Introduction to Computer Vision

Michael J. Black

Object Recognition

News

- No class on Wednesday. Have a good Thanksgiving break.
- ½ of project proposals have received feedback. The other ½ should get it today by tonight (maybe late).
- Assignment 3 back soon.
- Final project due Dec 16.
  - Keep it simple!

Goals

Focused on foundations of computer vision.

One big thing missing.

Today and next week

* Object Recognition.
* Final thoughts, what's left to do.
* Course evaluations.

Object Recognition

In Assignment 2 you learned a model of mouths.

- purely based on appearance
- aligned data
- one class (with some variation)
- un-occluded
- standard viewpoint

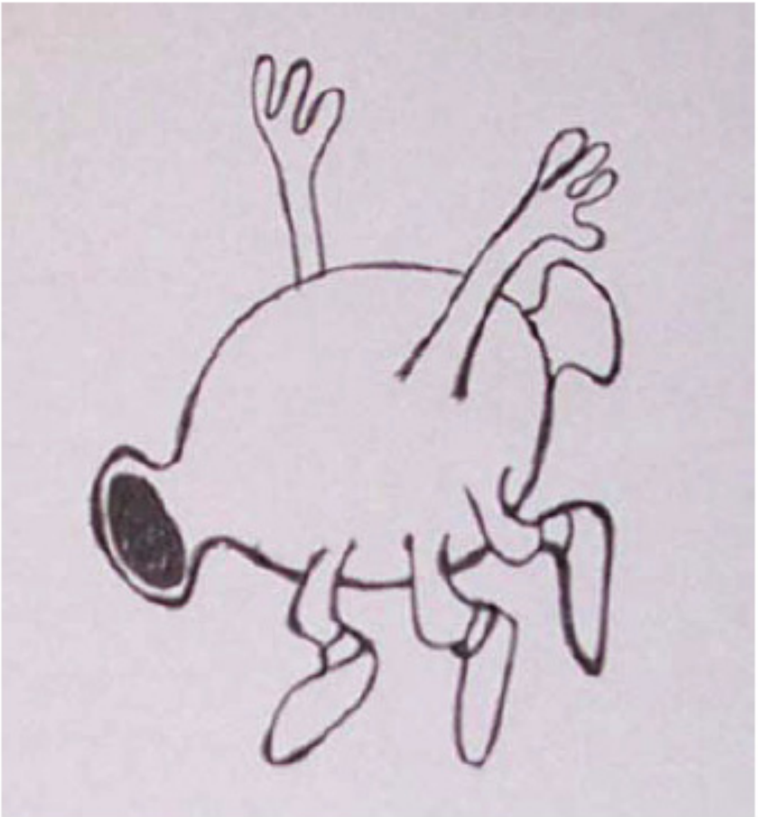
Recognizing and Learning Object Categories: Year 2007

