Recognition

Subhransu Maji

CMPSCI 670: Computer Vision

October 25, 2016
Agenda for the next few lectures

- Overview of recognition
- Image representations
- Machine learning
- Deep learning
Scene categorization

- outdoor/indoor
- city/forest/factory/etc.
Image annotation/tagging

- street
- people
- building
- mountain
- …
Object detection

- find pedestrians
Activity recognition

- walking
- shopping
- rolling a cart
- sitting
- talking
- …
Image parsing

sky

mountain

building

tree

building

banner

street lamp

market

people
Visual question answering

How many people are waking on the street?
Where was this picture taken? (external knowledge)
How many visual object categories?

~10,000 to 30,000

Biederman 1987

~10,000 to 30,000
Categorization spectrum

#categories

basic categories  individuals

Figure credit: Ryan Farrell
History of ideas in recognition

1960s – early 1990s: the geometric era
1990s: appearance-based models
Late 1990s: local features
Early 2000s: parts-and-shape models
Mid-2000s: bags-of-features, learning-based techniques
Present trends: big data, recognition + X (X=geometry, robotics, language), deep learning, getting AI to work, many applications: health care, autonomous driving, face recognition, image/video search, etc.
Recognition by learning
The machine learning framework

Apply a prediction function to a feature representation of the image to get the desired output:

\[ f(\text{apple}) = \text{“apple”} \]
\[ f(\text{tomato}) = \text{“tomato”} \]
\[ f(\text{cow}) = \text{“cow”} \]
The machine learning framework

\[ y = f(x) \]

**Training:** given a *training set* of labeled examples \( \{(x_1, y_1), \ldots, (x_N, y_N)\} \), estimate the prediction function \( f \) by minimizing the prediction error on the training set.

**Testing:** apply \( f \) to a never before seen *test example* \( x \) and output the predicted value \( y = f(x) \).
Ingredients for learning

- **Whole idea:** Inject *your* knowledge into a learning system

- **Sources of knowledge:**
  1. Feature representation
     - Not typically a focus of machine learning
     - Typically seen as “problem specific”
     - However, it’s hard to learn from bad representations
  2. Training data: labeled examples
     - Often expensive to label lots of data
     - Sometimes data is available for “free”
  3. Model
     - No single learning algorithm is always good ("no free lunch")
     - Different learning algorithms work with different ways of representing the learned classifier
Features (examples)

Raw pixels (and simple functions of raw pixels)

GIST descriptors

Gradient histograms

bags of features
Recognition task and 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
How well does a learned model *generalize* from the data it was trained on to a new test set?

**Training set (labels known)**

**Test set (labels unknown)**
Datasets

Circa 2001: five categories, hundreds of images per category
Circa 2004: 101 categories
Today: up to thousands of categories, millions 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
Caltech-101: Intra-class variability
PASCAL Visual Object Classes Challenge (2005-12)

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

• **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

• **Dataset size (by 2012):**
  11.5K training/validation images, 27K bounding boxes, 7K segmentations
LabelMe Dataset [http://labelme.csail.mit.edu/]

Russell, Torralba, Murphy, Freeman, 2008
ImageNet is an image database organized according to the WordNet hierarchy (currently only the nouns), in which each node of the hierarchy is depicted by hundreds and thousands of images. Currently we have an average of over five hundred images per node. We hope ImageNet will become a useful resource for researchers, educators, students and all of you who share our passion for pictures. Click here to learn more about ImageNet, Click here to join the ImageNet mailing list.

What do these images have in common? Find out!

Check out the ImageNet Challenge 2014!