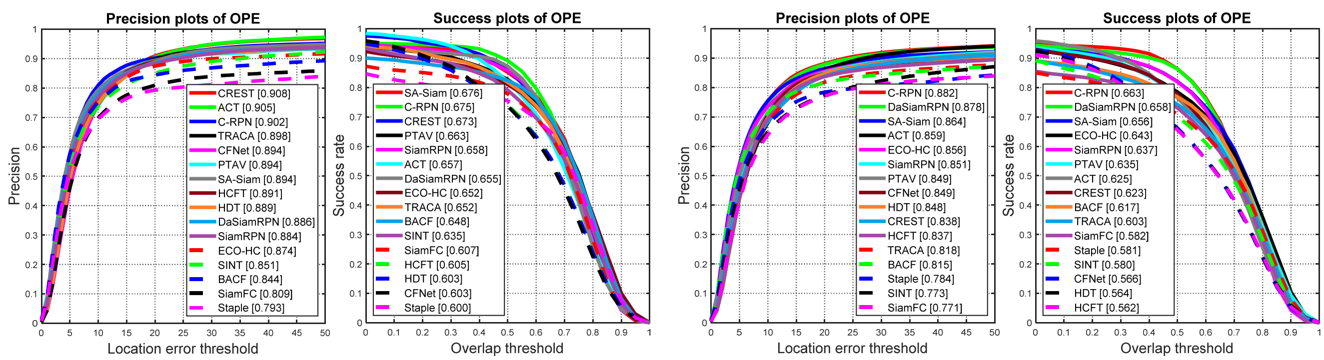
# Siamese Cascaded Region Proposal Networks for Real-Time Visual Tracking
# — Supplementary Material —

Heng Fan     Haibin Ling
Department of Computer and Information Sciences, Temple University, Philadelphia, PA USA
{hengfan,hbling}temple.edu

## 1. Detailed experimental results on OTB-2013 and OTB-2015

In this section we show the detailed results on OTB-2013 [19] and OTB-2015 [20]. Following [20], we report the results in *one-pass evaluation* (OPE) using *precision* and *success* plots. The comparison of C-RPN with 15 trackers (SiamRPN [12], DaSiamRPN [21], TRACA [4], ACT [3], BACF [8], ECO-HC [5], CREST [16], SiamFC [2], Staple [1], PTAV [7], SINT [17], SA-Siam [9], CFNet [18], HDT [15] and HCFT [13]) is shown in Fig. 1. Tab. 1 summarizes the comparison.




(a) Comparison on OTB-2013 (b) Comparison on OTB-2015

Figure 1. Comparison of C-RPN with 14 state-of-the-art trackers on OTB-2013 [19] and OTB-2015 [20]. Best viewed in color.

Table 1. Performance of different trackers on OTB-2013 [19] and OTB-2015 [20]. The best two results are in red and blue font, respectively.

| Tracker | Where | OTB-2013 | | OTB-2015 | | Speed (*fps*) |
|---|---|---|---|---|---|---|
| | | PRE | SUC | PRE | SUC | |
| C-RPN | - | 0.902 | 0.675 | 0.882 | 0.663 | 36 |
| DaSiamRPN [21] | ECCV'2018 | 0.886 | 0.655 | 0.878 | 0.658 | 160 |
| ACT [3] | ECCV'2018 | 0.905 | 0.657 | 0.859 | 0.625 | 30 |
| SiamRPN [12] | CVPR'2018 | 0.884 | 0.658 | 0.856 | 0.637 | 160 |
| SA-Siam [9] | CVPR'2018 | 0.894 | 0.676 | 0.864 | 0.656 | 50 |
| TRACA [4] | CVPR'2018 | 0.898 | 0.652 | 0.851 | 0.603 | 101 |
| PTAV [7] | ICCV'2017 | 0.894 | 0.663 | 0.849 | 0.635 | 25 |
| BACF [8] | ICCV'2017 | 0.844 | 0.648 | 0.815 | 0.617 | 35 |
| CREST [16] | ICCV'2017 | 0.908 | 0.673 | 0.838 | 0.623 | 5 |
| ECO-HC [5] | CVPR'2017 | 0.874 | 0.652 | 0.856 | 0.643 | 60 |
| CFNet [18] | CVPR'2017 | 0.785 | 0.603 | 0.777 | 0.566 | 75 |
| SiamFC [2] | ECCVW'2016 | 0.809 | 0.607 | 0.771 | 0.582 | 86 |
| SINT [17] | CVPR'2016 | 0.851 | 0.635 | 0.773 | 0.580 | 2 |
| Staple [1] | CVPR'2016 | 0.793 | 0.600 | 0.784 | 0.581 | 80 |
| HDT [15] | CVPR'2016 | 0.889 | 0.603 | 0.848 | 0.564 | 10 |
| HCFT [13] | ICCV'2015 | 0.891 | 0.605 | 0.837 | 0.562 | 11 |

In addition, we also show the attribute evaluation on OTB-2013 in Fig. 2 and on OTB-2015 in Fig. 3.

