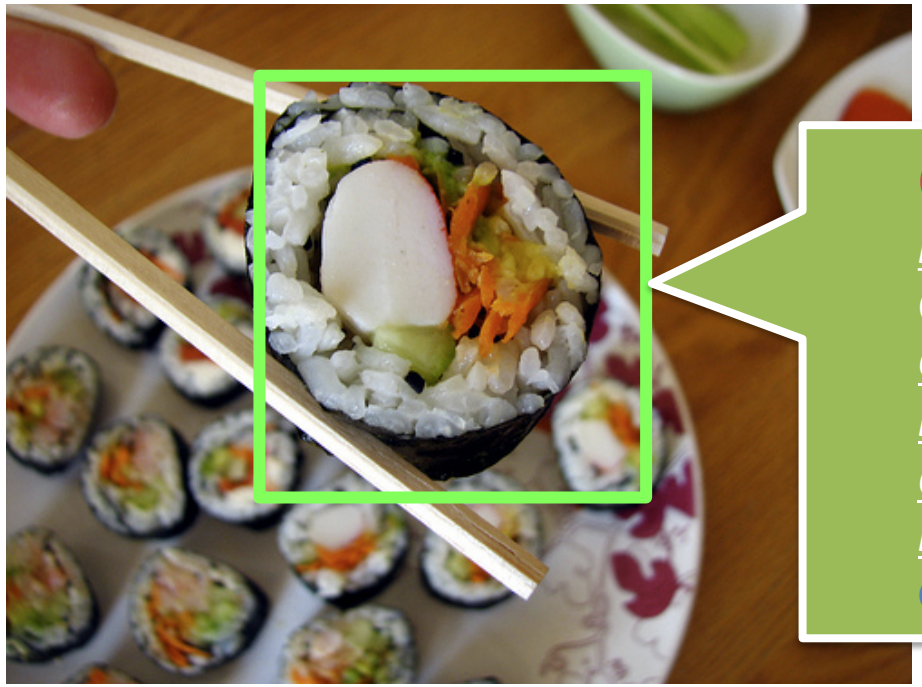
Lecture: Large Scale Recognition & Visual Ontologies

Juan Carlos Niebles and Ranjay Krishna
Stanford Vision and Learning Lab

CS 131 Roadmap

Pixels	Segments	Images	Videos	Web
Convolutions Edges Descriptors	Resizing Segmentation Clustering	Recognition Detection Machine learning	Motion Tracking	Neural networks Convolutional neural networks



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

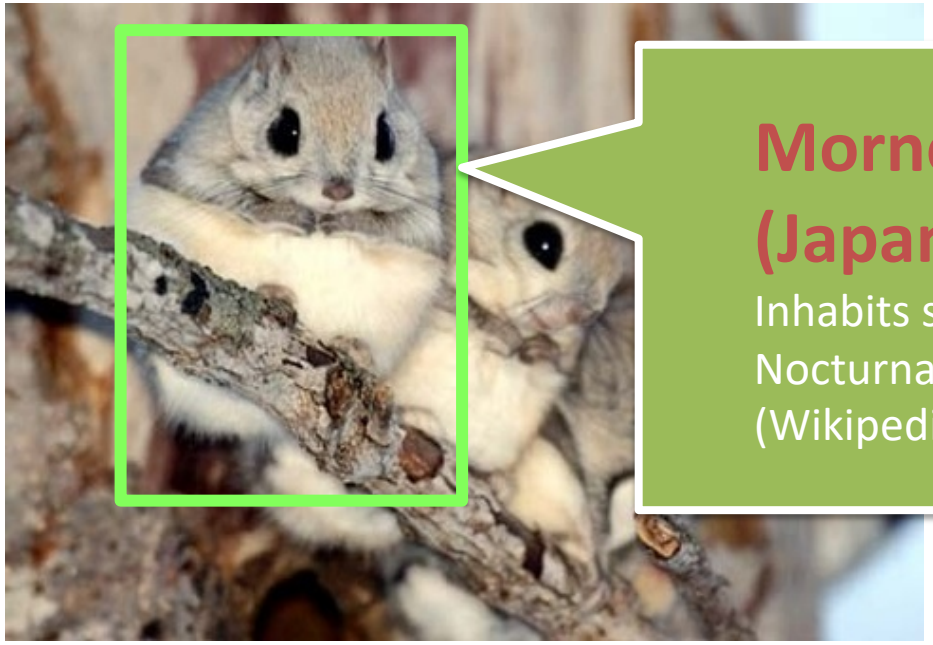
DO NOT RUN

Raise arms to appear larger.
Show your teeth

IKEA POANG Chair

ON SALE

$29.00 at ikea.com



Mornonga
(Japanese flying squirrel)

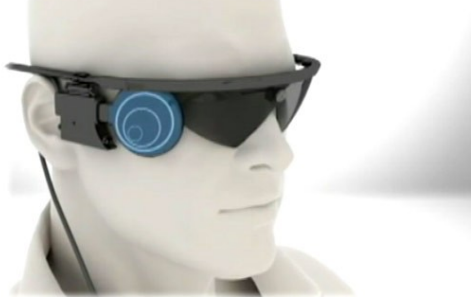
Inhabits sub-alpine forests in Japan.
Nocturnal. Eats seeds, fruit, tree leaves
(Wikipedia)

I wish my computer could recognize

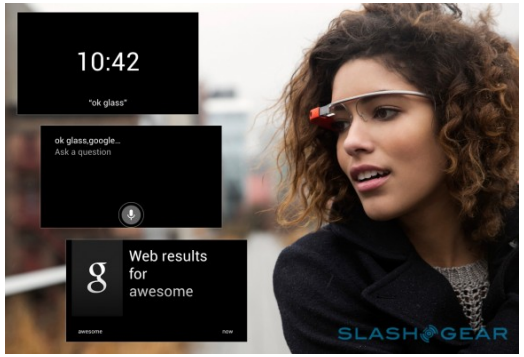
EVERYTHING

Surveillance

Robotics



Assistive tools



Wearable devices

Recognition Engine

Driverless cars

Smart photo album

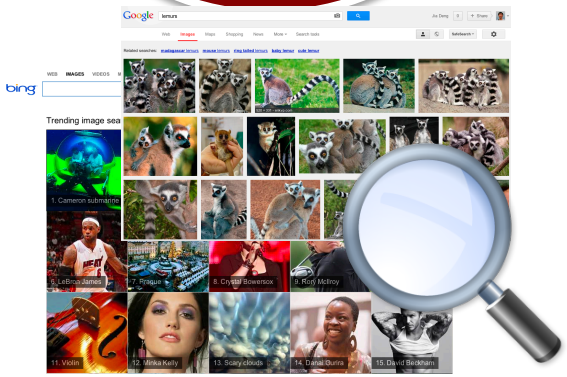


Image search

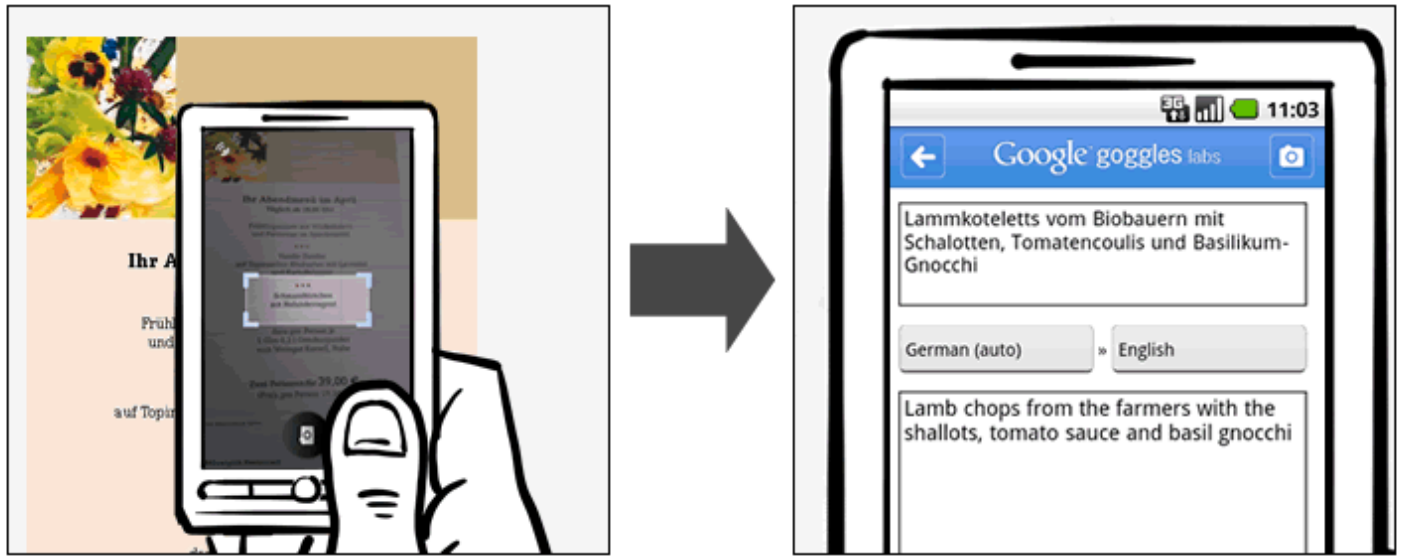
Mining social media

What can computers already recognize?

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

Google Goggles

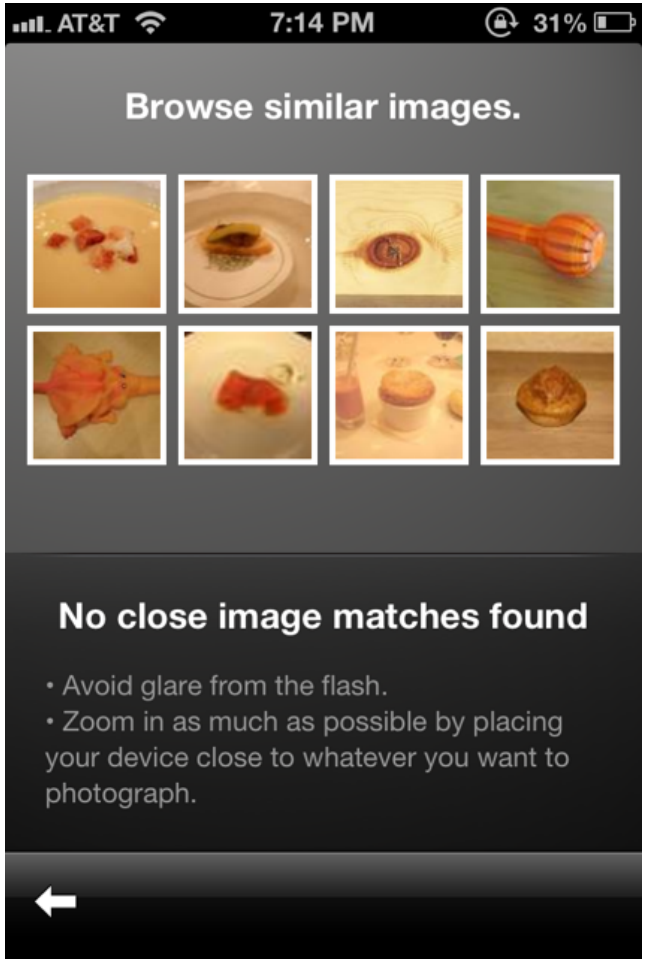
Use pictures to search the web.



What’s the next to work on?

Coffee Mugs!

BOSCH
TASSIMO

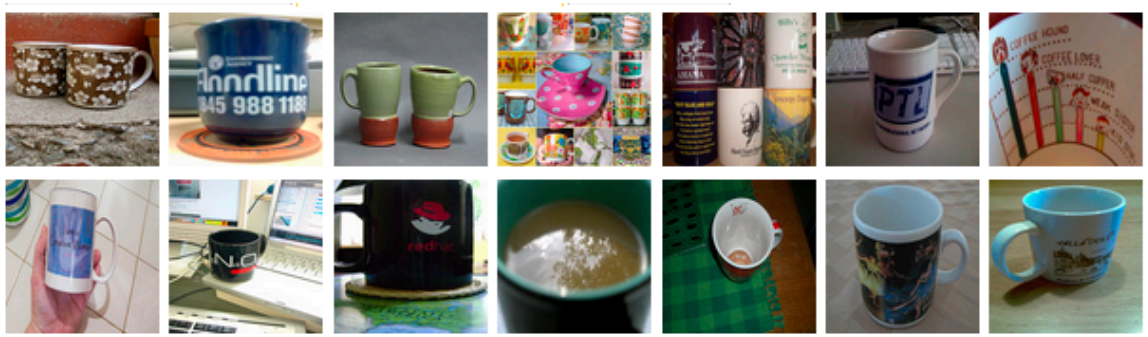

AT&T
7:14 PM
31%
Browse similar images.
No close image matches found
• Avoid glare from the flash.
• Zoom in as much as possible by placing your device close to whatever you want to photograph.