Li Fei-Fei, Princeton

Rob Fergus, MIT

Antonio Torralba, MIT

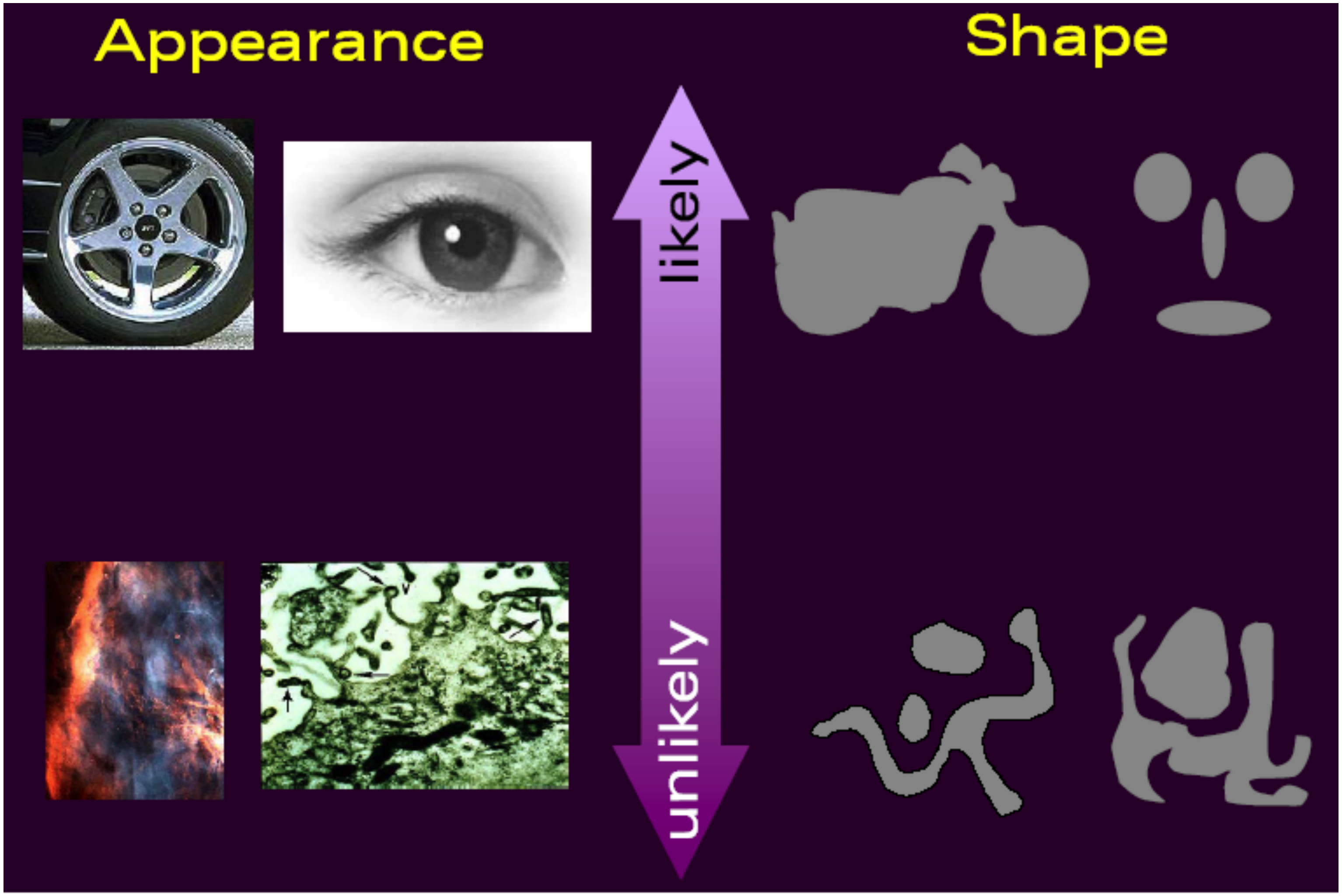
One-Shot Learning

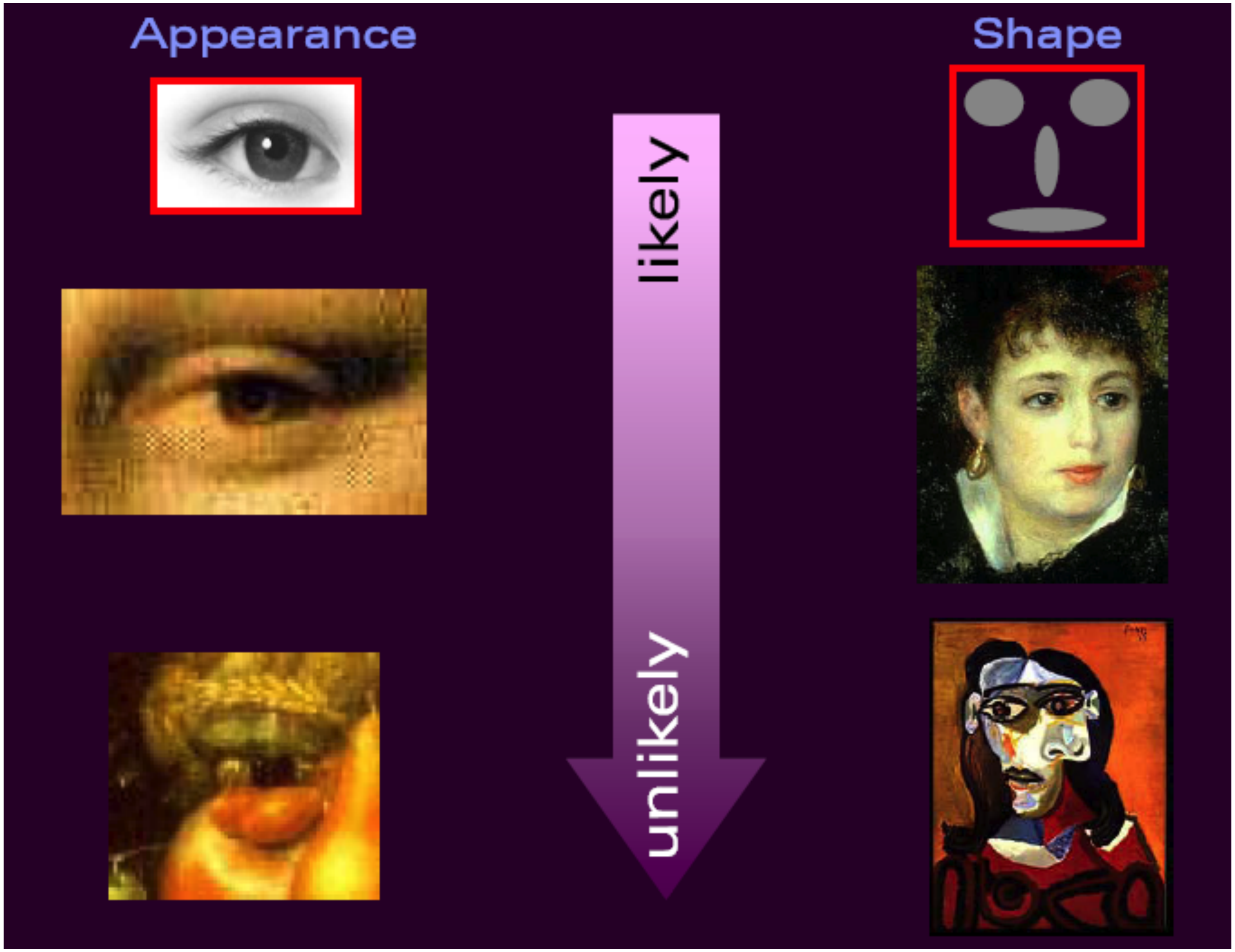


S. Savarese, 2003


P. Buegel, 1562


~10,000 to 30,000


Appearance
Shape
likely
unlikely


Appearance
Shape
likely
unlikely

Object categorization: the statistical viewpoint

$$p(zebra \mid image)$$

vs.

$$p(no\ zebra \mid image)$$

- Bayes rule:

$$\underbrace{\frac{p(zebra \mid image)}{p(no\ zebra \mid image)}}_{\text{posterior ratio}} = \underbrace{\frac{p(image \mid zebra)}{p(image \mid no\ zebra)}}_{\text{likelihood ratio}} \cdot \underbrace{\frac{p(zebra)}{p(no\ zebra)}}_{\text{prior ratio}}$$

