

CVPR 2007 Minneapolis, Short Course, June 17

Recognizing and Learning Object Categories: Year 2007

Li Fei-Fei, Princeton

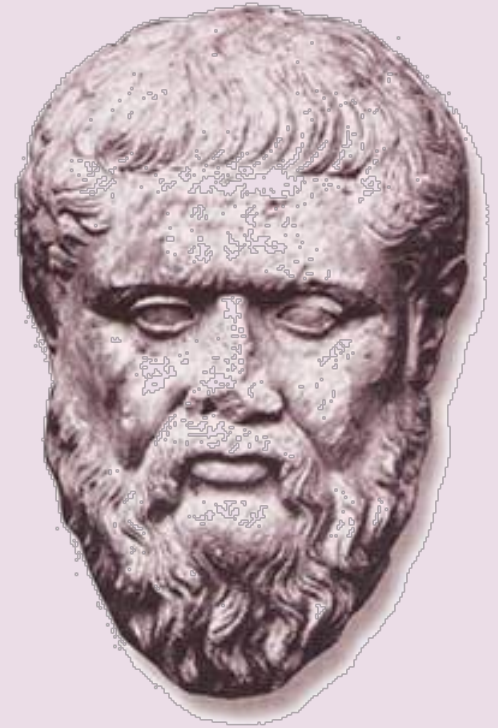
Rob Fergus, MIT

Antonio Torralba, MIT

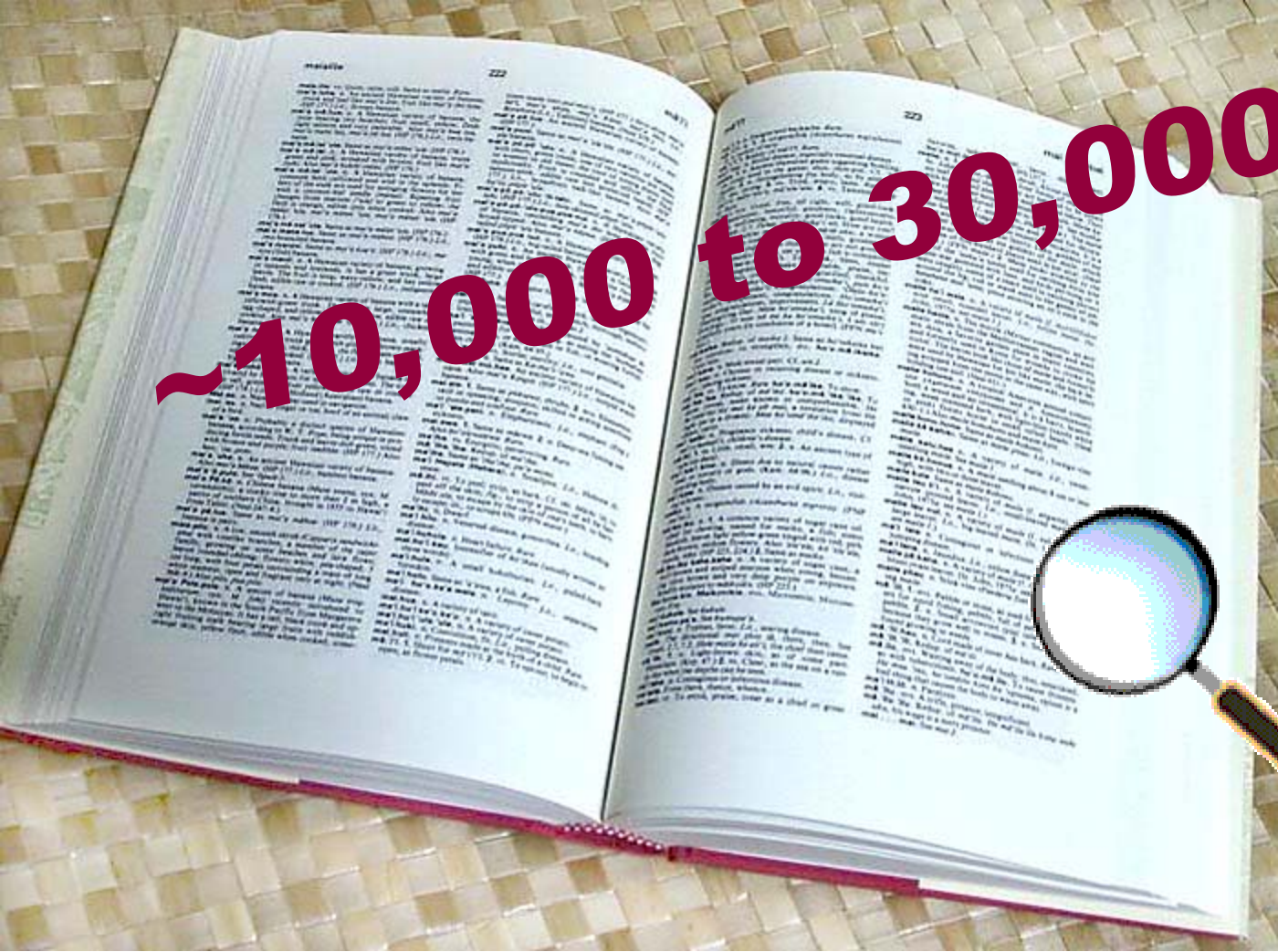


Plato said...

- Ordinary objects are classified together if they 'participate' in the same abstract Form, such as the Form of a Human or the Form of Quartz.
- Forms are proper subjects of philosophical investigation, for they have the highest degree of reality.
- Ordinary objects, such as humans, trees, and stones, have a lower degree of reality than the Forms.
- Fictions, shadows, and the like have a still lower degree of reality than ordinary objects and so are not proper subjects of philosophical enquiry.



