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

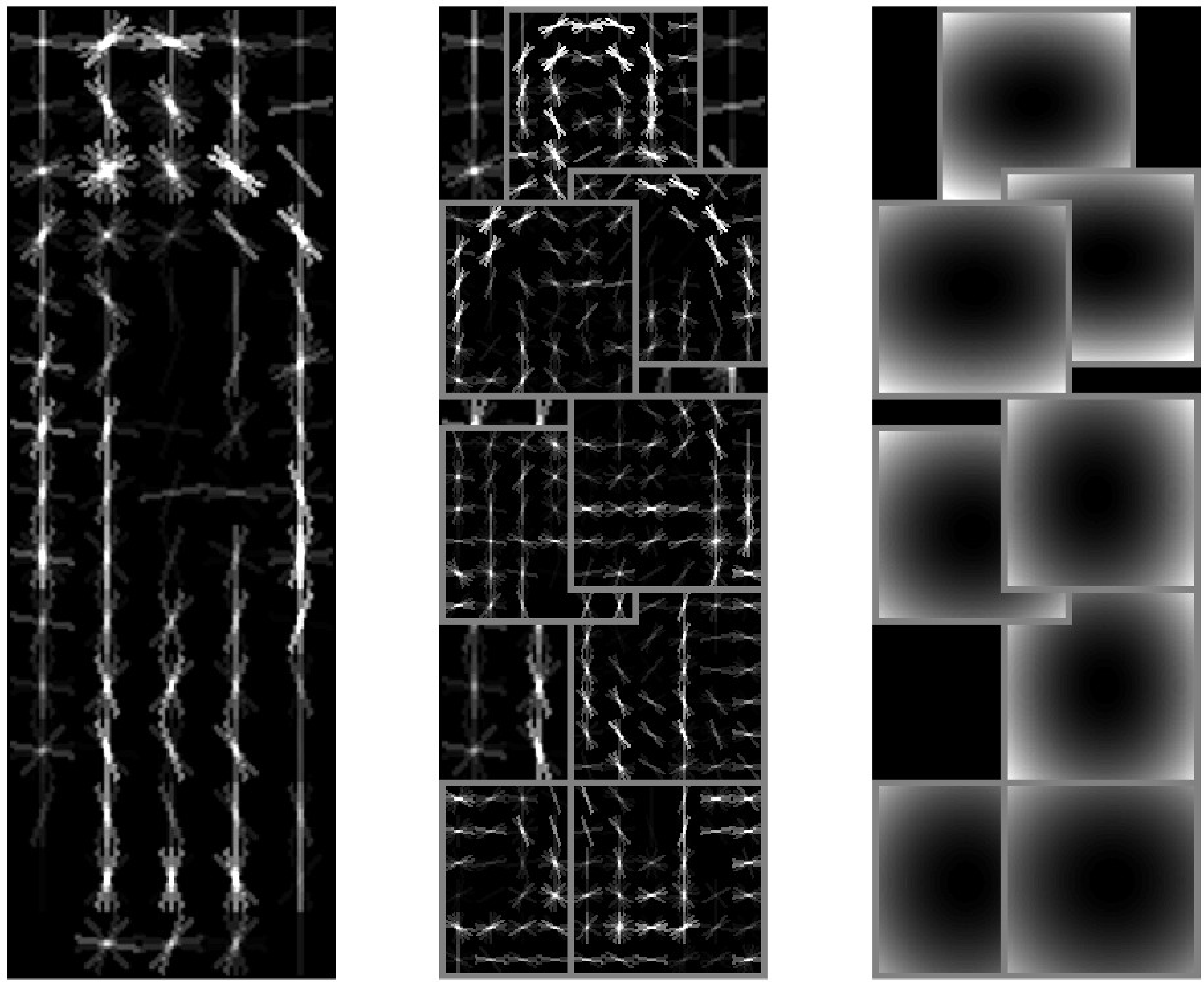
- Partial matching
- Non-maximum suppression
- Train image results
- Live demo

Partial Matching

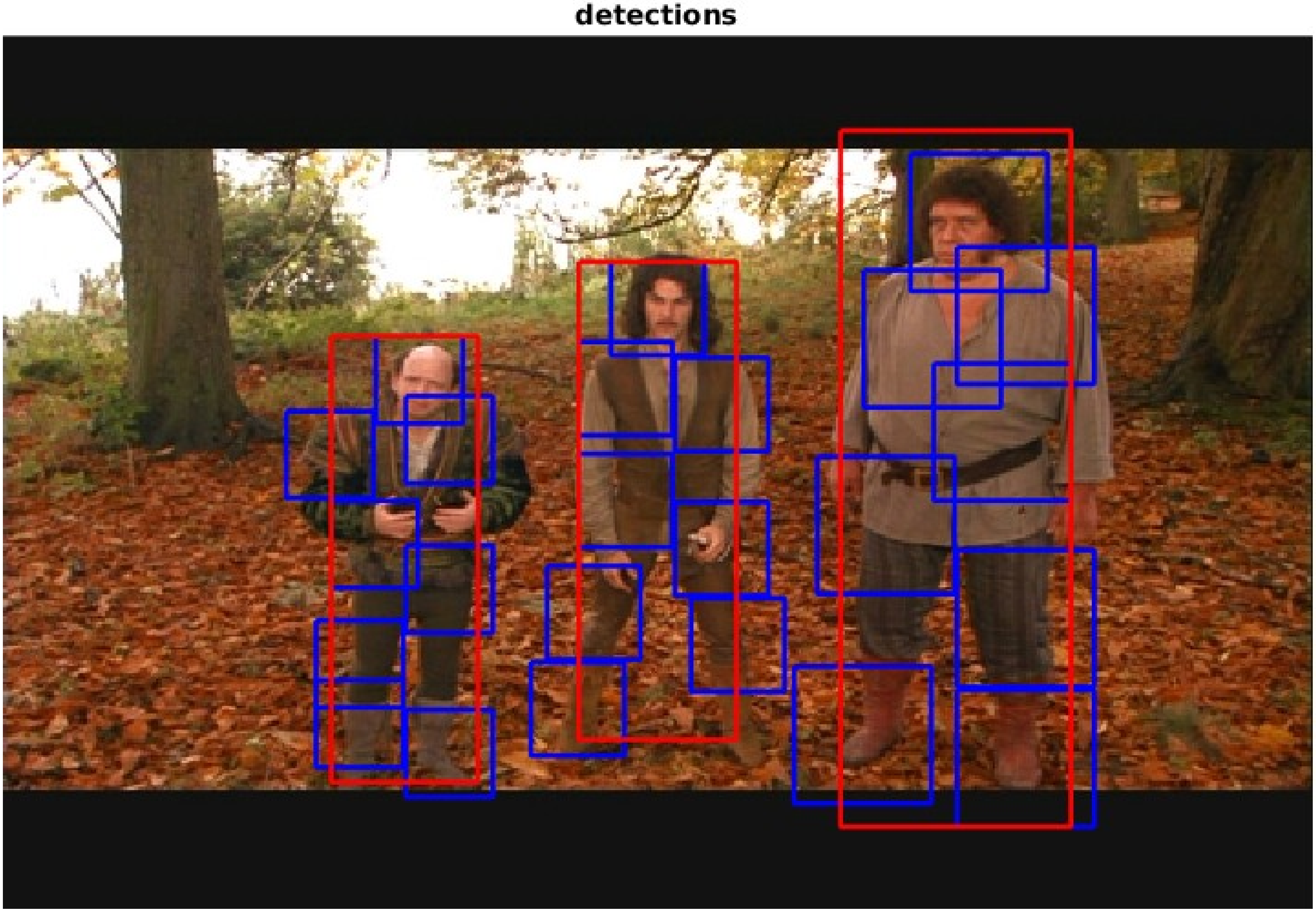
- Deformable Part Models allows parts of objects to shift around
- What happens when one of the parts is completely missing?
- What happens when the images are hacked to move parts of them around?

