Supplementary Material of LSOTB-TIR

Qiao Liu
Xin Li
Harbin Institute of Technology, Shenzhen
liuqiao.hit@gmail.com

Zhenyu He*
Harbin Institute of Technology, Shenzhen
Peng Cheng Laboratory
zhenyuhe@hit.edu.cn

Chenglong Li
Anhui University
chenglongli@ahu.edu.cn

Jun Li
Zikun Zhou
Di Yuan
Harbin Institute of Technology, Shenzhen

Jing Li
Kai Yang
Nana Fan
Harbin Institute of Technology, Shenzhen

Feng Zheng
Southern University of Science and Technology
zfeng02@gmail.com

ACM Reference Format:
Qiao Liu, Xin Li, Zhenyu He, Chenglong Li, Jun Li, Zikun Zhou, Di Yuan, Jing Li, Kai Yang, Nana Fan, and Feng Zheng. 2020. Supplementary Material of LSOTB-TIR. In *Proceedings of the 28th ACM International Conference on Multimedia (MM '20), October 12–16, 2020, Seattle, WA, USA.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3394171.3413922

1 PROPOSED LSOTB-TIR BENCHMARK

Comparison with more tracking benchmarks. Table 1 compares the proposed LSOTB-TIR with the existing TIR and RGB tracking benchmarks. It shows that our benchmark is not only the largest and the most diverse in all TIR tracking benchmarks but also larger and more diverse than most RGB tracking benchmarks.

Statistic information of LSOTB-TIR. Table 2 shows some statistic data, such as the object class, number of instances of each class, and number of bounding boxes of each class, of the proposed LSOTB-TIR benchmark.

The designed label tool. Fig. 1 shows the proposed semi-automatic label tool (Codes are available in 'Codes' folder). We just need to adjust the bounding box position at current frame and then run the ECO-HC tracker to automatic obtain the annotations of next ten frames. We suggest that the label tool makes the annotation more accurate, smoother, and faster than the annotation in each frame manually.

Training dataset examples. As shown in Fig. 2, we show some annotated frames of all the 47 object classes of the proposed TIR training dataset. We also show some annotated sequences in a video, i.e., 'video-demo.mp4'.

*Corresponding author.

MM '20, October 12–16, 2020, Seattle, WA, USA

ACM ISBN 978-1-4503-7988-5/20/10…$15.00
https://doi.org/10.1145/3394171.3413922

2 EXPERIMENTS

Evaluated trackers. We choose publicly available 33 TIR tracking methods and RGB tracking methods for evaluation on LSOTB-TIR. These methods includes hand-crated feature based sparse representation trackers, such as L1APG [2] and ASLA [48]; hand-crafted feature based correlation filter trackers, such as KCF [17], DSST [8], SRDCF [9], BACF [20], Staple [4], ECO-HC [7], and MCCT [43]; other hand-crafted feature based trackers, such as MIL [1], TGPR [14], LCT [33], Struck [16], and RPT [28]; deep feature based correlation filter trackers, such as HDT [38], ECO [7], HCF [32], DeepSTRCF [25], MCFTS [31], CREST [39], and ECO-stir [49]; matching based deep trackers, such as SiamFC [5], CFNet [41], DSiam [15], SiamFC-tri [11], TADT [27], SiamRPN++ [23], SiamMask [44], UDT [42], HSSNet [26], and MLSSNet [30]; classification based deep trackers, such as MDNet [36], VITAL [40], and ATOM [6]. We do not change the parameters of these trackers provided by the authors in the experiment. Furthermore, we re-train several deep trackers on the proposed training dataset for evaluation, such as ECO-TIR, SiamFC-TIR, CFNet-TIR. We use the backbone network of SiamFC-TIR as the feature extractor in the ECO-TIR tracker.

Attribute-based evaluated results. Fig. 3 shows the evaluated results of all the trackers on the 4 scenario subsets. Figs. 4, 5, and 6 show the success plots, precision plots, and normalized precision plots of all the trackers on the 12 challenge subsets, respectively.

REFERENCES

[1] B. Babenko, Ming Hsuan Yang, and S. Belongie. 2009. Visual tracking with online Multiple Instance Learning. In *IEEE Conference on Computer Vision and Pattern Recognition.*
[2] Chenglong Bao, Yi Wu, Haibin Ling, and Hui Ji. 2012. Real time robust L1 tracker using accelerated proximal gradient approach. In *IEEE Conference on Computer Vision and Pattern Recognition.*
[3] A. Berg, J. Ahlberg, and M. Felsberg. 2015. A thermal Object Tracking benchmark. In *IEEE International Conference on Advanced Video and Signal Based Surveillance.*
[4] Luca Bertinetto, Jack Valmadre, Stuart Golodetz, Ondrej Miksik, and Philip HS Torr. 2016. Staple: Complementary learners for real-time tracking. In *IEEE Conference on Computer Vision and Pattern Recognition.*
[5] Luca Bertinetto, Jack Valmadre, Joao F Henriques, Andrea Vedaldi, and Philip HS Torr. 2016. Fully-Convolutional Siamese Networks for Object Tracking. In *European Conference on Computer Vision Workshops.*
[6] Martin Danelljan, Goutam Bhat, Fahad Shahbaz Khan, and Michael Felsberg. 2019. Atom: Accurate tracking by overlap maximization. In *IEEE Conference on Computer Vision and Pattern Recognition.*
[7] Martin Danelljan, Goutam Bhat, Fahad Shahbaz Khan, and Michael Felsberg. 2017. ECO: efficient convolution operators for tracking. In *IEEE Conference on Computer Vision and Pattern Recognition.*

Table 1: Comparison of the proposed LSOTB-TIR benchmark with existing TIR and RGB tracking benchmarks. The first eight rows are TIR tracking benchmarks and last nine rows are RGB tracking benchmarks.

