Target-Aware Deep Tracking Supplementary Document

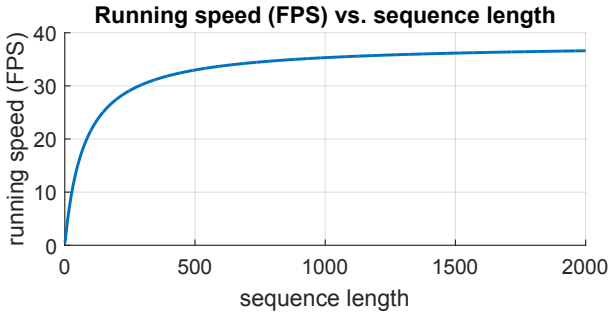
Xin Li[1] Chao Ma[2] Baoyuan Wu[3] Zhenyu He[1*] Ming-Hsuan Yang[4,5]
[1]Harbin Institute of Technology, Shenzhen
[2]MoE Key Lab of Artificial Intelligence, AI Institute, Shanghai Jiao Tong University
[3]Tencent AI Lab [4]University of California, Merced [5]Google Cloud AI

In this document, we present more experimental results and analyses to complement the manuscript:

- Running speed analysis in Section 1.
- More comparisons. We compare the tracking accuracy vs. speed performance on both the OTB-2013 [23] and OTB-2015 [24] datasets with more trackers in Section 2.
- More ablation tests. We present the experiments by directly fine-tuning a deep network with the initial information in Section 3.

1. Running Speed Analysis

The initialization time of the proposed model is about 2 seconds on the machine specified on L575. The final average running speed is related to the length of test sequences as shown in the figure below. The average running speed on the OTB-2015 dataset is 33.7 FPS.



2. Tracking Accuracy vs. Speed

Figure 1 and Figure 2 show the tracking accuracy vs. speed performance on the OTB-2013 and OTB-2015 datasets, respectively. We add more trackers (annotated with gray square) using the reported tracking speed as their source code is not available. While the comparisons based on the reported results without using the source code on the same machine do not truly reveal the run time performance, we add these results for completeness.

The trackers with the reported speed include DRT [19], LSART [18], FlowT [28], SACF [25], SA-Siam [9], RASNet [22], SiamRPN [11], RT-MDNet RT-MDNet, and CFnet2-tri [7]. The other trackers including CCOT [6], VITAL [17], DAT [15], MetaSDNet [14], ECO [5], STRCF [12], DSLT [13], MCPF [26], CREST [16], MCCTH [21], ACT [3], DSiamM [8], ECO-HC [5], BACF [10], CFNet [20], SiamFC [2], Staple [1], Trace [4], and DaSiamRPN [27] are evaluated on the same machine.

Compared with all these state-of-the-art trackers, the proposed algorithm achieves a favorable performance in terms of speed and accuracy. The proposed tracker achieves the best performance among the real-time trackers, which can be attributed to the proposed target-active and scale-sensitive features. The proposed features are generated by the filters which are active to specific patterns of the target and sensitive to scale changes. As such, the target-specific and scale-sensitive features distinguish the target from the background well and are sensitive to scale changes. On the other hand, the proposed algorithm uses a simple tracking framework directly comparing the features of the target template and the search region without a complicated inference model or an online update model. Therefore, the proposed tracker achieve fast performance and prevents overfitting and model drifting.

3. More Ablation Tests

According to suggestions of reviewers, we compare the proposed algorithm with the model which fine-tunes a deep feature network directly. We conduct the experiments by directly fine-tuning a deep network with the initial information and concatenate 3 or 4 CNN layers as features. Table 1 shows that directly fine-tuning performs poorly and concatenating more layers does not improve tracking performance. Concatenating low-level layers does not always improve tracking performance as it makes the network model complicated. In addition, using more layer features does not solve the uncertainty problem because uncertainty comes

*Corresponding author.