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

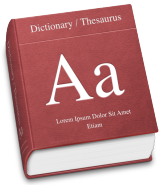
Spacesuit

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

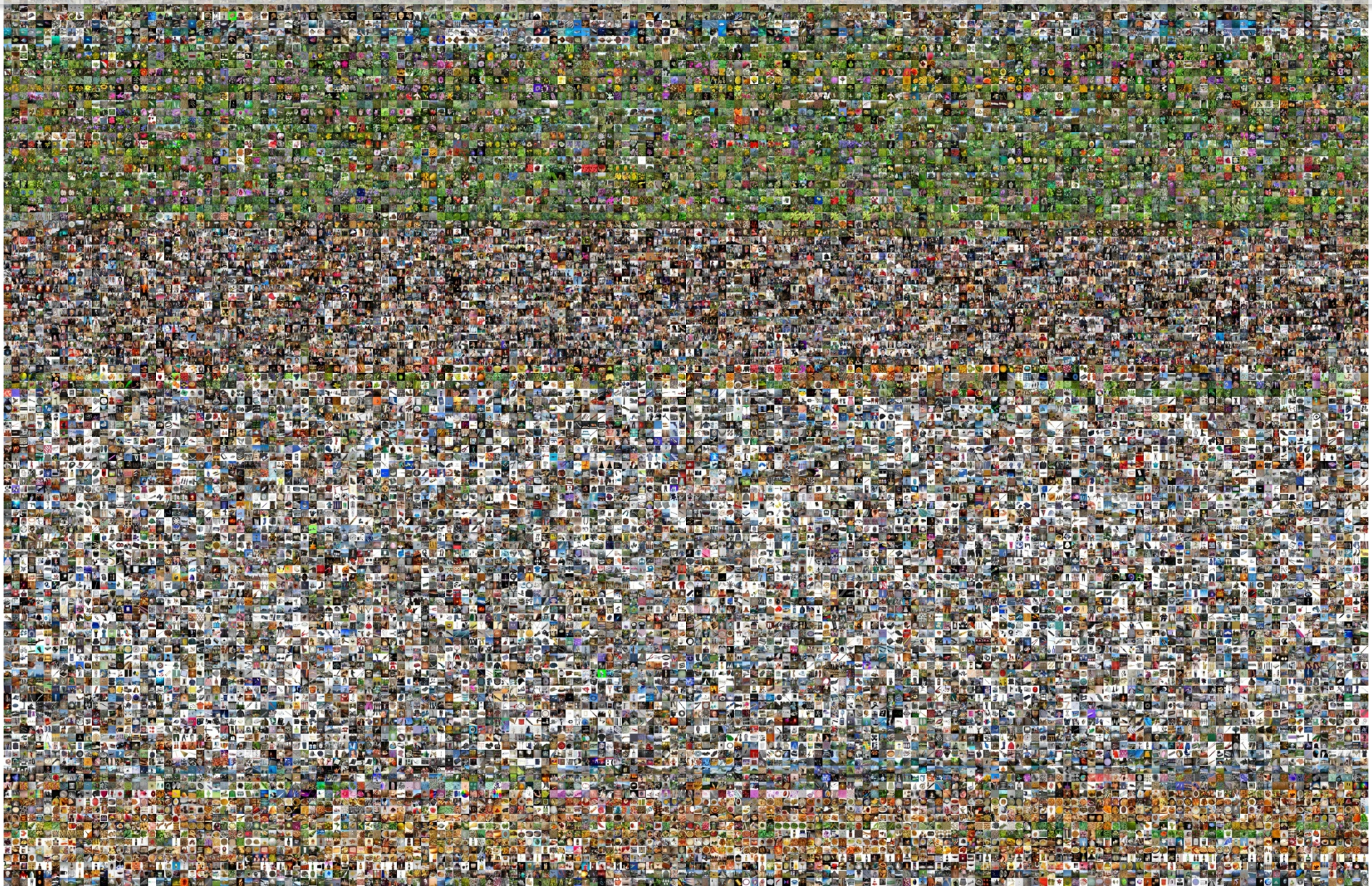
WordNet

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

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

4.1M+
articles

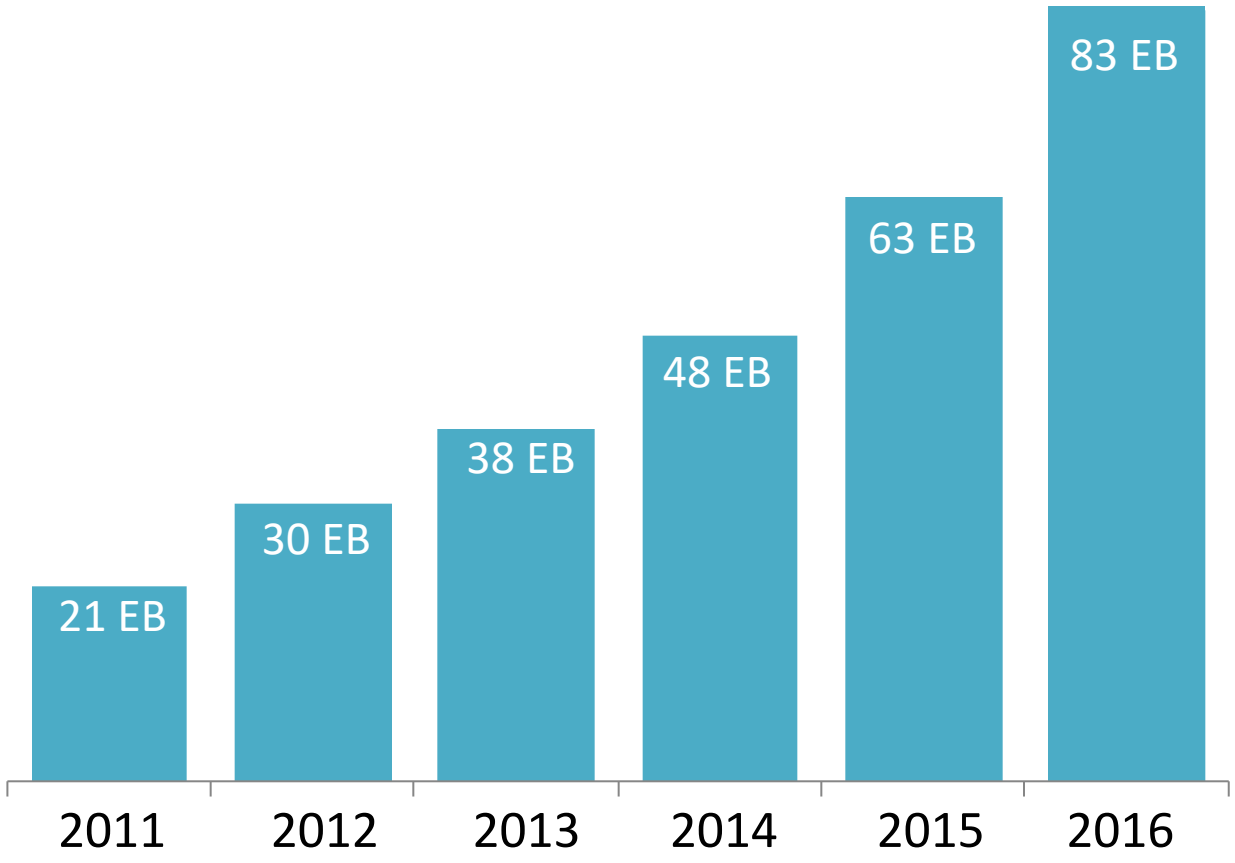
From 20 classes to Millions?



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

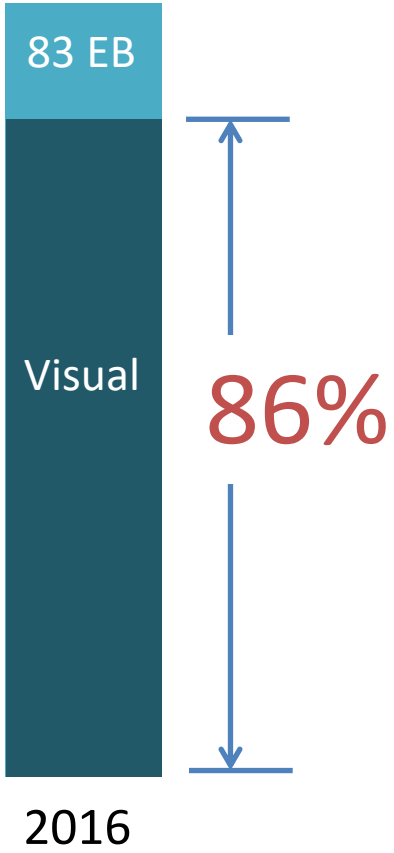
Global Consumer Internet Traffic Per Month



Source: Cisco

YouTube 72 hours of videos / min

facebook 300 million images / day



Source: Cisco

Big Data from the Internet

→ The Internet can teach EVERYTHING

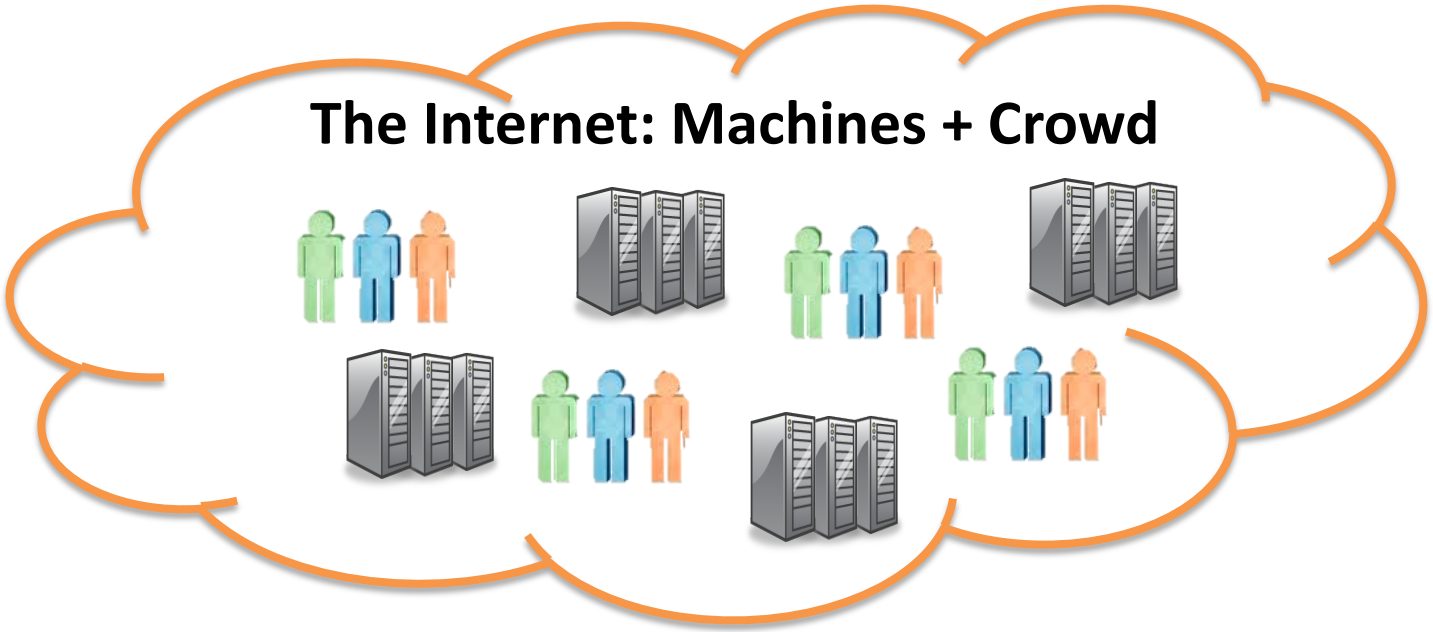
Google pitbullfrog

Evolution Gone Wild

Future plants and animals

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

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

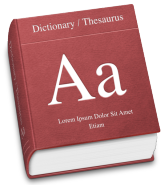


Teach machines to recognize EVERYTHING

PASCAL VOC

20

[Everingham et al.'06-'12]



10K+

[Biederman '87]

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

10K classes

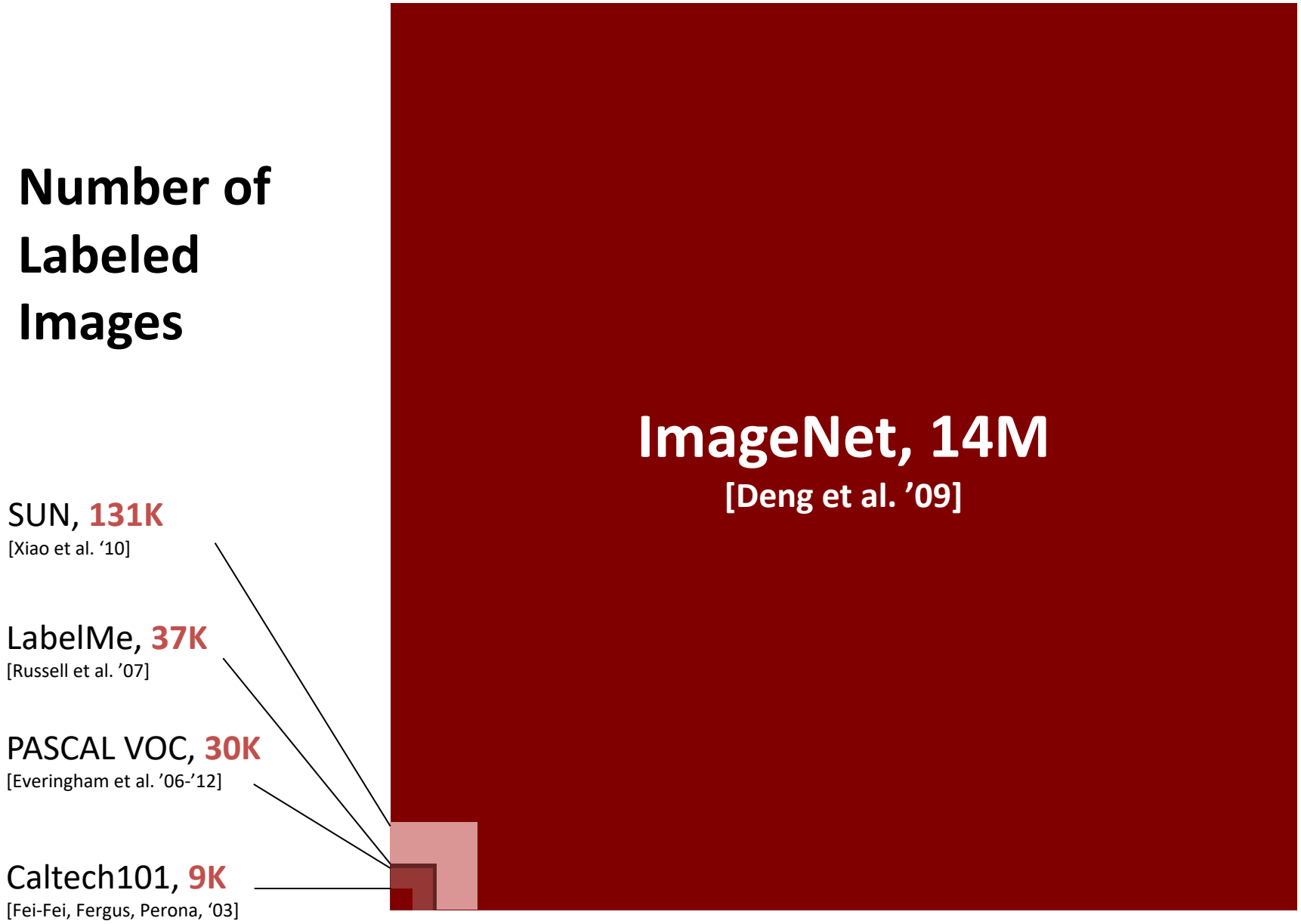
IMAGENET [Deng et al. 2009]

www.image-net.org

22K categories and **14M** images

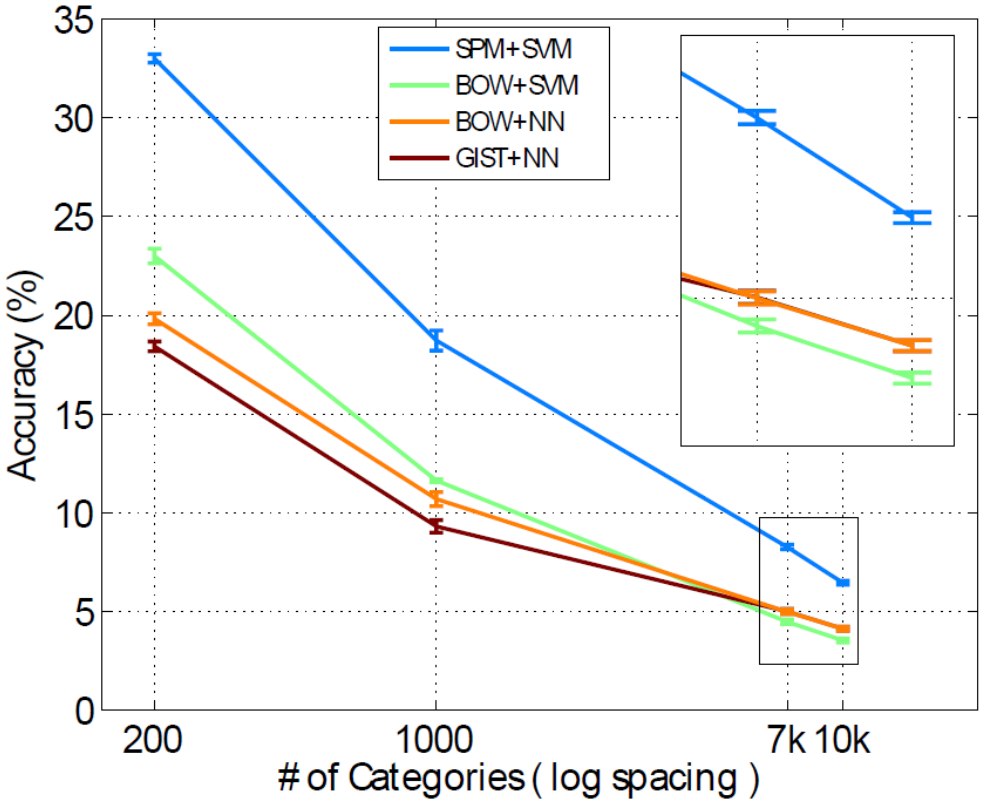
- Animals
  - Bird
  - Fish
  - Mammal
  - Invertebrate
- Plants
  - Tree
  - Flower
- Food
- Materials
- Structures
- Artifact
  - Tools
  - Appliances
  - Structures
- Person
- Scenes
  - Indoor
  - Geological Formations
- Sport Activities

