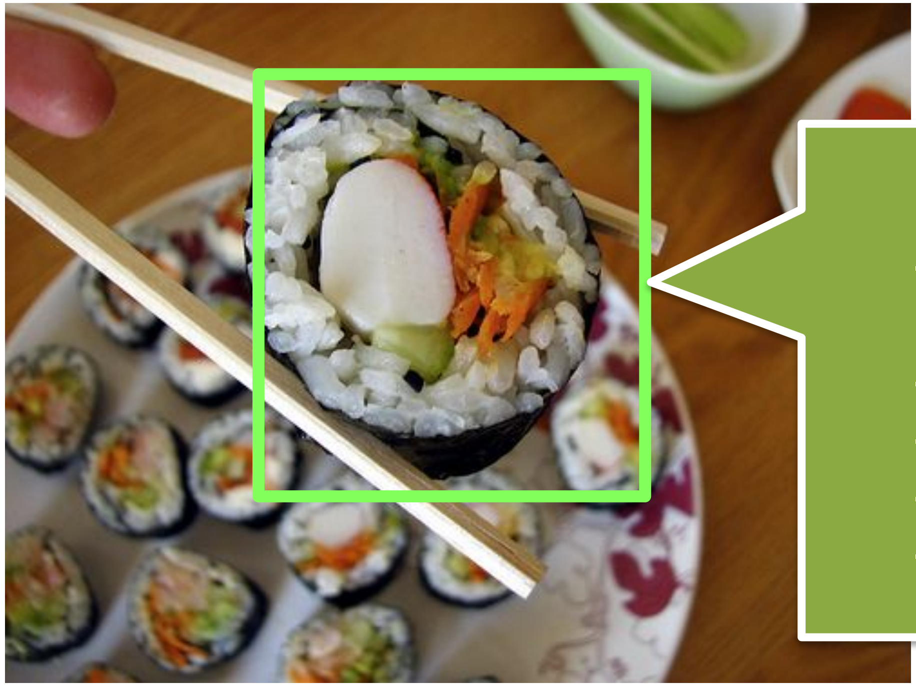
# Recitation

Large Scale Recognition & Visual Ontologies



**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**

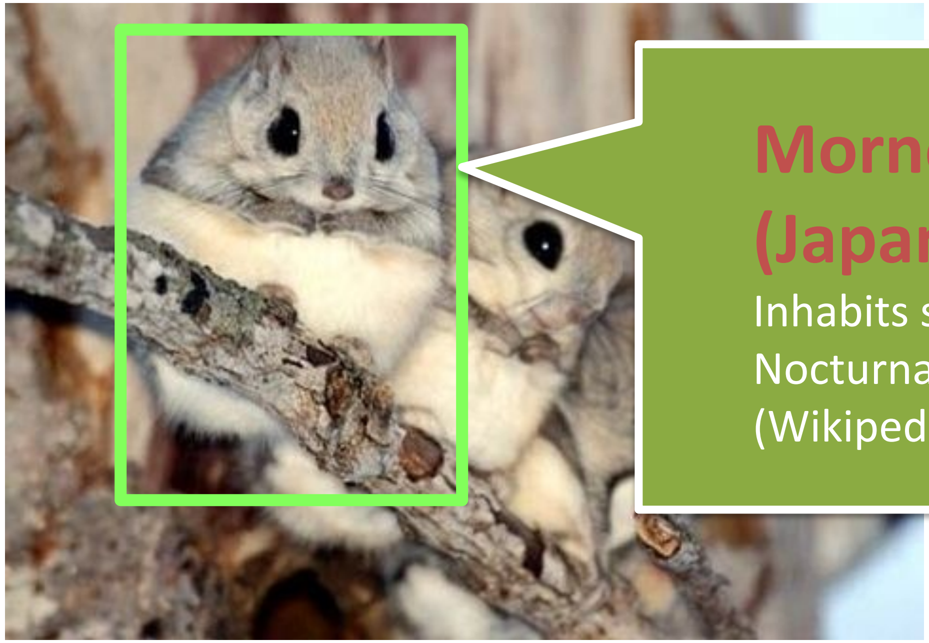
Raise arms to appear larger.

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 could recognize

**EVERYTHING**

Surveillance

Robotics

Assistive tools


Wearable devices

**Recognition Engine**

Driverless cars

Smart photo album

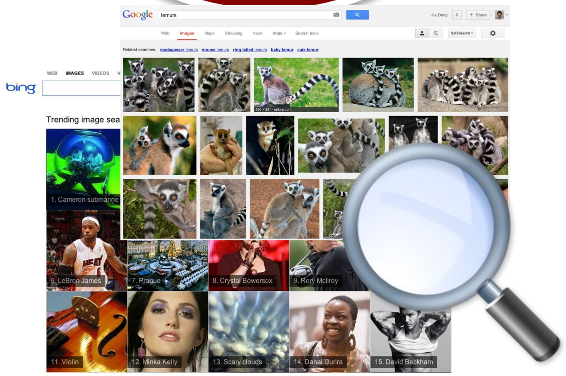



Image search

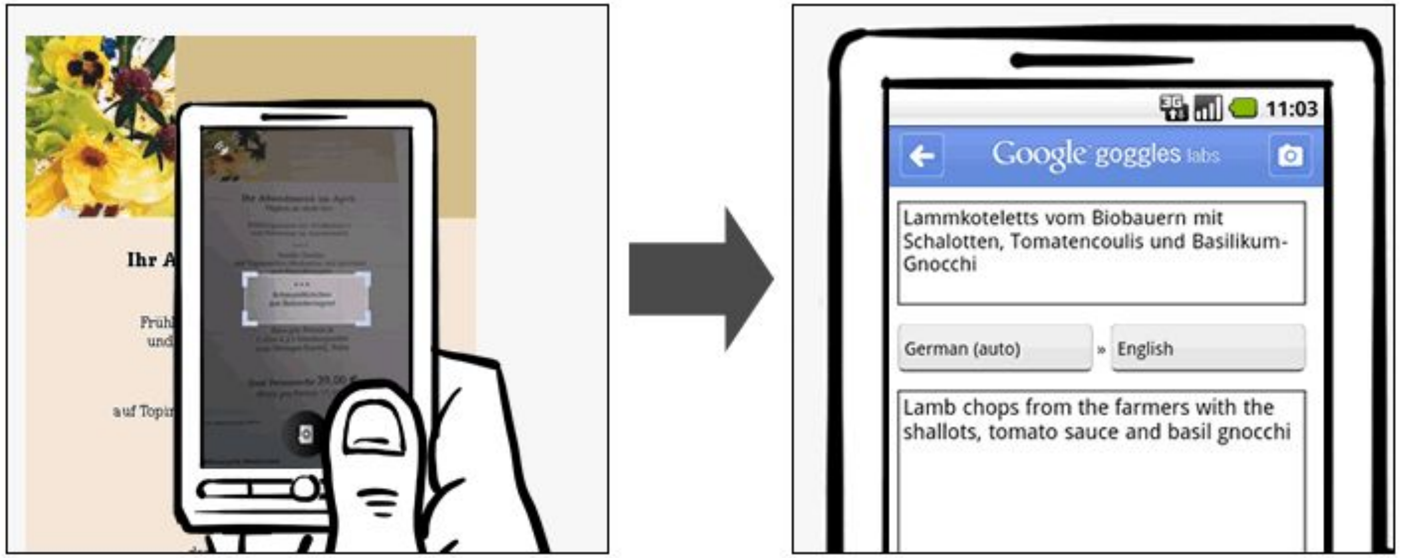

Mining social media

## What can computers already recognize?


AUTO
10M
IN
39
Nikon
The Nikon S60. Detects up to 12 faces.

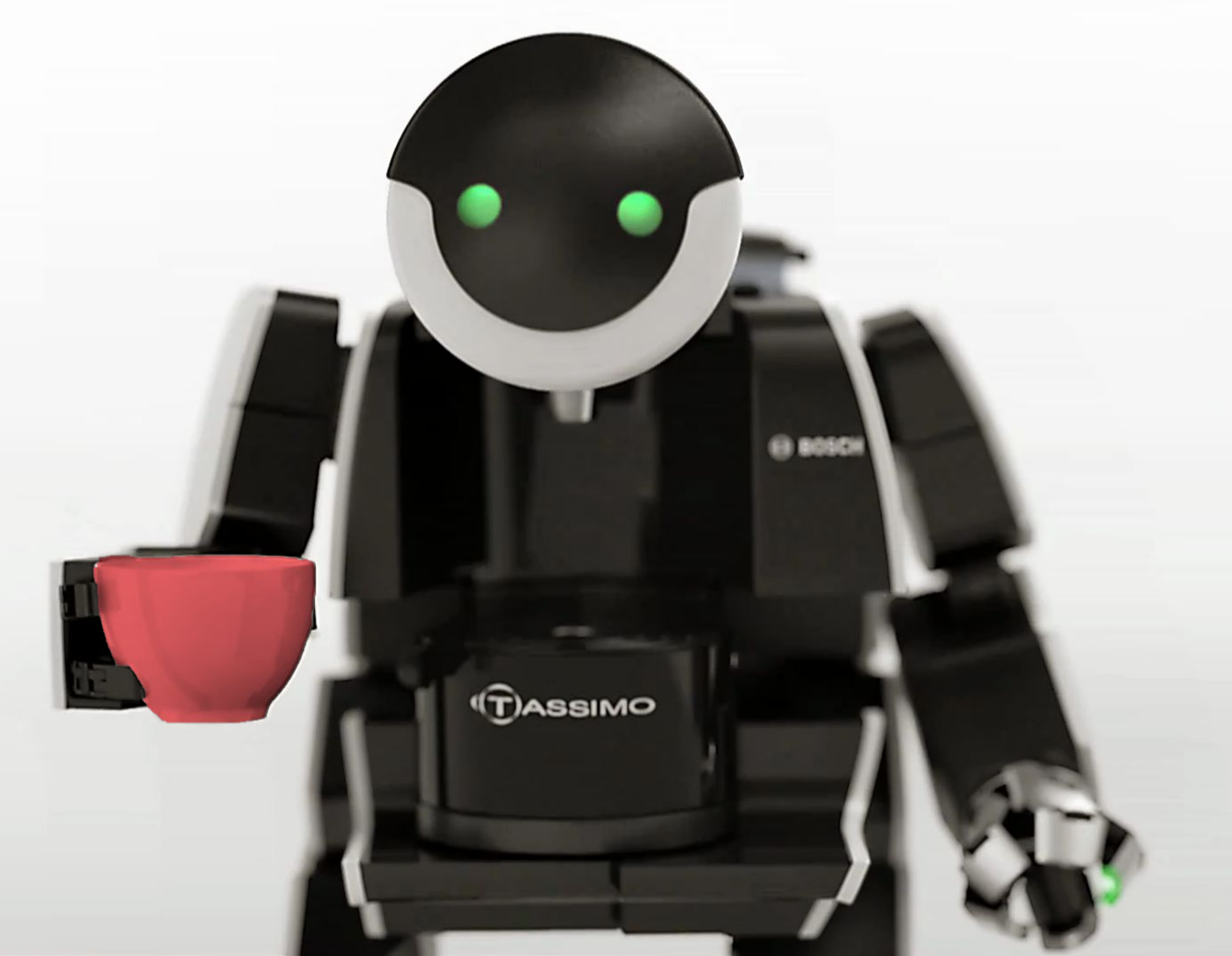
Google Goggles

Use pictures to search the web.

