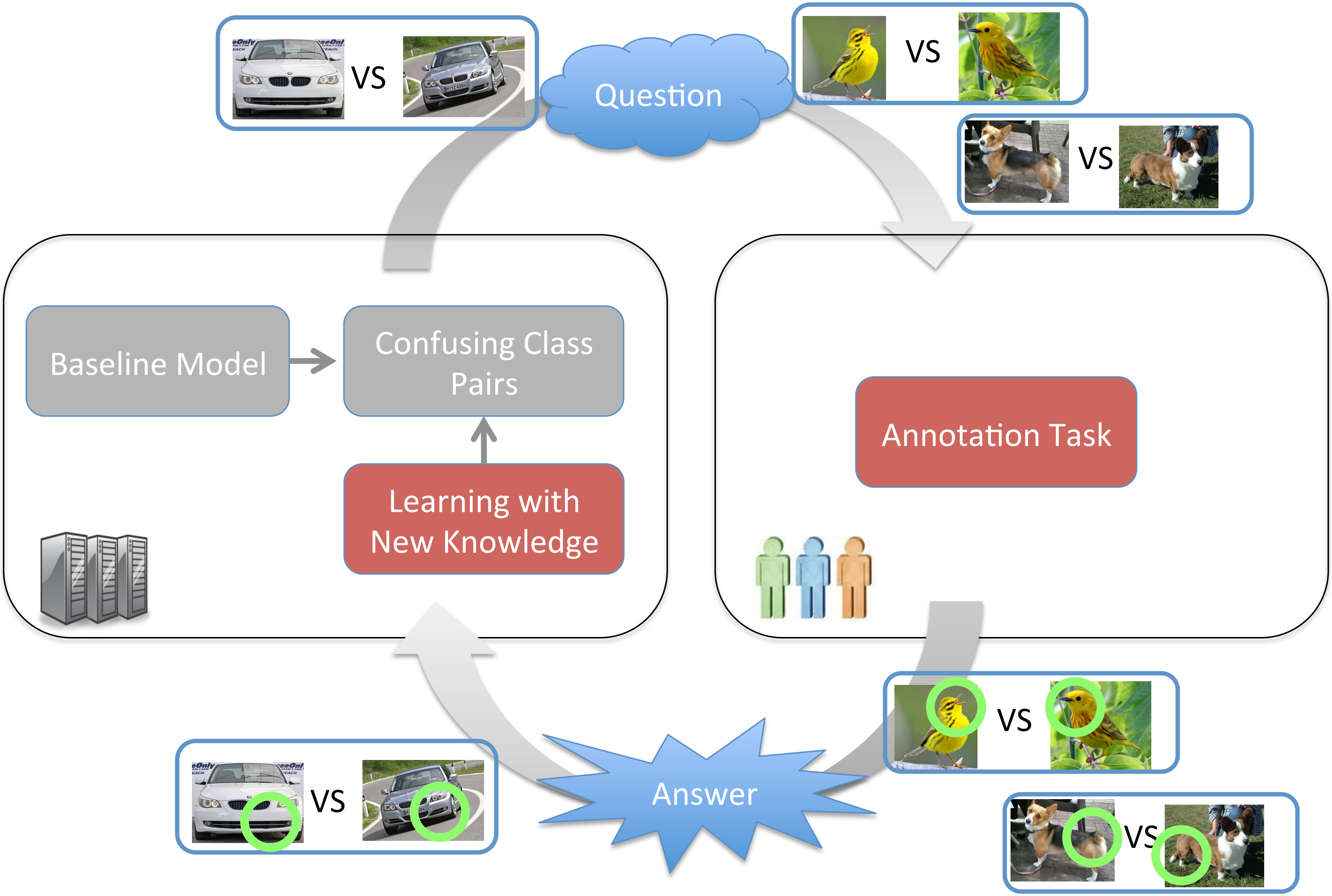


How to help computers select features?

