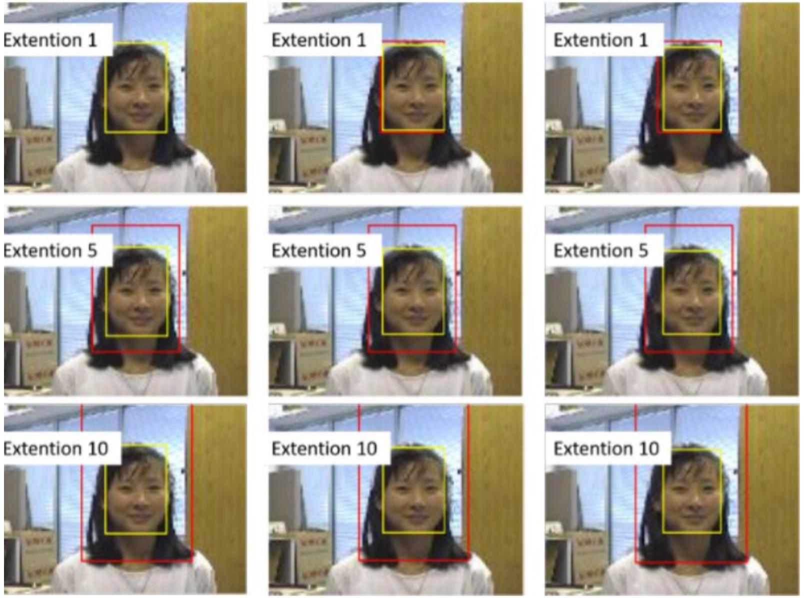
Model-Decay in Long-Term Tracking

Efstratios Gavves[1], Ran Tao[2], Deepak K. Gupta [1]and Arnold Smeulders[1]
[1]QUVA Lab and Informatics Institute, University of Amsterdam, The Netherlands
[2]Kepler Vision Technologies
Paper Id: 839

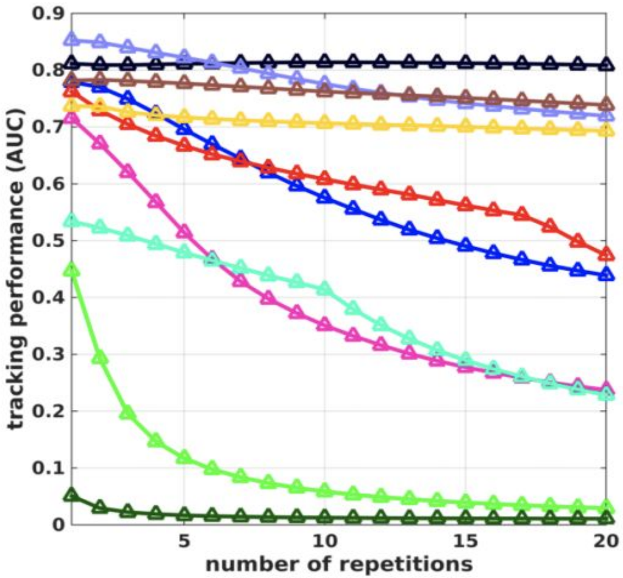
UNIVERSITY OF AMSTERDAM

QUVA Deep Vision Lab

Model decay on long videos



Example from OTB100 showing 3 frames from 3 different repetitions. Note that the predictions in the similar frames worsens over longer durations due to model decay..



AUC scores for ECO tracker on multiple repetitions of 10 videos from OTB100 showing model decay.

Small error in tracker prediction can accumulate over a large number of frames to eventually cause model drift.

How can model updates be performed cautiously?

Understanding tracker model decay

Mathematical model for model decay

Popular learning strategy for trackers:

$$\phi_{t+1} = \arg\min_{\phi} \mathcal{L}(x_{1:t}, y_{1:t}) \qquad y_{t+1} = f(x_{t+1}; \phi_{t+1})$$

where f s the ϕ-parameterized tracker that minimizes the tracker loss $\mathcal{L}$ over the dataset $D = [x_{1:t}, y_{1:t}]$

Assuming Gaussian noise with variance σ_i^2, we can state:

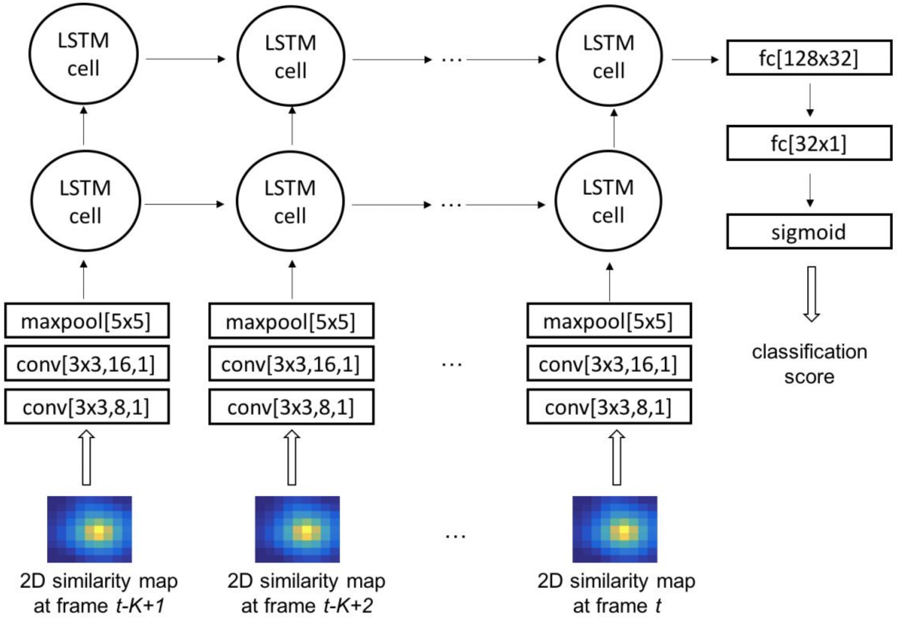
$$y_i = y_i^* + \delta_i, \text{ and } \delta_i \sim N(0, \sigma_i^2)$$