How many object categories are there?



So what does object recognition involve?



Verification: is that a lamp?



Detection: are there people?



Identification: is that Potala Palace?



Object categorization



mountain

tree

building

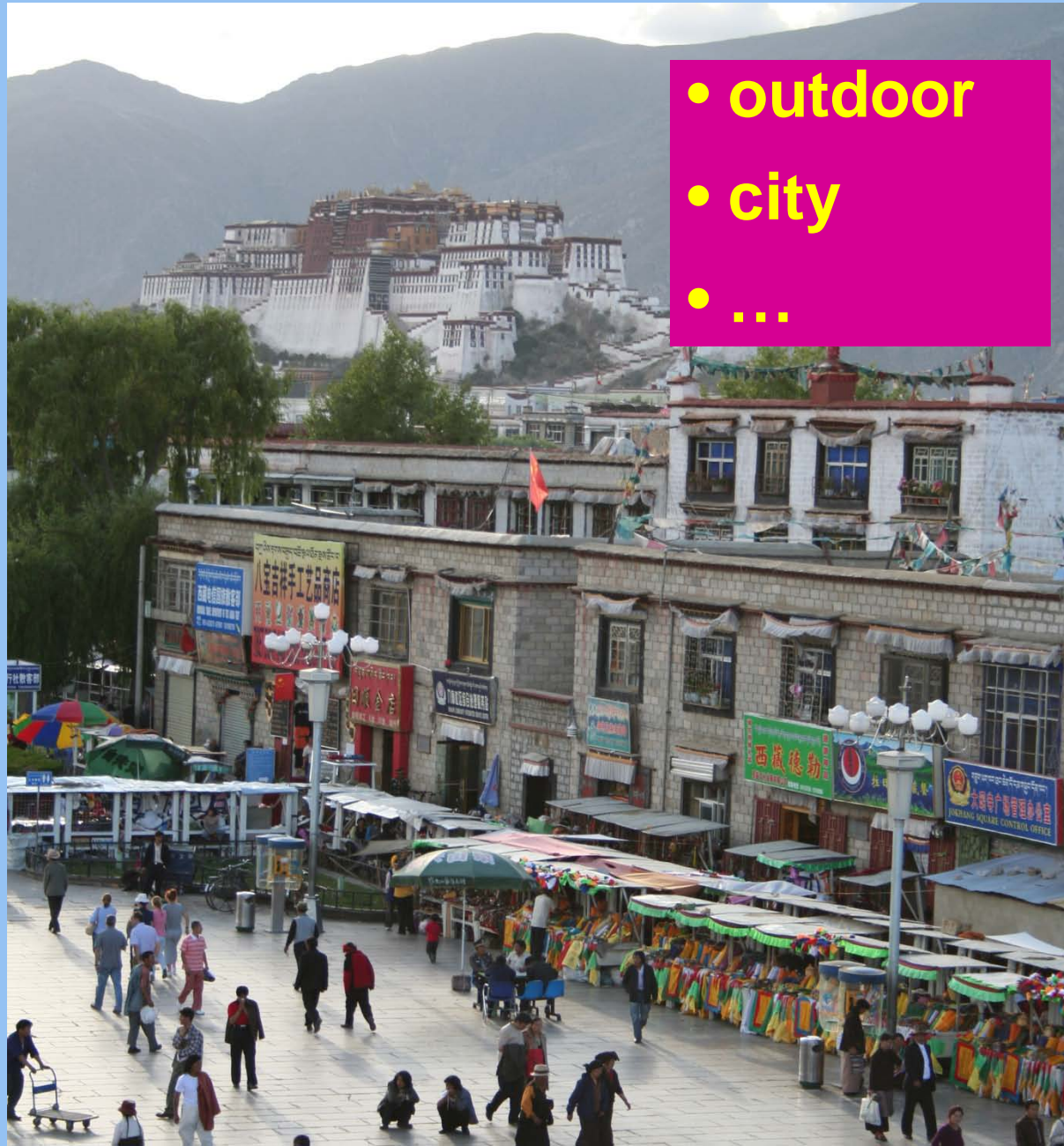
banner

street lamp

vendor

people

Scene and context categorization



- outdoor
- city
- ...

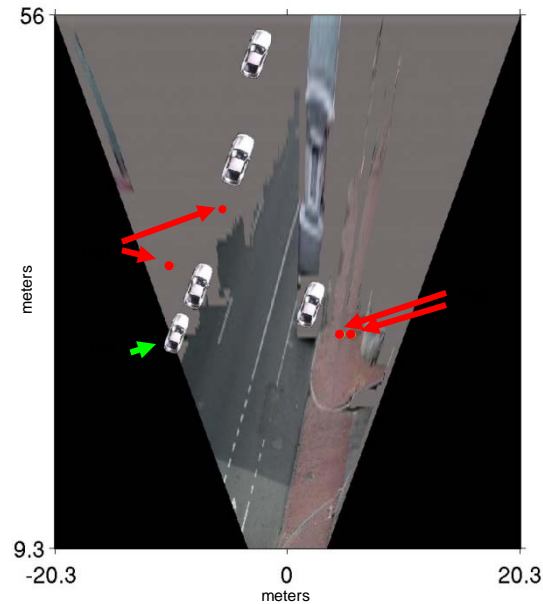
Computational photography



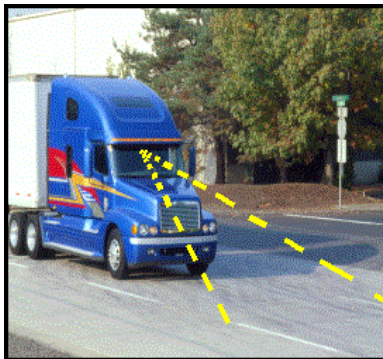
[Face priority AE] When a bright part of the face is too bright