Benchmarks	Num. of sequences	Max frames	Min frames	Mean frames	Total frames	Frame rates	Object classes	Num. of challenges	Scenario attribute	Training dataset
OSU [10]	6	2,031	601	1,424	8K	30 fps	1	n/a	✗	✗
PDT-ATV [37]	8	775	77	486	4K	20 fps	3	n/a	✗	✗
BU-TIV [47]	16	26,760	150	3,750	60K	30 fps	5	n/a	✗	✗
LTIR [3]	20	1,451	71	563	11K	30 fps	6	6	✗	✗
VOT-TIR2016 [13]	25	1,451	71	555	14K	30 fps	8	6	✗	✗
PTB-TIR [29]	60	1,451	50	502	30K	30 fps	1	9	✗	✗
RGB-T [24]	234	4,000	45	500	117K	30 fps	6	12	✗	✗
LSOTB-TIR (Ours)	1,400	3,056	47	428	600K	30 fps	47	12	✓	✓
OTB13 [45]	51	3,872	71	578	29K	30 fps	10	11	✗	✗
OTB15 [46]	100	3,872	71	590	59K	30 fps	16	11	✗	✗
VOT14 [22]	25	1,210	164	409	10K	30 fps	11	n/a	✗	✗
VOT17 [21]	60	1,500	41	356	21K	30 fps	24	n/a	✗	✗
UAV123 [34]	123	3,085	109	915	113K	30 fps	9	12	✗	✗
Nfs [19]	100	20,665	169	3,830	383K	240 fps	17	9	✗	✗
TrackingNet [35]	30,643	-	-	498	14.43M	30 fps	21	15	✗	✓
LaSOT [12]	1,400	11,397	1,000	2,506	3.52M	30 fps	70	14	✗	✓
GOT-10K [18]	10,000	-	-	150	1.5M	30 fps	563	6	✗	✓

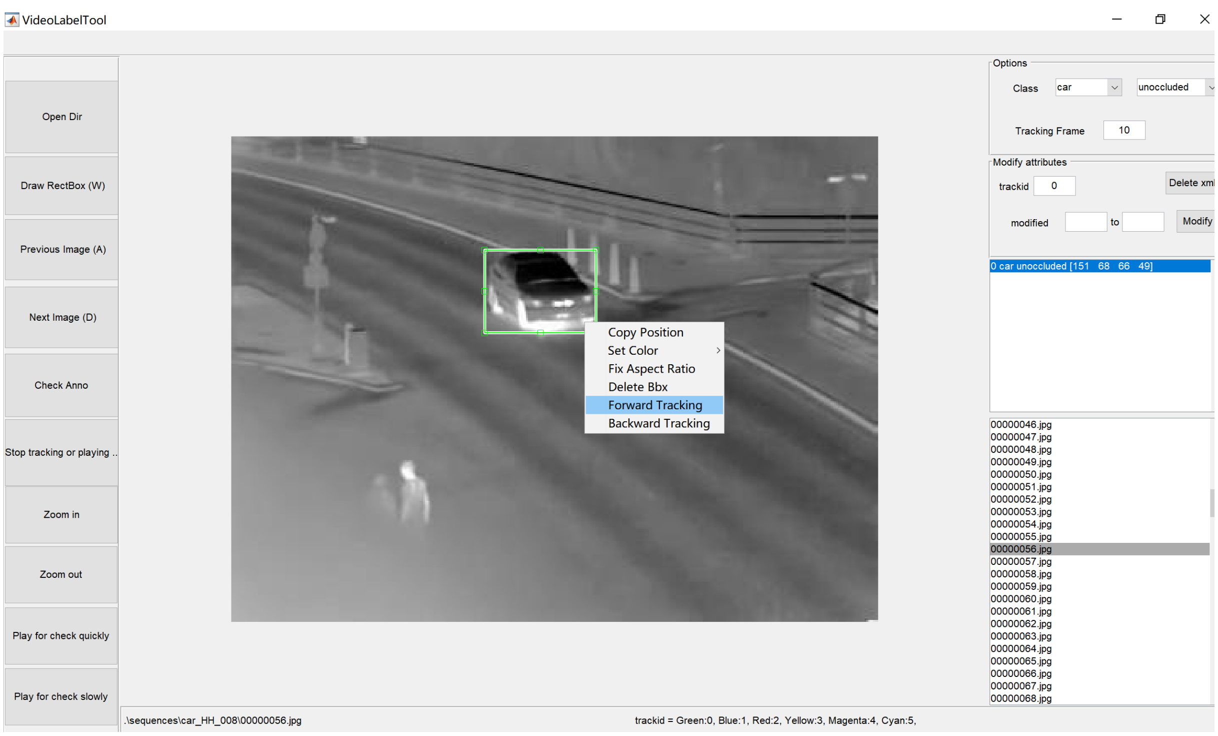


Figure 1: The main window of our designed semi-automatic label tool.

[8] Martin Danelljan, Gustav Häger, Fahad Khan, and Michael Felsberg. 2014. Accurate scale estimation for robust visual tracking. In *British Machine Vision Conference.*

[9] Martin Danelljan, Gustav Hager, Fahad Shahbaz Khan, and Michael Felsberg. 2015. Learning spatially regularized correlation filters for visual tracking. In *IEEE International Conference on Computer Vision.*

[10] James W Davis and Vinay Sharma. 2007. Background-subtraction using contour-based fusion of thermal and visible imagery. *Computer Vision and Image Understanding* 106, 2 (2007), 162–182.

