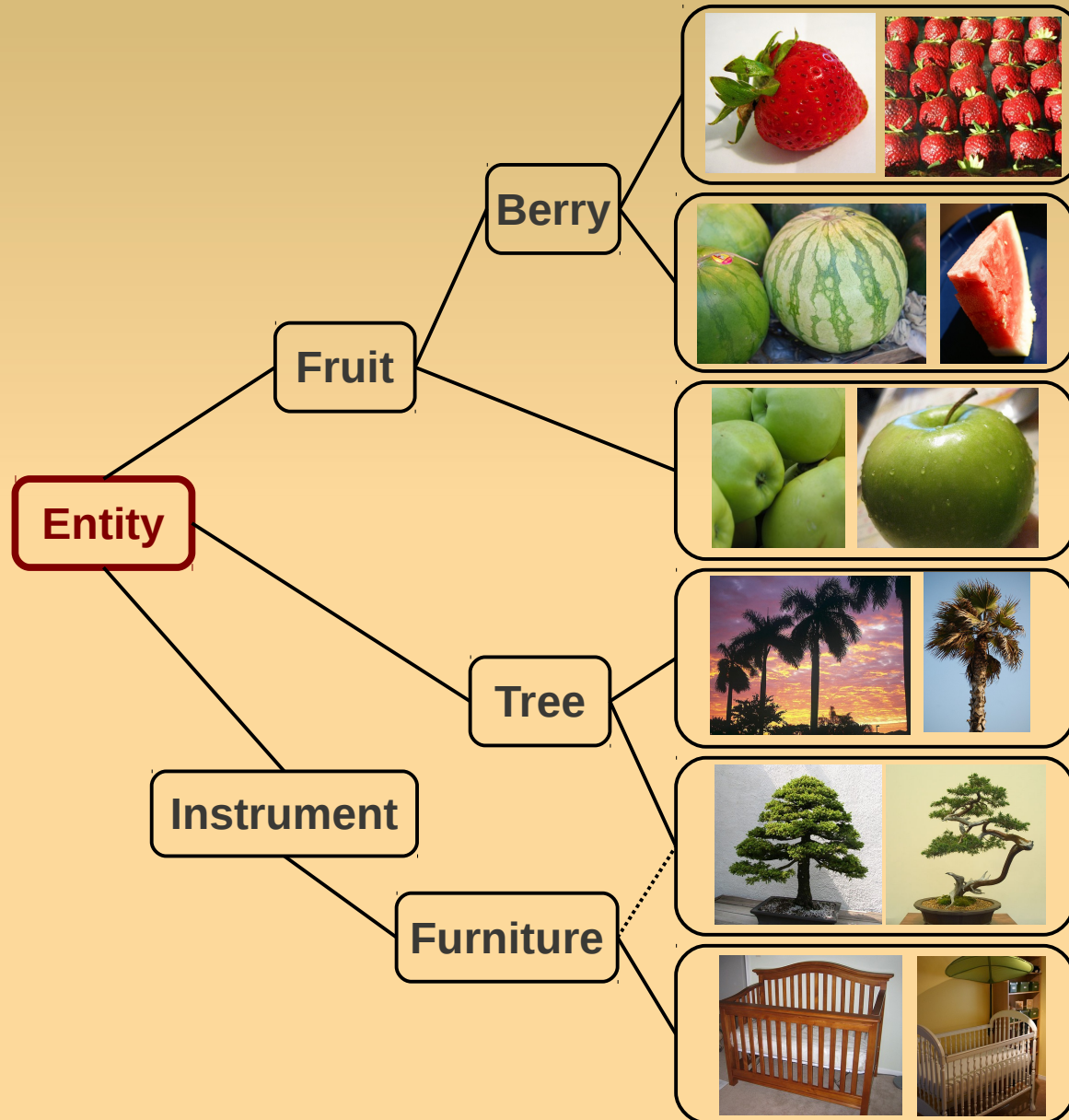


Attribute learning in large-scale datasets

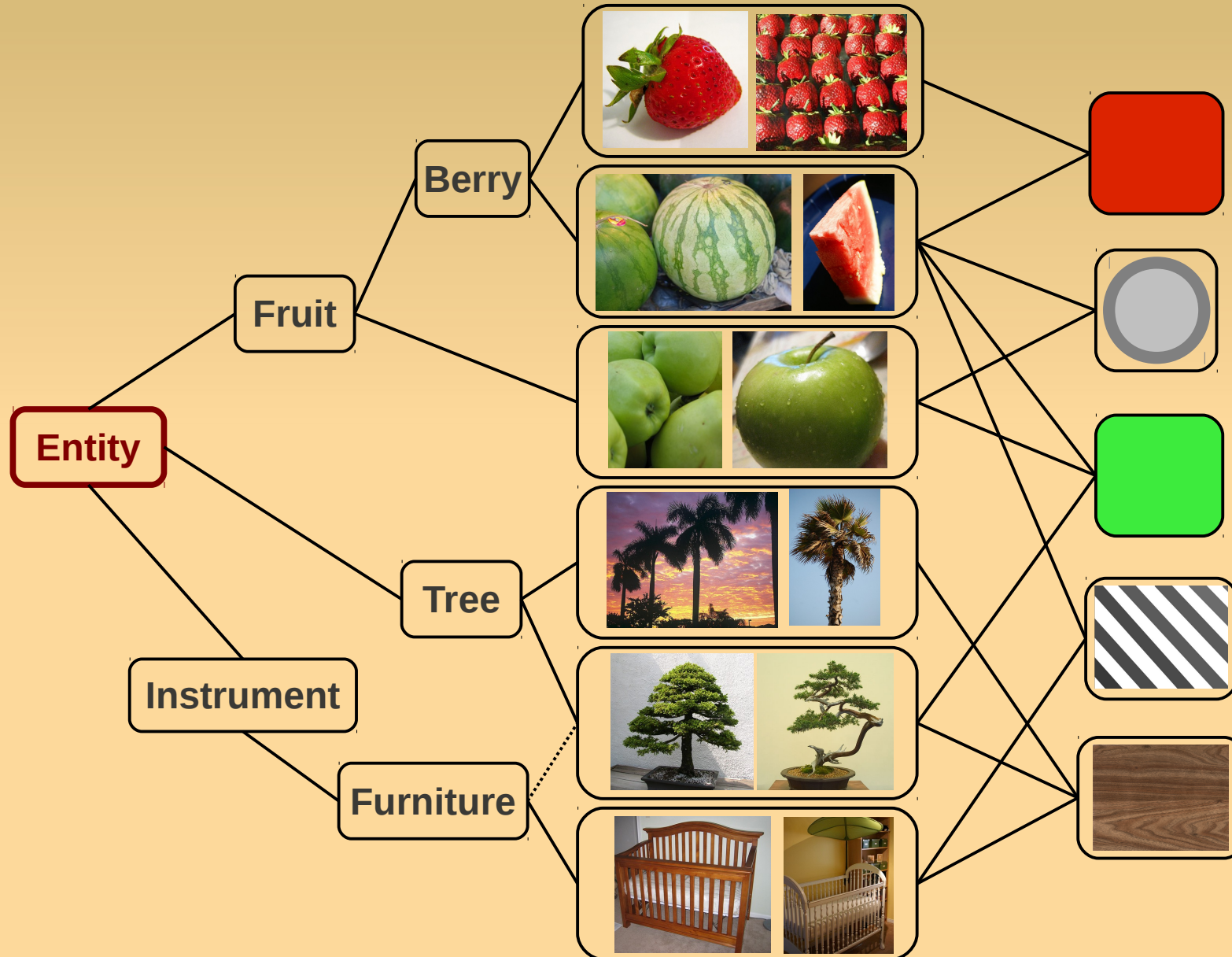
Olga Russakovsky and Li Fei-Fei



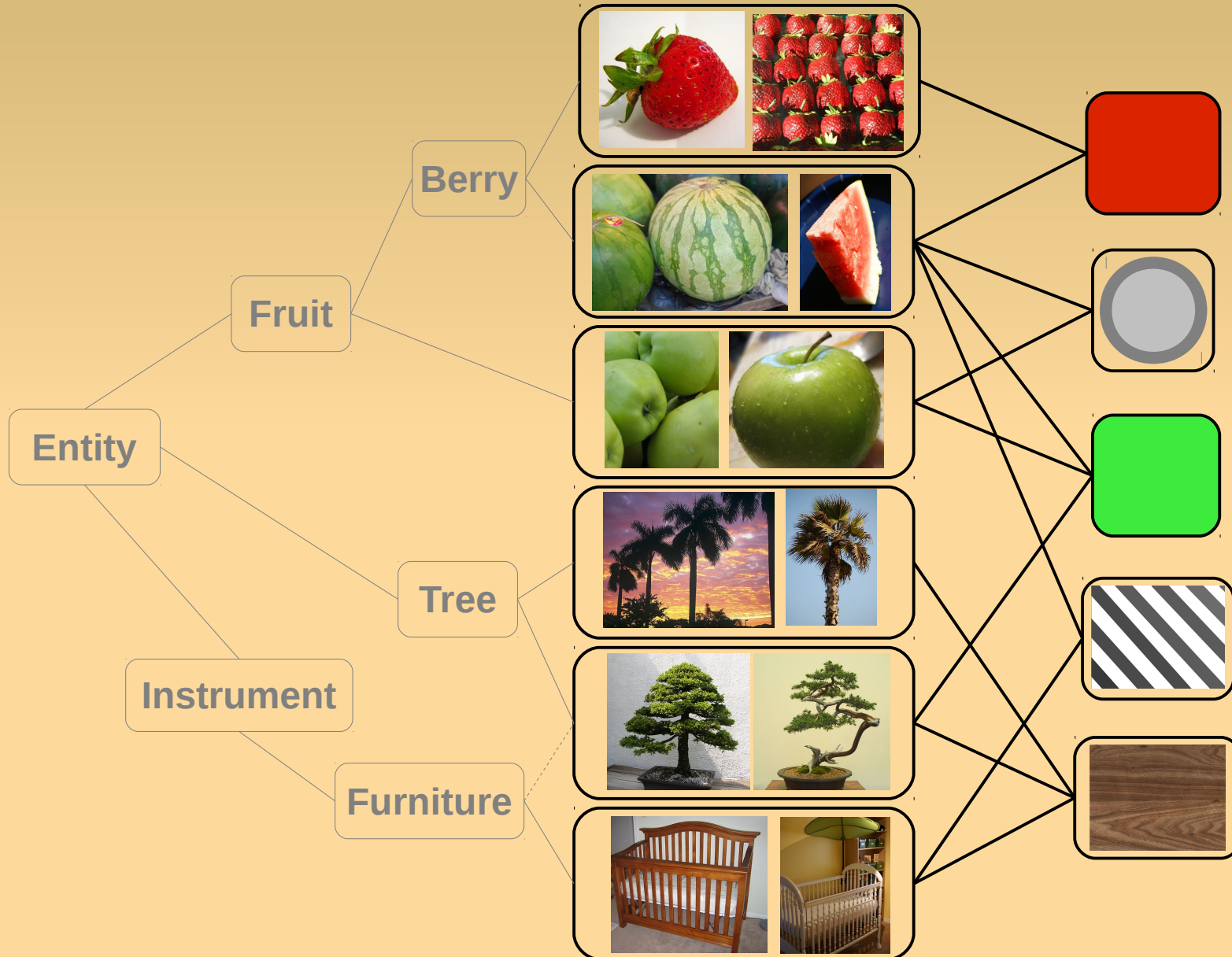
Categorization of the visual world



Categorization of the visual world



Categorization of the visual world



Scale of prior work on attributes

Berg '10



4

10

Lampert '09



50

100

Current work



384

1000

1

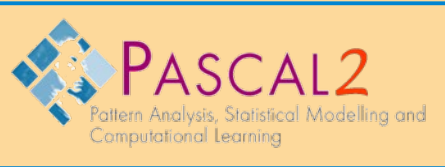
32

Number of object classes

10,000s



Kumar '09

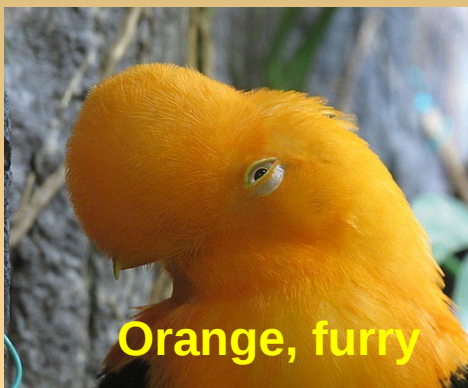


Farhadi '09, '10, Wang '10

Future goal

Why use attributes?

