FusionSeg: Learning to combine motion and appearance for fully automatic segmentation of generic objects in videos

Supplementary material

Suyog Dutt Jain*  Bo Xiong*  Kristen Grauman
University of Texas at Austin
suyog@cs.utexas.edu, bxiong@cs.utexas.edu, grauman@cs.utexas.edu
http://vision.cs.utexas.edu/projects/fusionseg/

This document supplements the main paper with per video results for the DAVIS and Segtrack-v2 datasets (referred in Table 1 and Table 3 from the main paper).

1. Per-video results for DAVIS and Segtrack-v2:

Table 1 shows the per video results for the 50 videos from the DAVIS dataset (referred in Table 1 of the main paper). We compare with several semi-supervised and fully automatic baselines. Our method outperforms the per-video best fully automatic and best semi-supervised baseline in 25 out of 50 videos.

Table 2 shows the per video results for the 14 videos from the Segtrack-v2 dataset (referred in Table 3 of the main paper). Our method outperforms the per-video best fully automatic method in 5 out of 14 cases. Our method also outperforms the semi-supervised HVS [1] method in 8 out of 14 cases.

*Both authors contributed equally to this work
Table 1: Video object segmentation results on DAVIS dataset. We show the results for all 50 videos. Table 1 in the main paper summarizes these results over all 50 videos. Our method outperforms several state-of-the-art methods, including the ones which actually require human annotation during segmentation. The best performing methods grouped by whether they require human-in-the-loop or not during segmentation are highlighted in bold. Metric: Jaccard score, higher is better.
Table 2: Video object segmentation results on Segtrack-v2. We show the results for all 14 videos. Table 3 in the main paper summarizes these results over all 14 videos. Our method outperforms several state-of-the-art methods, including the ones which actually require human annotation during segmentation. For NLC results are averaged over 12 videos as reported in their paper [4]. The best performing methods grouped by whether they require human-in-the-loop or not during segmentation are highlighted in bold. Metric: Jaccard score, higher is better.
References