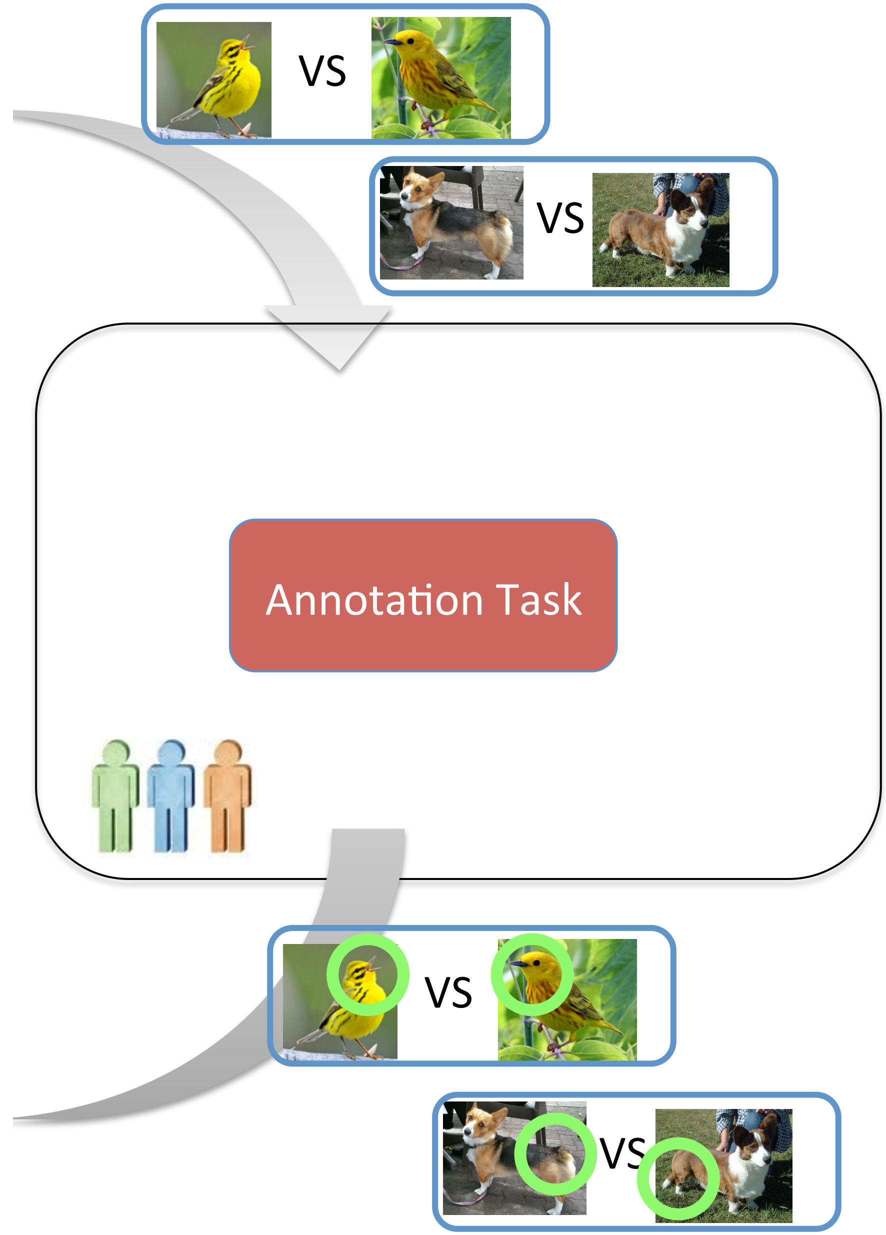
Machine-Crowd Collaboration



Machine-Crowd Collaboration

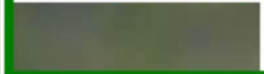


Machine-Crowd Collaboration