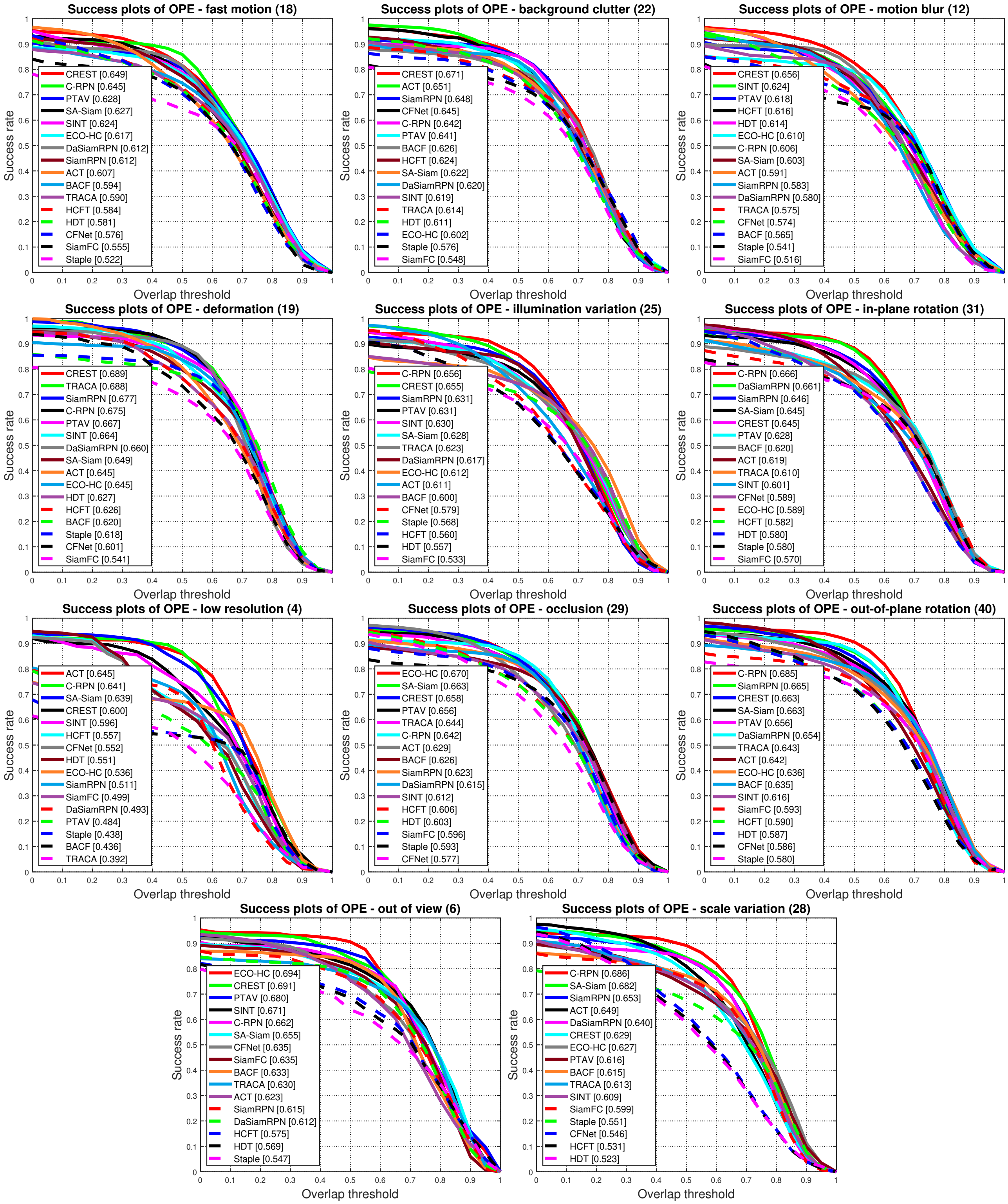



Figure 2. Evaluation on 11 attributes using SUC on OTB-2013 [19], including fast motion (FM), background cluttered (BC), motion blur (MB), deformation (DEF), illumination variation (IV), in-plane rotation (IPR), low resolution (LR), occlusion (OCC), out-of-plane rotation (OPR), out-of-view (OV) and scale variation (SV).