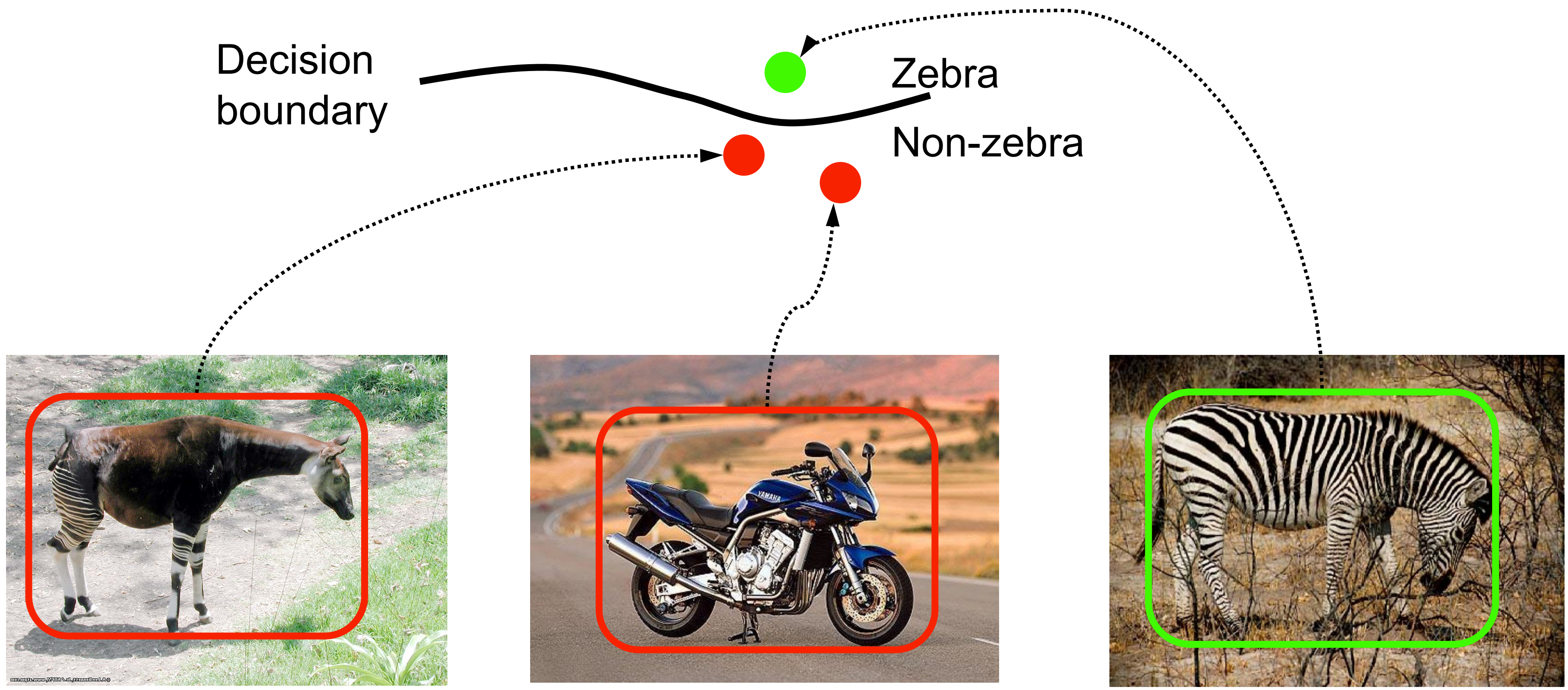
Object categorization: the statistical viewpoint

$$\underbrace{\frac{p(zebra \mid image)}{p(no\ zebra \mid image)}}_{\text{posterior ratio}} = \underbrace{\frac{p(image \mid zebra)}{p(image \mid no\ zebra)}}_{\text{likelihood ratio}} \cdot \underbrace{\frac{p(zebra)}{p(no\ zebra)}}_{\text{prior ratio}}$$

- **Discriminative methods model posterior**
- **Generative methods model likelihood and prior**

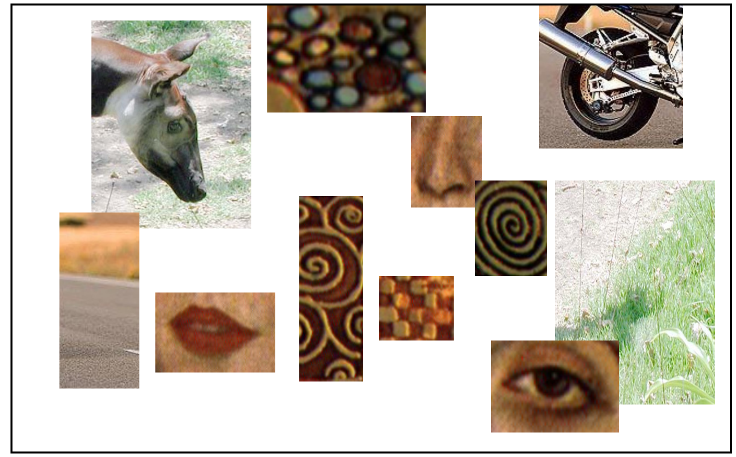
Discriminative

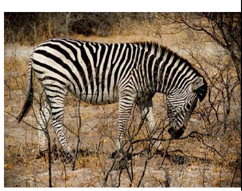
- Direct modeling of $\dfrac{p(zebra \mid image)}{p(no\ zebra \mid image)}$



Generative

- Model $p(image \mid zebra)$ and $p(image \mid no\ zebra)$



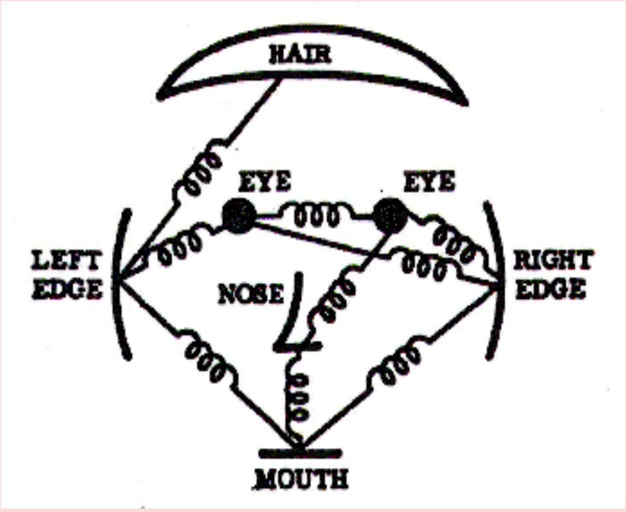
	$p(image \mid zebra)$	$p(image \mid no\ zebra)$
	Low	Middle
	High	Middle→Low

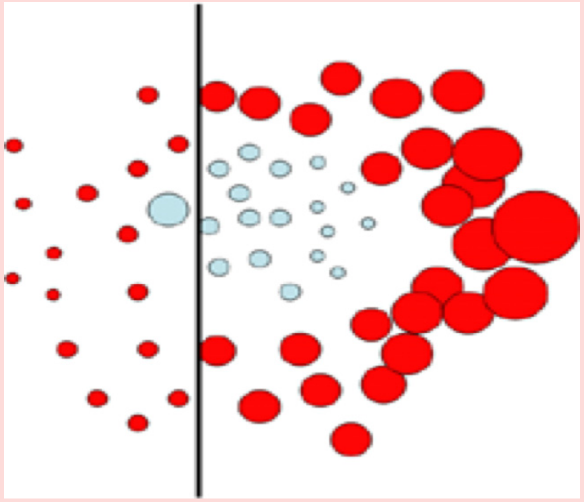
Three main issues

- Representation
  – How to represent an object category
- Learning
  – How to form the classifier, given training data
- Recognition
  – How the classifier is to be used on novel data