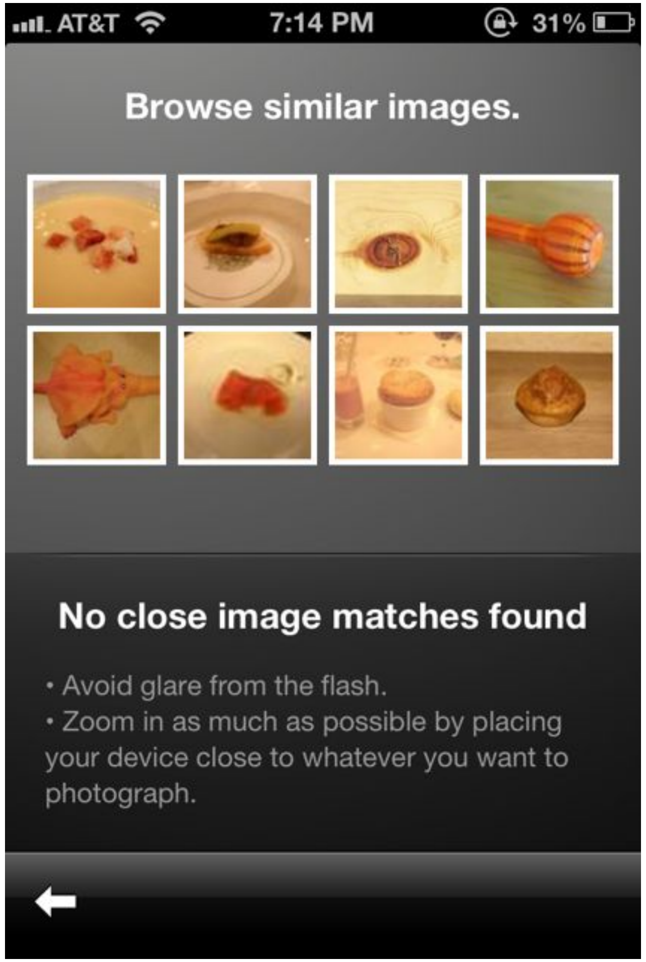



**What’s the next to work on?**

**Coffee Mugs!**


BOSCH
TASSIMO


EMY


AT&T
7:14 PM
31%
Browse similar images.
No close image matches found
• Avoid glare from the flash.
• Zoom in as much as possible by placing your device close to whatever you want to photograph.

## 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!**

## The rest of the talk will be about Coffee Mugs!

# What about Gas Pumps!

**The rest of the talk will be about Coffee Mugs**

**And Gas Pumps**

**And Solar arrays**

**Radio**

**First aid kit**

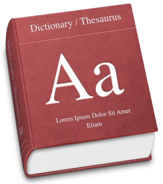
**Spacesuit**

**Oxygen Cylinder**

What do they have in common?

Let's work on recognizing **EVERYTHING**

# How many things are there?





10K+
[Biederman ’87]

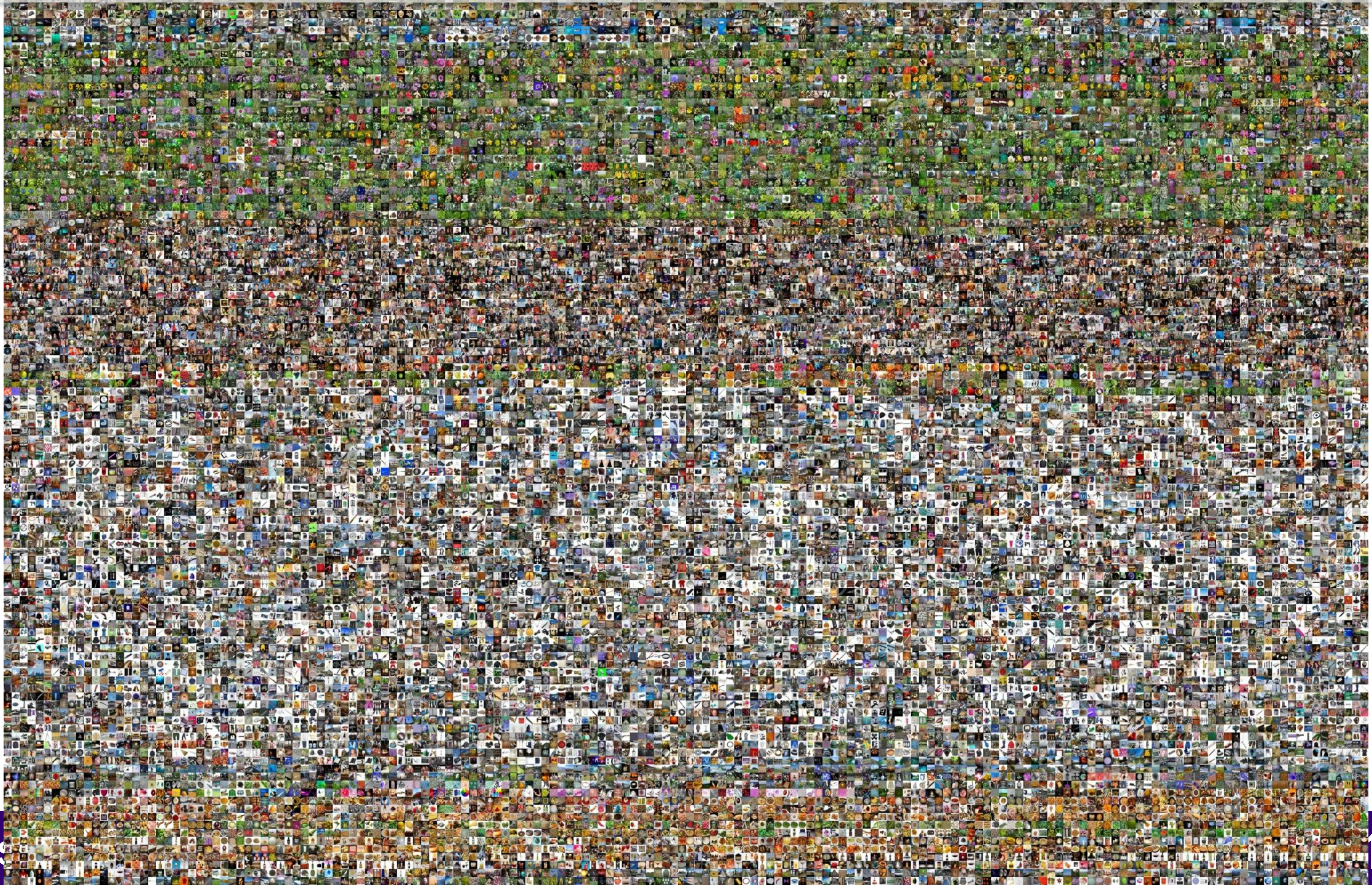
60K+
product categories

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

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

4.1M+
articles

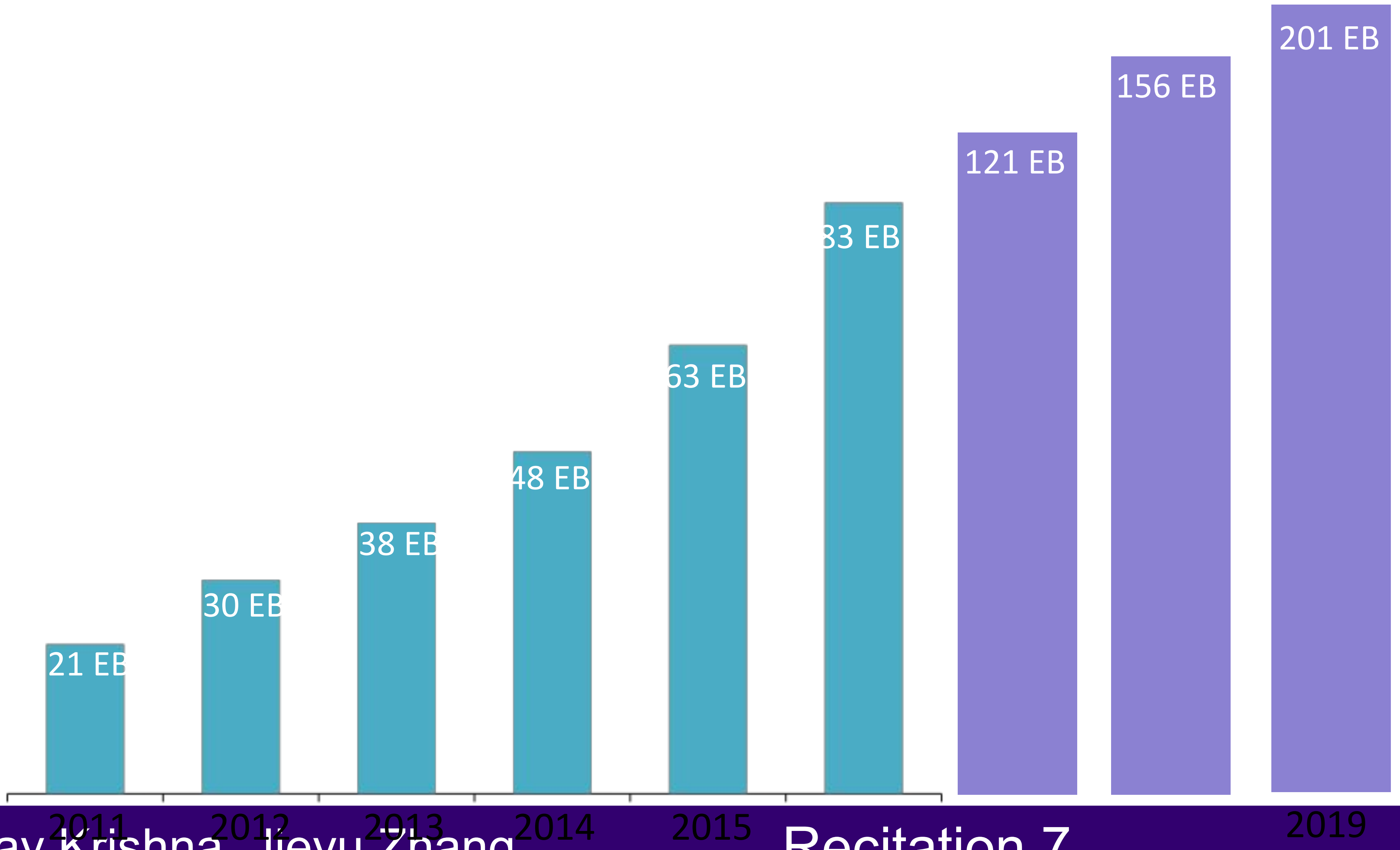
# From 20 classes to Millions?