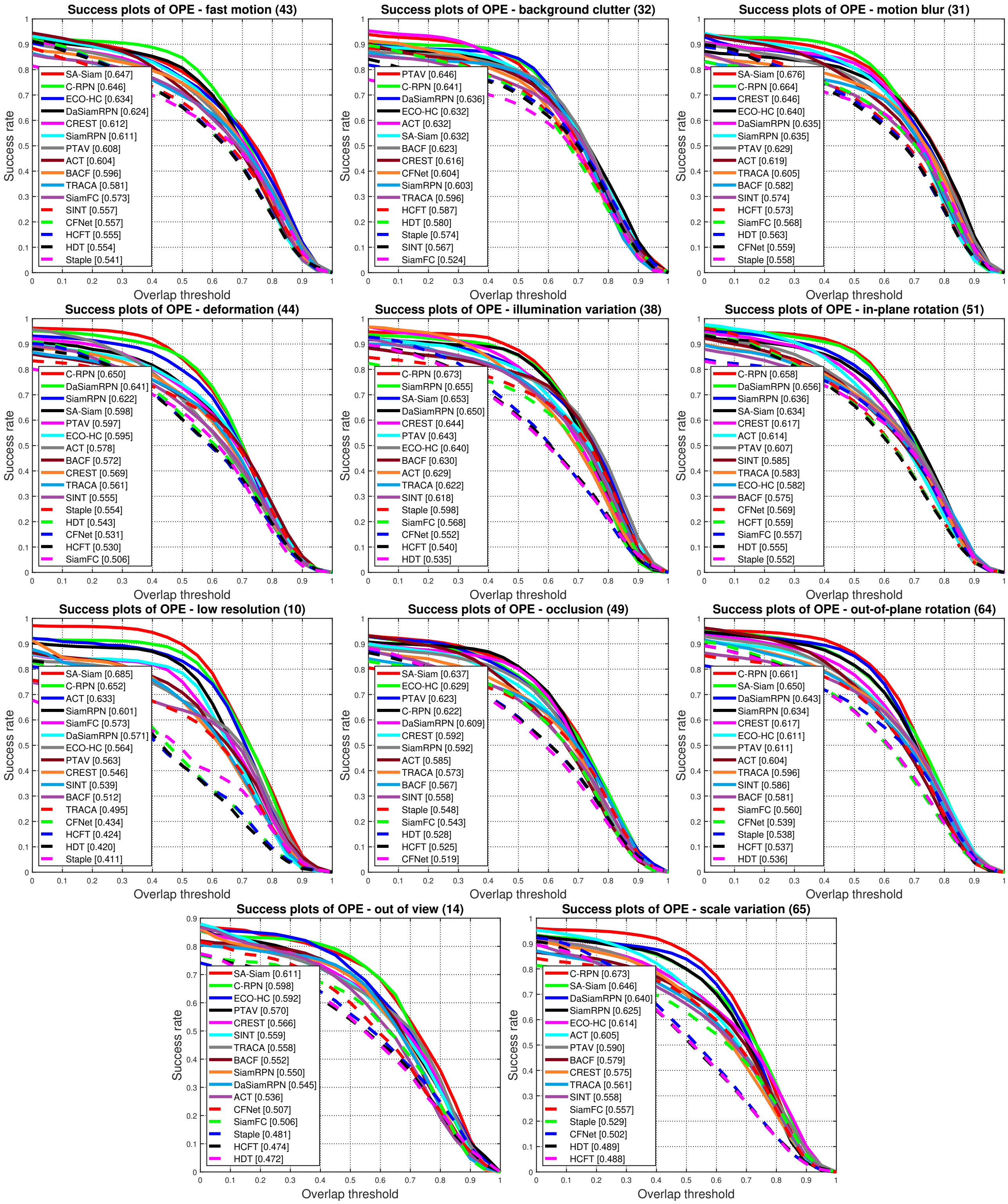



Figure 3. Evaluation on 11 attributes using SUC on OTB-2015 [20], including fast motion (FM), background cluttered (BC), motion blur (MB), deformation (DEF), illumination variation (IV), in-plane rotation (IPR), low resolution (LR), occlusion (OCC), out-of-plane rotation (OPR), out-of-view (OV) and scale variation (SV).

## 2. Detailed experimental results on VOT-2016 and VOT-2017

In this section, we show the full results for baseline experiment on VOT-2016 [10] in Fig. 4. Fig. 5 and Fig. 6 show the results for baseline and real-time experiments on VOT-2017 [11], respectively.

