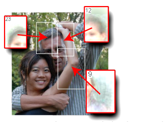
Detecting People Using Mutually Consistent Poselet Activations*

Lubomir Bourdev[1,2], Subhransu Maji[1], Thomas Brox[1] and Jitendra Malik[1]
[1]University of California, Berkeley [2] Adobe Systems, Inc.

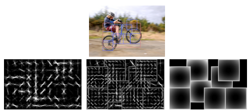
Goals and Contributions

- Best person detection/segmentation on PASCAL VOC 07-09
- New poselet selection algorithm to maximize coverage on the training examples
- Improved detections using neighboring detections of other poselets
- Saliency based agglomerative clustering for generating hypothesis
- Integrating both top down and bottom up information for segmentation
- Large scale 2D annotations done on Amazon Mechanical Turk

Comparison to Felzenszwalb et.al.[1]



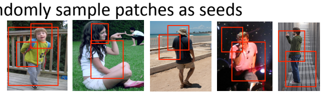
We use ~100 parts that are tightly clustered in pose space. Poselets always have visual meaning ("frontal face", "hand next to hip")



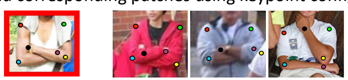
Felzenszwalb et. al. have one root and 6-8 higher-resolution parts trained in an unsupervised manner.

From annotations to poselets

1. Randomly sample patches as seeds



2. Find corresponding patches using keypoint configurations



Query Match 1 Match 2 Weaker Match

3. Train poselets (linear SVMs based on HOG features)

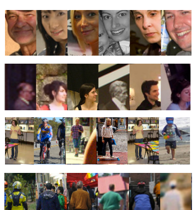
4. Select poselets based on maximizing coverage of the training examples

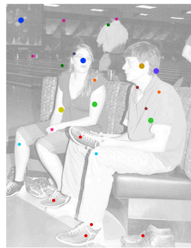
increasing coverage	person 1	person 2	person 3	person 4	person 5
1	X		X		X
2		X		X	
3	X	X			
4					X

Poselet coverage table

Poselet 4 activates on person 5

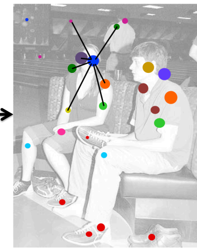
Selected poselets





1. Finding poselet activations

Dots represent centers of poselet activations with size proportional to the detection score

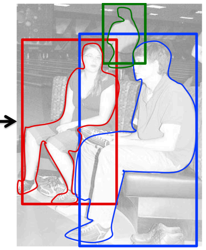


2. Rescoring activations

An example of an activation and its consistent neighbors

3. Clustering activations

Each cluster represents a person hypothesis



4. Hypothesis generation

Example bounds and segmentation

1. Collecting poselet activations

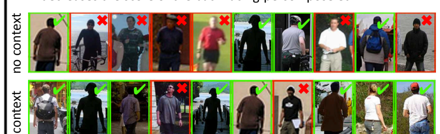
We find all poselet detections above a threshold over multiple scales and locations followed by non-max suppression

2. Improved detections using context

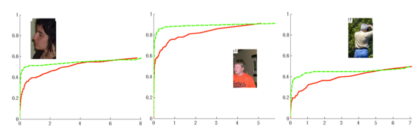
- Local detections can often be wrong:

Front and back look similar | Face false positive | Left or right leg?

- Presence and absence of consistent poselet activations is used to disambiguate activations
- In the picture below the presence of a face poselet activation decreases the score of the back-facing-person poselet



Top 10 activations of the back-facing-person poselet



ROC curves of poselets with (green) and without context (red)

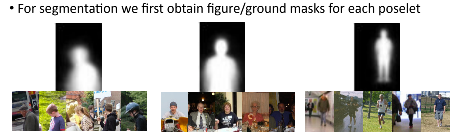
3. Clustering poselet activations

- We cluster consistent poselet activations into person hypotheses
- We use greedy bottom up clustering sorted by the poselet detection scores

Highest prob. activation creates the first cluster | Consistent activation falls in the same cluster | Inconsistent activation creates a new cluster

4. Scoring and Segmenting the hypotheses

- The detection score of the hypothesis is the weighted combination of the scores of the poselet detections
- The predicted bounds are the combination of the bounds predicted by each activation weighted by its detection score
- For segmentation we first obtain figure/ground masks for each poselet



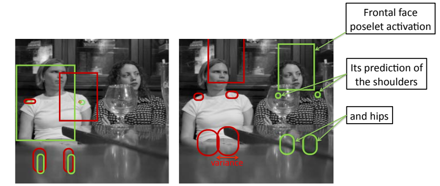
Various poselets with their figure/ground masks

- We obtain an initial segmentation by averaging the masks of individual poselets weighted by their detection scores
- The initial mask g is aligned to the image boundaries f by estimating a smooth deformation field (u,v):

$$E(u,v) = \int_{\mathbb{R}^2} |f(x,y) - g(x+u, y+v)| + \alpha \left(|\nabla u|^2 + |\nabla v|^2\right) dx dy.$$

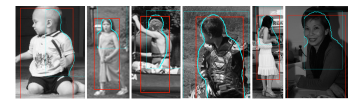
Which poselet activations are consistent?

- Consistent activations refer to the same object
- We measure consistency by thresholding the KL-divergence



Consistent | Not Consistent

Detection/Segmentation Results



Predicted bounding box (red) with the poselet with the highest detection score (cyan)

Predicted segmentation masks

VOC	POSELETS	Felzenszwalb et al.[1]
2009	**47.8%**	43.8%
2008	**54.1%**	43.1%

Person detection

VOC	POSELETS	Yang et al.[2]
2009	**40.5%**	38.9%
2008	**43.1%**	41.3%

Person segmentation

References

1. P. Felzenszwalb, R. Girshick, D. McAllester, D. Ramanan. Object Detection with Discriminatively Trained Part Based Models, PAMI'09
2. Yang, Y., Hallman, S., Ramanan, D., Fowlkes, C.: Layered object detection for multi-class segmentation. CVPR (2010)

* This work was supported by Adobe Systems, Inc., a grant from Hewlett Packard and the MICRO program, a Google Graduate Fellowship, a fellowship from the German Academic Exchange Service (DAAD), as well as ONR MURI N00014-06-1-0734