[11] Xingping Dong and Jianbing Shen. 2018. Triplet Loss in Siamese Network for Object Tracking. In *European Conference on Computer Vision.*

[12] Heng Fan, Liting Lin, Fan Yang, Peng Chu, Ge Deng, Sijia Yu, Hexin Bai, Yong Xu, Chunyuan Liao, and Haibin Ling. 2019. Lasot: A high-quality benchmark

Table 2: Statistic information of the proposed LSOTB-TIR, which includes object class, number of instances (Num. of ins.) of each class, and number of bounding boxes (Num. of BBs) of each class.

Category	Training dataset			Evaluation dataset		
	Object class	Num. of ins.	Num. of BBs	Object class	Num. of ins.	Num. of BBs
Aircraft	airplane	31	16,871	airplane	2	1,782
	drone	2	996	drone	1	559
	helicopter	49	23,211	helicopter	3	1,800
Vehicle	bus	13	1,741	bus	6	7,124
	car	116	36,954	car	20	19,835
	motobike	16	6,512	motobike	2	1,760
	tricycle	8	1,361	-	-	-
	truck	9	2,616	truck	1	450
	tank	6	475	-	-	-
Animal	bird	40	12,514	bird	3	1,335
	bat	20	4,305	bat	1	270
	bear	13	2,956	-	-	-
	badger	32	20,224	badger	1	600
	buffalo	48	6,295	-	-	-
	cat	57	17,024	cat	3	1,480
	coyote	85	23,762	coyote	1	475
	cow	51	15,354	cow	1	340
	chicken	26	5,527	-	-	-
	crocodile	4	566	-	-	-
	deer	75	43,244	deer	1	330
	dog	56	23,093	dog	3	1,585
	duck	7	3,243	-	-	-
	elephant	44	17,692	-	-	-
	ericius	8	1,481	-	-	-
	fox	46	36,477	fox	1	620
	goose	10	1,842	-	-	-
	giraffe	31	5,641	-	-	-
	hog	44	15,685	hog	6	3,572
	horse	50	21,023	-	-	-
	hedgehog	4	2,416	-	-	-
	hippo	10	3,754	-	-	-
	hyenas	11	2,060	-	-	-
	lion	69	16,056	-	-	-
	leopard	40	9,645	leopard	1	208
	rat	32	13,886	-	-	-
	monkey	6	1,310	-	-	-
	penguin	4	2,822	-	-	-
	rabbit	24	11,812	-	-	-
	raccoon	31	9,322	-	-	-
	roo	13	6,995	-	-	-
	rhino	21	6,775	-	-	-
	sheep	55	16,543	-	-	-
	squirrel	5	894	-	-	-
	snake	5	2,049	-	-	-
	seal	3	1,049	-	-	-
Boat	boat	63	20,908	boat	2	865
Person	person	362	157,348	person	62	37,348
	-	-	-	head	2	1,135
	-	-	-	face	2	1,740
	Total classes	Total ins.	Total BBs.	Total classes	Total ins.	Total BBs
	47	1,755	654,330	22	120	82,133