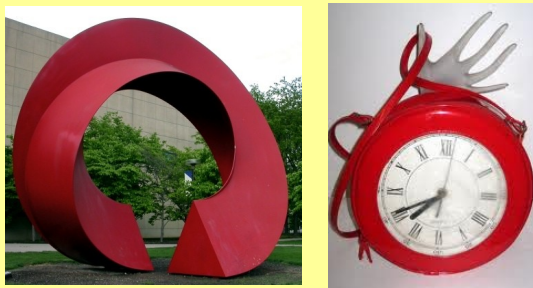
Object description



Farhadi et al. '09

Targetted retrieval

Find red round objects.



Zero-shot learning

Frogs are green, have heads and legs. What is this?



Lampert et al. '09

Object classification



Always
yellow
Never
blue



Better model

Wang and Mori '10

Outlier discovery



Farhadi et al. '09

Visual categorizaion

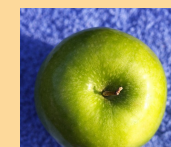


Sharp



Green

Metallic



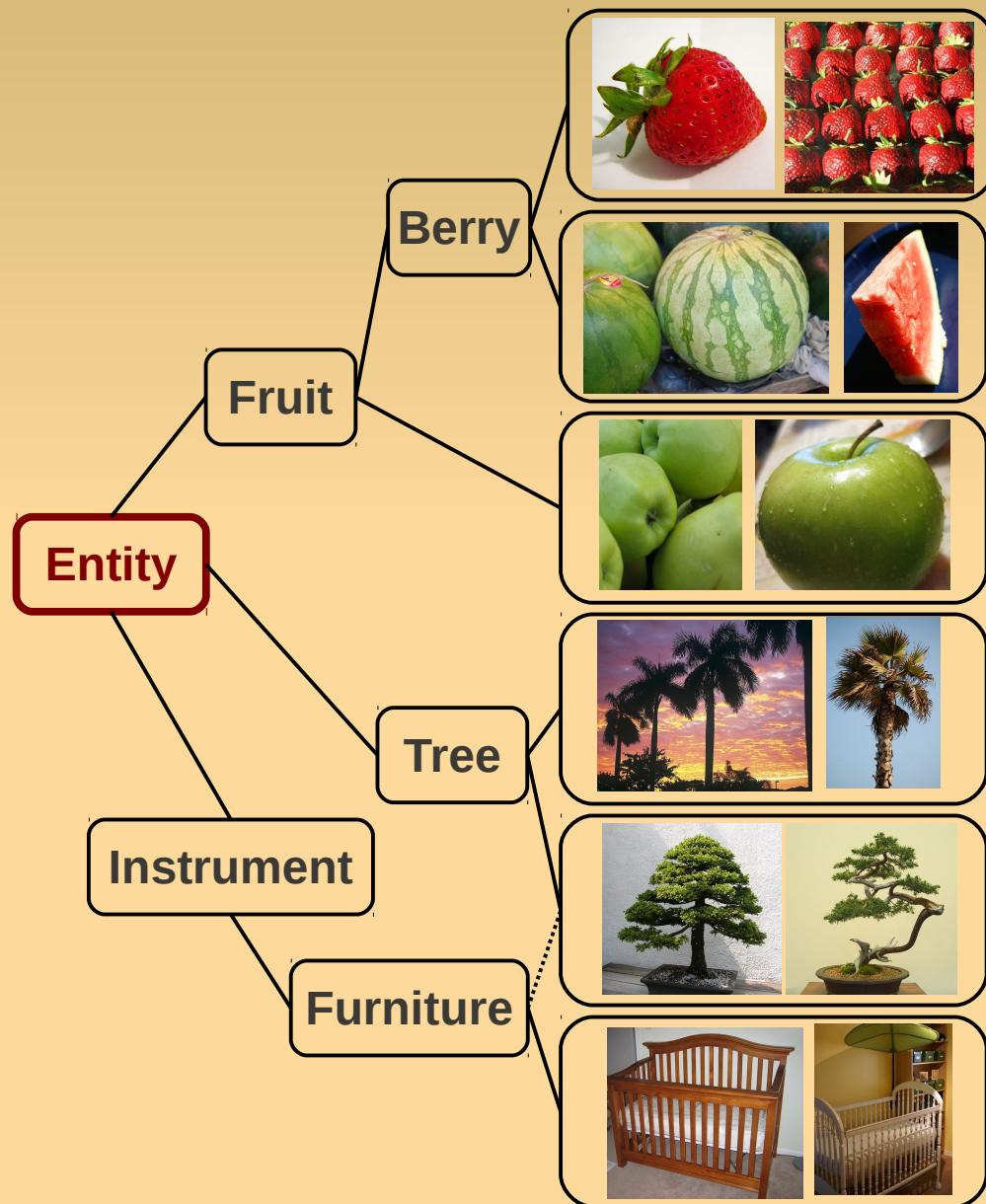
Round



Overview

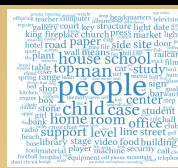
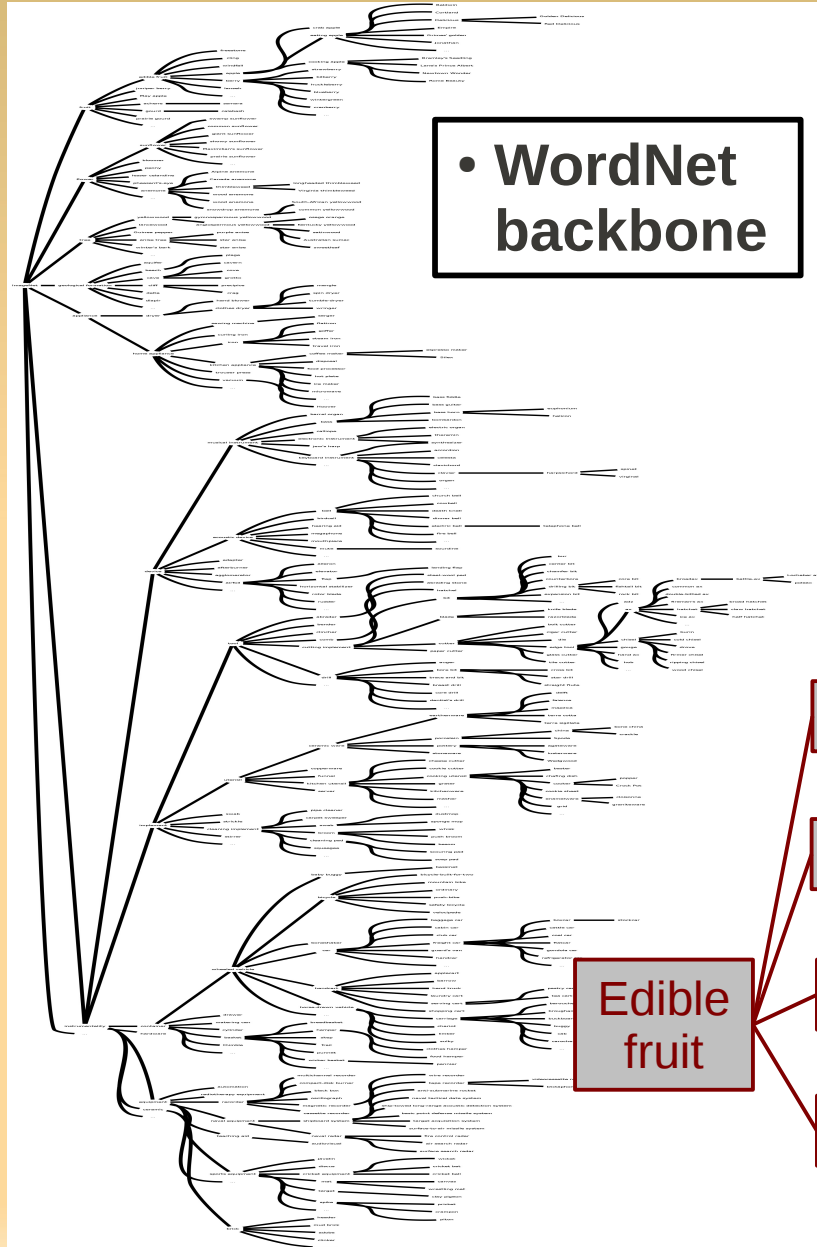
- **Obtaining large-scale training data**
 - **Images**
 - **Attribute labels**
- Training and evaluating attribute classifiers
- Performing higher-level tasks:
 - Targetted retrieval
 - Zero-shot learning
- Future directions

Obtaining images:

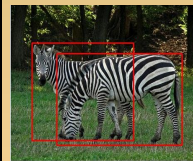


J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. In CVPR 2009.

Obtaining images:



- **Lots of data**
> 15,500 synsets, > 11M images



- **Bounding box annotations**
1,000 synsets

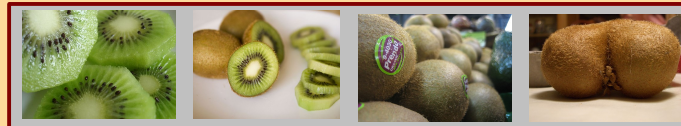


- **Definitions**
“Sweet fleshy red fruit”

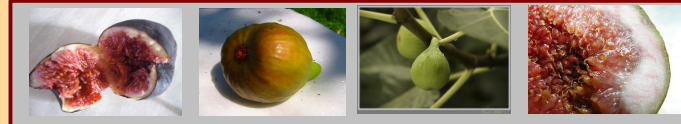
Mango



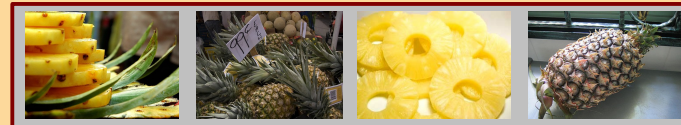
Kiwi



Fig



Pineapple



Edible fruit

Obtaining **semantic attribute** labels

A semantic attribute is an attribute which can be described using language.

	Class-level: all frogs are green	Image-level: this particular frog is green
Text mining	Rohrbach '10	Ferrari '07, Berg '10
Hand labeling	Lampert '10	Kumar '09, Wang '09, Farhadi '09, '10

Text mining in ImageNet

False negative

Firetruck: Any of various large trucks that carry firemen and equipment to the site of a fire