Source Image

Learned HOG Features from INRIA



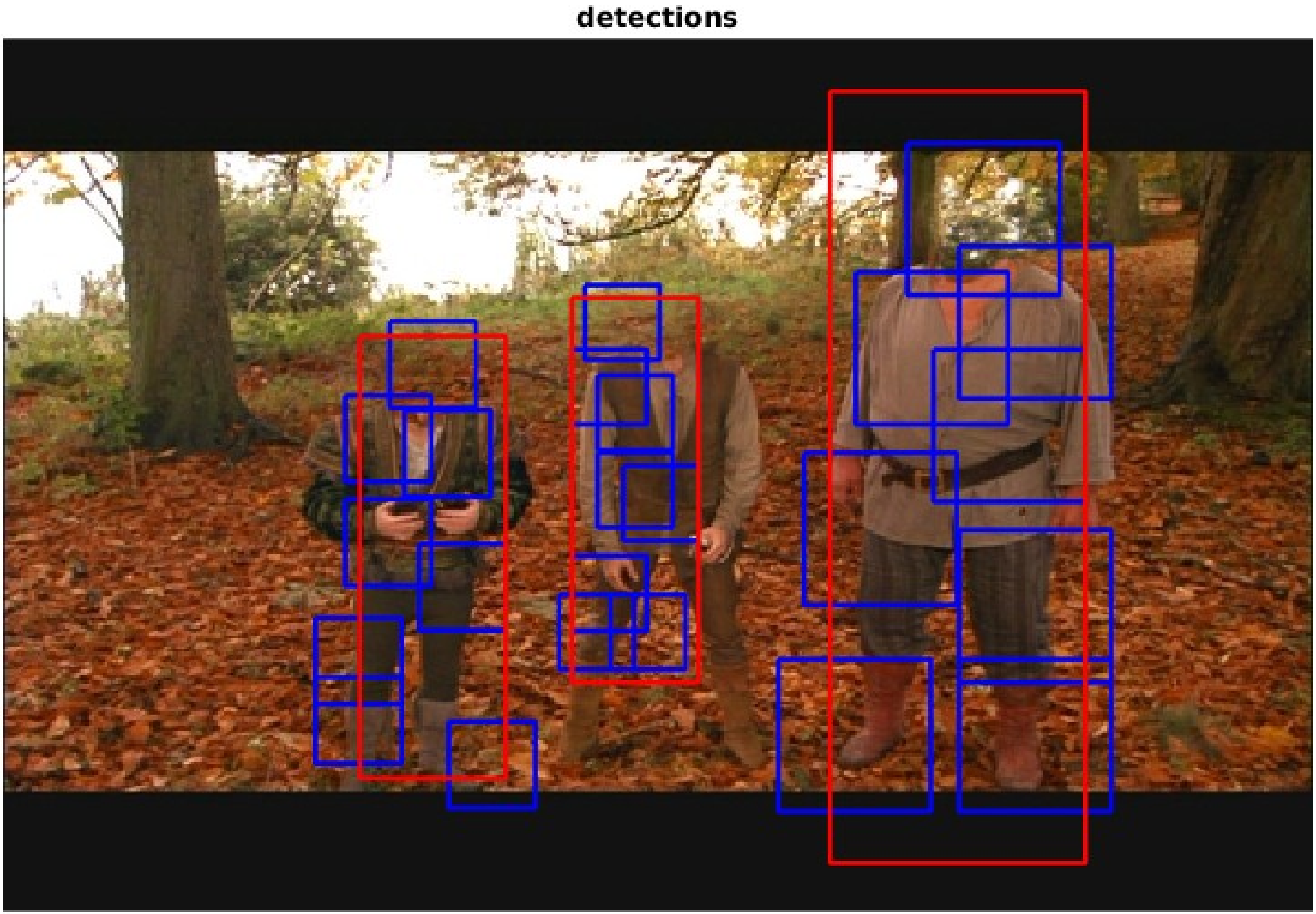
INRIA Person Dataset Matches



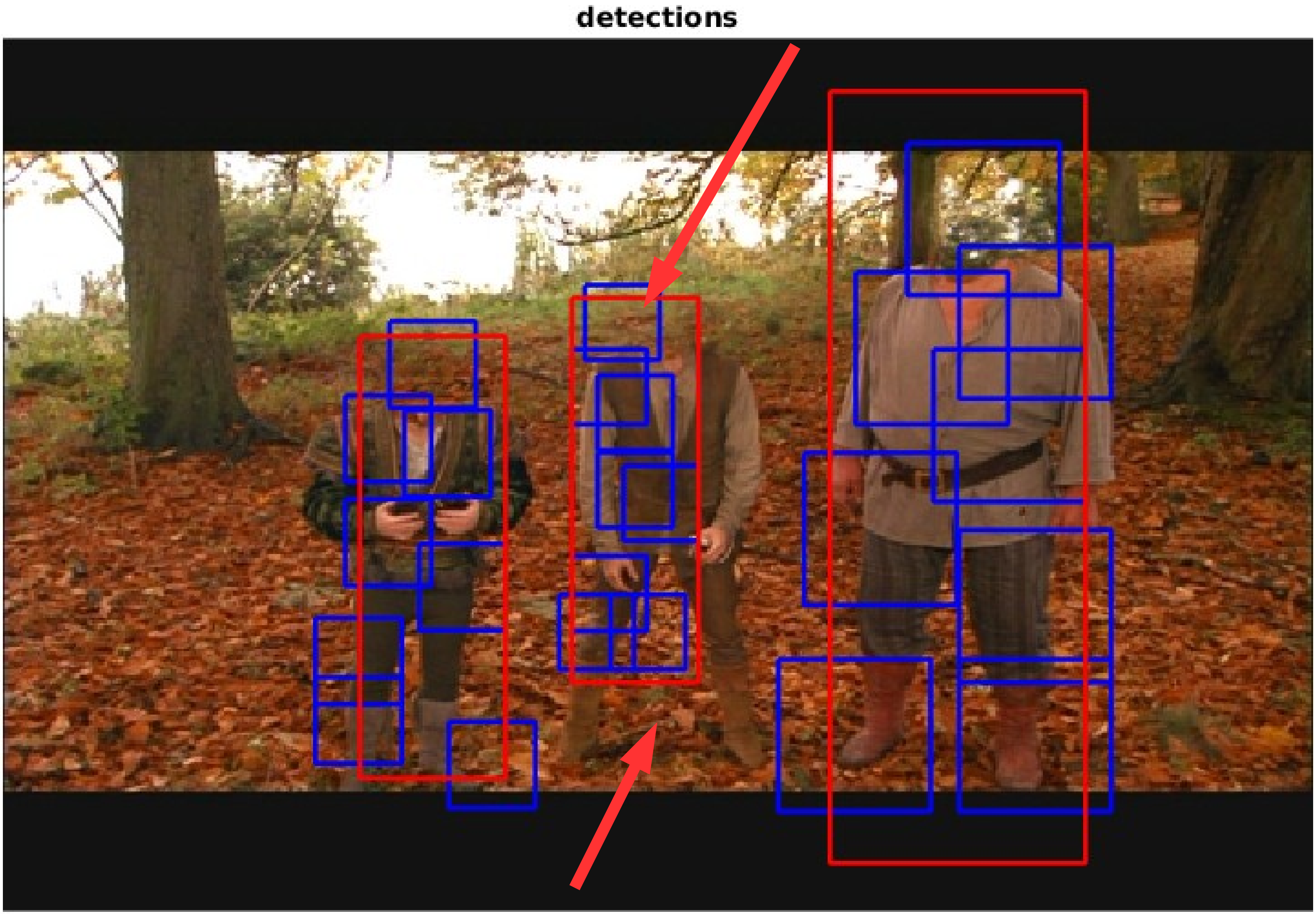
Source Image

Modified Source Image

INRIA Person Dataset Matches



Bad Background = Bad Detection



Blocked Parts

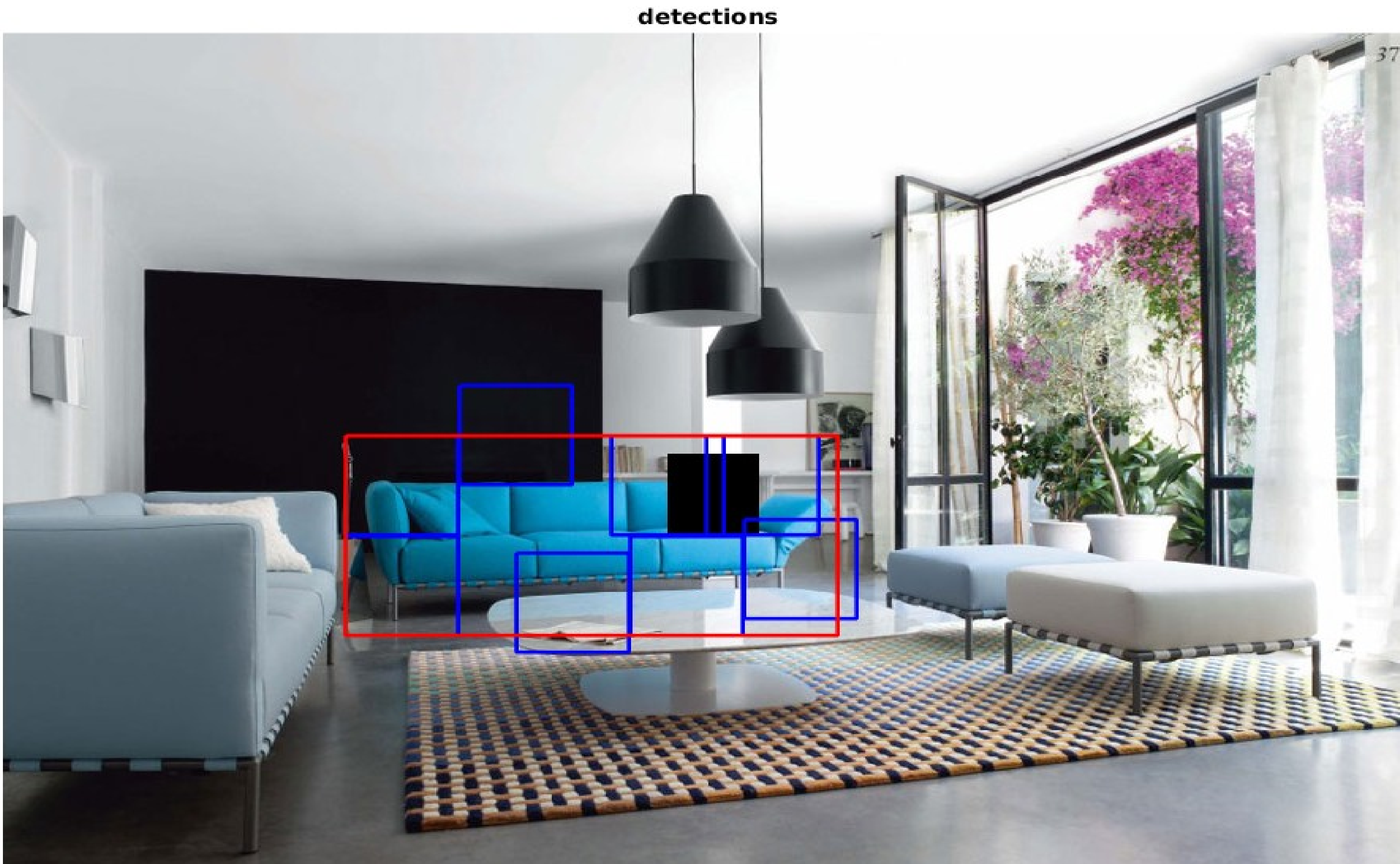
- Take the list of part filter responses in a detection
- One by one, replace their area with black pixels
- Test intersection over union against ground truth

Source Image

Detection

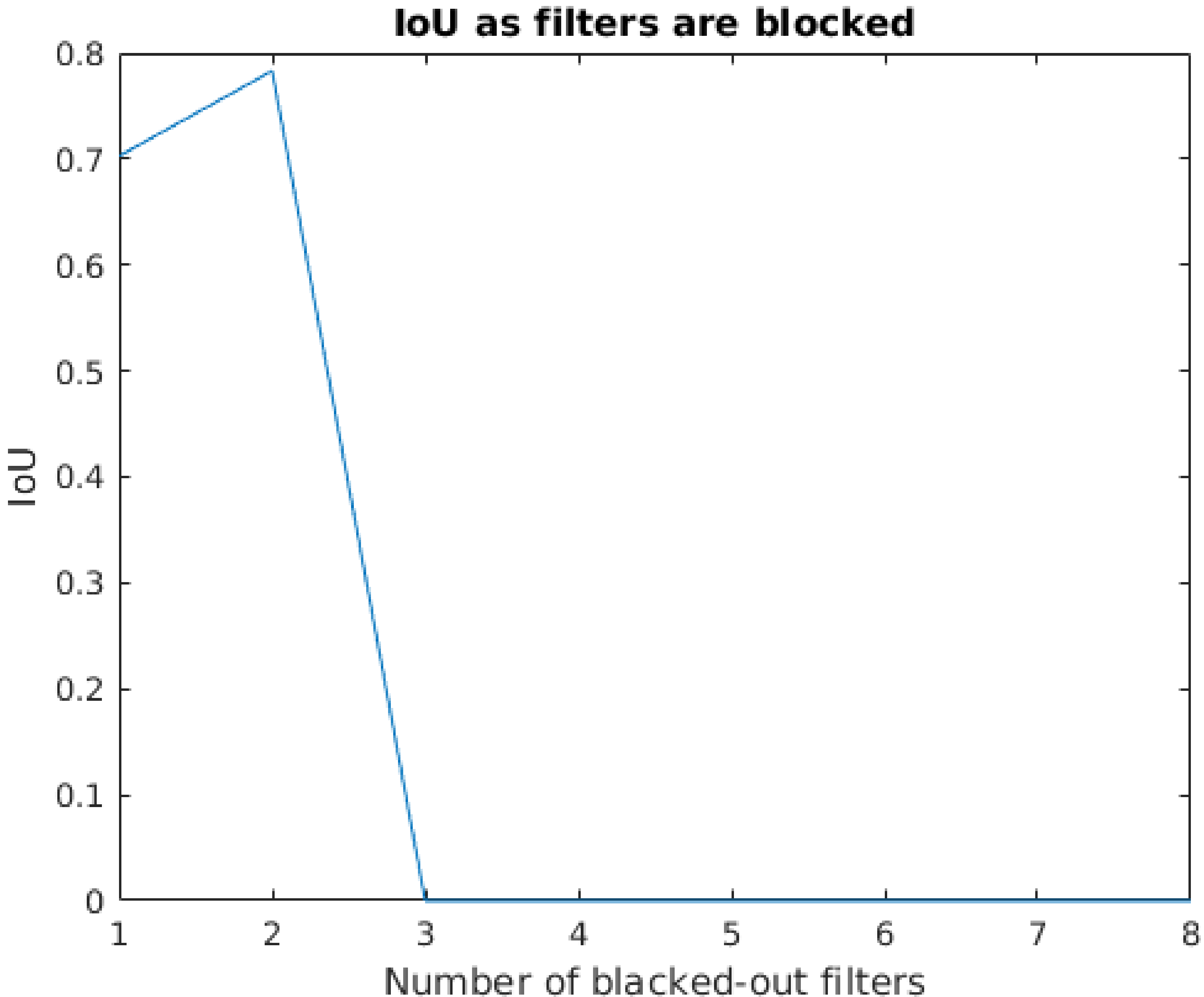
detections

1 Filter Blocked



3 Filters Blocked

Degradation (VOC 2010 Detector)



