

# Introduction to Computer Vision

Michael J. Black

Object Recognition

# News

- No class on Wednesday. Have a good Thanksgiving break.
- $\frac{1}{2}$  of project proposals have received feedback. The other  $\frac{1}{2}$  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

CVPR 2007 Minneapolis, Short Course, June 17

# Recognizing and Learning Object Categories: Year 2007

Li Fei-Fei, Princeton

Rob Fergus, MIT

Antonio Torralba, MIT



# One-Shot Learning



*S. Savarese, 2003*



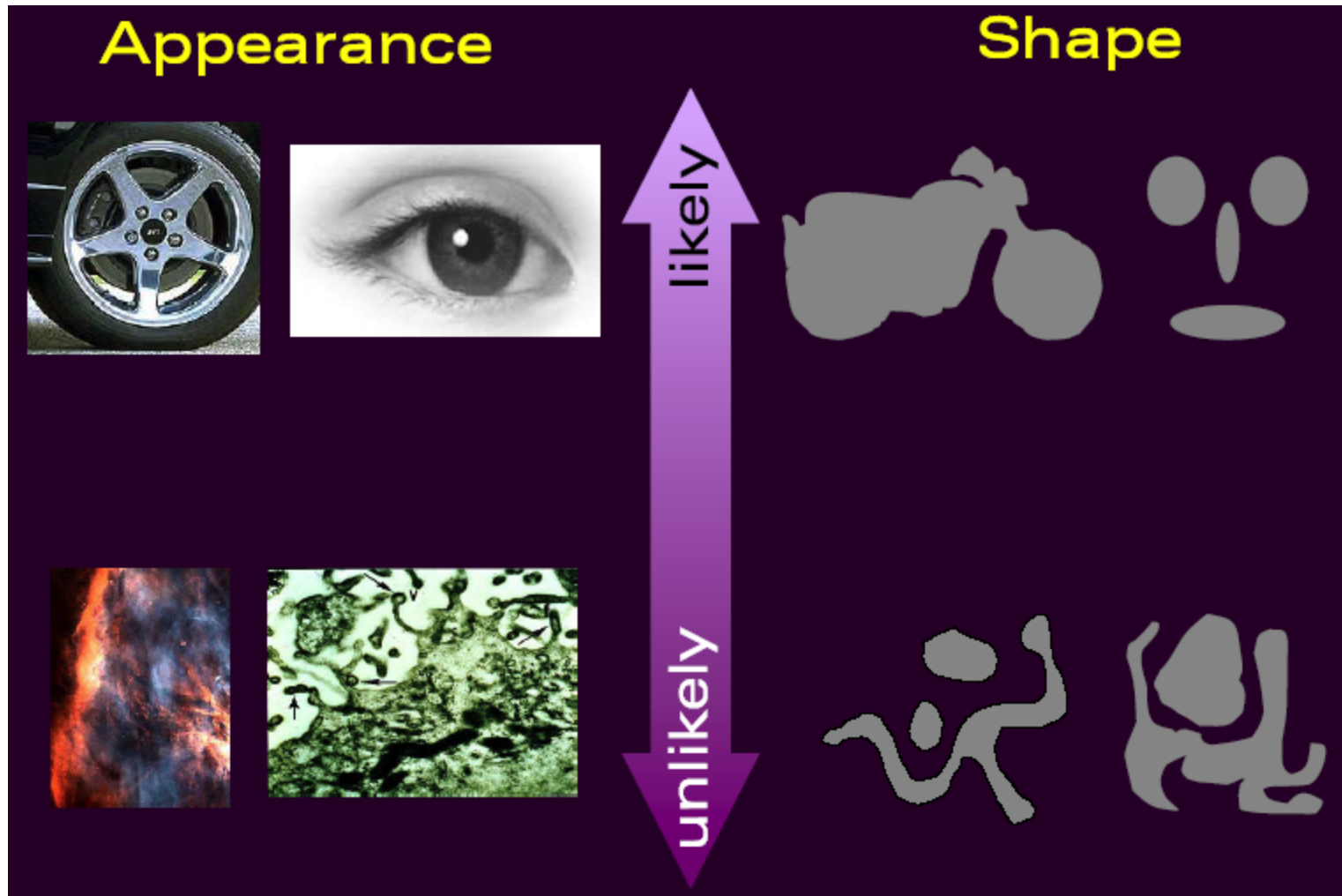
Fei-Fei Li.  
CS143 Intro to  
Computer Vision

