Object Segmentation by Alignment of Poselet Activations to Image Contours

Thomas Brox[1], Lubomir Bourdev[23], Subhransu Maji[2], Jitendra Malik[2]

[1]University of Freiburg, Germany [2]University of California at Berkeley [3]Adobe Systems Inc.

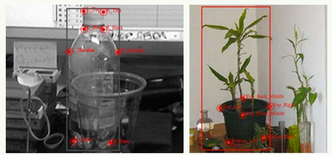
Pose specific part classifiers: Poselets beyond people

Definition of subcategories

boat → sailboat, ocean liner, motorboat
aeroplane → propeller plane, jet, military aircraft
bird → flying bird, non-flying bird

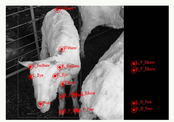
Separate definition of keypoints and separate classifiers for each subcategory

Keypoints for symmetric objects

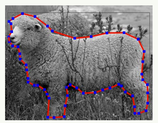


Bottles and potted plants (among others) are rotation symmetric and require a viewpoint dependent definition of keypoints

Annotation with Amazon Mechanical Turk



Keypoint annotation

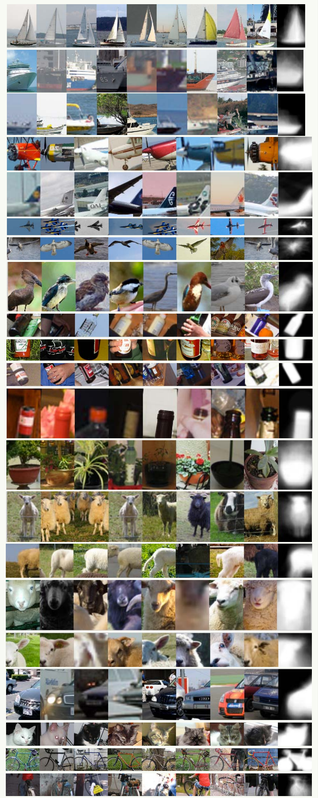


Segmentation

Annotation of the complete PASCAL VOC training set within 2 weeks and for about $3000.

Generation of object hypotheses

Clustering of mutually consistent poselet activations in the same manner as in Bourdev et al. ECCV 2010.



Alignment to image contours

Alignment of each poselet activation to the image contours

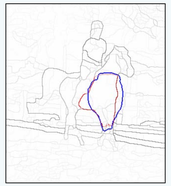
- Extract contour from the poselet's average mask
- Extract image edges with UCM (Arbelaez et al. PAMI 2011)
- Align poselet contour to the image edges with variational optical flow

edges — poselet contour

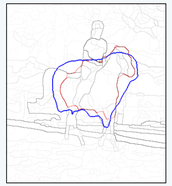
$$E(u,v) = \int_{\mathbb{R}^2} |f(x,y) - g(x+u, y+v)|\, dxdy + \alpha \int_{\mathbb{R}^2} (|\nabla u|^2 + |\nabla v|^2)\, dxdy.$$

Alignment vector field

Input image

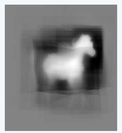


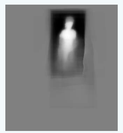
UCM with a poselet activation before and after alignment



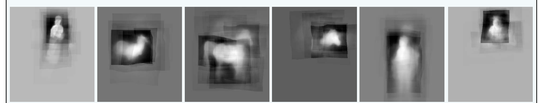
UCM with another poselet activation before and after alignment

Summation of all aligned contours for the two highest ranked hypotheses





Summation of all aligned masks for the two highest ranked hypotheses



6 more hypotheses (out of 20)

L. Bourdev, S. Maji, T. Brox, J.Malik: Detecting people using mutually consistent poselet activations, ECCV 2010.
P. Arbelaez, M. Maire, C. Fowlkes, J. Malik: Contour detection and hierarchical image segmentation, IEEE Trans. on Pattern Analysis and Machine Intelligence, 33(5):898-916, 2011.

Competitive spatial integration

Dealing with overlapping hypotheses

$$M'_j(x,y) = \begin{cases} M_j(x,y), \text{if } M_j(x,y) = \max_k M_k(x,y) \\ M_j(x,y) - \max_k M_k(x,y), \text{otherwise} \end{cases}$$

Winner keeps its score
Winner suppresses all losers

Losers of the winner's category contribute their score to the winner for not losing object evidence in case of erroneous poselet clustering.

Removing false positive hypotheses

$$M''_j(x,y) = \frac{M'_j(x,y)}{\lambda + \max_{x,y} M'_j(x,y)}$$

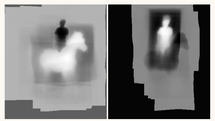
Global normalization of the score

Thresholding after normalization keeps only hypotheses with high scores. This also removes local areas with a low score.