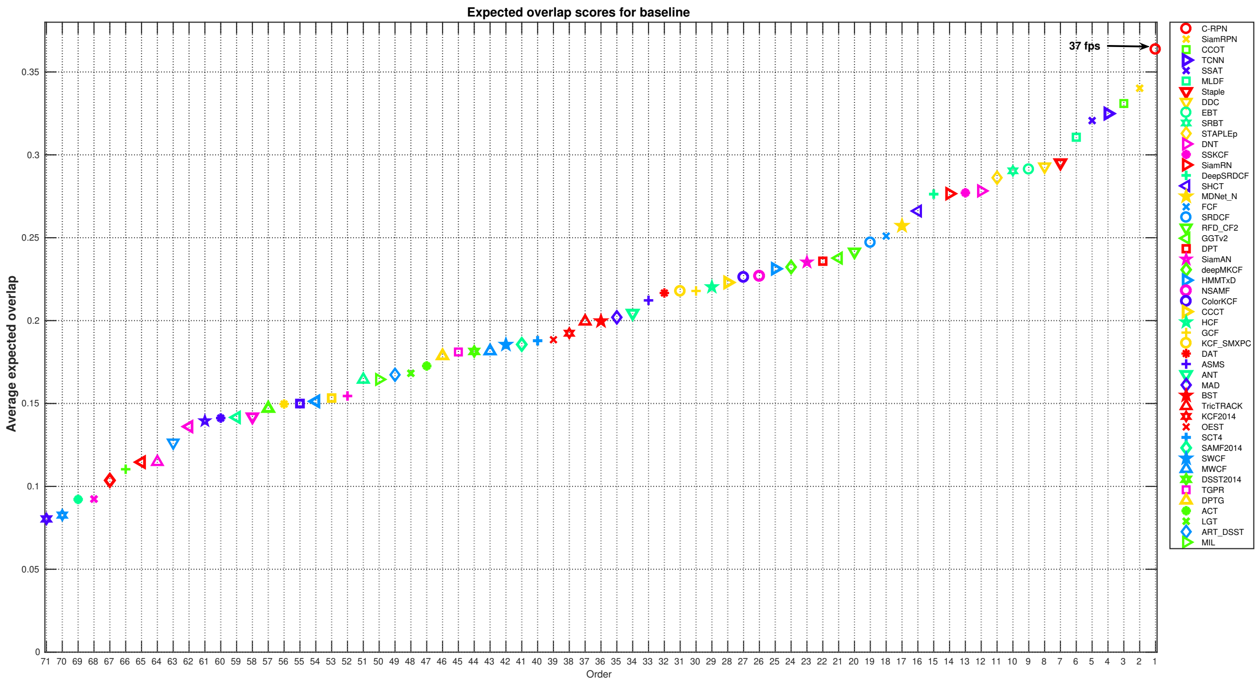



Figure 4. Results for baseline experiment on VOT-2016 [10]. Best viewed in color and when zoomed-in. Our C-RPN ranks the first.

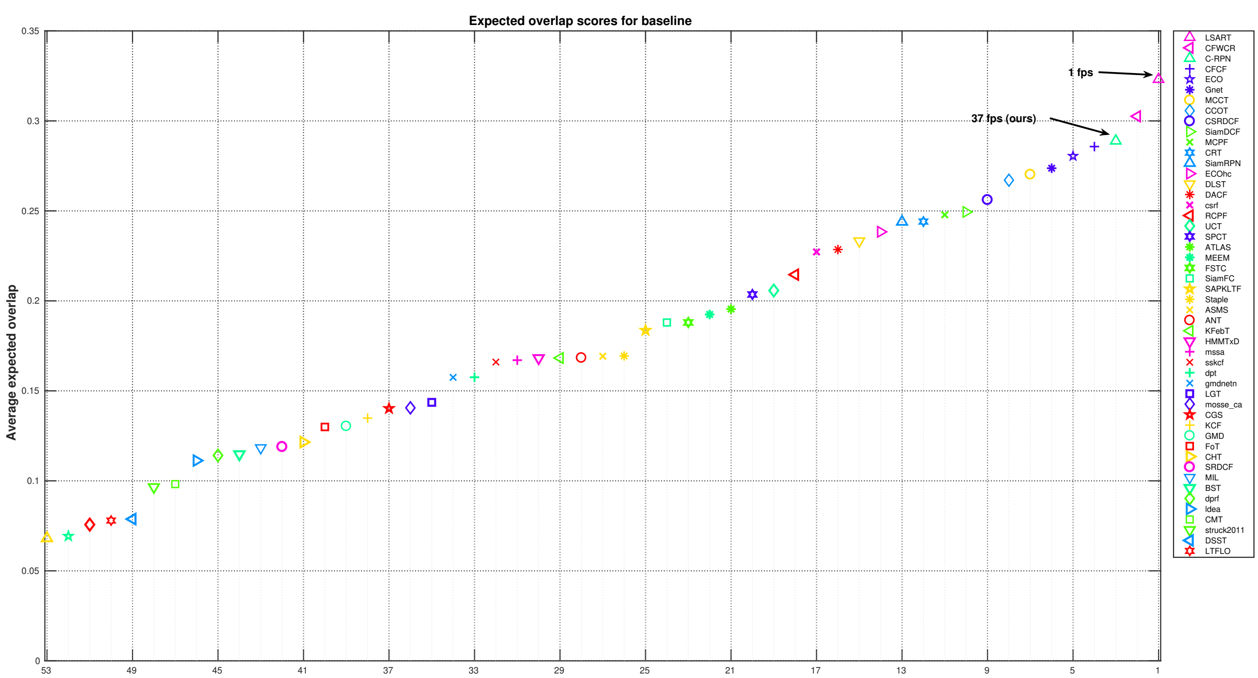



Figure 5. Results for baseline experiment on VOT-2017 [11]. Best viewed in color and when zoomed-in. Our C-RPN shows competitive performance.