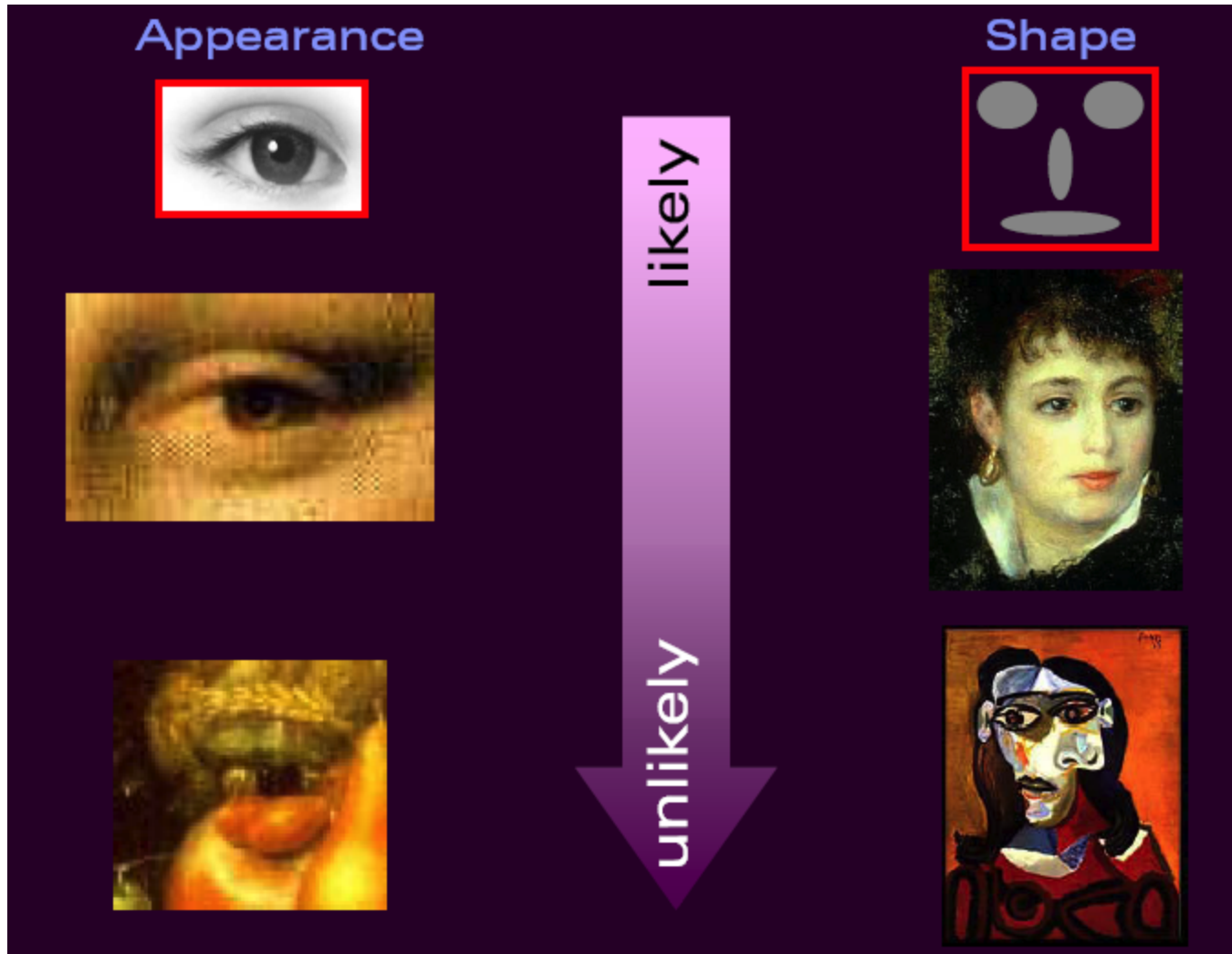
Brown University



~10,000 to 30,000







# Object categorization: the statistical viewpoint



$$p(\textit{zebra} | \textit{image})$$

vs.

$$p(\textit{no zebra} | \textit{image})$$

- Bayes rule:

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

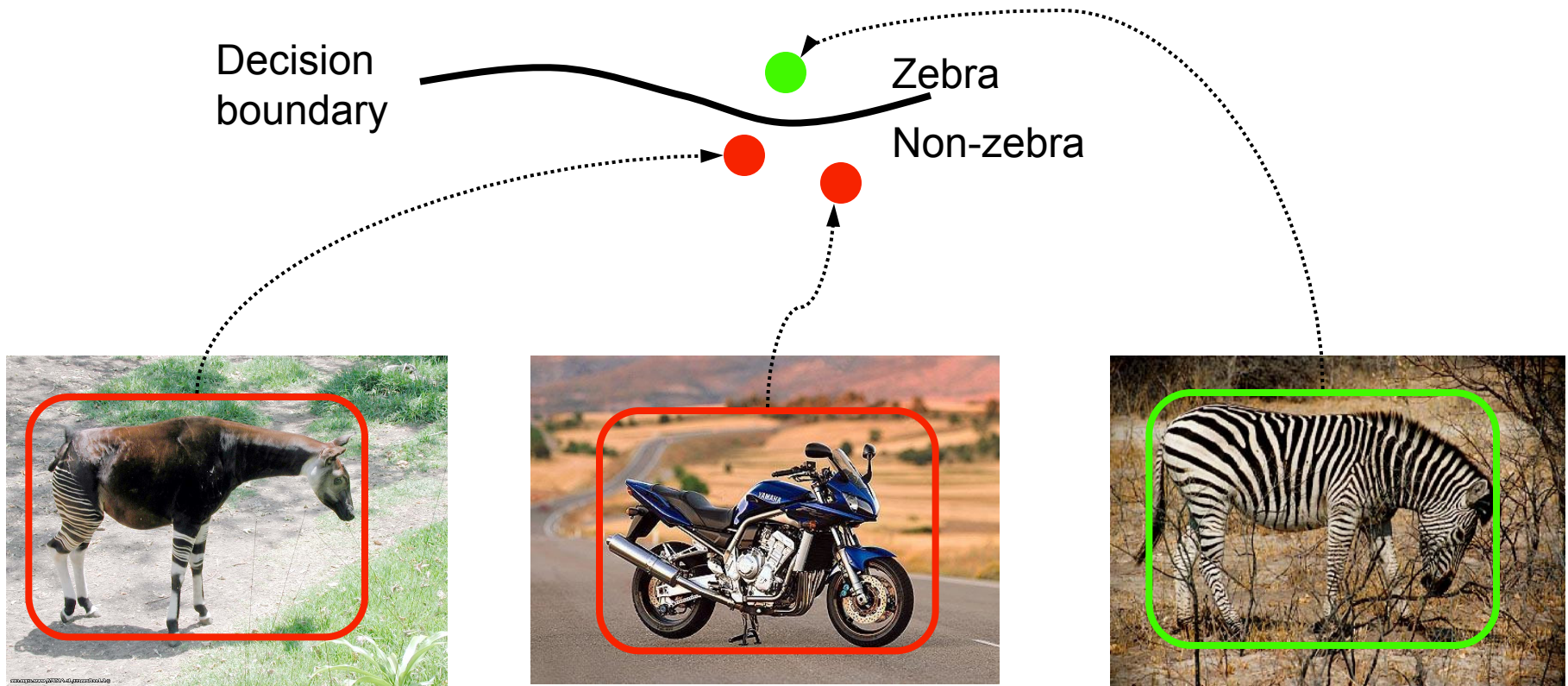
# Object categorization: the statistical viewpoint