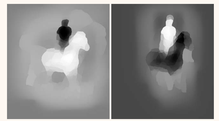
Creating spatially consistent segmentations by joint variational smoothing

$$E(u_1, \ldots, u_K) = \sum_j \int (u_j - M''_j)^2 |M''_j| + \frac{2}{C_j + 1} |\nabla u_j|\, dxdy$$

Object evidence weighted by certainty
Smoothness weighted by object edges



Remaining hypotheses **before** variational smoothing



Remaining hypotheses **after** variational smoothing

Zero level sets

Patch based refinement

Texture similarity defined by 7x7 image patches. Each pixel in a UCM superpixel votes for a label based on the majority label among its 100 nearest neighbors.

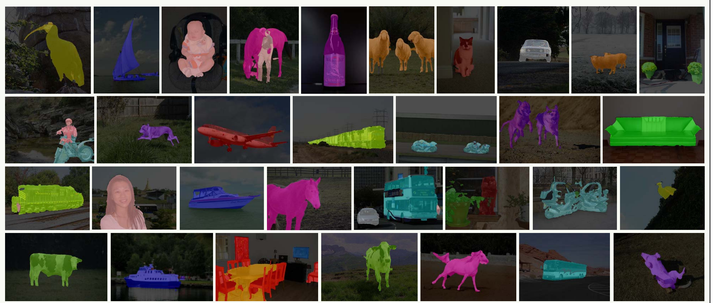
Segmentation results on Pascal datasets

	full model	alignment+ smoothing	alignment	baseline model
background	79.23	78.77	78.76	78.58
aeroplane	36.26	33.25	26.14	26.63
bicycle	38.54	36.02	32.98	32.14
bird	16.57	15.78	13.46	12.70
boat	12.14	12.38	13.18	12.74
bottle	30.40	30.45	32.94	31.40
bus	33.20	32.28	28.43	29.24
car	42.15	41.88	39.84	39.25
cat	44.99	42.87	38.67	38.19
chair	10.33	8.99	8.27	7.89
cow	37.21	34.80	29.77	29.24
diningtable	10.69	9.90	11.61	11.37
dog	23.15	21.64	18.04	17.61
horse	43.92	40.71	36.34	35.41
motorbike	32.59	31.53	28.52	27.90
person	49.64	47.78	44.92	44.00
pottedplant	17.60	18.91	18.12	17.07
sheep	37.38	34.23	27.63	26.68
sofa	9.49	9.22	9.97	9.72
train	23.55	22.63	20.23	20.34
tvmonitor	47.50	47.19	38.87	43.51
average	32.21	31.01	28.41	28.17

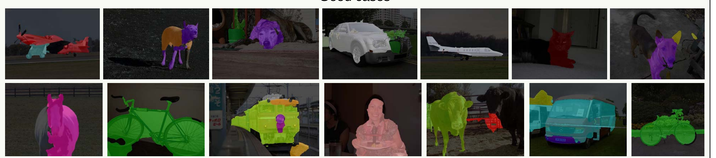
Impact of each component, VOC 2007 dataset

	ours	Barce-lona	Bonn	Chicago	Oxford Brookes
background	82.2	81.1	84.2	80.0	70.1
aeroplane	43.8	58.3	52.5	36.7	31.0
bicycle	23.7	23.1	27.4	23.9	18.8
bird	30.4	39.0	32.3	20.9	19.5
boat	22.2	37.8	34.5	18.8	23.9
bottle	45.7	36.4	47.4	41.0	31.3
bus	56.0	63.2	60.6	62.7	53.5
car	51.9	62.4	54.8	49.0	45.3
cat	30.4	31.9	42.6	21.5	24.4
chair	**9.2**	9.1	9.0	8.3	8.2
cow	27.7	36.8	32.9	21.1	31.0
diningtable	6.9	24.6	25.2	7.0	16.4
dog	29.6	29.4	27.1	16.4	16.4
horse	**42.8**	37.5	32.4	28.2	27.3
motorbike	37.0	60.6	47.1	42.5	48.1
person	**47.1**	44.9	38.3	40.5	31.1
pottedplant	15.1	30.1	36.8	19.6	31.0
sheep	35.1	36.8	50.3	33.6	27.5
sofa	**23.0**	19.4	21.9	13.3	19.8
train	37.7	44.1	35.2	34.1	34.8
tvmonitor	36.5	35.9	40.9	48.5	26.4
average	34.9	40.1	39.7	31.8	30.3

Comparison to state-of-the-art, VOC 2010 dataset



Good cases



Failure cases

This project was funded by the German Academic Exchange Service (DAAD), Adobe Systems Inc., Google Inc., and ONR MURI N00014-06-1-0734