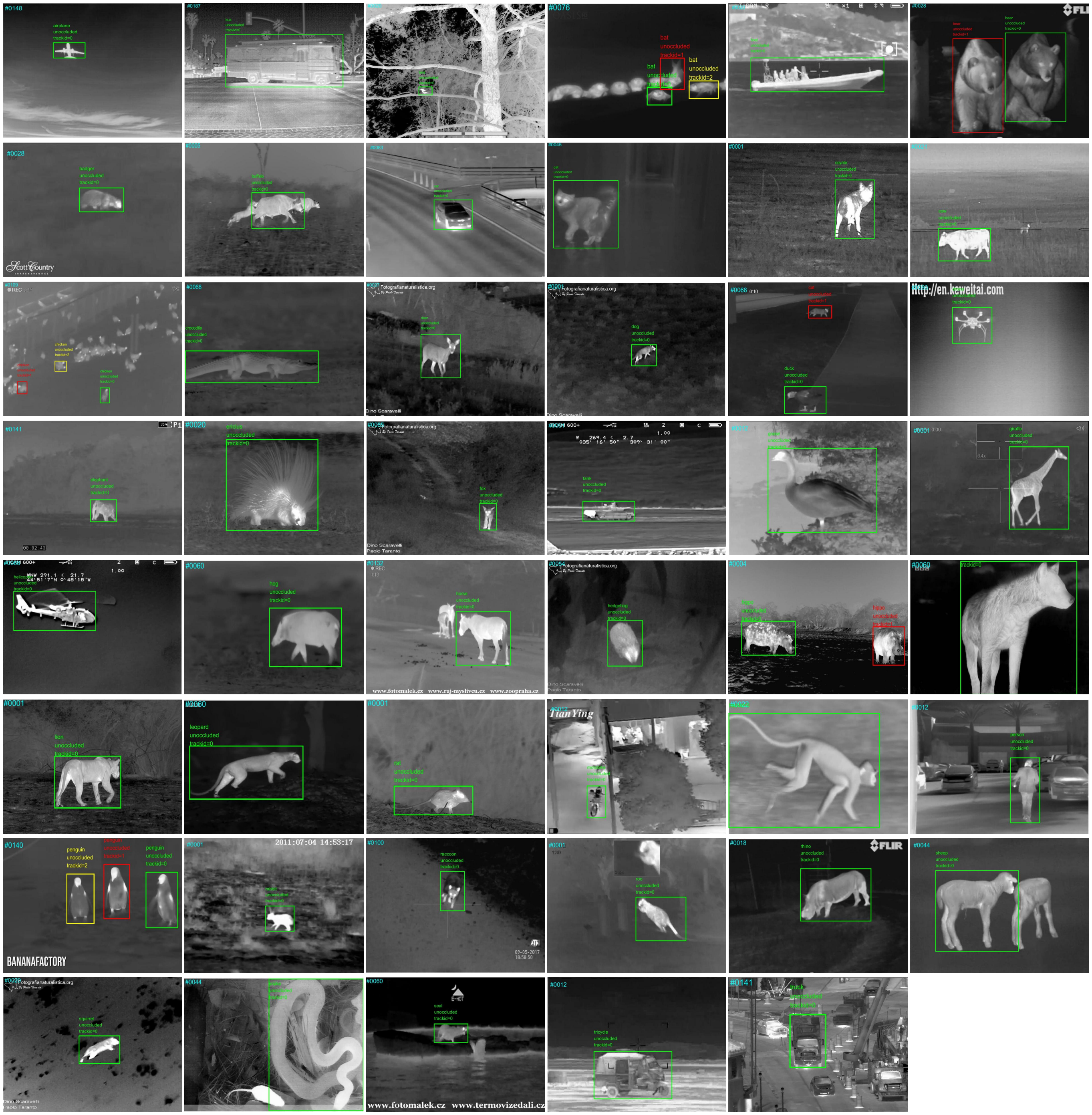


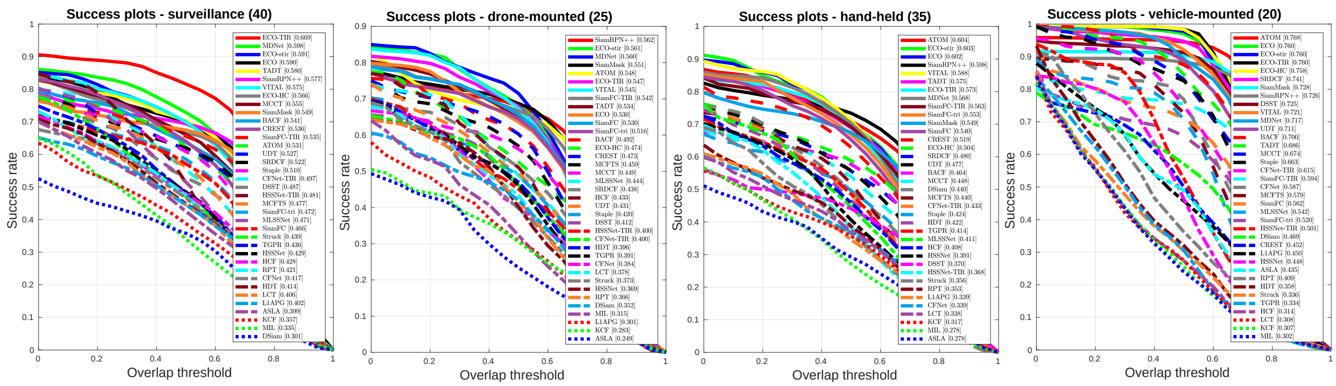
Figure 2: Some samples of the proposed TIR training dataset which consists of 47 classes and 1,280 image sequences. We annotate the classname and bounding box of objects in every frame of each sequence.

for large-scale single object tracking. In *IEEE Conference on Computer Vision and Pattern Recognition.*

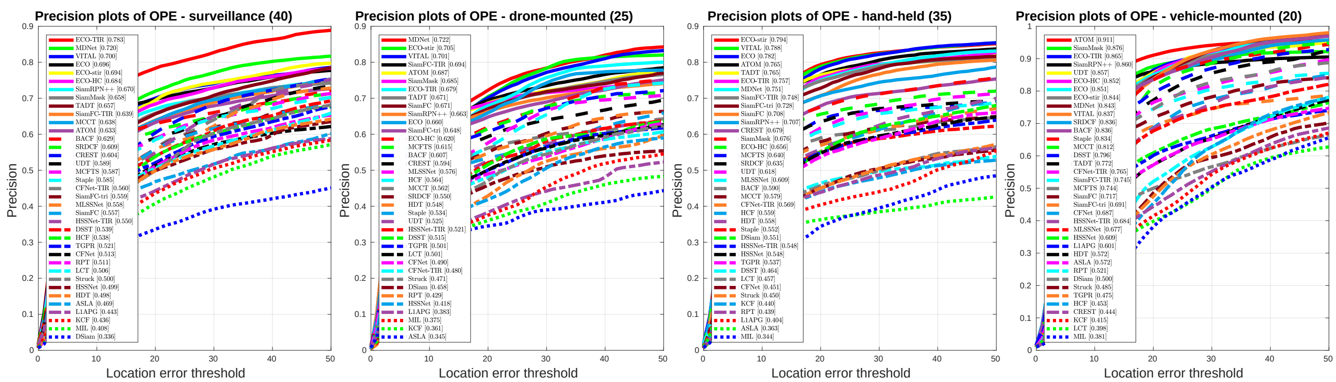
[13] Michael Felsberg, Matej Kristan, Jiři Matas, Aleš Leonardis, et al. 2016. The Thermal Infrared Visual Object Tracking VOT-TIR2016 Challenge Results. In *European Conference on Computer Vision Workshops.*