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

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

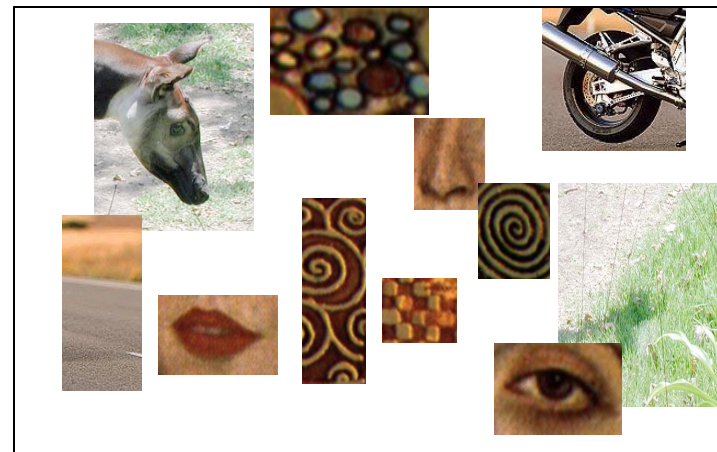
# Discriminative



- Direct modeling of  $\frac{p(\text{zebra} | \text{image})}{p(\text{no zebra} | \text{image})}$



# Generative

- Model  $p(\text{image} | \text{zebra})$  and  $p(\text{image} | \text{no zebra})$



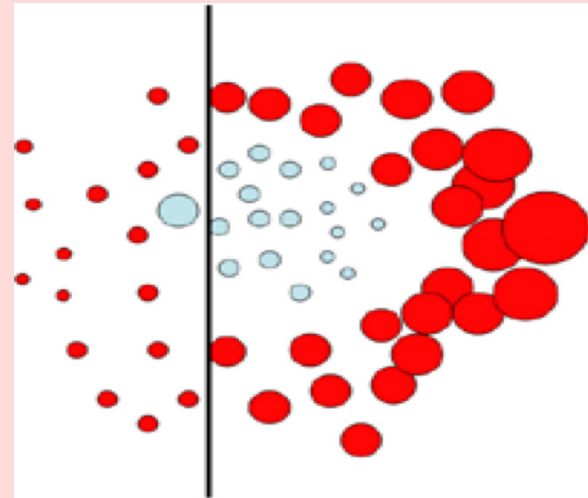
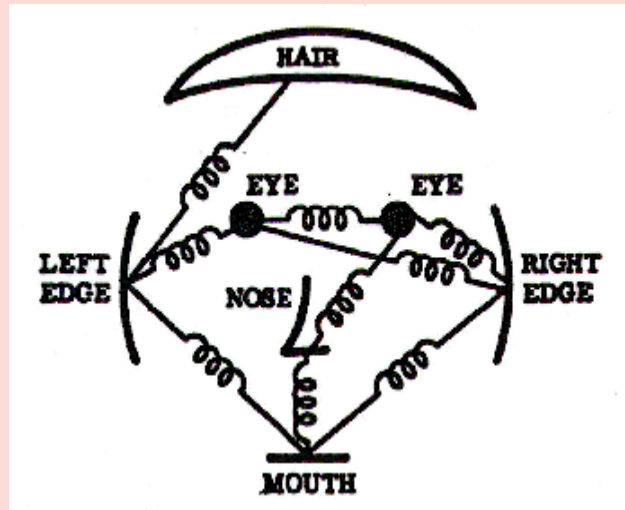
	$p(\text{image}   \text{zebra})$	$p(\text{image}   \text{no zebra})$
	Low	Middle
	High	Middle $\rightarrow$ Low

