Supplemental Materials on
Video Object Segmentation Using Global and Instance Embedding Learning

Wenbin Ge[1]*, Xiankai Lu[2*†], Jianbing Shen[3,1]

[1] Beijing Institute of Technology   [2] School of Software, Shandong University   [3] Inception Institute of Artificial Intelligence

gewenbin292@bit.edu.cn, carrierlxk@gmail.com

1. Results on YouTube-VOS 2019 validation dataset

We report complete results of our proposed method on YouTube-VOS 2019 validation dataset [4]. Table 1 shows results of STM [2], LWL [1] and our proposed method.

		STM [2]	LWL [1]	Ours_fast	**Ours**
	Overall	79.2	**81.0**	79.1	80.5
seen	Mean$\mathcal{J}$ ↑	79.6	**79.6**	78.2	79.5
	Mean$\mathcal{F}$ ↑	83.6	**83.8**	82.3	83.6
unseen	Mean$\mathcal{J}$ ↑	73.0	**76.4**	73.8	76.0
	Mean$\mathcal{F}$ ↑	80.6	**84.2**	82.1	82.9

Table 1: **Evaluation of O-VOS on Youtube-VOS$_{19}$ val set**, with region similarity $\mathcal{J}$ and boundary accuracy $\mathcal{F}$. "Overall": averaged over the four metrics.

2. Visualization

We present a qualitative evaluation of our proposed method on each sequence of DAVIS$_{17}$ *val* dataset [3], DAVIS$_{17}$ *test-dev* dataset [3] and Youtube-VOS *val* dataset [4] including Youtube-VOS$_{18}$ and Youtube-VOS$_{19}$. Specifically, Figure. 1 shows the visualization results of Ours on DAVIS$_{17}$ *val* dataset [3]. Figure. 2 shows the visualization results of Ours_fast on DAVIS$_{17}$ *val* dataset [3]. Figure. 3 shows the visualization results of Ours on DAVIS$_{17}$ *test-dev* dataset [3]. Figure. 4 shows the visualization results of Ours_fast on DAVIS$_{17}$ *test-dev* dataset [3]. Figure. 5 shows the visualization results of Ours on Youtube-VOS *val* dataset [4]. Figure. 6 shows the visualization results of Ours_fast on Youtube-VOS *val* dataset [4].

References

[1] Goutam Bhat, Felix Jremo Lawin, Martin Danelljan, Andreas Robinson, Michael Felsberg, Luc Van Gool, and Radu Timofte. Learning what to learn for video object segmentation. In *ECCV*, 2020. 1

[2] Seoung Wug Oh, Joon-Young Lee, Ning Xu, and Seon Joo Kim. Video object segmentation using space-time memory networks. In *ICCV*, 2019. 1

[3] Jordi Pont-Tuset, Federico Perazzi, Sergi Caelles, Pablo Arbeláez, Alex Sorkine-Hornung, and Luc Van Gool. The 2017 davis challenge on video object segmentation. *arXiv preprint arXiv:1704.00675*, 2017. 1, 2, 3

[4] Ning Xu, Linjie Yang, Yuchen Fan, Jianchao Yang, Dingcheng Yue, Yuchen Liang, Brian Price, Scott Cohen, and Thomas Huang. Youtube-vos: Sequence-to-sequence video object segmentation. In *ECCV*, 2018. 1, 4

