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

DAVIS: Densely Annotated Video Segmentation dataset (50 videos)									
Methods	FST [2]	KEY [3]	NLC [4]	HVS [1]	FCP [5]	BVS [6]	Ours-A	Ours-M	Ours-Joint
Human in loop?	No	No	No	Yes	Yes	Yes	No	No	No
Bear	89.8	89.1	90.7	93.8	90.6	95.5	91.52	86.30	90.66
Blackswan	73.2	84.2	87.5	91.6	90.8	94.3	89.54	61.71	81.10
Bmx-Bumps	24.1	30.9	63.5	42.8	30	43.4	38.77	26.42	32.97
Bmx-Trees	18	19.3	21.2	17.9	24.8	38.2	34.67	37.08	43.54
Boat	36.1	6.5	0.7	78.2	61.3	64.4	63.80	59.53	66.35
Breakdance	46.7	54.9	67.3	55	56.7	50	14.22	61.80	51.10
Breakdance-Flare	61.6	55.9	80.4	49.9	72.3	72.7	54.87	62.09	76.21
Bus	82.5	78.5	62.9	80.9	83.2	86.3	80.38	77.70	82.70
Camel	56.2	57.9	76.8	87.6	73.4	66.9	76.39	74.19	83.56
Car-Roundabout	80.8	64	50.9	77.7	71.7	85.1	74.84	84.75	90.15
Car-Shadow	69.8	58.9	64.5	69.9	72.3	57.8	88.38	81.03	89.61
Car-Turn	85.1	80.6	83.3	81	72.4	84.4	90.67	83.92	90.23
Cows	79.1	33.7	88.3	77.9	81.2	89.5	87.96	82.22	86.82
Dance-Jump	59.8	74.8	71.8	68	52.2	74.5	10.32	64.22	61.16
Dance-Twirl	45.3	38	34.7	31.8	47.1	49.2	46.23	55.39	70.42
Dog	70.8	69.2	80.9	72.2	77.4	72.3	90.41	81.90	88.92
Dog-Agility	28	13.2	65.2	45.7	45.3	34.5	68.94	67.88	73.36
Drift-Chicane	66.7	18.8	32.4	33.1	45.7	3.3	46.13	44.14	59.86
Drift-Straight	68.3	19.4	47.3	29.5	66.8	40.2	67.24	69.08	81.06
Drift-Turn	53.3	25.5	15.4	27.6	60.6	29.9	85.09	72.09	86.30
Elephant	82.4	67.5	51.8	74.2	65.5	85	86.18	77.51	84.35
Flamingo	81.7	69.2	53.9	81.1	71.7	88.1	44.46	63.80	75.67
Goat	55.4	70.5	1	58	67.7	66.1	84.11	74.99	83.09
Hike	88.9	89.5	91.8	87.7	87.4	75.5	82.54	58.30	76.90
Hockey	46.7	51.5	81	69.8	64.7	82.9	66.03	44.89	70.05
Horsejump-High	57.8	37	83.4	76.5	67.6	80.1	71.09	54.10	64.93
Horsejump-Low	52.6	63	65.1	55.1	60.7	60.1	70.23	55.20	71.20
Kite-Surf	27.2	58.5	45.3	40.5	57.7	42.5	47.71	18.54	38.98
Kite-Walk	64.9	19.7	81.3	76.5	68.2	87	52.65	39.35	49.00
Libby	50.7	61.1	63.5	55.3	31.6	77.6	67.70	35.34	58.48
Lucia	64.4	84.7	87.6	77.6	80.1	90.1	79.93	49.18	77.31
Mallard-Fly	60.1	58.5	61.7	43.6	54.1	60.6	74.62	42.64	68.46
Mallard-Water	8.7	78.5	76.1	70.4	68.7	90.7	83.34	25.31	79.43
Motocross-Bumps	61.7	68.9	61.4	53.4	30.6	40.1	83.78	56.56	77.15
Motocross-Jump	60.2	28.8	25.1	9.9	51.1	34.1	80.43	59.02	77.50
Motorbike	55.9	57.2	71.4	68.7	71.3	56.3	28.67	45.71	41.15
Paragliding	72.5	86.1	88	90.7	86.6	87.5	17.68	60.76	47.42
Paragliding-Launch	50.6	55.9	62.8	53.7	57.1	64	58.88	50.34	57.00
Parkour	45.8	41	90.1	24	32.2	75.6	79.39	58.51	75.81
Rhino	77.6	67.5	68.2	81.2	79.4	78.2	77.56	83.03	87.52
Rollerblade	31.8	51	81.4	46.1	45	58.8	63.27	57.73	69.01
Scooter-Black	52.2	50.2	16.2	62.4	50.4	33.7	36.07	62.18	68.47
Scooter-Gray	32.5	36.3	58.7	43.3	48.3	50.8	73.22	61.69	73.40
Soapbox	41	75.7	63.4	68.4	44.9	78.9	49.70	53.24	62.57
Soccerball	84.3	87.9	82.9	6.5	82	84.4	29.27	73.56	79.72
Stroller	58	75.9	84.9	66.2	59.7	76.7	63.91	54.40	66.55
Surf	47.5	89.3	77.5	75.9	84.3	49.2	88.78	73.00	88.41
Swing	43.1	71	85.1	10.4	64.8	78.4	73.75	59.41	74.05
Tennis	38.8	76.2	87.1	57.6	62.3	73.7	76.88	47.19	70.75
Train	83.1	45	72.9	84.6	84.1	87.2	42.50	80.33	75.56
Avg. IoU	57.5	56.9	64.1	59.6	63.1	**66.5**	64.69	60.18	**71.51**