Challenges

Red-winged blackbird: blackbird with scarlet patches on the wings



Two-spotted ladybug: **Red** ladybug with a black spot on each wing

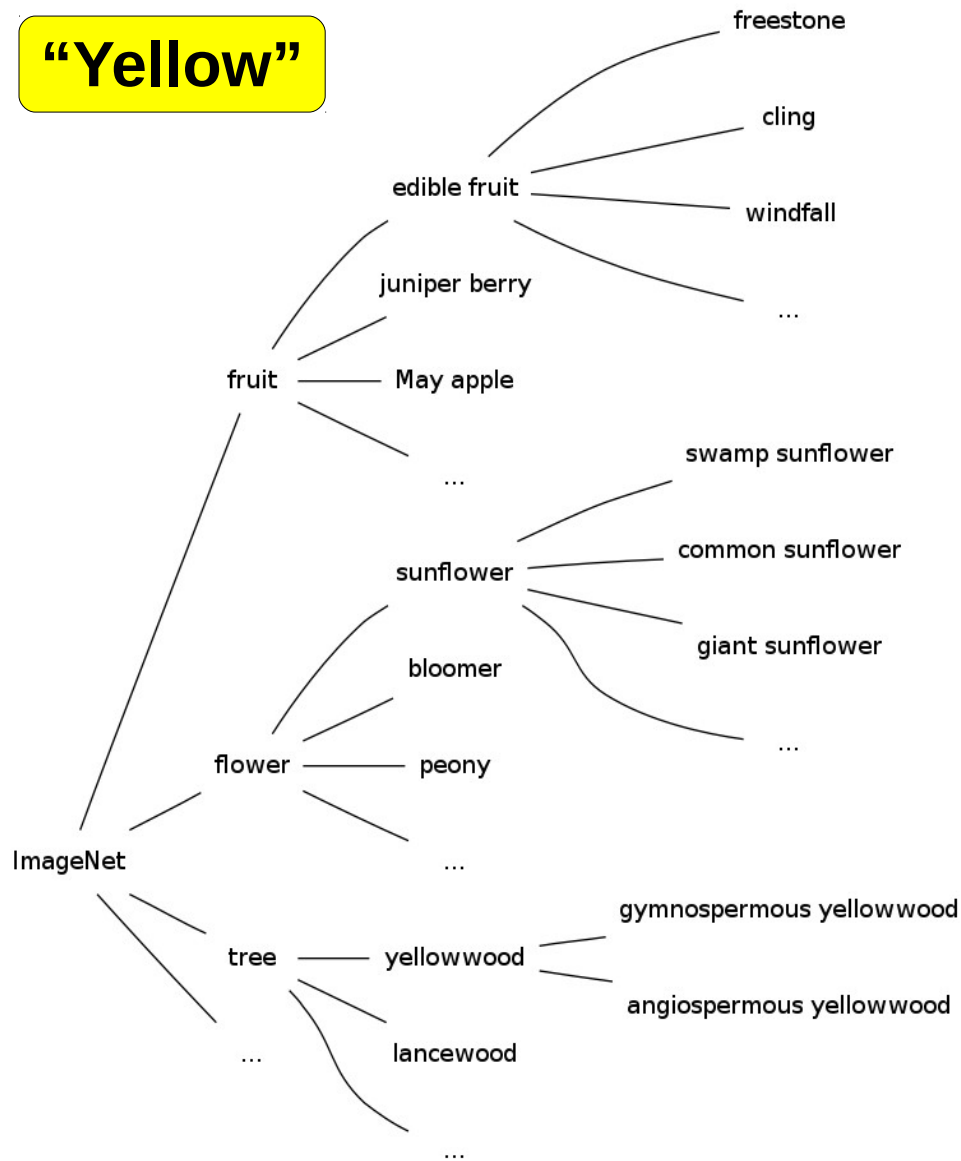


Coq-au-vin: Chicken and onions and mushrooms braised in **red** wine and seasonings



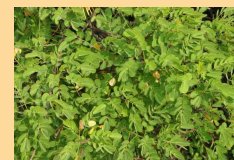
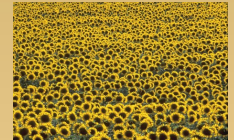
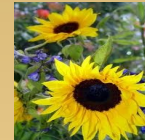
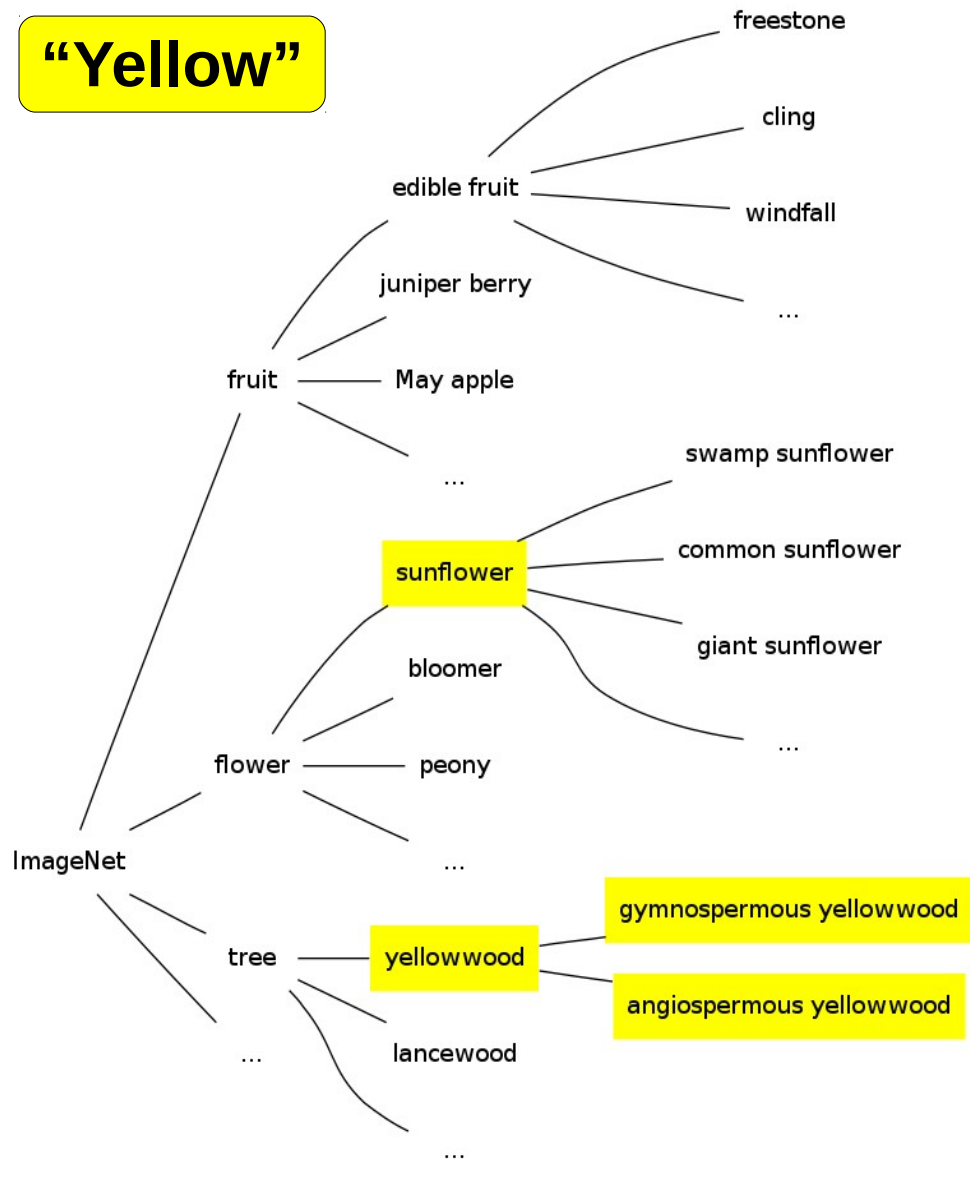
Semantic attribute labeling

“Yellow”



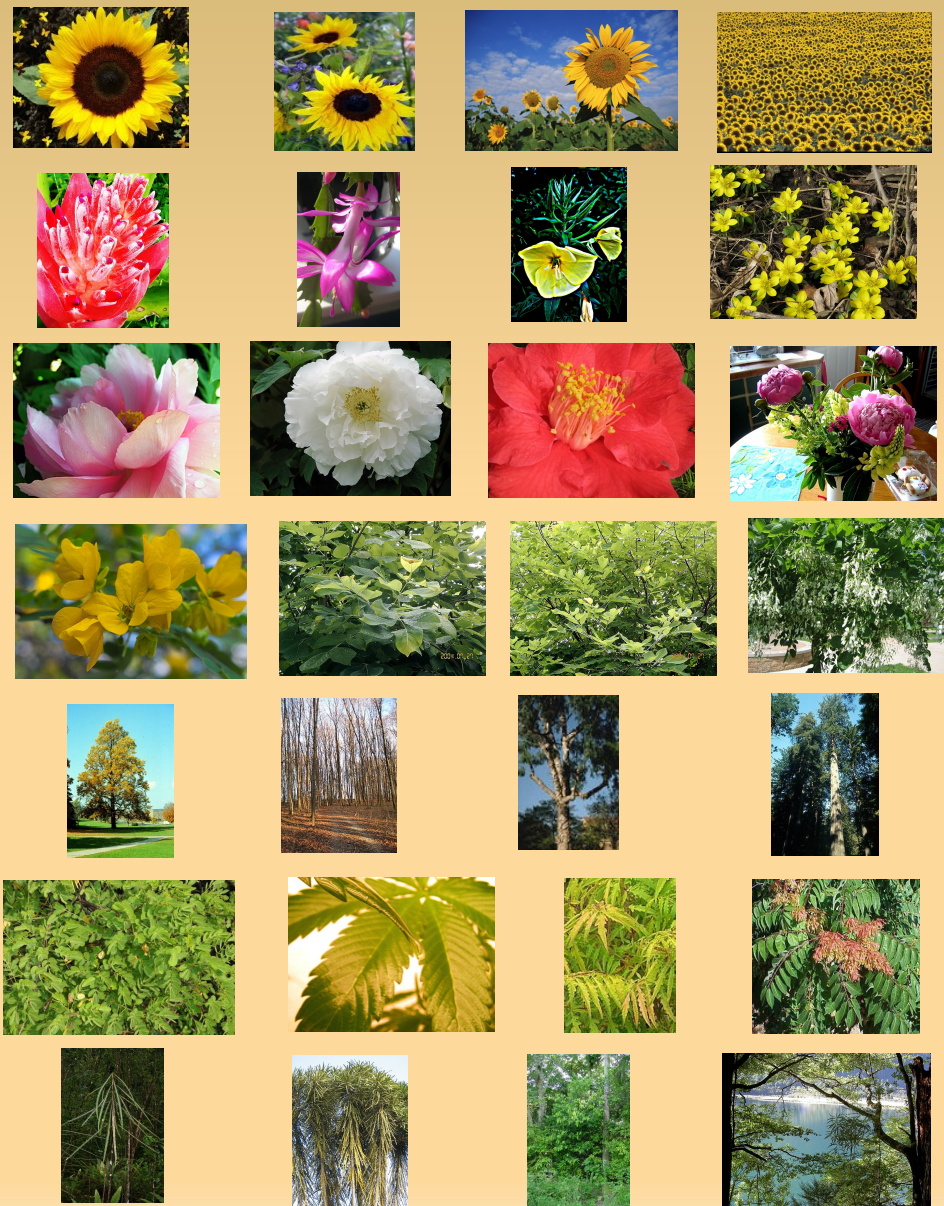
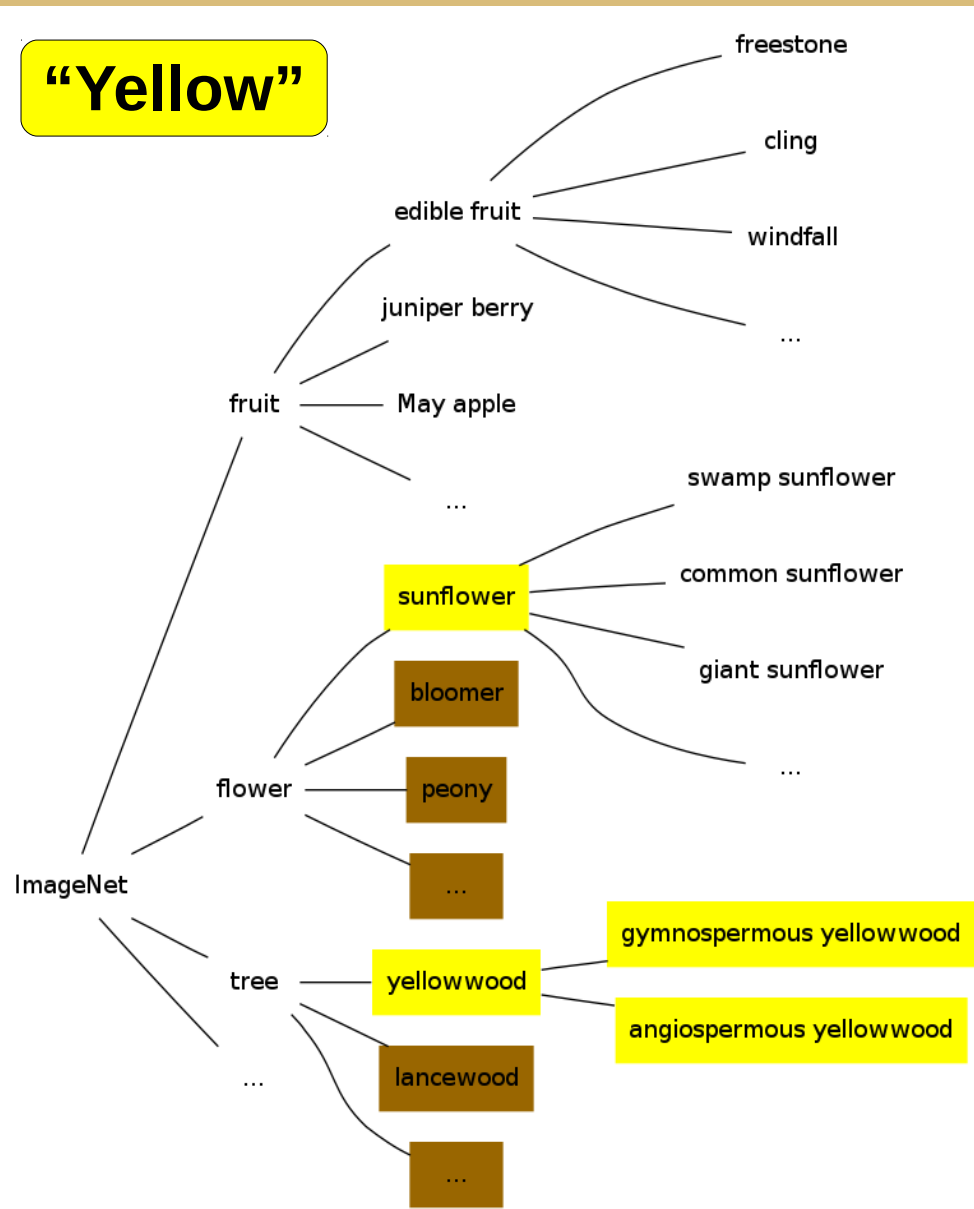
Semantic attribute labeling

