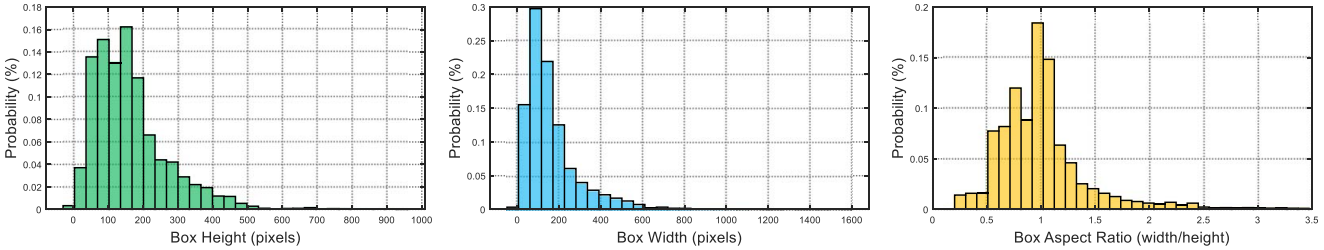
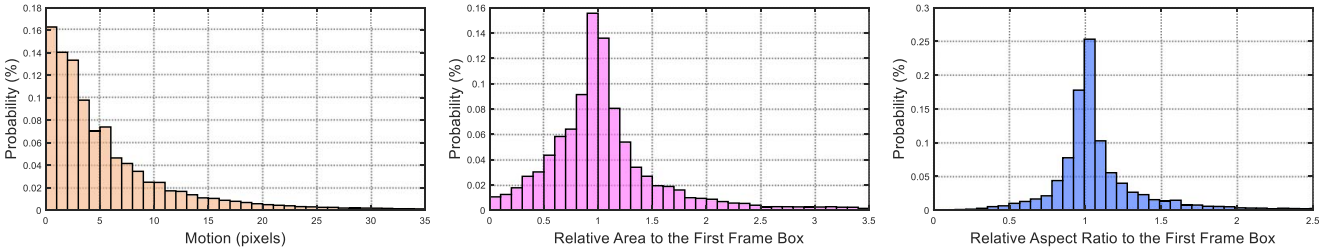
Transparent Object Tracking Benchmark
— Supplementary Material —

1. Statistics of Annotation Boxes

In this section, we demonstrate more statistics of annotation boxes in TOTB. In particular, we show the bounding box distribution over all sequences of TOTB, including box height, width and aspect ration (*i.e.*, width/height), in Figure 1 (a). From Figure 1 (a), we can see that our benchmark covers objects with various scales, showing its diversity. In addition, we also demonstrate the bounding box variation over time, as shown in Figure 1 (b). In specific, we display the distance of center points of targets in two consecutive frames, relative area and aspect ratio to the first frame bounding box. We observe that the target varies rapidly in videos of TOTB.



(a) Bounding box distribution over all sequences



(b) Bounding box variation over time

Figure 1. Statistics of annotations in TOTB, including bounding box distribution over all sequences and bounding box variation over time. *Best viewed in pdf and by zooming in.*

2. Details of Transparent Object Segmentation Network

In our proposed TransATOM, we propose to exploit transparency features for improving transparent object tracking. To this end, we develop a transparent object segmentation network to extract the transparent features. In specific, we adopt fully convolutional network (FCN) [2] for segmentation. The feature extraction backbone network is borrowed from the powerful ResNet-18 [1]. Figure 2 illustrates the architecture of our transparent object segmentation network.

We use the training data from [3]. Note that, we only use the images from the *thing* category in which the images are small and movable. In specific, we utilize 2,844 static images for training. The initial parameters of the backbone are from pre-trained ResNet-18. We use the standard cross-entropy loss to train the whole network. For optimization, we apply stochastic gradient descent (SGD) method with momentum of 0.9 and a weight decay of 0.0005. The training batch size is set to 8. The initial learning rate is 0.02 and decayed by poly strategy with the power of 0.9 for 30 epochs. Note that, one can leverage more complex architecture for the segmentation network, which is expected to bring improvements. Nevertheless, considering the