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

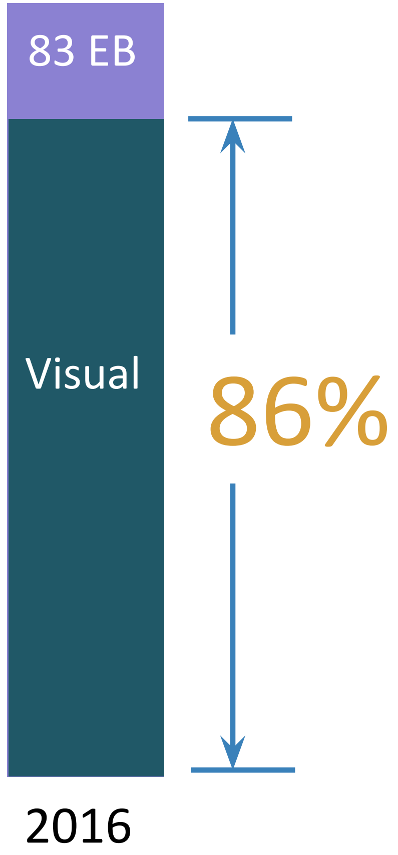
# Big Data from the Internet

# Global Consumer Internet Traffic Per Month



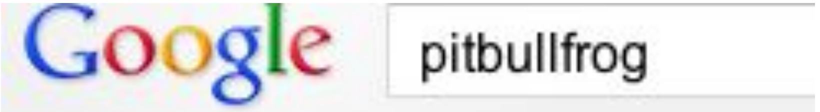
72 hours of videos / min

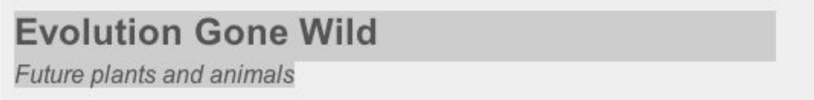
300 million images / day



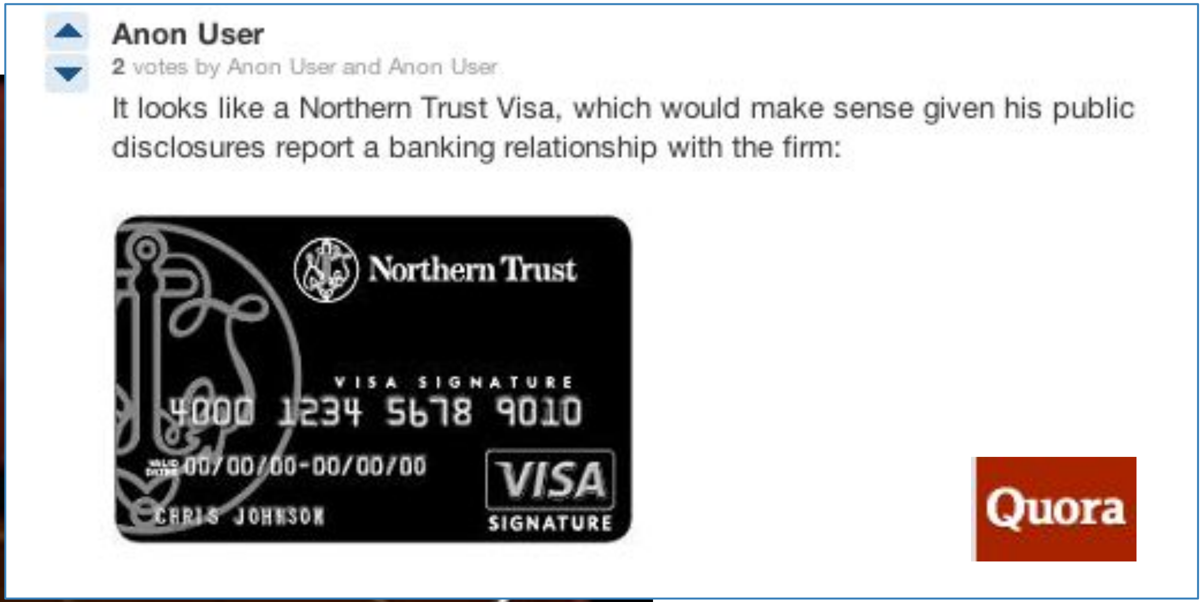
Big Data from the Internet

 The Internet can teach **EVERYTHING**




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



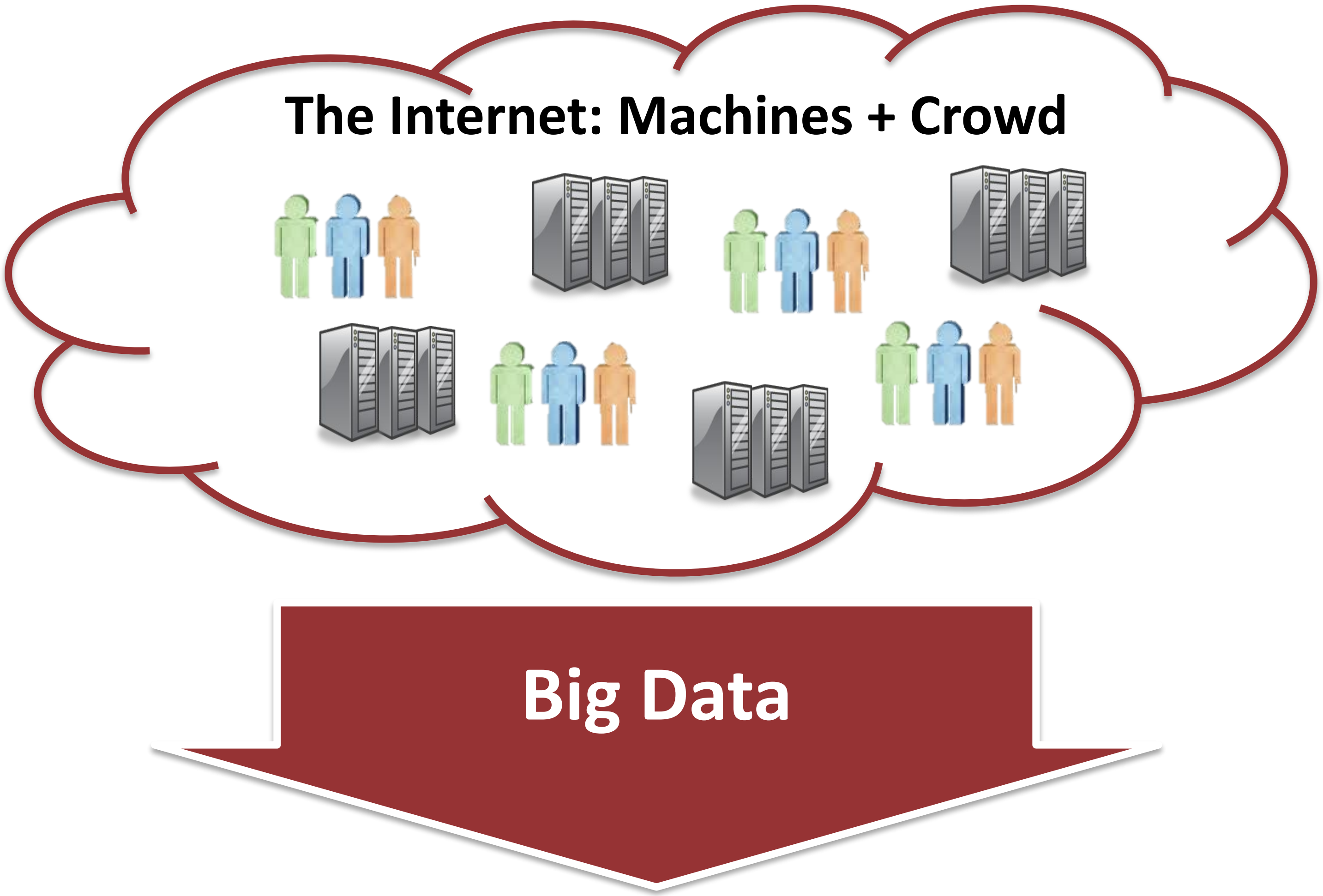
**Anon User**

2 votes by Anon User and Anon User

It looks like a Northern Trust Visa, which would make sense given his public disclosures report a banking relationship with the firm:



**What kind of credit card is President Obama using in this video of him donating to his campaign?**

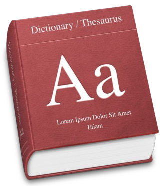



**Teach machines to recognize EVERYTHING**

**PASCAL VOC**

20

[Everingham et al.'06-'12]




10K+

[Biederman '87]

**Goal: Build a recognition engine on ~~EVERYTHING~~**

**10K classes**

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


SUN, **131K**
[Xiao et al. '10]

LabelMe, **37K**
[Russell et al. '07]

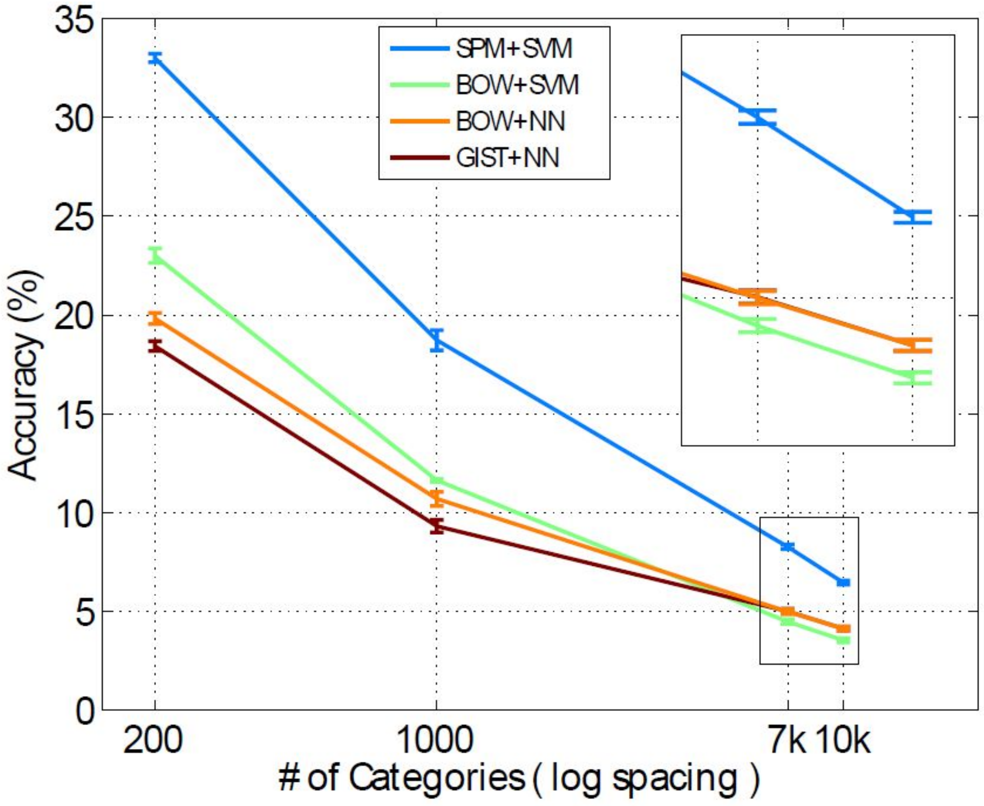
PASCAL VOC, **30K**
[Everingham et al. '06-'12]