Click Me or Press 1

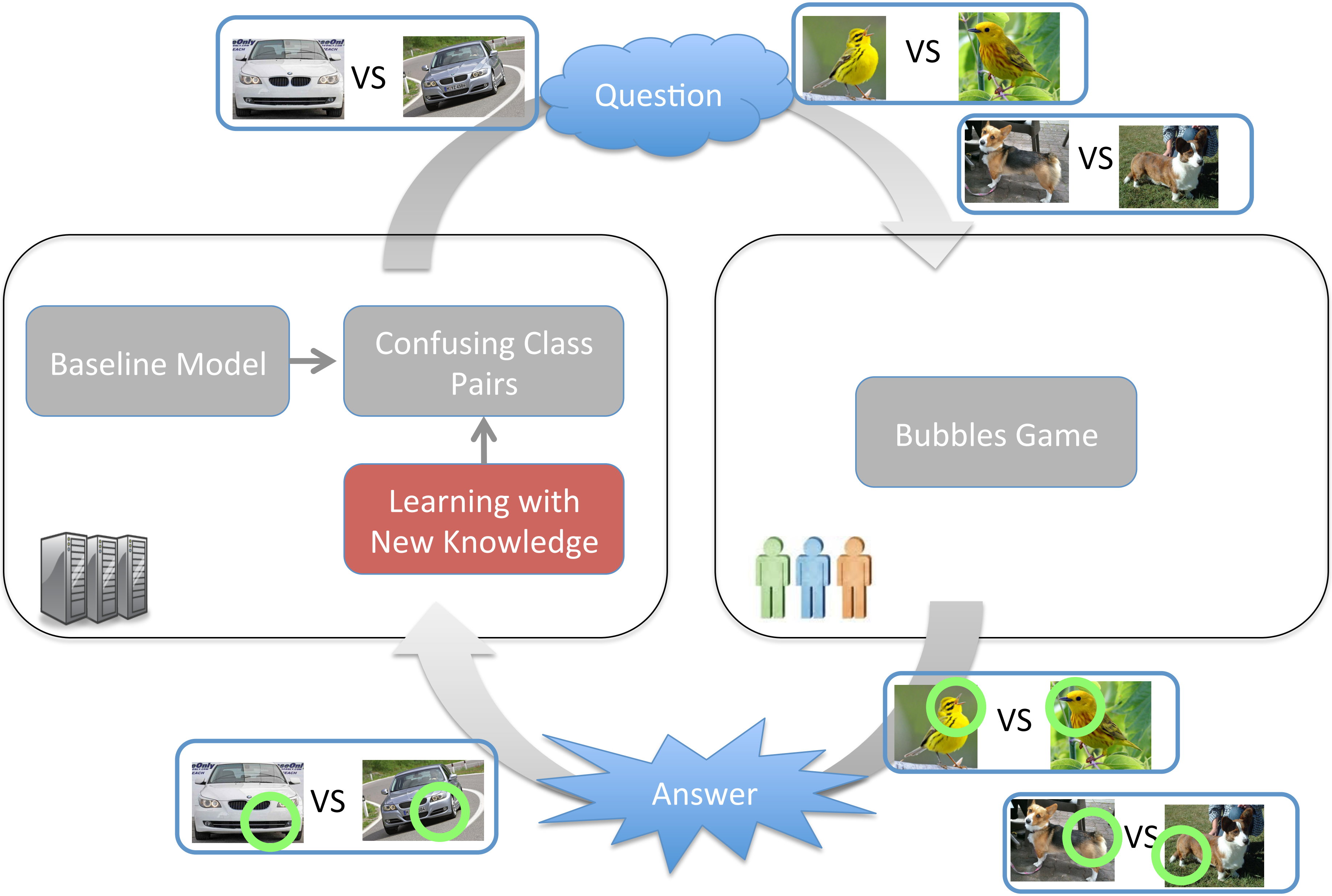
[Prairie Warbler (wikipedia)]



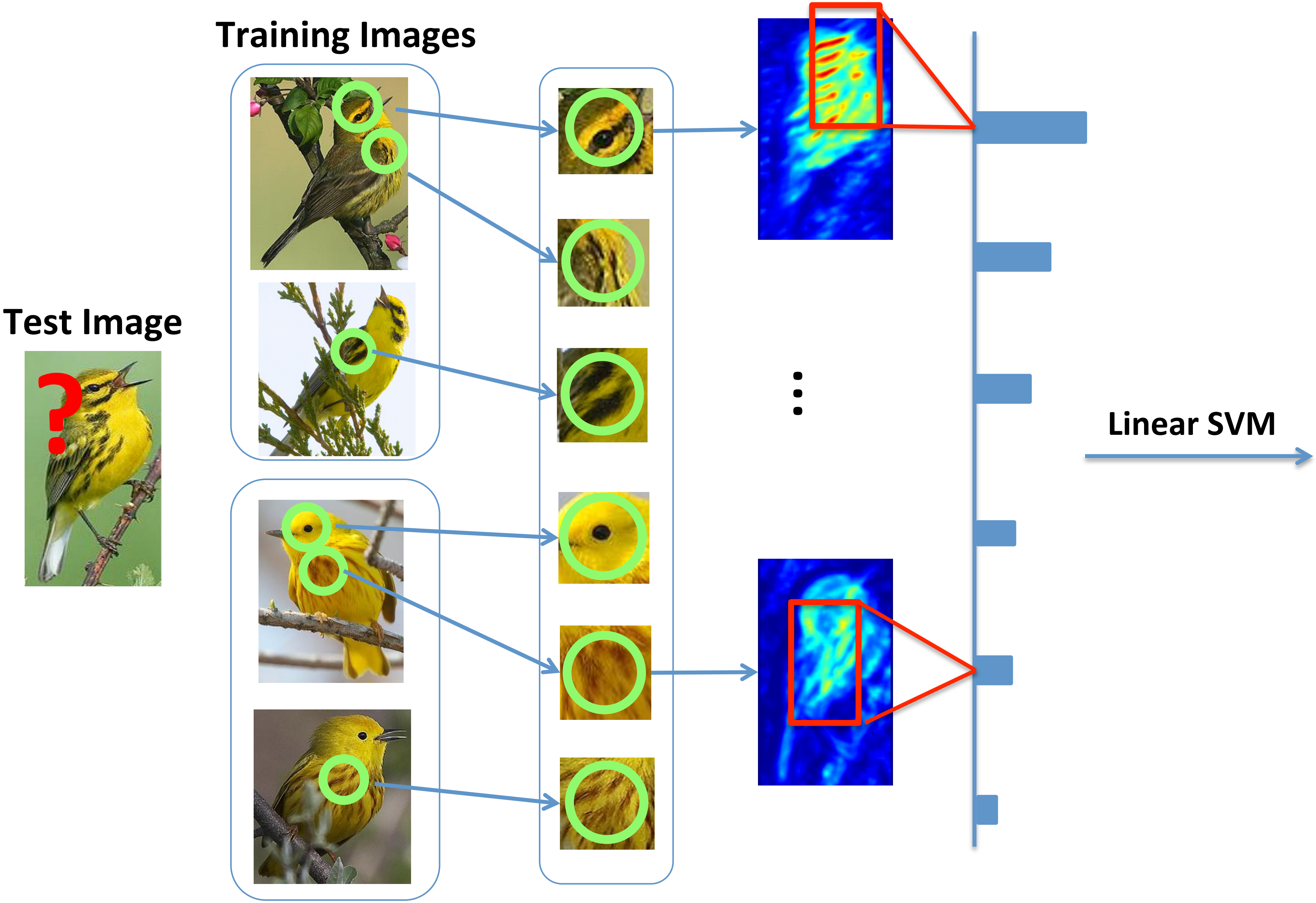
Click Me or Press 2

[Yellow Warbler (wikipedia)]