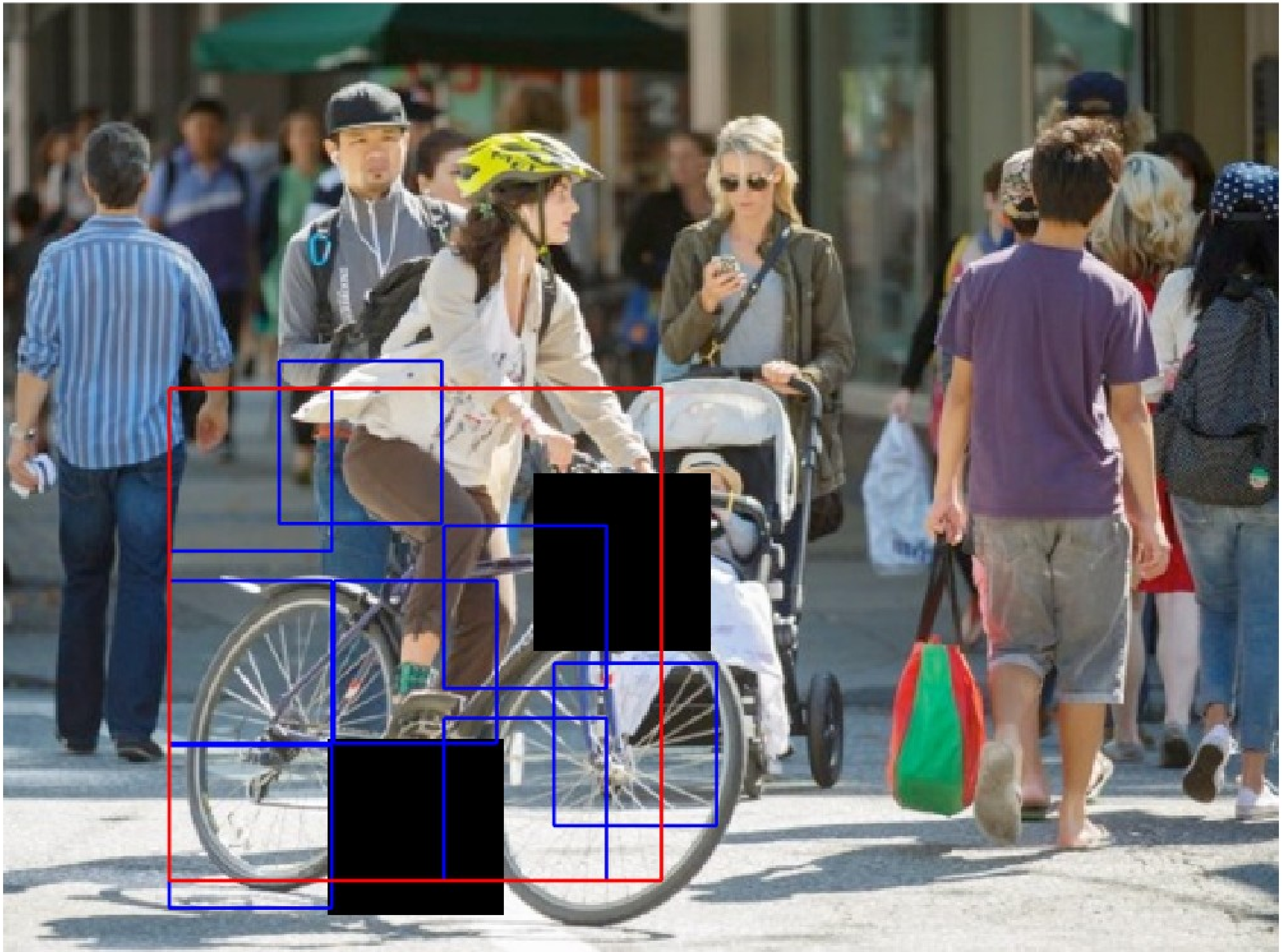
Source Image

0 Blocked Filters

1 Blocked Filters

2 Blocked Filters

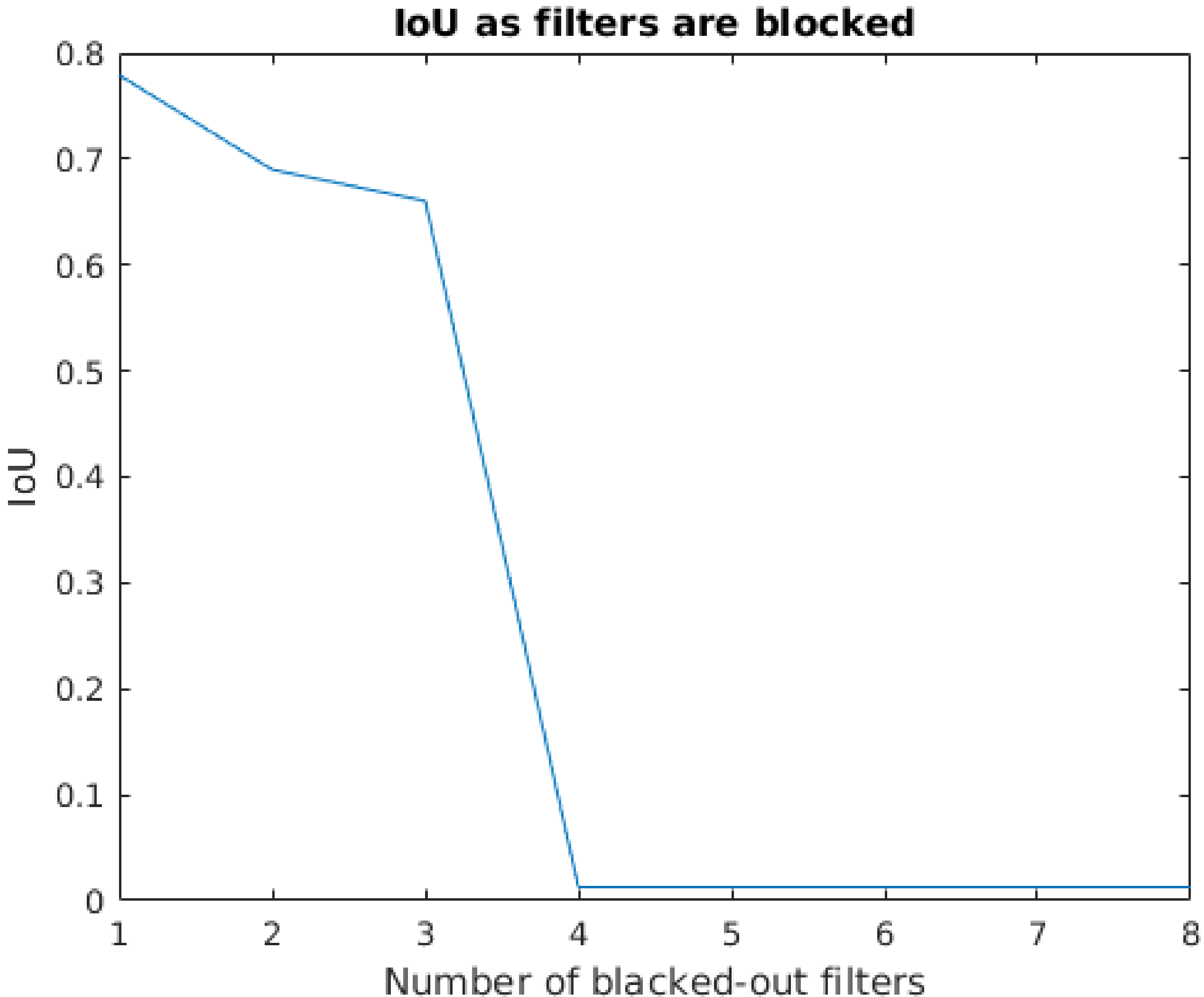
detections



3 Blocked Filters

detections

Degradation (VOC 2007 Detector)