# 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

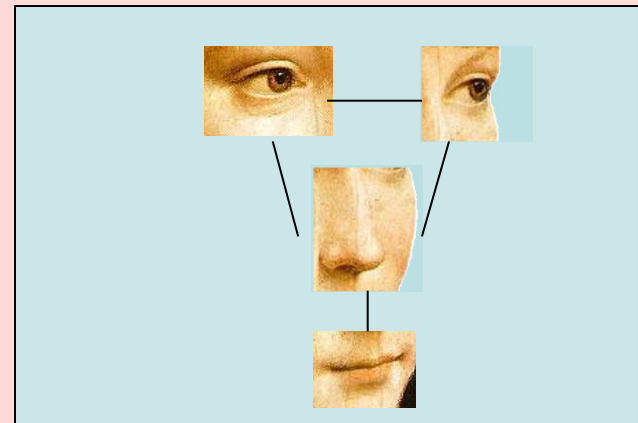
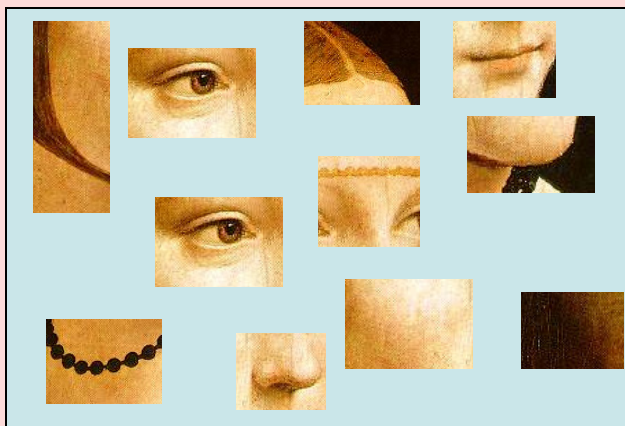
- Generative /  
discriminative / hybrid





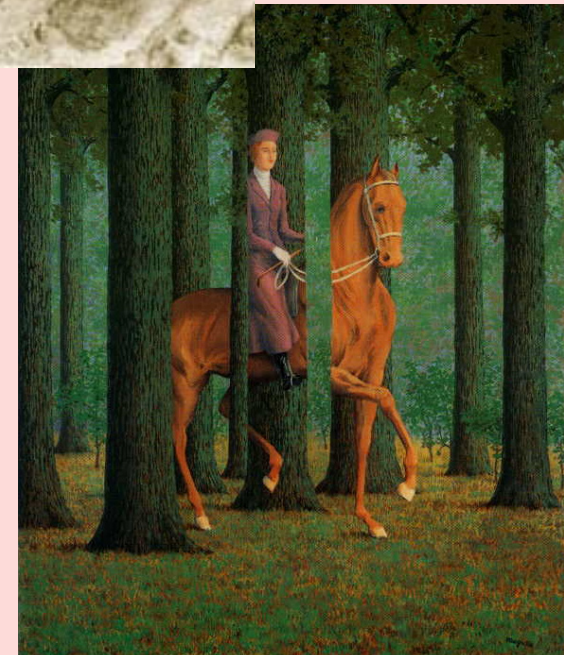
# Representation

- Generative / discriminative / hybrid
- Appearance only or location and appearance