---

*The first two authors contribute equally to this work.
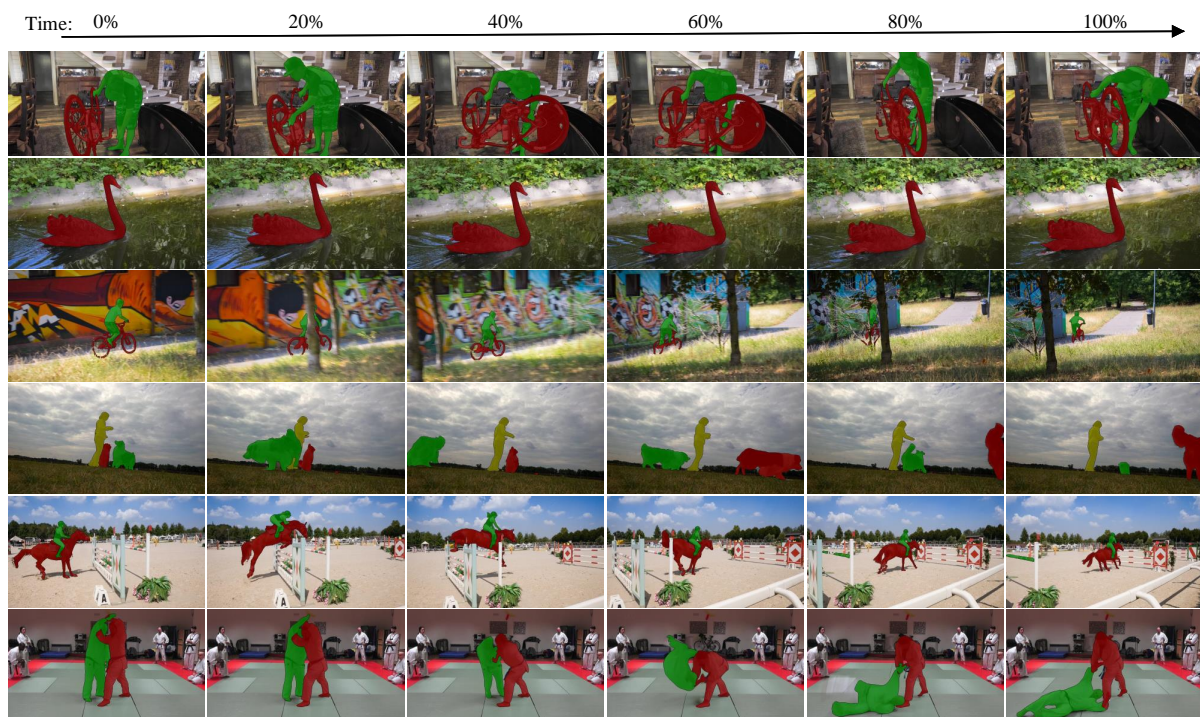
†Corresponding author: *Xiankai Lu*.

Figure 1: Qualitative results of Ours on DAVIS$_{17}$ *val* dataset [3]. From top to bottom are *bike-packing, blackswan, bmx-trees, dogs-jump, horsejump-high* and *judo*.
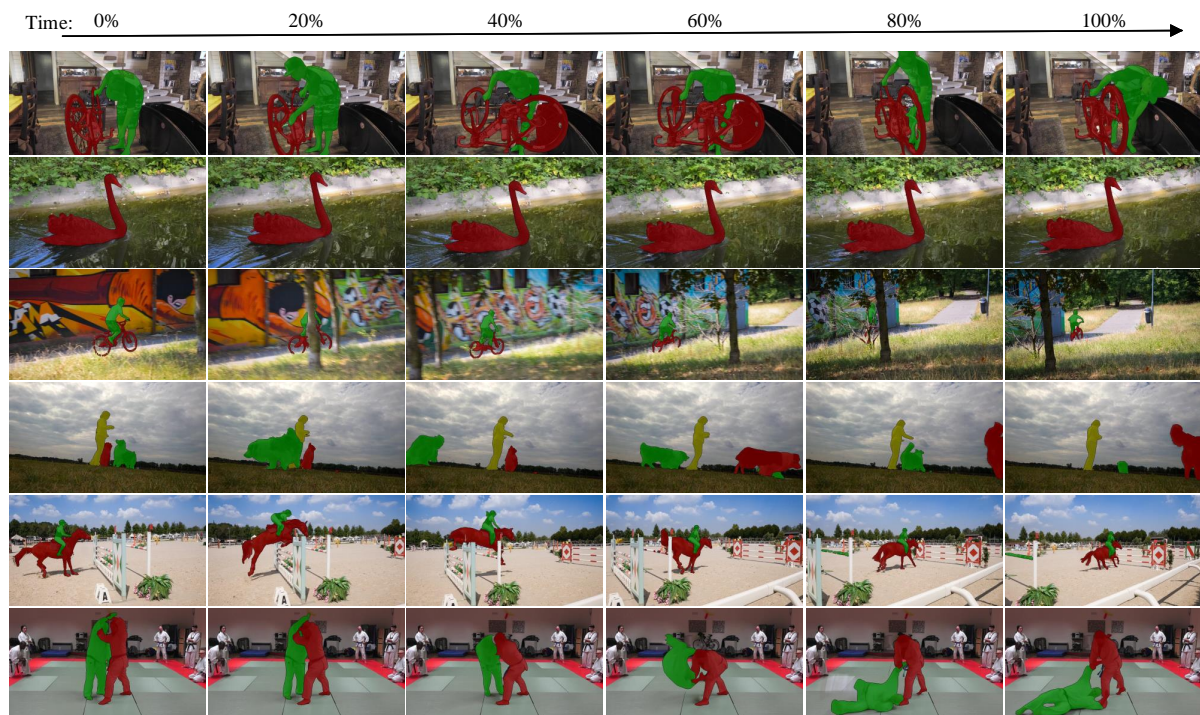

