Unsupervised Domain Adaptation for Nighttime Aerial Tracking (Supplementary Material)

Junjie Ye†, Changhong Fu†,*, Guangze Zheng†, Danda Pani Paudel‡, and Guang Chen†
†Tongji University, China ‡ETH Zürich, Switzerland
{ye.jun.jie, changhongfu, mmlp, guangchen}@tongji.edu.cn, paudel@vision.ee.ethz.ch

Abstract

This supplementary material provides more details of the proposed nighttime aerial tracking 2021 (NAT2021) benchmark, definitions of evaluation metrics, and more experimental results.

1. More details of NAT2021

1.1. Attribute definitions

Definitions of 12 attributes are displayed in Tab. 1. Among them, ARC, FM, SV, IV, and LAI are annotated automatically by analyzing the variation of the ground truth boxes across the timeline, while the rest attributes are manually labelled by visually analyzing.

Table 1. Abbreviations and definitions of 12 attributes.

Attribute	Abbreviation	Definition
Aspect Ratio Change	**ARC**	Aspect ratio of at least one bounding box to the initial one is outside the range [0.5, 2]
Background Clutter	**BC**	The background near the target has similar appearance as the target
Camera Motion	**CM**	Abrupt motion of the camera
Fast Motion	**FM**	The motion of the target is larger than the size of its bounding box
Partial Occlusion	**OCC**	The target is partially occluded by the background
Full Occlusion	**FOC**	The target is fully occluded by the background
Out-of-View	**OV**	The target completely leaves the view in at least one frame
Scale variation	**SV**	The ratio of at least one bounding box to the initial one is outside the range [0.5, 2]
Similar Object	**SOB**	There are objects of similar shape or same type near the target
Viewpoint Change	**VC**	Viewpoint affects target appearance significantly
Illumination Variation	**IV**	The difference between the maximum and the minimum of target ambient intensity is over 40
Low Ambient Intensity	**LAI**	The average target ambient intensity is below 20

1.2. Statistic summary

More statistics of NAT2021-$test$ are shown in Fig. 1. From the bounding box size distribution statistic, we can see that the objects in our benchmark mainly stand at a small size, which underlines the particularity of tracking from an aerial perspective. Normalized bounding box motion change distribution is illustrated in the top-right of Fig. 1, while the relative aspect ratio to the initial bounding box is displayed in the bottom-left. The statistics show that the tracked object varies rapidly. The attribute distribution depicted in the bottom-right of Fig. 1 also demonstrates that the proposed benchmark involves diverse challenging scenes.