Representation

– Generative / discriminative / hybrid

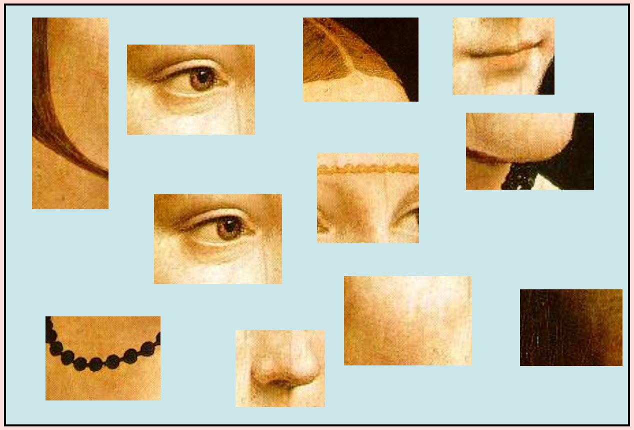


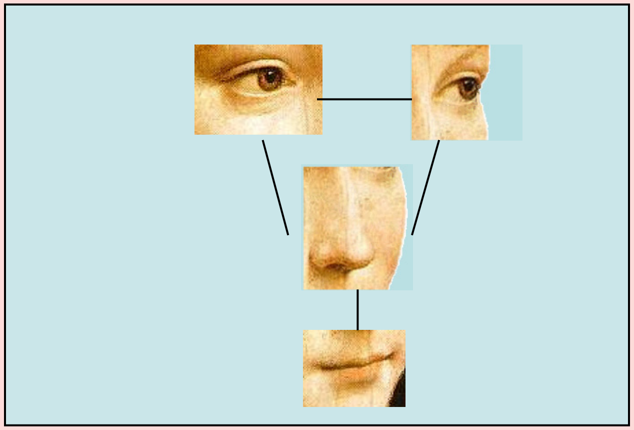


Representation

– Generative / discriminative / hybrid

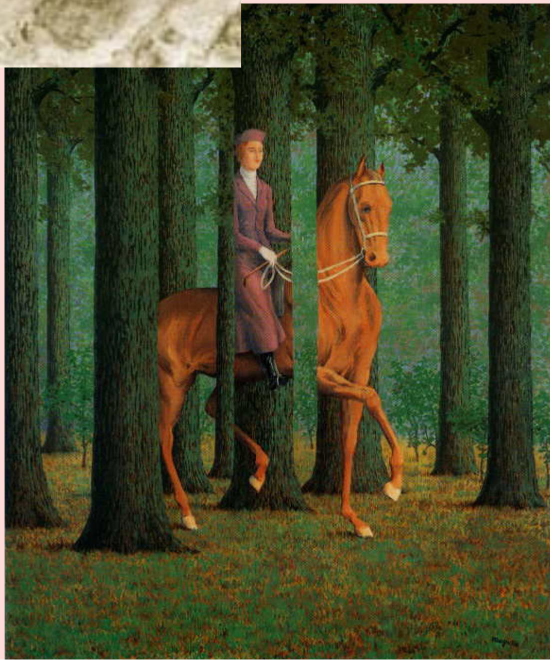
– Appearance only or location and appearance





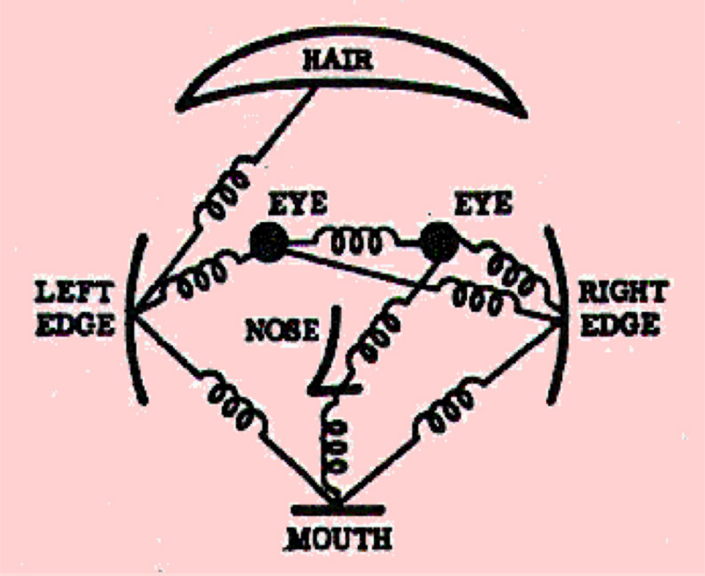
Representation

- Generative / discriminative / hybrid
- Appearance only or location and appearance
- Invariances
  - View point
  - Illumination
  - Occlusion
  - Scale
  - Deformation
  - Clutter
  - etc.



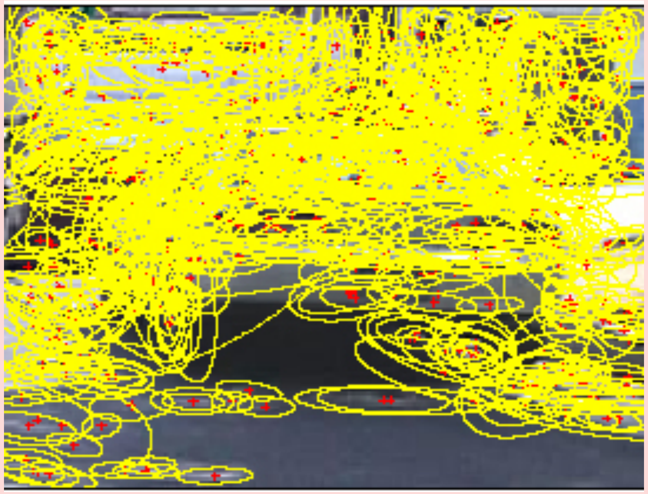
Representation

- Generative / discriminative / hybrid
- Appearance only or location and appearance
- invariances
- Part-based or global w/ sub-window