Figure 2: Qualitative results of Ours_fast on DAVIS$_{17}$ *val* dataset [3]. From top to bottom are *bike-packing, blackswan, bmx-trees, dogs-jump, horsejump-high* and *judo*.
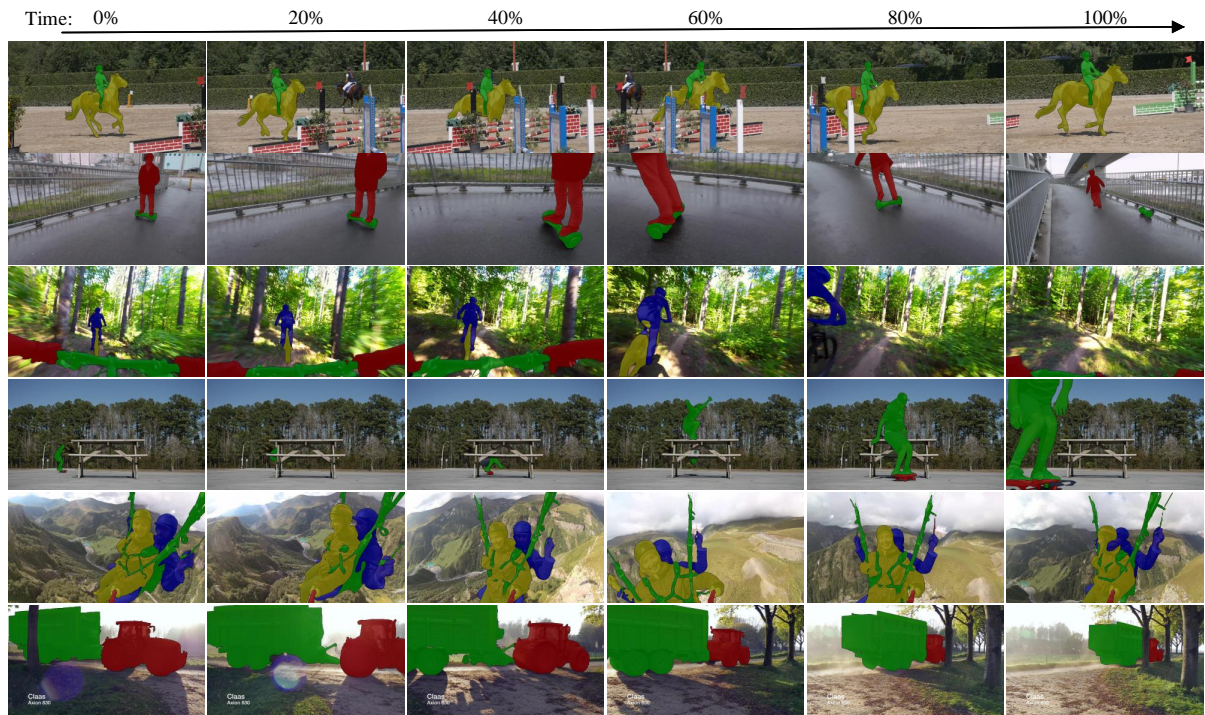
Figure 3: Qualitative results of Ours on DAVIS$_{17}$ *test-dev* dataset [3]. From top to bottom are *horsejump-stick, hoverboard, mtb-race, skate-jump, tandem* and *tractor*.
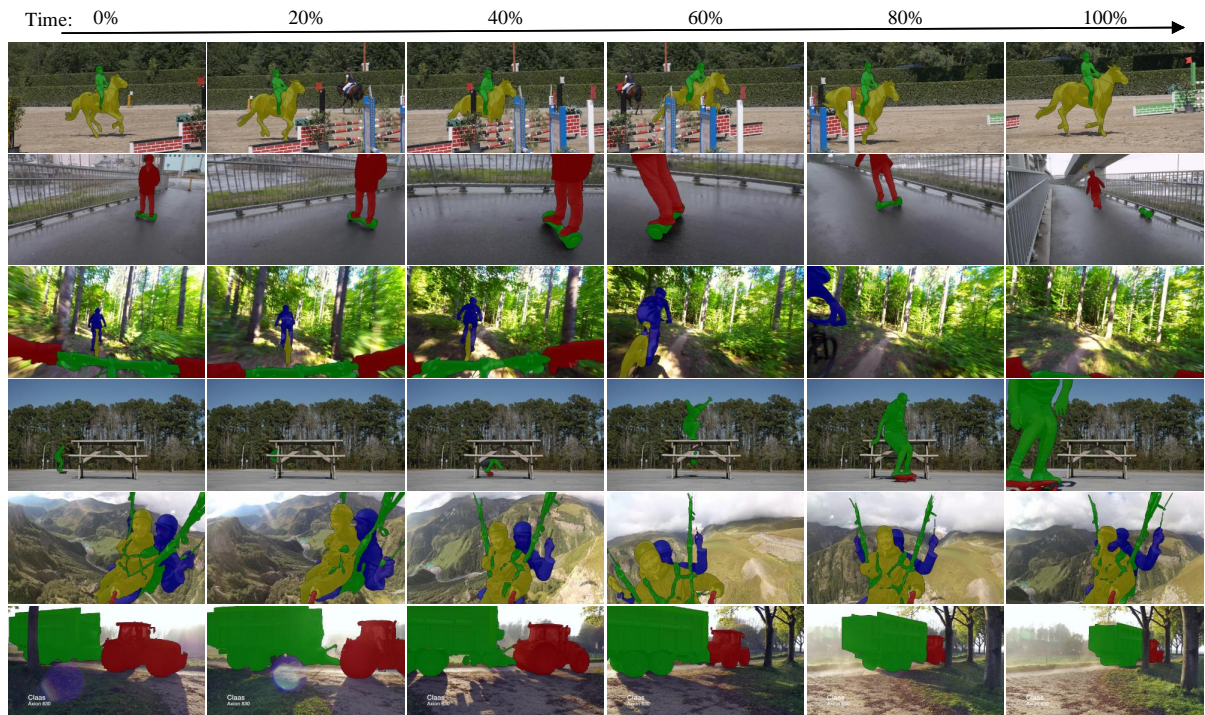

