Supplementary file for: Collect-Cut: Segmentation with Top-Down Cues Discovered in Multi-Object Images

Yong Jae Lee and Kristen Grauman

This document contains additional supporting results, as referenced in the main paper.

1. Results

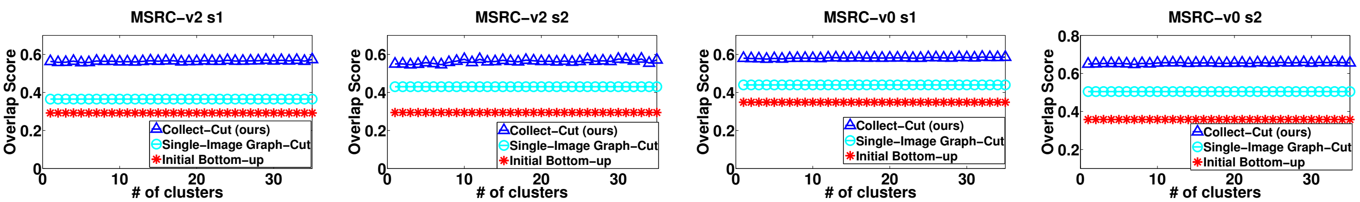
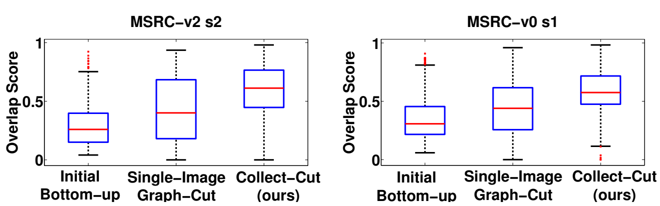
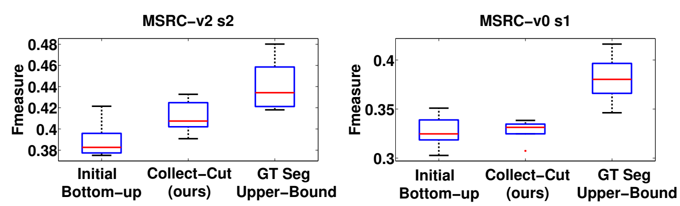


Figure 1. Mean segmentation overlap scores as a function of k. Higher values are better. We compare our method's results with the initial bottom-up multiple segmentations and the single-image graph cuts baseline. Our method produces the best segmentations.

We first show segmentation results as a function of the number of clusters k (see Figure 1). We compare against the initial bottom-up multiple segmentations baseline and the single-image graph cuts baseline. For each k, we compute a single score by averaging the scores of all unknown instances. Our results are stable over k (they improve slightly as k increases, i.e., as the groups get more coherent), and are significantly better than either baseline. The baseline results are constant over k, since their segmentations are independent of the number of clusters. These results confirm the advantage of discovering shared structure in the unlabeled set of images when computing segmentations.



(a) Segmentation Overlap Scores



(b) Discovery Accuracy

Figure 2. (a) Segmentation overlap scores. Higher values are better—a score of 1 would mean 100% pixel-for-pixel agreement with ground truth object segmentation. We compare our method's results (right box-plots) with the initial bottom-up multiple segmentations (left box-plots), and the single-image graph cuts baseline (middle box-plots). (b) Impact of collective segmentation on discovery accuracy, as evaluated by the F-measure (higher values are better). We compare against the initial bottom-up multiple segmentations, and with ground truth object segments, which provides an upper bound on accuracy.

We next show additional segmentation and discovery results on four more splits which we could not fit in the main paper. These are the same experiments as described in Sections 4.1 and 4.2, and the results show the same outcome as those shown in Figures 3 and 5 in the main paper.

By leveraging the shared structure in the *collection* of images, our method produces significantly better segmentations than the bottom-up segmentation baseline and the single-image graph-cuts baseline (see Figure 2 (a)). Our method also yields a significant gain in clustering accuracy over the initial bottom-up segmentations (see Figure 2 (b)). This can be attributed to the fact that the spatial extent of the refined regions more closely matches that of the true objects, thereby allowing more complete appearance features to be extracted per region, and then clustered.