A Discriminatively Trained, Multiscale, Deformable Part Model

February 24, 2016
Adam Allevato
CS 381V
University of Texas at Austin
Outline

- Partial matching
- Non-maximum suppression
- Train image results
- Live demo
Outline

- Partial matching
- Non-maximum suppression
- Train image results
- Live demo
Partial Matching

- Deformable Part Models allows parts of objects to shift around
- What happens when one of the parts is completely missing?
- What happens when the images are hacked to move parts of them around?
Learned HOG Features from INRIA
INRIA Person Dataset Matches
INRIA Person Dataset Matches
Bad Background = Bad Detection
Blocked Parts

- Take the list of part filter responses in a detection
- One by one, replace their area with black pixels
- Test intersection over union against ground truth
Detection
1 Filter Blocked
3 Filters Blocked
Degradation (VOC 2010 Detector)

IoU as filters are blocked

Number of blacked-out filters

IoU
0 Blocked Filters
1 Blocked Filters
2 Blocked Filters
3 Blocked Filters
Degradation (VOC 2007 Detector)

IoU as filters are blocked

Number of blacked-out filters

IoU
0 Blocked Filters
1 Blocked Filter
2 Blocked Filters
3 Blocked Filters
Degradation (VOC 2007 Detector)
Blocked Filters

- DPM is great against this, especially with canonical views
- Shows robustness to occlusion
Random Window Shifts

- Window is shifted by random amount
- The pixels covered are moved to the gap left behind
- All pixel information is maintained
VOC 2010 Bicycle Detector
Ground Truth
No Shifts
One Shift
Static Parts to the Rescue
One Shift
Two Shifts
Three Shifts
Four Shifts
Does how far we shift affect performance?

Averaged across 30 trials!
10 10-Pixel Shifts

Ground truth
Does how many times we shift affect performance?

**IoU based on the distance and number of window shifts**

- Red line: distance = 10px
- Green line: distance = 20px
- Blue line: distance = 40px

Intersection over union vs. number of shifts.
Does how many times we shift affect performance?

The diagram shows IoU based on the distance and number of window shifts. The graph plots the intersection over union (IoU) against the number of shifts for three different distances: 10px, 20px, and 40px. The diagram illustrates how IoU changes as the number of shifts increases, indicating different performance levels for each distance.
Window Shifts

- DPM is robust to small number of window shifts because some part filters still fire correctly
- More shifts give worse performance
- The shift distance does not have appreciable effect on the detection score loss
Partial Matching

- DPM is robust to object parts moving around
- It can also infer positions of hidden or missing object parts
- Sometimes, IoU can actually increase with occlusion
Outline

- Partial matching
- **Non-maximum suppression**
- Train image results
- Live demo
Without NMS, $N = 10$
Without NMS, $N = 50$

predicted bounding boxes

$N = 3$  $N = 4$  $N = 44$
With NMS, N = 3

predicted bounding boxes
Overlap = $|B_i \cap B_j| / |B_j|$
NMS Overlap

- 30 closely correlated matches are detected before the second person is detected
- 42 matches before third person is detected
- Repeated detections for similar objects rank similarly
- NMS helps highlight the weaker matches
- Asymmetric overlap metric allows good windows to subsume smaller windows that lie inside
Non-Maximum Supression

- Helps avoid duplicates
- Also helps let the weaker data show itself when a limit is imposed on the total number of matches
Outline

• Partial matching
• Non-maximum suppression
• **Train image results**
• Live demo
Chicago Elevated Train
VOC 2007 Train Model
VOC 2007 Train Results, N = 1
Without NMS, N=30
Without NMS, N=30

- Many different modes
- Overall high confusion
- Some lonesome outliers
Chicago Elevated Train

- Most detected windows contain mostly train
- No single canonical detection window - “lots of trains”
- No window captures the entire train
- No learned DPM for “train” is long enough to capture this shape
Outline

- Partial matching
- Non-maximum suppression
- Train image results
- Live demo
Live Demo

- INRIA person dataset
- VOC 2010 dataset - “chair”
- Can we fool it?
Summary

- Tested matches with parts of objects missing
- Surveyed non-max suppression effects
- Results on train image: technically correct, but still did not capture entire object
- Girshick's library is mature and can be easily integrated into live application
References

- Original code available on GitHub: https://github.com/rbgirshick/voc-dpm
- My code available on GitHub: https://github.com/Kukanani/voc-dpm
- Images
  - http://cdn.collider.com/wp-content/image-base/Movies/P/Princess_Bride/the_princess_bride_movie_image__1_.jpg
  - http://i.telegraph.co.uk/multimedia/archive/01947/B084FX_1947399c.jpg
Live Cam Examples
VOC 2010 Person Detector

predicted bounding boxes
VOC 2010 Person Detector
Chair Detector
Chair Detector