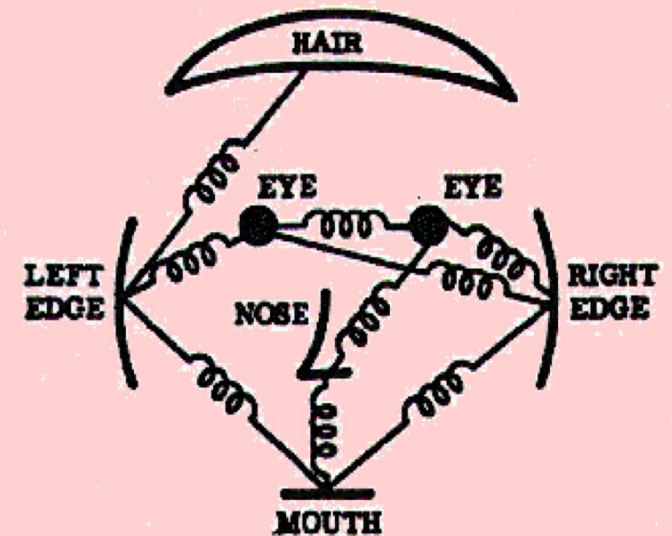
# Representation

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



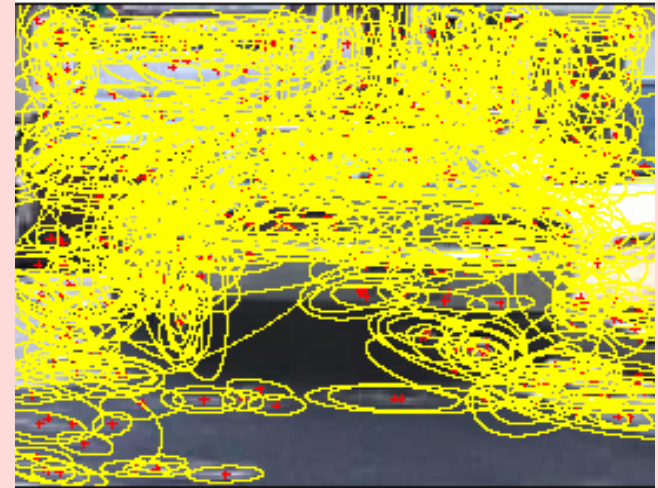
# Representation

- 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



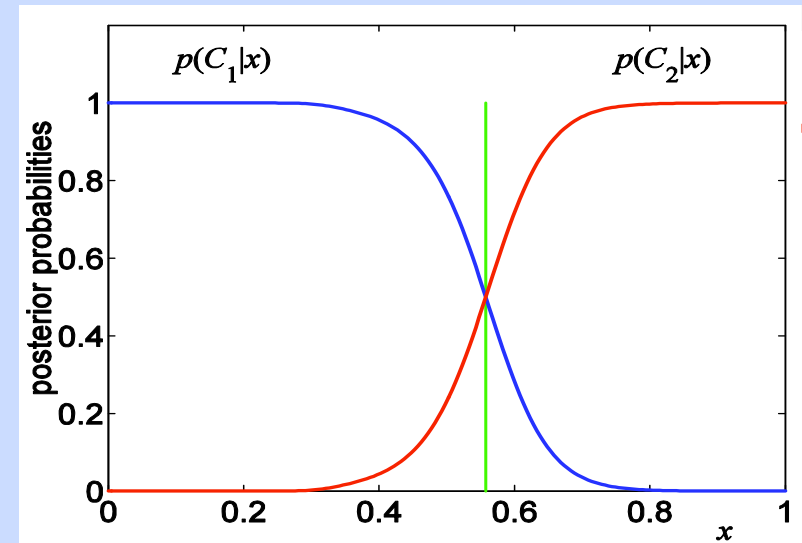
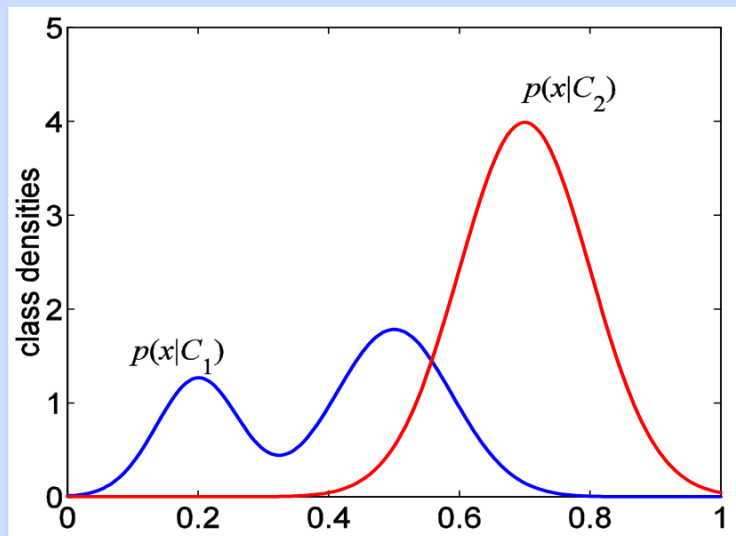
# Learning

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



# Learning

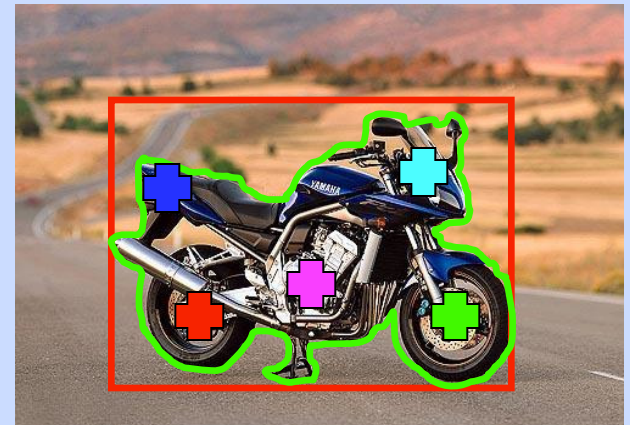
- Unclear how to model categories, so we learn what distinguishes them rather than manually specify the difference -- hence current interest in machine learning)
- Methods of training: generative vs. discriminative