Assisted driving

Pedestrian and car detection



Lane detection



- Collision warning systems with adaptive cruise control,
- Lane departure warning systems,
- Rear object detection systems,

Improving online search

flickr GAMMA

webshots beta

Ask TM Images

Cydral TM
Image & Site Search

picsearch TM











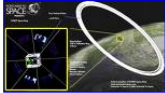

Google TM
Image Search

altavista TM

Query:
STREET

Google web images video news maps more >
street
Search Images Search the Web [Advanced Image Search Preferences](#)
Moderate SafeSearch is on

Images Showing: All image sizes Results 19 - 36 of about 44,200,000 for street [definition] (0.04 seconds)

 <p>Street sweeper 345 x 352 - 17k - jpg www.town.telluride.co.us</p>	 <p>Street Maintenance 407 x 402 - 18k - jpg www.town.telluride.co.us</p>	 <p>Main Street Station 360 x 392 - 30k - jpg www.rmaonline.org</p>	 <p>SHPO Wayne Donaldson at Main Street ... 410 x 314 - 41k - jpg ohp.parks.ca.gov</p>	 <p>Lombard Street, worlds crookedest See ... 500 x 387 - 59k - jpg www.inetours.com</p>	 <p>Street Bike (BS70-4A) Details 360 x 360 - 38k - jpg bashan.en.alibaba.com</p>
 <p>Street Lamps 360 x 360 - 18k - jpg syi.en.alibaba.com [More from img.alibaba.com]</p>	 <p>Washington D.C. Laminated Street Map 500 x 500 - 114k - jpg www.dcgiftshop.com</p>	 <p>street-riders-ss-3.jpg 550 x 309 - 53k - jpg www.pspworld.com</p>	 <p>Visually Street Riders is not nearly ... 550 x 309 - 52k - jpg www.pspworld.com</p>	 <p>STREET space ring Postcards To Space ... 1000 x 563 - 87k - jpg www.postcardstospace.com</p>	 <p>17 Fleet Street 492 x 681 - 74k - jpg www.pepysdiary.com</p>

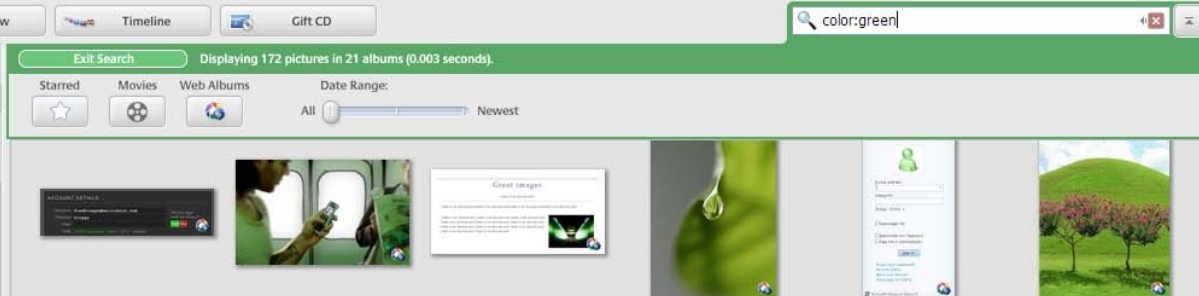
Organizing photo collections

Timeline Gift CD

color:green

Exit Search Displaying 172 pictures in 21 albums (0.003 seconds).