[14] Jin Gao, Haibin Ling, Weiming Hu, and Junliang Xing. 2014. Transfer learning based visual tracking with gaussian processes regression. In *European Conference on Computer Vision.*

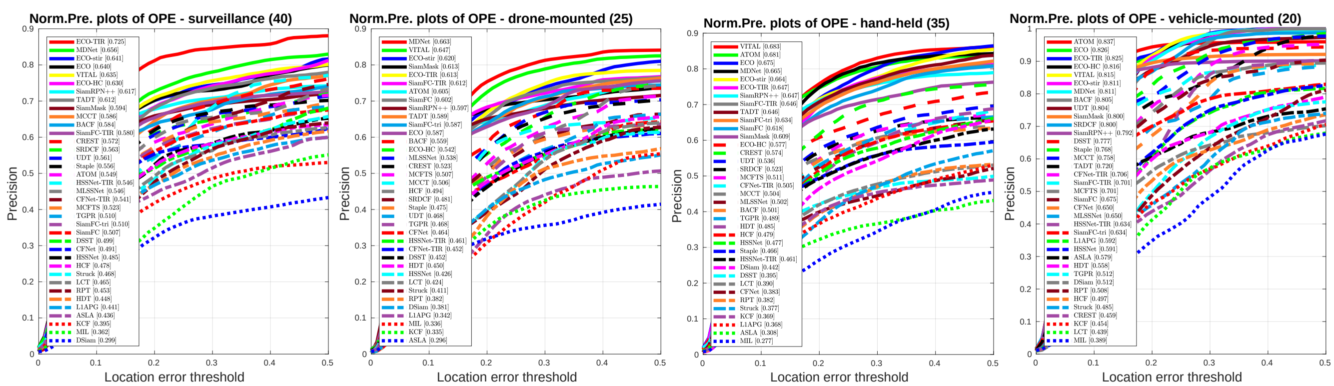
[15] Qing Guo, Wei Feng, Ce Zhou, et al. 2017. Learning dynamic siamese network for visual object tracking. In *IEEE International Conference on Computer Vision.*



(a) Success plots.



(b) Precision plots.



(c) Normalized precision plots.

Figure 3: Evaluation results of all the evaluated trackers on the 4 scenario attribute subsets.

[16] Sam Hare, Stuart Golodetz, Amir Saffari, Vibhav Vineet, et al. 2015. Struck: Structured output tracking with kernels. *IEEE Transactions of Pattern Analysis and Machine Intelligence* 38, 10 (2015), 2096–2109.

[17] João F Henriques, Rui Caseiro, Pedro Martins, and Jorge Batista. 2015. High-speed tracking with kernelized correlation filters. *IEEE Transactions of Pattern Analysis and Machine Intelligence* 37, 3 (2015), 583–596.

[18] Lianghua Huang, Xin Zhao, and Kaiqi Huang. 2018. Got-10k: A large high-diversity benchmark for generic object tracking in the wild. *arXiv preprint arXiv:1810.11981* (2018).

[19] Hamed Kiani Galoogahi, Ashton Fagg, Chen Huang, Deva Ramanan, and Simon Lucey. 2017. Need for speed: A benchmark for higher frame rate object tracking. In *IEEE International Conference on Computer Vision*.

[20] Hamed Kiani Galoogahi, Ashton Fagg, and Simon Lucey. 2017. Learning background-aware correlation filters for visual tracking. In *IEEE International Conference on Computer Vision*.

[21] Matej Kristan, Ales Leonardis, Jiri Matas, Michael Felsberg, et al. 2017. The visual object tracking vot2017 challenge results. In *IEEE International Conference on Computer Vision*.

[22] Matej Kristan, Roman Pflugfelder, Aleš Leonardis, Jiri Matas, et al. 2015. The Visual Object Tracking VOT2014 Challenge Results. In *European Conference on Computer Vision Workshops*.