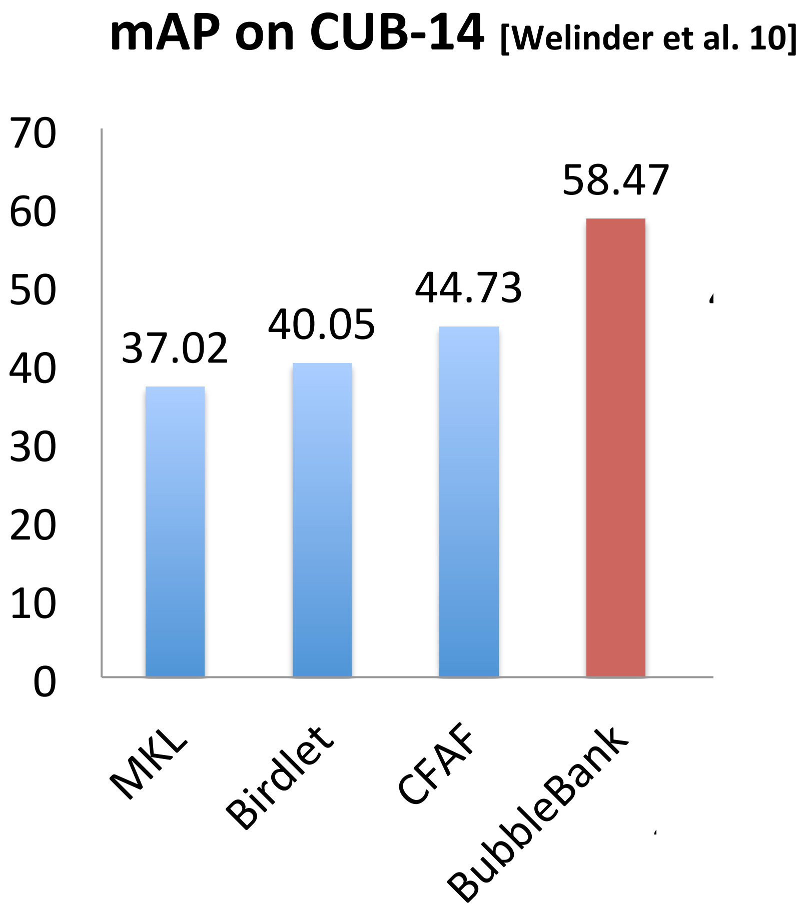
Machine-Crowd Collaboration

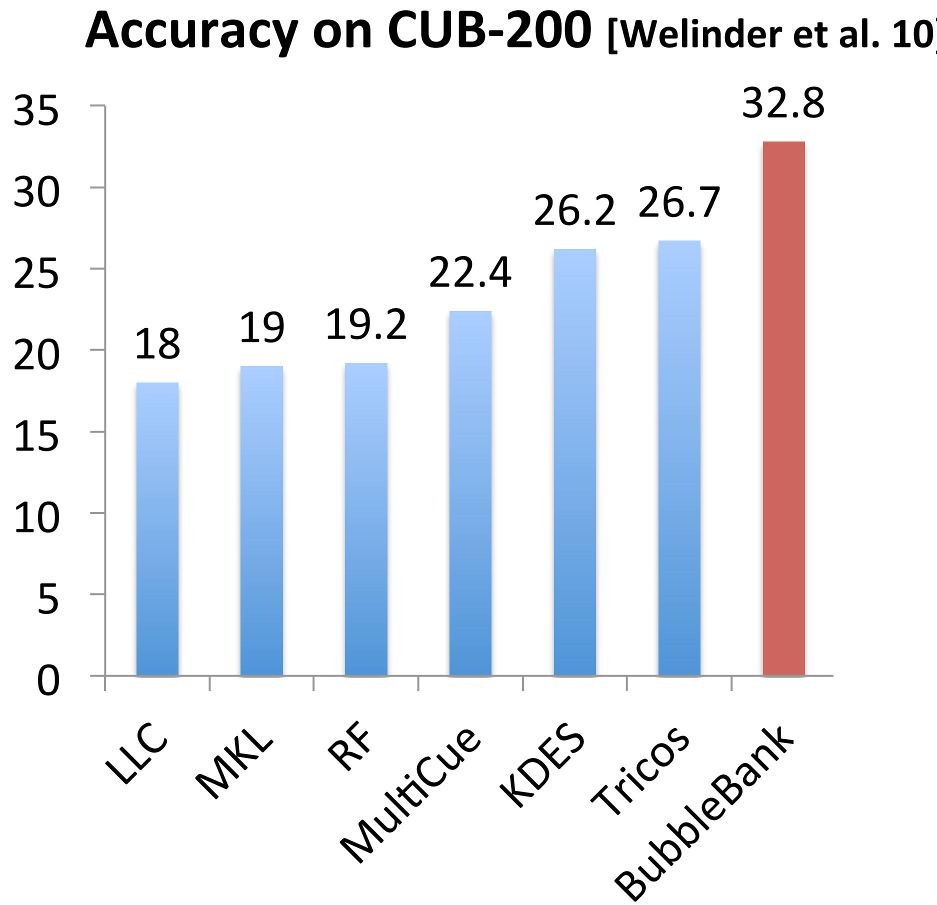


The BubbleBank Representation



Deng, Krause, & Fei-Fei, *CVPR2013*



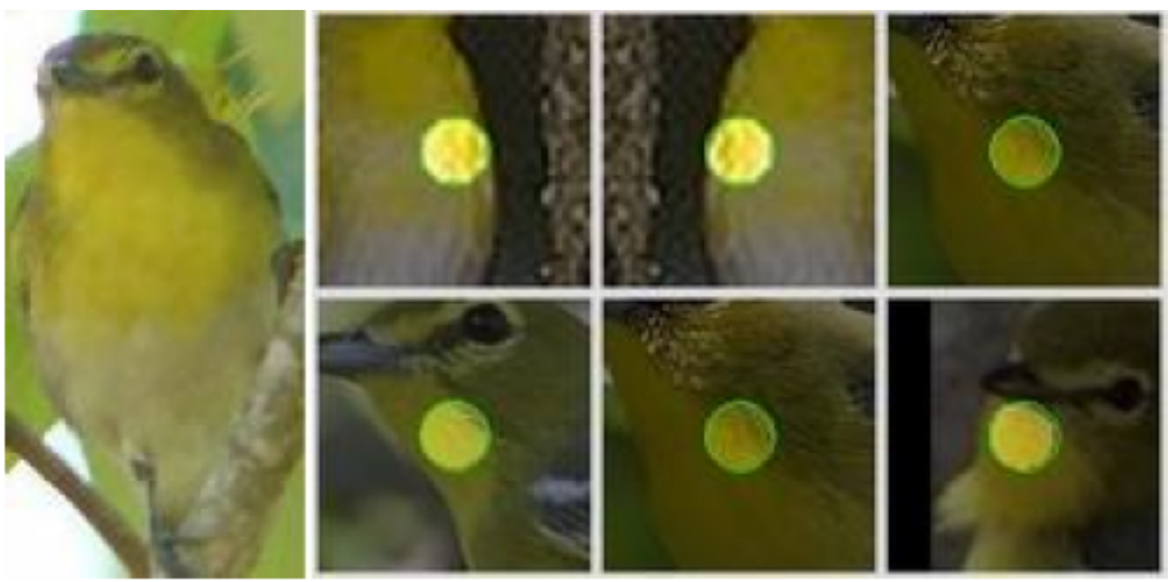


MKL [Branson et al. '10]
Birdlet [Farrell et al. '11]
CFAF [Yao et al.'12]

MKL [Branson et al. '10]
LLC [Wang et al. '09]
RF [Yao et al. '11]
MultiCue [Khan et al.'11]
KDES [Bo et al. '10]
Tricos [Chai '12]

Deng, Krause, & Fei-Fei, *CVPR2013*

Top Activated Bubbles (successful predictions)



Deng, Krause, & Fei-Fei, *CVPR2013*

Agenda

How to build a large-scale recognition engine using big data

STEP 1: Build a Large Knowledge Base (ImageNet)

STEP 2: Fine-Grained Recognition (Bubbles)

STEP 3: ?

Agenda

How to build a large-scale recognition engine using big data

STEP 1: Build a Large Knowledge Base (ImageNet)

STEP 2: Fine-Grained Recognition (Bubbles)

STEP 3: Putting a label on "everything"

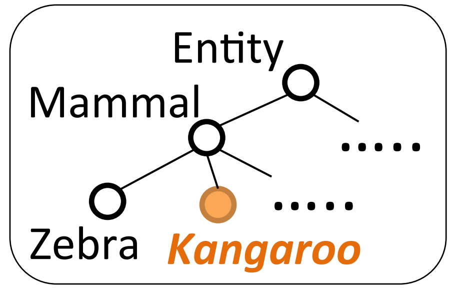
The Current State of the Art

10K classes	32.6%	Krizhevsky et al. NIPS 2012
20K classes	15%	Le et al. NIPS 2012

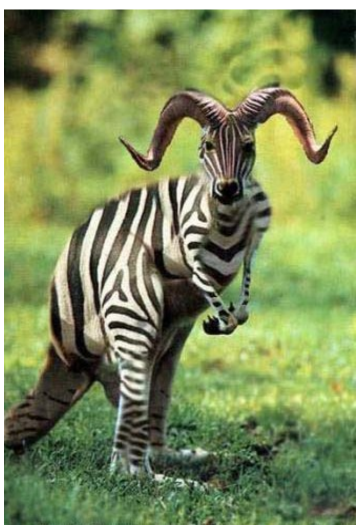
Not quite practical yet…

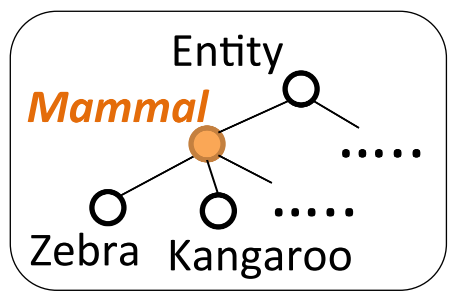
But we are measuring the very fine-grained level