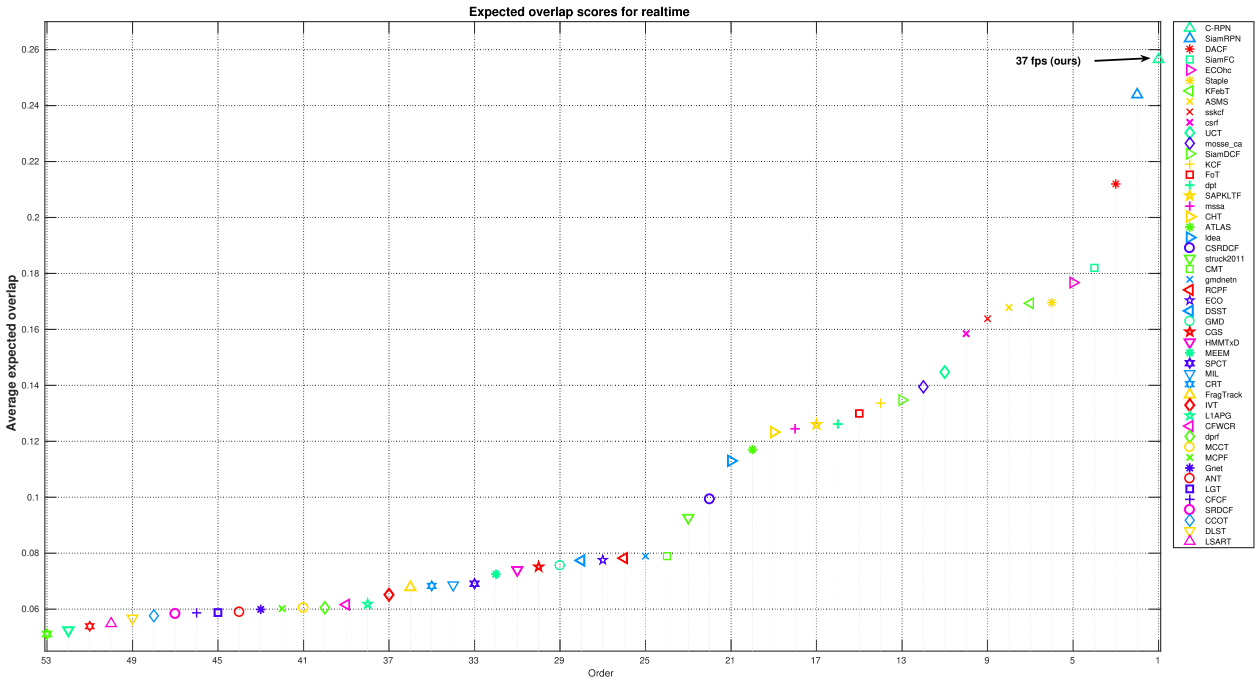
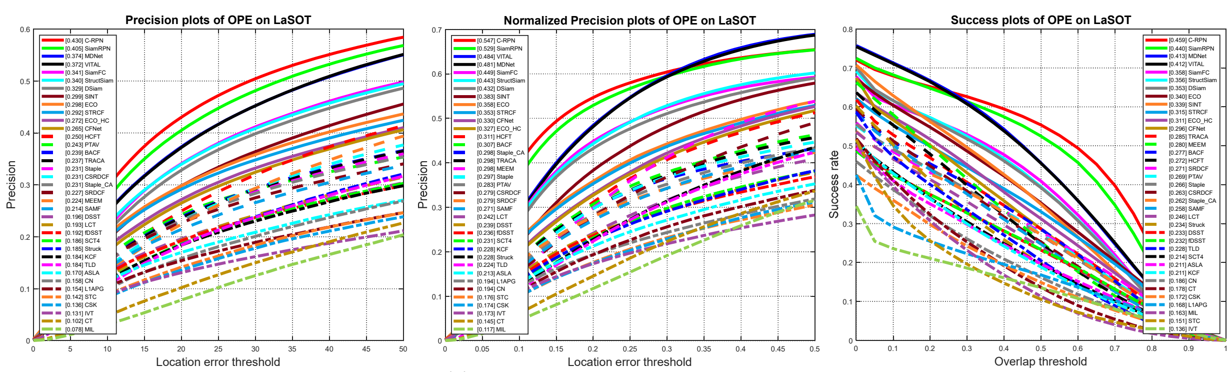


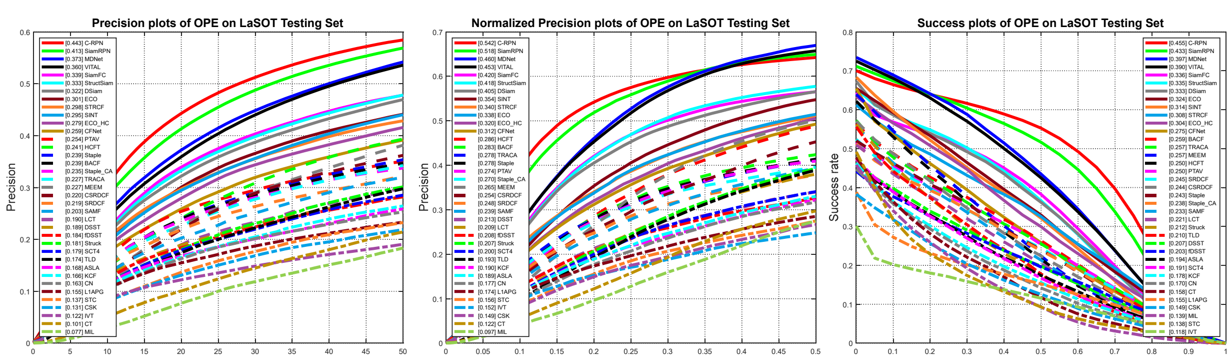

Figure 6. Results for real-time experiment on VOT-2017 [11]. Best viewed in color and when zoomed-in. Our C-RPN ranks the first.

## 3. Detailed experimental results on LaSOT

Fig. 7 demonstrates the results under Protocol I and II. Fig. 8 and 9 show the attribution evaluation under two protocols.