“Yellow”



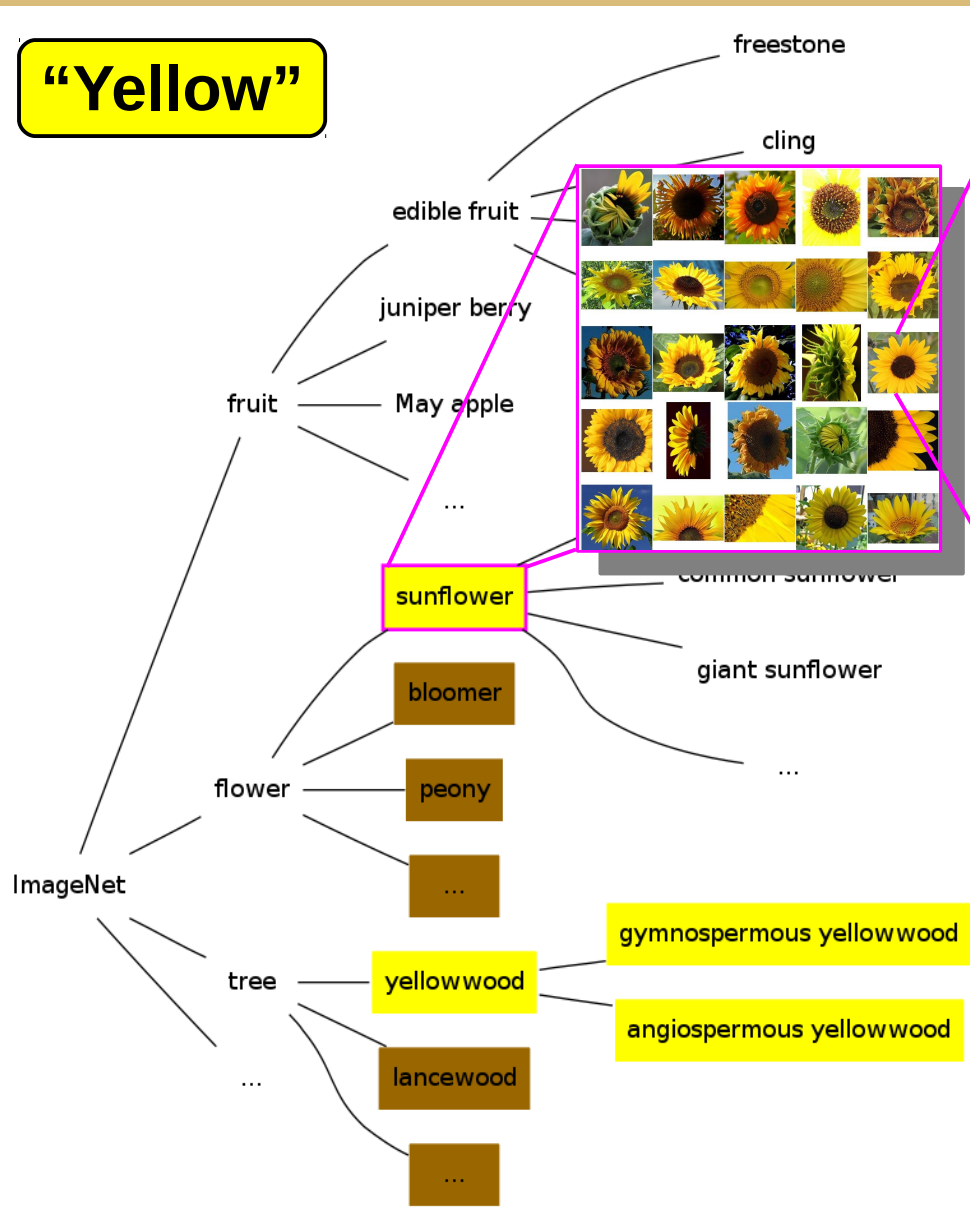
Semantic attribute labeling

“Yellow”



Semantic attribute labeling

“Yellow”



Find yellow

Consider the object in the image (not the background).
Is a significant part of the object **yellow**?

Image 15 out of 100



Yes, at least a quarter (25%) of this object is yellow

No, this object contains little or no yellow

If this page does not work properly, please try again with either Mozilla **Firefox** or Google **Chrome**.
Be sure to accept the HIT before starting.
You can submit after you're finished with all images.

Submit



ImageNet attribute dataset

384 synsets x 25 images each = **9600 images**
x **20 attributes**
x 3 or 4 people label each

Labels: 10% positive, 78% negative

<http://ai.stanford.edu/~olga>

20 semantic visual attributes

Color

Black, brown, gray, green, orange, red, white, yellow:



Texture

Furry, metallic, rough, shiny, smooth, wet, wooden:



Pattern

Spotted, striped:



Shape

Long, rectangular, round:



Overview

- Obtaining large-scale training data
 - Images
 - Attribute labels
- **Training and evaluating attribute classifiers**
- Performing higher-level tasks:
 - Targetted retrieval
 - Zero-shot learning
- Future directions

Training **striped** classifier



					...				
0	0.15	0	0.3	0	0	...	0.15	0.3	0.1

RGB codebook, size 50

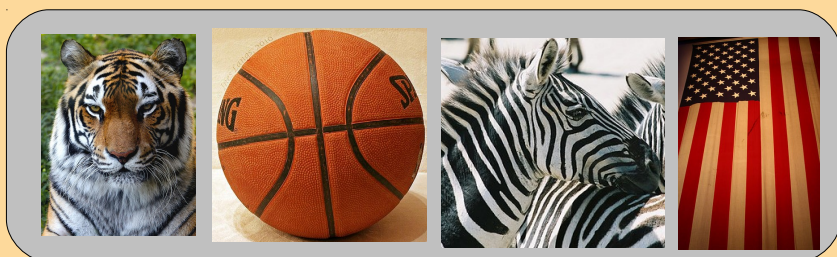
						...			
0.05	0.15	0.13	0.1	0.01	0	...	0.02	0	0

SIFT codebook, size 1000

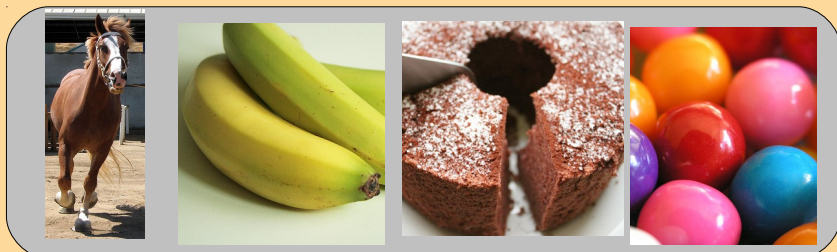
						...			
0.18	0	0.01	0.05	0.16	0.03	...	0.02	0	0.2

Shape context codebook, size 500

+



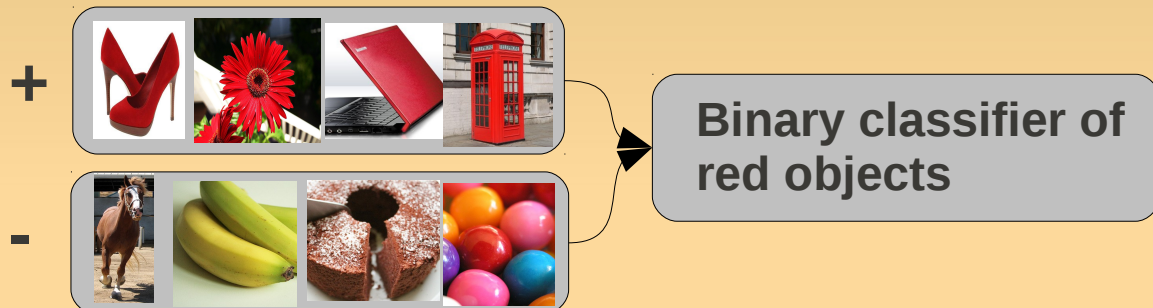
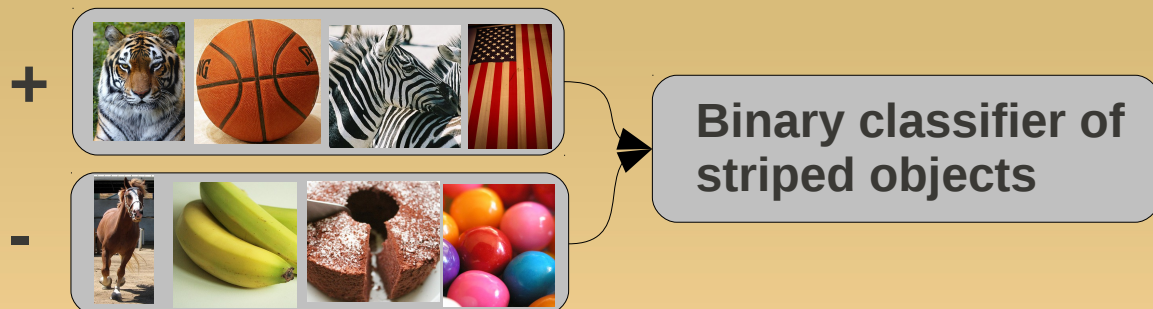
-



