Object detection with grammar models

Ross Girshick
University of Chicago

Pedro Felzenszwalb
Brown University

David McAllester
TTI Chicago

helmet, occluded left side

ski cap, no face truncated

pirate hat, dresses, long hair

truncation, holding glass, heavy occlusion

Objects from rich categories have diverse structural variation
Object detection with grammar models

Ross Girshick
University of Chicago

Pedro Felzenszwalb
Brown University

David McAllester
TTI Chicago

Dalal & Triggs
CVPR 2005
AP 0.12

Felzenszwalb, McAllester & Ramanan
CVPR 2008
AP 0.27

Felzenszwalb, Girshick, McAllester & Ramanan
PAMI 2010
AP 0.36

Felzenszwalb, Girshick & McAllester
voc-release4
AP 0.42

More mixture components?

There are too many combinations!
Instead...

... compositional models defined by grammars
Object detection with grammar models

Ross Girshick
University of Chicago

Pedro Felzenszwalb
Brown University

David McAllester
TTI Chicago

Localizing people with an object detection grammar

✸ Fine-grained occlusion  ✺ Part sharing
✸ Non-trivial model of the stuff that causes occlusion
✸ Part subtypes  ✺ Subparts at multiple resolutions

AP 0.47
Parts 1-6 (no occlusion)
Parts 1-4 & occluder
Parts 1-2 & occluder
Discriminative training when the label space \( \neq \) output space

Weak-label structural SVM
Generalizes latent structural SVM

Top performance on PASCAL VOC 2010 ‘person’