# 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

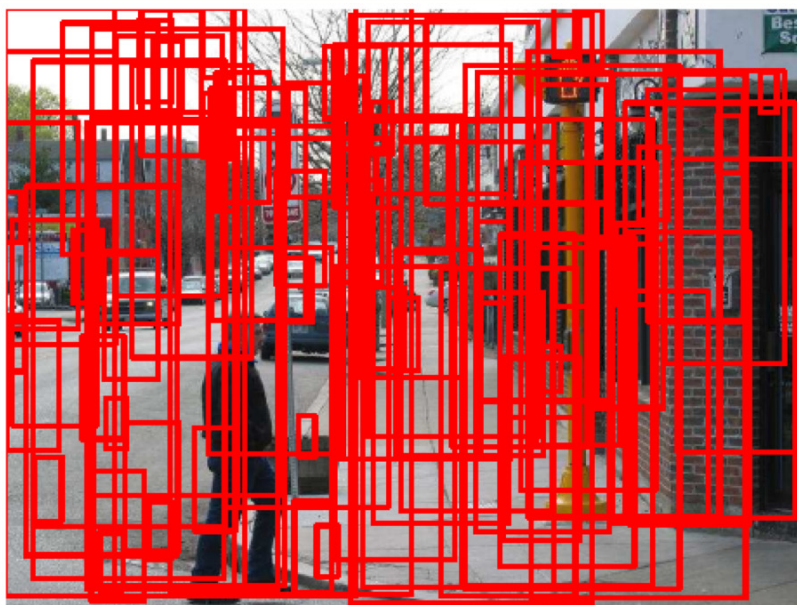
# Learning

- 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**

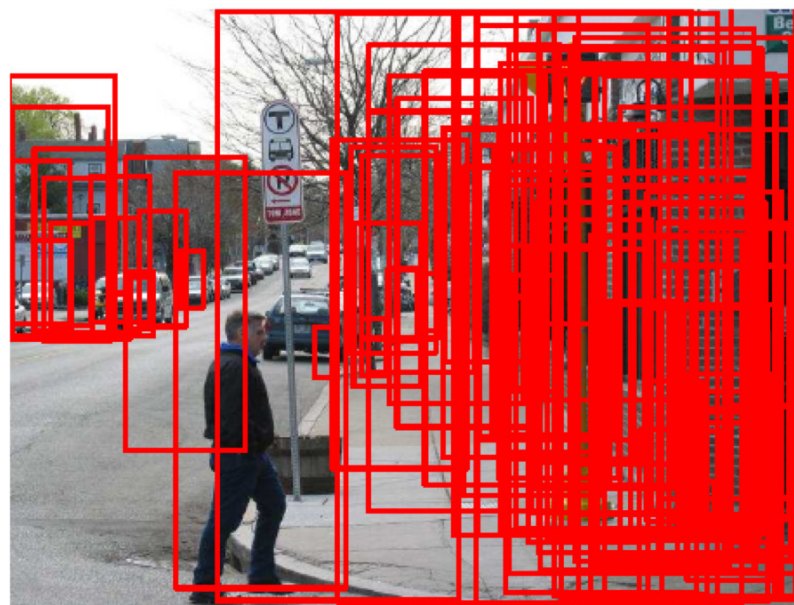
# Recognition

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

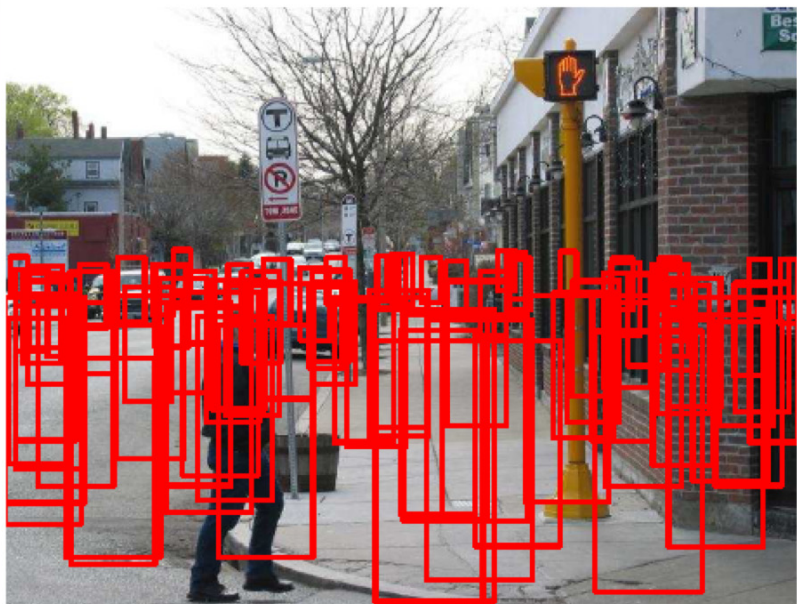




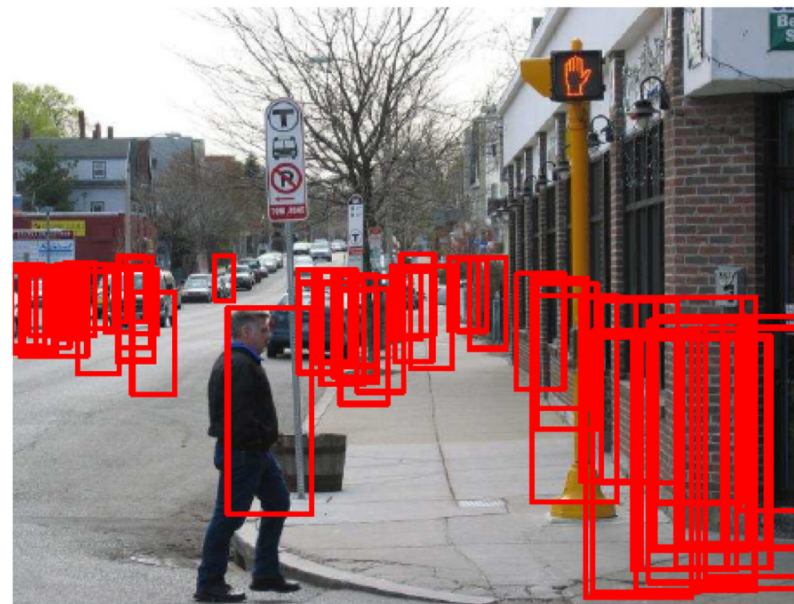
(b)  $P(\text{person}) = \text{uniform}$



