- Updating target with predicted BB can accumulate false info (bias) & cause loss of target in long-term. Mainly caused by heavy updating done to account for rapid changes.
- Short term tracking benchmark datasets cover hard cond^{n}s but can't be used to judge LT . An OTB vid was artificially extended to ensure that any change in performance is due to inc in len.ECO acc decreases significantly for little app change => erroneous updates.
- General trackers update model at step t+1 minimizing loss (L2) for frames 1:t by GD. $\nabla_{\varphi} L_t = \nabla_{\varphi} E[(y_i - f_{i,t})^2] = 2E[f_{i,t} \nabla_{\varphi} f_{i,t}] - 2E[y_t \nabla_{\varphi} f_{i,t}]$. (Assumed that y independent of φ) But y_i used for retrain & fine-tuning are not perfect for all i>1. If they were perfect => no need to retrain (i=1 to t). y_i are noisy preds of y_i^*, $y_i = y_i^* + \delta_i$.(Gausssian noise = bias)
- On deriving relation b/n $f_{i,t}$ and $f_{i,t+1}$, at each time step, predictions are drifted linearly proportional to past errors (Decay). This decay recursively increases for longer vids. CF based trackers are expected to develop huge bias in LT, but not Siamese (local search).
- To approximate the bias term, target has to be redetected for all frames with new params which is not feasible. A simplified version of bias was used to create LT-SINT in which global search is done at every T frames as it has better chances of not losing target. But even it can perform worse if global search coincides with large appearance change.
- Long OTB - extended vids of OTB. 3 extreme challenges causing decay : sampling drift (fast motion & scale variation caused by imperfect local space search), target disappearance (assumption of motion continuity compromised) and pixel level errors (caused by pseudo +ves generated which add some backg pixels as +ves).
- ECO decays largely by occlusion, but less by low resolution as it does multi search. SINT is affected only by backg clutter and oop rotation. But SINT scores significantly less than ECO. This dataset has same frames repeated so it is easier.
- ω_t(1/0) was used to limit updates to some frames. Decay recognition network was designed to predict ω_t(perfect or biased update). It should not share any params with tracker to be unaffected by $\nabla_{\varphi} f_{i,t}$'s bias. Overly correlated input should not be given.
- LT-SINT : DRN estimates if next update will add bias using last K frames. It is an LSTM trained with +ve (>0.5 IoU), -ve samples. Global search is done for N locs at M scales, also finer scales are used to improve, at every T frames. It does not decay with reps.
- DRN first encodes similarity maps by small convnet and then sends them to 2-layer LSTM. BB for absent annotations are given IoU 0. DRN outperforms blind update, no update and IoU>0.5 updates.
- Global search performs better by eliminating sampling drift. LT-SINT outperforms all in long-term but has low F-score in short term due to additional constraint.