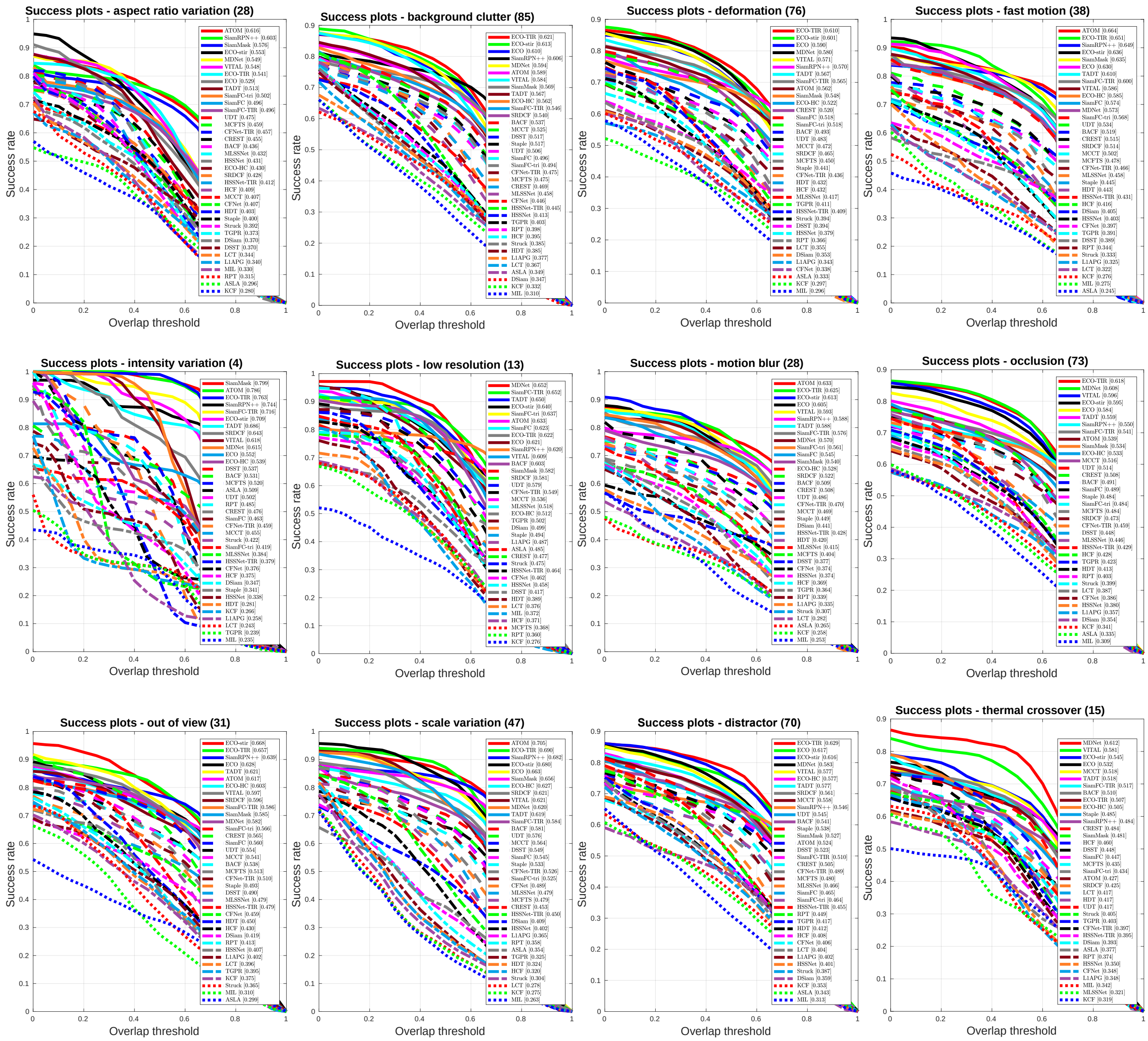


Figure 4: Success plots of all the evaluated trackers on the 12 challenge attribute subsets.

[23] Bo Li, Wei Wu, Qiang Wang, Fangyi Zhang, Junliang Xing, and Junjie Yan. 2019. Siamrpn++: Evolution of siamese visual tracking with very deep networks. In *IEEE Conference on Computer Vision and Pattern Recognition.*

[24] Chenglong Li, Xinyan Liang, Yijuan Lu, Nan Zhao, and Jin Tang. 2019. RGB-T object tracking: benchmark and baseline. *Pattern Recognization* 96 (2019), 106977.

[25] Feng Li, Cheng Tian, Wangmeng Zuo, Lei Zhang, and Ming-Hsuan Yang. 2018. Learning spatial-temporal regularized correlation filters for visual tracking. In *IEEE Conference on Computer Vision and Pattern Recognition.*

[26] Xin Li, Qiao Liu, Nana Fan, et al. 2019. Hierarchical spatial-aware Siamese network for thermal infrared object tracking. *Knowledge-Based Systems* 166 (2019), 71–81.

[27] Xin Li, Chao Ma, Baoyuan Wu, Zhenyu He, and Ming-Hsuan Yang. 2019. Target-Aware Deep Tracking. In *IEEE Conference on Computer Vision and Pattern Recognition.*

[28] Yang Li, Jianke Zhu, and Steven CH Hoi. 2015. Reliable patch trackers: Robust visual tracking by exploiting reliable patches. In *IEEE Conference on Computer Vision and Pattern Recognition.*

[29] Qiao Liu, Zhenyu He, Xin Li, and Yuan Zheng. 2019. PTB-TIR: A Thermal Infrared Pedestrian Tracking Benchmark. *IEEE Transaction on Multimedia* (2019). https://doi.org/10.1109/TMM.2019.2932615

[30] Qiao Liu, Xin Li, Zhenyu He, Nana Fan, Di Yuan, and Hongpeng Wang. 2019. Learning Deep Multi-Level Similarity for Thermal Infrared Object Tracking. *arXiv preprint arXiv:1906.03568* (2019).

[31] Qiao Liu, Xiaohuan Lu, Zhenyu He, et al. 2017. Deep convolutional neural networks for thermal infrared object tracking. *Knowledge-Based Systems* 134 (2017), 189–198.

[32] Chao Ma, Jia-Bin Huang, Xiaokang Yang, and Ming-Hsuan Yang. 2015. Hierarchical convolutional features for visual tracking. In *IEEE International Conference on Computer Vision.*