*Corresponding author

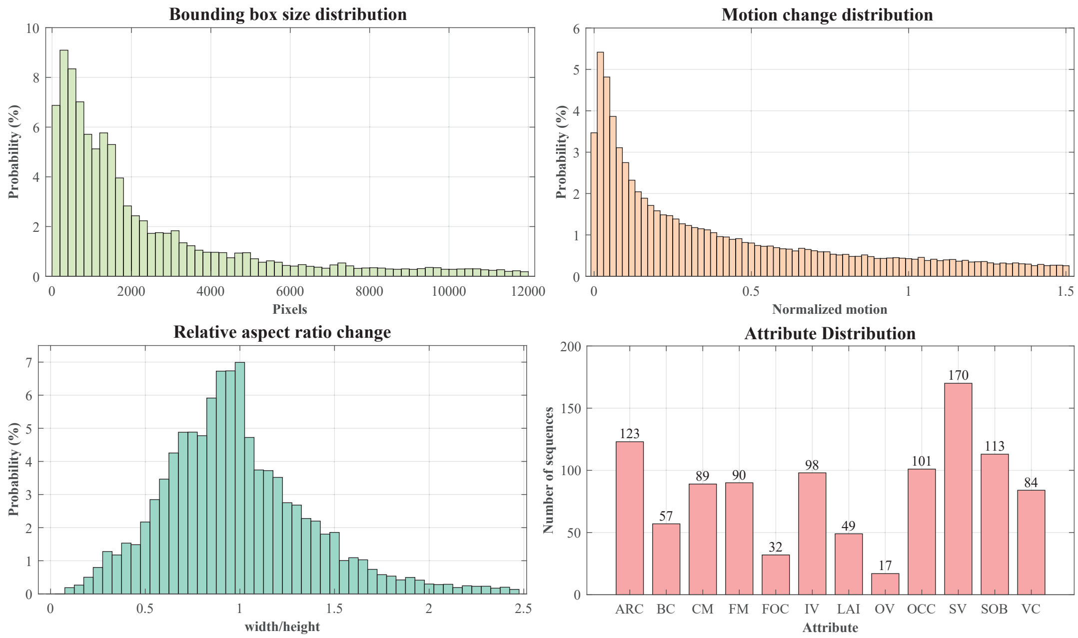


Figure 1. Statistics of NAT2021-*test* in terms of bounding box size distribution, normalized motion change distribution, relative aspect ratio change to the initial box, and attributes distribution.

2. Evaluation metrics

We adopt the one-pass evaluation (OPE) and measure the success rate, precision, and normalized precision to rank trackers. Concretely, the success rate is measured by the intersection over union (IoU) between the tracking result and the ground truth bounding box. The percentage of frames whose IoUs are beyond the preset threshold is drawn as the success plot. In general, the area under the curve (AUC) of the success plot is reported. The distance between the predicted box and the ground truth one is utilized to compute the precision. The percentage of frames whose distances are within a threshold is illustrated as the precision plot. The precision score is reported at the threshold of 20 pixels. Since the precision metric is sensitive to the image resolution and object size, normalized precision is introduced following [4], which is computed as the percentage of frames where the normalized distance between the estimated and the ground truth positions within a threshold. The normalized precision is plotted over a range of 0 to 0.5. The AUC of the normalized precision plot is used to rank trackers.

3. Supplemental experiments

3.1. Can low-light enhancement facilitate nighttime aerial tracking?

In the data preprocessing stage, we found that low-light enhancement [3] facilitates the saliency detection model works better in nighttime scenes, which inspires us to investigate whether nighttime tracking can also benefit from low-light enhancers. An experiment of tracking with enhancement is presented in Tab. 2, though low-light enhancement does raise the visibility of images, results show that the nighttime aerial tracking performance is degraded. Besides, as shown in Fig. 2, the low-light enhancement method [3] is not very effective for narrowing domain discrepancy at the feature level. We conjecture the degradation caused by low-light enhancement comes from the gap in optimization objectives and the weak collaboration of the enhancer and tracker.

Table 2. Performance comparison on NAT2021-*test* of the bare tracker, tracker with enhancer, and the proposed UDAT.

Trackers	Prec.	Succ.	Norm. Prec.
SiamCAR	0.663	0.453	0.542
w/ low-light enhancer	0.648 (-2.26%)	0.430 (-5.08%)	0.520 (-4.06%)
UDAT-CAR	**0.687 (+3.62%)**	**0.483 (+6.62%)**	**0.564 (+4.06%)**

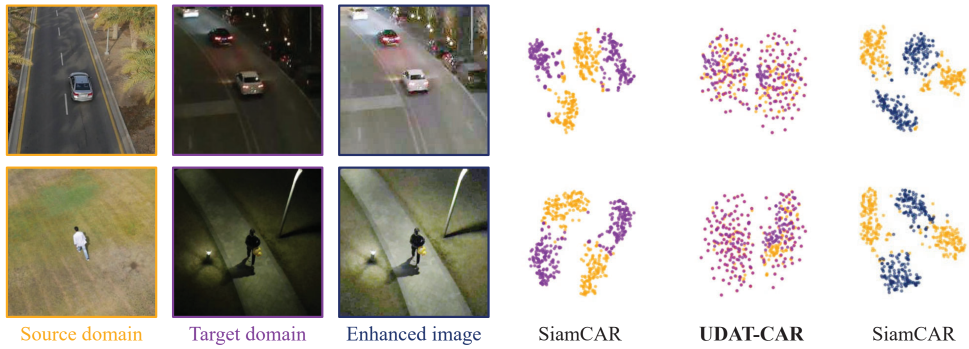


Figure 2. Features visualizations by t-SNE [5] of day/night similar scenes. Gold, purple, and cyan indicate source domain, target domain, and enhancement of target domain by [3], respectively. The first two columns of scattergrams from left to right depict day/night features extracted by SiamCAR [1] and the proposed UDAT-CAR, while the third column illustrates features extracted by SiamCAR [1] from daytime image and enhanced nighttime image.