Number of Labeled Images



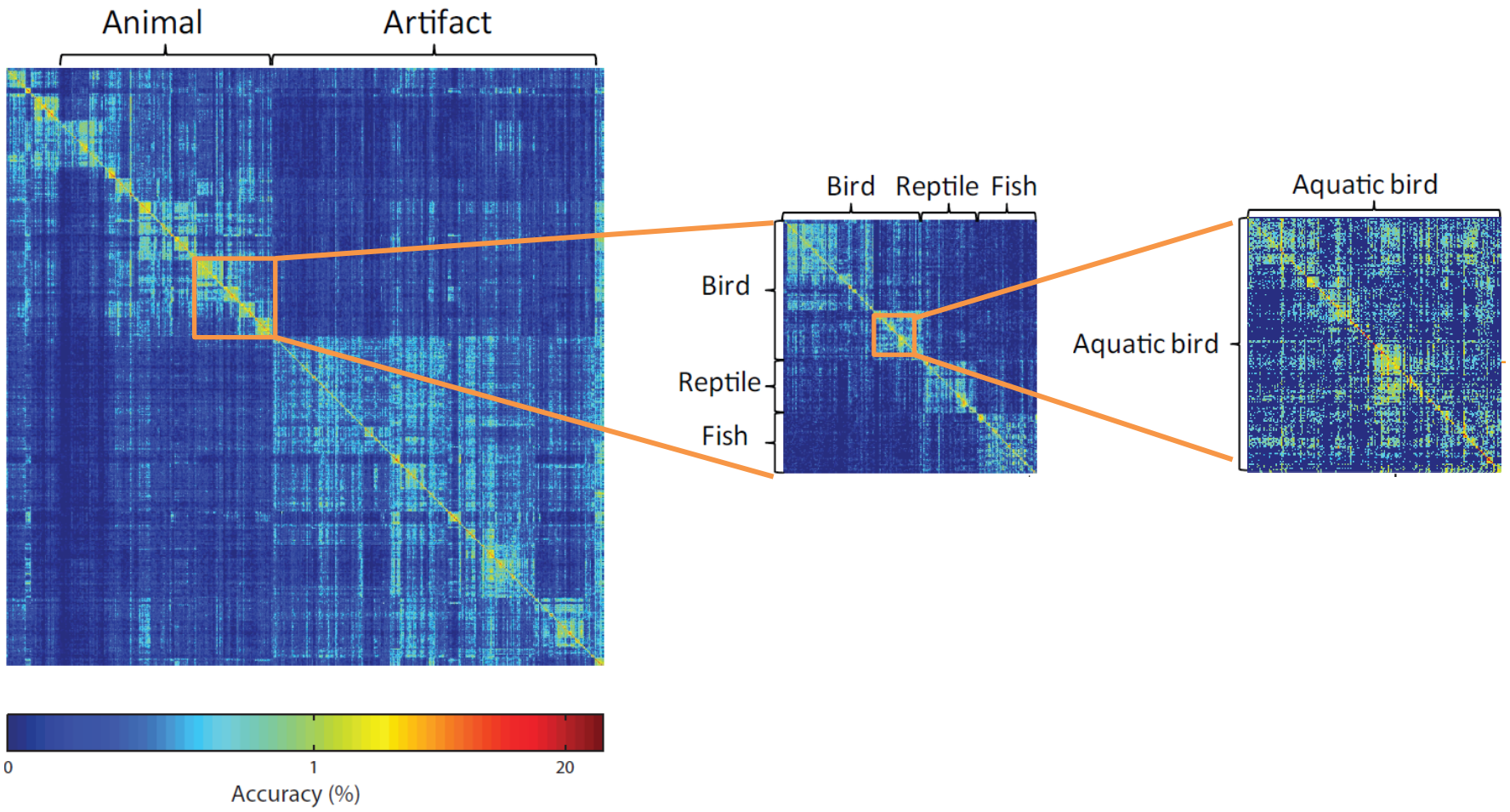
Learn to Classify 10K Classes

- 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



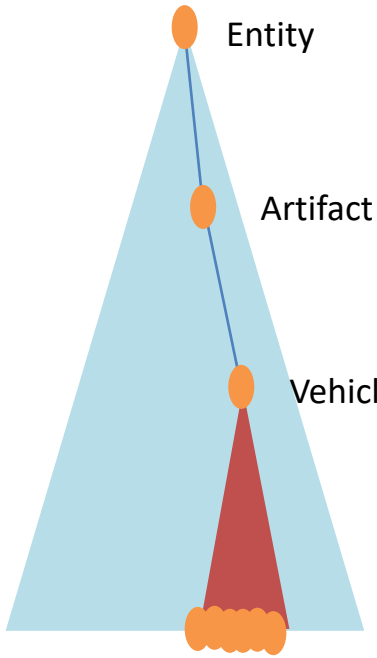
Deng, Berg, Li, & Fei-Fei, *ECCV2010*

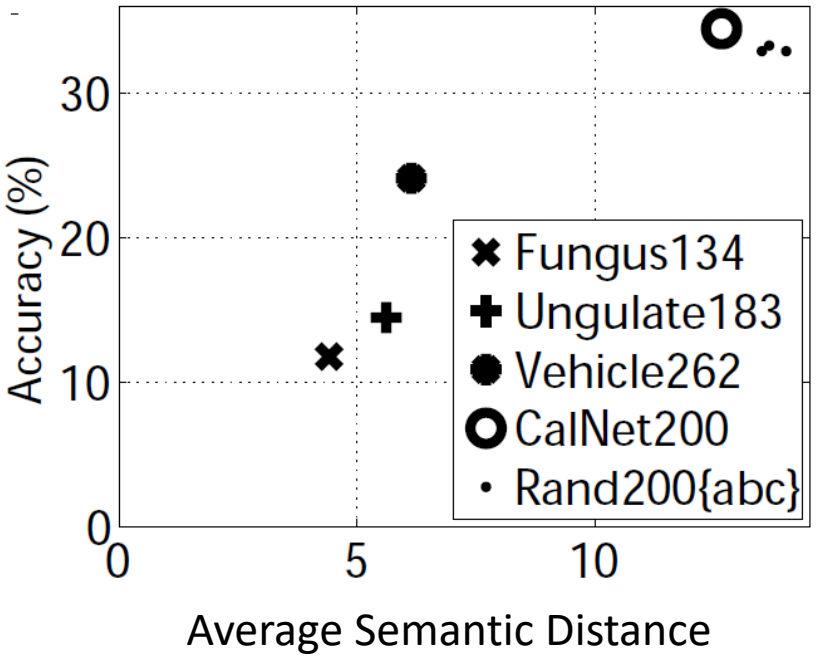
Learn to Classify 10K Classes

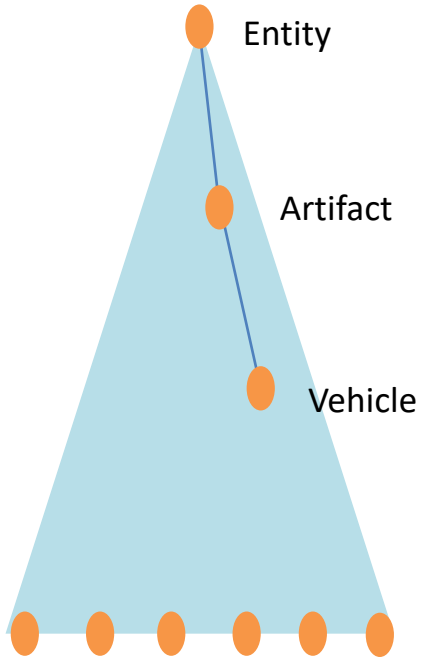


Deng, Berg, Li, & Fei-Fei, *ECCV2010*

Fine-grained categories are a lot harder





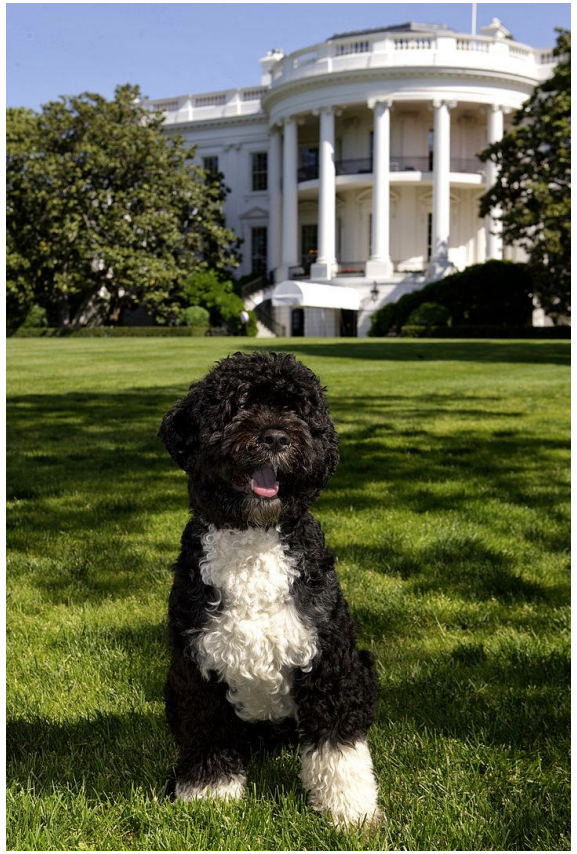


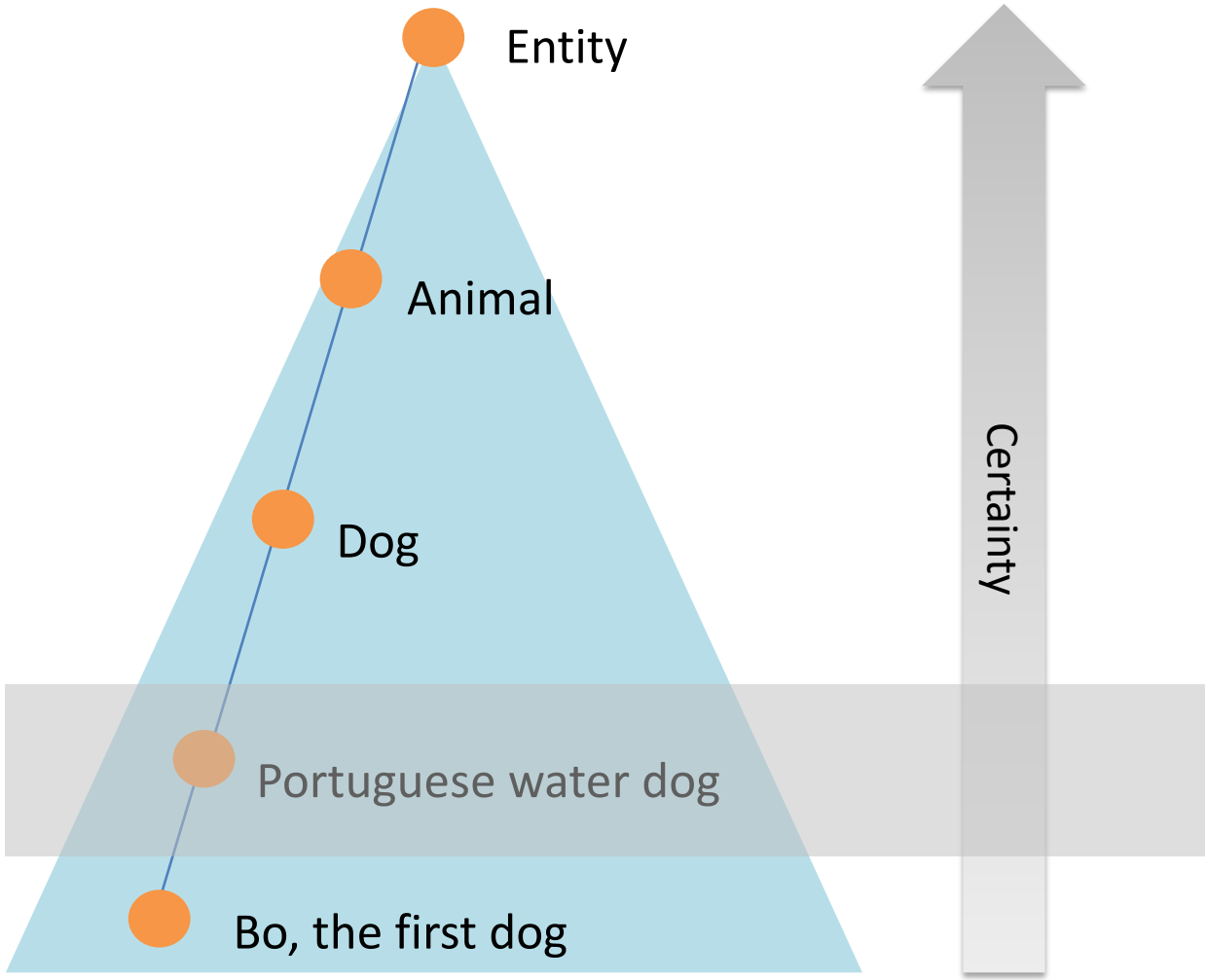
Finer **Coarser**

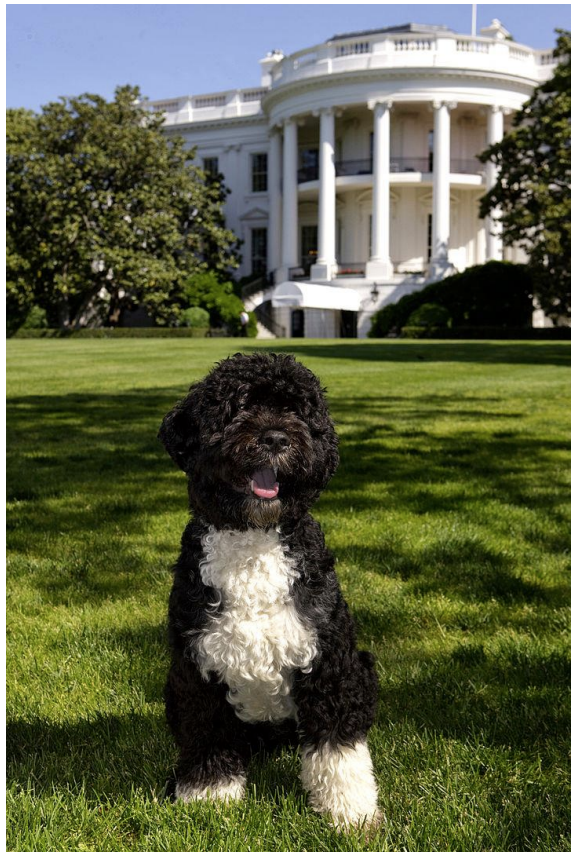
Deng, Berg, Li, & Fei-Fei, *ECCV2010*

Challenges

- Semantic hierarchy
- Fine-grained classes
- Large-scale Learning



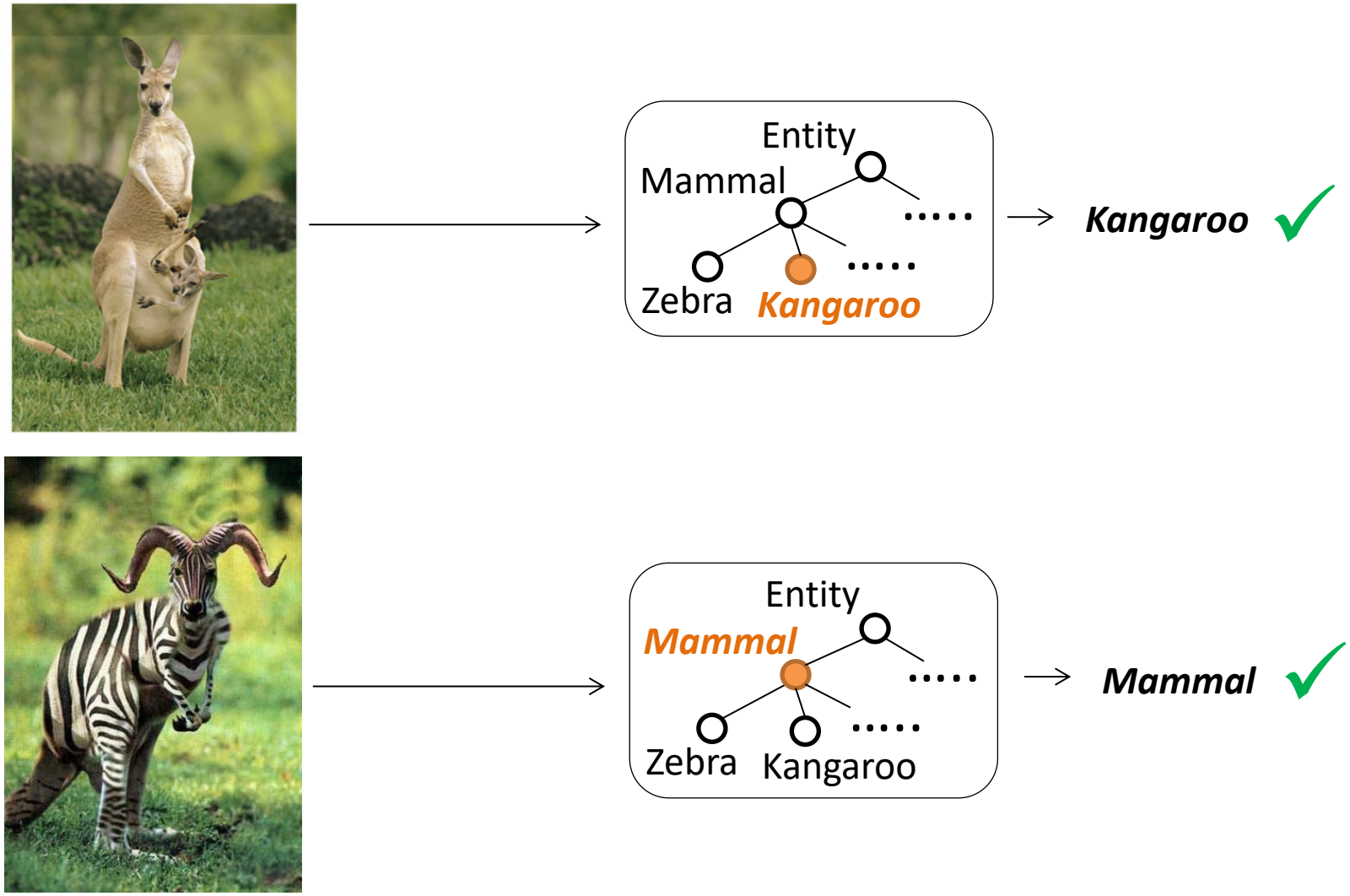

Entity
Animal
Dog
Portuguese water dog
Bo, the first dog
Certainty



Let's describe at this level

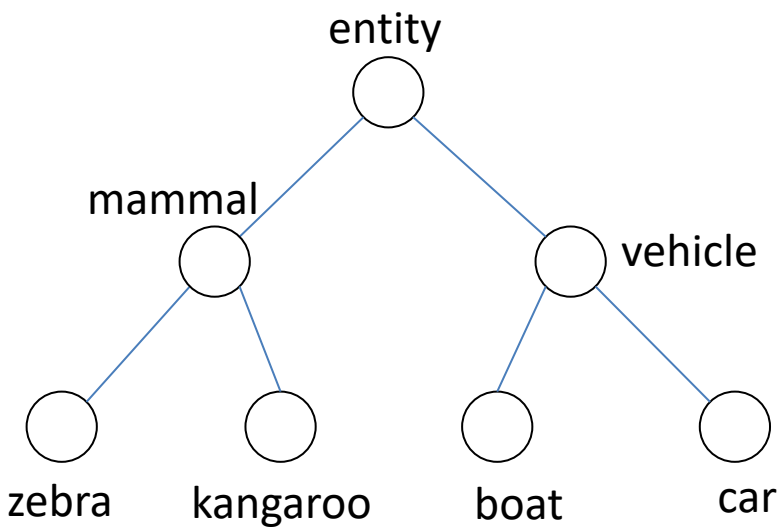
But **it wastes capacity**

Hedging: Be as informative as possible with few mistakes



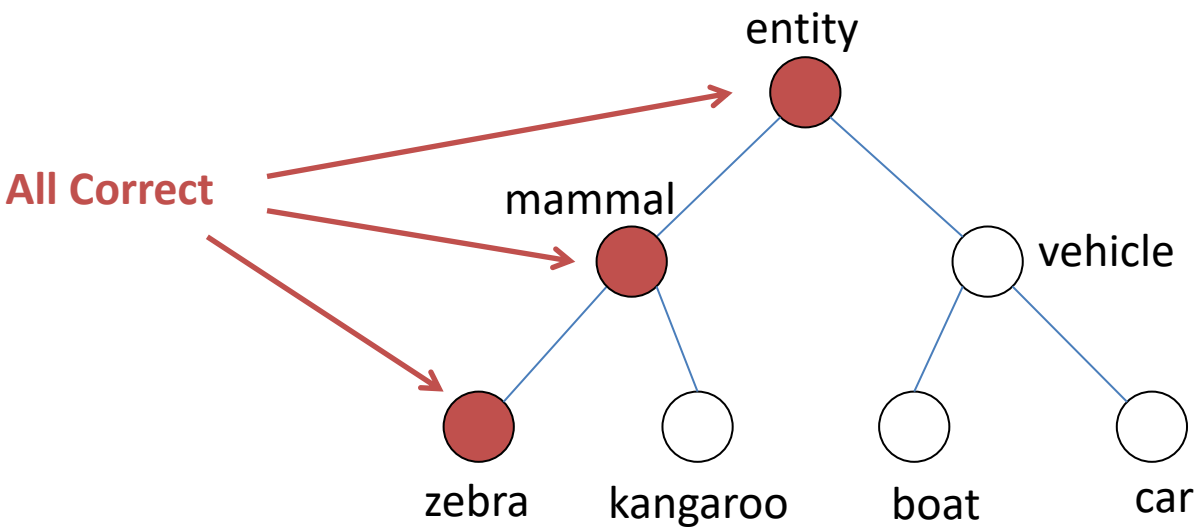
Deng, Krause, Berg, Fei-Fei, CVPR2012

Formal Problem Statement



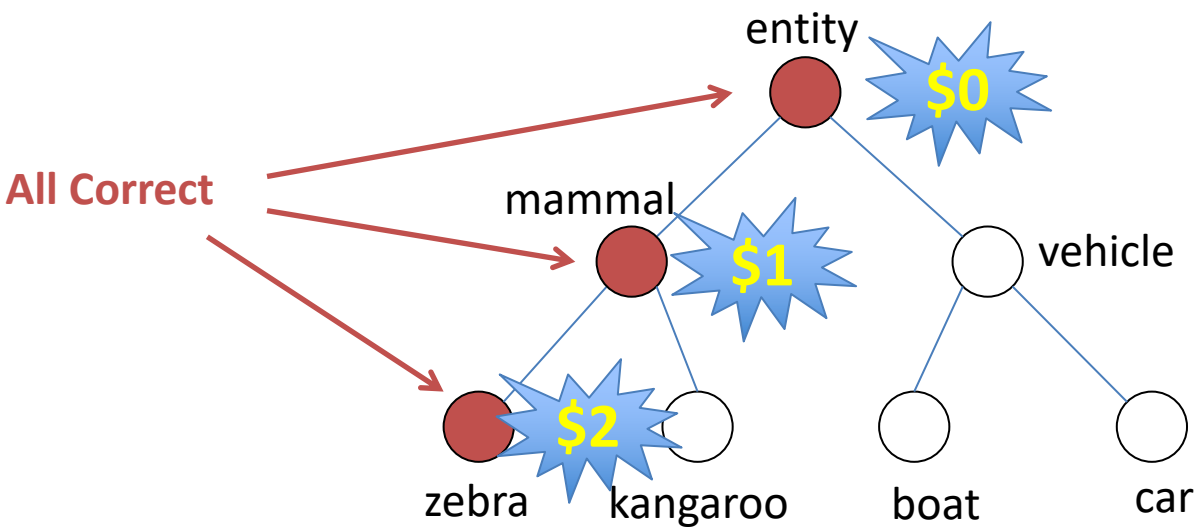
Deng, Krause, Berg, Fei-Fei, CVPR2012

Formal Problem Statement