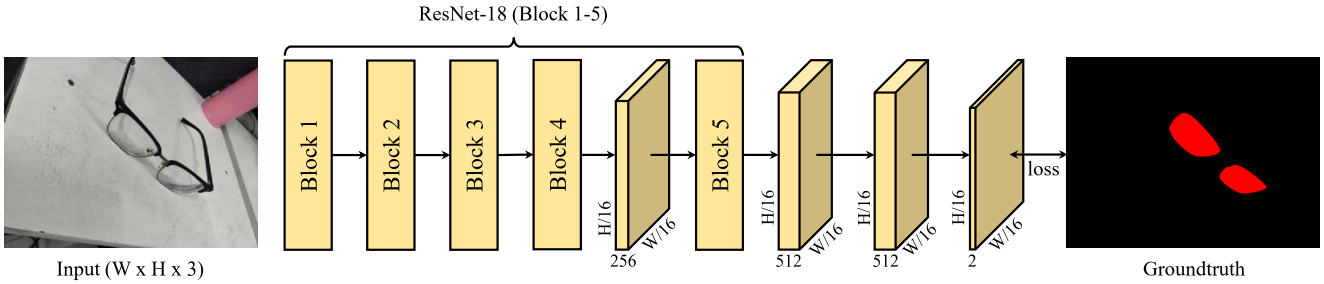
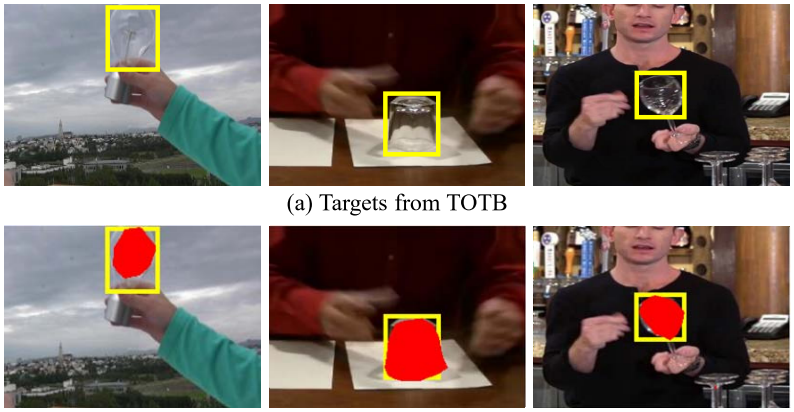


Figure 2. Illustration of transparent object segmentation network. We apply the standard ResNet-18 [1] as the feature extraction backbone. The parameters of Block 1-5 are directly borrowed and initialized from ResNet-18. We refer readers to [1] for detailed structures of the backbone network. *Best viewed in pdf and by zooming in*.

importance of efficiency for object tracking, we utilize simple FCN with ResNet-18. After completing training, we apply Block 1-4 in the segmentation to extract transparency features.

It is worth noticing that, there is no overlap between the training images and sequences in TOTB. However, since transparency features are common and transferable between different transparent instances, we can directly utilize the trained transparent segmentation network for feature extraction of transparent object for tracking. We display some segmentation results of targets in TOTB using trained model as shown in Figure 3. We observe that the targets can be well segmented from background, which shows the transferability of transparency features.



(a) Targets from TOTB

(b) Segmentation results

Figure 3. Segmentation results of targets in TOTB using transparent object segmentation networks. *Best viewed in pdf and by zooming in*.

3. Full Results of Overall Performance for All Tracking Algorithms

Figure 4 shows the overall performance of 25 state-of-the-art trackers and TransATOM in terms of precision, normalized precision and success.

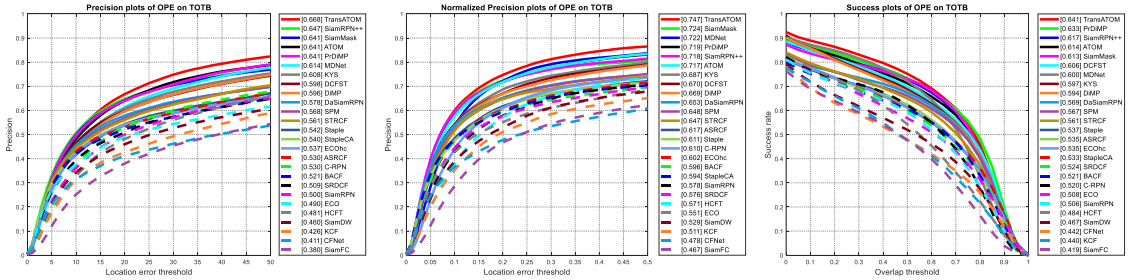


Figure 4. Overall performance of 25 state-of-the-art trackers and TransATOM on TOTB in terms of precision, normalized precision and success. *Best viewed in pdf and by zooming in*.

