Supplementary File for:
Efficient Region Search for Object Detection

Sudheendra Vijayanarasimhan
University of Texas at Austin
svnaras@cs.utexas.edu

Kristen Grauman
University of Texas at Austin
grauman@cs.utexas.edu

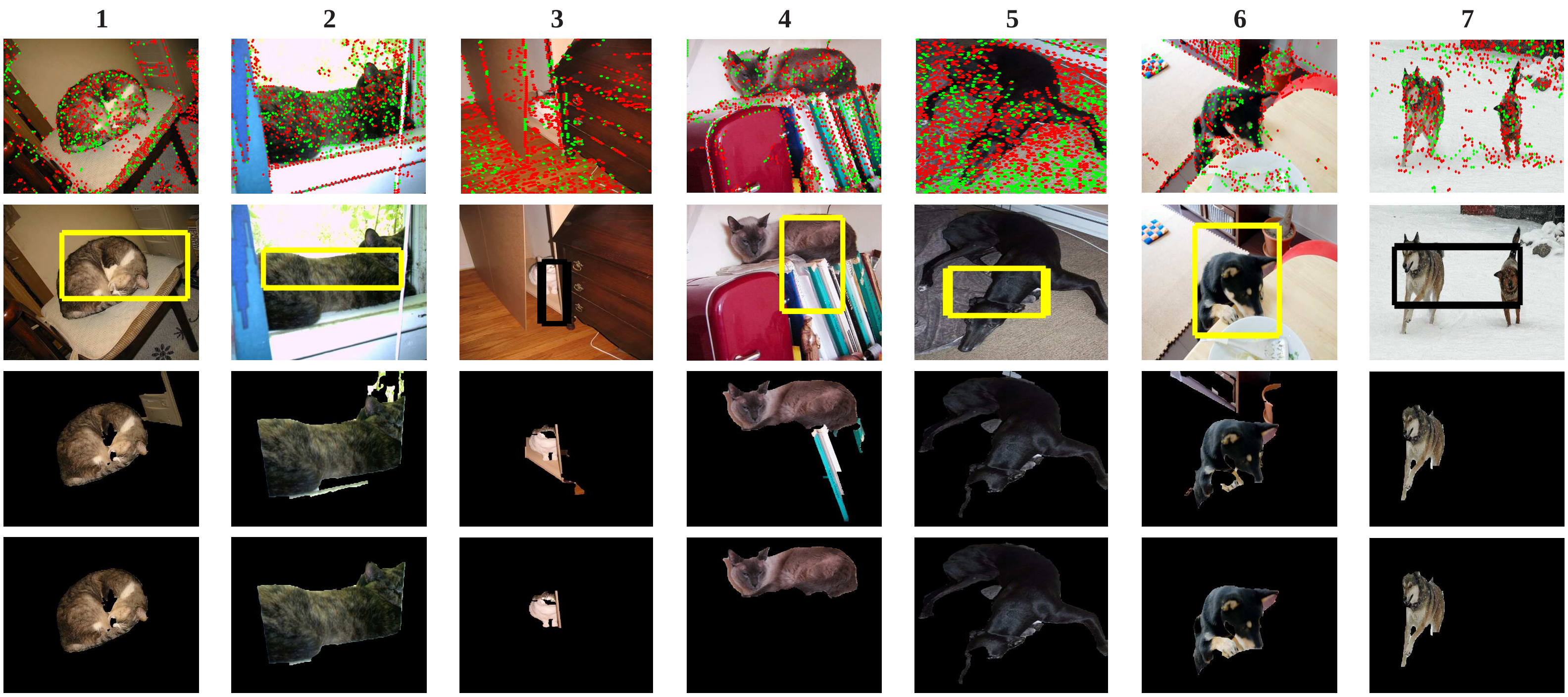

Figure 1. Example PASCAL detections. First row shows images with the sign of the point feature scores (sign$(w_v^{c_i})$) superimposed: red dots denote negatively weighted features, green dots denote positive features (*best viewed in color*). Remaining rows show detections returned by ESS and ERS. Both methods seek the region that will accumulate the most green points while avoiding including excessive red ones. However, since ESS is restricted to finding the max scoring *rectangle*, it often over/underestimates the object's extent. Our method provides precise arbitrarily shaped detections. Last row illustrates how ERS-C can avoid including spurious background regions.