(a) Comparison on LaSOT under Protocol I




(b) Comparison on LaSOT under Protocol II

Figure 7. Comparison of C-RPN with 35 state-of-the-art trackers on LaSOT [6]. Best viewed in color and when zoomed-in.

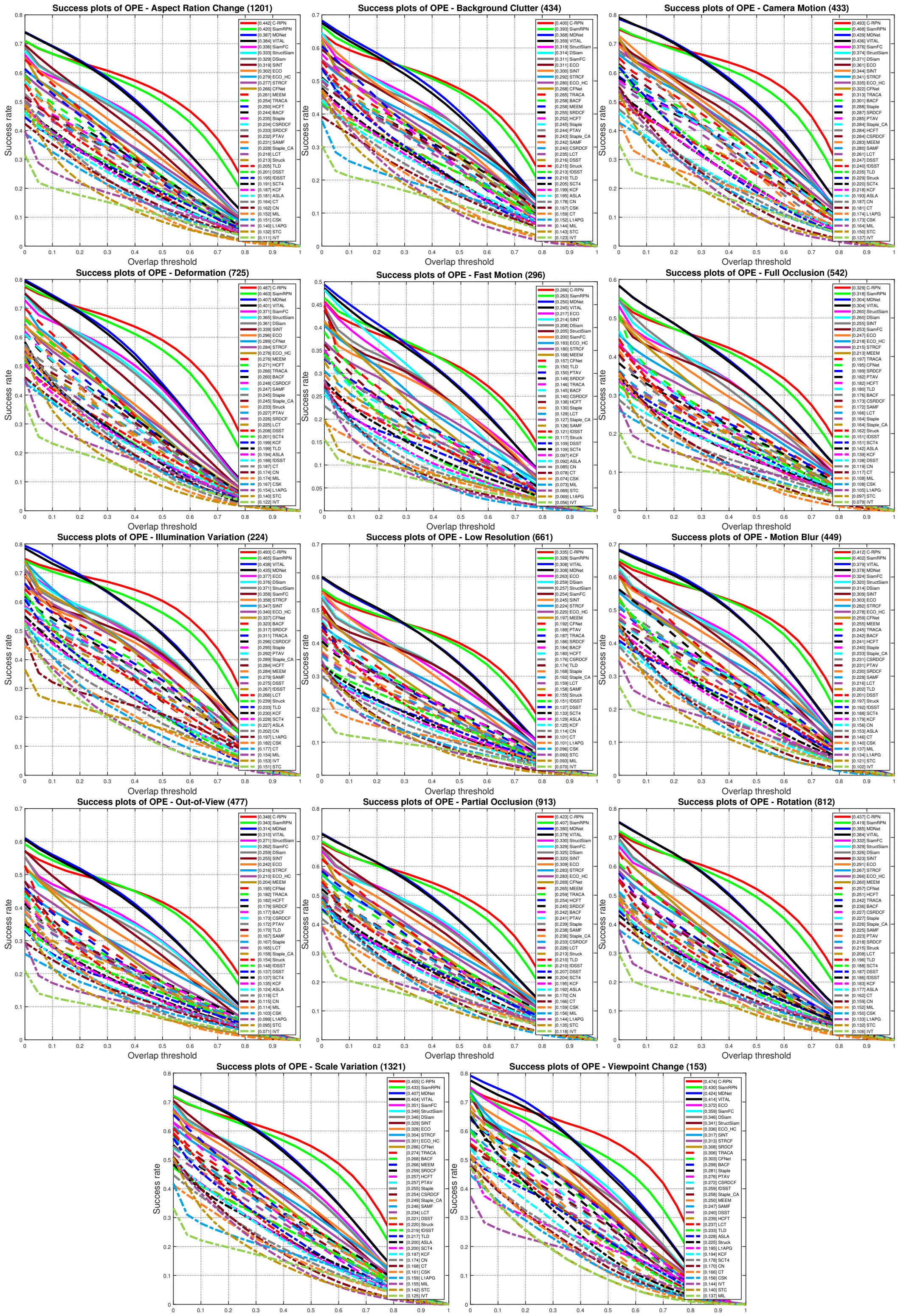



Figure 8. Evaluation on 14 attributes under Protocol I using SUC on LaSOT [20], including illumination variation (IV), full occlusion (FOC), partial occlusion (POC), deformation (DEF), motion blur (MB), fast motion (FM), scale variation (SV), camera motion (CM), rotation (ROT), background clutter (BC), low resolution (LR), viewpoint change (VC), out-of-view (OV) and aspect ration change (ARC).