Source Image

0 Blocked Filters

1 Blocked Filter

detections

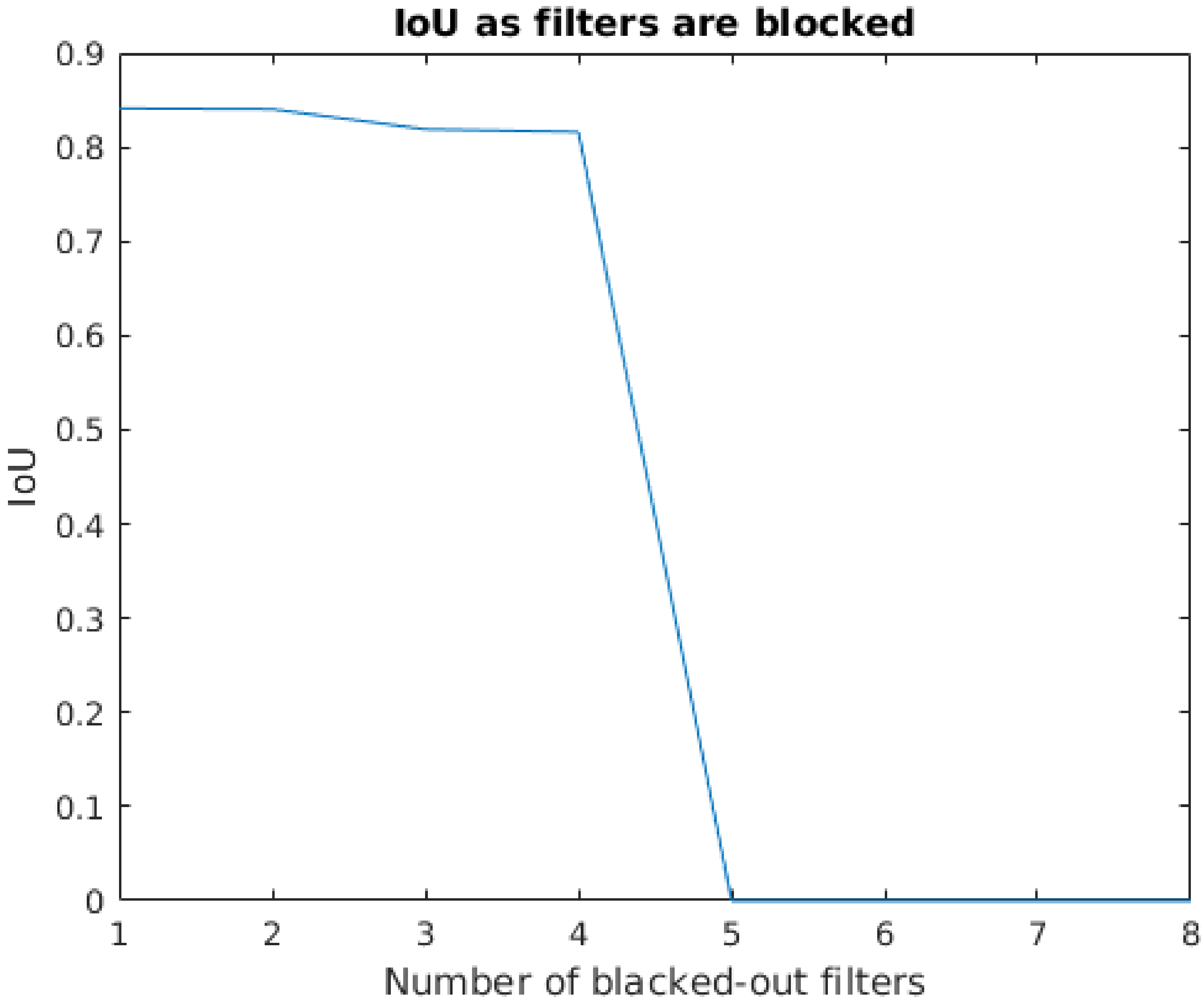
2 Blocked Filters

detections

3 Blocked Filters

detections

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

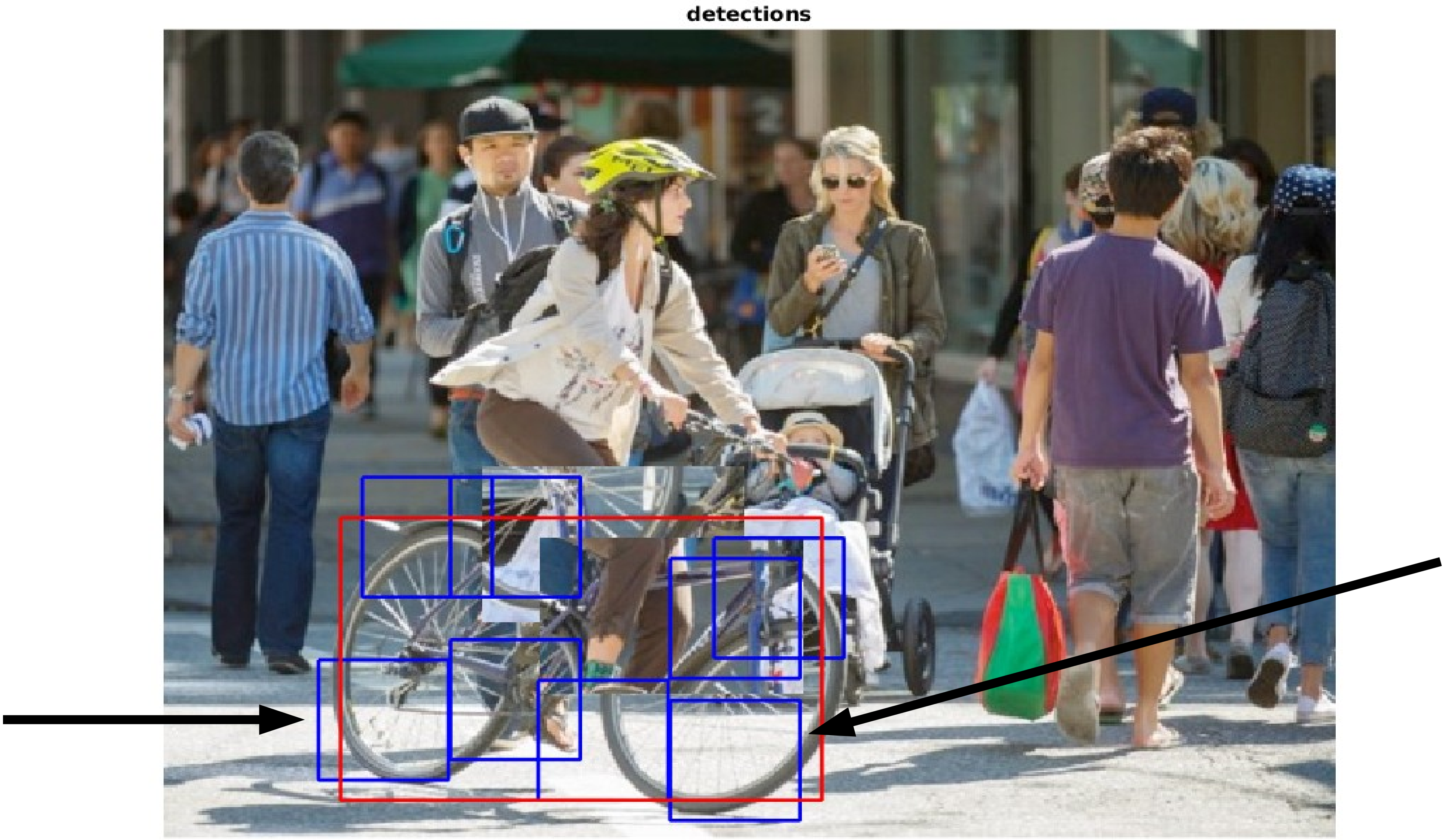
VOC 2010 Bicycle Detector

Ground Truth

No Shifts

One Shift

Static Parts to the Rescue

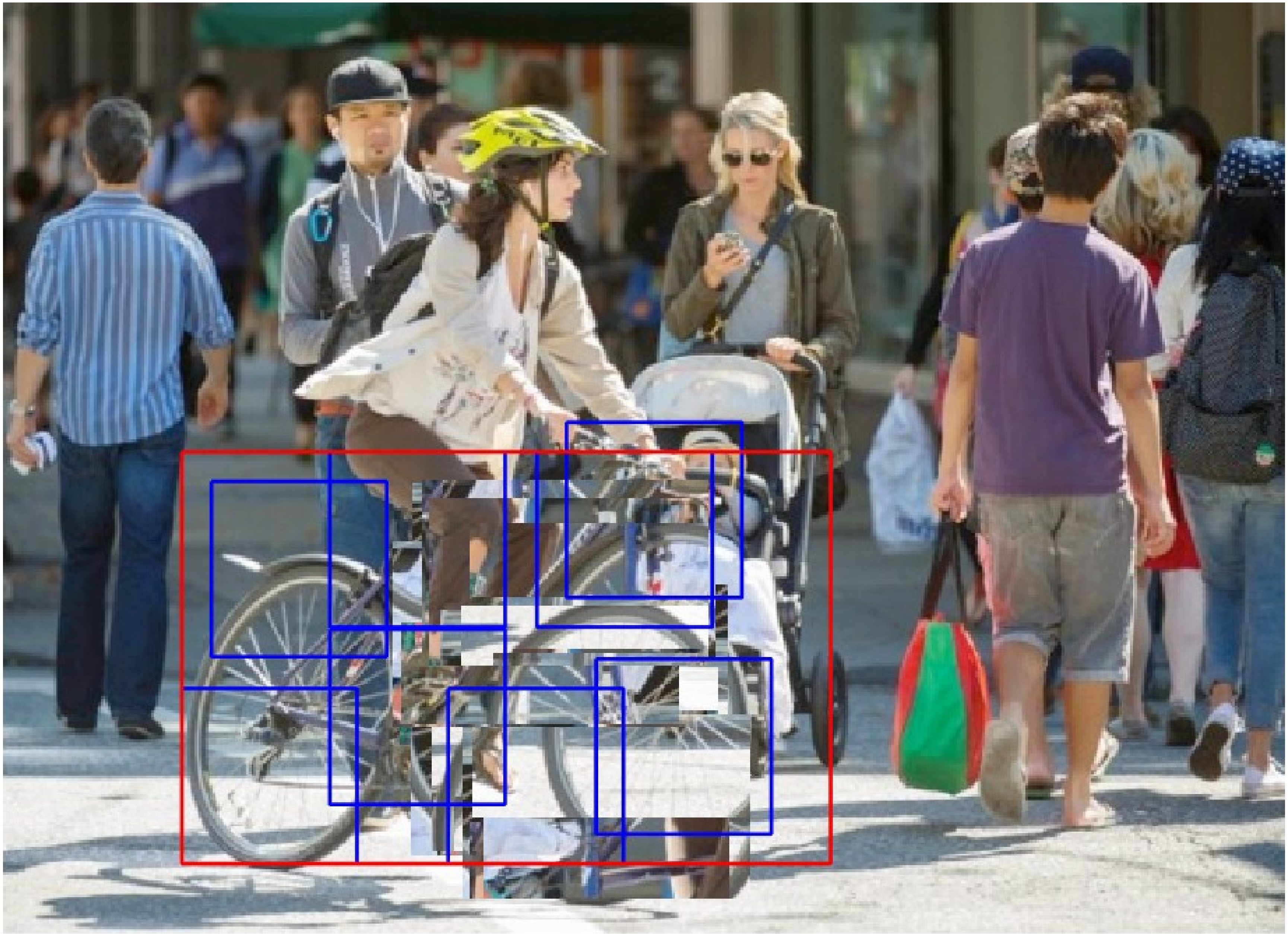


One Shift

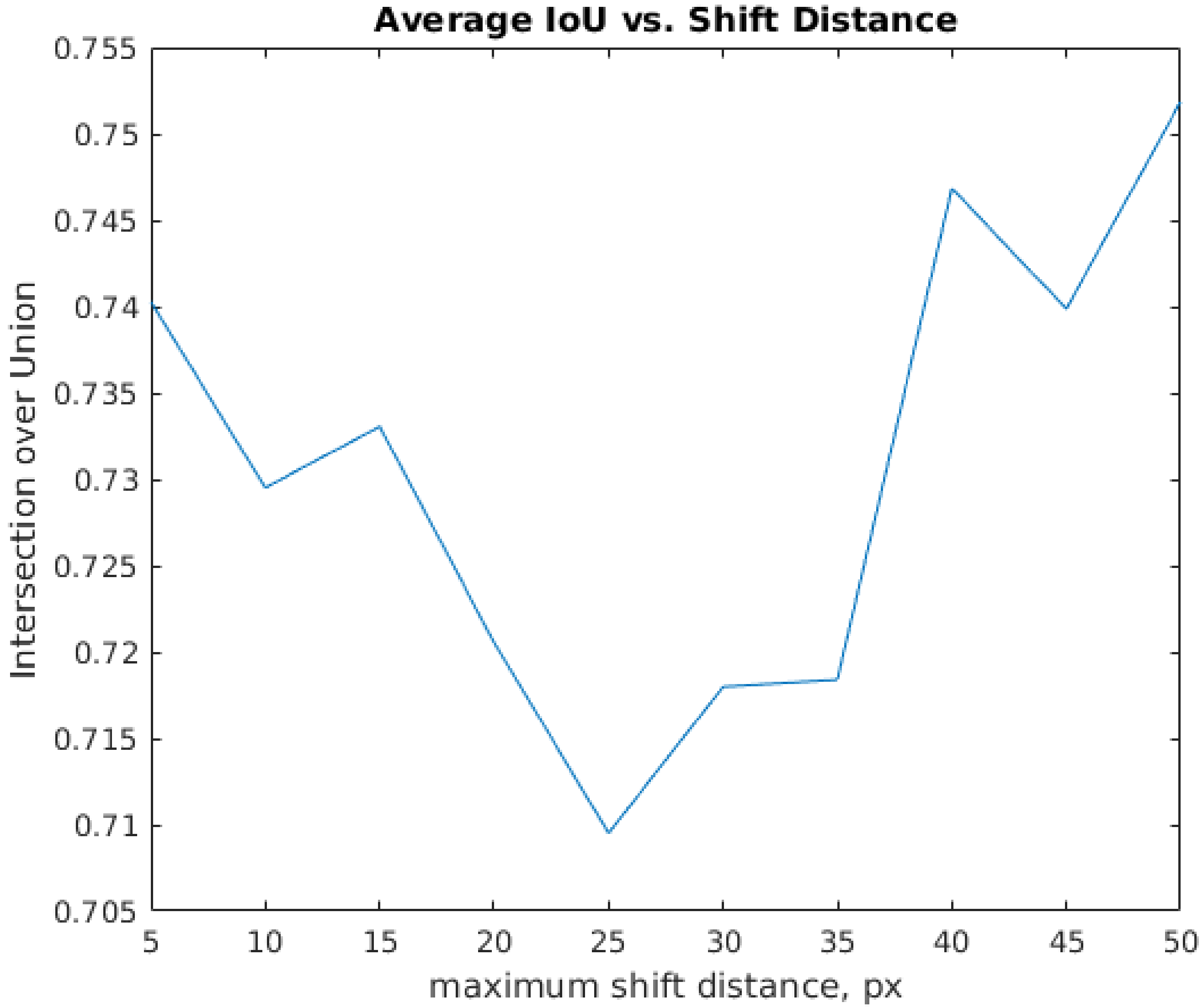
Two Shifts

Three Shifts

Four Shifts



Does how far we shift affect performance?

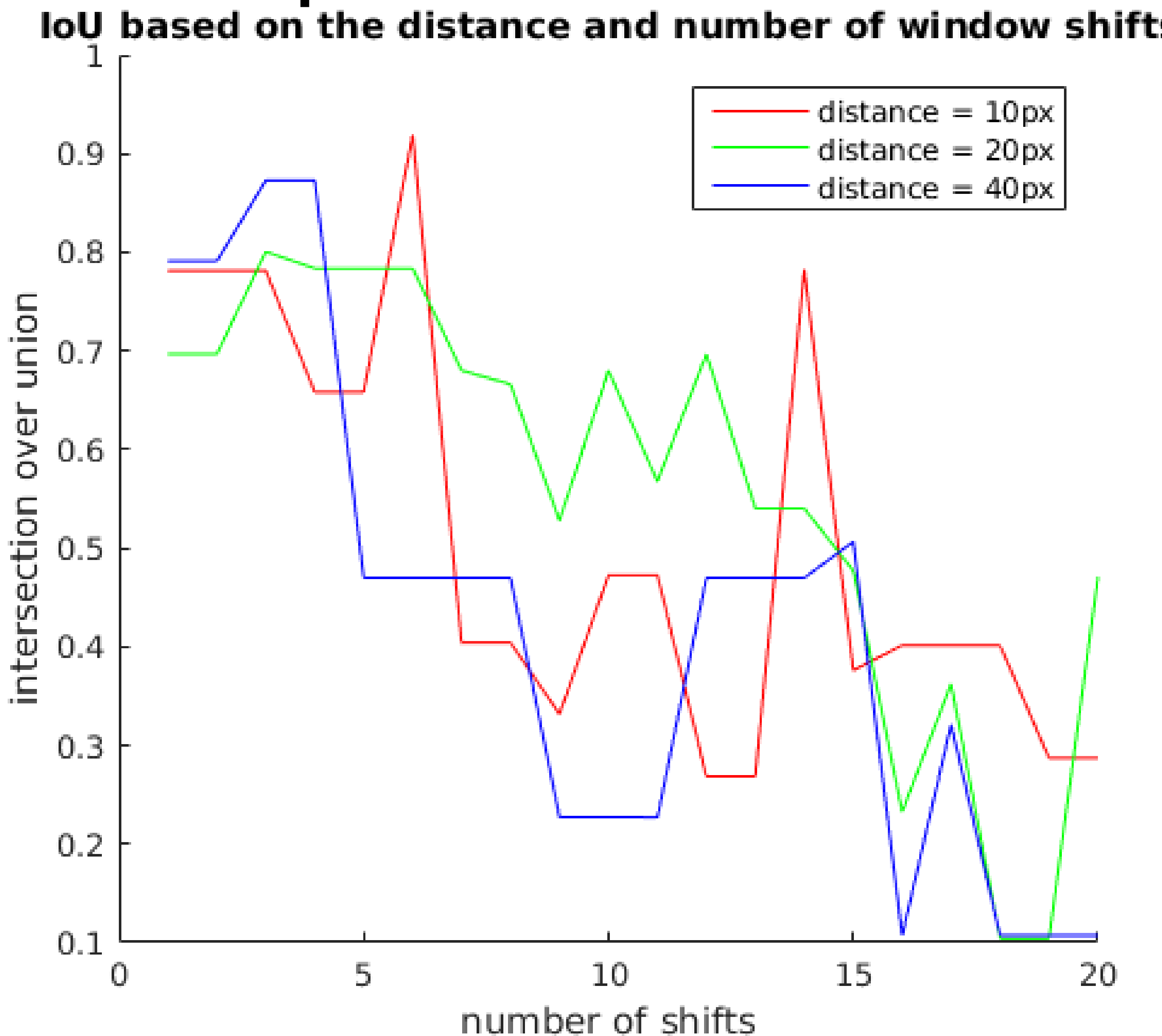


Averaged across 30 trials!