Deng, Krause, Berg, Fei-Fei, CVPR2012

Formal Problem Statement



Deng, Krause, Berg, Fei-Fei, CVPR2012

Formal Problem Statement

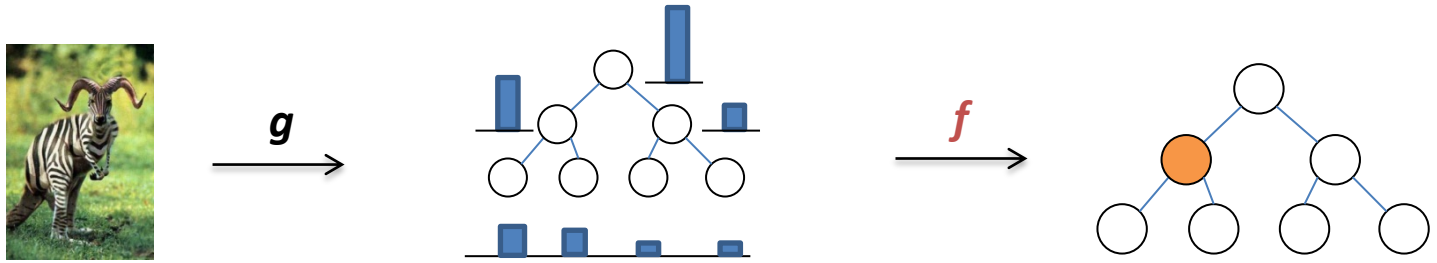
Assumptions

- 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)***

$$\underset{f}{Maximize}\ \ R(f)$$
$$Subject\ to\ \ A(f) \geq 1-\varepsilon$$

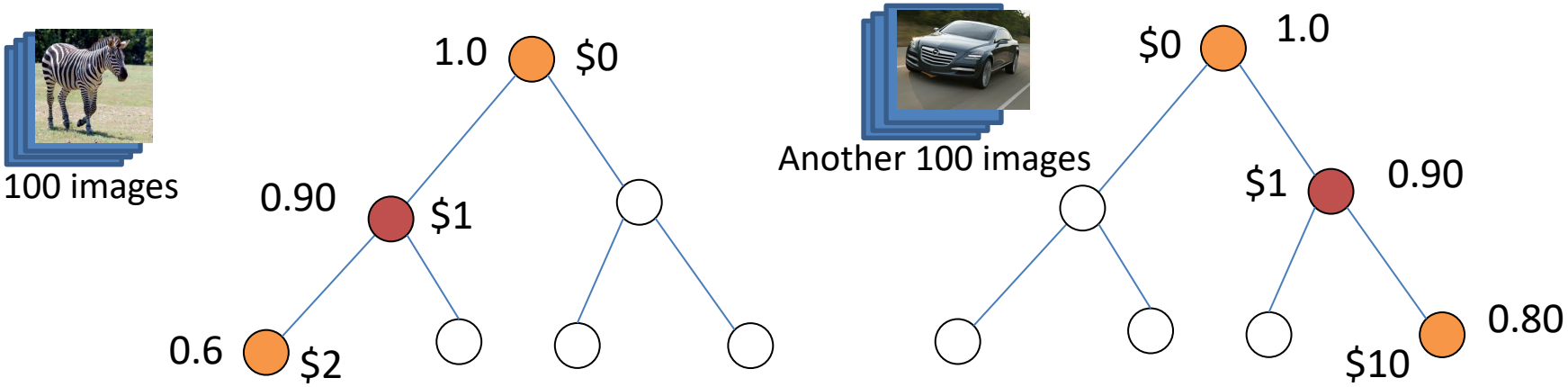


Test image

posterior for all nodes

Deng, Krause, Berg, Fei-Fei, CVPR2012

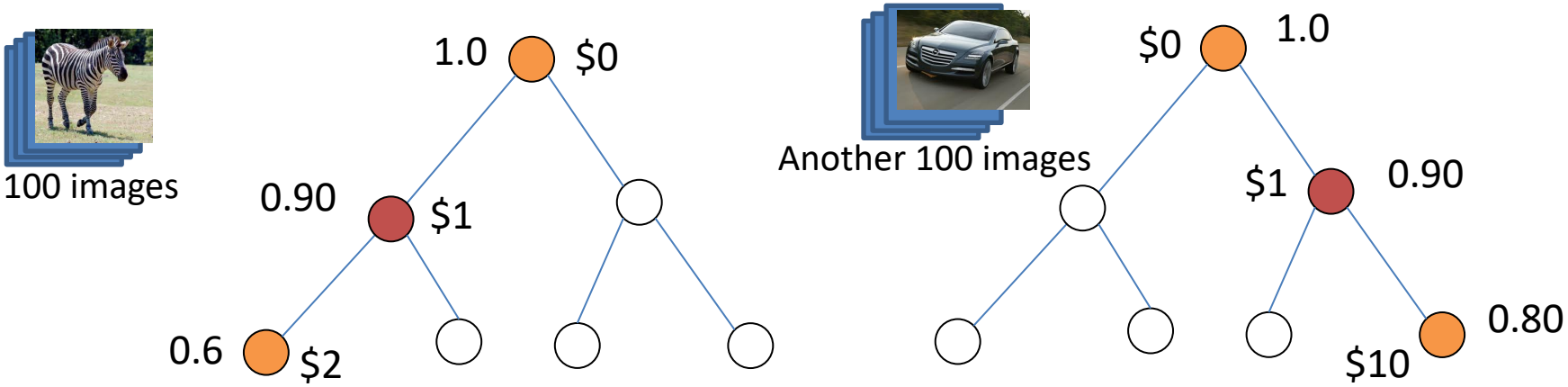
Pick a global confidence threshold T=0.9 [Vailaya et al. '99]



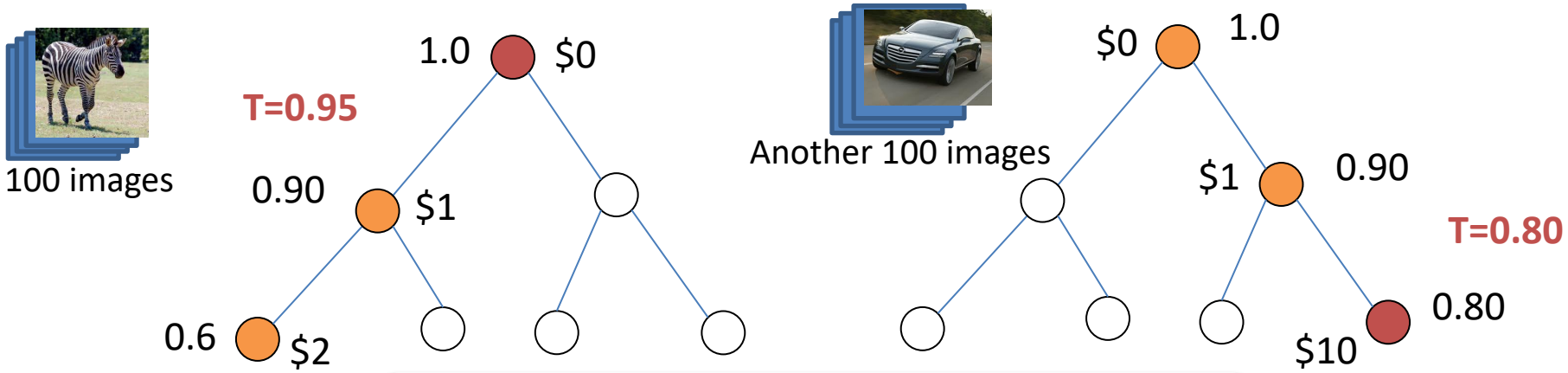
Reward = ($1 * 0.90 + $1 * 0.90) / 2 = $0.90
Accuracy = (0.90 + 0.90) / 2 = 0.90

Deng, Krause, Berg, Fei-Fei, CVPR2012

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

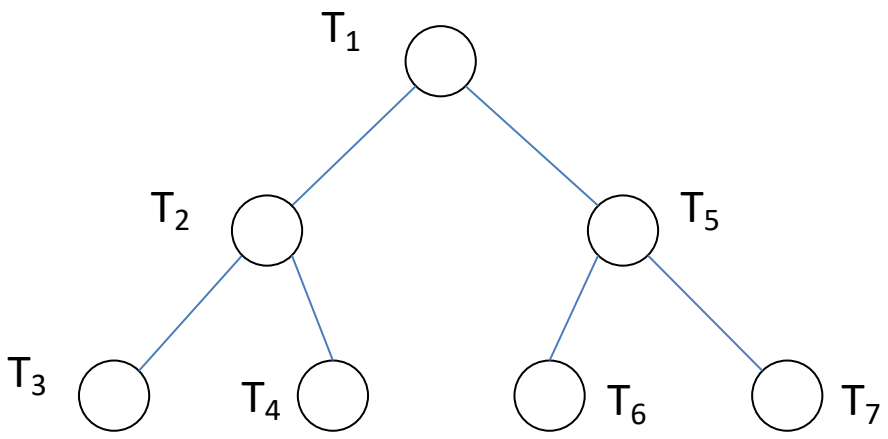


Reward = ($1 * 0.90 + $1 * 0.90) / 2 = $0.90
Accuracy = (0.90 + 0.90) / 2 = 0.90



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

Deng, Krause, Berg, Fei-Fei, CVPR2012

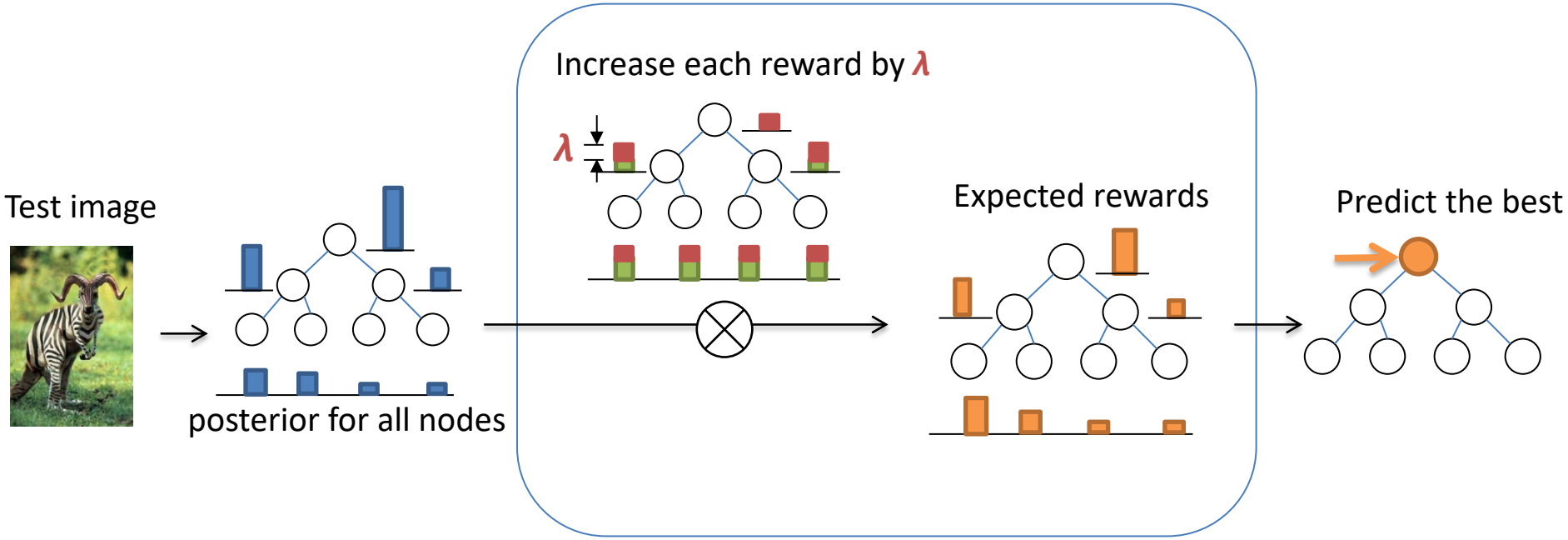


We can optimize individual thresholds…

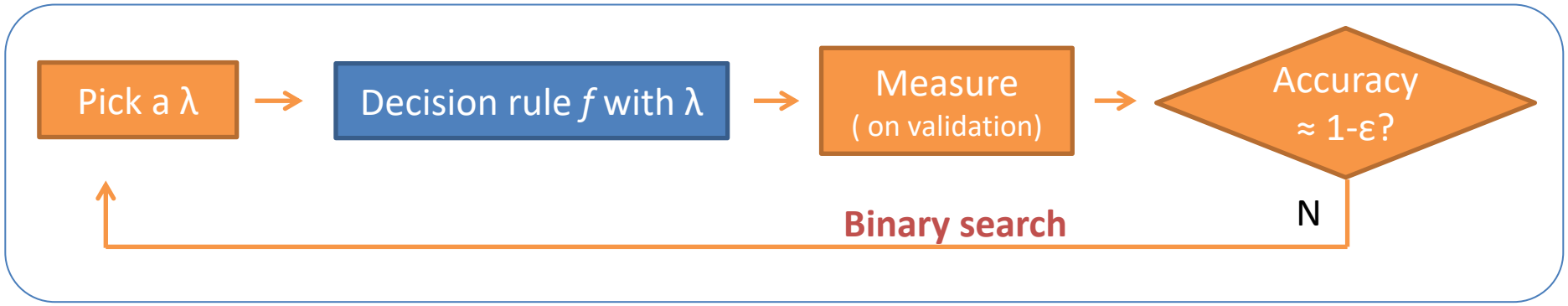
But actually we don't need to.
There is a simpler and provably optimal solution