Starred Movies Web Albums Date Range: All [slider] Newest



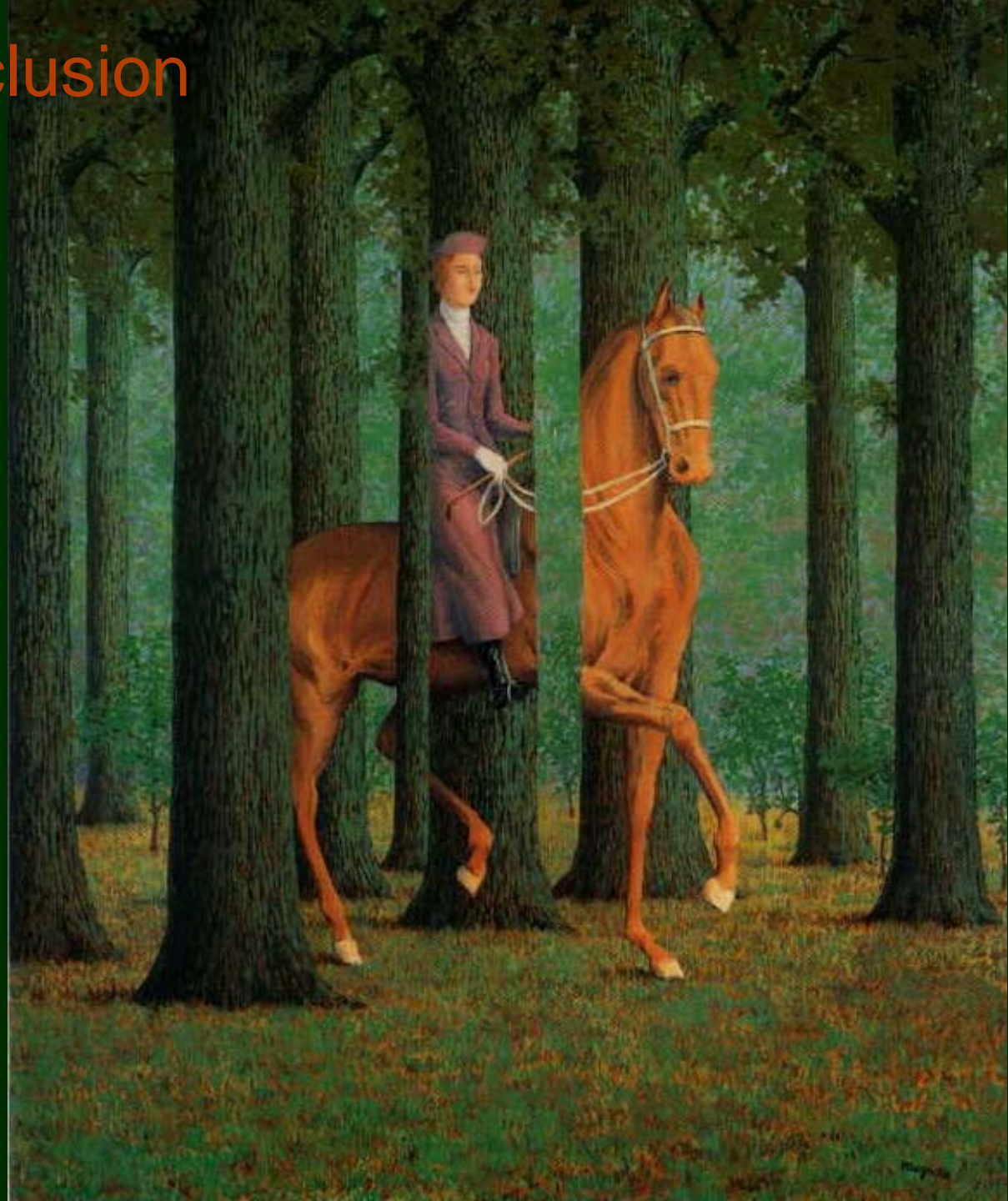
Challenges 1: view point variation



Challenges 2: illumination



Challenges 3: occlusion

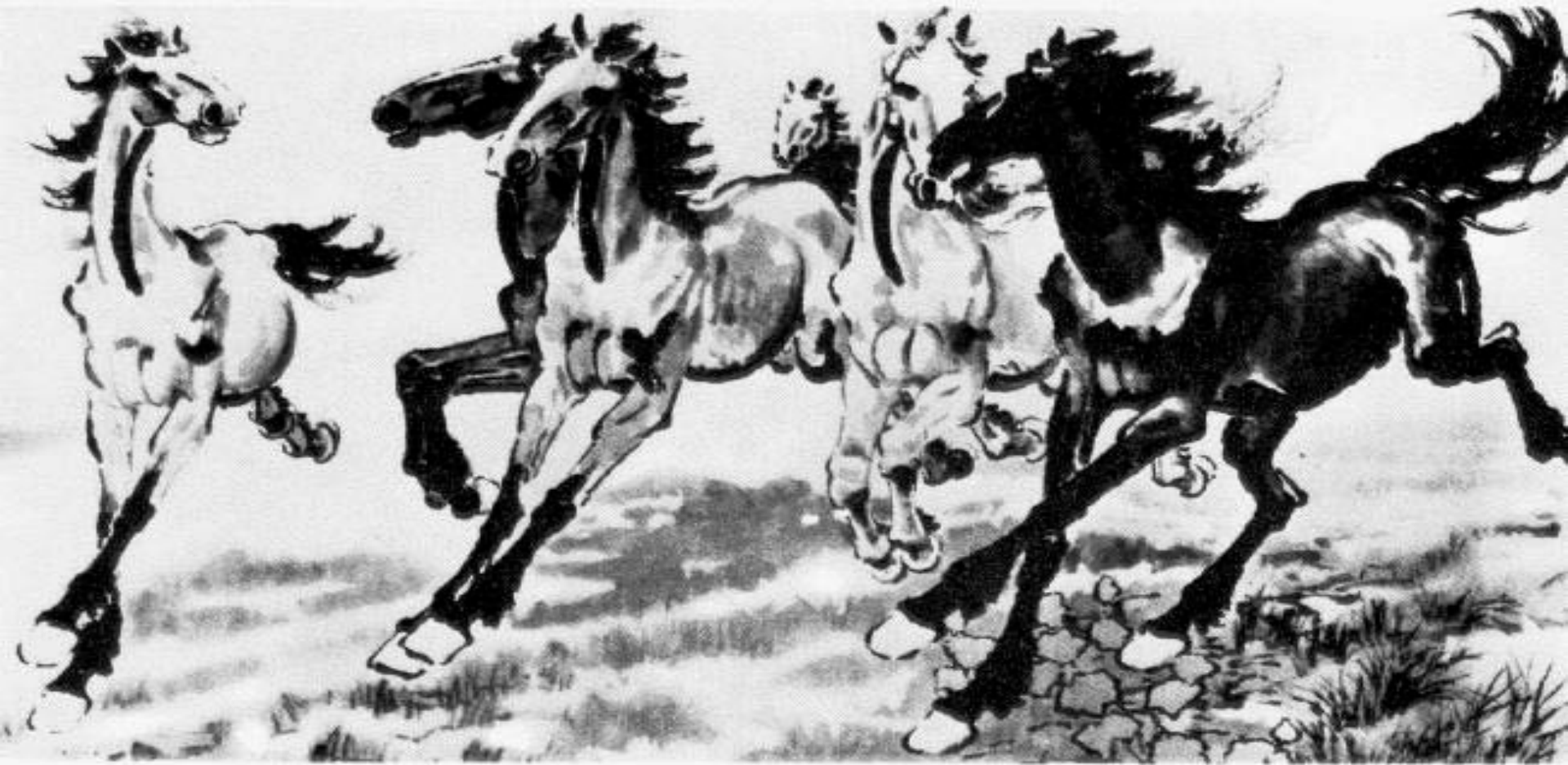


Magritte, 1957

Challenges 4: scale



Challenges 5: deformation



Challenges 6: background clutter

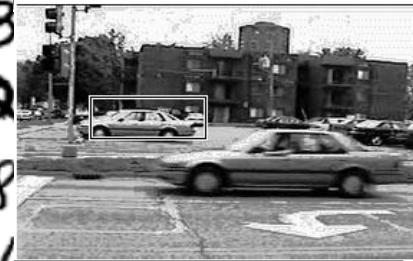
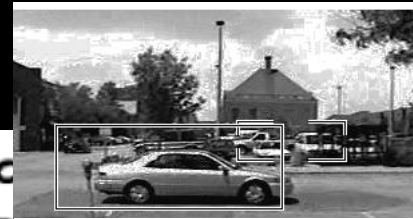


Klimt, 1913

Challenges 7: intra-class variation



History: early object categorization



1 7 9 6
7 8 6 3
2 1 7 9 7 1 2
4 8 1 9 0 1 8
7 6 1 8 6 4 1
7 5 9 2 6 5 8 1 9 7
2 2 2 2 2 3 4 4 8 0
0 2 3 8 0 7 3 8 5 7
0 1 4 6 4 6 0 2 4 3
7 1 2 8 7 6 9 8 6 1



- Turk and Pentland, 1991
