Statistical Recognition

Slides adapted from Fei-Fei Li, Rob Fergus, Antonio Torralba, and Kristen Grauman
Object categorization: the statistical viewpoint

• MAP decision:  \( p(zebra \mid image) \)
  vs.
  \( p(no\ zebra \mid image) \)
Object categorization: the statistical viewpoint

- MAP decision: \( p(\text{zebra} \mid \text{image}) \)
  \[ \text{vs.} \]
  \( p(\text{no zebra} \mid \text{image}) \)

- Bayes rule:

\[
p(\text{zebra} \mid \text{image}) \propto p(\text{image} \mid \text{zebra}) p(\text{zebra})
\]

\( p(\text{image} \mid \text{zebra}) \) \quad \text{posterior}
\( p(\text{image} \mid \text{zebra}) \) \quad \text{likelihood}
\( p(\text{zebra}) \) \quad \text{prior}
Object categorization: the statistical viewpoint

\[ p(\text{zebra} \mid \text{image}) \propto p(\text{image} \mid \text{zebra}) p(\text{zebra}) \]

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

- Direct modeling of $p(\text{zebra} \mid \text{image})$
Generative methods

- 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           |
Generative vs. discriminative learning

Generative

Discriminative

Class densities

$p(x|C_1)$

$p(x|C_2)$

$p(C_1|x)$

$p(C_2|x)$

Posterior probabilities

$x$
Generative vs. discriminative methods

- Generative methods
  + Can sample from them / compute how probable any given model instance is
  + Can be learned using images from just a single category
  – Sometimes we don’t need to model the likelihood when all we want is to make a decision

- Discriminative methods
  + Efficient
  + Often produce better classification rates
  – Require positive and negative training data
  – Can be hard to interpret
Steps for statistical recognition

• Representation
  – Specify the model for an object category
  – Bag of features, part-based, global, etc.

• Learning
  – Given a *training set*, find the parameters of the model
  – Generative vs. discriminative

• Recognition
  – Apply the model to a new *test image*
Generalization

• How well does a learned model generalize from the data it was trained on to a new test set?

• **Underfitting**: model is too “simple” to represent all the relevant class characteristics
  – High training error and high test error

• **Overfitting**: model is too “complex” and fits irrelevant characteristics (noise) in the data
  – Low training error and high test error

• Occam’s razor: given two models that represent the data equally well, the simpler one should be preferred
Supervision

• Images in the training set must be annotated with the “correct answer” that the model is expected to produce

Contains a motorbike
Spectrum of supervision

Unsupervised

“Weakly” supervised

Fully supervised

Definition depends on task
What task?

• Classification
  – Object present/absent in image
  – Background may be correlated with object

• Localization / Detection
  – Localize object within the frame
  – Bounding box or pixel-level segmentation
Datasets

• Circa 2001: 5 categories, 100s of images per category
• Circa 2004: 101 categories
• Today: thousands of categories, tens of thousands of images
Caltech 101 & 256

http://www.vision.caltech.edu/Image_Datasets/Caltech101/
http://www.vision.caltech.edu/Image_Datasets/Caltech256/

Griffin, Holub, Perona, 2007

Fei-Fei, Fergus, Perona, 2004

http://pascallin.ecs.soton.ac.uk/challenges/VOC/

2008 Challenge classes:

*Person*: person

*Animal*: bird, cat, cow, dog, horse, sheep

*Vehicle*: aeroplane, bicycle, boat, bus, car, motorbike, train

*Indoor*: bottle, chair, dining table, potted plant, sofa, tv/monitor

http://pascallin.ecs.soton.ac.uk/challenges/VOC/

• Main competitions
  – **Classification**: For each of the twenty classes, predicting presence/absence of an example of that class in the test image
  – **Detection**: Predicting the bounding box and label of each object from the twenty target classes in the test image

http://pascallin.eecs.soton.ac.uk/challenges/VOC/

- **“Taster” challenges**
  - **Segmentation:**
    Generating pixel-wise segmentations giving the class of the object visible at each pixel, or "background" otherwise
  - **Person layout:**
    Predicting the bounding box and label of each part of a person (head, hands, feet)
Lotus Hill Research Institute image corpus

http://www.imageparsing.com/

Figure 5: Two examples of the parse trees (cat and car) in the Lotus Hill Research Institute image corpus. From [87].

Labeling with games

http://www.gwap.com/gwap/

Figure 1. Partners agreeing on an image in the ESP Game. Neither player can see the other’s guesses.

Figure 2. Peekaboom. “Peek” tries to guess the word associated with an image slowly revealed by “Boom.”

LabelMe

http://labelme.csail.mit.edu/

Russell, Torralba, Murphy, Freeman, 2008
80 Million Tiny Images

http://people.csail.mit.edu/torralba/tinyimages/
Dataset issues

• How large is the degree of intra-class variability?
• How “confusabble” are the classes?
• Is there bias introduced by the background? I.e., can we “cheat” just by looking at the background and not the object?
Caltech-101
Summary

- Recognition is the “grand challenge” of computer vision
- History
  - Geometric methods
  - Appearance-based methods
  - Sliding window approaches
  - Local features
  - Parts-and-shape approaches
  - Bag-of-features approaches
- Statistical recognition concepts
  - Generative vs. discriminative models
  - Generalization, overfitting, underfitting
  - Supervision
- Tasks, datasets