

Figure 1: Video object segmentation on DAVIS16 or DAVIS17 datasets. Our tracker outperforms the leading segmentation-based trackers and predicts accurate masks under challenging scenes.



Figure 2: Examples of sequences with similar distractors and extreme complex background. It shows our tracker is robust to the similar objects even the distractors are closed to the targets. Our tracker can also handle with the complex background such as dark, raining and blur.





Figure 3: Examples of sequences with dim targets and appearance variations. As is shown in the sequence Bird, our tracker can predict accurate segmentation results even when target is extremely small. Moreover, our tracker achieves marvelous performance when fast-moving targets have extreme appearance deformation.

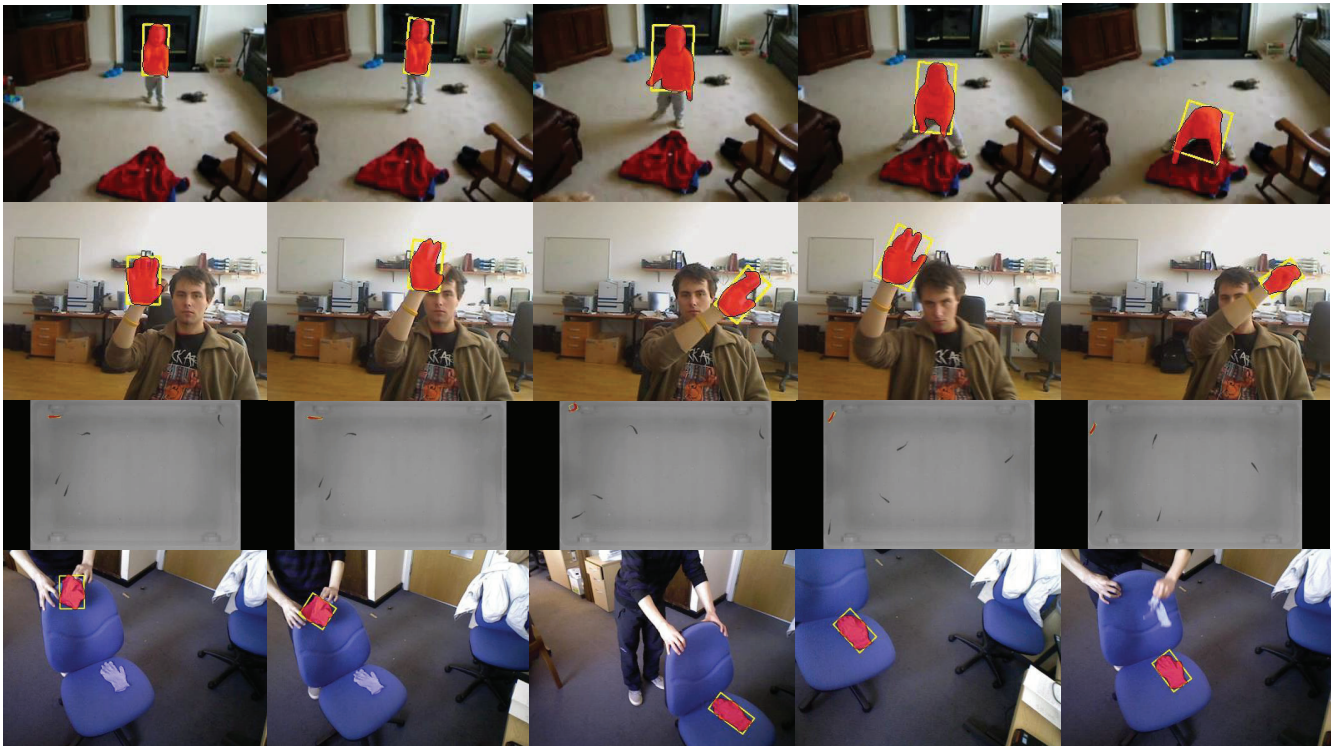


Figure 4: Examples of sequences with unseen targets. DNN-based trackers heavily rely on the pre-trained knowledge of targets. Our tracker utilizes the spatiotemporal information of every frame to enhance the generalization ability. From the examples of hand, blanket and glove, our tracker still tackles with those unseen targets well.