Hedging: Be as informative as possible with few mistakes



→ ***Kangaroo*** ✓

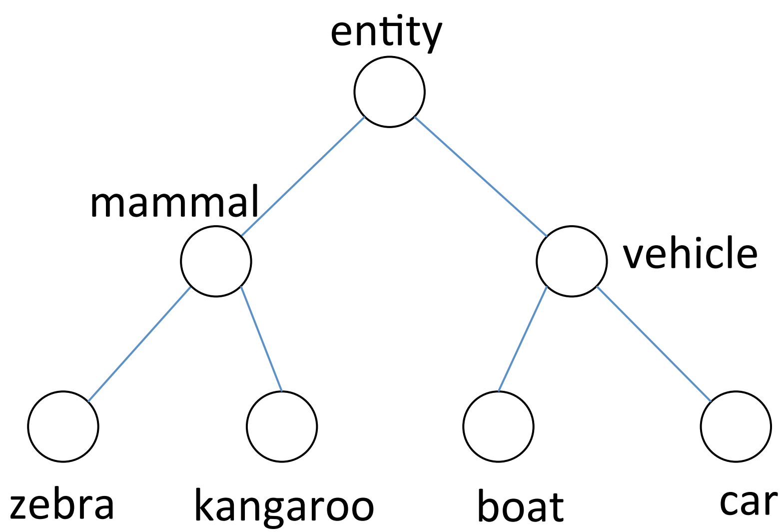




→ ***Mammal*** ✓

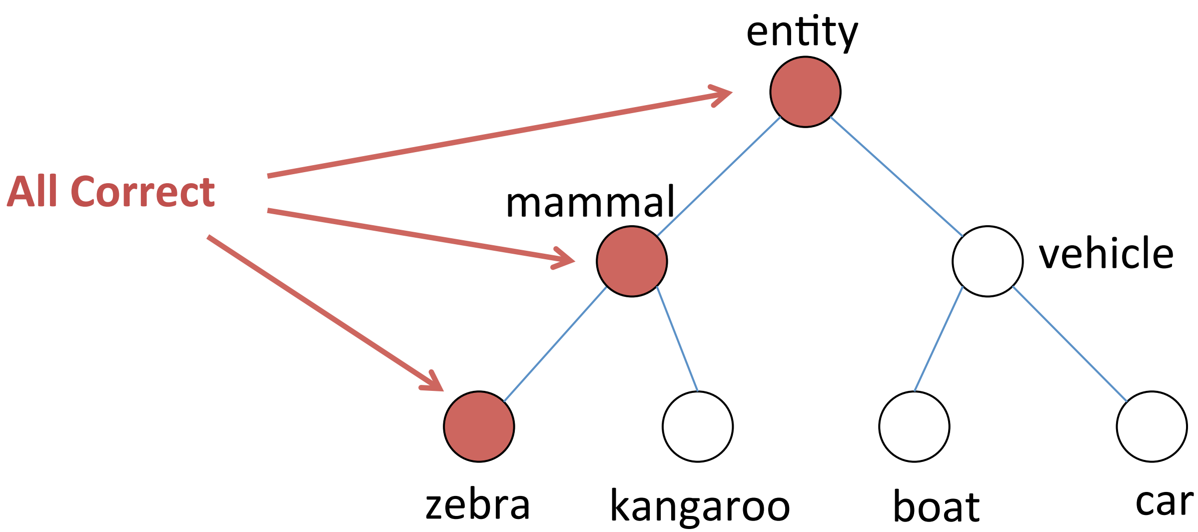
Deng, Krause, Berg, Fei-Fei, CVPR2012

Formal Problem Statement



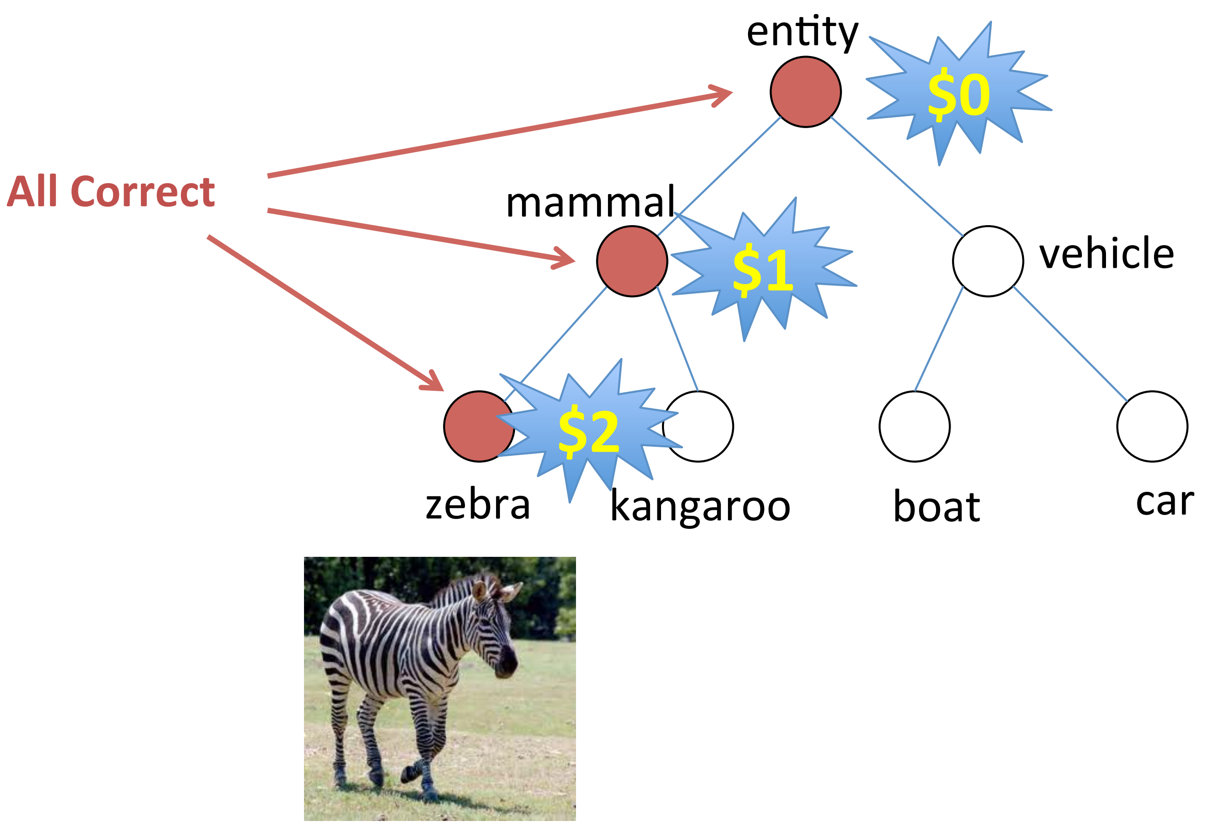
Deng, Krause, Berg, Fei-Fei, CVPR2012

Formal Problem Statement



Deng, Krause, Berg, Fei-Fei, CVPR2012

Formal Problem Statement



Deng, Krause, Berg, Fei-Fei, CVPR2012

Formal Problem Statement

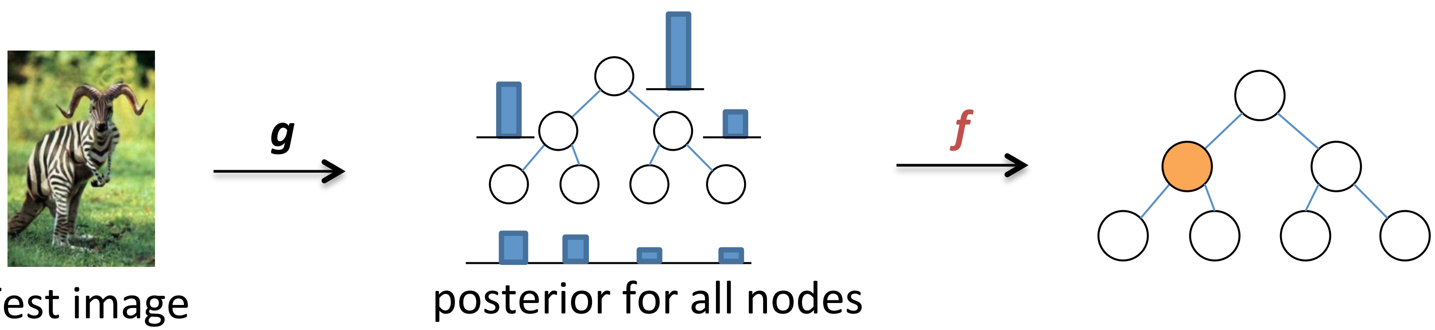
Assumptions

- Same distribution for training and test.
- A base classifier ***g*** that gives posterior probability on the hierarchy.

Goal

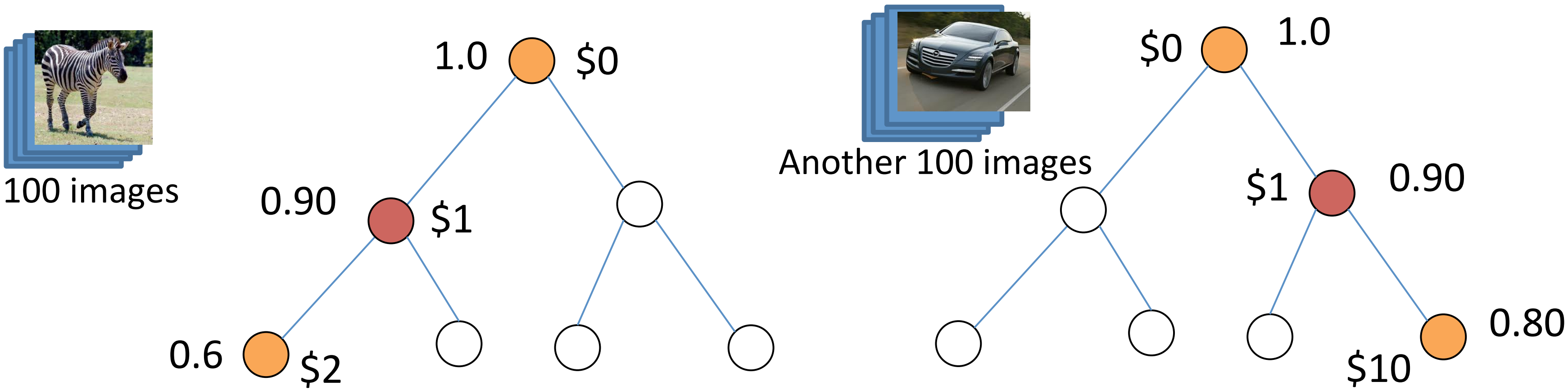
- Find a ***decision rule f***
  - Expected accuracy ***A(f)*** is at least ***1-ε***
  - Maximize expected reward ***R(f)***

$$\begin{aligned} &\underset{f}{Maximize} \quad R(f) \\ &Subject\ to \quad A(f) \geq 1 - \varepsilon \end{aligned}$$