Figure 4: Qualitative results of Ours_fast on DAVIS$_{17}$ *test-dev* dataset [3]. From top to bottom are *horsejump-stick, hover-board, mtb-race, skate-jump, tandem* and *tractor*.

Figure 5: Qualitative results of Ours on Youtube-VOS *val* dataset [4]. From top to bottom are *3dd327ab4e, 45d898acc4, 63ca8970d9, 83a5056a16, 193aa74f36* and *3674b2c70a*.
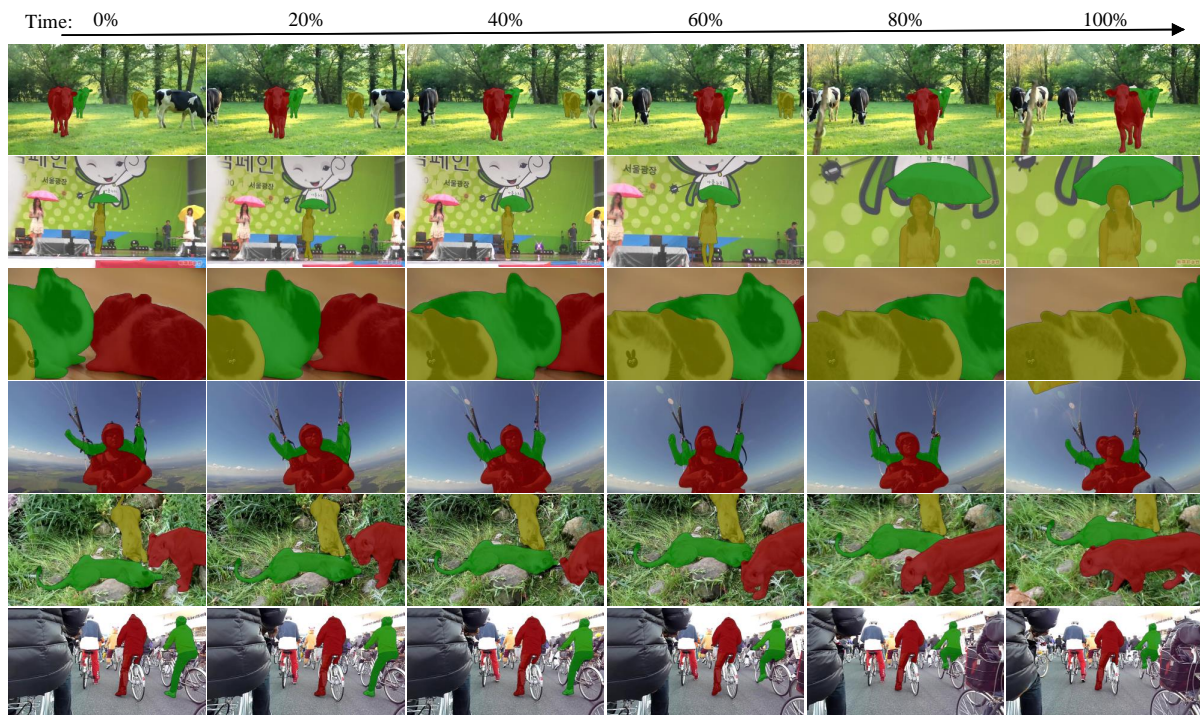

Figure 6: Qualitative results of Ours_fast on Youtube-VOS *val* dataset [4]. From top to bottom are *3dd327ab4e, 45d898acc4, 63ca8970d9, 83a5056a16, 193aa74f36* and *3674b2c70a*.