Object detection as supervised classification

Announcements
• A4 due today
• A5 out, due May 2
• Exam May 10, 2-5 pm

Last time
• Introduction to object categorization
• Window-based object detection
  – boosting classifiers
  – face detection as case study

Today
• Recap of boosting + face detection
• Pros/cons of window-based detectors
• Mosaic examples
• Support vector machines
See slides / handout from lecture 22

Boosting: pros and cons

• Advantages of boosting
  - Integrates classification with feature selection
  - Complexity of training is linear in the number of training examples
  - Flexibility in the choice of weak learners, boosting scheme
  - Testing is fast
  - Easy to implement

• Disadvantages
  - Needs many training examples
  - Other discriminative models may outperform in practice (SVMs, CNNs, …)
    - especially for many-class problems

Window-based detection: strengths

• Sliding window detection and global appearance descriptors:
  - Simple detection protocol to implement
  - Good feature choices critical
  - Past successes for certain classes
Window-based detection: Limitations

- High computational complexity
  - For example: 250,000 locations x 30 orientations x 4 scales = 30,000,000 evaluations!
  - If training binary detectors independently, means cost increases linearly with number of classes
- With so many windows, false positive rate better be low

Limitations (continued)

- Not all objects are “box” shaped

Limitations (continued)

- Non-rigid, deformable objects not captured well with representations assuming a fixed 2d structure; or must assume fixed viewpoint
- Objects with less-regular textures not captured well with holistic appearance-based descriptions
Limitations (continued)

• If considering windows in isolation, context is lost

<table>
<thead>
<tr>
<th>Sliding window</th>
<th>Detector’s view</th>
</tr>
</thead>
</table>

Figure credit: Derek Hoiem
Slide: Kristen Grauman

Limitations (continued)

• In practice, often entails large, cropped training set (expensive)
• Requiring good match to a global appearance description can lead to sensitivity to partial occlusions

Image credit: Adam, Rivlin, & Shimshoni
Slide: Kristen Grauman

Summary so far

• Basic pipeline for window-based detection
  – Model/representation/classifier choice
  – Sliding window and classifier scoring
• Boosting classifiers: general idea
• Viola-Jones face detector
  – Exemplar of basic paradigm
  – Plus key ideas: rectangular features, Adaboost for feature selection, cascade
• Pros and cons of window-based detection
Object proposals

Main idea:
- Learn to generate category-independent regions/boxes that have object-like properties.
- Let object detector search over "proposals", not exhaustive sliding windows

Alexe et al. Measuring the objectness of image windows, PAMI 2012

Multi-scale saliency
Color contrast

Object proposals

Edge density Superpixel straddling

(a) (b) (a) (b)

Alexe et al. Measuring the objectness of image windows, PAMI 2012
Object proposals

More proposals

Alexe et al. Measuring the objectness of image windows, PAMI 2012

Region-based object proposals


MOSAIC EXAMPLES
Window-based models:
Three case studies

- SVM + person detection (e.g., Dalal & Triggs)
- Boosting + face detection (Viola & Jones)
- NN + scene Gist classification (e.g., Hays & Efros)

Linear classifiers

- Find linear function to separate positive and negative examples

\[ x \cdot w + b \geq 0 \]
\[ x \cdot w + b < 0 \]

Which line is best?
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.

\[ w \cdot x + b = 1 \]

For support vectors, \( w \cdot x + b = \pm 1 \)

C. Burges, A Tutorial on Support Vector Machines for Pattern Recognition, Data Mining and Knowledge Discovery, 1998
Support vector machines

- Want line that maximizes the margin.

\[ x, \text{positive} (y_i = 1): \quad x_i \cdot w + b \geq 1 \]
\[ x, \text{negative} (y_i = -1): \quad x_i \cdot w + b \leq -1 \]

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

Distance between point and line:
\[ \frac{|x_i \cdot w + b|}{||w||} \]

Therefore, the margin is \[ \frac{2}{||w||} \]

Finding the maximum margin line

1. Maximize margin \( \frac{2}{||w||} \)
2. Correctly classify all training data points:
   - \[ x, \text{positive} (y_i = 1): \quad x_i \cdot w + b \geq 1 \]
   - \[ x, \text{negative} (y_i = -1): \quad x_i \cdot w + b \leq -1 \]

Quadratic optimization problem:

Minimize \( \frac{1}{2} w^T w \)
Subject to \( y_i(w \cdot x_i + b) \geq 1 \)

Finding the maximum margin line

- Solution: \( w = \sum \alpha_i y_i x_i \)
  - learned weight
  - Support vector

C. Burges, A Tutorial on Support Vector Machines for Pattern Recognition, Data Mining and Knowledge Discovery
Finding the maximum margin line

- Solution: \( \mathbf{w} = \sum \alpha_i y_i \mathbf{x}_i \)
  \( b = y_i - \mathbf{w} \cdot \mathbf{x}_i \) (for any support vector)
  \( \mathbf{w} \cdot \mathbf{x} + b = \sum \alpha_i y_i \mathbf{x}_i \cdot \mathbf{x} + b \)

- Classification function:
  \( f(x) = \text{sign} (\mathbf{w} \cdot \mathbf{x} + b) \)
  \( = \text{sign} \left( \sum \alpha_i y_i \mathbf{x}_i \cdot \mathbf{x} + b \right) \)

  If \( f(x) < 0 \), classify as negative.
  If \( f(x) > 0 \), classify as positive.


Person detection
with HoG’s & linear SVM’s

- Histogram of oriented gradients (HoG): Map each grid cell in the input window to a histogram counting the gradients per orientation.
- Train a linear SVM using training set of pedestrian vs. non-pedestrian windows.

Dalal & Triggs, CVPR 2005

Person detection
with HoGs & linear SVMs

- Histograms of Oriented Gradients for Human Detection, Navneet Dalal, Bill Triggs, International Conference on Computer Vision & Pattern Recognition - June 2005
Summary

- Object recognition as classification task
  - Boosting (face detection ex)
  - Support vector machines and HOG (person detection ex)
- Sliding window search paradigm
  - Pros and cons
  - Speed up with attentional cascade
  - Object proposals, proposal regions as alternative

Next time

- What if the data are not linearly separable?
- What about the multi-class case?
- Nearest neighbors
- Convolutional neural networks