The qualitative segmentation results of the four unsupervised VOS methods, including FSEG [1], ARP [2], LVO [3] and AnDiff [4]. For the seven challenging sequences, FSEG, ARP and LVO miss parts of objects and incorrectly segment background regions to the foreground in some cases. AnDiff can generate relatively good results but still with slight segmentation errors.

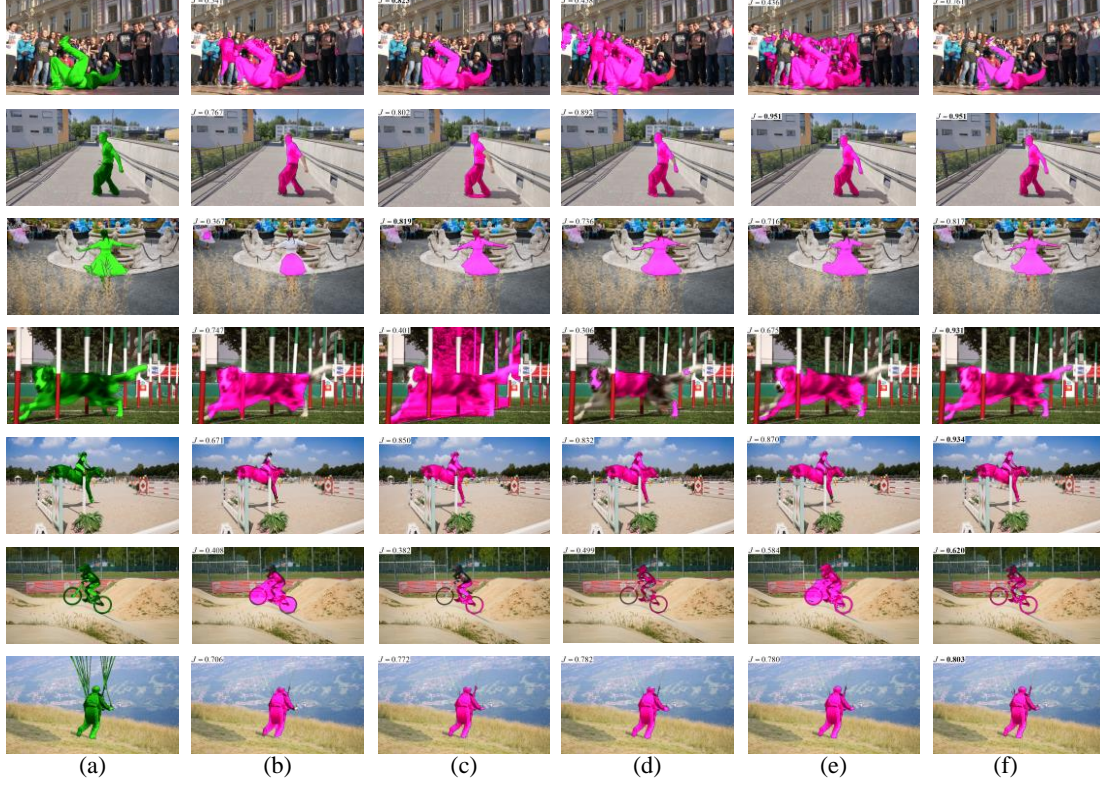| (a) | (b) | (c) | (d) | (e) | (f) |

Fig. 1 Comparisons of segmentation results on DAVIS-2016 dataset. (a) video frames with ground truth; (b)-(e) results from traditional methods, including FSEG [61], ARP [2], LVO [3] and AnDiff [4]; (f) results generated by the proposed method. The value on the left corner of each image is its quantitative result in terms of region similarity J. Bold indicates the best result.

References:

[1] S. Jain, B. Xiong, K. Grauman. FusionSeg: Learning to combine motion and appearance for fully automatic segmentation of generic objects in videos. In CVPR, pp. 2117-2126, 2017.

[2] Y. J. Koh and C.-S. Kim. Primary Object Segmentation in Videos Based on Region Augmentation and Reduction. In CVPR, pp. 7417-7425, 2017.

[3] P. Tokmakov, K. Alahari, C. Schmid. Learning Video Object Segmentation with Visual Memory. In ICCV, pp. 4491-4500, 2017.

[4] Z. Yang, Q. Wang, L. Bertinetto, W. M. Hu, S. Bai, P. H.S. Torr. Anchor Diffusion for Unsupervised Video Object Segmentation. In ICCV, pp. 931-940, 2019.