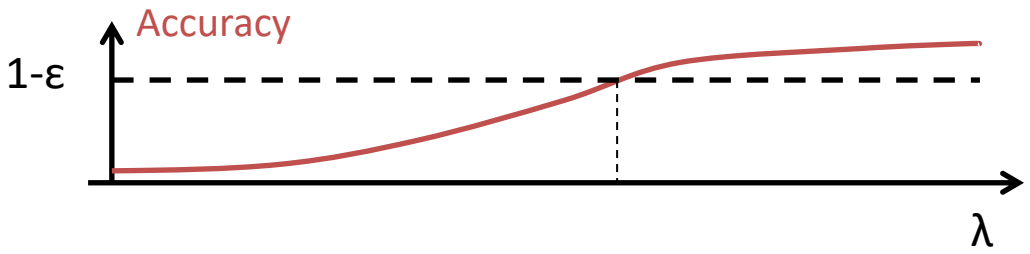
Deng, Krause, Berg, Fei-Fei, CVPR2012

A global, fixed scalar parameter $\lambda \geq 0$



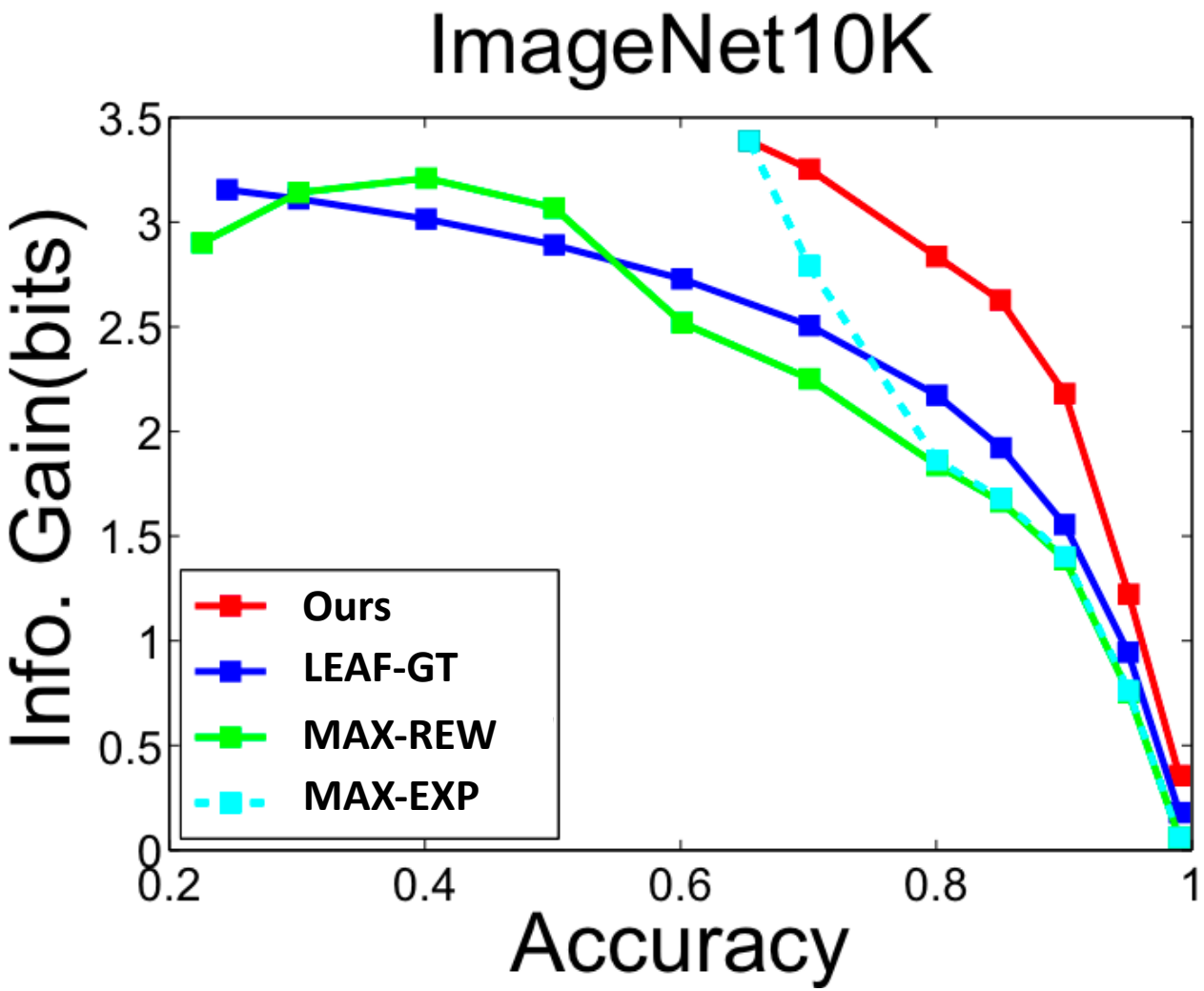
Deng, Krause, Berg, Fei-Fei, CVPR2012





Theorem: Under very mild conditions, this is optimal.

Deng, Krause, Berg, Fei-Fei, CVPR2012



Deng, Krause, Berg, Fei-Fei, CVPR2012

www.image-net.org/eva

EVA

Engine for Visual Annotation

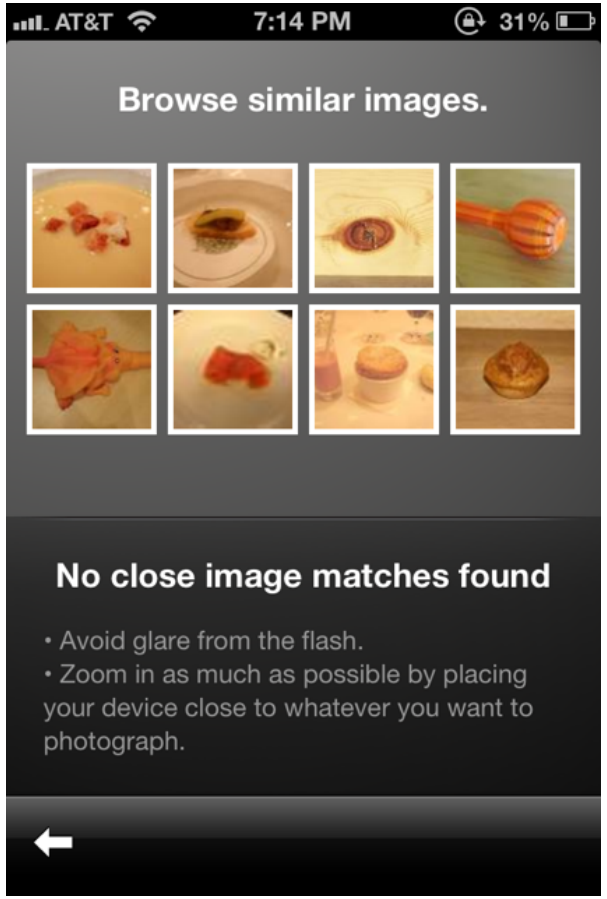
The EVA system, powered by **ImageNet**, can annotate images with guaranteed accuracies. It currently recognizes over **10,000** visual categories. See the project page to find out more.

Paste a URL | Upload an image

ANNOTATE

Google Goggles

Use pictures to search the web.



EVA
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

EVA

Engine for Visual Annotation

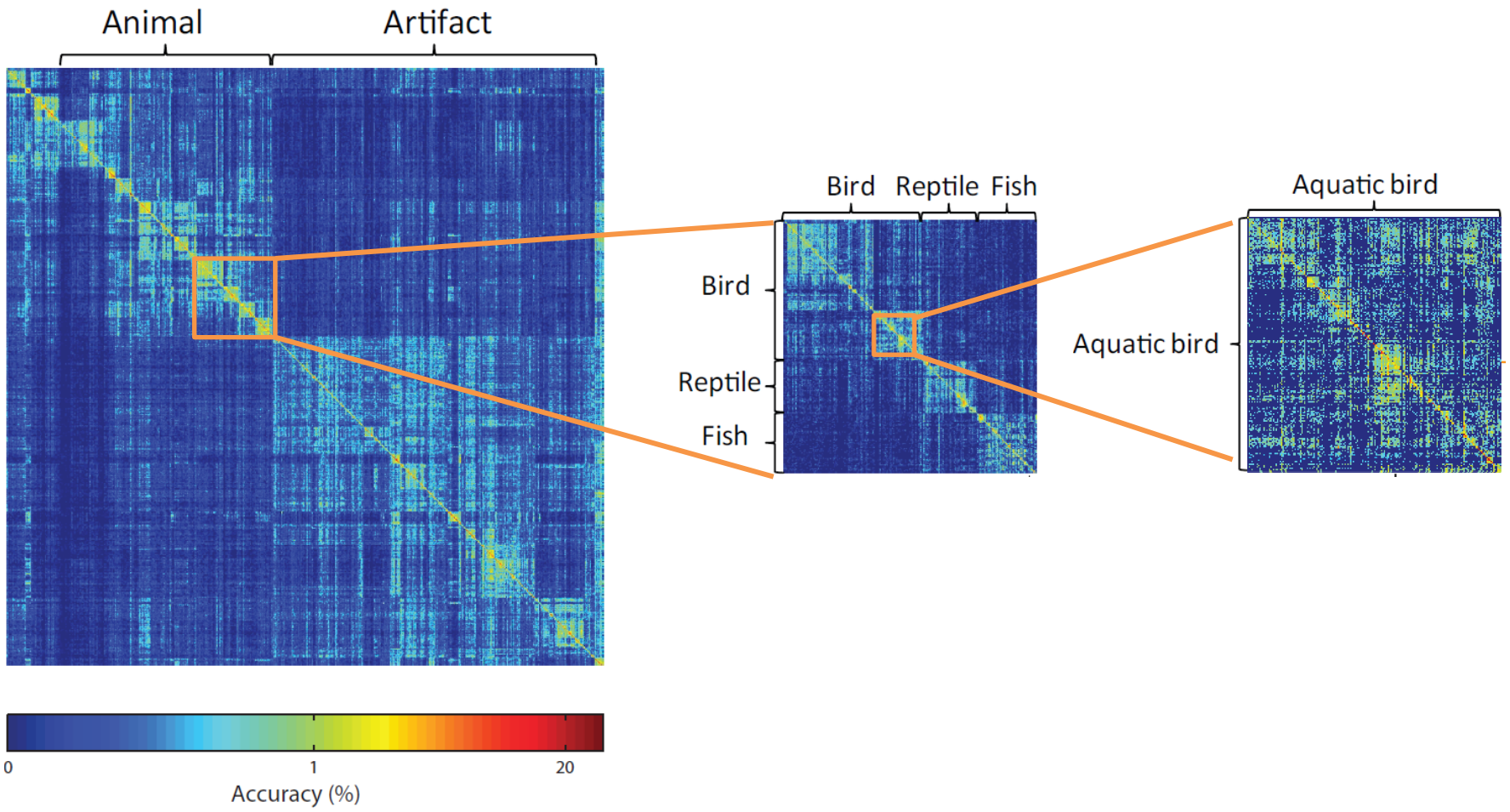
0.78 person, instrument

0.84 person

Challenges

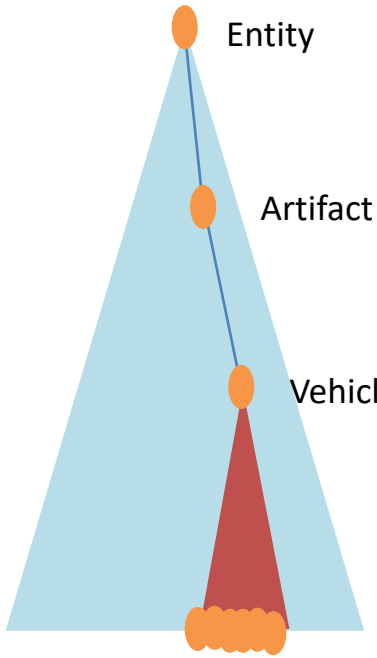
- Semantic hierarchy
- Fine-grained classes
- Large-scale Learning

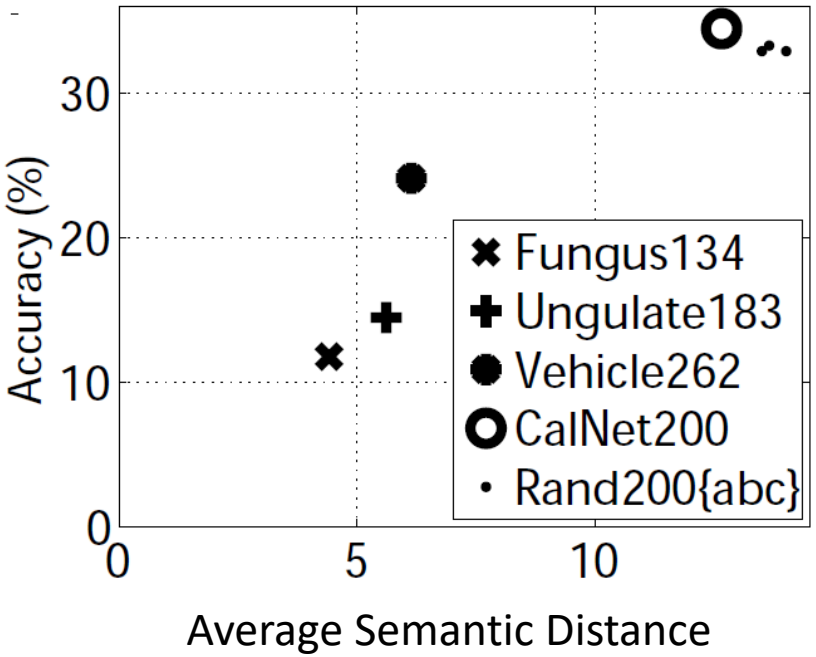
Learn to Classify 10K Classes

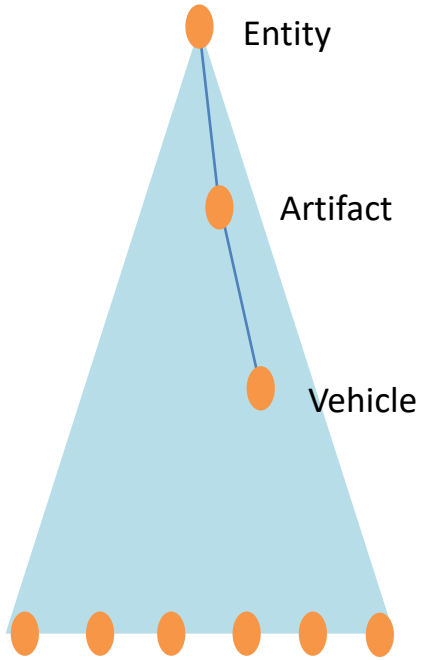


Deng, Berg, Li, & Fei-Fei, *ECCV2010*

Fine-grained categories are a lot harder



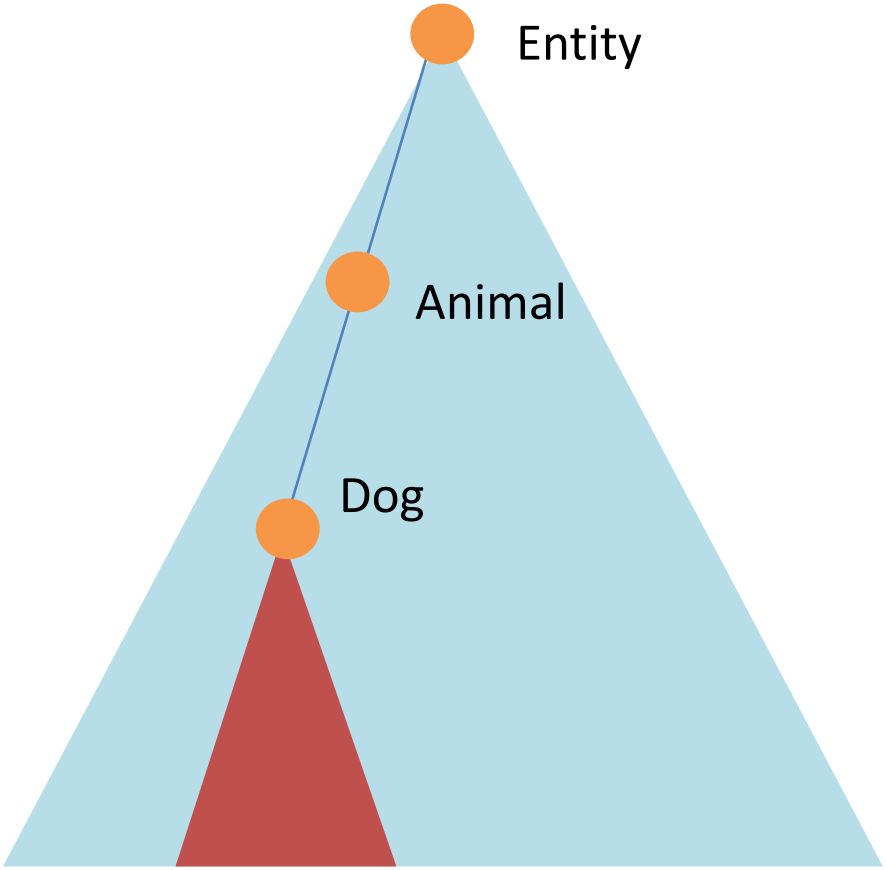




Finer **Coarser**

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

Fine-Grained Recognition



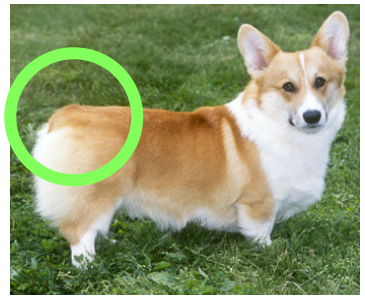
What breed is this dog?