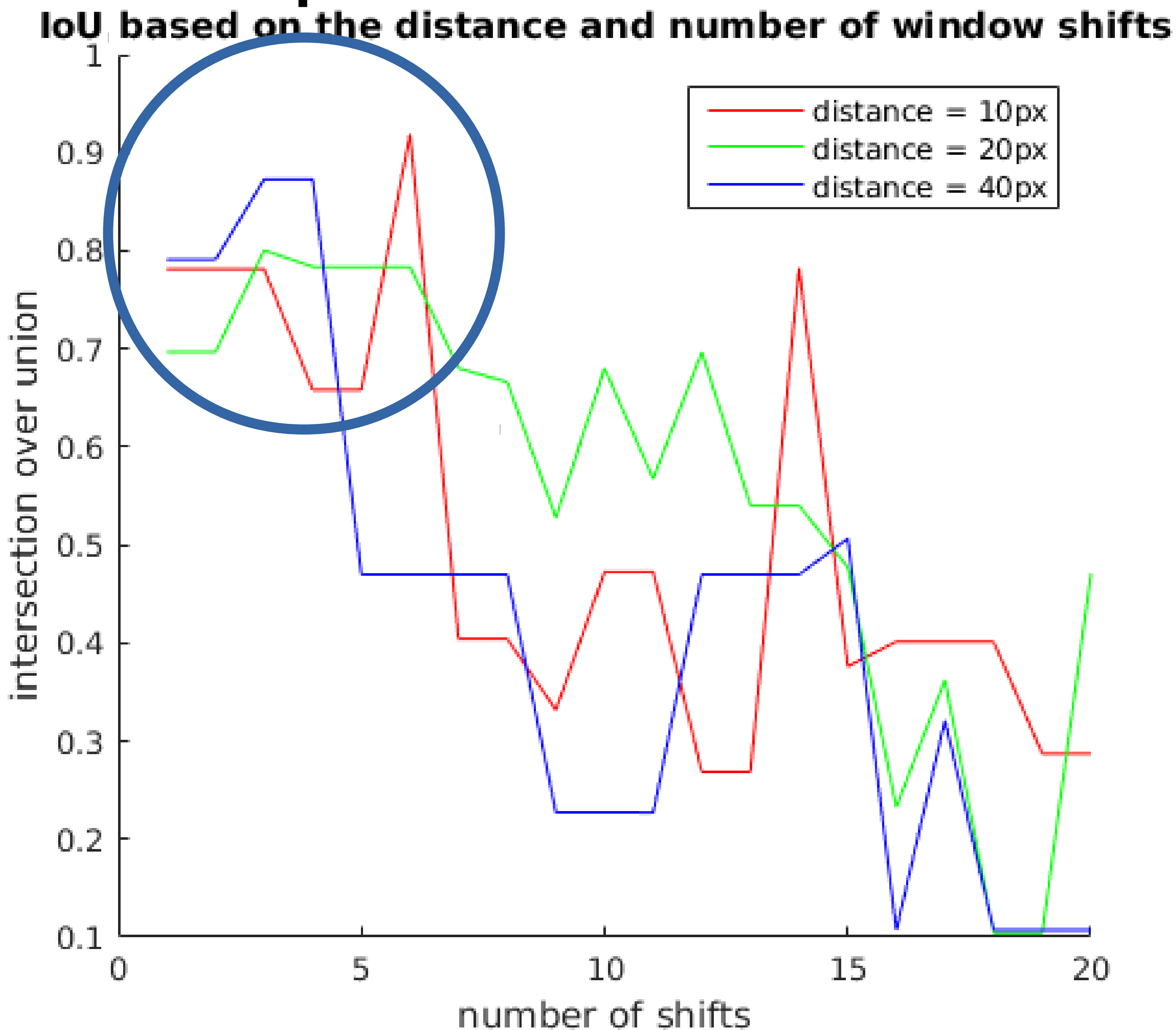
10 10-Pixel Shifts

Ground truth

Does how many times we shift affect performance?



Does how many times we shift affect performance?



