Learning the Model Update for Siamese Trackers

Lichao Zhang, Abel Gonzalez-Garcia, Joost van de Weijer, Martin Danelljan, Fahad Shahbaz Khan

Computer Vision Center, Universitat Autònoma de Barcelona,
Computer Vision Laboratory, ETH Zürich
Inception Institute of Artificial Intelligence, UAE & Computer Vision Laboratory, Linköping University

Introduction

Standard formulation for Siamese trackers:

$$\widetilde{T}_i = (1-\gamma)\widetilde{T}_{i-1} + \gamma T_i$$

Drawbacks of standard update:

- a constant update rate for every video, regardless of attributes such as camera motion and motion blur.
- access to the initial appearance template is lost.
- limited flexibility of linear combination.

UpdateNet learns to update the template automatically. It is a compact neural network that includes as inputs:

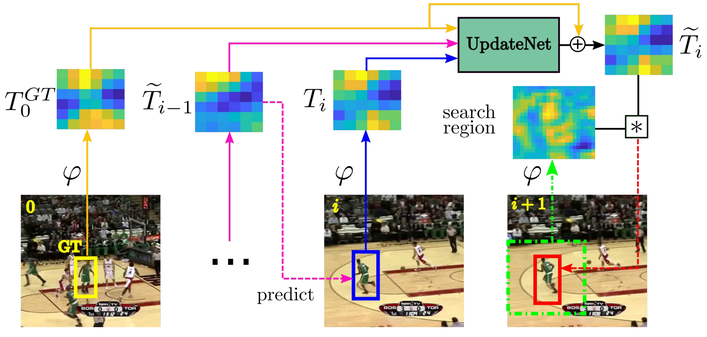
- the initial ground-truth template T_0^{GT}
- the accumulated template from all previous frames $\widetilde{T}_{i-1}$
- the feature template predicted in the current frame T_i

Qualitative results:

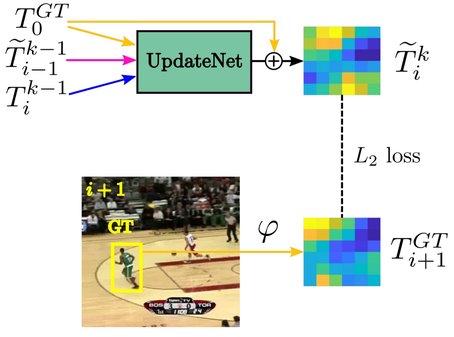
UpdateNet-SiamFC **SiamFC** **GT**

Model and Training

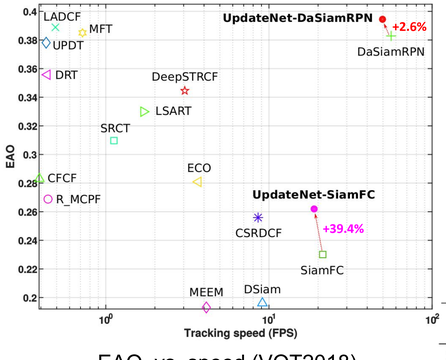
Learning to update: $\widetilde{T}_i = \phi(T_0^{GT}, \widetilde{T}_{i-1}, T_i)$



Loss: $\mathcal{L}_2 = \left\| \phi(T_0^{GT}, \widetilde{T}_{i-1}, T_i) - T_{i+1}^{GT} \right\|_2$



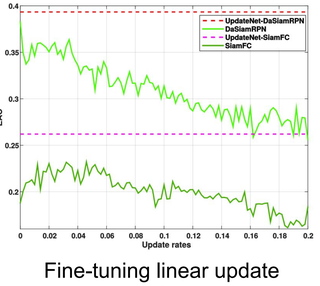
VOT Results



EAO vs. speed (VOT2018).

Update for SiamFC	Skip	EAO (↑)	A (↑)	R (↓)
Linear	-	0.188	0.50	0.59
UpdateNet ($K=1$)	-	0.205	0.48	0.58
UpdateNet ($K=1$)	T_i	0.207	0.47	0.57
UpdateNet ($K=1$)	$\widetilde{T}_{i-1}$	0.214	0.49	0.58
UpdateNet ($K=1$)	T_0^{GT}	0.250	0.50	0.53
UpdateNet ($K=2$)	T_0^{GT}	0.257	0.51	0.50
UpdateNet ($K=3$)	T_0^{GT}	**0.262**	**0.52**	**0.49**

Ablation study (VOT2018), skip connection with T_0^{GT} achieves best results.



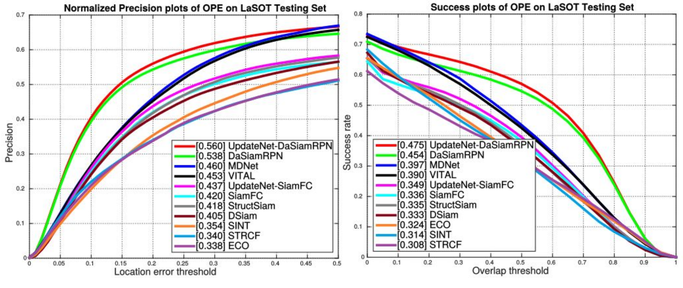
Fine-tuning linear update rates (VOT2018).

	DSiam	MemTrack	SiamFC		DaSiamRPN	
			Linear	UpdateNet	Linear	UpdateNet
EAO	0.181	0.273	0.235	0.289	**0.439**	**0.481**
A	0.492	0.533	0.529	0.543	**0.619**	**0.610**
R	2.934	1.441	1.908	1.320	**0.262**	**0.206**

Comparison with other update strategies (VOT2016).

More Datasets

LaSOT testing set [Fan *et al.* CVPR 2019]:

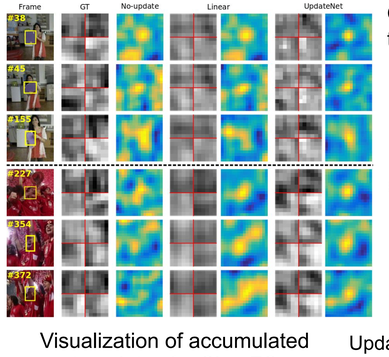


TrackingNet dataset [Muller *et al.* ECCV 2018]:

	ATOM	ECO	CFNet	MDNet	SiamFC		DaSiamRPN	
					Linear	UpdateNet	Linear	UpdateNet
Precision (%)	**64.8**	49.2	53.3	56.5	53.3	53.1	59.1	**62.5**
Norm. Prec. (%)	**77.1**	61.8	65.4	70.5	66.3	67.4	73.3	**75.2**
Success (%)	**70.3**	55.4	57.8	60.6	57.1	58.4	63.8	**67.7**

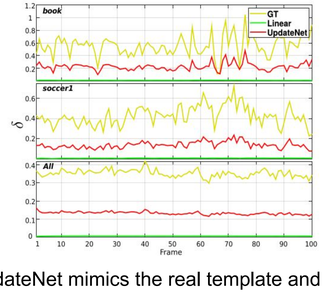
UpdateNet leads to significant performance improvements on the two datasets.

Analysis



Visualization of accumulated templates for SiamFC

Quantifying change rates between templates of contiguous frames:



UpdateNet mimics the real template and predicts a clear peak in the response map.