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 grountruth bounding boxes are coded with different colors.

• GT (red): Ground truth bounding box of the target object.
• CD-BHRL-ASFT (yellow) : Aaaugmented-shot fine-tuning by BHRL.
• CD-BHRL-V1st (pink): The baseline for CD-BHRL-ASFT.
• CD-DeFRCN-ASFT (green): Augmented-shot finetuning by DeFRCN.
• CD-BHRL-V1st (blue): The baseline for CD-DeFRCN-ASFT.

Table 1 presents the data augmentation policies applied in our Augmented-Shot Finetuning process. It includes one original image along with four images generated using different augmentation techniques, illustrating the transformations introduced by each policy.

Table 1 The policies and their explanation used in our Augmented-Shot FineTuning.

	Operation 1	Operation 2	Explanation
Policy 1	No Operation	No Operation	Original Query Frame
Policy 2	TranslateX	Equalize	The query is shifted in X direction after histogram equalization
Policy 3	BBox_Only_TranslateY	Cutout	Only the query novel object BB is translated in Y direction, subsequently a random cutout is applied on the frame to mimic occlusion
Policy 4	ShearY	BBox_Only_TranslateY	Shearing is applied on Y axis, subsequently only the query novel object BB is translated in Y direction
Policy 5	Rotate	Color	Original query frame is rotated, subsequently illumination change is applied on