(d)  $P(\text{person} \mid \text{geometry})$

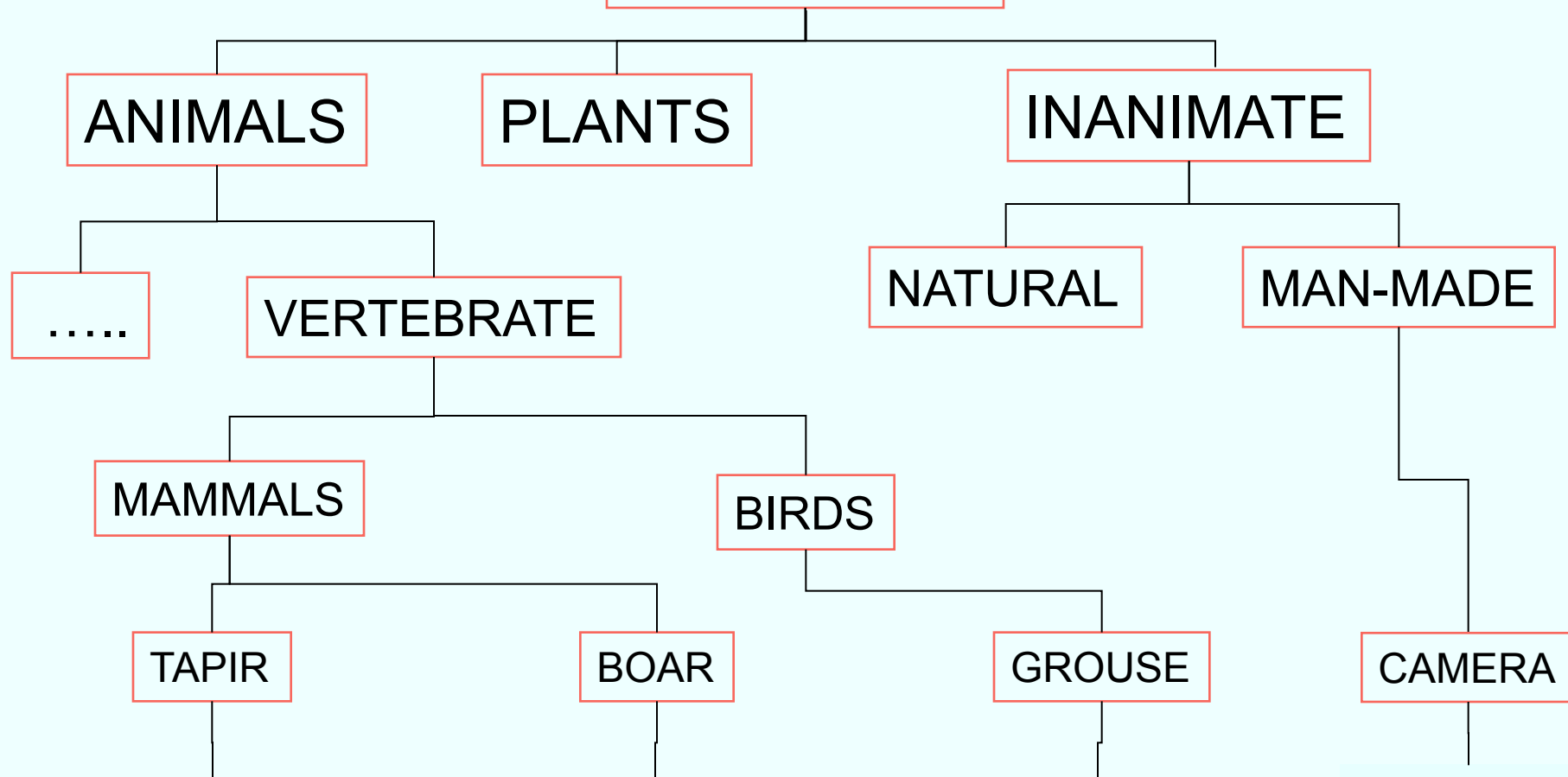


(f)  $P(\text{person} \mid \text{viewpoint})$



(g)  $P(\text{person} \mid \text{viewpoint, geometry})$

# OBJECTS





# Part 1: Bag-of-words models

by Li Fei-Fei (Princeton)

**Object**



**Bag of 'words'**



# Analogy to documents

Of all the sensory impressions proceeding to the brain, the visual experiences are the dominant ones. Our perception of the world around us is based essentially on the messages that reach our eyes. For a long time, the retinal image was considered as a movie screen. It is now known that the image is processed in a more complex way following the path to the various centers of the cortex. Hubel and Wiesel have demonstrated that the *message about the image falling on the retina undergoes a point-by-point analysis in a system of nerve cells stored in columns. In this system each cell has its specific function and is responsible for a specific detail in the pattern of the retinal image.*

**sensory, brain,  
visual, perception,  
retinal, cerebral cortex,  
eye, cell, optical  
nerve, image  
Hubel, Wiesel**

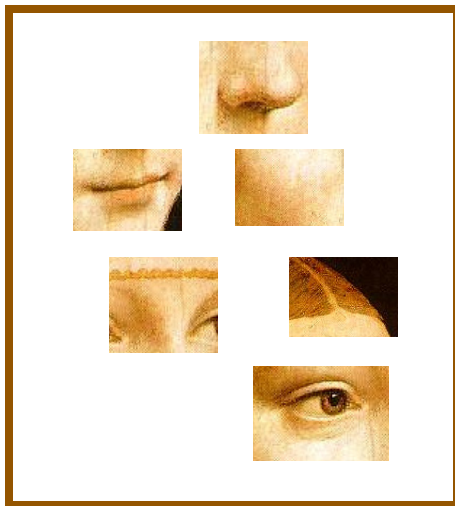
China is forecasting a trade surplus of \$90bn (£51bn) to \$100bn this year, a threefold increase on 2004's \$32bn. The Commerce Ministry said the surplus would be created by a predicted 30% increase in exports to \$750bn, compared with \$580bn in 2004. The surplus will annoy the US because it will reduce the trade deficit. China's government has deliberately kept the yuan's value low to encourage exports. The government agrees that the yuan is undervalued. The government also needs to increase the demand for its goods in the rest of the world. China has been increasing the value of the yuan against the dollar. The government has permitted it to trade within a narrow range but the US wants the yuan to be allowed to trade freely. However, Beijing has made it clear that it will take its time and tread carefully before allowing the yuan to rise further in value.