4. Full Results of Attribute-based Evaluation

Figure 5 shows the performance of trackers on each attribute using success.

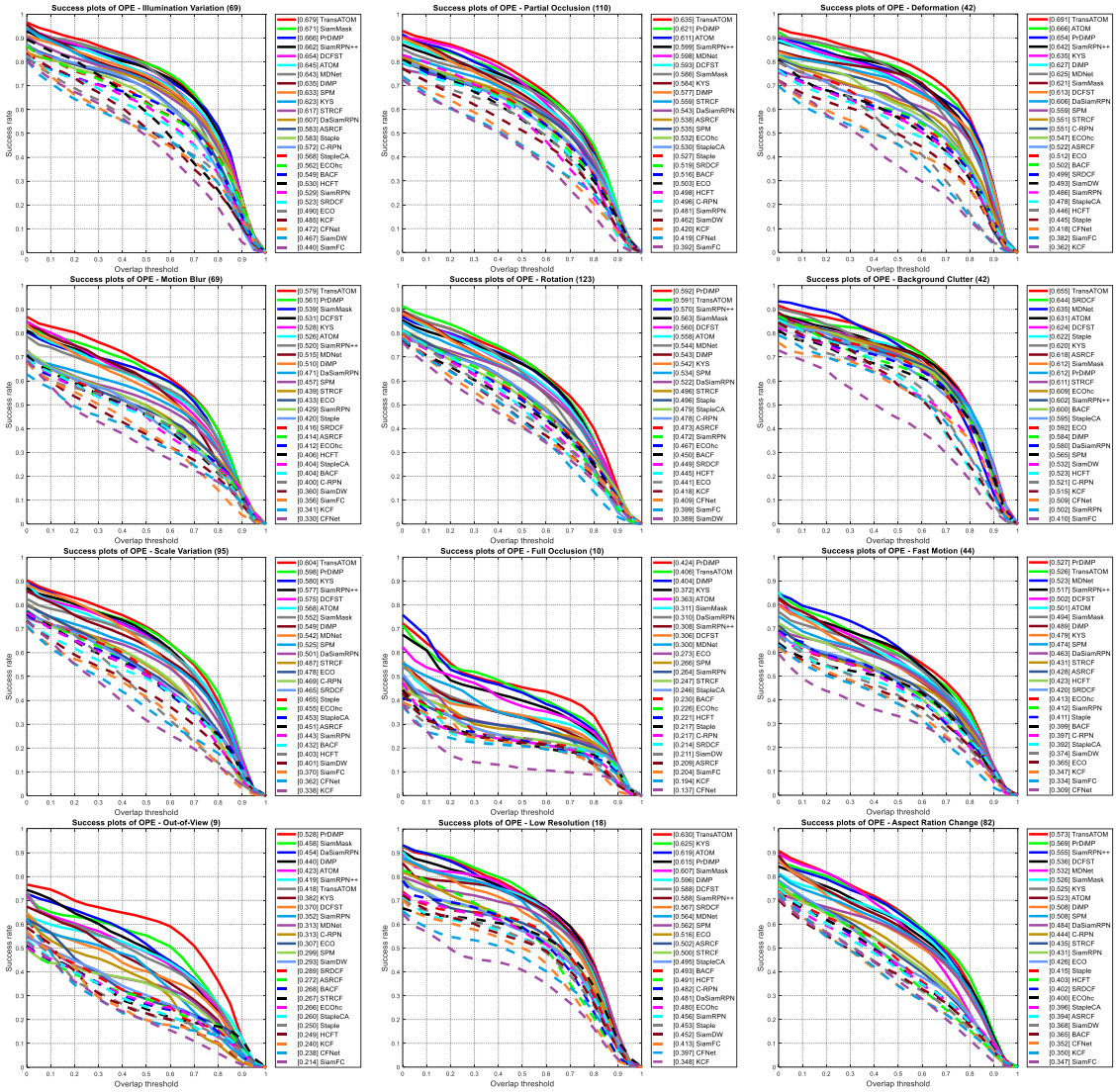


Figure 5. Performance of trackers on 12 attributes using success. *Best viewed in pdf and by zooming in*.

References

[1] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, 2016. 1, 2

[2] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *CVPR*, 2015. 1

[3] Enze Xie, Wenjia Wang, Wenhai Wang, Mingyu Ding, Chunhua Shen, and Ping Luo. Segmenting transparent objects in the wild. In *ECCV*, 2020. 1