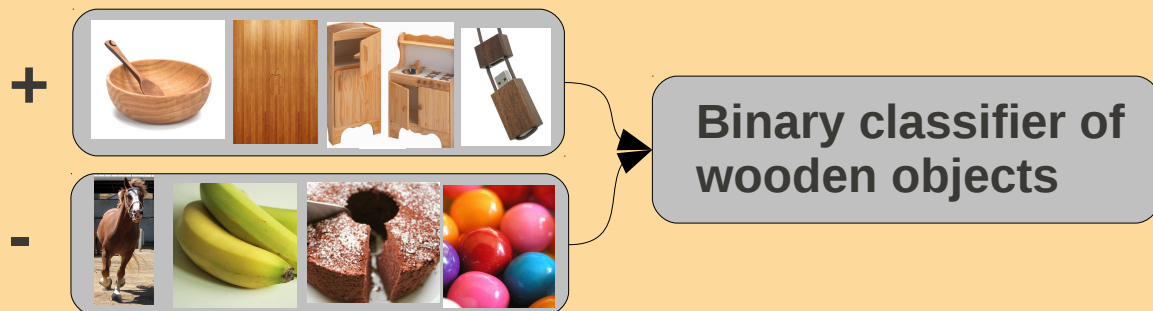
Binary classifier
of striped objects

**Histogram
intersection
kernel SVM**

Training attribute classifiers

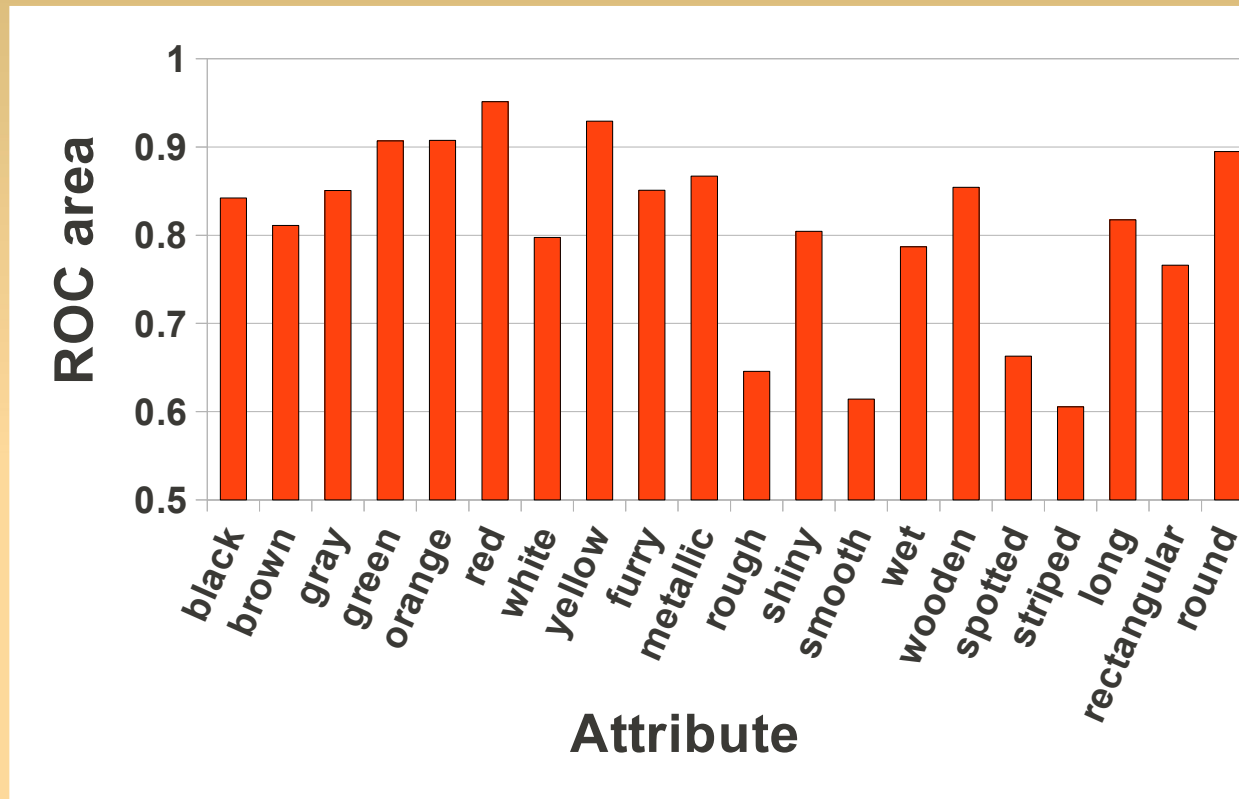


...



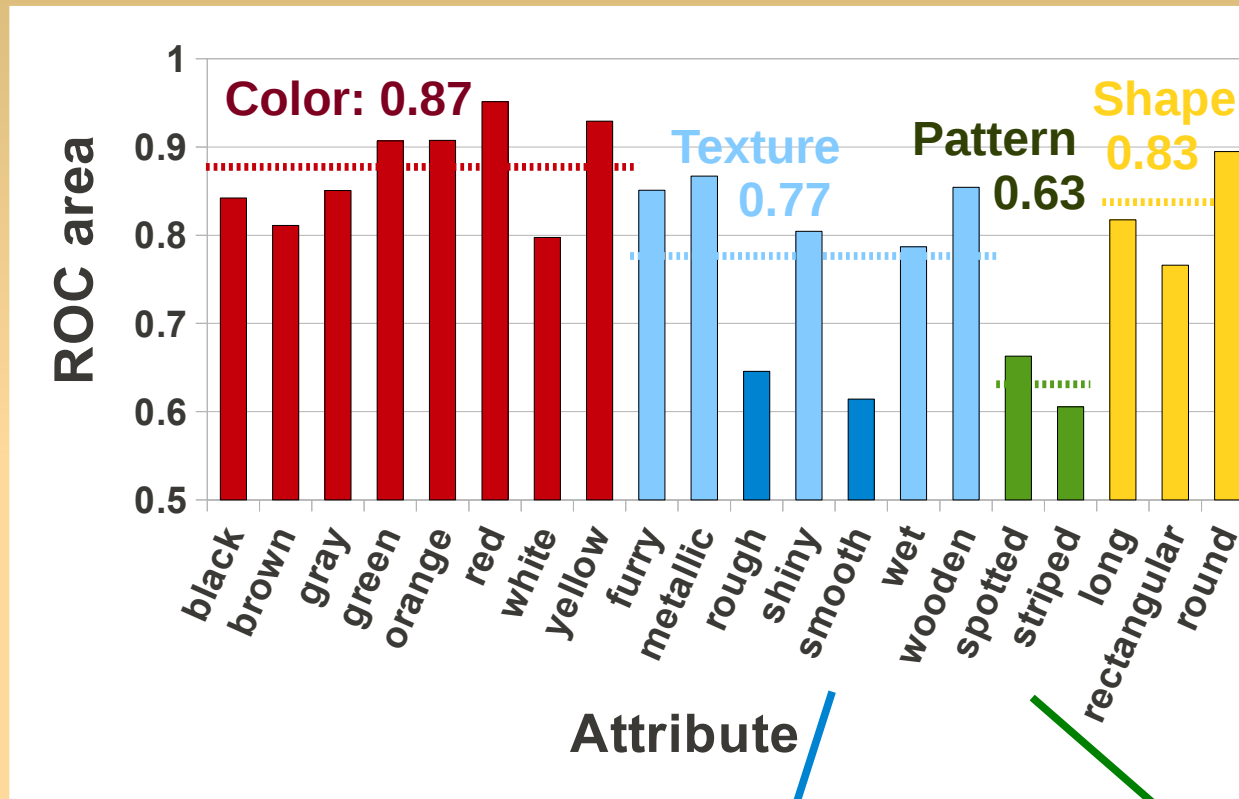
Binary classifiers for each of the 20 semantic attributes

Evaluating attribute classifiers



- Histogram intersection kernel SVM
- 5-fold cross-validation
- 100-2000 positive training example
- 2000-8000 negative training examples
- Regularization chosen on holdout set

Evaluating attribute classifiers



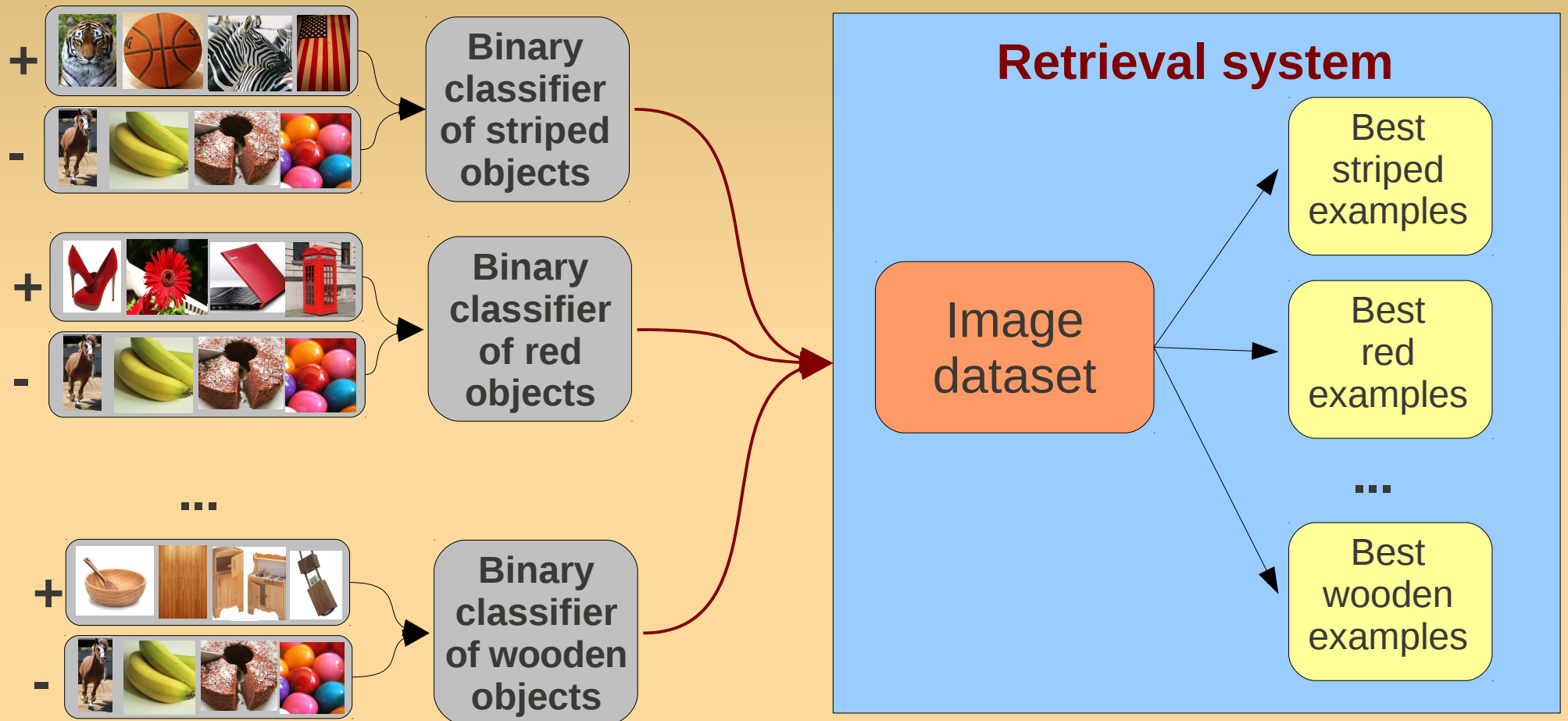
- **Lack of labeling consensus:**
34% images ambiguously labeled for smooth,
28% images ambiguously labeled for rough,
≤ 21% images for any other attribute