Caltech101, **9K**
[Fei-Fei, Fergus, Perona, '03]

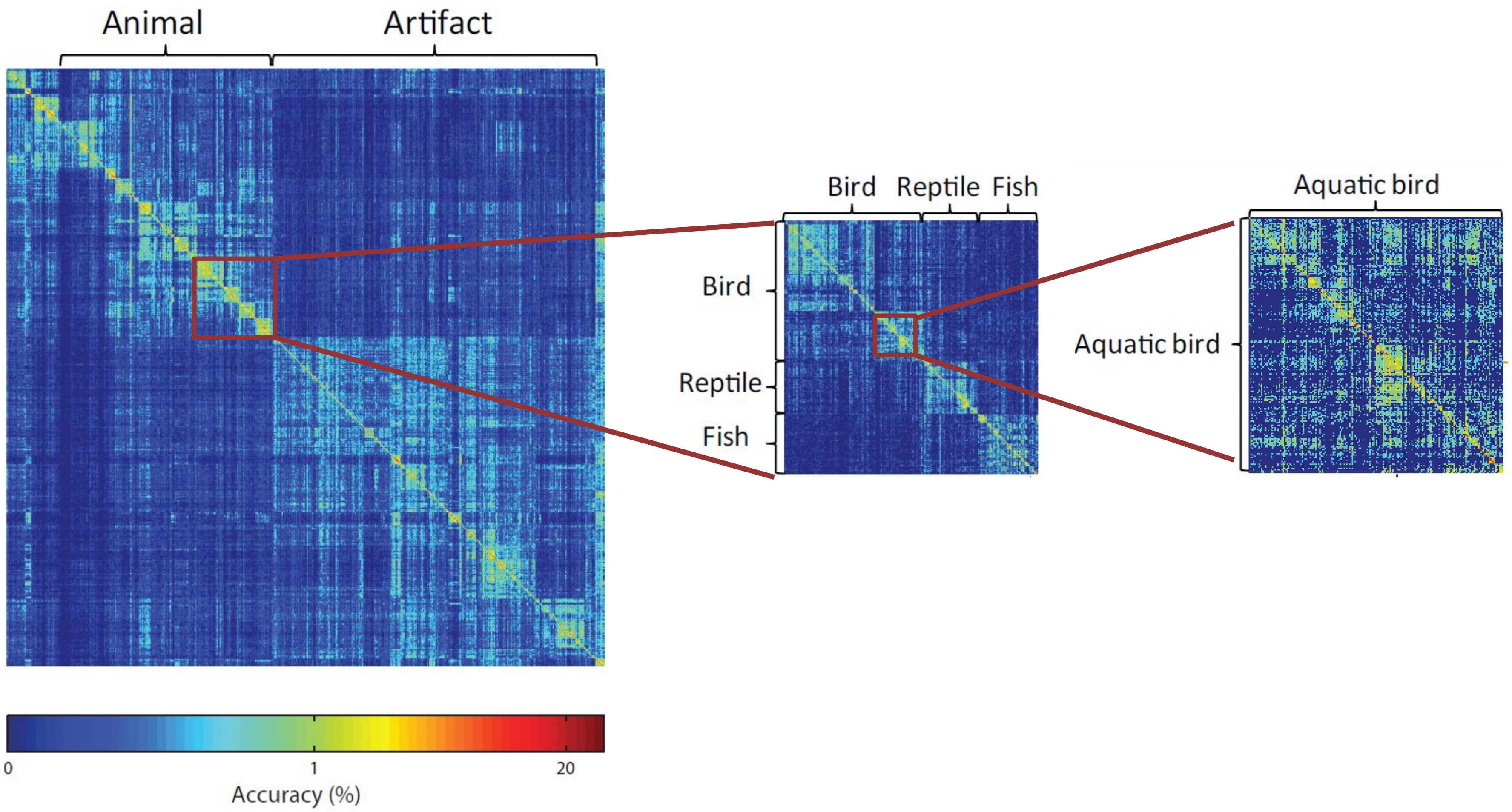
# Learn to Classify 10K Classes

- 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]