Representation

– Generative / discriminative / hybrid
– Appearance only or location and appearance
– invariances
– Parts or global w/sub-window
– Use set of features or each pixel in image

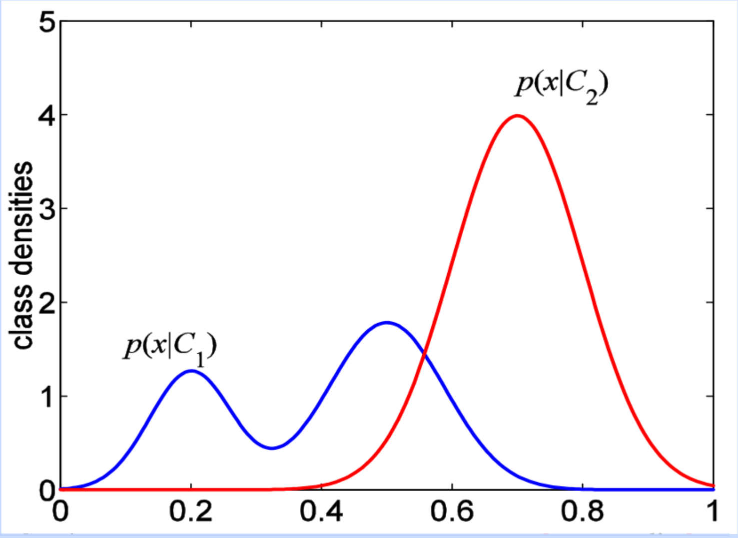


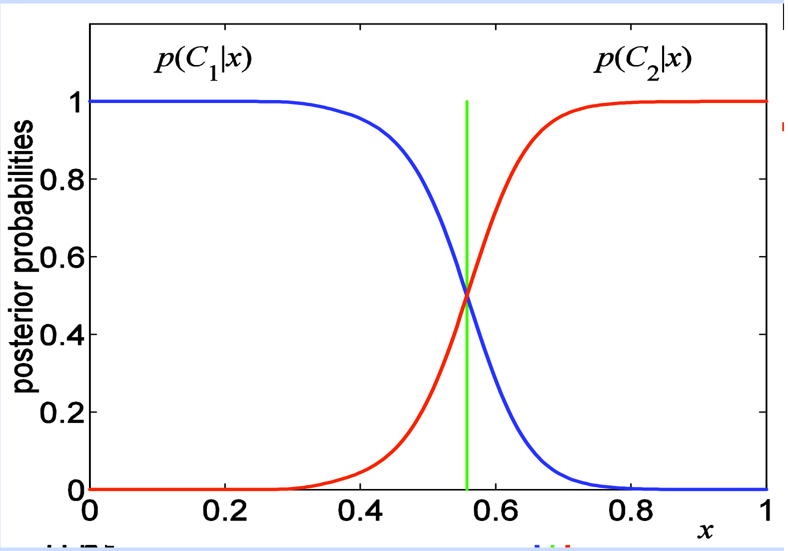
Learning

– Unclear how to model categories, so we learn what distinguishes them rather than manually specify the difference -- hence current interest in machine learning

Learning

– Unclear how to model categories, so we learn what distinguishes them rather than manually specify the difference -- hence current interest in machine learning)
– Methods of training: generative vs. discriminative

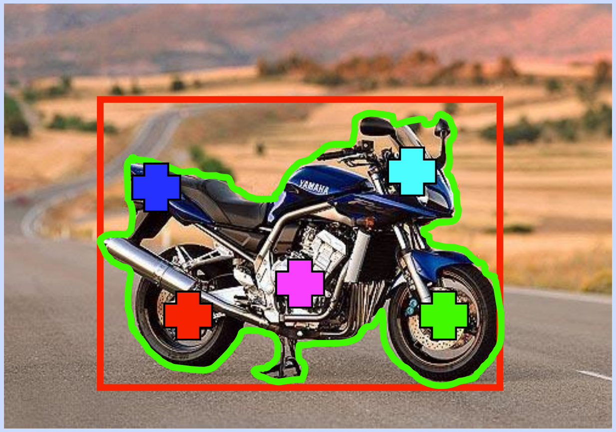




Learning

- Unclear how to model categories, so we learn what distinguishes them rather than manually specify the difference -- hence current interest in machine learning)
- What are you maximizing? Likelihood (Gen.) or performances on train/validation set (Disc.)
- Level of supervision
  - Manual segmentation; bounding box; image labels; noisy labels

Contains a motorbike



Learning

- Unclear how to model categories, so we learn what distinguishes them rather than manually specify the difference -- hence current interest in machine learning)
- What are you maximizing? Likelihood (Gen.) or performances on train/validation set (Disc.)
- Level of supervision
  - Manual segmentation; bounding box; image labels; noisy labels
- Batch/incremental (on category and image level; user-feedback)

Learning

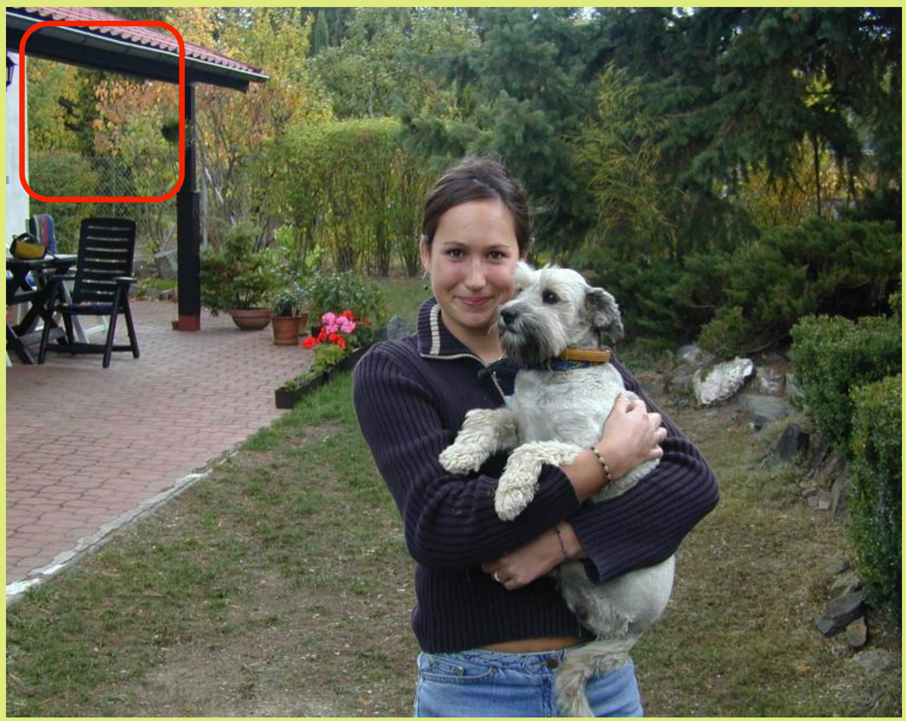
- Unclear how to model categories, so we learn what distinguishes them rather than manually specify the difference -- hence current interest in machine learning)
- What are you maximizing? Likelihood (Gen.) or performances on train/validation set (Disc.)
- Level of supervision
  - Manual segmentation; bounding box; image labels; noisy labels
- Batch/incremental (on category and image level; user-feedback)
- Training images:
  - Issue of overfitting
  - Negative images for discriminative methods Priors