- Belhumeur, Hespanha, & Kriegman, 1997
- Schneiderman & Kanade 2004
- Viola and Jones, 2000



- Amit and Geman, 1999
- LeCun et al. 1998
- Belongie and Malik, 2002



- Schneiderman & Kanade, 2004
- Argawal and Roth, 2002
- Poggio et al. 1993



~10,000 to 30,000



Object categorization: the statistical viewpoint



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

vs.

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

- Bayes rule:

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

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 $\frac{p(\text{zebra} | \text{image})}{p(\text{no zebra} | \text{image})}$

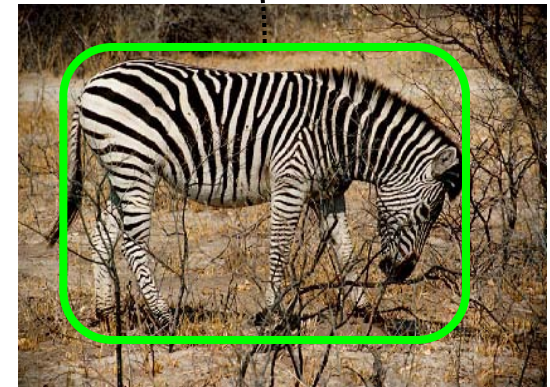
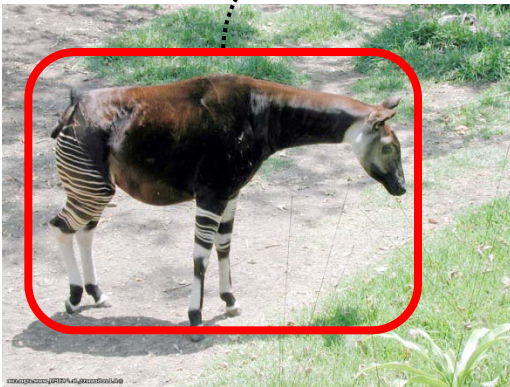
Decision
boundary