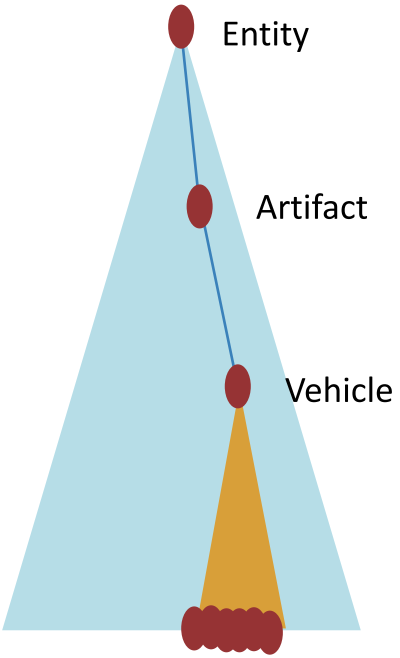
    - 6.4% for 10K categories

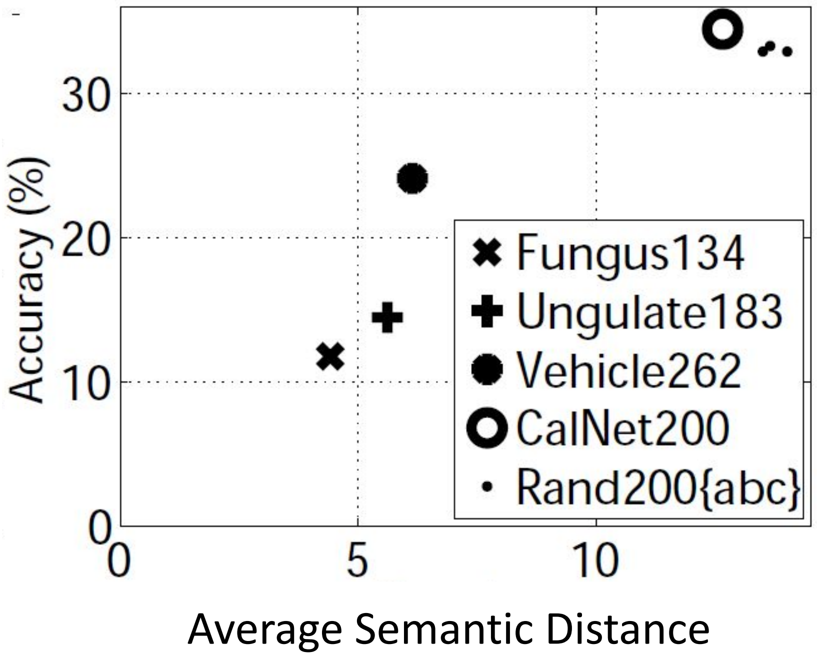


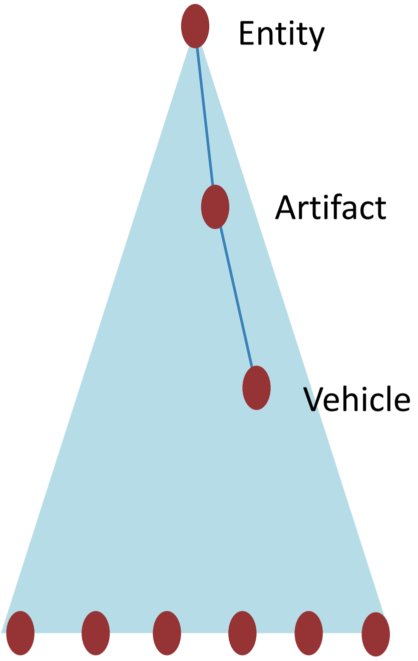
# Learn to Classify 10K Classes

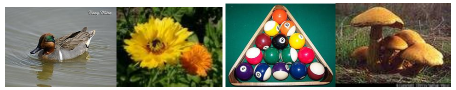


# Fine-grained categories are a lot harder



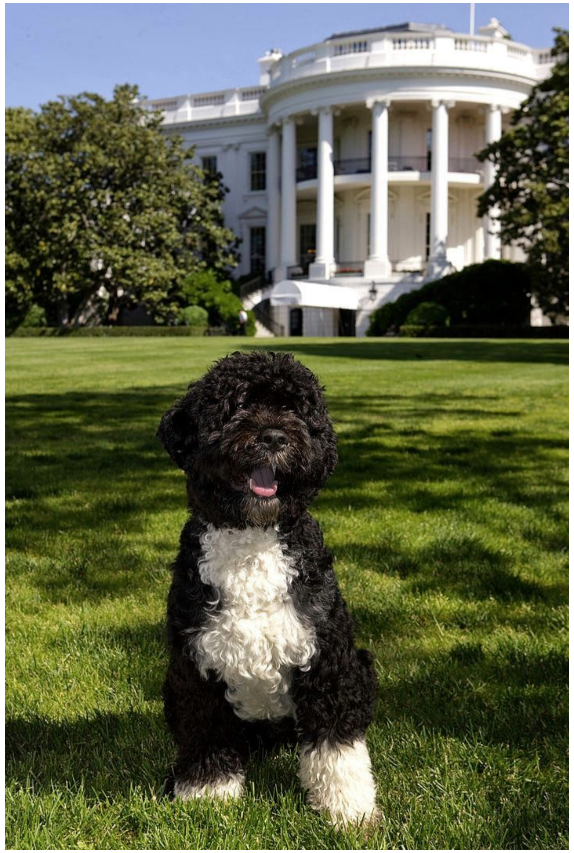



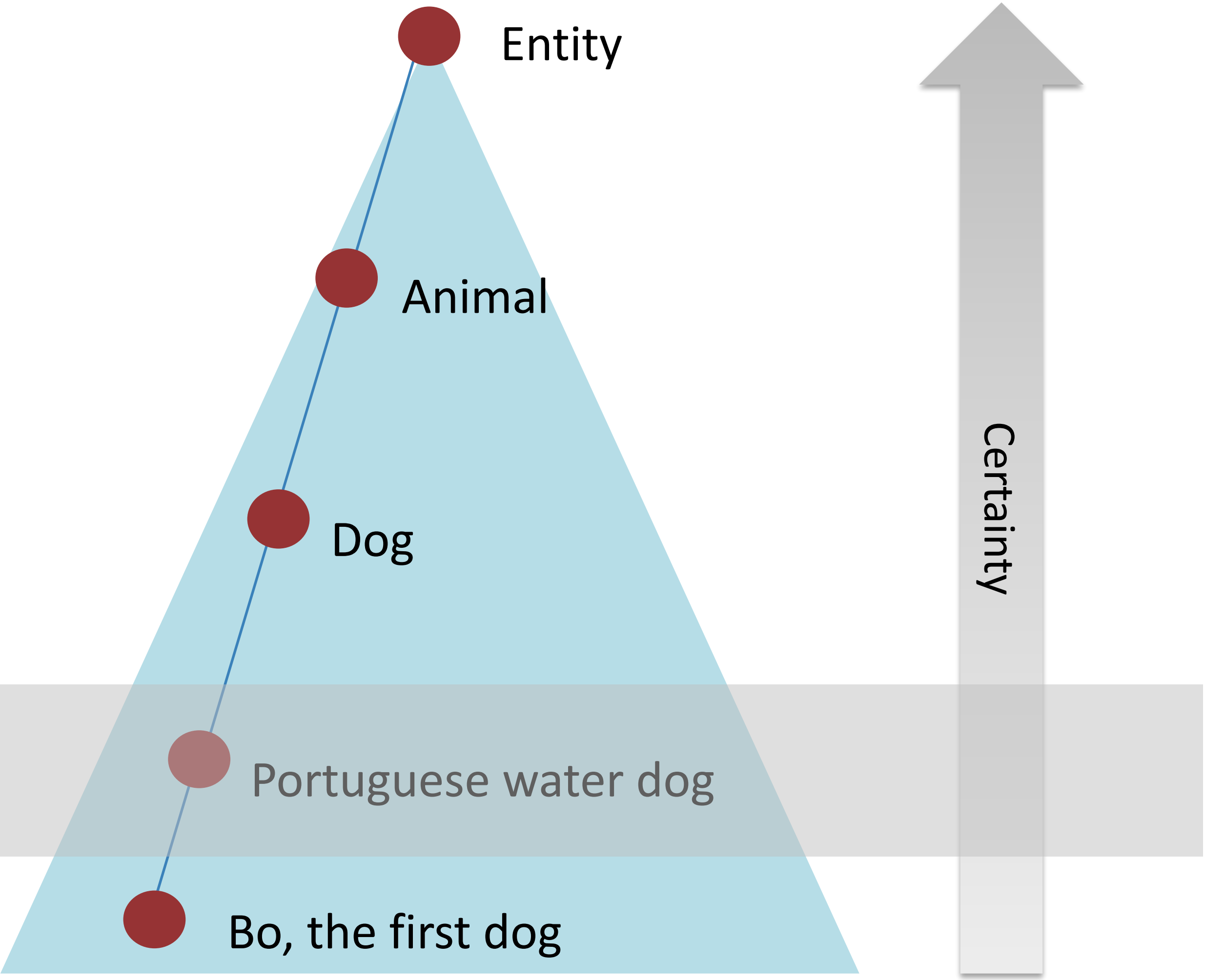



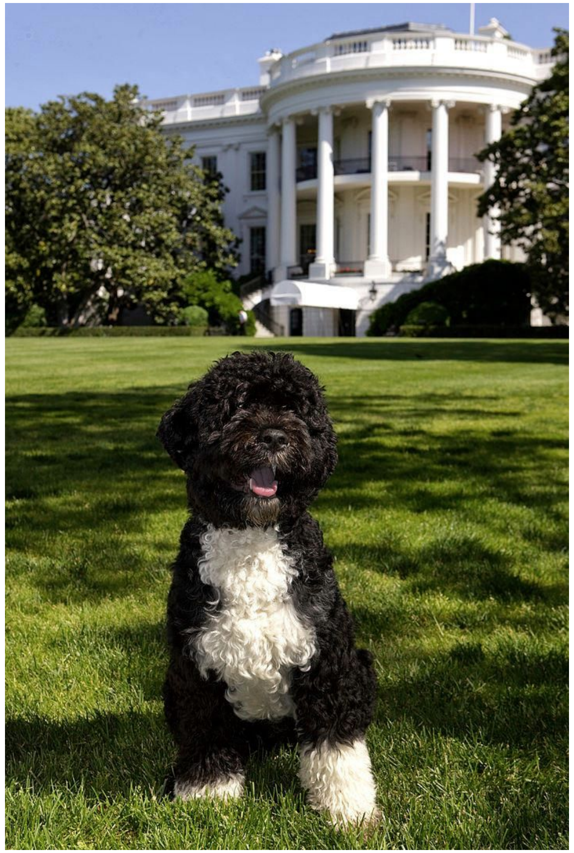



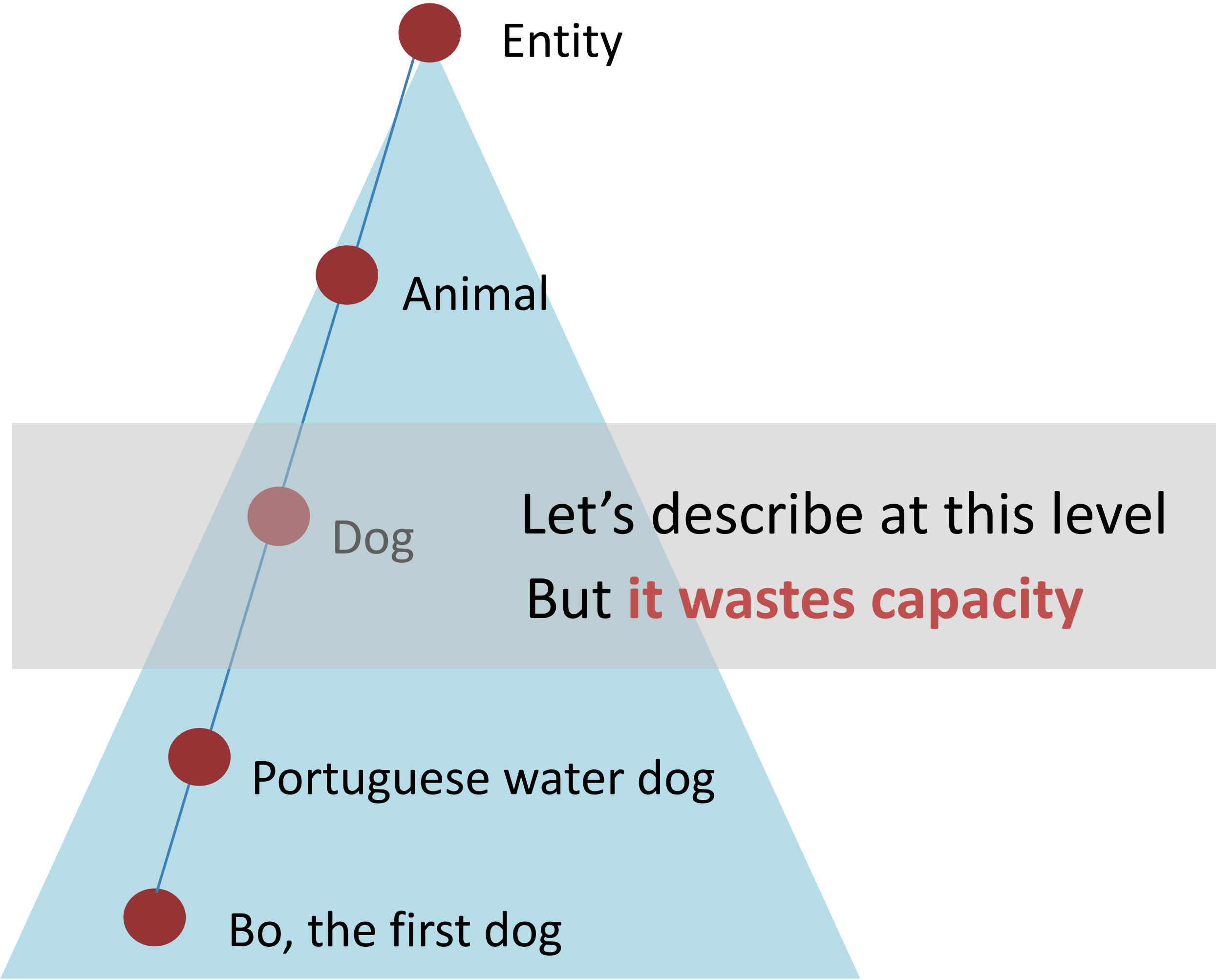
## Challenges

- Semantic hierarchy
- Fine-grained classes

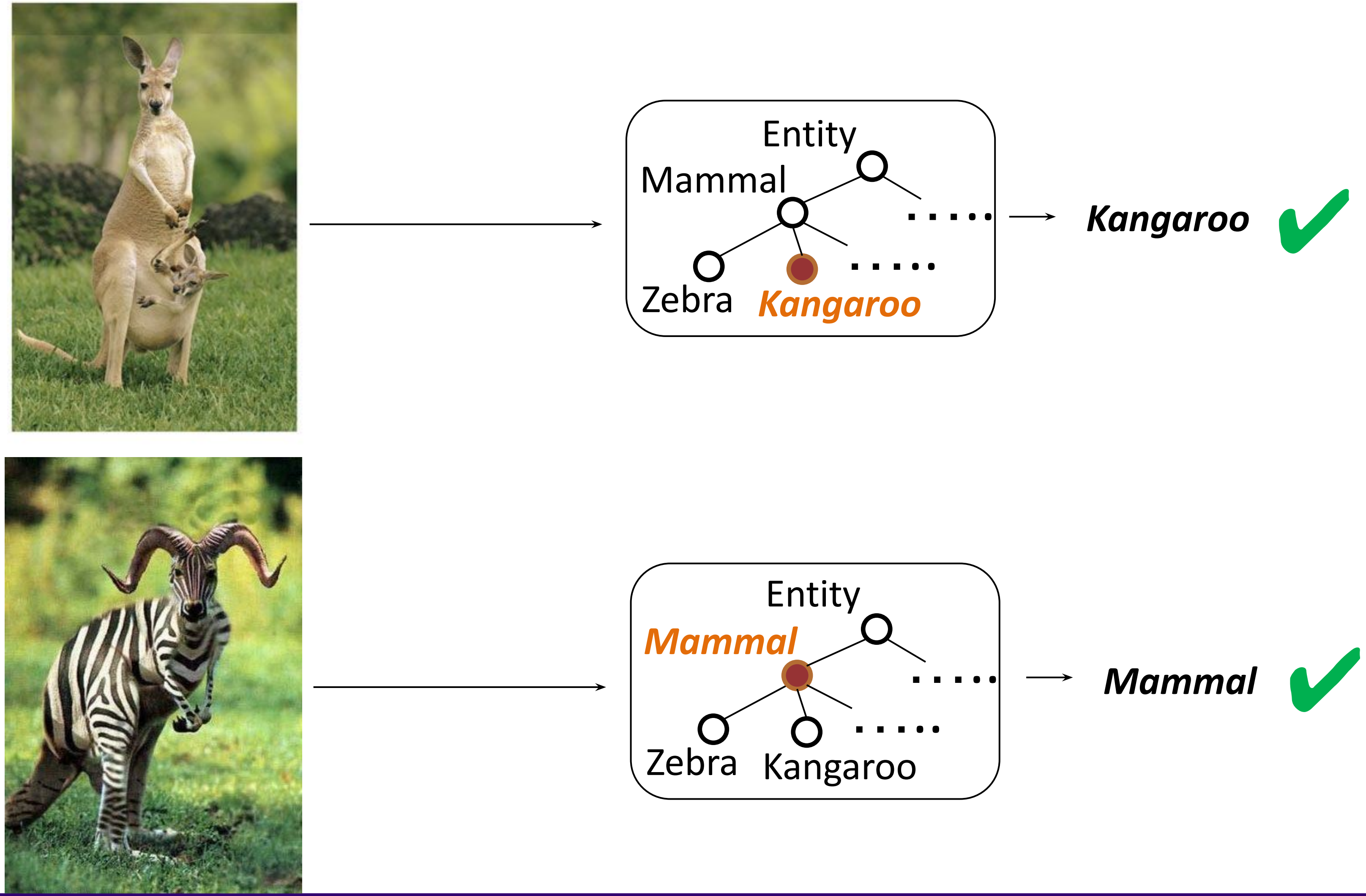



Entity
Animal
Dog
Portuguese water dog
Bo, the first dog
Certainty

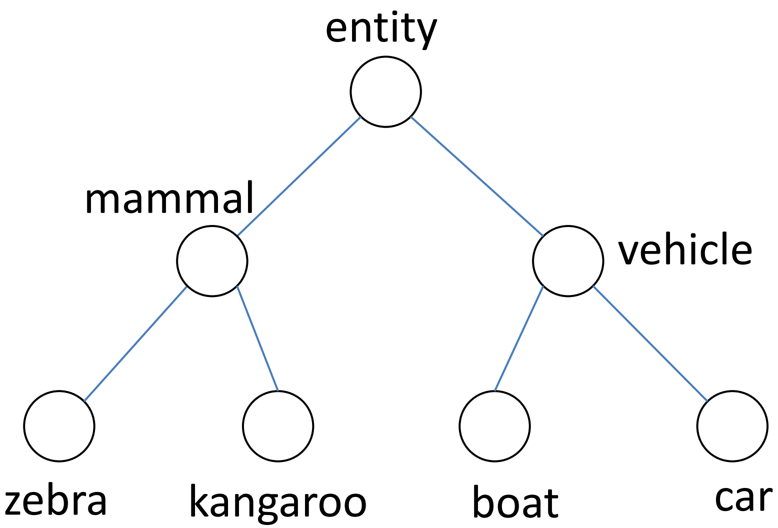