Deng, Krause, Berg, Fei-Fei, CVPR2012

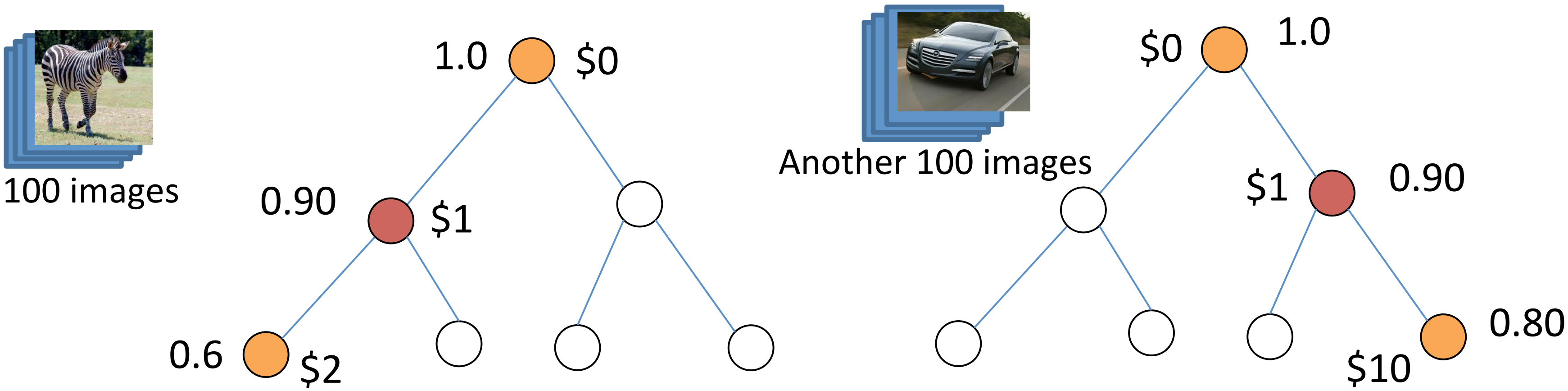
Pick a global confidence threshold T=0.9 [Vailaya et al. '99]



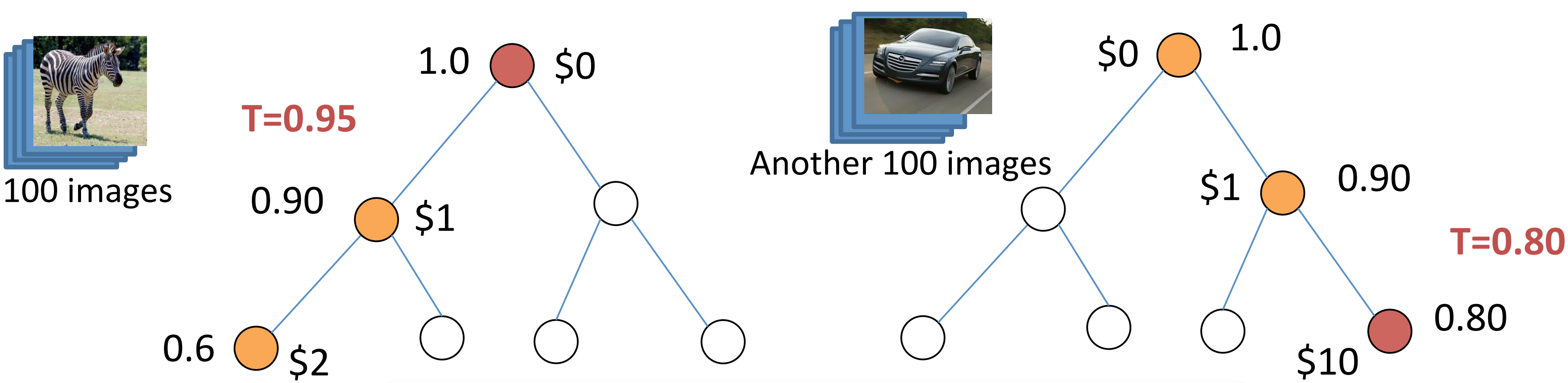
Reward = ($1 * 0.90 + $1 * 0.90) / 2 = $0.90
Accuracy = (0.90 + 0.90) / 2 = 0.90

Deng, Krause, Berg, Fei-Fei, CVPR2012

Pick a global confidence threshold T=0.9 [Vailaya et al. '99]