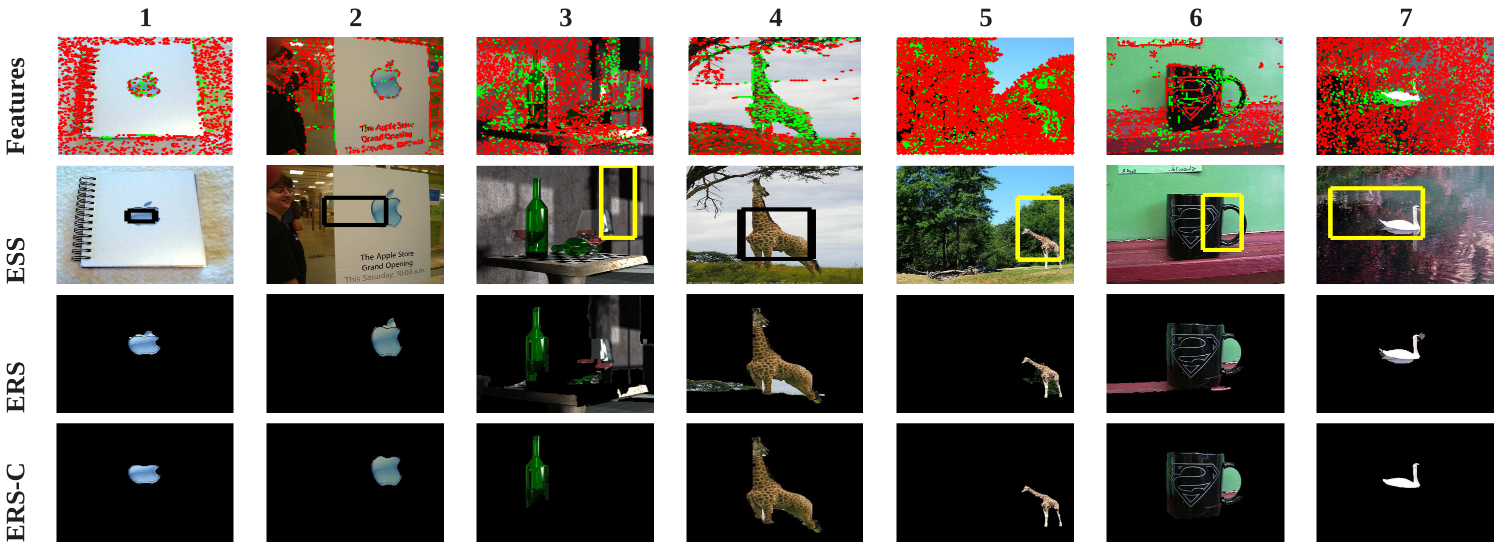

Figure 2. Example detections on the ETHZ images using point features, in the same format as Fig. 4 in the paper. ESS underestimates the object's extent when spurious negatively weighted features appear on the object (see cols 1, 4, 6), while it overestimates due to positively weighted background features near the object (see cols 2, 5, 7). Both variants of our ERS method can model the complex shapes via the region-graph. Further, the intermediate use of regions to sum up feature scores makes the results more coherent. We find that ERS can wrongly include some background regions (see cols 3, 4, 6 in ERS row) when nearby regions have some positively weighted points. However, introducing edge costs based on learned contour weights further excludes those regions that cross strong contour boundaries (see ERS-C row). Best viewed in color.

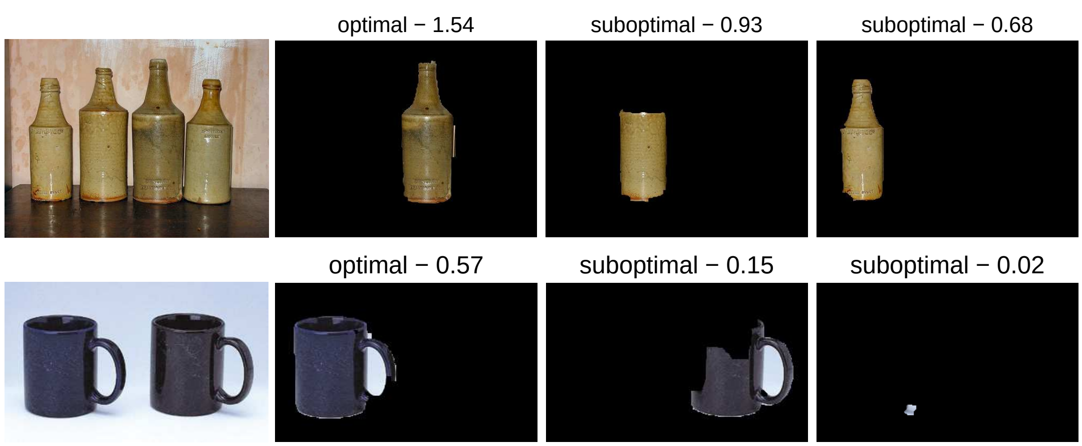

Figure 3. Multiple sub-optimal solutions per test image. The branch-and-cut algorithm can produce multiple sub-optimal solutions in addition to the optimal one, allowing one to detect multiple objects per image or sample among the classifier's most confident regions. Numbers above the images denote $f(R)$ scores, using point features.

Figure 4. Example ERS detections using shape features on the ETHZ shape dataset. Interior white lines show component superpixels selected for the optimal subgraph.

	Average Precision (Point features)					
	Applelogos	Bottles	Giraffes	Mugs	Swans	Avg.
ERS-C	25.5	**41.2**	47.1	**30.4**	37.5	**36.4**
ERS	**27.6**	32.5	**50.0**	29.3	**41.4**	36.2
ESS	25.4	38.0	26.3	27.9	24.2	28.3

	Mean Overlap (Point features)					
	Applelogos	Bottles	Giraffes	Mugs	Swans	Avg.
ERS-C	46.8	**55.4**	**48.9**	**43.4**	**48.2**	**48.5**
ERS	**47.2**	32.9	38.5	41.3	42.8	40.5
ESS	26.3	35.6	22.2	31.2	25.9	28.3

Figure 5. These tables compare the mAP and overlap scores of the three approaches (ERS and ERS-C are ours, ESS is the baseline) using the same set of point features on the ETHZ dataset. Left: average precision scoring. Right: mean overlap scoring. The mAP of ERS is from 9% to 90% better than ESS. Our ERS and ERS-C variants yield fairly similar average precision, but ERS-C produces higher mean overlaps (right table), avoiding more regions that stretch across object boundaries.