**China, trade,  
surplus, commerce,  
exports, imports, US,  
yuan, bank, domestic,  
foreign, increase,  
trade, value**



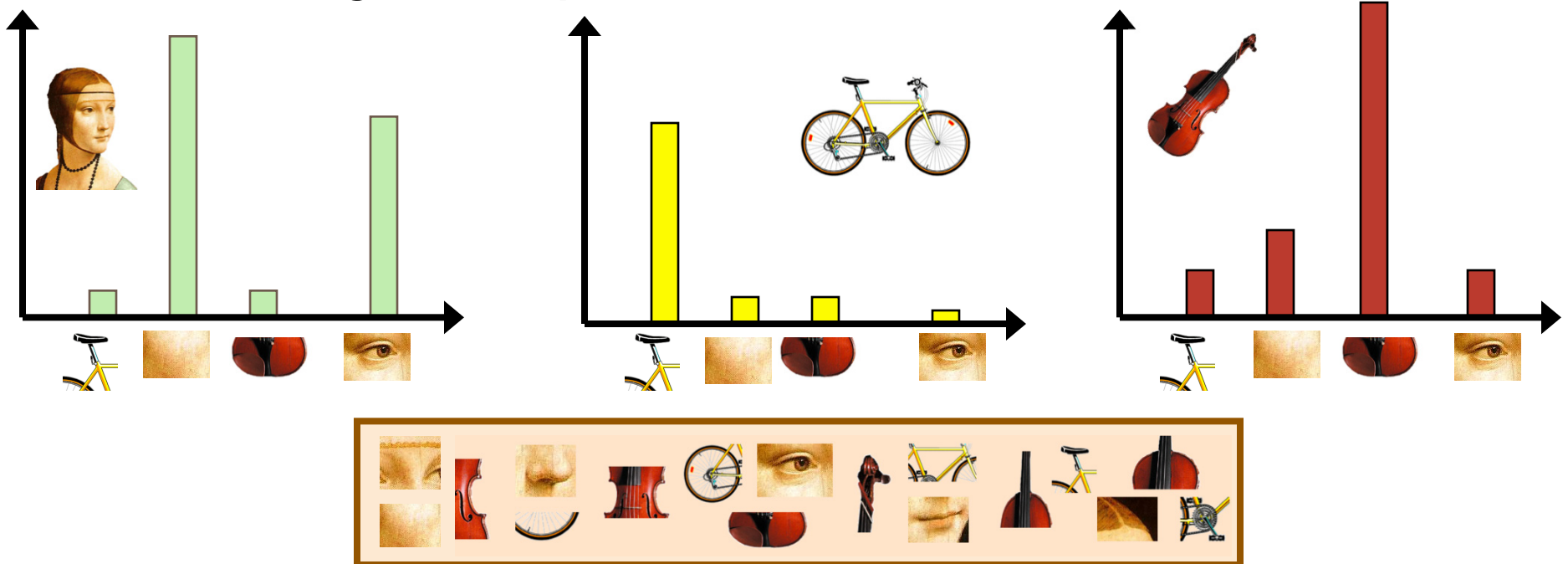
# A clarification: definition of “BoW”

- Looser definition
  - Independent features

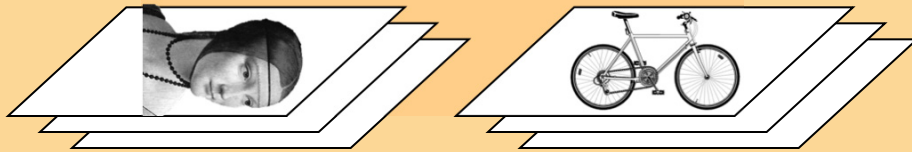


# A clarification: definition of “BoW”

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



# learning



feature detection & representation

codewords dictionary

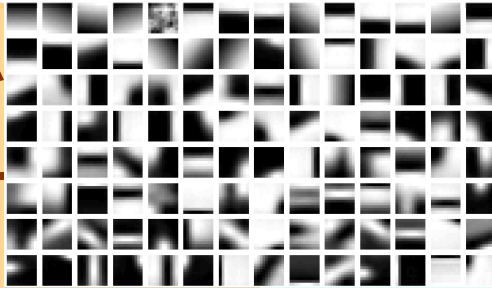
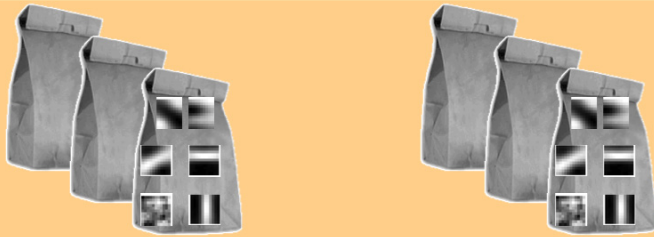


image representation



**category models  
(and/or) classifiers**

# recognition



**category  
decision**

