

SUPPLEMENTARY MATERIAL

You can see the performance of the proposed methods on the -bull- video of VOT-LT data set. The target object is specified at the first frame of the video and it is detected through 1930 frames. For comparison, bounding boxes detected by the baseline and the proposed methods along with the ground truth bounding boxes are coded with different colors.

- GT (red): Ground truth bounding box of the target object.
- CD-BHRL-ASFT (yellow) : Aaugmented-shot fine-tuning by BHRL.
- CD-BHRL- V_{1st} (pink): The baseline for CD-BHRL-ASFT.
- CD-DeFRCN-ASFT (green) : Augmented-shot fine-tuning by DeFRCN.
- CD-BHRL- V_{1st} (blue): The baseline for CD-DeFRCN-ASFT.