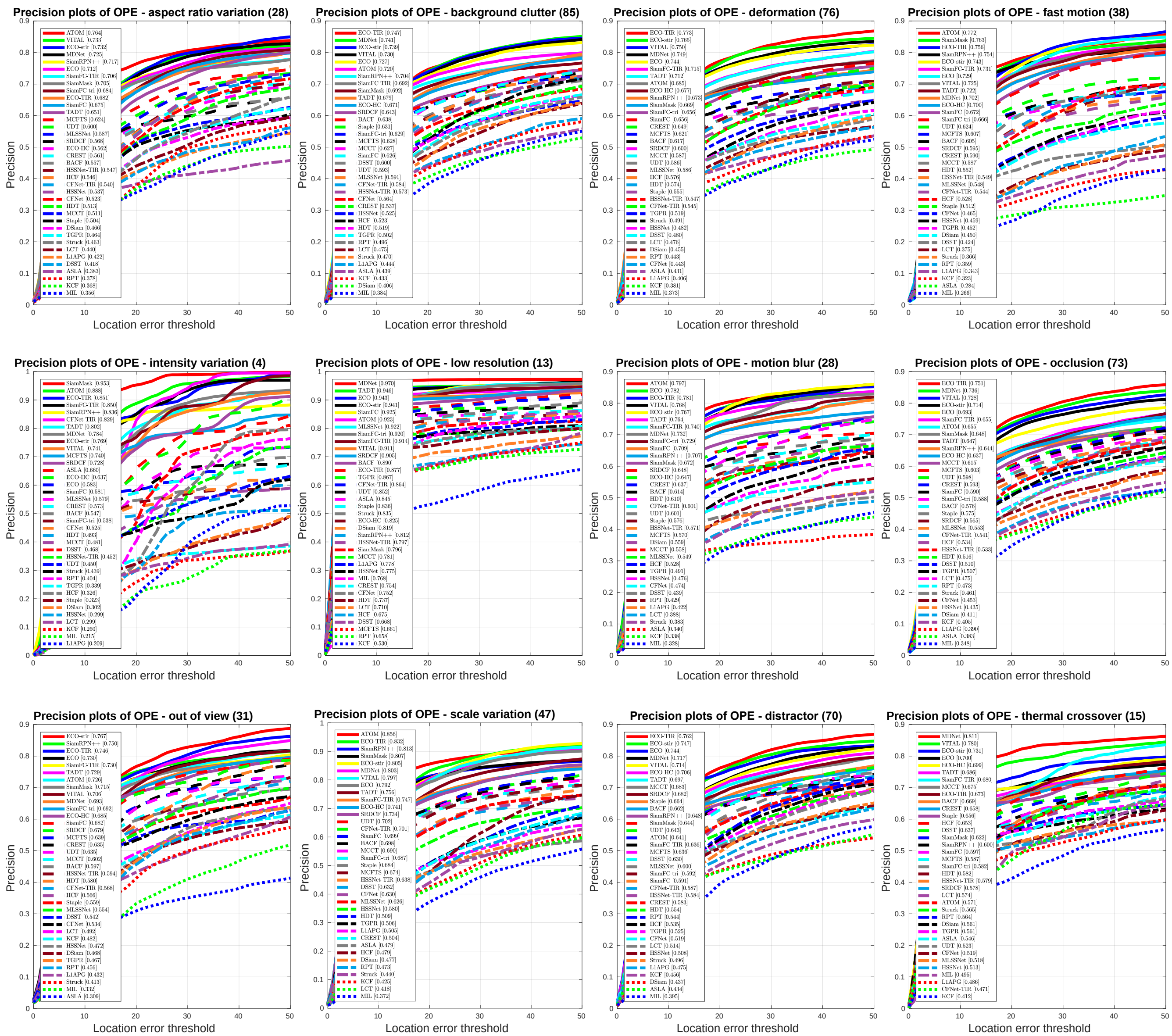


Figure 5: Precision plots of all the evaluated trackers on the 12 challenge attribute subsets.

[33] Chao Ma, Xiaokang Yang, Chongyang Zhang, and Ming Hsuan Yang. 2015. Long-term correlation tracking. In *IEEE Conference on Computer Vision and Pattern Recognition*.

[34] Matthias Mueller, Neil Smith, and Bernard Ghanem. 2016. A benchmark and simulator for uav tracking. In *European Conference on Computer Vision*.

[35] Matthias Muller, Adel Bibi, Silvio Giancola, Salman Alsubaihi, and Bernard Ghanem. 2018. Trackingnet: A large-scale dataset and benchmark for object tracking in the wild. In *European Conference on Computer Vision*.

[36] Hyeonseob Nam and Bohyung Han. 2016. Learning multi-domain convolutional neural networks for visual tracking. In *IEEE Conference on Computer Vision and Pattern Recognition*.

[37] Jan Portmann, Simon Lynen, Margarita Chli, and Roland Siegwart. 2014. People detection and tracking from aerial thermal views. In *IEEE Robotics and Automation Society*.

[38] Yuankai Qi, Shengping Zhang, Lei Qin, et al. 2016. Hedged Deep Tracking. In *IEEE Conference on Computer Vision and Pattern Recognition*.

[39] Yibing Song, Chao Ma, Lijun Gong, et al. 2017. Crest: Convolutional residual learning for visual tracking. In *IEEE International Conference on Computer Vision*.

[40] Yibing Song, Chao Ma, Xiaohe Wu, Lijun Gong, et al. 2018. Vital: Visual tracking via adversarial learning. In *IEEE Conference on Computer Vision and Pattern Recognition*.

[41] Jack Valmadre, Luca Bertinetto, João Henriques, Andrea Vedaldi, and Philip HS Torr. 2017. End-to-end representation learning for correlation filter based tracking. In *IEEE Conference on Computer Vision and Pattern Recognition*.

[42] Ning Wang, Yibing Song, Chao Ma, Wengang Zhou, Wei Liu, and Houqiang Li. 2019. Unsupervised Deep Tracking. In *IEEE Conference on Computer Vision and Pattern Recognition*.