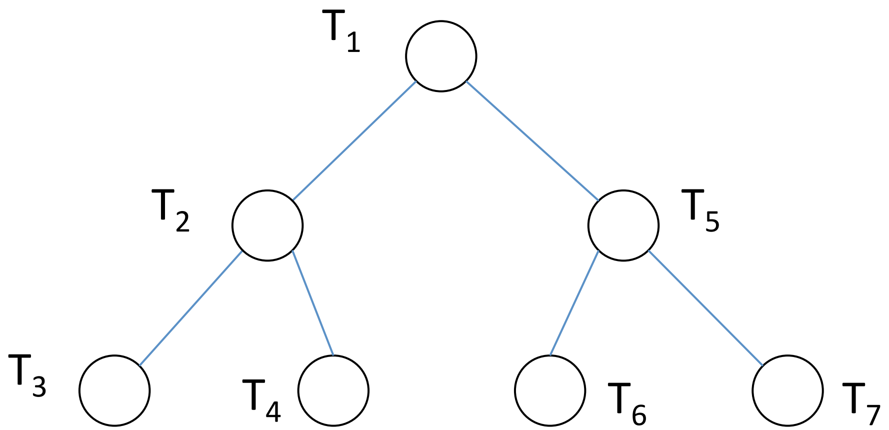


Reward = ($1 * 0.90 + $1 * 0.90) / 2 = $0.90
Accuracy = (0.90 + 0.90) / 2 = 0.90



Reward = ($0 * 1.0 + $10 * 0.80) / 2 = $4
Accuracy = (1.0 + 0.80) / 2 = 0.90

Deng, Krause, Berg, Fei-Fei, CVPR2012

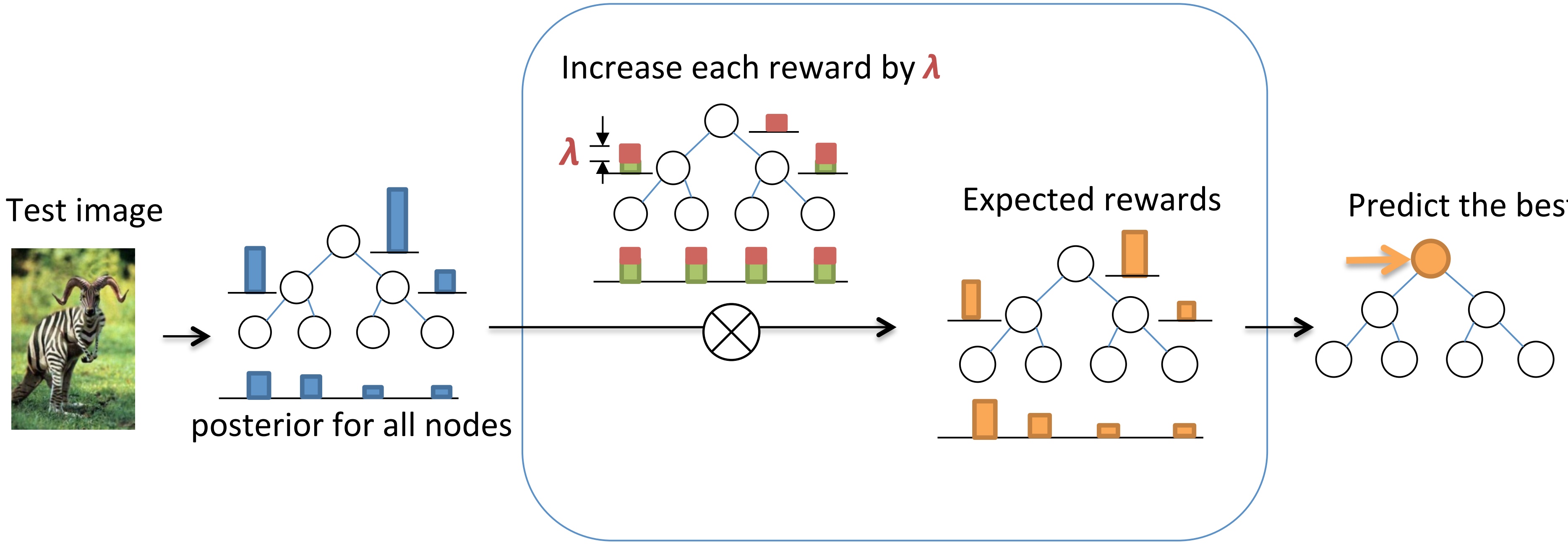


We can optimize individual thresholds…

But actually we don't need to.
There is a simpler and provably optimal solution

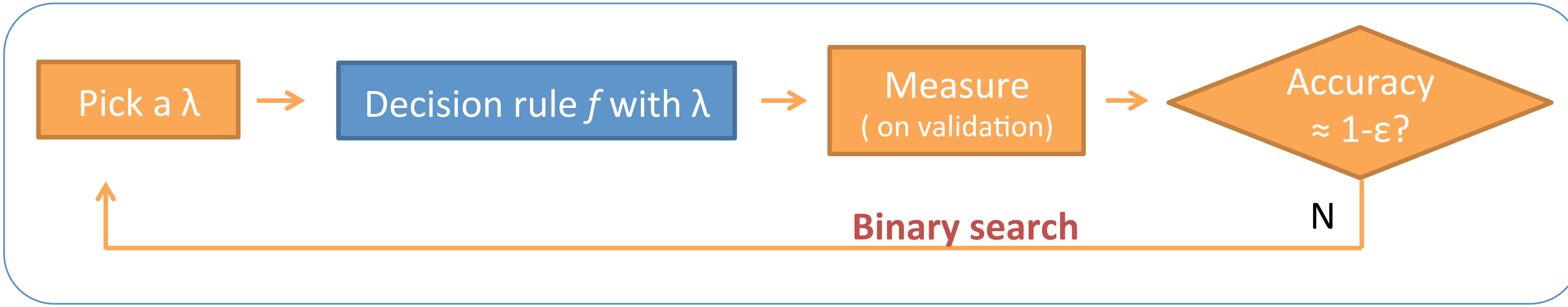
Deng, Krause, Berg, Fei-Fei, CVPR2012

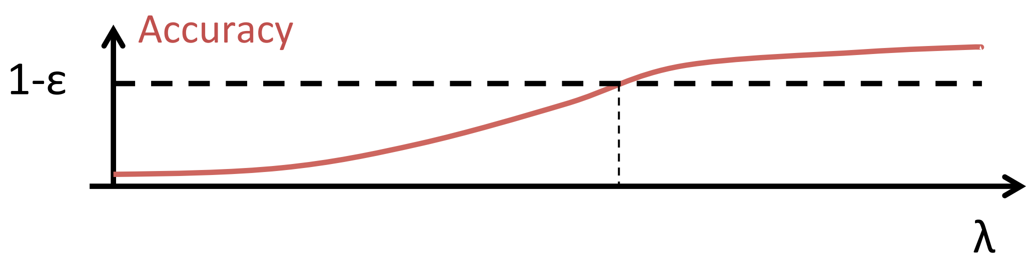
A global, fixed scalar parameter *λ≥0*



Deng, Krause, Berg, Fei-Fei, CVPR2012

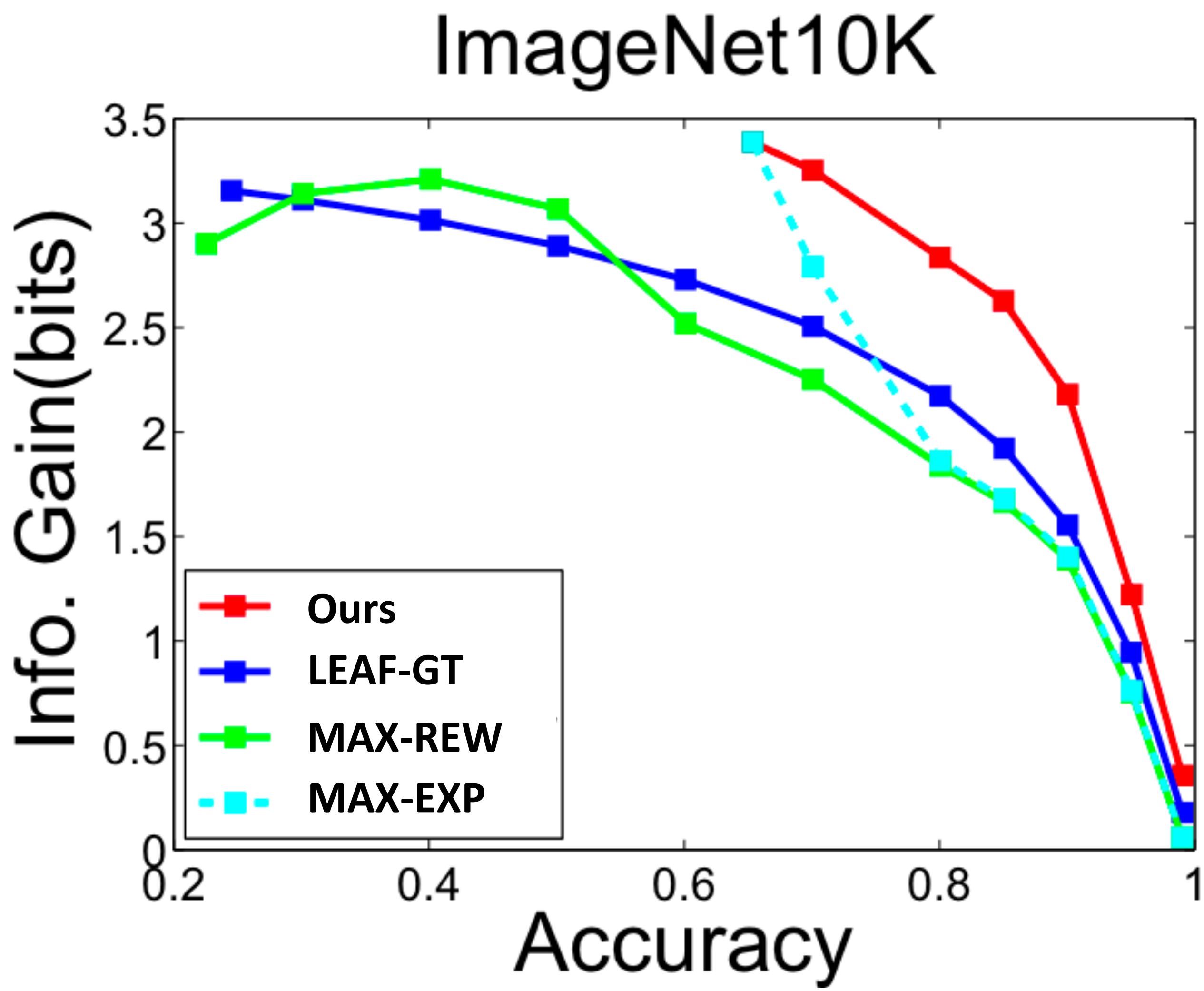
The DARTS algorithm





Theorem: Under very mild conditions, this is optimal.

Deng, Krause, Berg, Fei-Fei, CVPR2012



Deng, Krause, Berg, Fei-Fei, CVPR2012

EVA

Engine for Visual Annotation

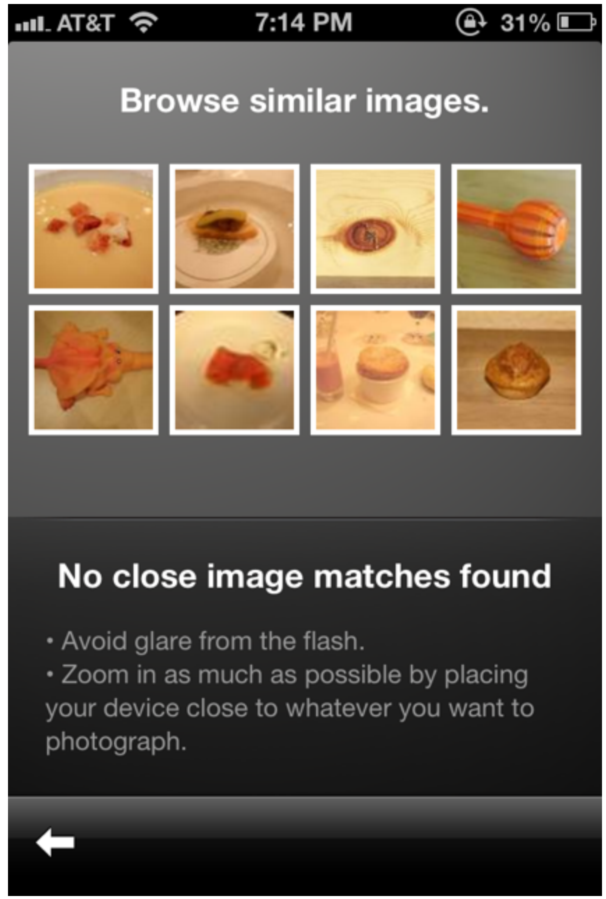
The EVA system, powered by **ImageNet**, can annotate images with guaranteed accuracies. It currently recognizes over **10,000** visual categories. See the project page to find out more.

Paste a URL | Upload an image

ANNOTATE

Google Goggles

Use pictures to search the web.



EV

Engine for Visual Annotation

0.95 coffee mug

0.97 mug

0.99 drinking vessel

Google images

EVA
Engine for Visual Annotation