Entity
Animal
Dog
Let’s describe at this level
But it wastes capacity
Portuguese water dog
Bo, the first dog

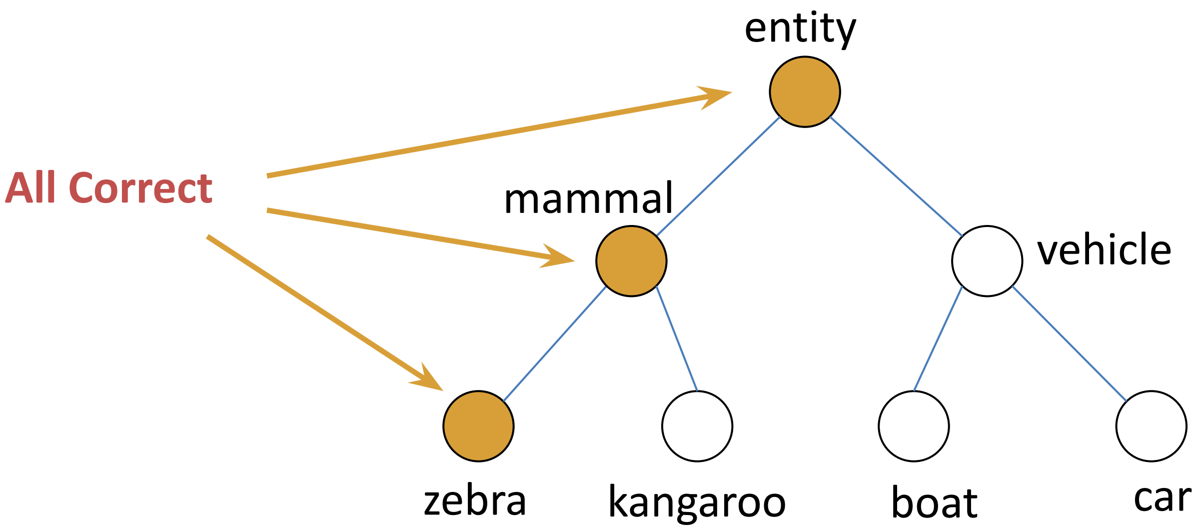
# Hedging: Be as informative as possible with few mistakes



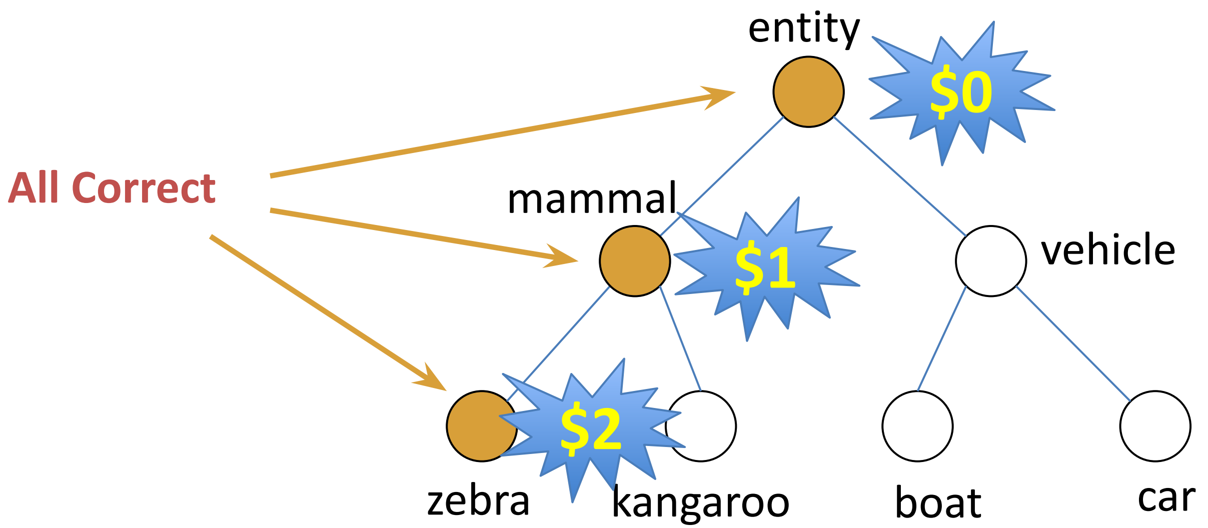
# Formal Problem Statement



# Formal Problem Statement




# Formal Problem Statement




# Formal Problem Statement

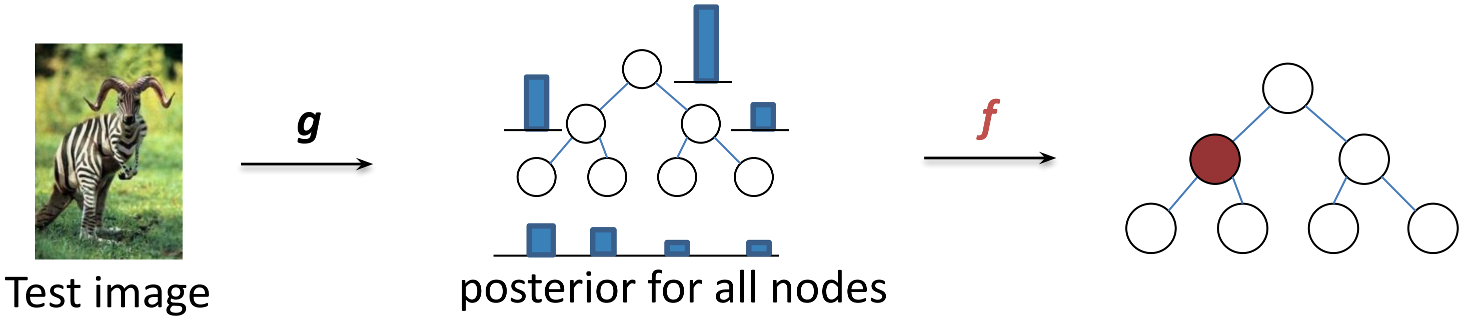
## Assumptions

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

## Goal

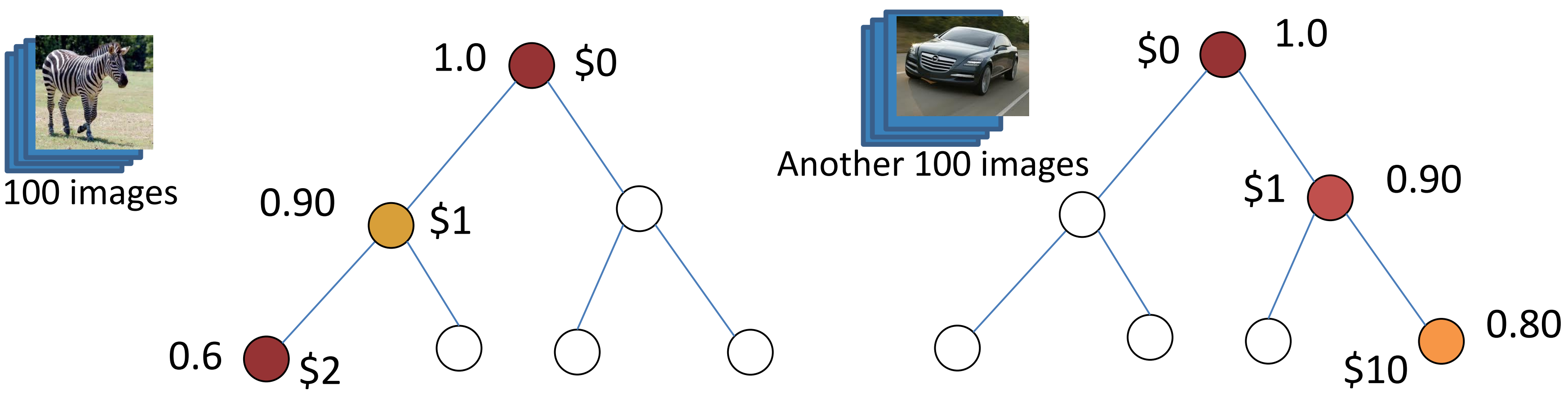
- Find a ***decision rule f***
  - Expected accuracy ***A(f)*** is at least ***1-ε***
  - Maximize expected reward ***R(f)***