Table 3. Normalized precision of top-10 trackers in aerial tracking-specific attributes on NAT2021-$test$. The first two places are bolded and underlined, respectively.

Trackers	ARC	BC	CM	FOC	POC	SV	SO	VC
LUDT	0.369	0.417	0.358	0.239	0.389	0.452	0.437	0.361
SiamFC++	0.454	0.414	0.428	0.320	0.459	0.504	0.497	0.460
Ocean	0.436	0.426	0.417	0.326	0.433	0.471	0.456	0.435
SiamRPN++	0.439	0.416	0.437	0.264	0.442	0.496	0.495	0.427
SiamAPN++	0.453	0.415	0.459	0.316	0.450	0.493	0.469	0.453
D3S	0.423	0.423	0.407	0.337	0.425	0.456	0.456	0.397
SiamBAN	0.466	0.439	0.449	0.295	0.445	0.522	0.485	0.444
SiamCAR	0.494	0.457	0.486	0.334	0.485	0.549	0.529	0.484
UDAT-BAN	0.513	0.471	0.510	**0.364**	0.487	0.556	0.533	0.484
UDAT-CAR	**0.520**	**0.494**	**0.515**	0.353	**0.513**	**0.574**	**0.541**	**0.517**
Δ_{BAN} (%)	10.09	7.29	13.59	23.39	9.44	6.51	9.90	9.01
Δ_{CAR} (%)	5.26	8.10	5.97	5.69	5.77	4.55	2.27	6.82

3.2. Aerial tracking-oriented evaluation

Apart from illumination-related evaluation, attributes that frequently appear in aerial tracking are also evaluated to analyze the performance of trackers in aerial view. The results are shown in Tab. 3. With domain adaptive training, UDAT trackers generalize well in nighttime aerial perspective and realize favorable performance.

3.3. Detailed attribute-based performance

To give an in-depth analysis of state-of-the-art trackers along with UDAT under various challenging tracking scenes, we provide detailed attribute-based performance on NAT2021-$test$ (see Figs. 3 and 4) and UAVDark70 [2] (see Fig. 5). The proposed UDAT trackers, *i.e*., UDAT-CAR and UDAT-BAN, rank first two places in most attributes of NAT2021-$test$ and UAVDark70.

References

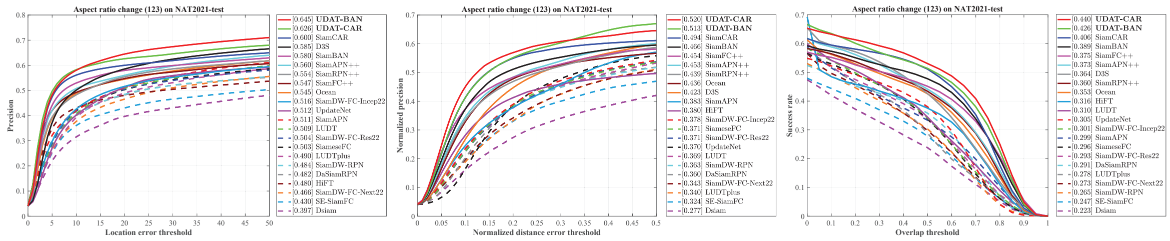
[1] Dongyan Guo, Jun Wang, Ying Cui, Zhenhua Wang, and Shengyong Chen. SiamCAR: Siamese Fully Convolutional Classification and Regression for Visual Tracking. In *CVPR*, pages 6268–6276, 2020. 3