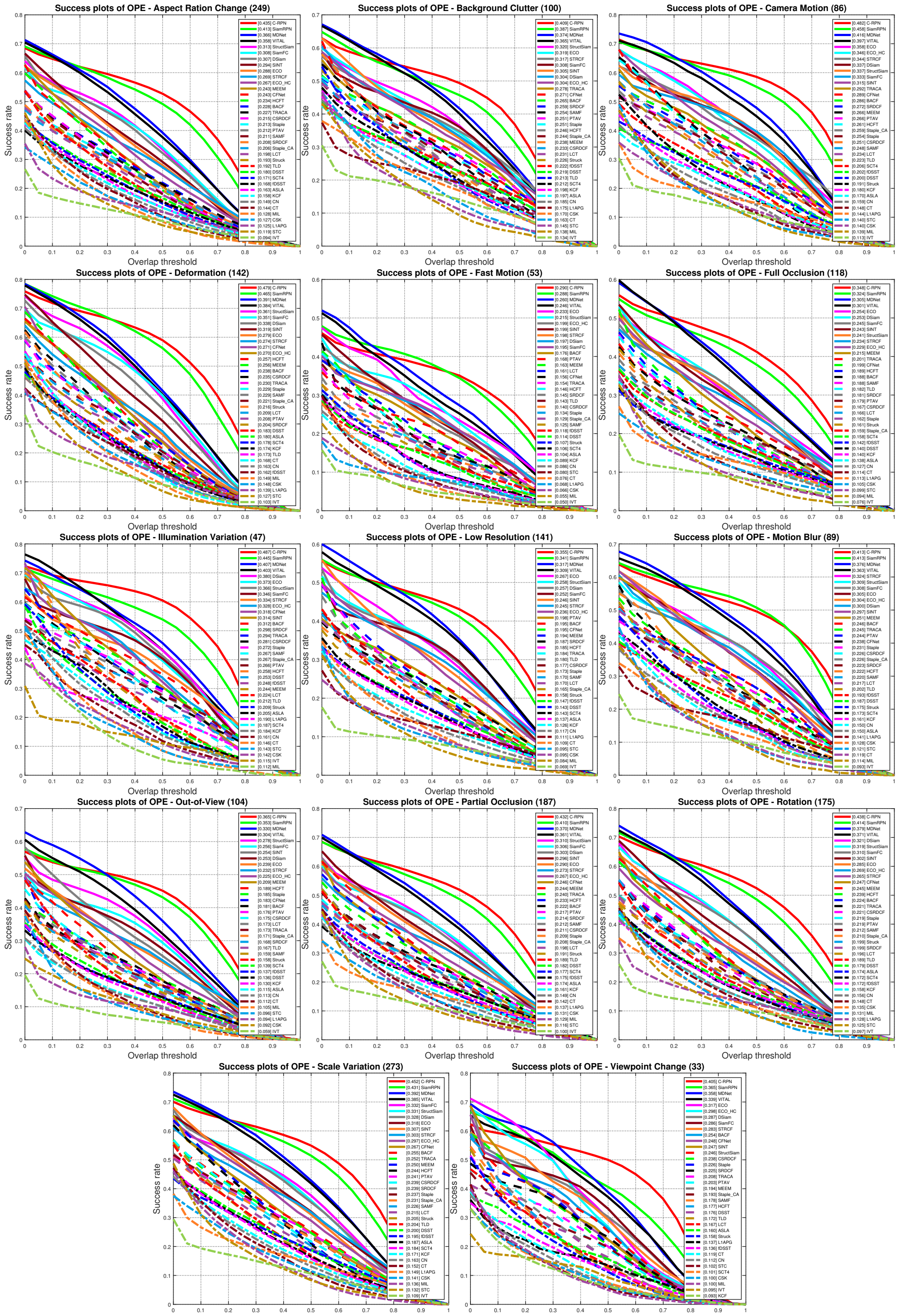



Figure 9. Evaluation on 14 attributes under Protocol II using SUC on LaSOT [20], including illumination variation (IV), full occlusion (FOC), partial occlusion (POC), deformation (DEF), motion blur (MB), fast motion (FM), scale variation (SV), camera motion (CM), rotation (ROT), background clutter (BC), low resolution (LR), viewpoint change (VC), out-of-view (OV) and aspect ration change (ARC).

Moreover, we also show some qualitative results of the proposed C-RPN and comparison to other state-of-the-art trackers on LaSOT [6], as shown in Fig. 10.


Figure 10. Qualitative comparison on ten challenging sequences (from top to bottom: *bicycle-18*, *bird-3*, *boat-3*, *bus-17*, *dog-7*, *horse-1*, *leopard-7*, *person-1*, *turtle-16* and *kite-4*) of LaSOT under the Protocol II. Best viewed in color.

## 4. Detailed experimental results on TrackingNet

This section shows the result of our C-RPN on TrackingNet [14]. Note that the result is evaluated by the server provided by the organizer at http://eval.tracking-net.org/web/challenges/challenge-page/39/leaderboard/

42. Here we list the full comparison of our method with state-of-the-art trackers in Tab. 2. The results of other trackers are reported from the original paper [14].