- Basic category known.
- Object already localized.

Fine-Grained Recognition

...

Cardigan Welsh Corgi



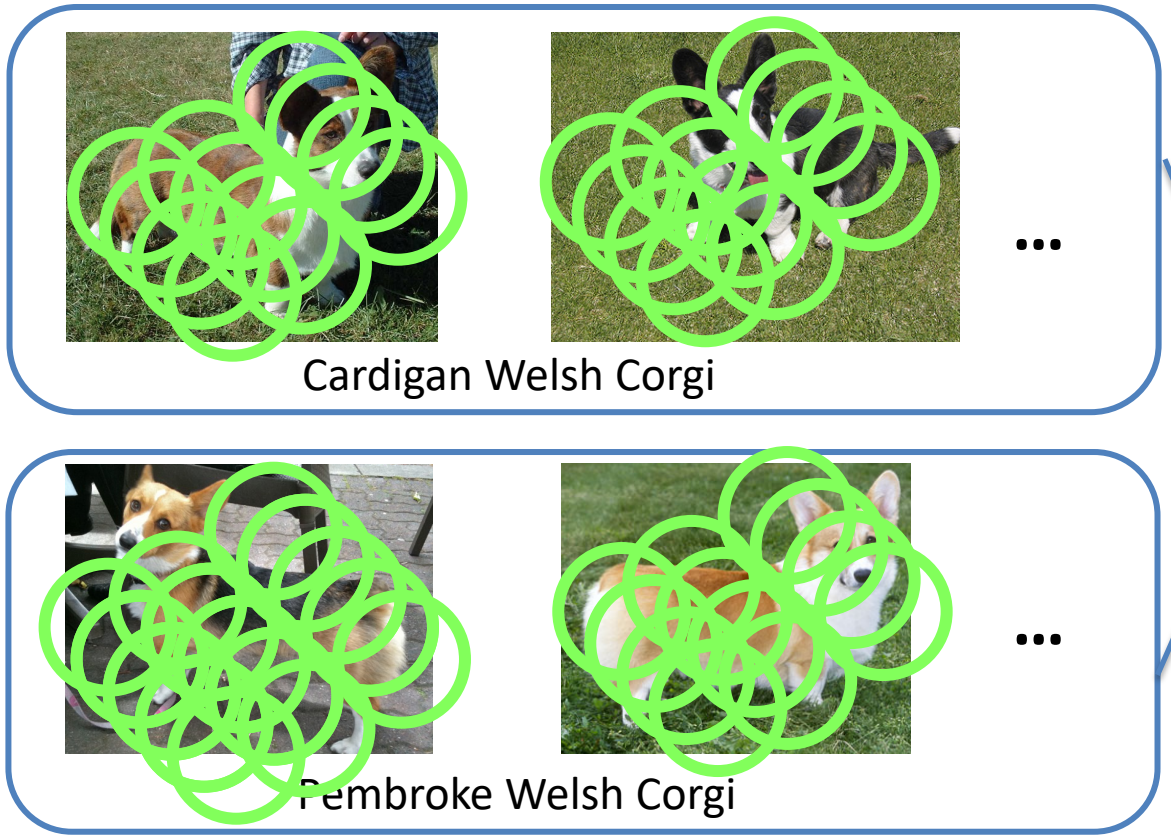
...

Pembroke Welsh Corgi

What breed is this dog?

Key: Find the right features.

Fine-Grained Recognition



Existing Work

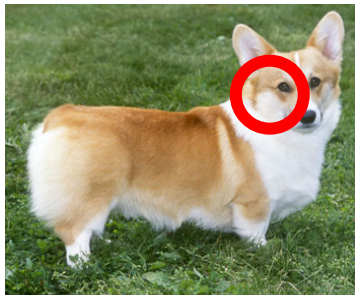
Feature Selection from all possible locations

[Branson et al. '10]
[Bo et al. '10]
[Farrell et al. '11]
[Yao et al. ’11]
[Yao et al. ’12]

Fine-Grained Recognition

...

Cardigan Welsh Corgi



...

Pembroke Welsh Corgi

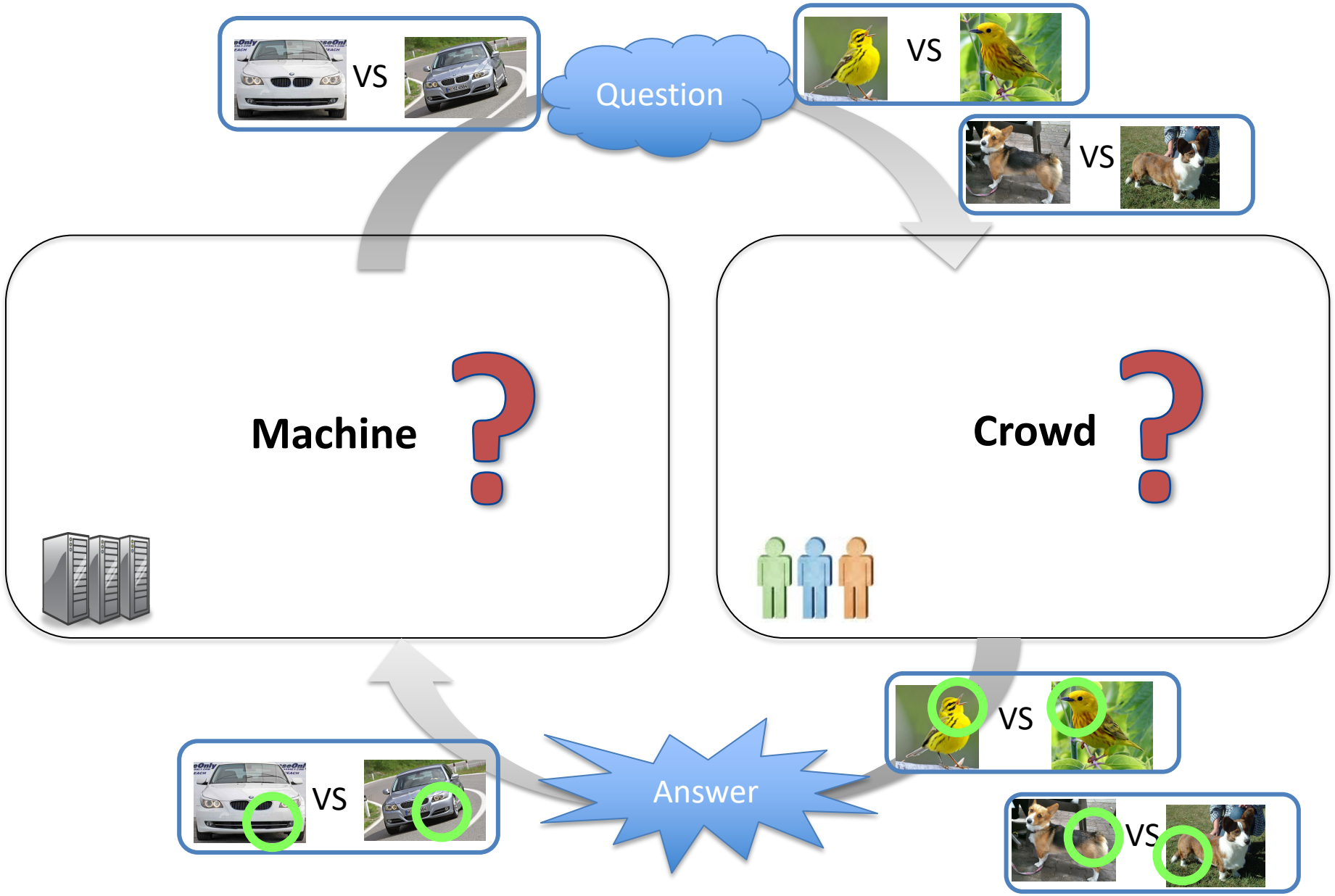
Existing Work

Feature Selection from all possible locations

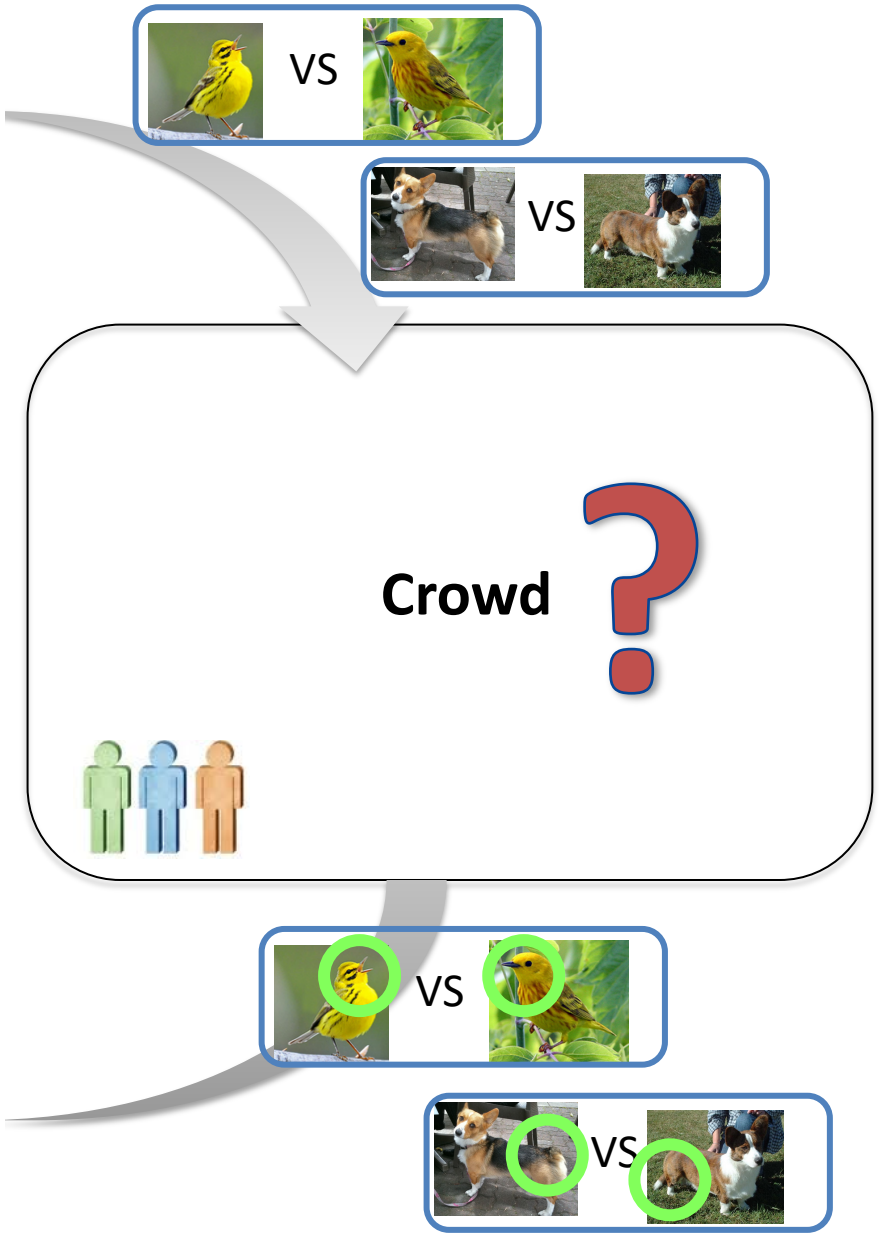
[Branson et al. '10]
[Bo et al. '10]
[Farrell et al. '11]
[Yao et al. ’11]
[Yao et al. ’12]

Can fail to find the right features.

Crowd-Machine Collaboration



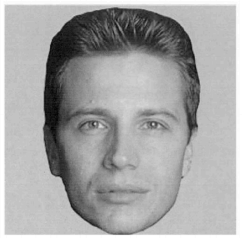
Crowd-Machine Collaboration



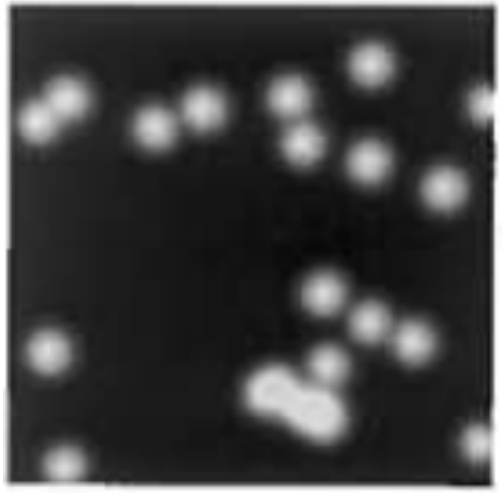
Bubbles [Gosselin & Schyns, ’01]

Smiling

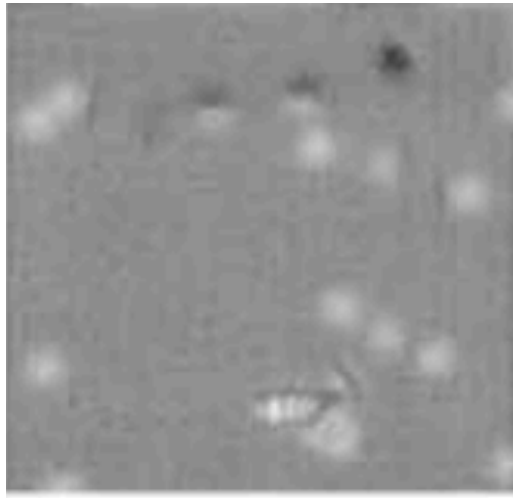
VS



Neutral



Random Bubble Mask

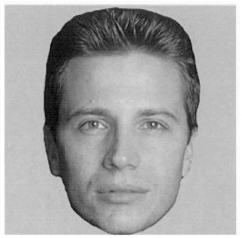


?