- Large variety
- Few training examples
- Global SIFT codebook

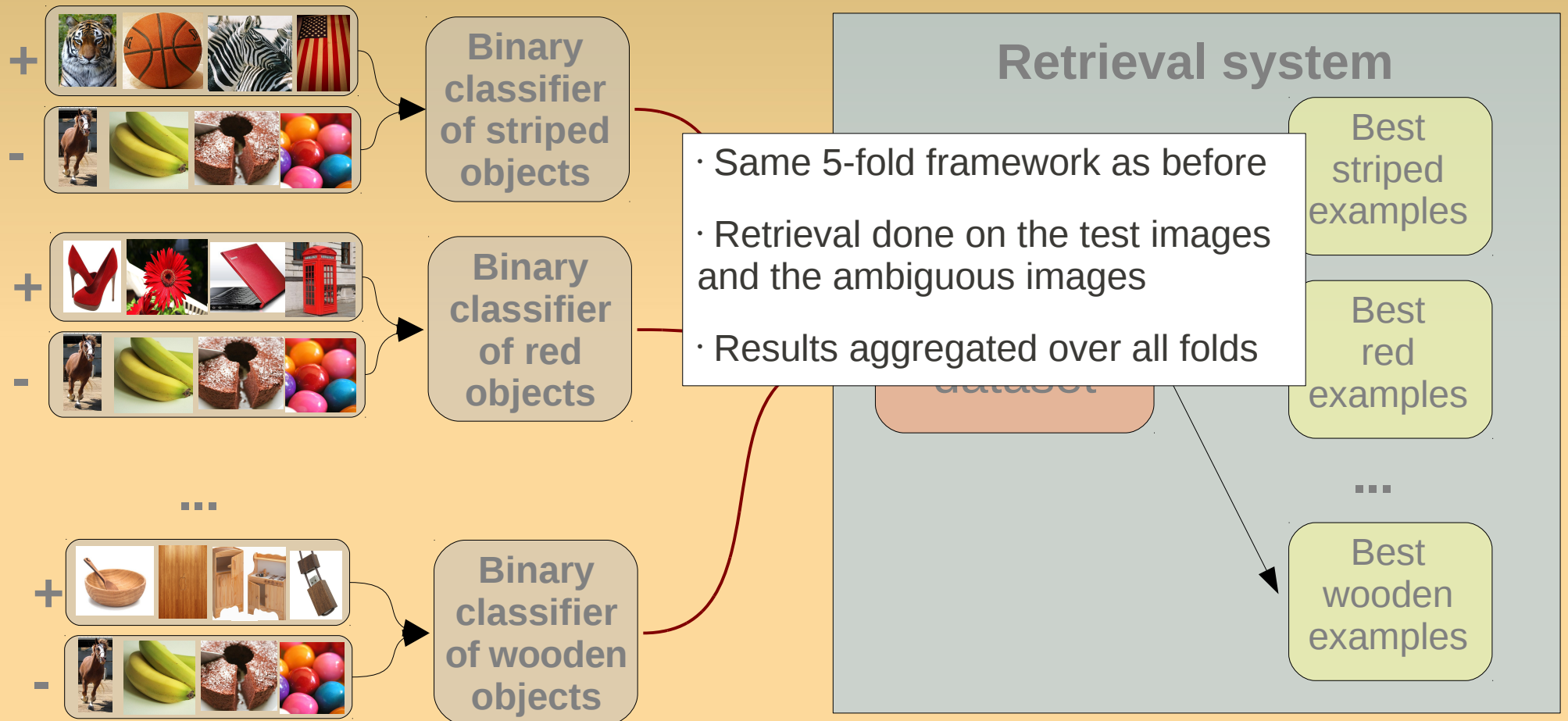
Overview

- Obtaining large-scale training data
 - Images
 - Attribute labels
- Training attribute classifiers
- **Performing higher-level tasks:**
 - **Targetted retrieval**
 - **Zero-shot learning**
- Future directions

Targetted retrieval



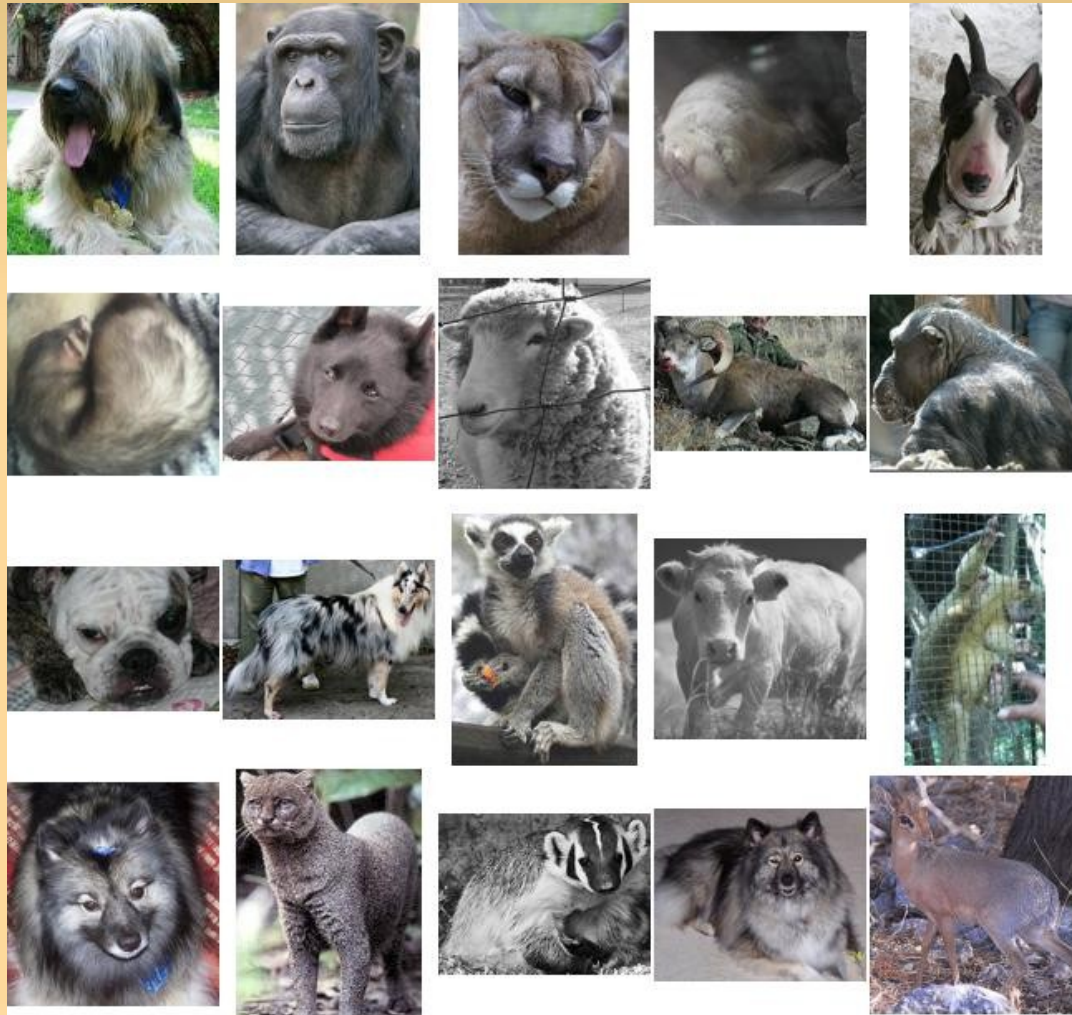
Targetted retrieval



Targetted retrieval

Gray objects

Top 20 retrieved images



To better interpret results, additionally train an L1-regularized logistic regression

Top 10 chosen features:



Legend

	Color
	SIFT
	Shape context

Targetted retrieval

Furry objects

Top 20 retrieved images



To better interpret results, additionally train an L1-regularized logistic regression

Top 10 chosen features:



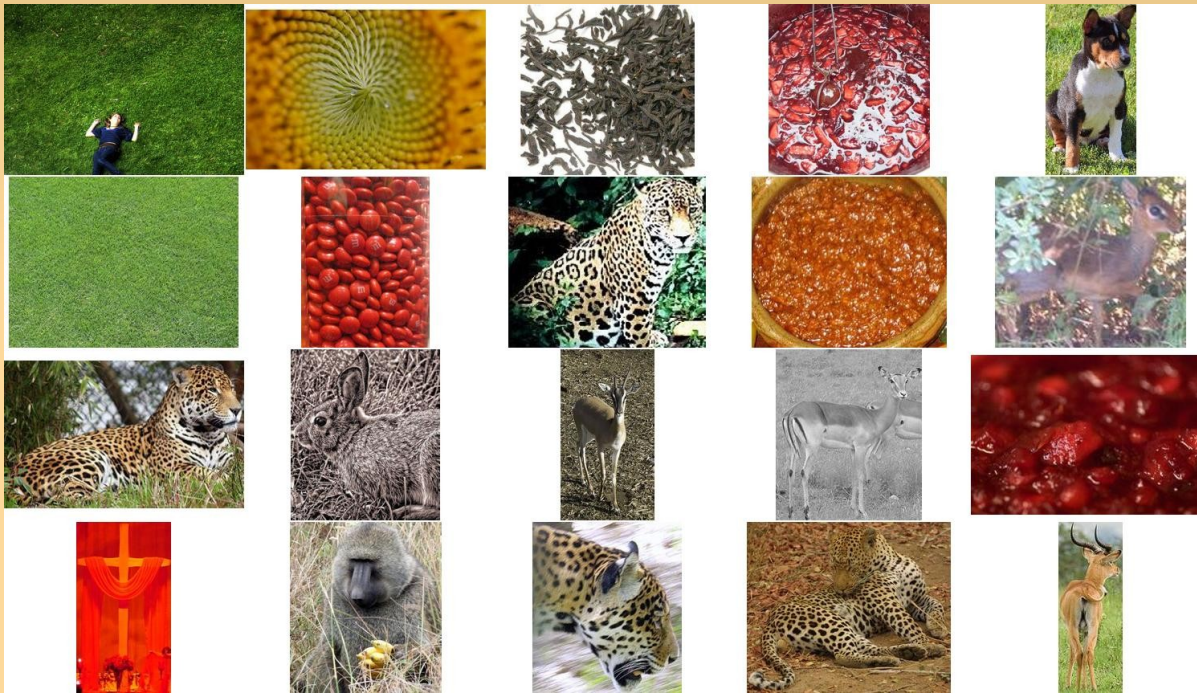
Legend

	Color
	SIFT
	Shape context

Targetted retrieval

Spotted objects

Top 20 retrieved images



To better interpret results, additionally train an L1-regularized logistic regression

Top 10 chosen features:



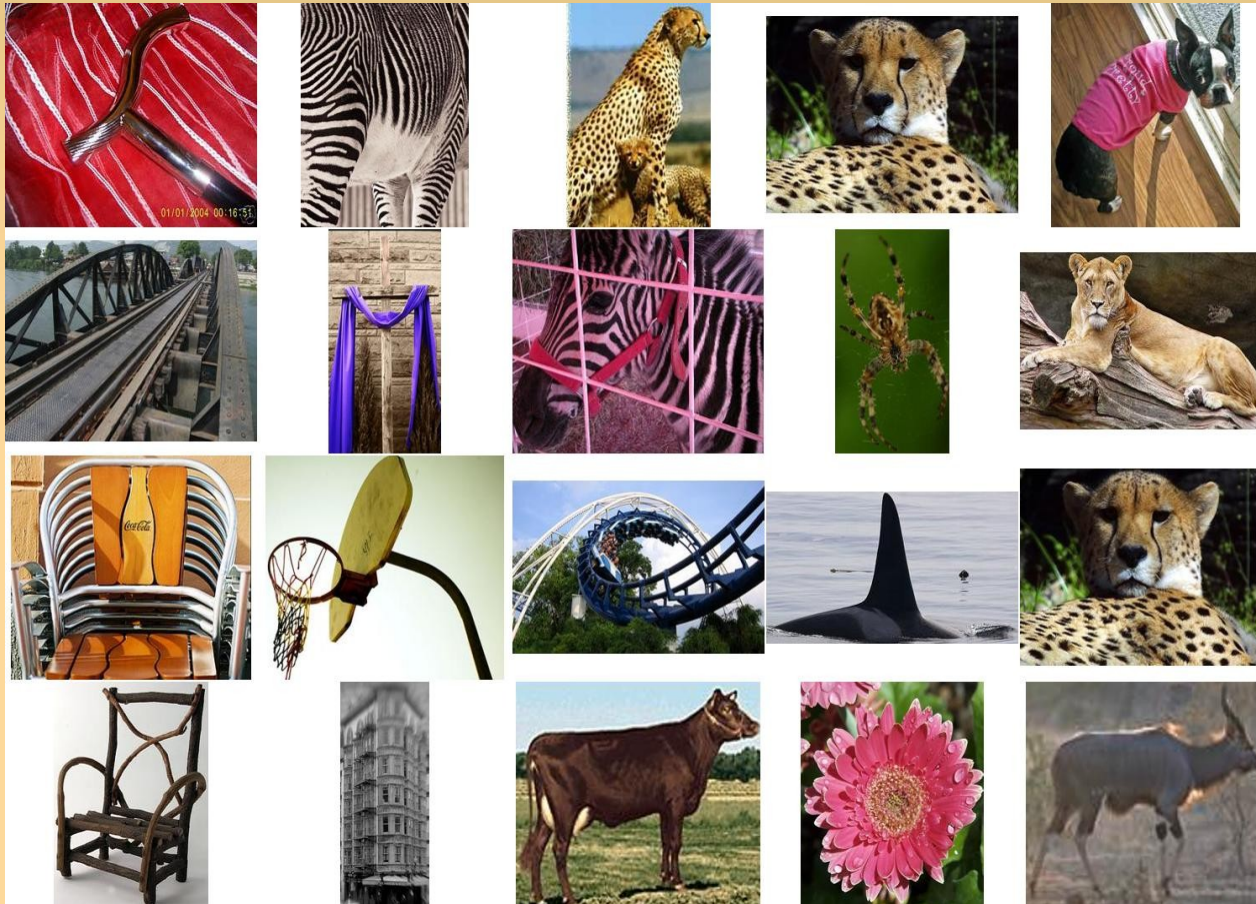
Legend

Color
SIFT
Shape context

Targetted retrieval

Striped objects

Top 20 retrieved images



To better interpret results, additionally train an L1-regularized logistic regression