Zebra

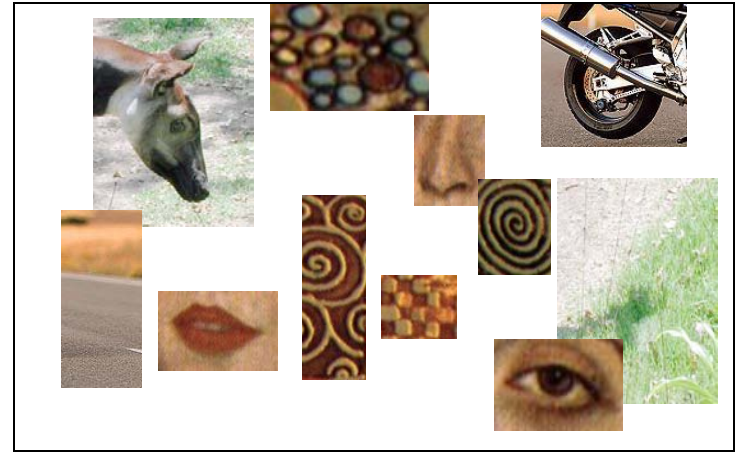




Non-zebra



Generative

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



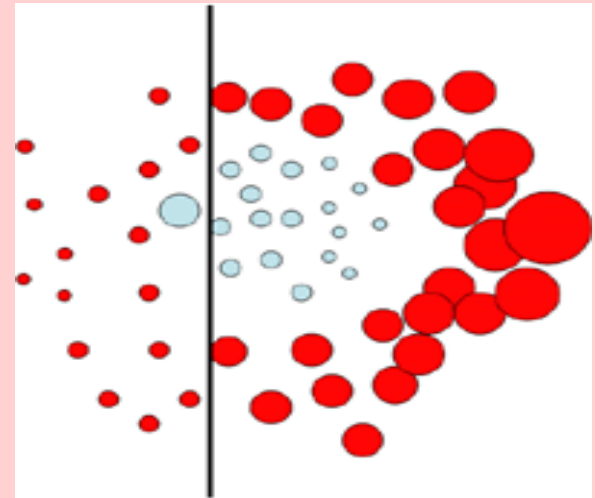
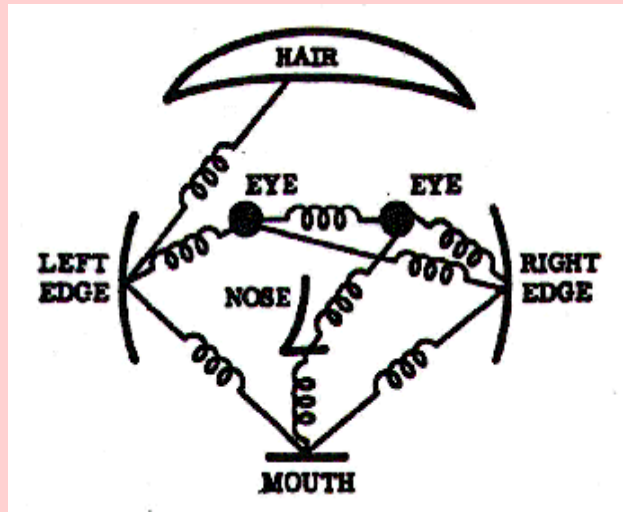
	$p(\text{image} \mid \text{zebra})$	$p(\text{image} \mid \text{no zebra})$
	Low	Middle
	High	Middle \rightarrow Low

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

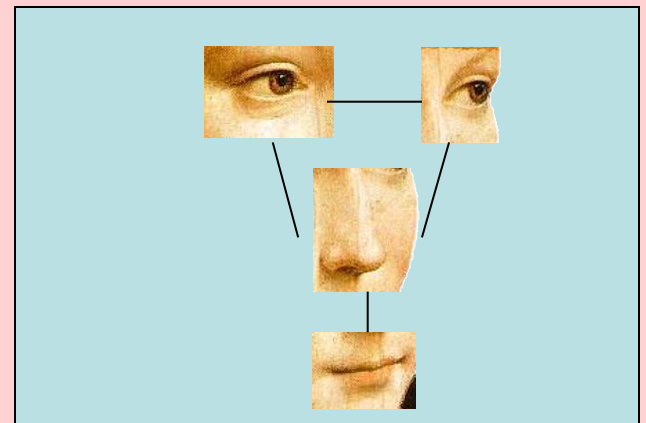
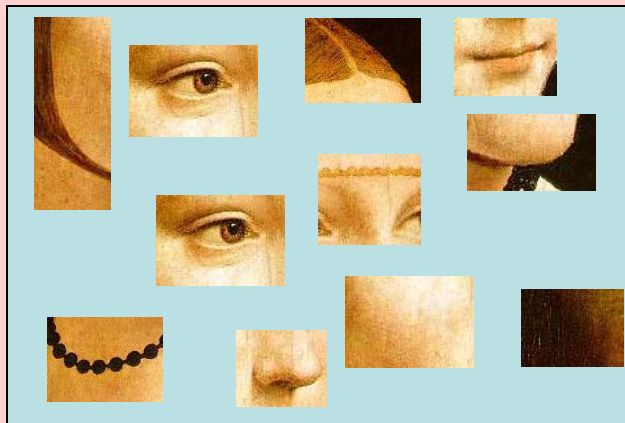
Representation

- Generative /
discriminative / hybrid



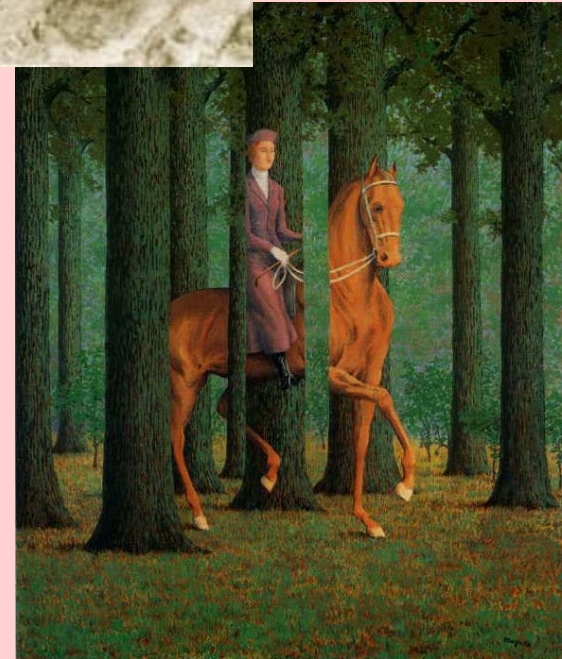
Representation

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



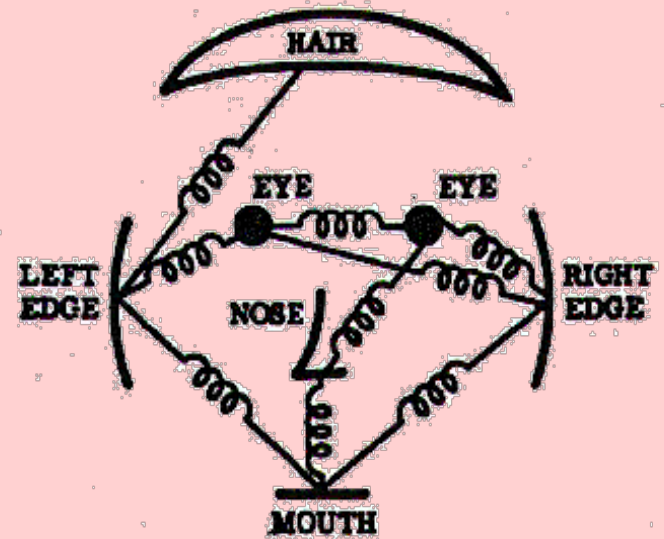
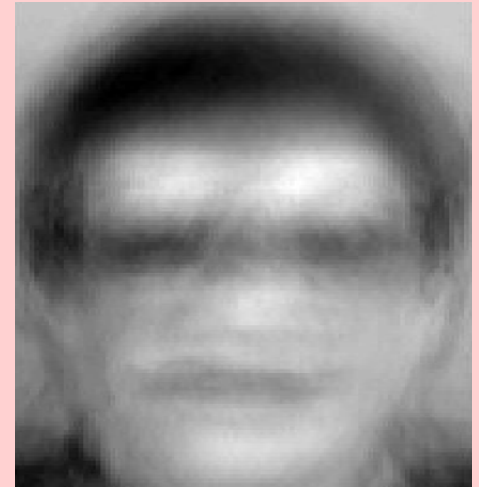
Representation