Table 2. Comparisons on TrackingNet [14] with the best two results highlighted in **red** and **blue** fonts, respectively.

| | Precision | Norm. Precision | Success |
|---|---|---|---|
| C-RPN (Ours) | **0.619** | **0.746** | **0.669** |
| MDNet | **0.565** | **0.705** | **0.606** |
| CFNet | 0.533 | 0.654 | 0.578 |
| SiamFC | 0.533 | 0.663 | 0.571 |
| ECO | 0.492 | 0.618 | 0.554 |
| CSRDCF | 0.48 | 0.622 | 0.534 |
| SAMF | 0.477 | 0.598 | 0.504 |
| ECO-HC | 0.476 | 0.608 | 0.541 |
| Staple | 0.470 | 0.603 | 0.528 |
| Staple_CA | 0.468 | 0.605 | 0.529 |
| BACF | 0.461 | 0.580 | 0.523 |
| DSST | 0.460 | 0.588 | 0.464 |
| SRDCF | 0.455 | 0.573 | 0.521 |
| SAMF_AT | 0.447 | 0.560 | 0.472 |
| DCF | 0.419 | 0.548 | 0.448 |
| KCF | 0.419 | 0.546 | 0.447 |
| DLSSVM | 0.418 | 0.562 | 0.470 |
| ASLA | 0.406 | 0.536 | 0.478 |
| Struck | 0.402 | 0.539 | 0.456 |
| MEEM | 0.386 | 0.545 | 0.460 |
| CSK | 0.368 | 0.503 | 0.429 |
| IVT | 0.336 | 0.460 | 0.417 |
| MOSSE | 0.326 | 0.442 | 0.388 |
| TLD | 0.292 | 0.438 | 0.400 |

## References


[1] Luca Bertinetto, Jack Valmadre, Stuart Golodetz, Ondrej Miksik, and Philip HS Torr. Staple: Complementary learners for real-time tracking. In *CVPR*, 2016. 1

[2] Luca Bertinetto, Jack Valmadre, Joao F Henriques, Andrea Vedaldi, and Philip HS Torr. Fully-convolutional siamese networks for object tracking. In *ECCVW*, 2016. 1

[3] Boyu Chen, Dong Wang, Peixia Li, Shuang Wang, and Huchuan Lu. Real-time actor-critictracking. In *ECCV*, 2018. 1

[4] Jongwon Choi, Hyung Jin Chang, Tobias Fischer, Sangdoo Yun, Kyuewang Lee, Jiyeoup Jeong, Yiannis Demiris, and Jin Young Choi. Context-aware deep feature compression for high-speed visual tracking. In *CVPR*, 2018. 1

[5] Martin Danelljan, Goutam Bhat, Fahad Shahbaz Khan, Michael Felsberg, et al. Eco: Efficient convolution operators for tracking. In *CVPR*, 2017. 1

[6] Heng Fan, Liting Lin, Fan Yang, Peng Chu, Ge Deng, Sijia Yu, Hexin Bai, Yong Xu, Chunyuan Liao, and Haibin Ling. Lasot: A high-quality benchmark for large-scale single object tracking. *arXiv*, 2018. 5, 8

[7] Heng Fan and Haibin Ling. Parallel tracking and verifying: A framework for real-time and high accuracy visual tracking. In *ICCV*, 2017. 1

[8] Hamed Kiani Galoogahi, Ashton Fagg, and Simon Lucey. Learning background-aware correlation filters for visual tracking. In *ICCV*, 2017. 1

[9] Anfeng He, Chong Luo, Xinmei Tian, and Wenjun Zeng. A twofold siamese network for real-time object tracking. In *CVPR*, 2018. 1

[10] Matej Kristan et al. The visual object tracking vot2016 challenge results. In *ECCVW*, 2016. 4

[11] Matej Kristan et al. The visual object tracking vot2016 challenge results. In *ICCVW*, 2017. 4, 5

[12] Bo Li, Junjie Yan, Wei Wu, Zheng Zhu, and Xiaolin Hu. High performance visual tracking with siamese region proposal network. In *CVPR*, 2018. 1

[13] Chao Ma, Jia-Bin Huang, Xiaokang Yang, and Ming-Hsuan Yang. Hierarchical convolutional features for visual tracking. In *ICCV*, 2015. 1

[14] Matthias Müller, Adel Bibi, Silvio Giancola, Salman Al-Subaihi, and Bernard Ghanem. Trackingnet: A large-scale dataset and benchmark for object tracking in the wild. In *ECCV*, 2018. 8, 9

[15] Yuankai Qi, Shengping Zhang, Lei Qin, Hongxun Yao, Qingming Huang, Jongwoo Lim, and Ming-Hsuan Yang. Hedged deep tracking. In *CVPR*, 2016. 1

[16] Yibing Song, Chao Ma, Lijun Gong, Jiawei Zhang, Rynson WH Lau, and Ming-Hsuan Yang. Crest: Convolutional residual learning for visual tracking. In *ICCV*, 2017. 1

[17] Ran Tao, Efstratios Gavves, and Arnold WM Smeulders. Siamese instance search for tracking. In *CVPR*, 2016. 1

[18] Jack Valmadre, Luca Bertinetto, João Henriques, Andrea Vedaldi, and Philip HS Torr. End-to-end representation learning for correlation filter based tracking. In *CVPR*, 2017. 1

[19] Yi Wu, Jongwoo Lim, and Ming-Hsuan Yang. Online object tracking: A benchmark. In *CVPR*, 2013. 1, 2

[20] Yi Wu, Jongwoo Lim, and Ming-Hsuan Yang. Object tracking benchmark. *TPAMI*, 37(9):1834–1848, 2015. 1, 3, 6, 7

[21] Zheng Zhu, Qiang Wang, Bo Li, Wei Wu, Junjie Yan, and Weiming Hu. Distractor-aware siamese networks for visual object tracking. In *ECCV*, 2018. 1