Top 10 chosen features:



Legend

	Color
	SIFT
	Shape context

Targetted retrieval

Round objects

Top 20 retrieved images



To better interpret results, additionally train an L1-regularized logistic regression

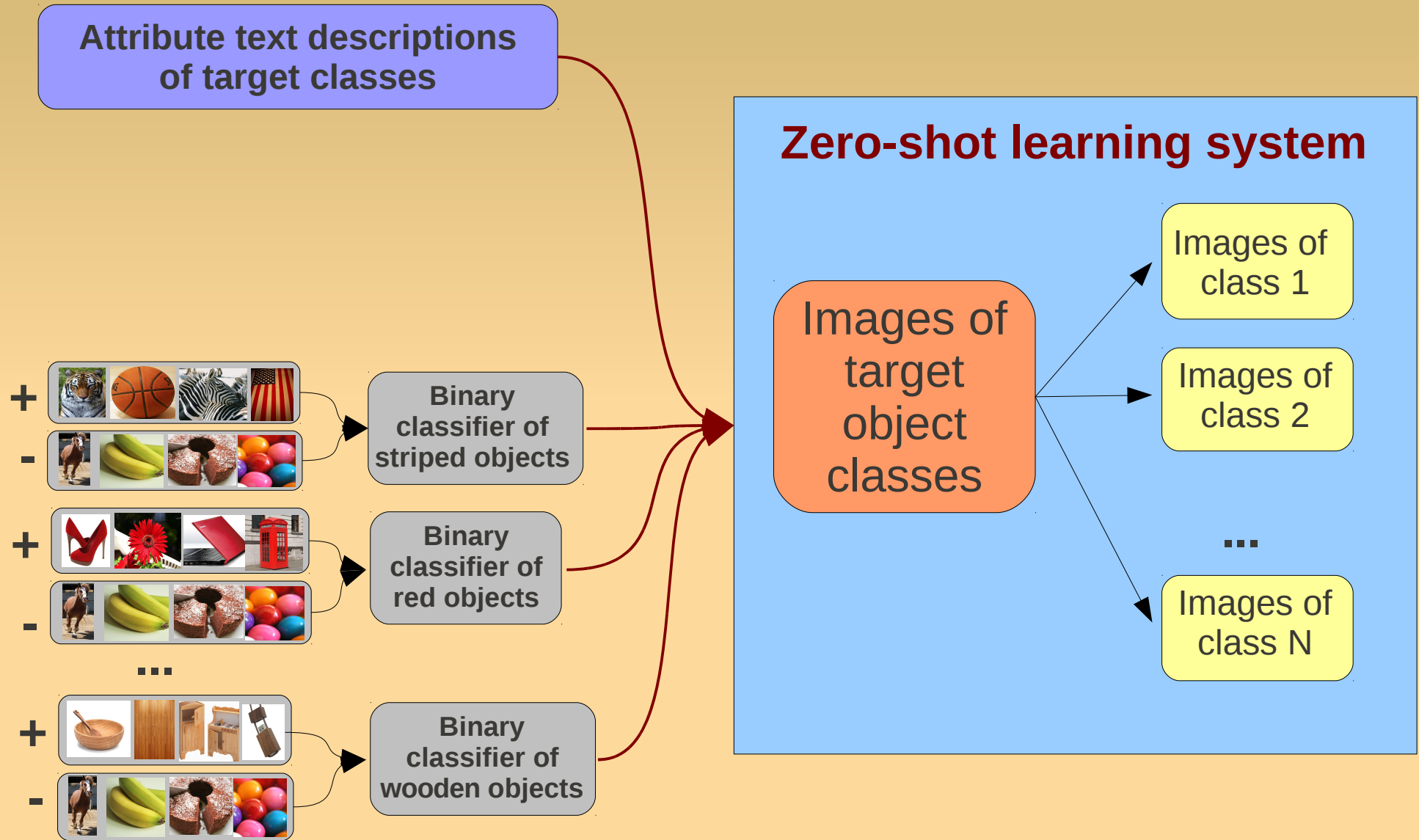
Top 10 chosen features:



Legend

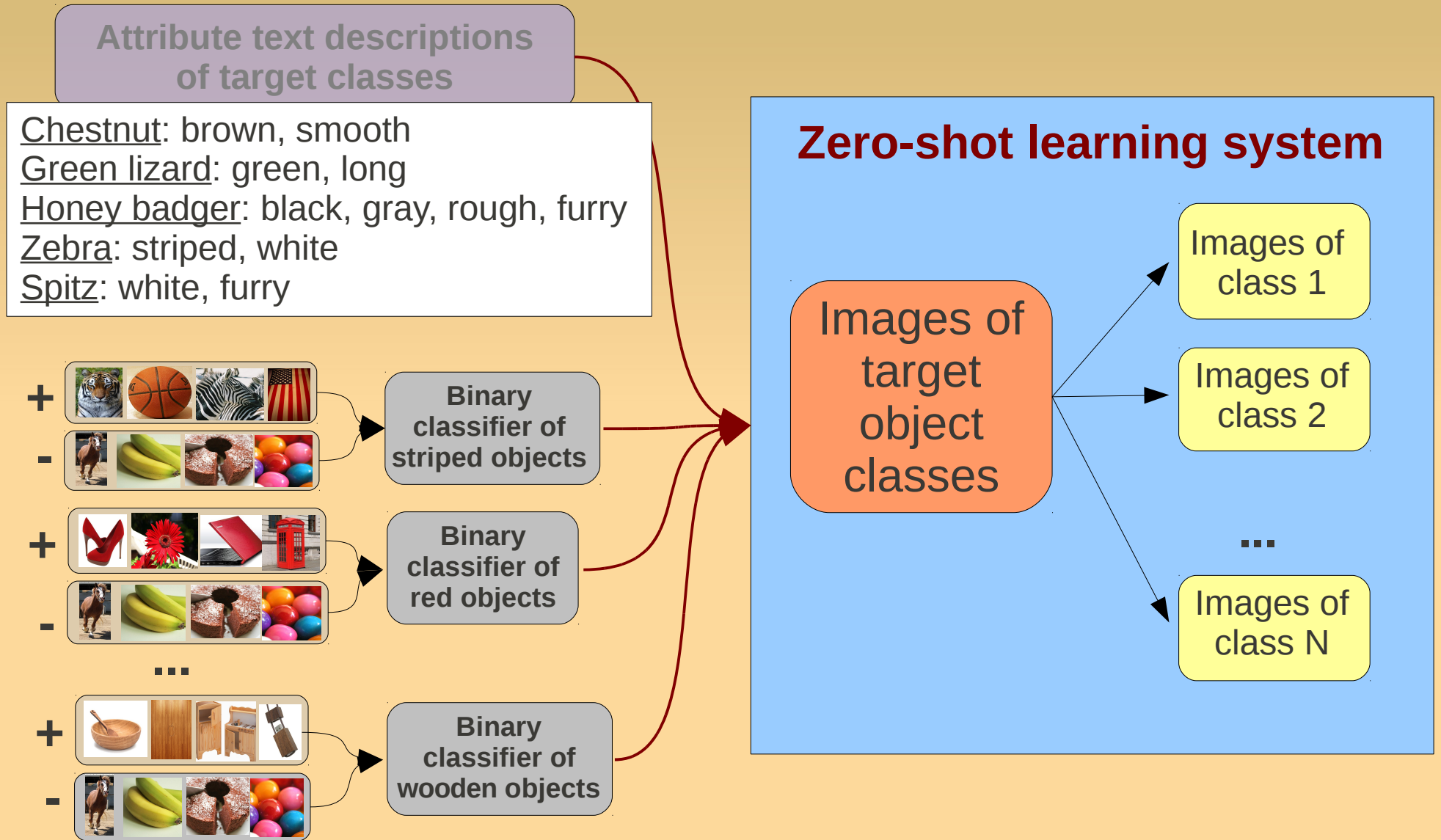
Red	Color
Blue	SIFT
Green	Shape context

Zero-shot learning



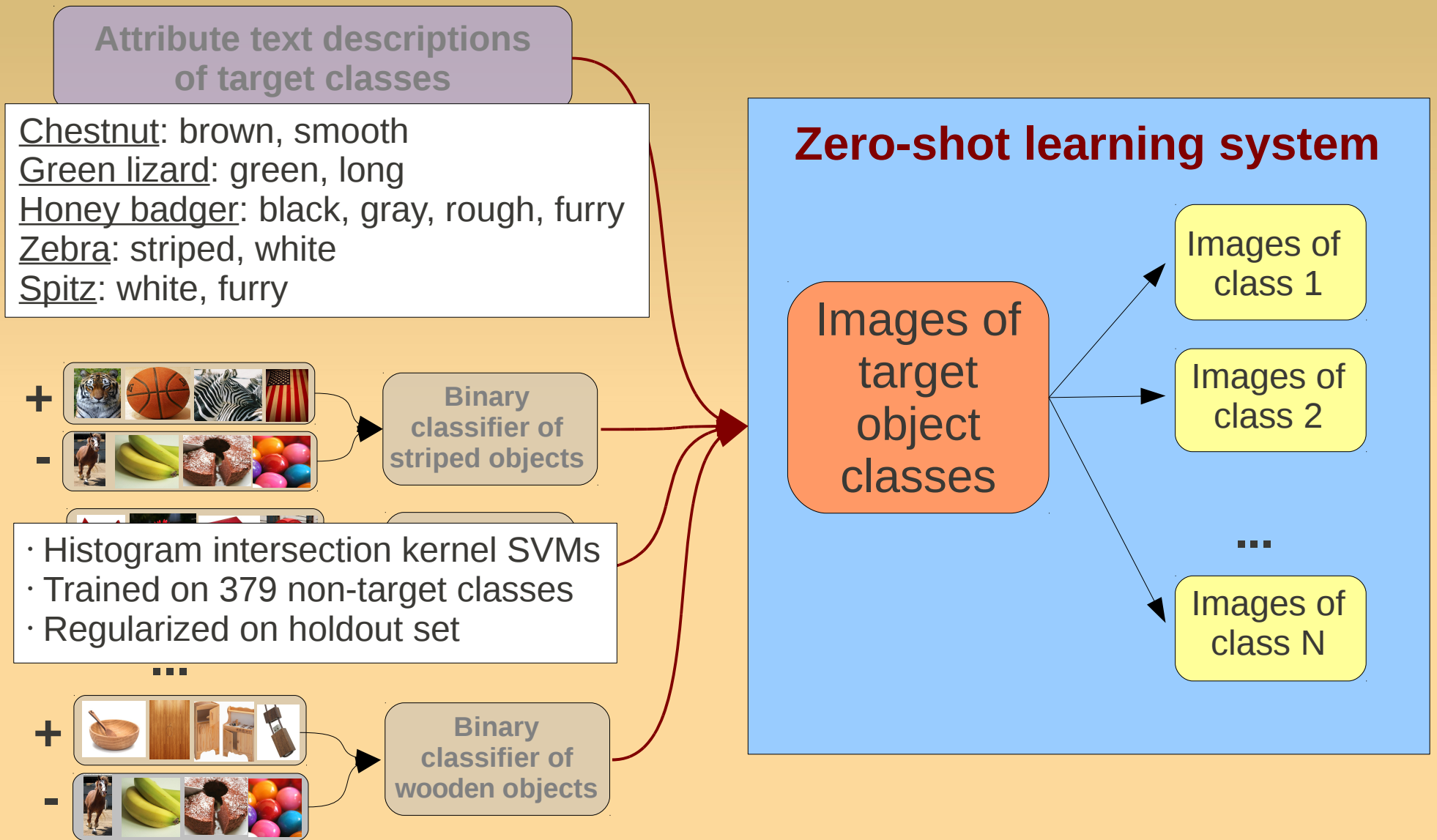
Classifiers trained on other object classes