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



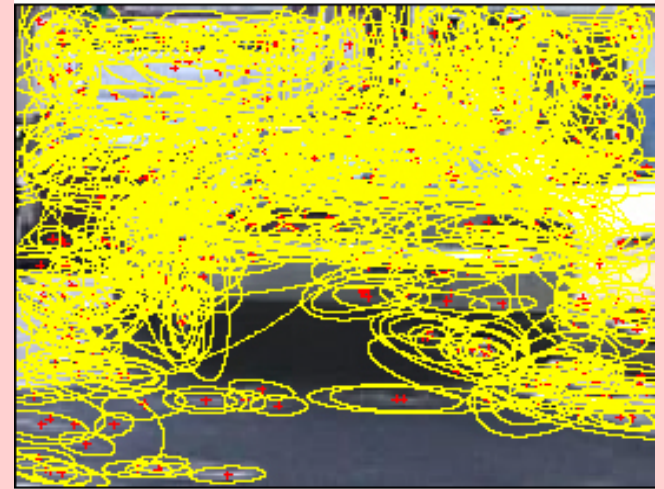
Representation

- 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



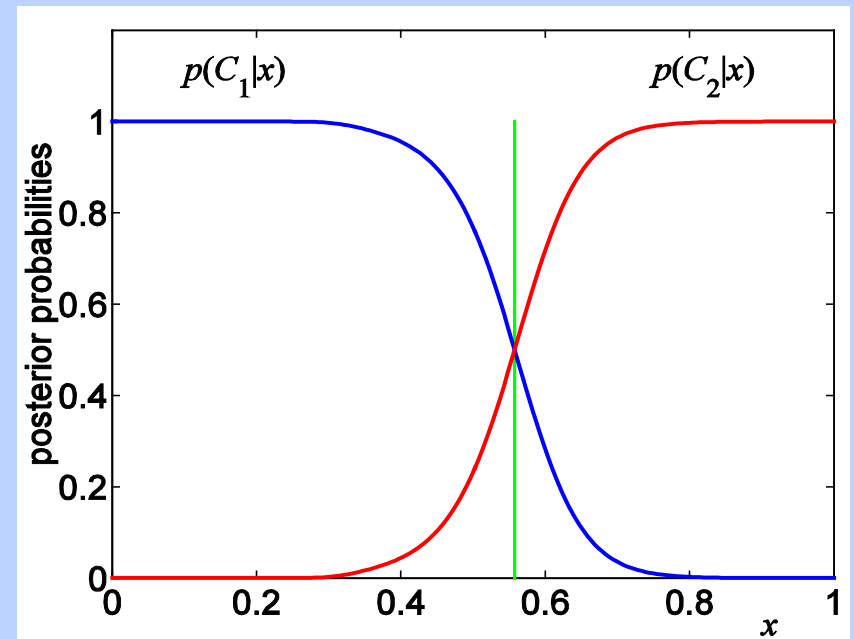
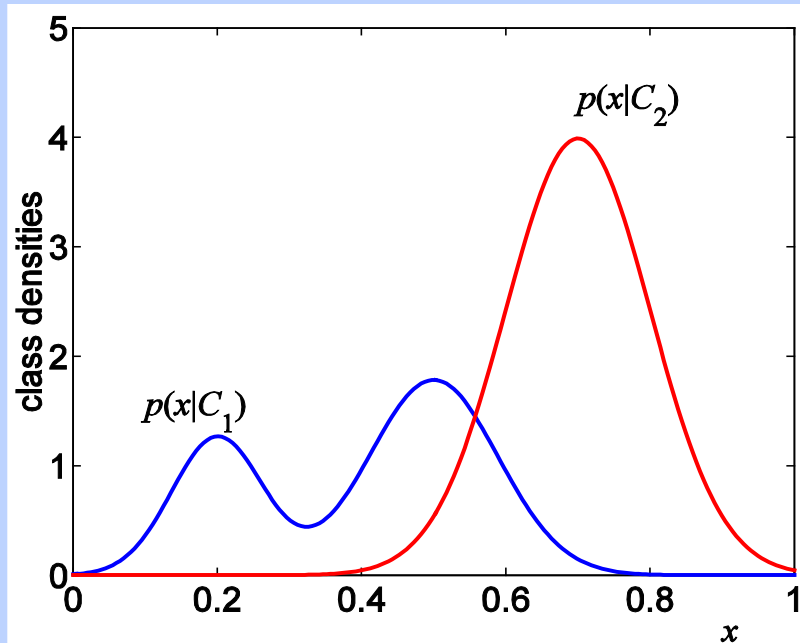
Learning

