Recognizing and Learning
Object Categories: Year 2007

Li Fei-Fei, Princeton
Rob Fergus, MIT
Antonio Torralba, MIT
Agenda

• Introduction
• Bag-of-words models
• Part-based models
• Discriminative methods
• Segmentation and recognition
• Datasets & Conclusions
object  n.

1. Something visible to one or more of the senses, especially sight or touch; a material thing.
2. A focus of attention, thinking, or action: an object of concern.
3. The purpose or effect of a specific action or effort: the object of a game.
4. Grammar. a. A noun, pronoun, or noun phrase that receives or is affected by the action of a verb or a preposition within a sentence.
   b. A noun or substantive governed by a preposition.
5. Philosophy. Something intangible or perceptible by the mind.
6. Computer Science. A discrete item that can be selected and maneuvered, such as an onscreen graphic. In object-oriented programming, objects include data and the procedures necessary to operate on that data.
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?

~10,000 to 30,000

Biederman 1987
So what does object recognition involve?
Verification: is that a lamp?
Detection: are there people?
Identification: is that Potala Palace?
Object categorization

- tree
- banner
- mountain
- building
- street lamp
- vendor
- people
Scene and context categorization

• outdoor
• city
• ...

[Image of a cityscape with a pink rectangle indicating scene and context categories]
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

Query: STREET

Organizing photo collections
Challenges 1: view point variation

Michelangelo 1475-1564
Challenges 2: illumination

slide credit: S. Ullman
Challenges 3: occlusion

Magritte, 1957
Challenges 4: scale
Challenges 5: deformation

Xu, Beihong 1943
Challenges 6: background clutter

Klimt, 1913
History: single object recognition
History: single object recognition

• Lowe, et al. 1999, 2003
• Mahamud and Herbert, 2000
• Ferrari, Tuytelaars, and Van Gool, 2004
• Rothganger, Lazebnik, and Ponce, 2004
• Moreels and Perona, 2005
• ...
Challenges 7: intra-class variation
History: early object categorization
~10,000 to 30,000
Object categorization: the statistical viewpoint

\[ p(\text{zebra} \mid \text{image}) \]
\[ p(\text{no zebra} \mid \text{image}) \]

• Bayes rule:

\[ \frac{p(\text{zebra} \mid \text{image})}{p(\text{no zebra} \mid \text{image})} = \frac{p(\text{image} \mid \text{zebra})}{p(\text{image} \mid \text{no zebra})} \cdot \frac{p(\text{zebra})}{p(\text{no zebra})} \]

- posterior ratio
- likelihood ratio
- prior ratio
Object categorization: the statistical viewpoint

\[
\frac{p(\text{zebra} \mid \text{image})}{p(\text{no zebra} \mid \text{image})} = \frac{p(\text{image} \mid \text{zebra})}{p(\text{image} \mid \text{no zebra})} \cdot \frac{p(\text{zebra})}{p(\text{no zebra})}
\]

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

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

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

<table>
<thead>
<tr>
<th>$p(\text{image} \mid \text{zebra})$</th>
<th>$p(\text{image} \mid \text{no zebra})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>Middle</td>
</tr>
<tr>
<td>High</td>
<td>Middle $\rightarrow$ Low</td>
</tr>
</tbody>
</table>
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 with 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
ANIMALS

VERTEBRATE

MAMMALS

TAPIR

BOAR

BIRDS

GROUSE

INANIMATE

NATURAL

MAN-MADE

OBJECTS

PLANTS

CAMERA

...