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

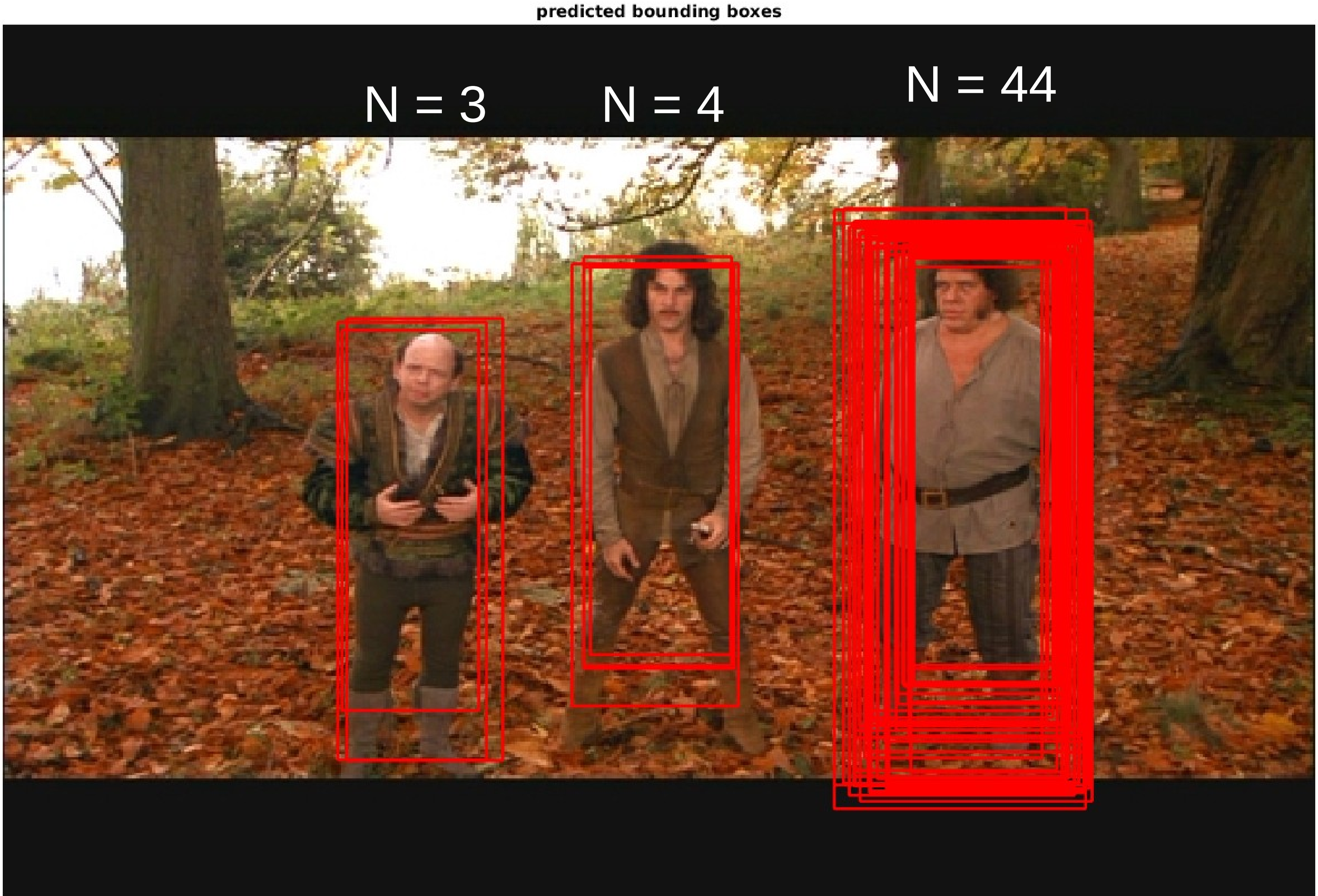
Outline

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

Size-Matched Image

Without NMS, N = 10

Without NMS, N = 50



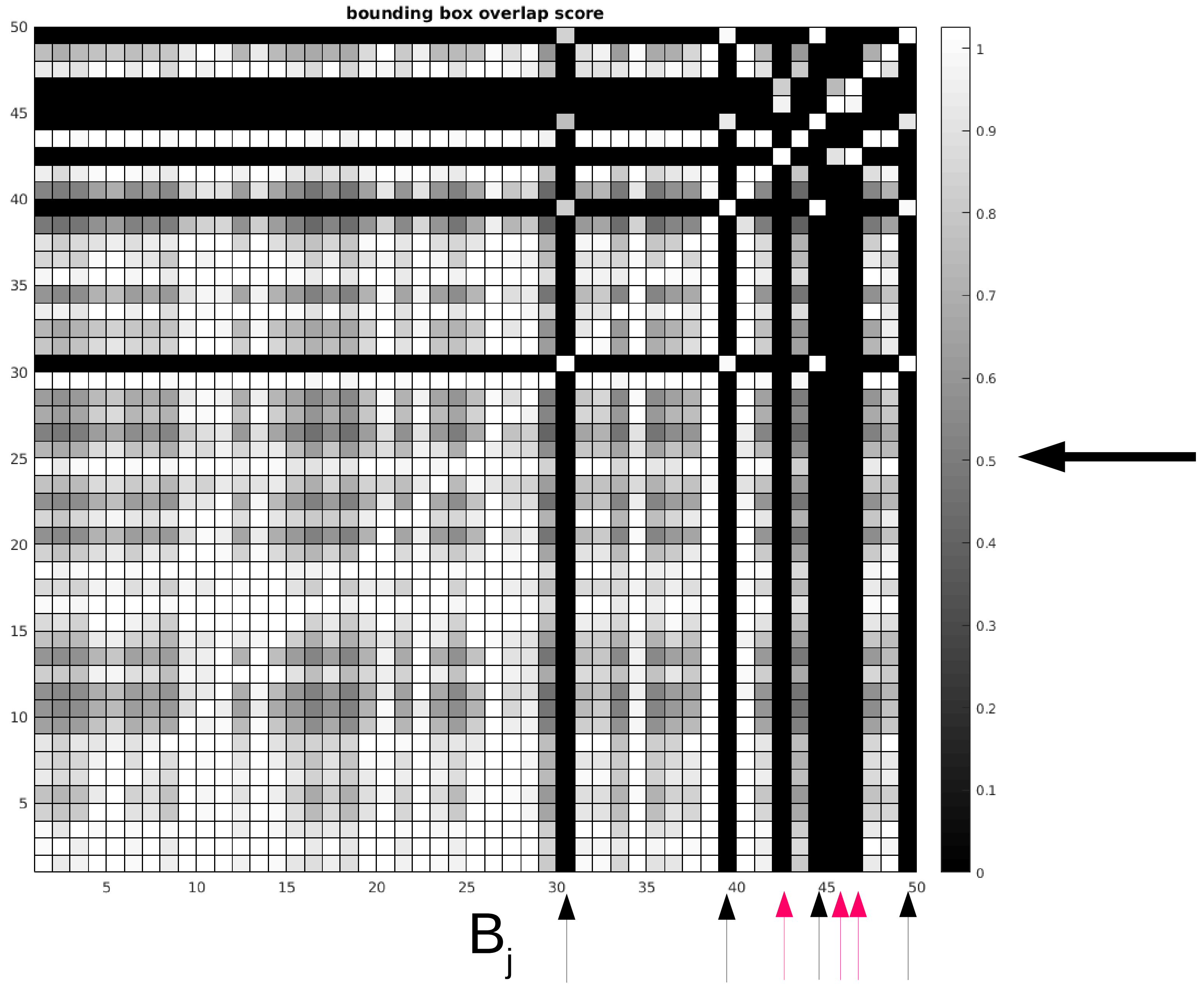
With NMS, N = 3

Overlap = $|B_i \cap B_j| / |B_j|$

Worse matches

B_i

Better matches



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

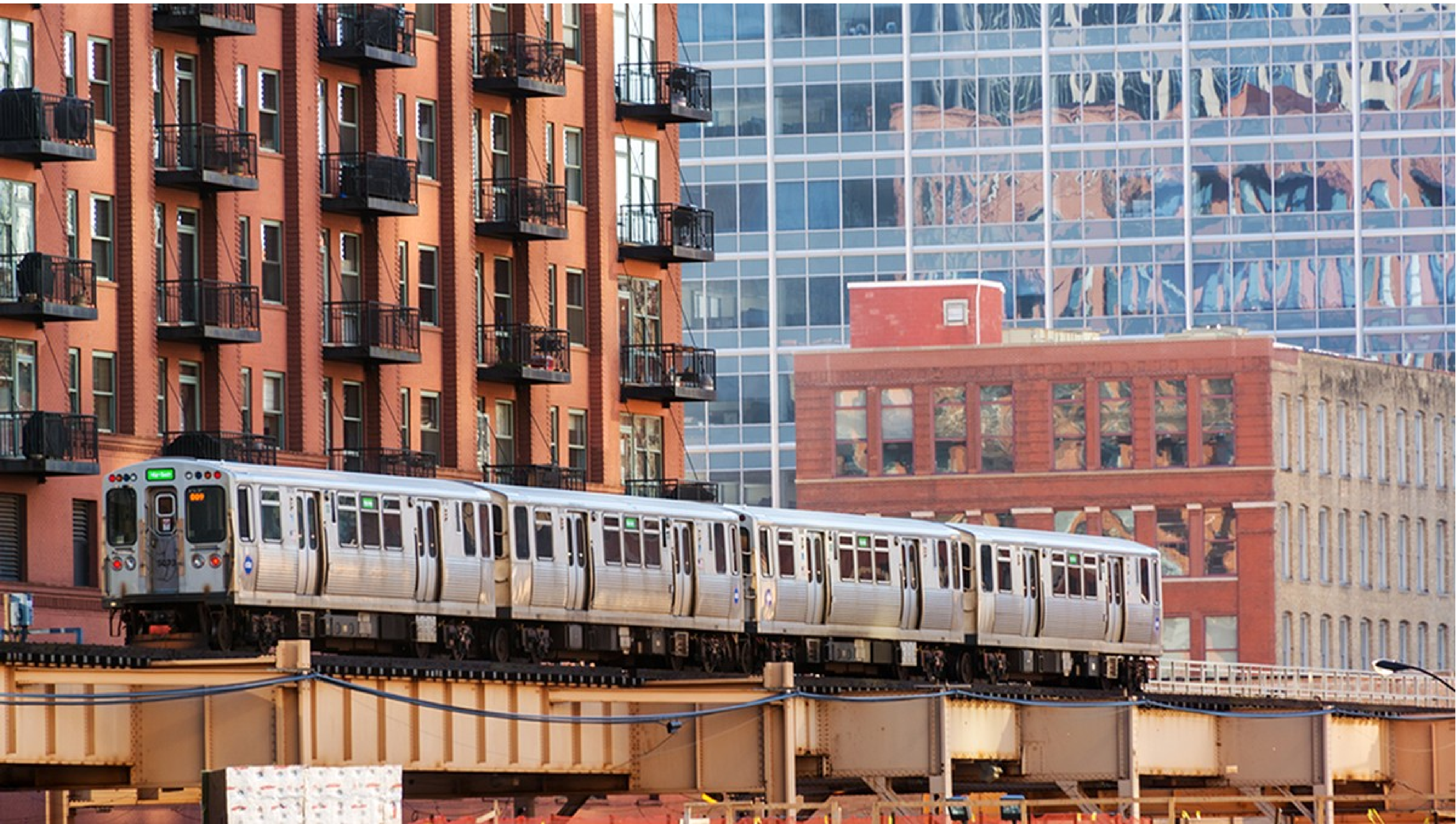
Non-Maximum Supression

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

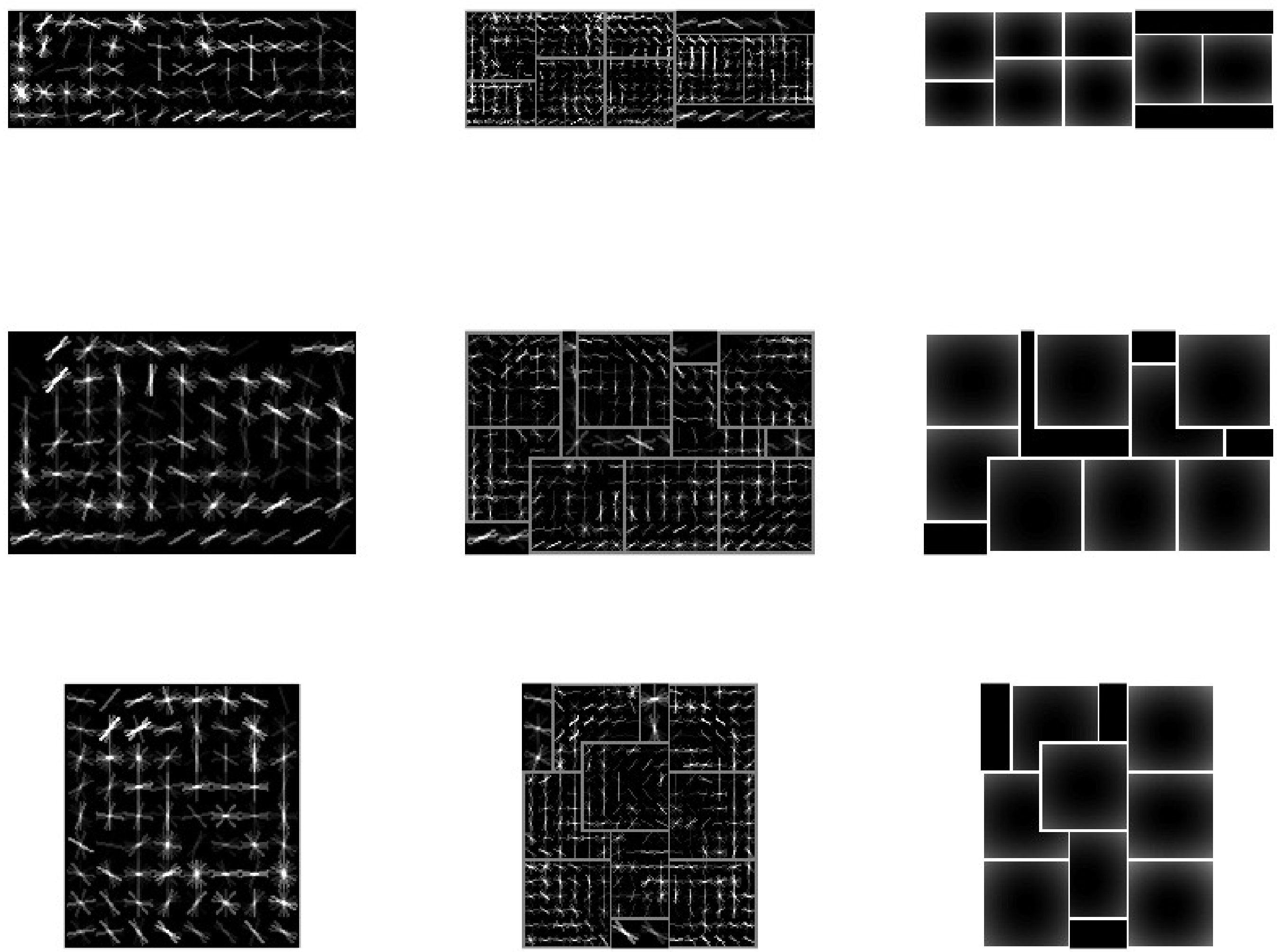
Outline

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

Chicago Elevated Train

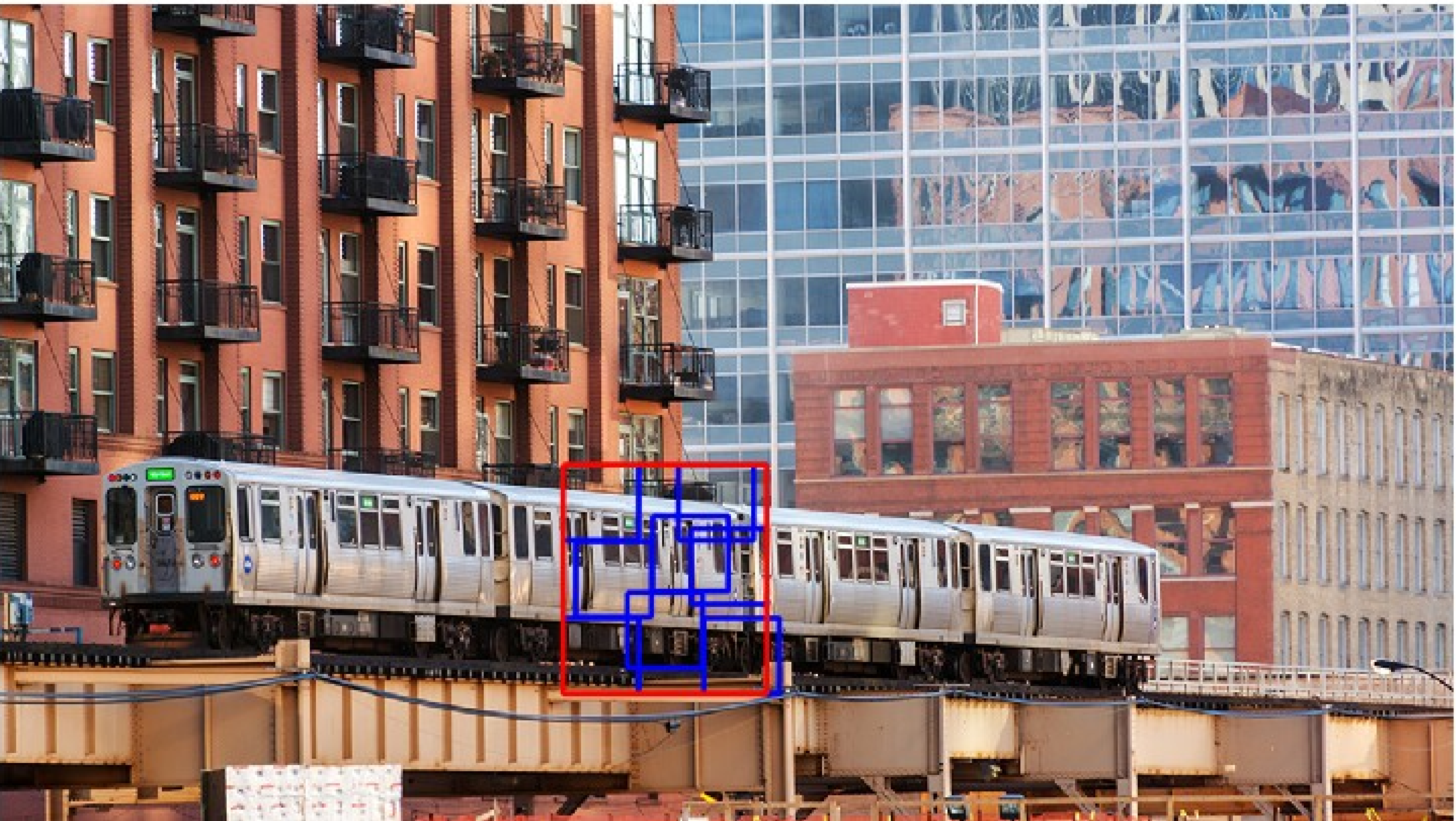


VOC 2007 Train Model