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



Learning

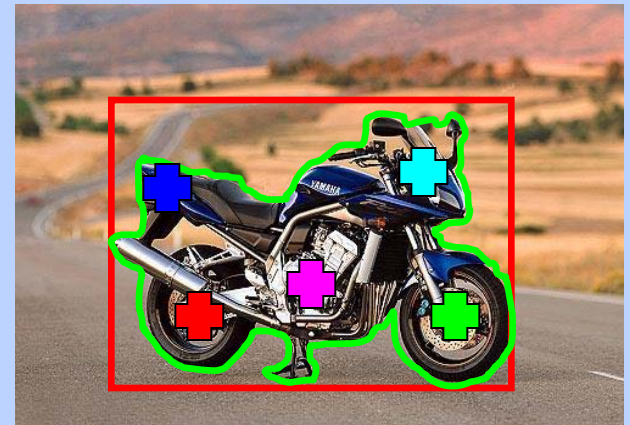
- Unclear how to model categories, so we learn what distinguishes them rather than manually specify the difference -- hence current interest in machine learning)
- Methods of training: generative vs. discriminative



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

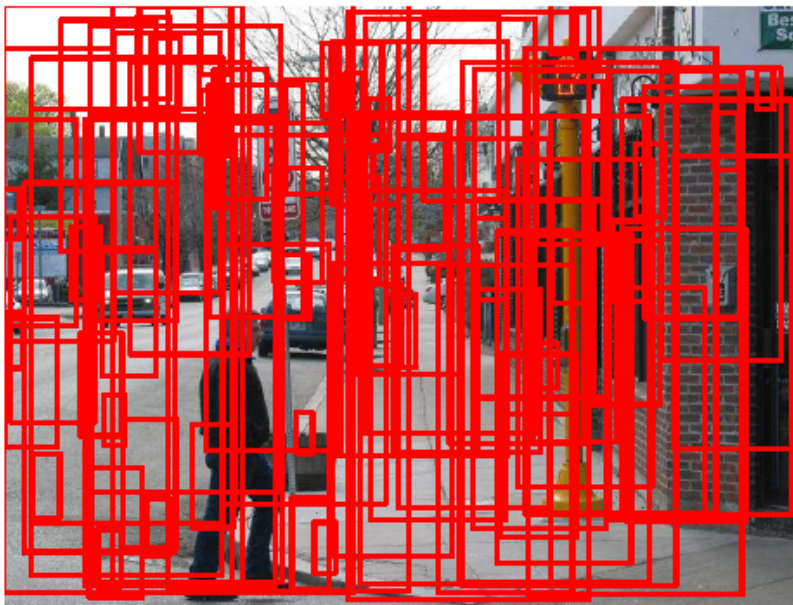
Learning

- 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

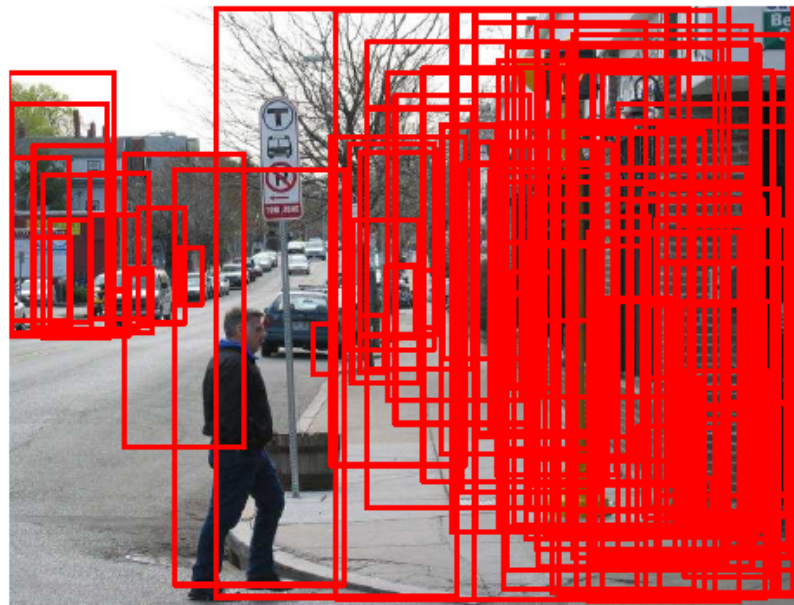
Recognition

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

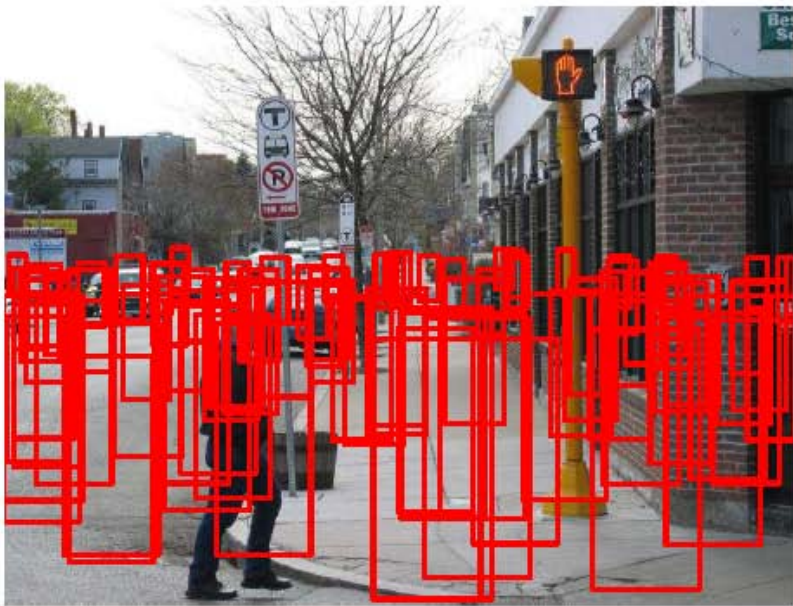




(b) $P(\text{person}) = \text{uniform}$



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



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



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