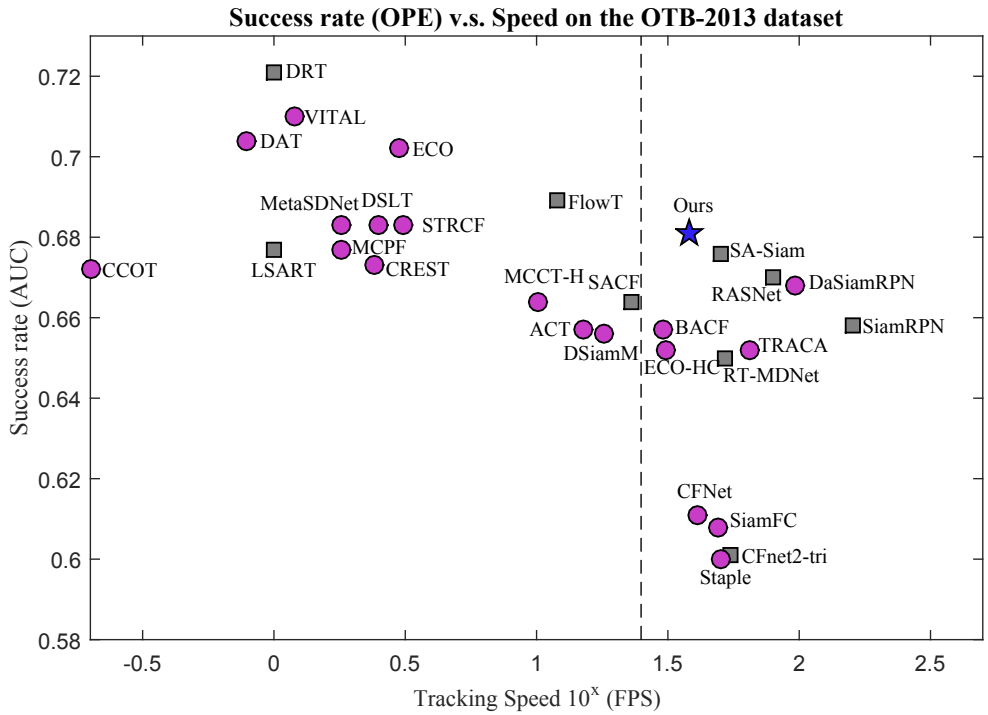


Figure 1. Tracking accuracy vs. speed on the OTB2013 dataset. The proposed algorithm achieves a favorable performance against the state-of-the-art trackers.

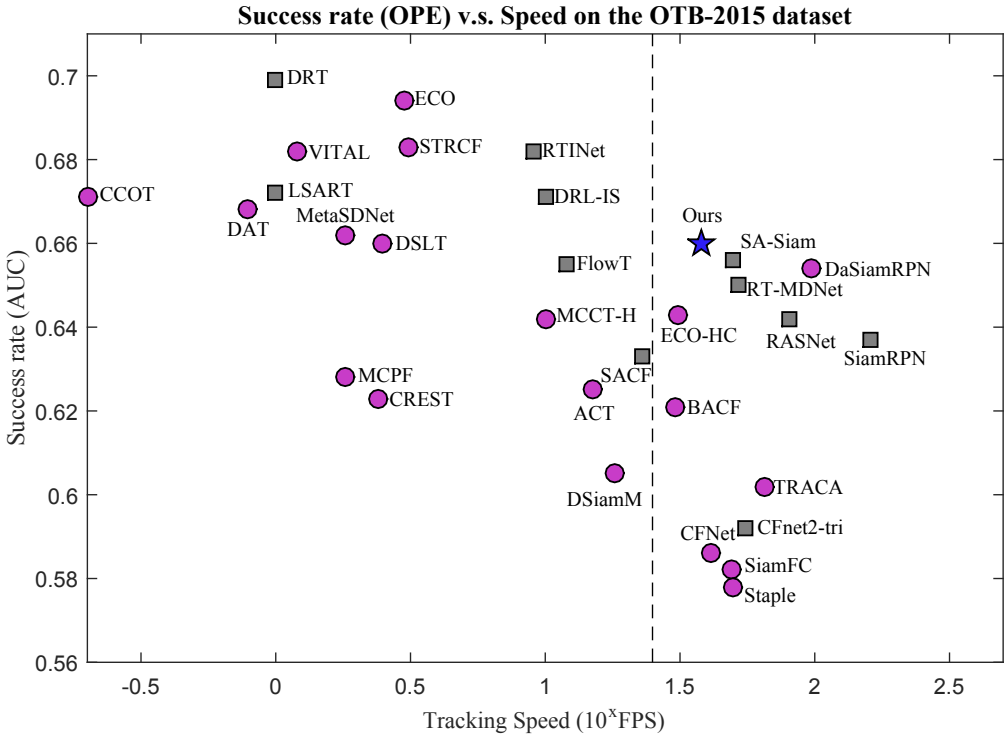


Figure 2. Tracking accuracy vs. speed on the OTB2015 dataset. The proposed algorithm achieves a favorable performance against the state-of-the-art trackers.

from an arbitrary target in the tracking task, which is unknown before tracking.

Table 1. **Comparison with the baseline by directly fine-tuning and concatenating more features.**

Tracker	OTB-2013		OTB-2015	
	Precision	AUC	Precision	AUC
Finetuning	0.848	0.653	0.826	0.639
Threelayers	0.850	0.656	0.827	0.642
Fourlayers	0.817	0.635	0.817	0.633
Ours	**0.896**	**0.680**	**0.866**	**0.660**

Our method significantly differs from the approach by directly fine-tuning a deep network with the initial information in two aspects. First, we do not modify or change the activations of existing pre-trained deep models. Instead, we only select the target-active and the scale-sensitive filters (channels). Second, the original deep models designed for image classification is not effective for tracking due to the use of spatial pooling, which decreases the valuable spatial information that helps to precisely localize target objects.

Note that fine-tuning a deep network with the initial frame would easily result in over-fitting due to limited samples in one image. In contrast, we aim to identify the target-aware features. We use the given target information to compute gradients of the proposed losses and identify the target-specific channels according to the gradients, rather than using the given target information to fine-tune a network directly. In addition, a classification network usually needs a large feature space to separate all possible (hundreds to thousands) classes, which is redundant for tracking as only two classes (foreground and background) are essential for a test sequence. As such, we propose to exploit features specific to the target of interest in a test sequence.