Zero-shot learning



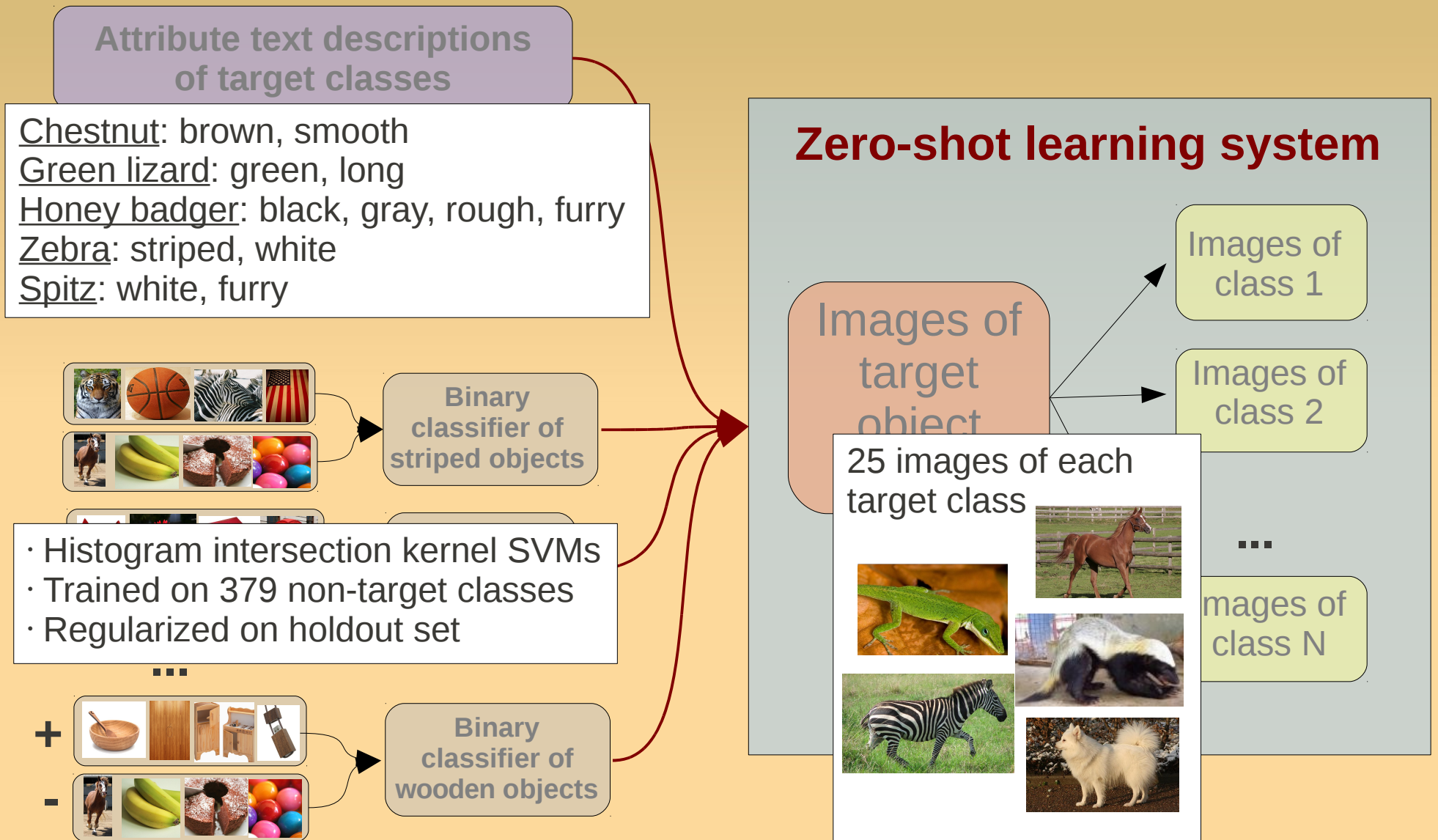
Classifiers trained on other object classes

Zero-shot learning



Classifiers trained on other object classes

Zero-shot learning



Classifiers trained on other object classes

Zero-shot learning

Chestnut: brown, smooth



Green lizard: green, long



Honey badger: black, gray, rough, furry



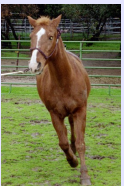
Zebra: black, white, striped, smooth



Spitz: white, furry



\mathbf{X} =



\mathbf{TS} =

Training Set

$$P(\text{chestnut} | \mathbf{X}) = \frac{P(\text{brown} | \mathbf{X}) \times P(\text{smooth} | \mathbf{X}) \times P(\text{not green} | \mathbf{X}) \times \dots}{P(\text{brown} | \mathbf{TS}) \times P(\text{smooth} | \mathbf{TS}) \times P(\text{not green} | \mathbf{TS}) \times \dots}$$

$$\text{Class}(\mathbf{X}) = \text{argmax} \{ P(\text{chestnut} | \mathbf{X}), P(\text{lizard} | \mathbf{X}), P(\text{zebra} | \mathbf{X}) \dots \}$$

Model of Lampert et al. "Learning to Detect Unseen Object Classes..." In CVPR, 2009.

Zero-shot learning

Chestnut: brown, smooth



Green lizard: green, long



Honey badger: black, gray, rough, furry



Zebra: black, white, striped, smooth

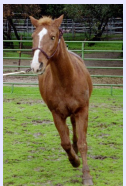


Spitz: white, furry



				
52	16	12	12	8
0	84	0	12	4
32	0	60	4	4
36	8	40	8	8
8	0	36	8	48

$\mathbf{X} =$



$\mathbf{TS} =$

Training Set

$$P(\text{chestnut} | \mathbf{X}) = \frac{P(\text{brown} | \mathbf{X}) \times P(\text{smooth} | \mathbf{X}) \times P(\text{not green} | \mathbf{X}) \times \dots}{P(\text{brown} | \mathbf{TS}) \times P(\text{smooth} | \mathbf{TS}) \times P(\text{not green} | \mathbf{TS}) \times \dots}$$

$$\text{Class}(\mathbf{X}) = \text{argmax} \{ P(\text{chestnut} | \mathbf{X}), P(\text{lizard} | \mathbf{X}), P(\text{zebra} | \mathbf{X}) \dots \}$$

Model of Lampert et al. "Learning to Detect Unseen Object Classes..." In CVPR, 2009.

Overview

- Obtaining large-scale training data
 - Images
 - Attribute labels
- Training attribute classifiers
- Performing higher-level tasks:
 - Targetted retrieval
 - Zero-shot learning
- **Future directions**

More attributes

Global appearance

Local parts

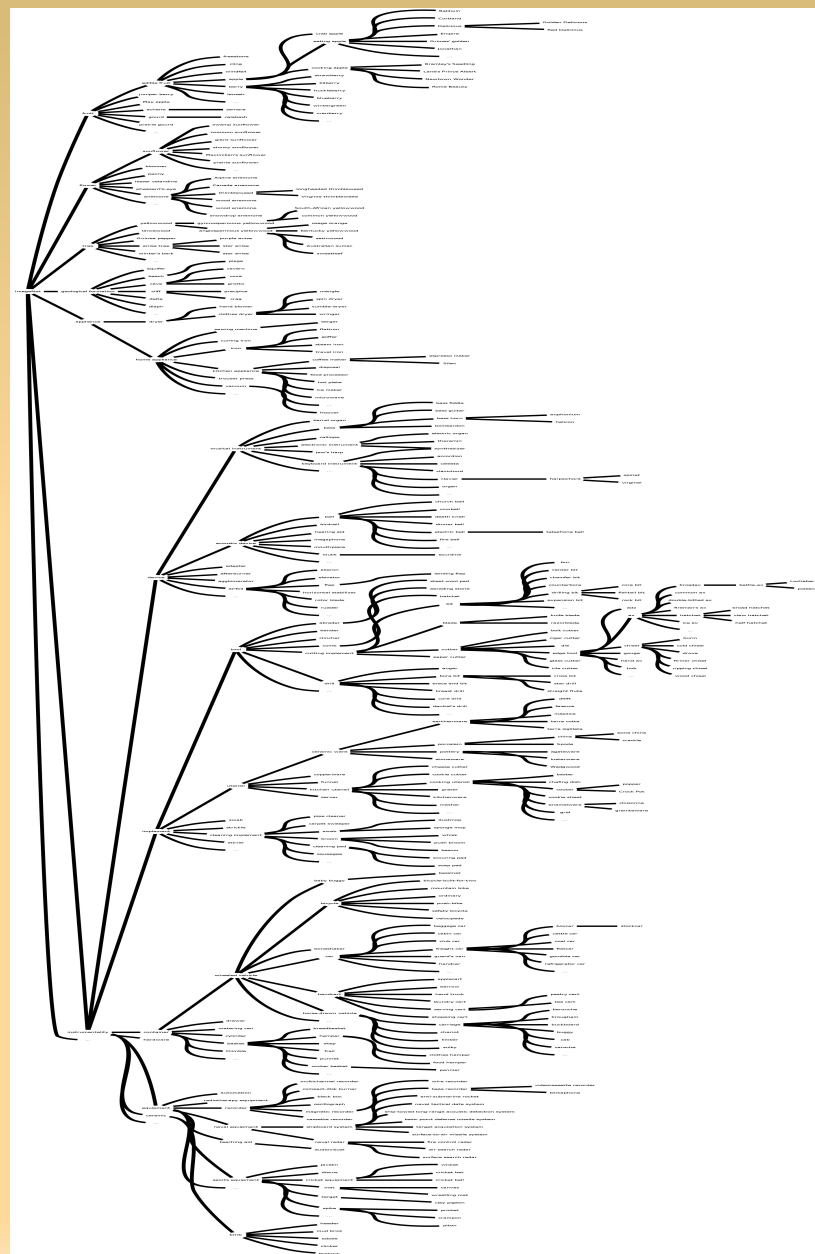
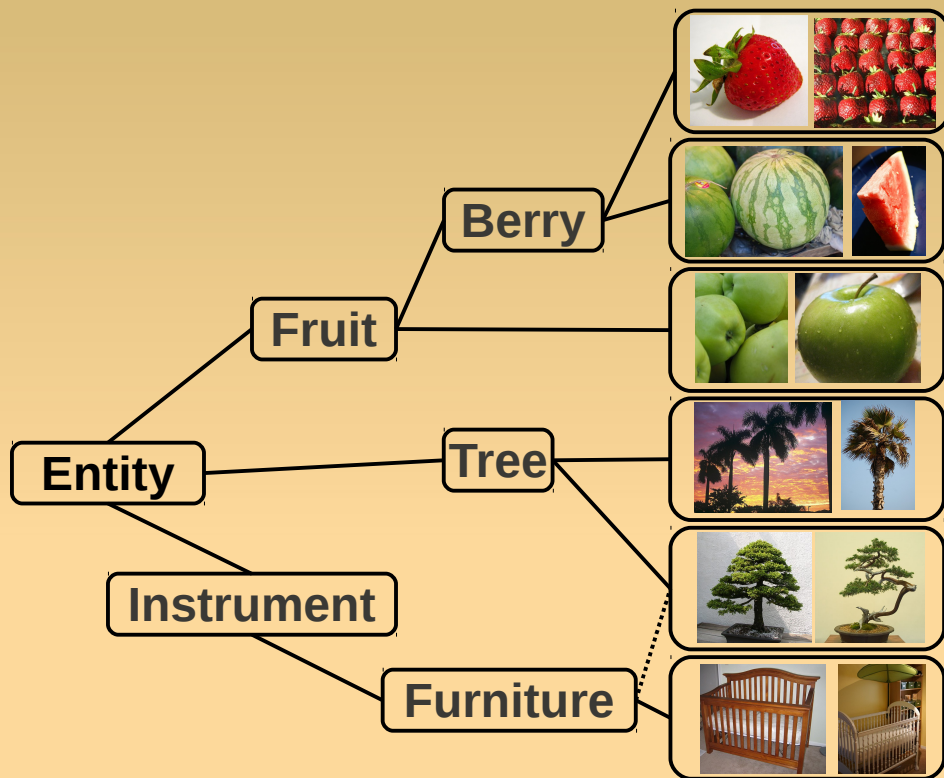
Semantic



Similarity-based



More object classes



Attribute learning in large-scale datasets

Olga Russakovsky and Li Fei-Fei



Many thanks to Alex Berg, Jia Deng, Li-Jia Li, Bangpeng Yao, Juan Carlos Niebles, and all of Stanford vision lab.