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



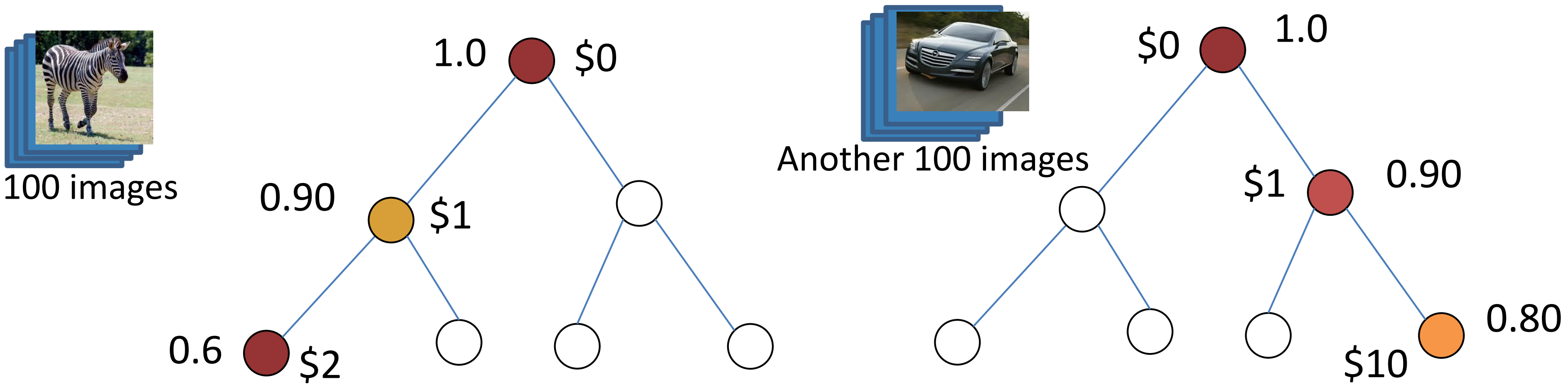

Deng, Krause, Berg, Fei-Fei, CVPR2012

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

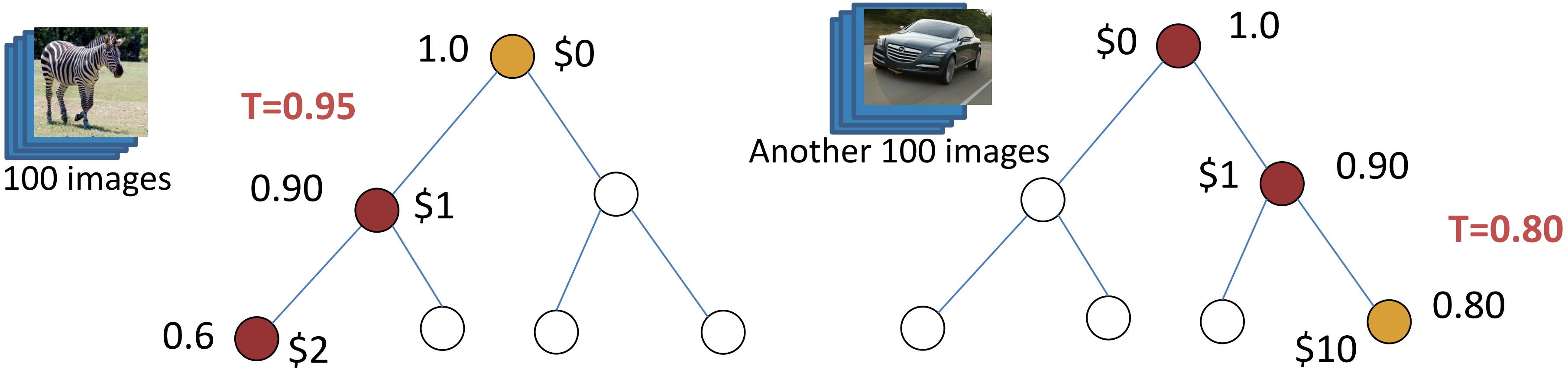



Reward = ($1 * 0.90 + $1 * 0.90) / 2 = $0.90
Accuracy = (0.90 + 0.90 ) / 2 = 0.90

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

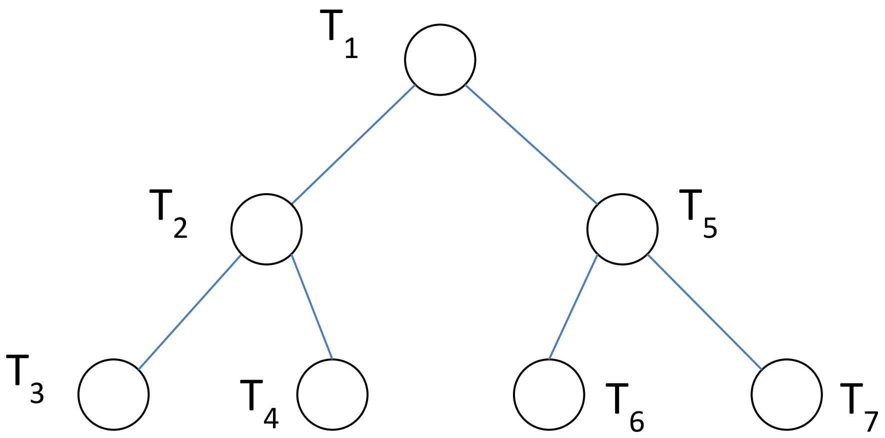



Reward = ($1 * 0.90 + $1 * 0.90) / 2 = $0.90
Accuracy = (0.90 + 0.90 ) / 2 = 0.90