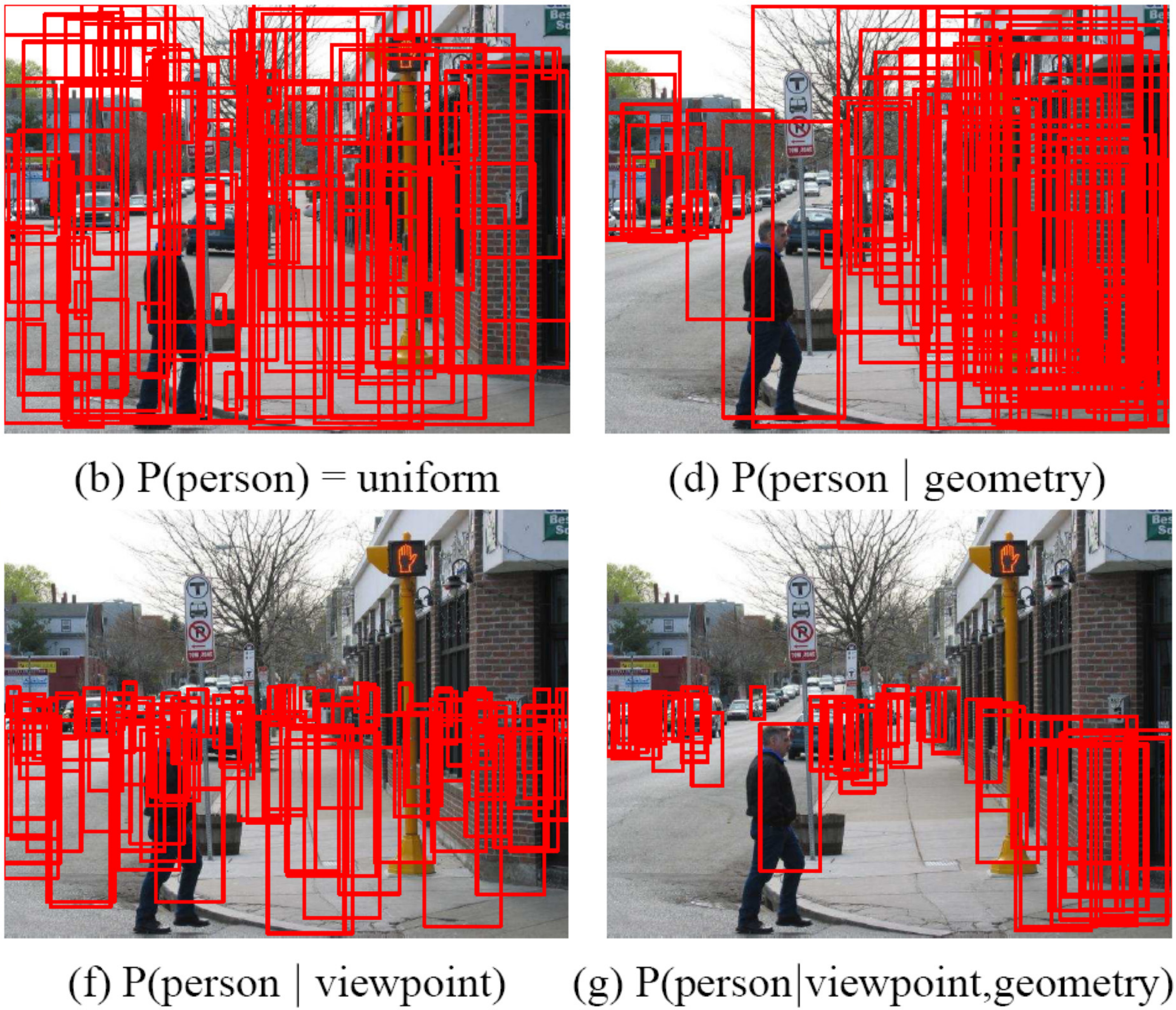
Learning

- Unclear how to model categories, so we learn what distinguishes them rather than manually specify the difference -- hence current interest in machine learning)
- What are you maximizing? Likelihood (Gen.) or performances on train/validation set (Disc.)
- Level of supervision
  - Manual segmentation; bounding box; image labels; noisy labels
- Batch/incremental (on category and image level; user-feedback)
- Training images:
  - Issue of overfitting
  - Negative images for discriminative methods
- Priors

Recognition

- Scale / orientation range to search over
- Speed
- Context





(b) P(person) = uniform

(d) P(person | geometry)

(f) P(person | viewpoint)

(g) P(person|viewpoint,geometry)