Model update can be represented as:

$$\boxed{\phi_{t+1} - \phi_t = \underbrace{-2\eta\mathbb{E}[(f_{i,t} - y_i^*) \cdot \nabla_\phi f_{i,t}]}_{\text{Perfect parameter update}} + \underbrace{2\eta\mathbb{E}[\delta_i \cdot \nabla_\phi f_{i,t}]}_{\text{Parameter bias}}}$$

Decay Recognition Network

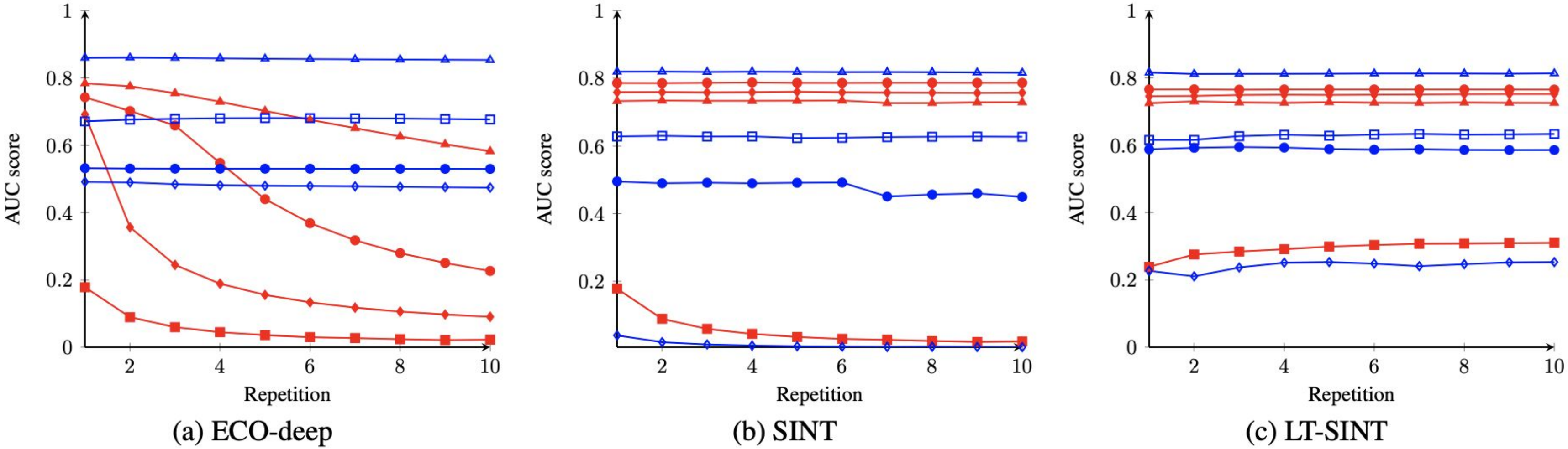
Performing cautious model updates



Network architecture for decay recognition network. Based on K-1 previous evaluations, it decides whether the current prediction can be used to update the model.

Experiments

Siamese tracker equipped with decay recognition network can better handle the decay of the tracker model.



Performance scores for three different trackers on 10 different videos of OTB100. We show the top 5 (blue) and bottom 5 (red) AUC scores among all the videos.

Experiments

	UAV20L	YouTube Long	YouTube Long *(20 sec)*	OxUvA
ECO-DEEP [4]	42.7	7.1	1.4	39.5
TLD [14]	22.8	22.4	20.2	20.8
LTCT [21]	25.5	2.2	0.2	29.2
SPL [31]	35.6	—	—	—
SRDCF [7]	34.3	—	—	—
MUSTer [13]	32.9	—	—	—
SiamFC+R [33]	—	—	—	42.7
SiamFC [34]	—	—	—	39.1
MDNet [27]	—	—	—	47.2
SINT [32]	49.4	37.6	—	42.6
EBT [37]	—	—	—	32.1
BACF [10]	—	—	—	31.6
Staple [1]	—	—	—	27.3
LT-SINT	**52.4**	**42.1**	**39.5**	**57.9**

Performance comparison of LT-SINT with other trackers on long videos.

Conclusions

- Too many model updates can lead to tracker drift in long videos.
- To avoid model drift, cautious updates should be performed.
- The weak temporal correspondence used by the decay recognition network helps to tackle model decay.