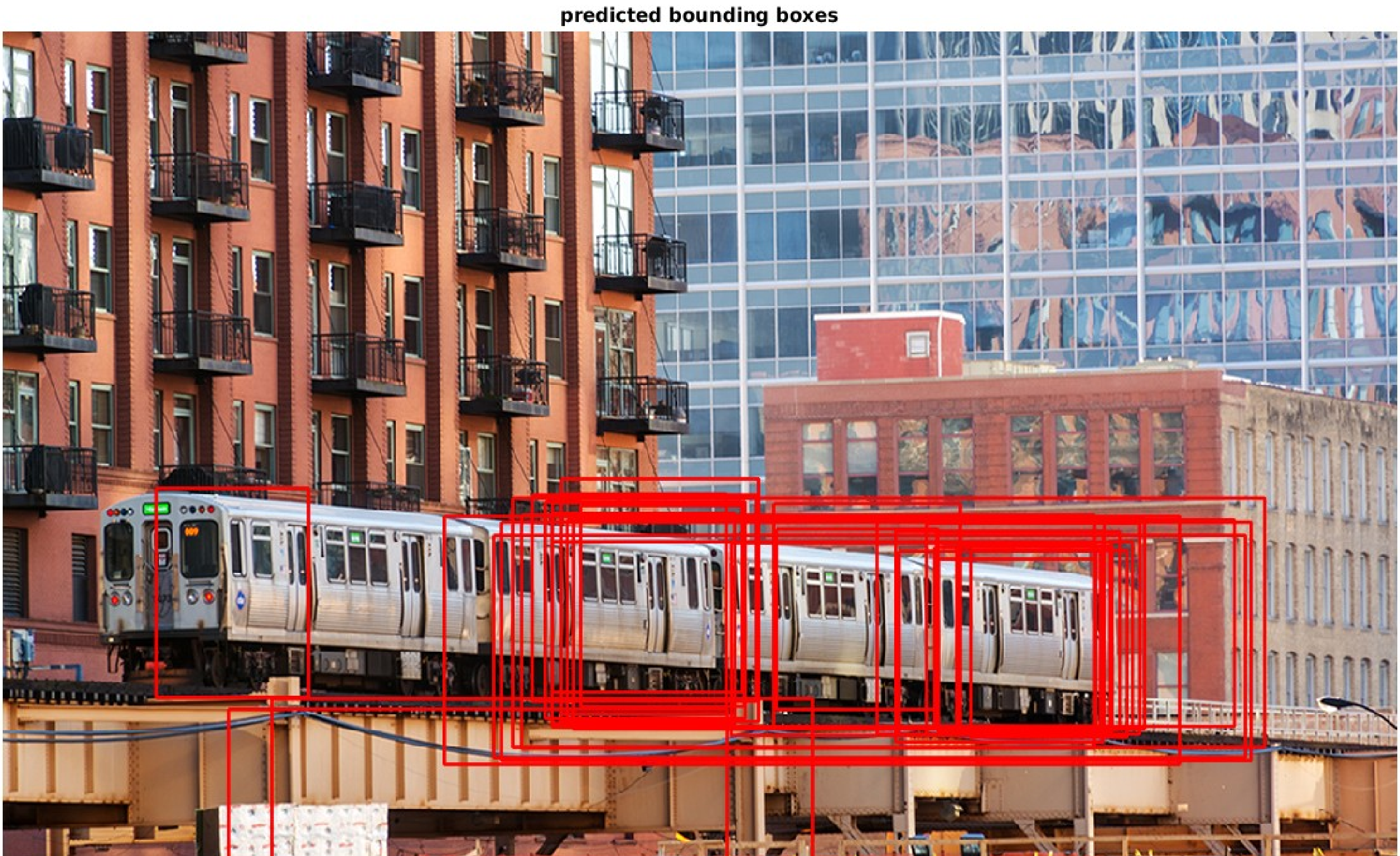


VOC 2007 Train Results, N = 1

detections

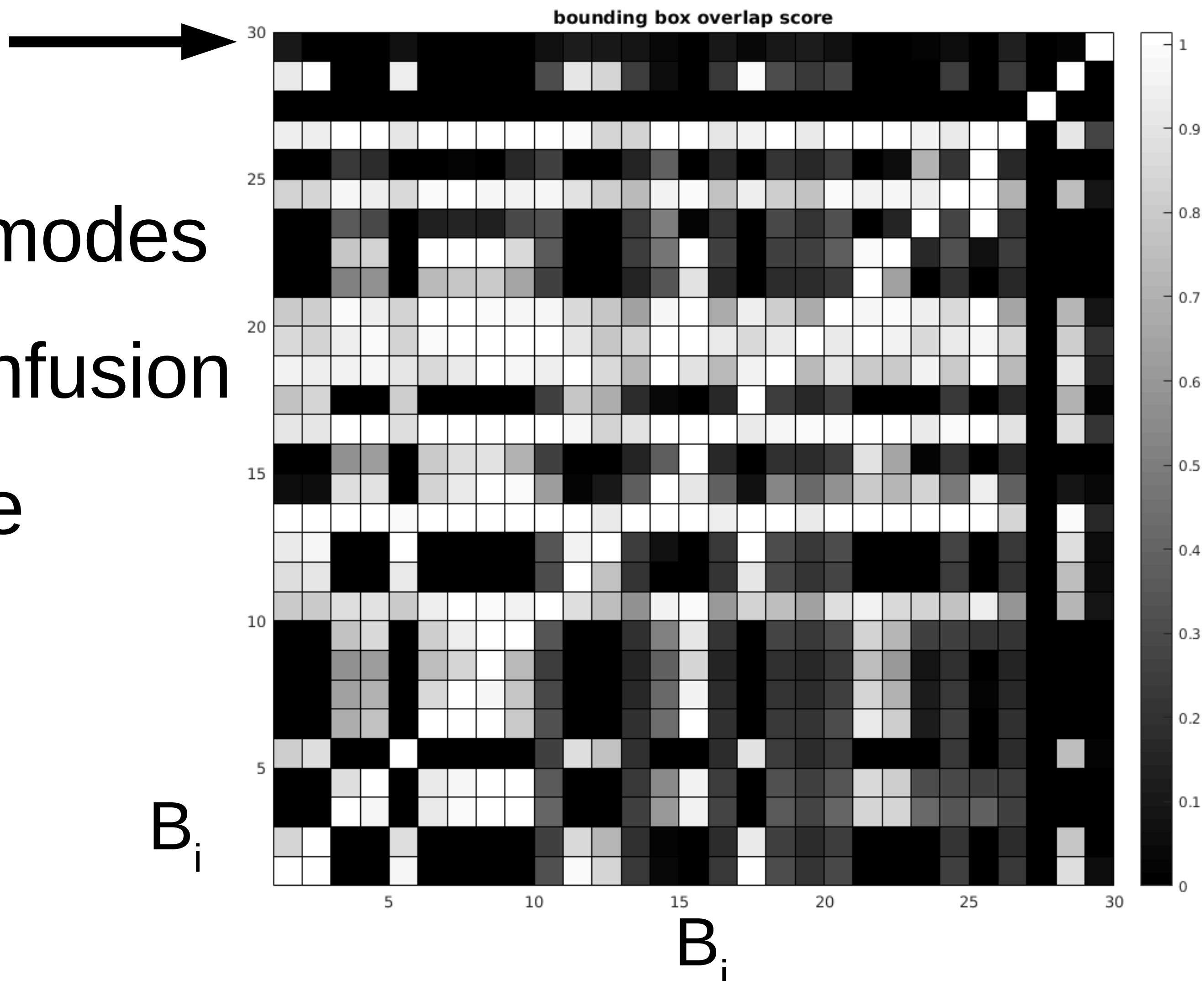


Without NMS, N=30



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

- A Discriminatively Trained, Multiscale, Deformable Part Model. P. Felzenszwalb, D. McAllester, D. Ramanan. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2008
- 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://www.planetizen.com/files/images/ChicagoEl.jpg

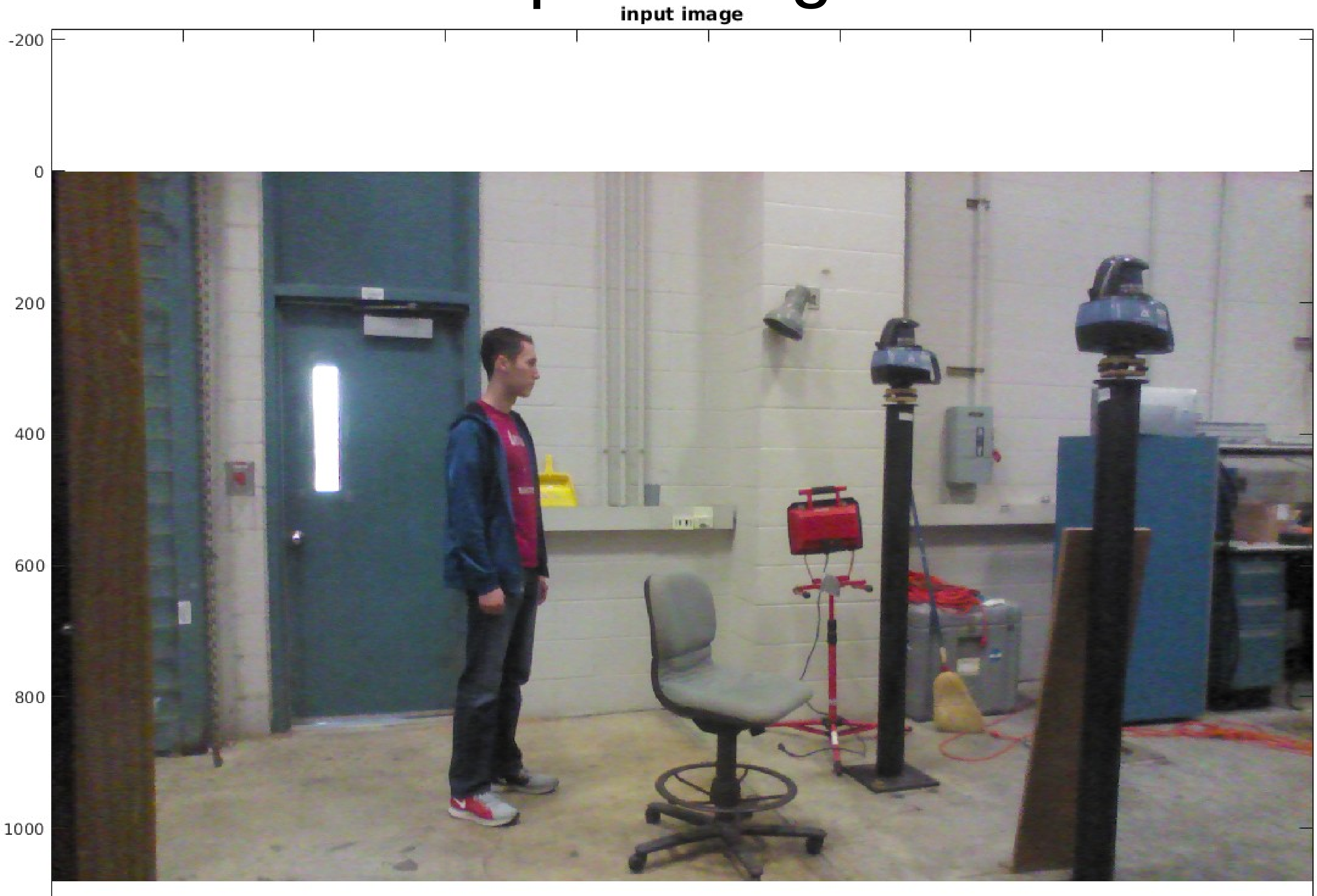
  http://www.brinoideas.xyz/wp-content/uploads/2015/11/open-design-living-room-ideas-with-black-drume-pendant-and-blue-sofa-and-unique-glass-coffee-table-and-lovely-black-white-area-rug-and-grey-cream-pouf-also-big-window.jpg