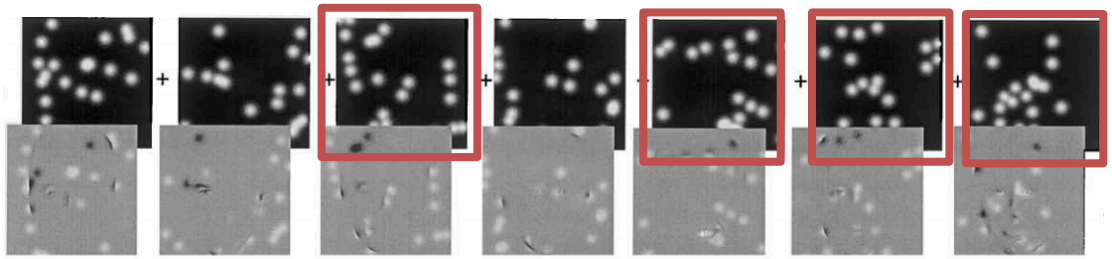
Bubbles [Gosselin & Schyns, ’01]

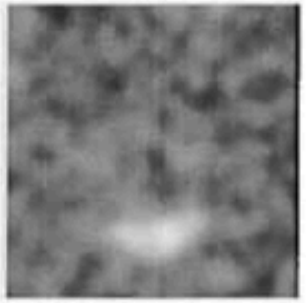
Smiling

VS



Neutral





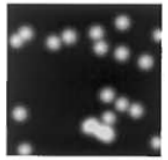
Too costly

Annotation Rationale

[Donahue & Grauman ’11]

No easy quality assurance

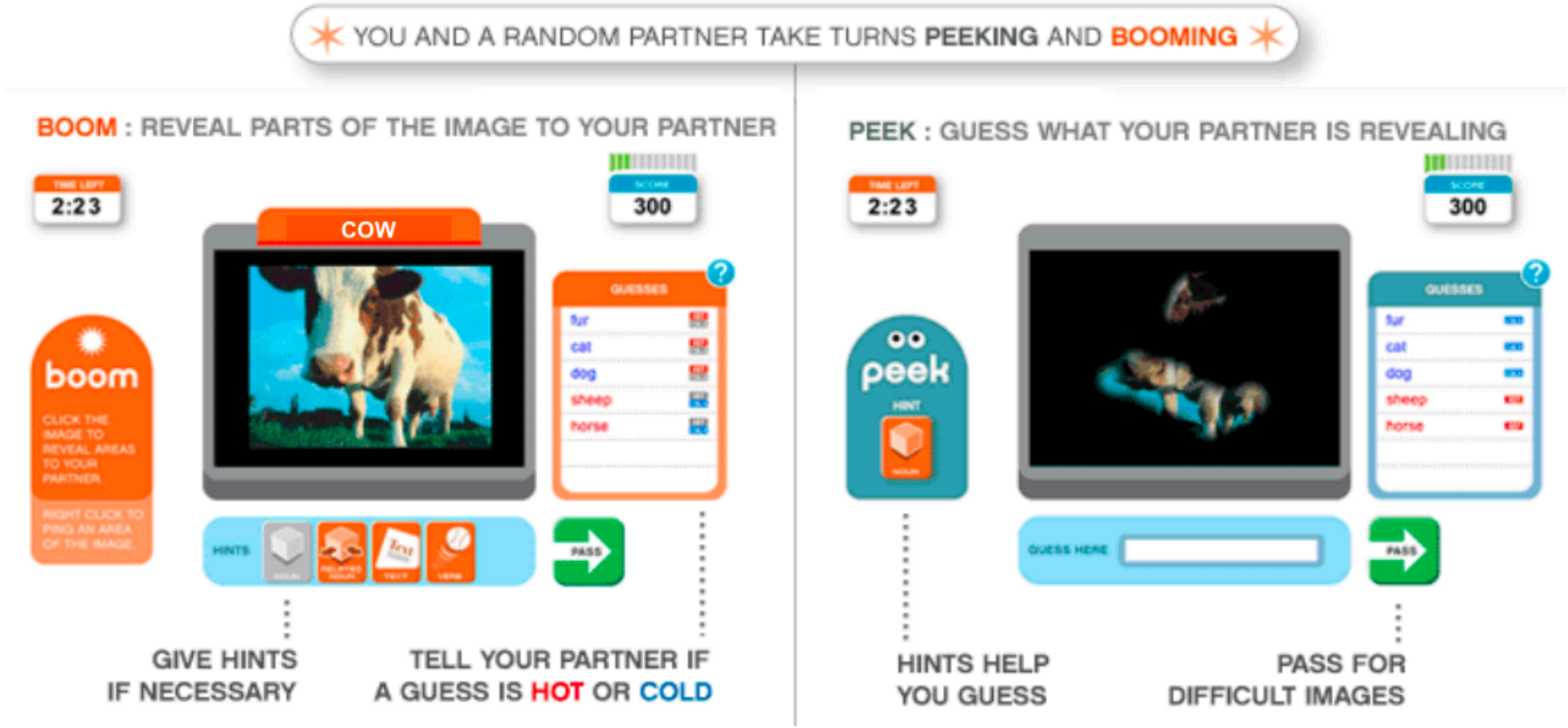
What makes the form of the skater good?



Gosselin & Schyns ’01

Donahue & Grauman ’11

	Gosselin & Schyns ’01	Donahue & Grauman ’11	?
Cost Effective	X		
Quality Assurance		X	

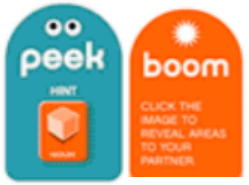


Peekaboom [Ahn, Liu, Blum '06]

Does not work for fine-grained classes

Gosselin & Schyns ’01

Donahue & Grauman ’11



Ahn, Liu, Blum ’06

	Gosselin & Schyns ’01	Donahue & Grauman ’11	Ahn, Liu, Blum ’06
Cost Effective	X		
Quality Assurance		X	
Fine Grained			X

The Bubbles Game

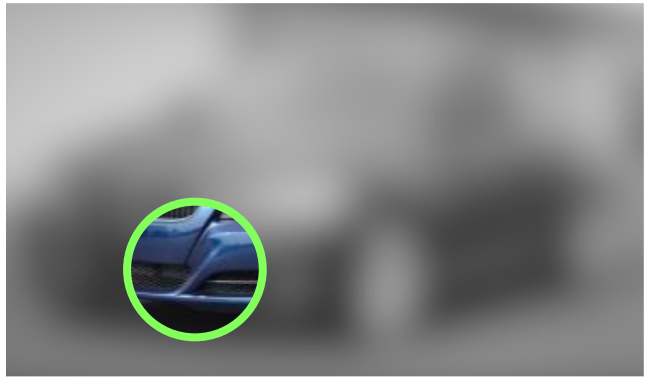
J. Deng, J. Krause, L. Fei-Fei**. Fine-Grained Crowdsourcing for Fine-Grained Recognition.** CVPR 2013**.**

Let’s play

?

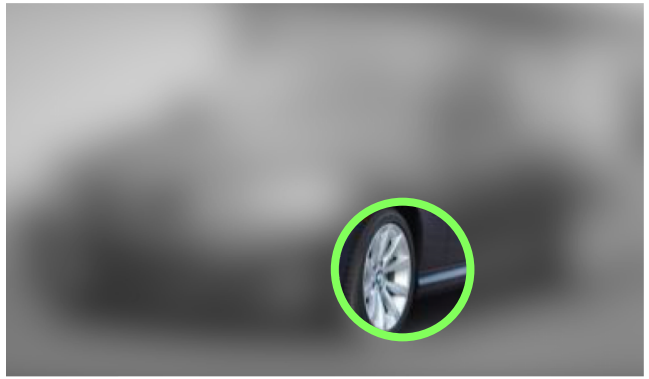
Let’s play

Let’s play

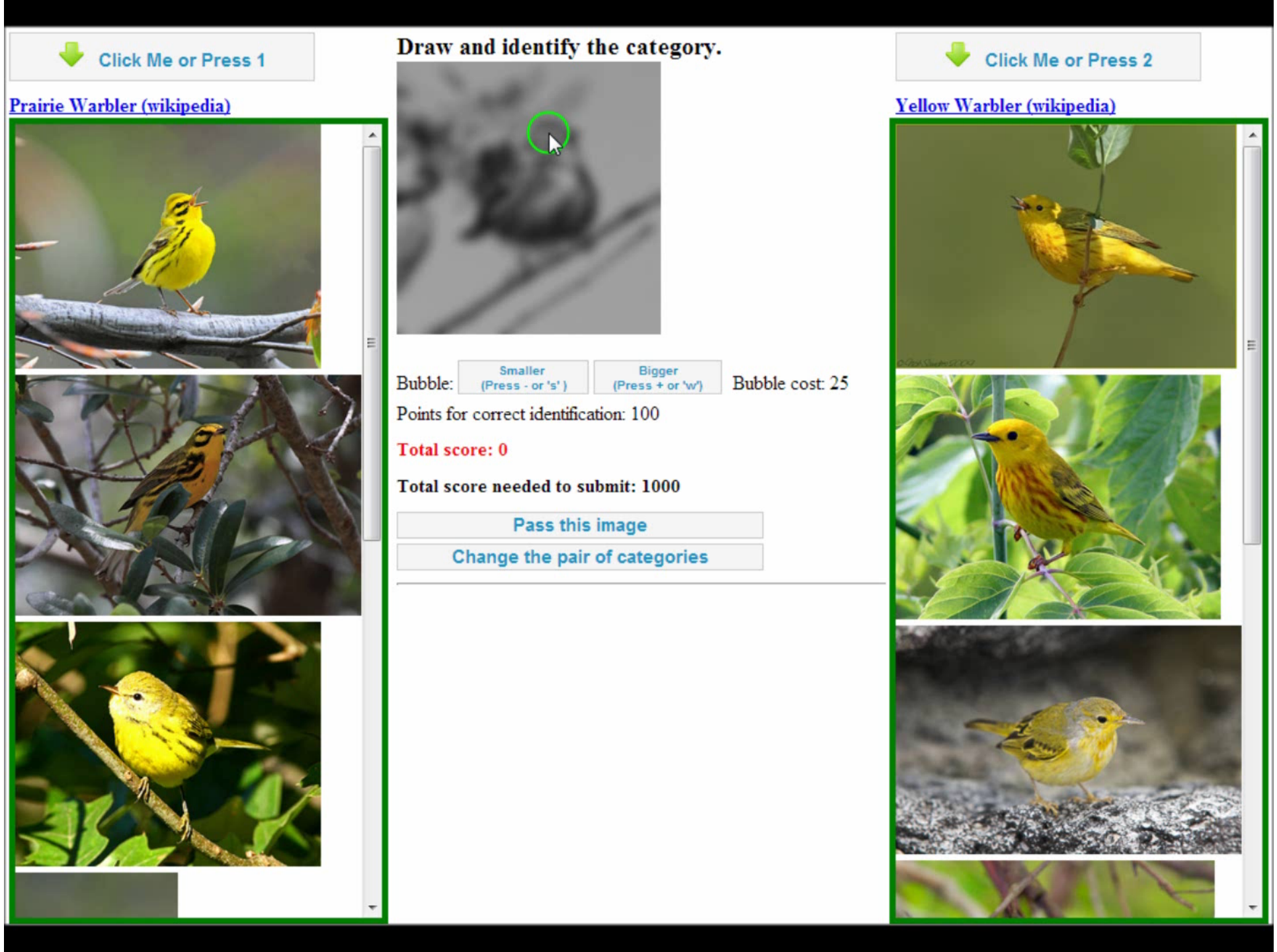


?

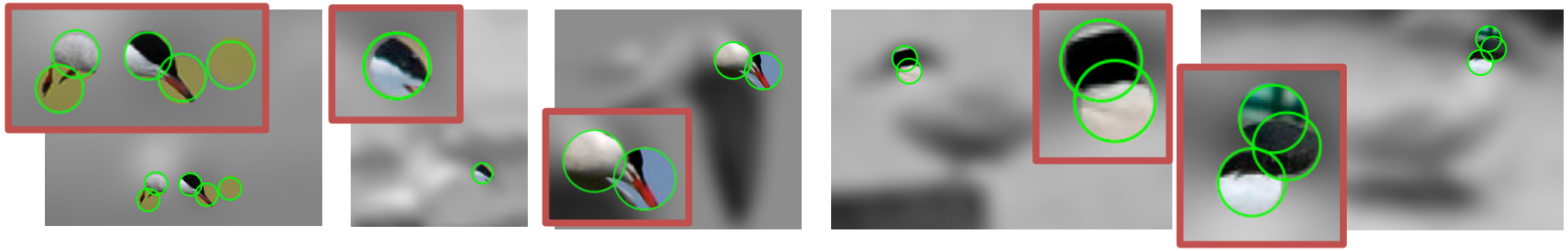
Let’s play

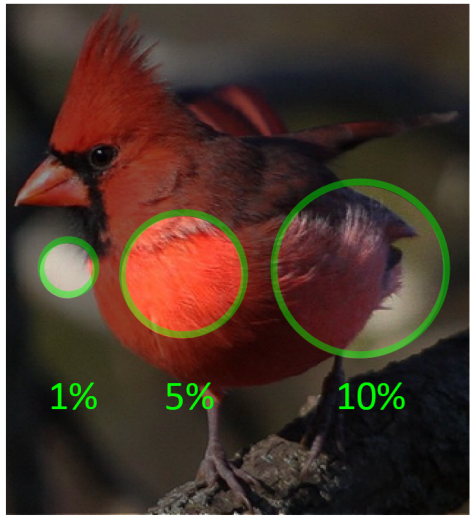


?


Click Me or Press 1
Draw and identify the category.
Click Me or Press 2
Prairie Warbler (wikipedia)
Yellow Warbler (wikipedia)
Bubble:
Smaller (Press - or 's')
Bigger (Press + or 'w')
Bubble cost: 25
Points for correct identification: 100
Total score: 0
Total score needed to submit: 1000
Pass this image
Change the pair of categories

Crowd Picked Bubbles (AMT for now)