Image size:
401 × 604

No other sizes of this image found.

Visually similar images - Report images

0.87 face , gas pump, person

0.90 face , gas pump

EVA

Engine for Visual Annotation

0.75 artifact, crater, matter, vertebrate

0.77 crater, matter, vertebrate

0.78 chordate, crater, matter

0.86 animal, matter

0.87 animal

Summary

How to build a large-scale recognition engine using big data

STEP 1: **Build a Large Knowledge Base (ImageNet)**

STEP 2: **Fine-Grained Recognition (Bubbles)**

STEP 3: **Putting a label on “everything” (Hedging)**

ImageNet Challenge 2013

- 1.2 Million images and 1000 classes
- New PASCAL-style Detection Task
  - Full annotation of 200 classes in test images.
- http://www.image-net.org/challenges/LSVRC/2013/

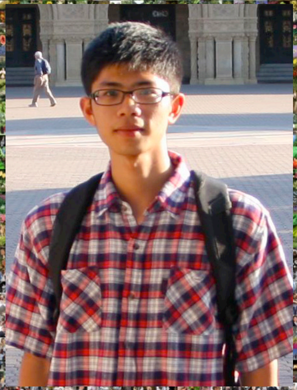
Fine-Grained Challenge 2013

- Competition on Fine-Grained Recognition
  - Airplanes, Birds, Cars, Dogs, Shoes.
- https://sites.google.com/site/fgcomp2013/

Thank you!








Prof. Kai Li
Princeton U.

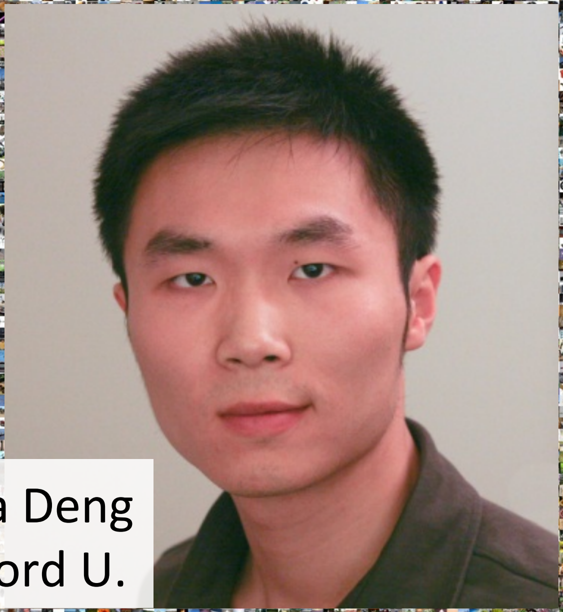
Prof. Alex Berg
Stony Brook U.

Sanjeev Satheesh
Stanford U.

Jonathan Krause
Stanford U.

Zhiheng Huang
Stanford U.

Olga Russakovsky
Stanford U.



Dr. Jia Deng
Stanford U.

Google

NSF

Microsoft Research

DEPARTMENT OF THE NAVY ONR Science & Technology