  http://i.telegraph.co.uk/multimedia/archive/01947/B084FX_1947399c.jpg

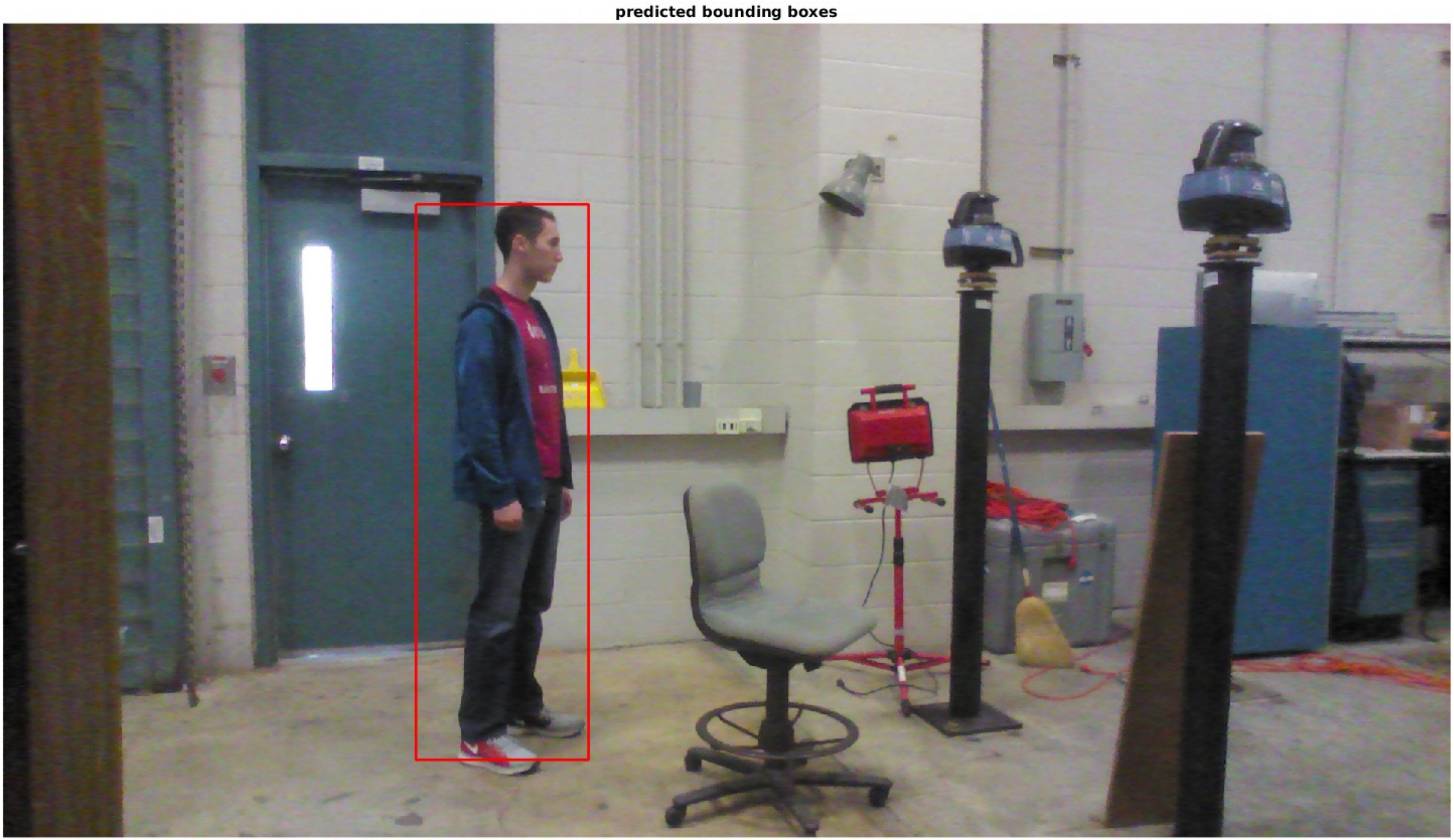
  http://images.glaciermedia.ca/polopoly_fs/1.1346352.1410102588!/fileImage/httpImage/image.jpg_gen/derivatives/landscape_563/10175643-1-jpg.jpg

Live Cam Examples

Input Image

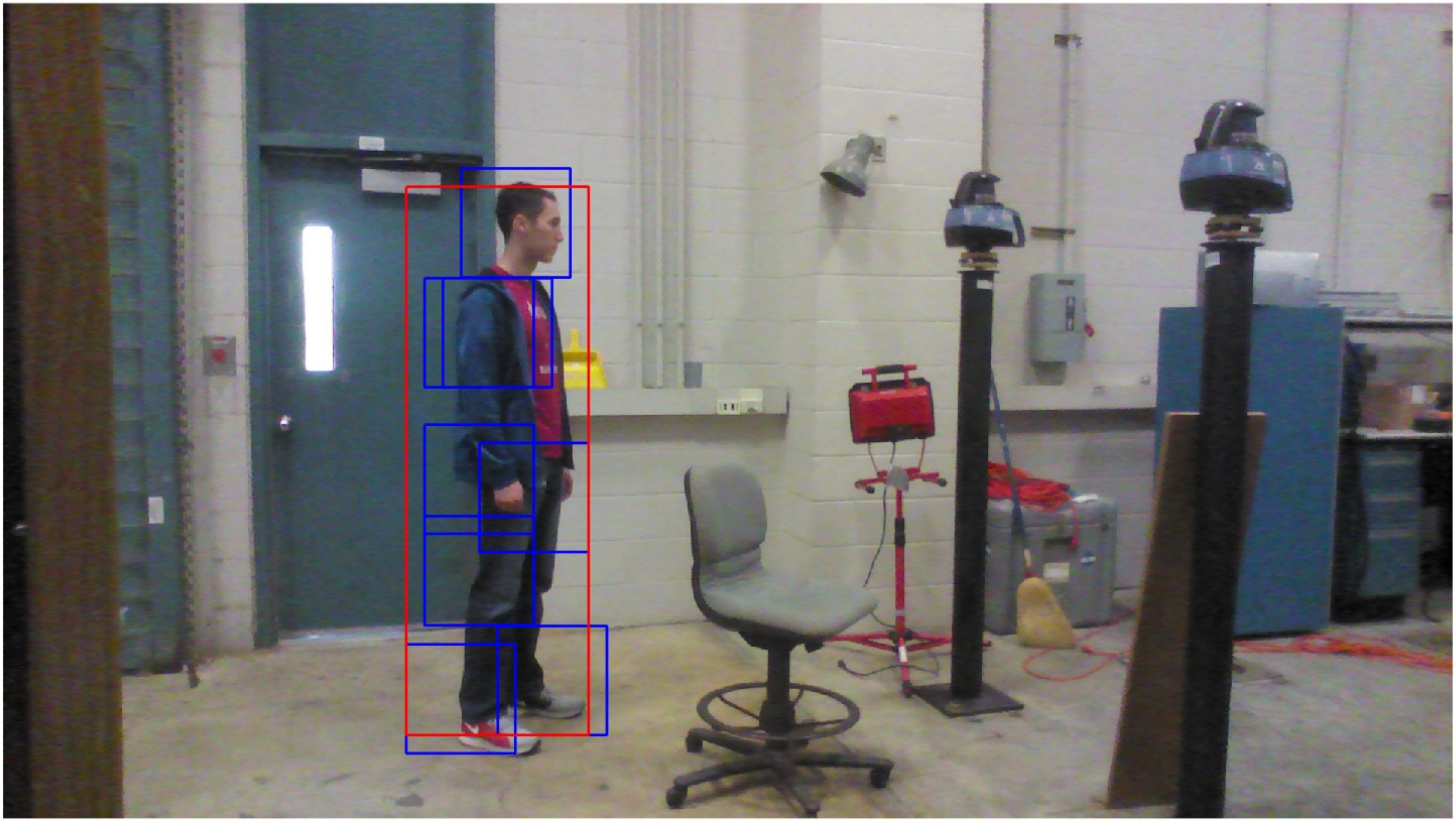


VOC 2010 Person Detector

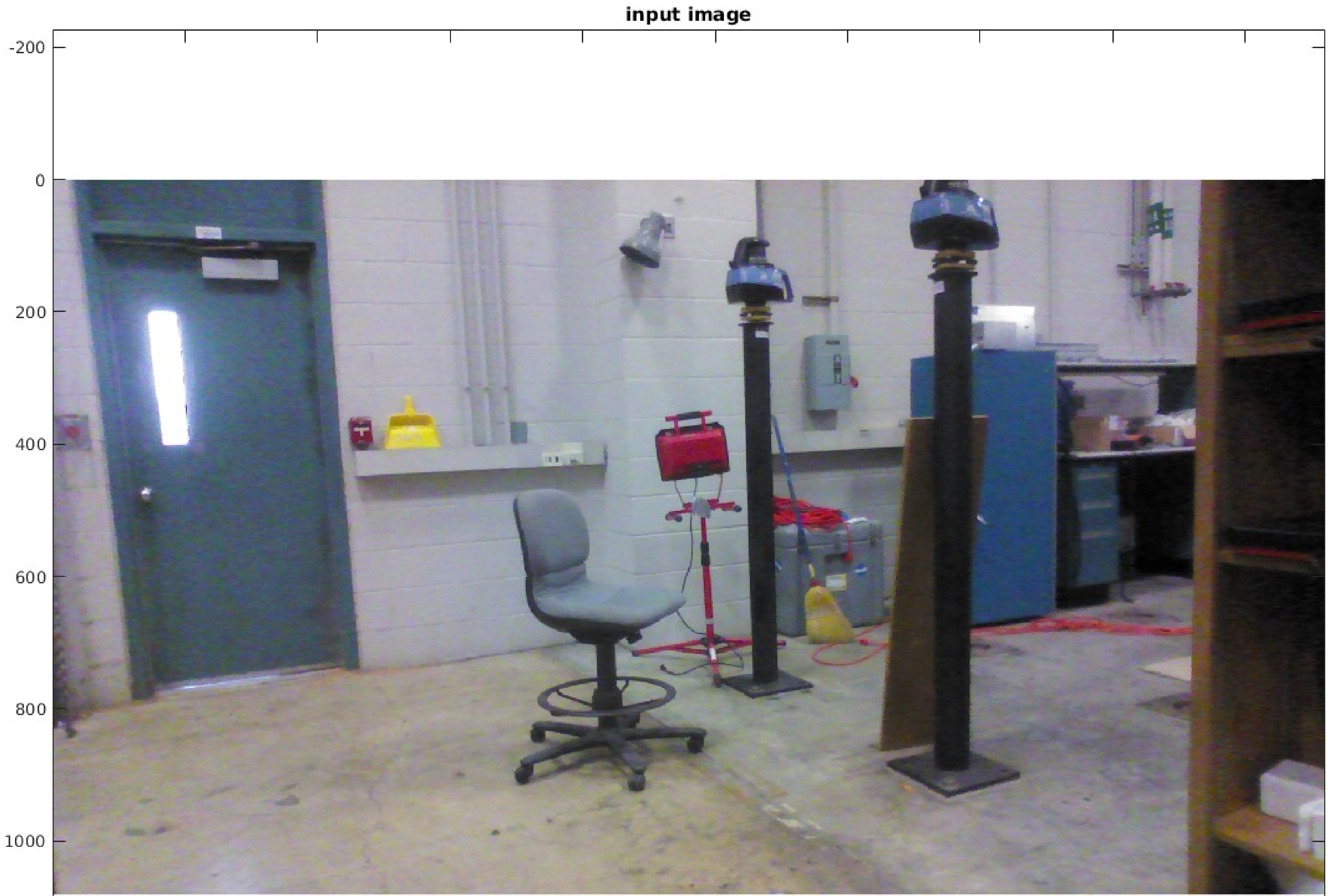


VOC 2010 Person Detector

detections



Chair Detector



Chair Detector

predicted bounding boxes