[43] Ning Wang, Wengang Zhou, Qi Tian, Richang Hong, Meng Wang, and Houqiang Li. 2018. Multi-cue correlation filters for robust visual tracking. In *IEEE Conference on Computer Vision and Pattern Recognition*.

[44] Qiang Wang, Li Zhang, Luca Bertinetto, Weiming Hu, and Philip HS Torr. 2019. Fast online object tracking and segmentation: A unifying approach. In *IEEE*

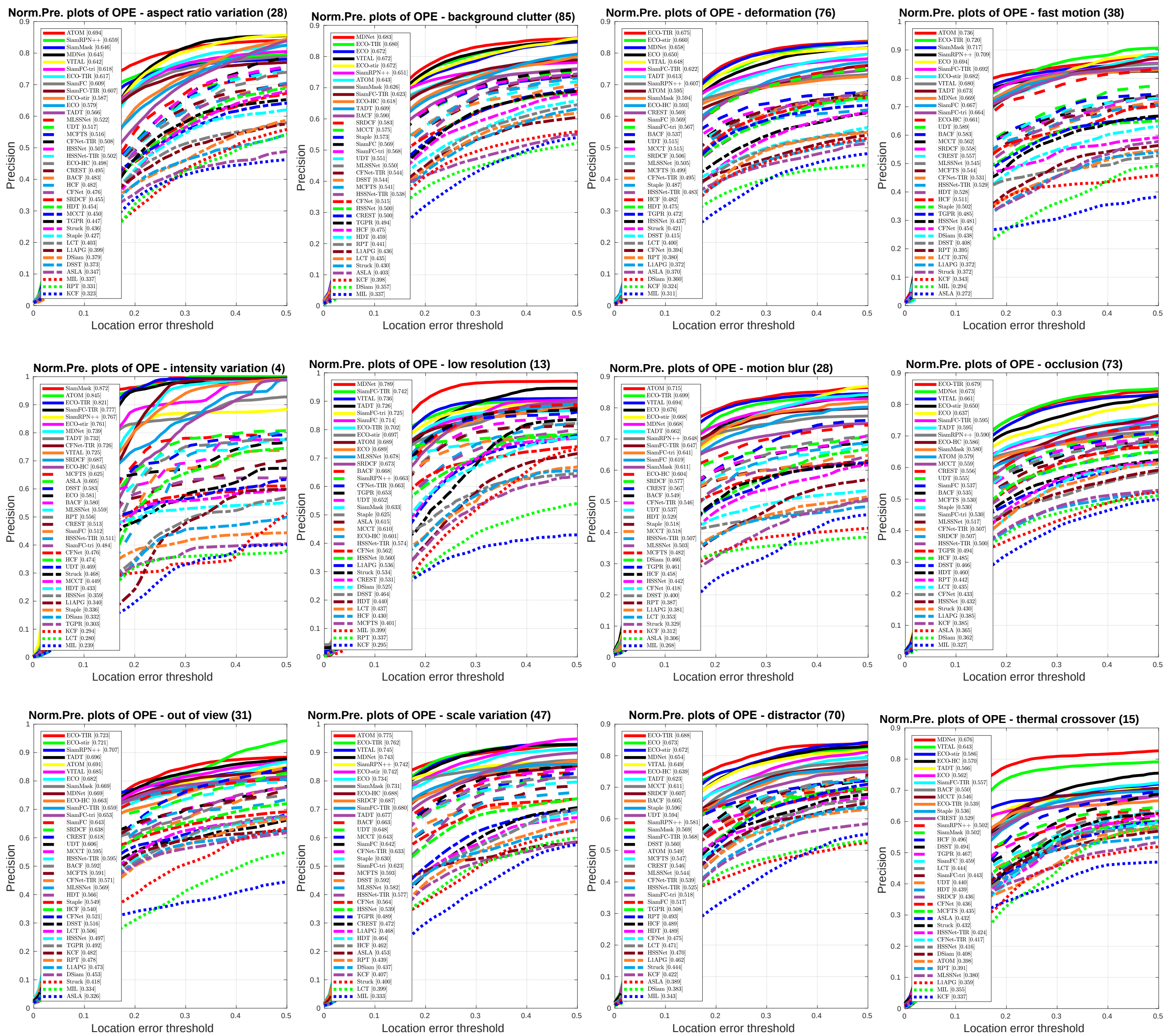


Figure 6: Normalized precision plots of all the evaluated trackers on the 12 challenge attribute subsets.

Conference on Computer Vision and Pattern Recognition.

[45] Yi Wu, Jongwoo Lim, and Ming-Hsuan Yang. 2013. Online object tracking: A benchmark. In *IEEE Conference on Computer Vision and Pattern Recognition.*

[46] Yi Wu, Jongwoo Lim, and Ming-Hsuan Yang. 2015. Object tracking benchmark. *IEEE Transactions of Pattern Analysis and Machine Intelligence* 37, 9 (2015), 1834–1848.

[47] Zheng Wu, Nathan Fuller, Diane Theriault, and Margrit Betke. 2014. A thermal infrared video benchmark for visual analysis. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops.*

[48] Jia Xu, Huchuan Lu, and Ming Hsuan Yang. 2012. Visual tracking via adaptive structural local sparse appearance model. In *IEEE Conference on Computer Vision and Pattern Recognition.*

[49] Lichao Zhang, Abel Gonzalez-Garcia, Joost van de Weijer, Martin Danelljan, and Fahad Shahbaz Khan. 2019. Synthetic data generation for end-to-end thermal infrared tracking. *IEEE Transactions on Image Processing* 28, 4 (2019), 1837–1850.