References

[1] L. Bertinetto, J. Valmadre, S. Golodetz, O. Miksik, and P. H. Torr. Staple: Complementary learners for real-time tracking. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2016. 1

[2] L. Bertinetto, J. Valmadre, J. F. Henriques, A. Vedaldi, and P. H. Torr. Fully-convolutional siamese networks for object tracking. In *European Conference on Computer Vision Workshops*, 2016. 1

[3] B. Chen, D. Wang, P. Li, S. Wang, and H. Lu. Real-time actor-critictracking. In *European Conference on Computer Vision*, 2018. 1

[4] J. Choi, H. J. Chang, T. Fischer, S. Yun, K. Lee, J. Jeong, Y. Demiris, and J. Y. Choi. Context-aware deep feature compression for high-speed visual tracking. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2018. 1

[5] M. Danelljan, G. Bhat, F. S. Khan, and M. Felsberg. Eco: Efficient convolution operators for tracking. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2017. 1

[6] M. Danelljan, A. Robinson, F. S. Khan, and M. Felsberg. Beyond correlation filters: Learning continuous convolution operators for visual tracking. In *European Conference on Computer Vision*, 2016. 1

[7] X. Dong and J. Shen. Triplet loss in siamese network for object tracking. In *European Conference on Computer Vision*, 2018. 1

[8] Q. Guo, W. Feng, C. Zhou, R. Huang, L. Wan, and S. Wang. Learning dynamic siamese network for visual object tracking. In *IEEE International Conference on Computer Vision*, 2017. 1

[9] A. He, C. Luo, X. Tian, and W. Zeng. A twofold siamese network for real-time object tracking. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2018. 1

[10] H. Kiani Galoogahi, A. Fagg, and S. Lucey. Learning background-aware correlation filters for visual tracking. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2017. 1

[11] B. Li, J. Yan, W. Wu, Z. Zhu, and X. Hu. High performance visual tracking with siamese region proposal network. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2018. 1

[12] F. Li, C. Tian, W. Zuo, L. Zhang, and M.-H. Yang. Learning spatial-temporal regularized correlation filters for visual tracking. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2018. 1

[13] X. Lu, C. Ma, B. Ni, X. Yang, I. Reid, and M.-H. Yang. Deep regression tracking with shrinkage loss. In *European Conference on Computer Vision*, 2018. 1

[14] E. Park and A. C. Berg. Meta-tracker: Fast and robust online adaptation for visual object trackers. In *European Conference on Computer Vision*, 2018. 1

[15] S. Pu, Y. Song, C. Ma, H. Zhang, and M.-H. Yang. Deep attentive tracking via reciprocative learning. In *Annual Conference on Neural Information Processing Systems*, 2018. 1

[16] Y. Song, C. Ma, L. Gong, J. Zhang, R. W. Lau, and M.-H. Yang. Crest: Convolutional residual learning for visual tracking. In *IEEE International Conference on Computer Vision*, pages 2574–2583, 2017. 1

[17] Y. Song, C. Ma, X. Wu, L. Gong, L. Bao, W. Zuo, C. Shen, R. W. Lau, and M.-H. Yang. Vital: Visual tracking via adversarial learning. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2018. 1

[18] C. Sun, H. Lu, and M.-H. Yang. Learning spatial-aware regressions for visual tracking. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2018. 1

[19] C. Sun, D. Wang, H. Lu, and M.-H. Yang. Correlation tracking via joint discrimination and reliability learning. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2018. 1

[20] J. Valmadre, L. Bertinetto, J. Henriques, A. Vedaldi, and P. H. Torr. End-to-end representation learning for correlation filter based tracking. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2017. 1

[21] N. Wang, W. Zhou, Q. Tian, R. Hong, M. Wang, and H. Li. Multi-cue correlation filters for robust visual tracking. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2018. 1

[22] Q. Wang, Z. Teng, J. Xing, J. Gao, W. Hu, and S. Maybank. Learning attentions: residual attentional siamese network for high performance online visual tracking. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2018. 1

[23] Y. Wu, J. Lim, and M.-H. Yang. Online object tracking: A benchmark. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2013. 1

[24] Y. Wu, J. Lim, and M.-H. Yang. Object tracking benchmark. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(9):1834–1848, 2015. 1

[25] M. Zhang, Q. Wang, J. Xing, J. Gao, P. Peng, W. Hu, and S. Maybank. Visual tracking via spatially aligned correlation filters network. In *European Conference on Computer Vision*, 2018. 1

[26] T. Zhang, C. Xu, and M.-H. Yang. Multi-task correlation particle filter for robust object tracking. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2017. 1

[27] Z. Zhu, Q. Wang, B. Li, W. Wu, J. Yan, and W. Hu. Distractor-aware siamese networks for visual object tracking. In *European Conference on Computer Vision*, 2018. 1

[28] Z. Zhu, W. Wu, W. Zou, and J. Yan. End-to-end flow correlation tracking with spatial-temporal attention. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2018. 1