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

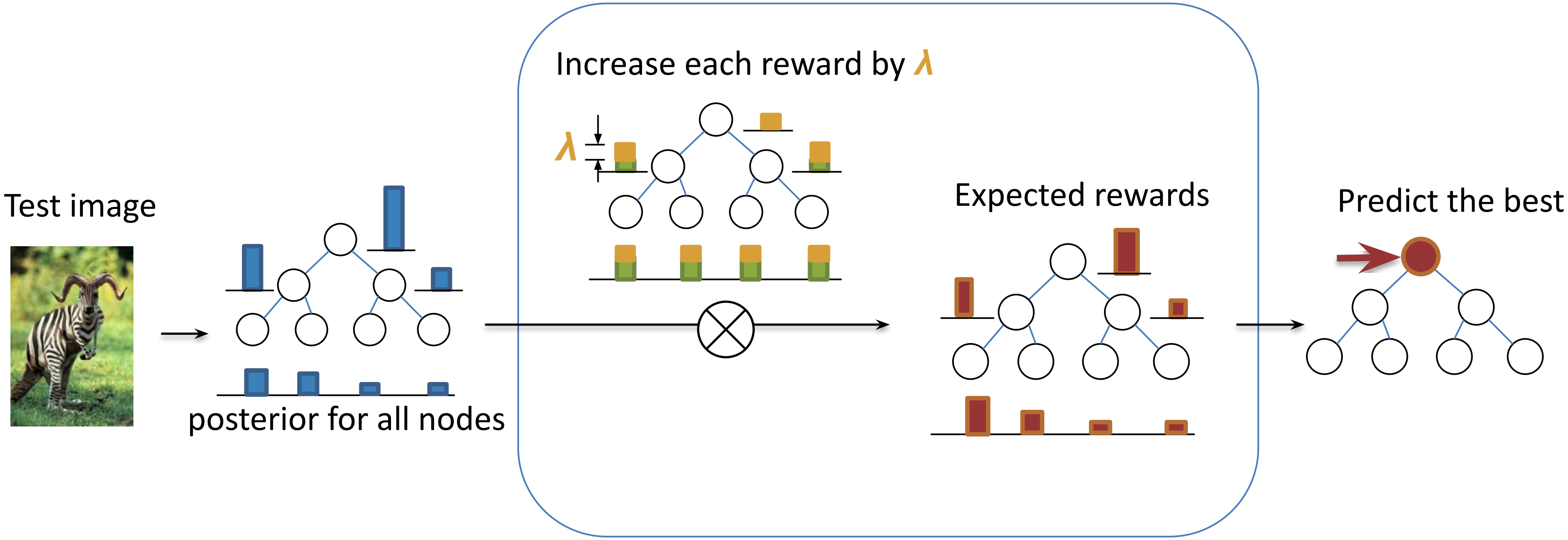
Deng, Krause, Berg, Fei-Fei, CVPR20

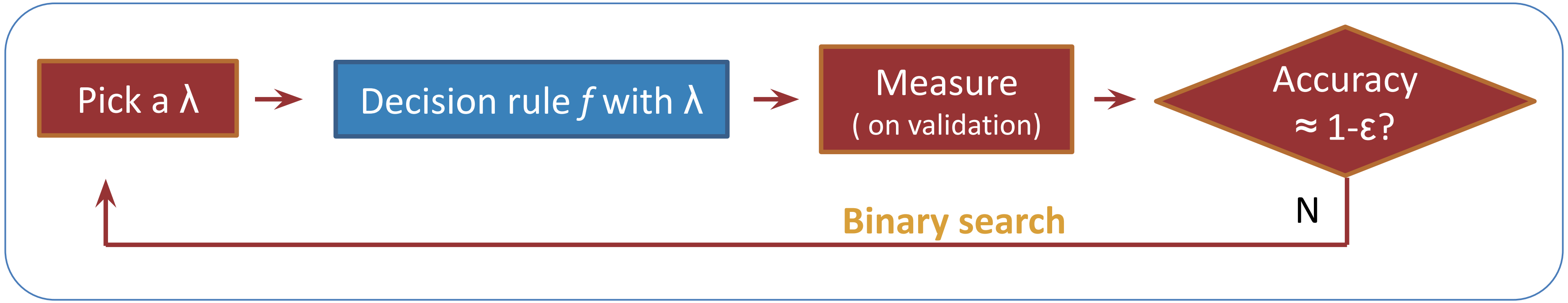


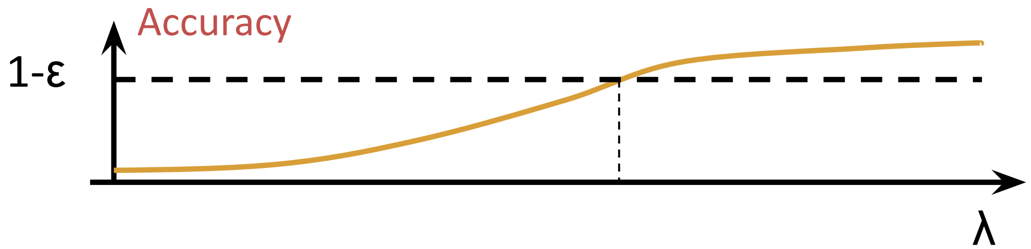

**We can optimize individual thresholds…**

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

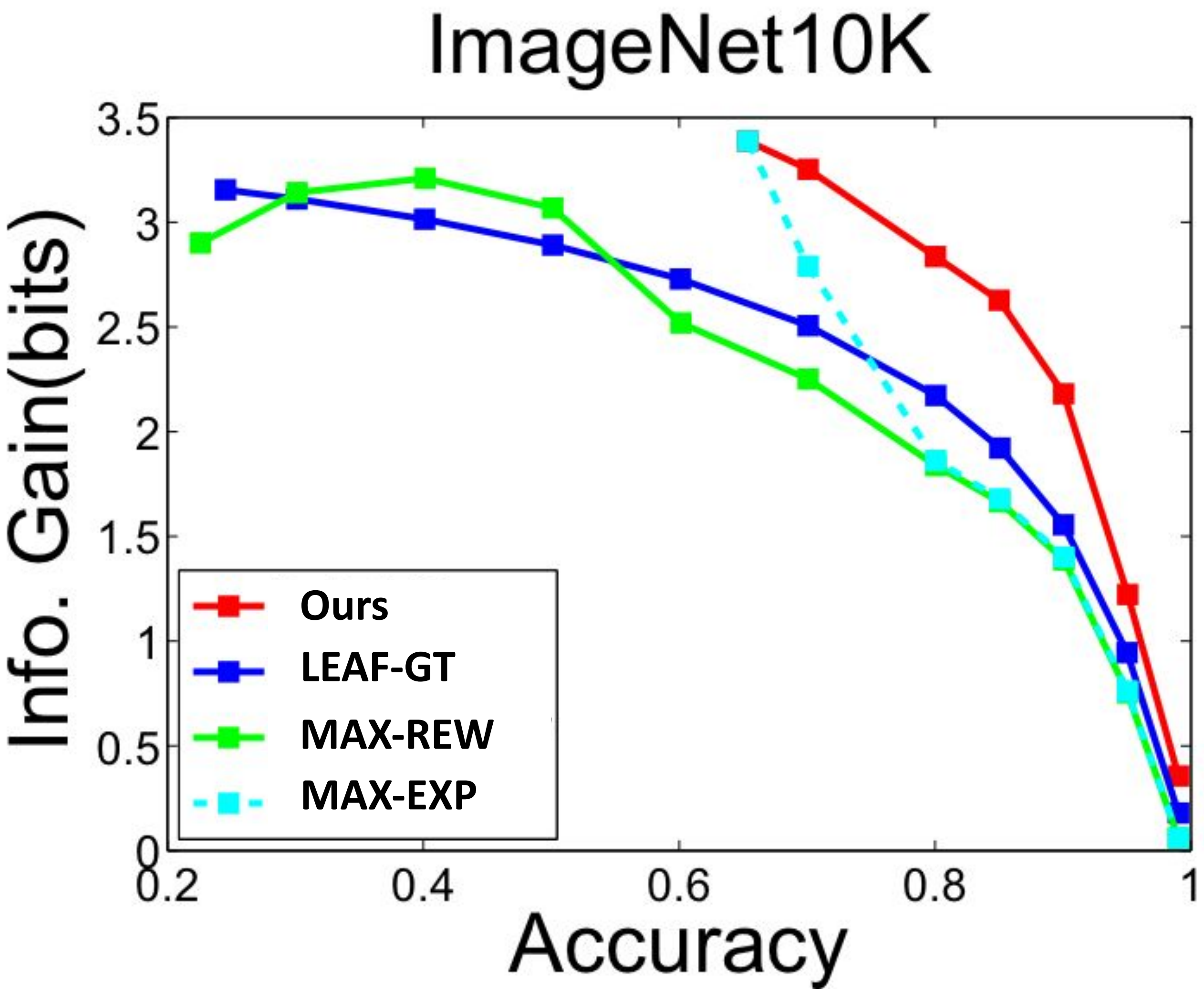
# A global, fixed scalar parameter λ≥0





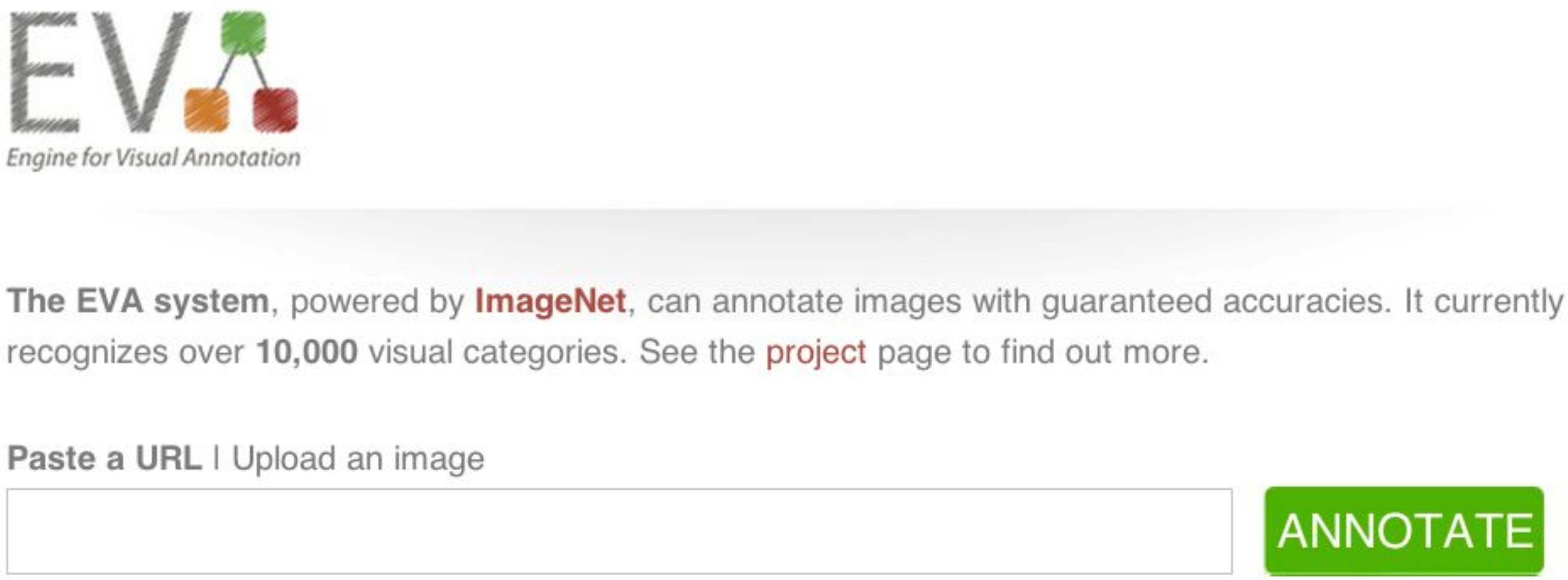




**Theorem: Under very mild conditions, this is optimal.**

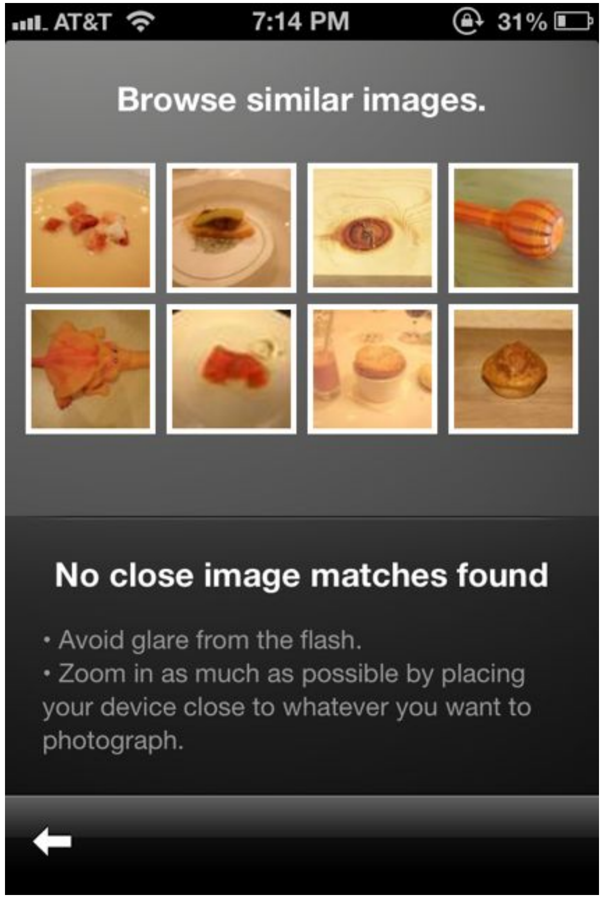



Deng, Krause, Berg, Fei-Fei, CVPR2012

www.image-net.org/eva



Google Goggles
Use pictures to search the web.

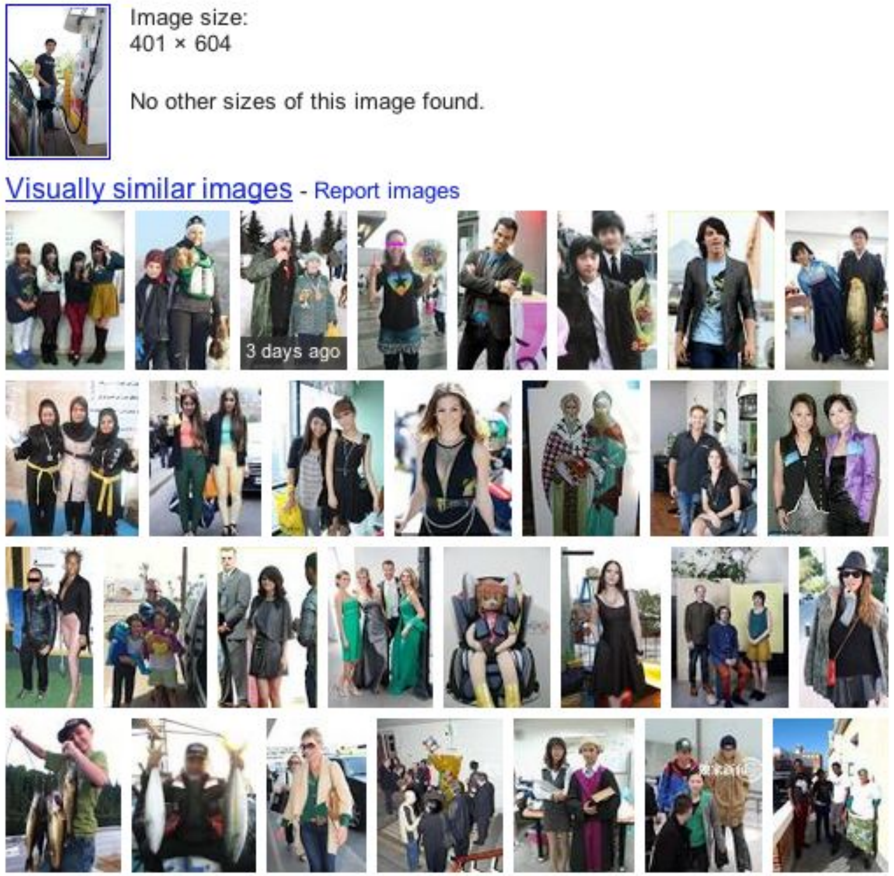



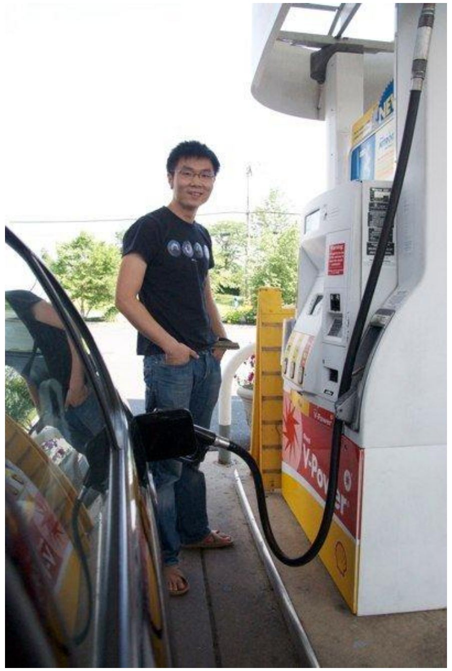

EVA
Engine for Visual Annotation

0.95 coffee mug
0.97 mug
0.99 drinking vessel

Google
images

EV
Engine for Visual Annotation



0.87 face , gas pump, person

0.90 face , gas pump

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

EV

Engine for Visual Annotation

0.78 person, instrument

0.84 person

## Challenges

- Semantic hierarchy
- Fine-grained classes

**Next Recitation!**