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.

Segtrack-v2 dataset (14 videos)							
Methods	FST [2]	KEY [3]	NLC [4]	HVS [1]	Ours-A	Ours-M	Ours-Joint
Human in loop?	No	No	No	Yes	No	No	No
birdfall2	17.50	49.00	74.00	57.40	6.94	55.50	38.01
bird of paradise	81.83	92.20	-	86.80	49.82	62.46	69.91
bmx	67.00	63.00	79.00	35.85	59.53	55.12	59.08
cheetah	28.00	28.10	69.00	21.60	71.15	36.00	59.59
drift	60.50	46.90	86.00	41.20	82.18	80.03	87.64
frog	54.13	0.00	83.00	67.10	54.86	52.88	57.03
girl	54.90	87.70	91.00	31.90	81.07	43.57	66.73
hummingbird	52.00	60.15	75.00	19.45	61.50	60.86	65.19
monkey	65.00	79.00	71.00	61.90	86.42	58.95	80.46
monkeydog	61.70	39.60	78.00	43.55	39.08	24.36	32.80
parachute	76.32	96.30	94.00	69.10	24.86	59.43	51.58
penguin	18.31	9.27	-	74.45	66.20	45.09	71.25
soldier	39.77	66.60	83.00	66.50	83.70	48.37	69.82
worm	72.79	84.40	81.00	34.70	29.13	59.94	50.63
Avg. IoU	53.5	57.3	**80***	**50.8**	56.88	53.04	61.40

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

[1] Grundmann, M., Kwatra, V., Han, M., Essa, I.: Efficient hierarchical graph based video segmentation. In: CVPR. (2010) 1, 2, 3

[2] Papazoglou, A., Ferrari, V.: Fast object segmentation in unconstrained video. In: ICCV. (2013) 2, 3

[3] Lee, Y.J., Kim, J., Grauman, K.: Key-segments for video object segmentation. In: ICCV. (2011) 2, 3

[4] Faktor, A., Irani, M.: Video segmentation by non-local consensus voting. In: Proceedings of the British Machine Vision Conference, BMVA Press (2014) 2, 3

[5] Perazzi, F., Wang, O., Gross, M., Sorkine-Hornung, A.: Fully connected object proposals for video segmentation. In: The IEEE International Conference on Computer Vision (ICCV). (2015) 2

[6] Märki, N., Perazzi, F., Wang, O., Sorkine-Hornung, A.: Bilateral space video segmentation. In: CVPR. (2016) 743–751 2