Hoiem, Efros, Herbert, 2006

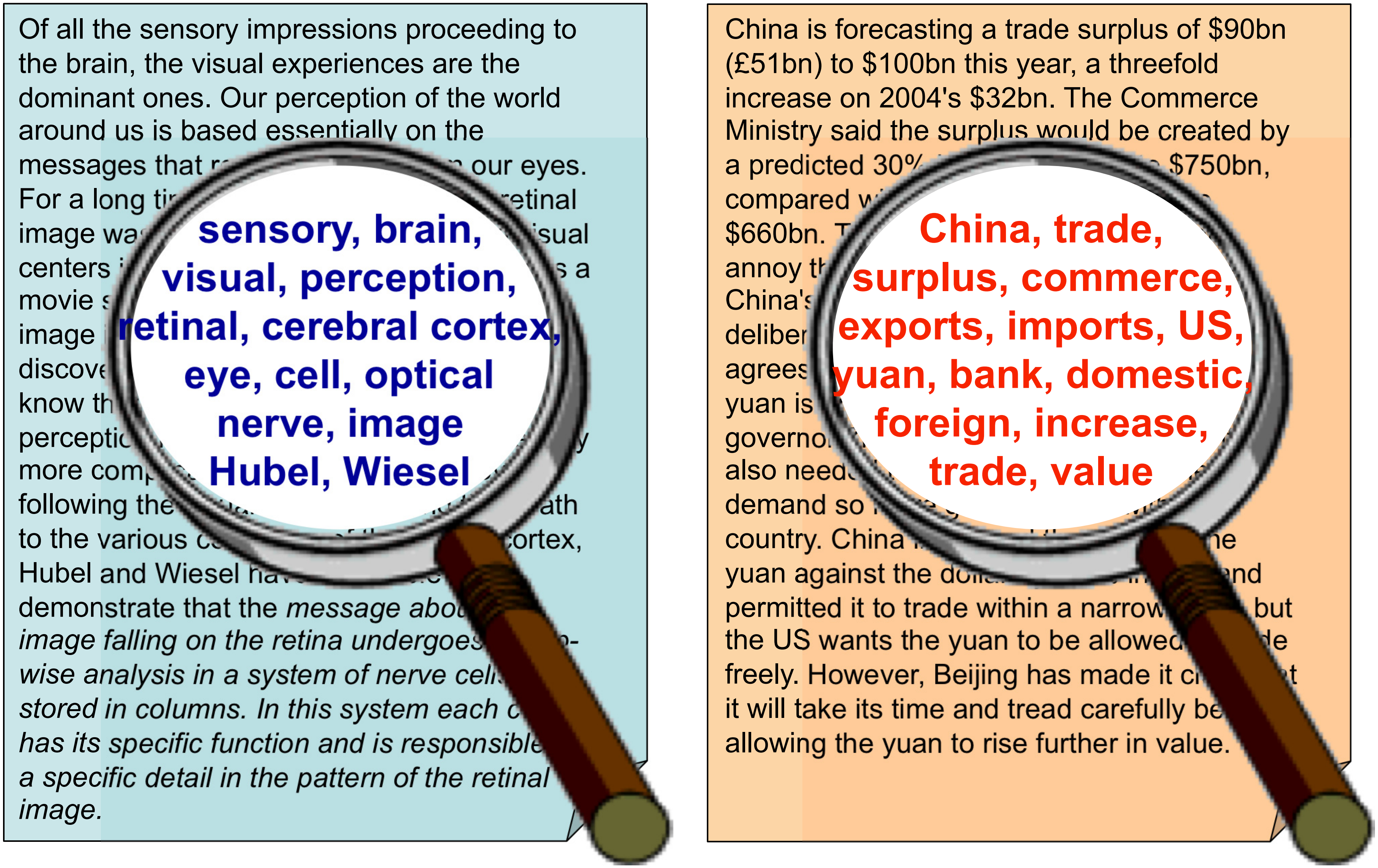
- OBJECTS
  - ANIMALS
    - .....
    - VERTEBRATE
      - MAMMALS
        - TAPIR
        - BOAR
      - BIRDS
        - GROUSE
  - PLANTS
  - INANIMATE
    - NATURAL
    - MAN-MADE
      - CAMERA

Part 1: Bag-of-words models

by Li Fei-Fei (Princeton)

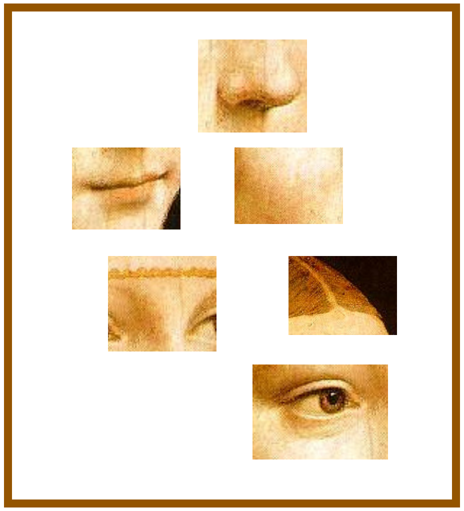
Object → **Bag of 'words'**

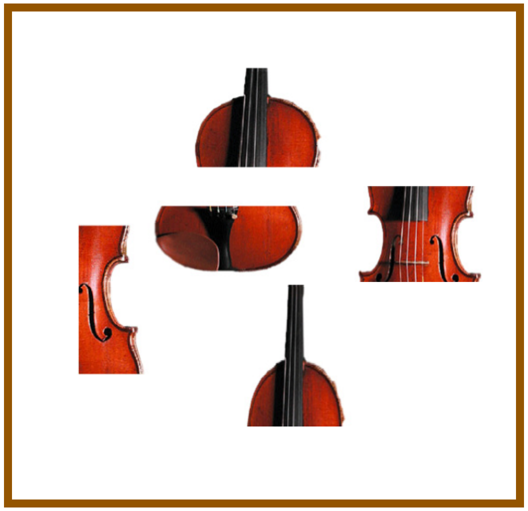
Analogy to documents



A clarification: definition of “BoW”