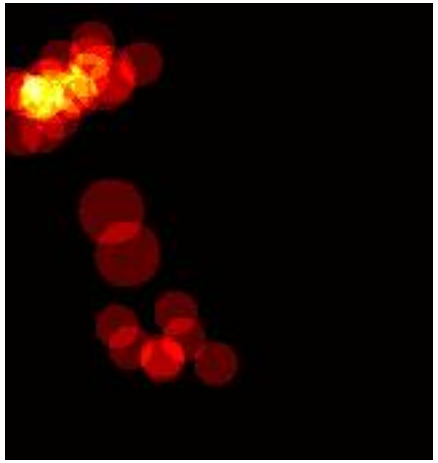
Bubble sizes as proportions of image

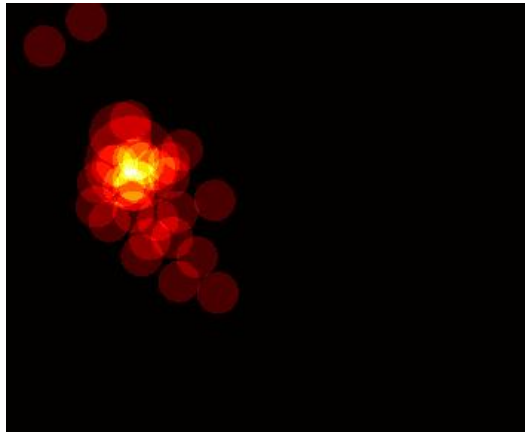
70% of games are successful

>90% of successful games use <10% of area

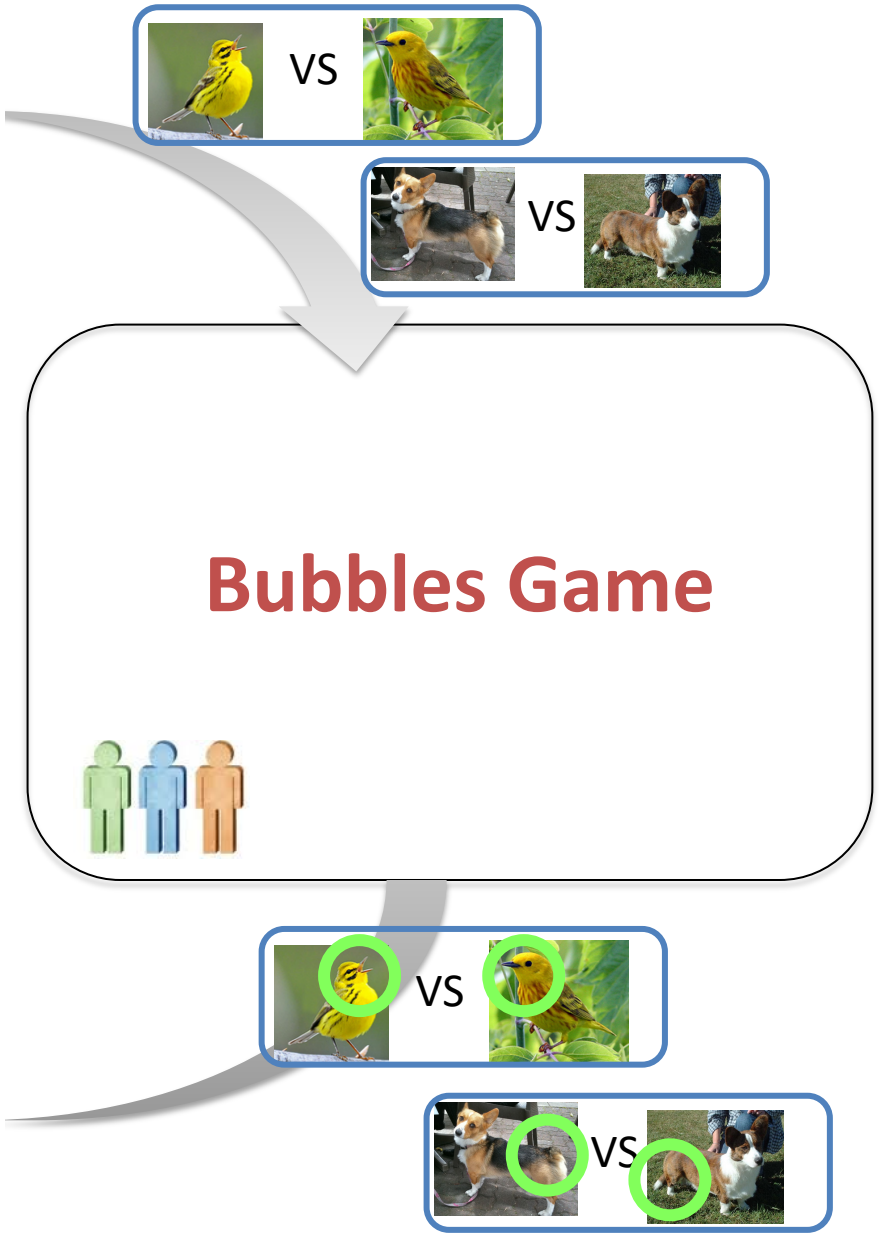
Deng, Krause, & Fei-Fei, *CVPR2013*

Bubble Heatmaps

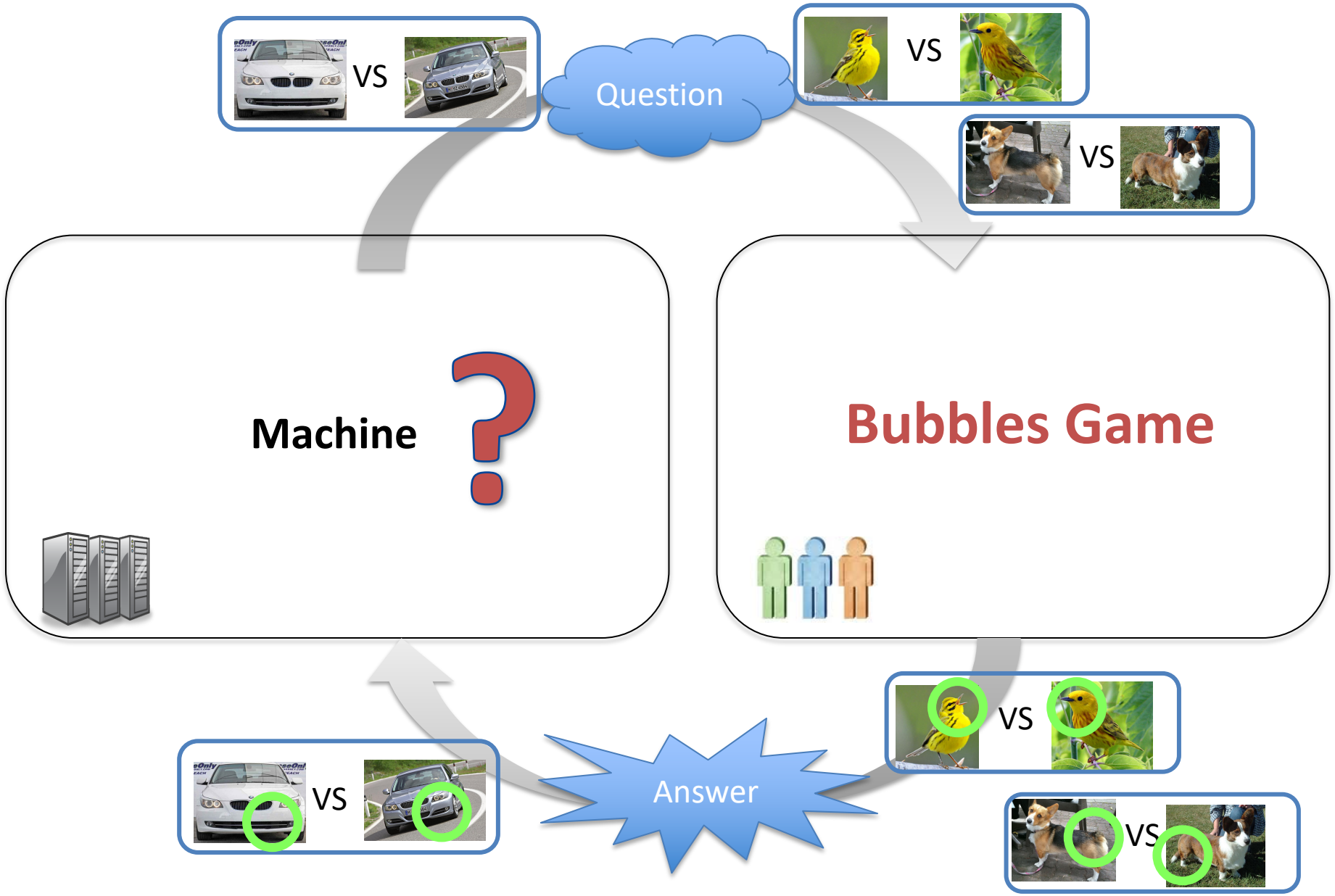




Crowd-Machine Collaboration



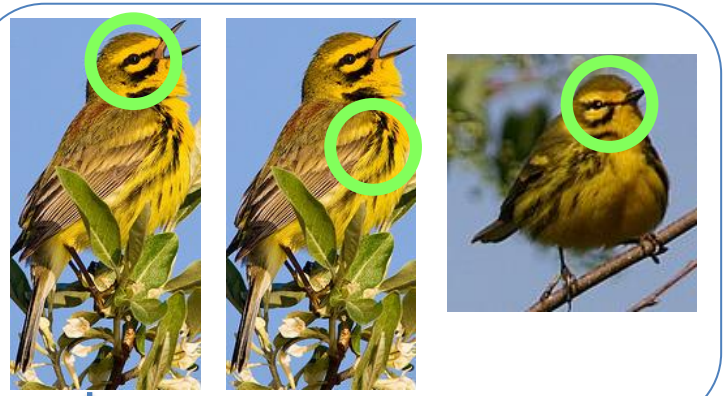
Crowd-Machine Collaboration

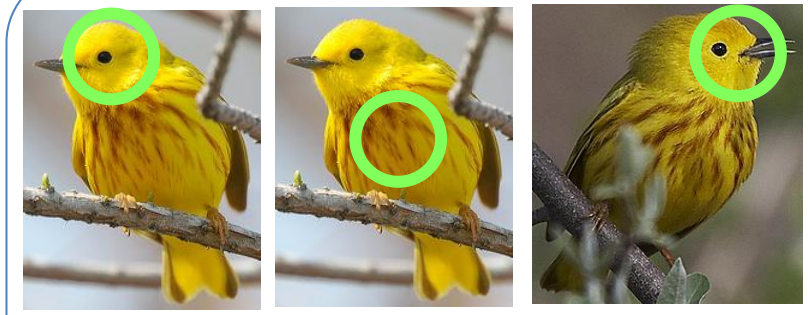


BubbleBank Representation

A test image

Crowd-picked Bubbles (on training)

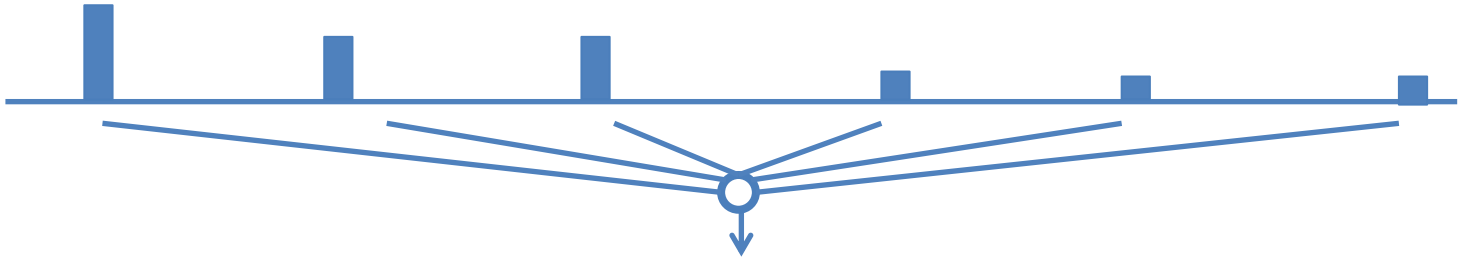


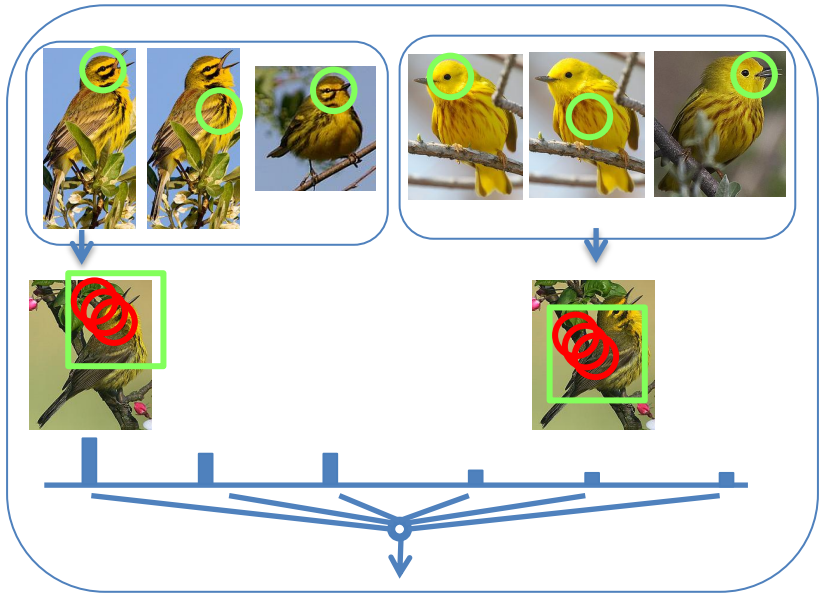


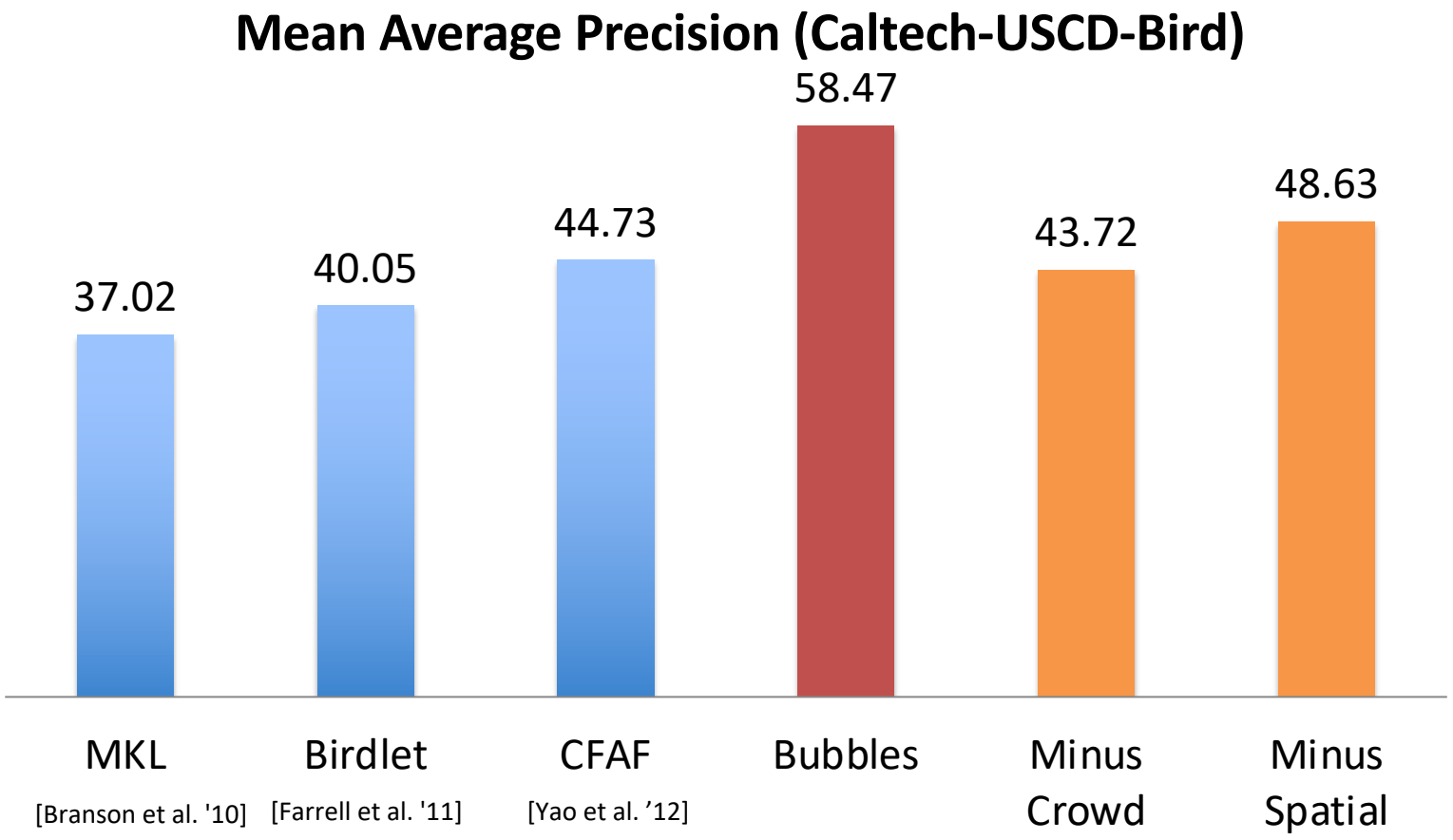
Bubble Matching

Max-pooling

Linear SVM



	V1-like models	BubbleBank	Prior Work
Templates		Crowd picked	Clustering random patches
Pooling		Pool over a single region (spatial prior)	Pool over multiple uniformly sampled regions (e.g. SPM)
		[Deng et al. '13]	[Lecun et al. '98] [Csurka et al. '04] [Lazebnik et al. '06] [Wang et al. '09] [Lee et al. '09] [Pinto et al. '09] [Perronnin et al. '10] [Li et al. '10] [Coates & Ng '11]. [Yao et al. '12]

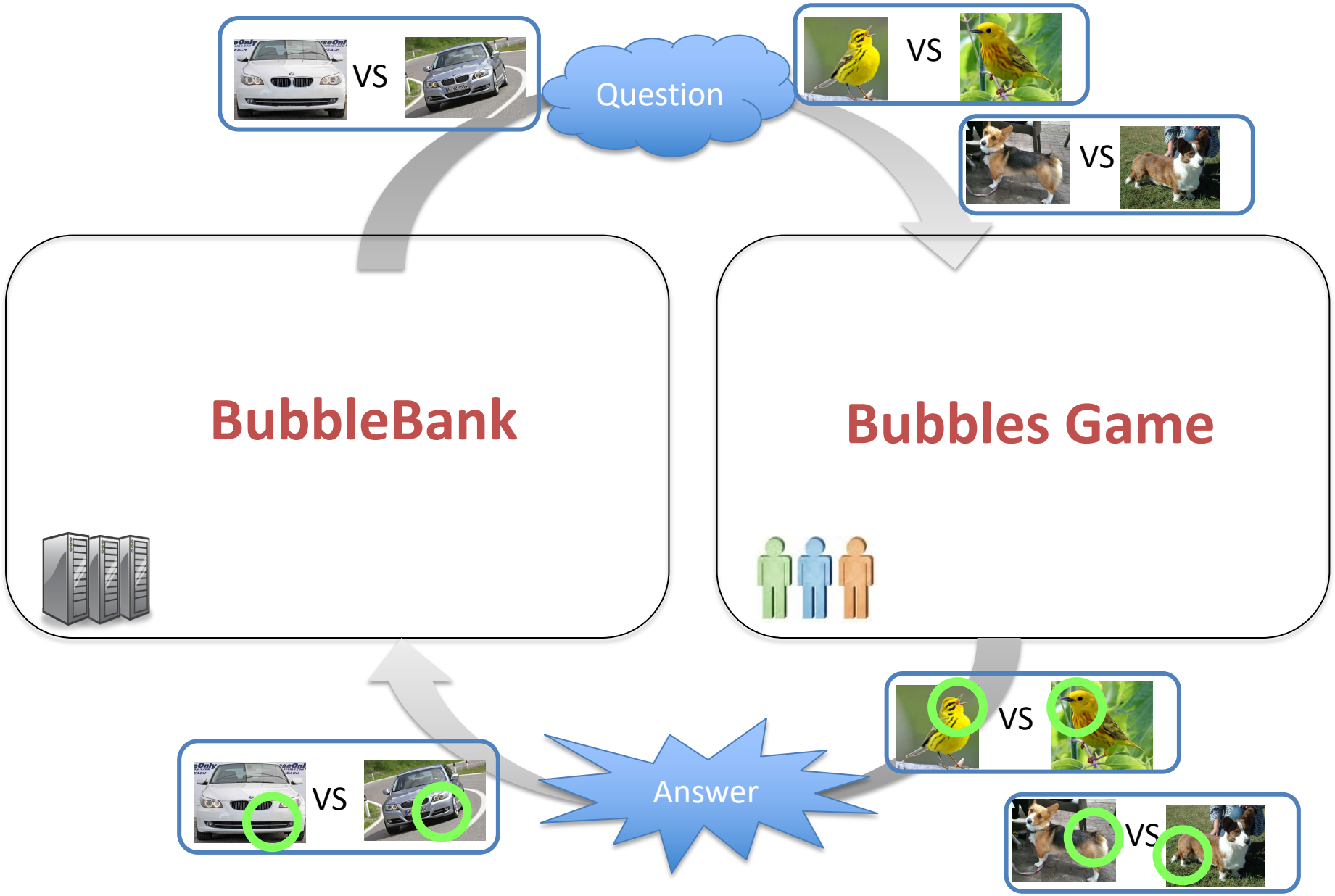


Deng, Krause, & Fei-Fei, *CVPR2013*

Top Activated Bubbles

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

Crowd-Machine Collaboration



Challenges

- Semantic hierarchy
- Fine-grained classes
- Large-scale Learning