Object Detection: Now and Next

Xiaoyu Wang
NEC Labs America
Introduction

What’s the task:
- Is there an object
- Where is the object

Related Topics:
- Action Recognition
  - A person is riding a motorbike
- Scene Understanding
  - A person is riding a motorbike on the road
- Identification
- …
Introduction

Literatures have been focused on:
- Features
- Handling Deformation
- Handling intra-class variation
- Improving Speed
- Handling Occlusion

Another way to separate them:
- Fix model size
- Fix image size

The third way: Sliding window vs selective search
Introduction

Most popular features:
- Haar features
  - Scale invariant
  - Relative weak feature
- Color Histogram
  - Scale invariant
  - Relative weak feature
Introduction

Most popular features:

- Histogram of Oriented Gradients
  - Somehow scale invariant
  - Relative strong feature

- Histogram of Local Binary Patterns
  - Not scale invariant
  - Relative strong feature
Introduction

Most popular features:

- **SIFT**
  - Scale invariant
  - Strong feature

- **Shapelet**
  - Somehow scale invariant

- **Covariance**
  - Scale invariant
  - Strong feature
Handling deformation:

- Histogram based features
  - Smooth features in a small patch
- Bag of Words (BoWs)
  - Encode histogram based features with a dictionary (usually obtained by k-means clustering)
- Spatial pyramid
  - Divide a detection window into 1x1, 3x3, 5x5 sub-regions,
  - Build BoWs inside each sub-region
Handling deformation:

- Deformable part based model\(^1\)
  - Composed of the root detector and several part detectors
  - A part detector could be anywhere with a deformation cost from the central position

\(^1\)A Discriminatively Trained, Multiscale, Deformable Part Model, *CVPR*, 2008
Introduction

Handling intra-class variation:

- Multiple components\(^1\)
  - Viewpoints
  - Appearance
- Sub-category learning

\(^1\)How important are “Deformable Parts” in the Deformable Parts Model?, *ECCV Workshops*, 2012
Introduction

Improving speed:
- SVM or Boosting based cascade\(^1,2\)
- Reduce number of evaluated sliding windows\(^3\)
- GPU and multiple threads

\(^1\)Cascade Object Detection with Deformable Part Models, *CVPR*, 2012
\(^2\)Rapid Object Detection using a Boosted Cascade of Simple Features, *CVPR*, 2001
\(^3\)Crosstalk Cascades for Frame-Rate Pedestrian Detection, *ECCV*, 2012
Introduction

Handling Occlusion:
- Include occluded training samples
- Part based models
- Occlusion reasoning\(^1,^2\)
- Latent occlusion handling\(^3\)

\(^1\) Occlusion Reasoning for Object Detection under Arbitrary Viewpoint, *CVPR*, 2012

\(^2\) An HOG LBP Human detector with Partial Occlusion Handling, *ICCV*, 2009

\(^3\) A Segmentation-aware Object Detection Model with Occlusion Handling, *CVPR*, 2011
Fix model vs fix image

Model is fixed
- Fix the model size, rescale the image to detect objects with different sizes.
Fix model vs fix image

Image is fixed:
- Do not resize the image, train a mode for each scale
- Do not resize the image, use single model with the capability of detecting objects of different sizes.
Sliding window vs selective search

**Sliding window**
- Search each possible position
- Most of the state of art approaches are based on sliding window 😊
- Complicate models are not applicable 😞

**Selective search:**
- Only search possible regions¹,² 😞
- Model appearance with sophisticated coding schemes 😊

¹Segmentation as Selective Search for Object Recognition, *ICCV*, 2011
²Measuring the Objectness of image windows, *PAMI*, 2012
Object Detection: Next

- Fine-grained object detection (vs general object detection)
  - It’s a person (vs “it’s a male person”)
  - Between object detection and instance recognition
Object Detection: Next

- Handling intra-class variation, sub-category learning
  - Unsupervised sample clustering
  - Unified framework for model learning and sub-category division
  - Interaction and component sharing between sub-categories
Object Detection: Next

- Selective candidate regions
  - Propose candidate regions
  - Scale invariant features
  - Speedup the testing procedure
Object Detection: Next

- Video object detection
  - Motion segmentation for selective search
  - Cross frame analysis
Object Detection: Next

- Object detection with deep learning
Thank you!