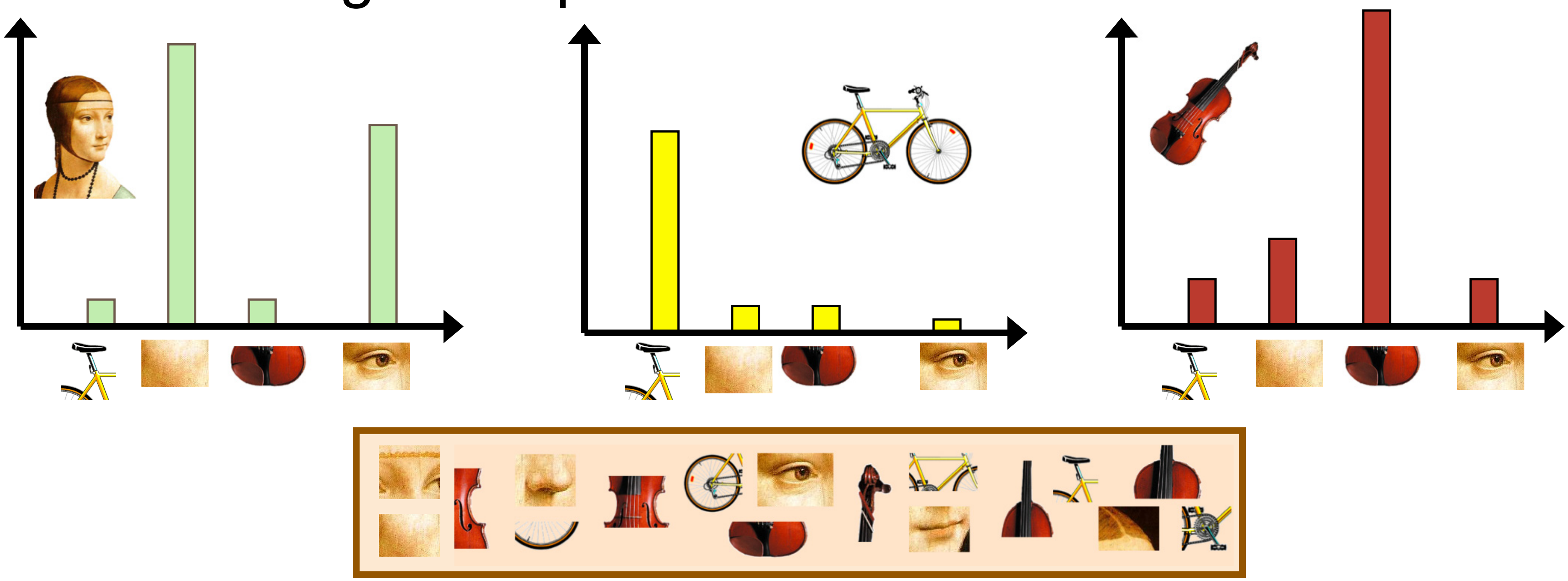
- Looser definition
  - Independent features

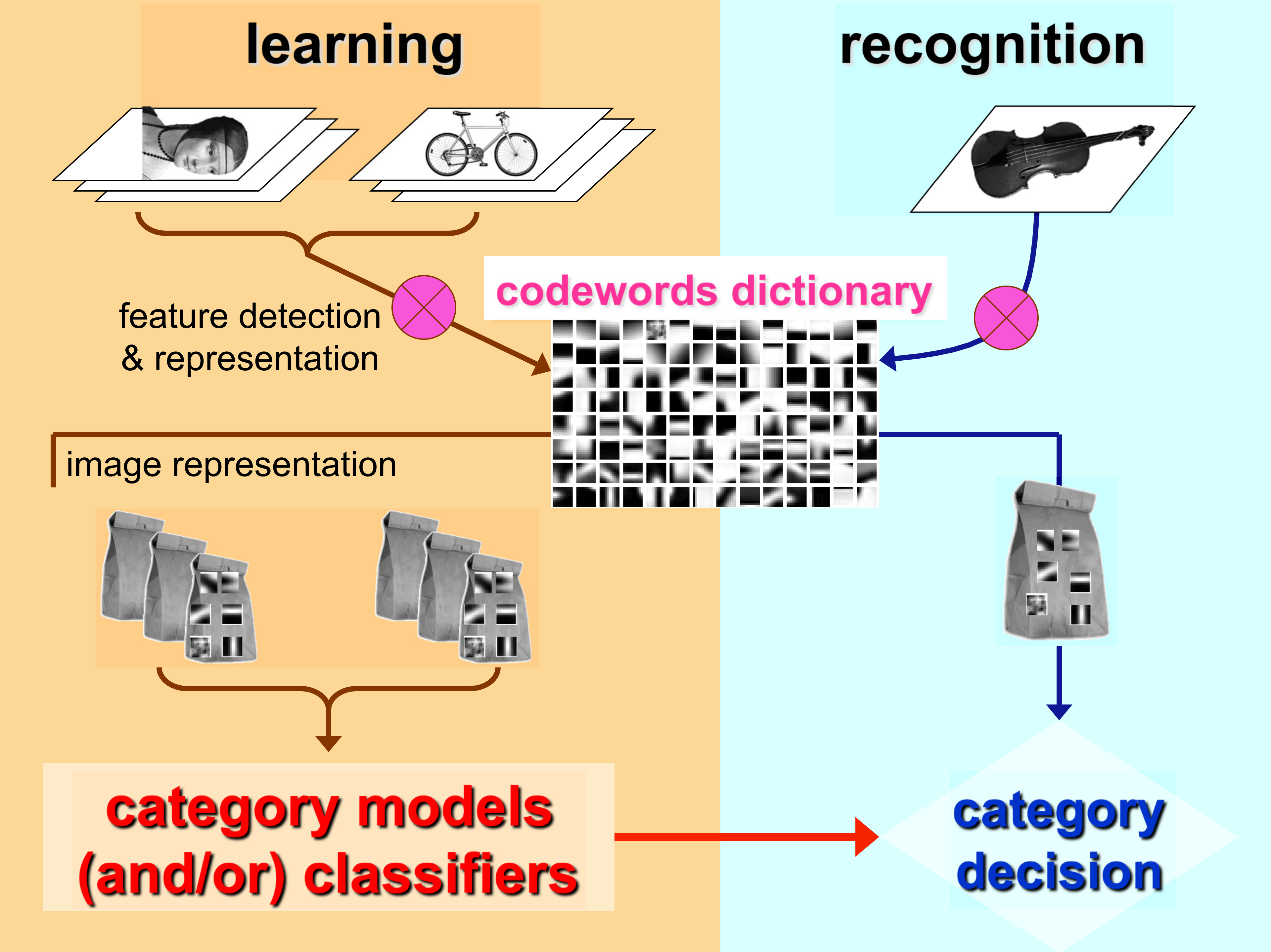




A clarification: definition of "BoW"

- Looser definition
  - Independent features
- Stricter definition
  - Independent features
  - histogram representation




learning
recognition
feature detection
& representation
codewords dictionary
image representation
category models
(and/or) classifiers
category
decision