Lecture 31: Modern object recognition

Visual Object Classes Challenge 2009 (VOC2009)

Announcements

• Project 4 due today
• Demos Monday
Object detection: where are we?

- Incredible progress in the last ten years
- Better features, better models, better learning methods, better datasets
- Combination of science and hacks

Vision Contests

- PASCAL VOC Challenge

- 20 categories
- Annual classification, detection, segmentation, ... challenges
Machine learning for object detection

• What features do we use?
  – intensity, color, gradient information, ...

• Which machine learning methods?
  – generative vs. discriminative
  – k-nearest neighbors, boosting, SVMs, ...

• What hacks do we need to get things working?

Histogram of Oriented Gradients (HoG)

[Dalal and Triggs, CVPR 2005]
Histogram of Oriented Gradients (HoG)

Like SIFT (Scale Invariant Feature Transform), but...
- Sampled on a dense, regular grid
- Gradients are contrast normalized in overlapping blocks

[Dalal and Triggs, CVPR 2005]
Histogram of Oriented Gradients (HoG)

- First used for application of person detection [Dalal and Triggs, CVPR 2005]
- Cited since in thousands of computer vision papers

Linear classifiers

- Find linear function to separate positive and negative examples

\[ x_i \text{ positive: } x_i \cdot w + b \geq 0 \]
\[ x_i \text{ negative: } x_i \cdot w + b < 0 \]

Which line is best?

[slide credit: Kristin Grauman]
Support Vector Machines (SVMs)

- Discriminative classifier based on *optimal separating line* (for 2D case)
- Maximize the *margin* between the positive and negative training examples

---

Support vector machines

- Want line that maximizes the margin.

For support, vectors, $x_i \cdot w + b = \pm 1$

---


[slide credit: Kristin Grauman]
Person detection, ca. 2005

1. Represent each example with a single, fixed HoG template

2. Learn a single [linear] SVM as a detector

Code available: http://pascal.inrialpes.fr/soft/olt/

Positive and negative examples

+ thousands more...

+ millions more...
HoG templates for person detection

Person detection with HoG & linear SVM

[Dalal and Triggs, CVPR 2005]
Are we done?

• Single, rigid template usually not enough to represent a category
  – Many objects (e.g. humans) are articulated, or have parts that can vary in configuration
  – Many object categories look very different from different viewpoints, or from instance to instance
Difficulty of representing positive instances

• Discriminative methods have proven very powerful
• But linear SVM on HoG templates not sufficient?

• Alternatives:
  – Parts-based models [Felzenszwalb et al. CVPR 2008]
  – Latent SVMs [Felzenszwalb et al. CVPR 2008]
  – Today’s paper [Exemplar-SVMs, Malisiewicz, et al. ICCV 2011]

Parts-based models

Our first innovation involves enriching the Dalal-Triggs model using a star-structured part-based model defined by a “root” filter (analogous to the Dalal-Triggs filter) plus a set of parts filters and associated deformation models.

Felzenszwalb, et al., Discriminatively Trained Deformable Part Models, 
http://people.cs.uchicago.edu/~pff/latent/
Latent SVMs

• Rather than training a single linear SVM separating positive examples...
• ... cluster positive examples into “components” and train a classifier for each (using all negative examples)

Two-component bicycle model
Six-component car model

- Side view
  - Root filters (coarse)
  - Part filters (fine)
  - Deformation models

Six-component person model
The PASCAL Visual Object Classes Challenge 2009 (VOC2009)

• Twenty object categories (aeroplane to TV/monitor)

• Three challenges:
  – Classification challenge (is there an X in this image?)
  – Detection challenge (draw a box around every X)
  – Segmentation challenge