[2] Bowen Li, Changhong Fu, Fangqiang Ding, Junjie Ye, and Fuling Lin. ADTrack: Target-Aware Dual Filter Learning for Real-Time Anti-Dark UAV Tracking. In *ICRA*, pages 496–502, 2021. 3

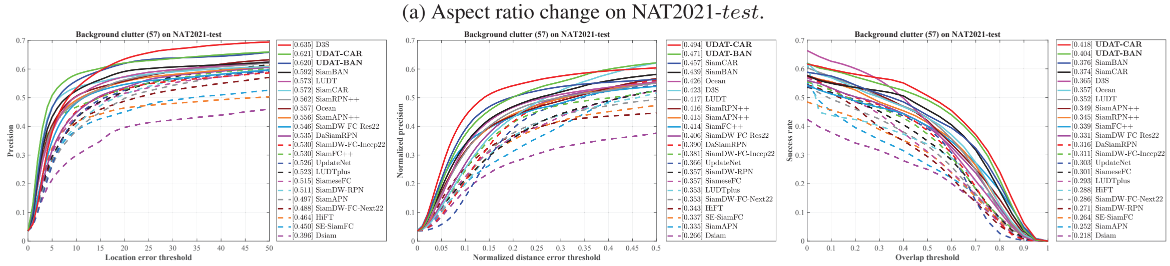
[3] Chongyi Li, Chunle Guo, and Change Loy Chen. Learning to Enhance Low-Light Image via Zero-Reference Deep Curve Estimation. *IEEE TPAMI*, pages 1–14, 2021. 2, 3

[4] Matthias Muller, Adel Bibi, Silvio Giancola, Salman Alsubaihi, and Bernard Ghanem. TrackingNet: A Large-Scale Dataset and Benchmark for Object Tracking in the Wild. In *ECCV*, pages 310–327, 2018. 2

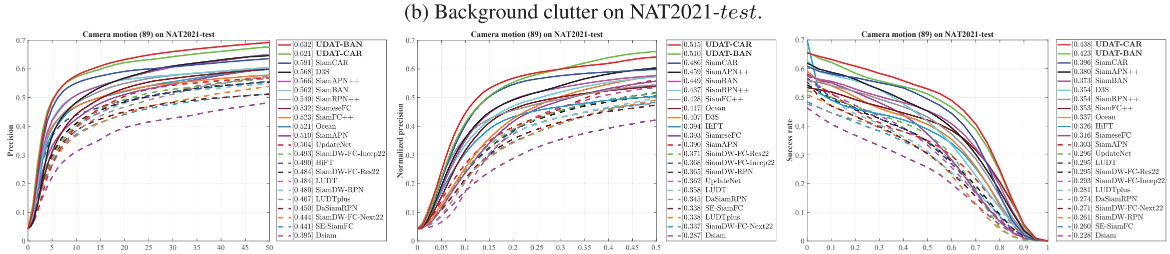
[5] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *JMLR*, 9(11):2579–2605, 2008. 3



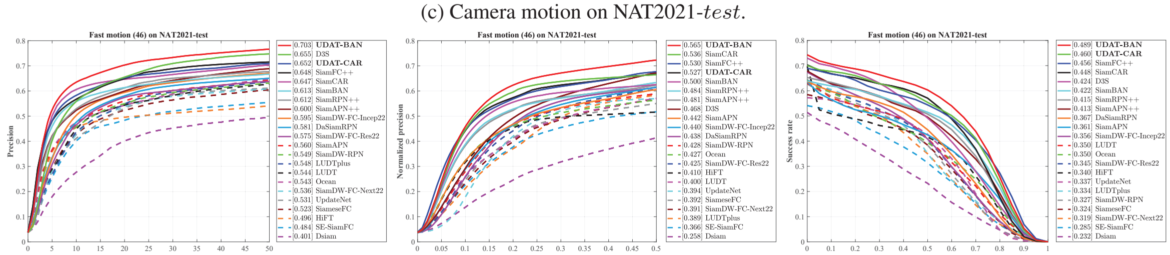
(a) Aspect ratio change on NAT2021-*test*.



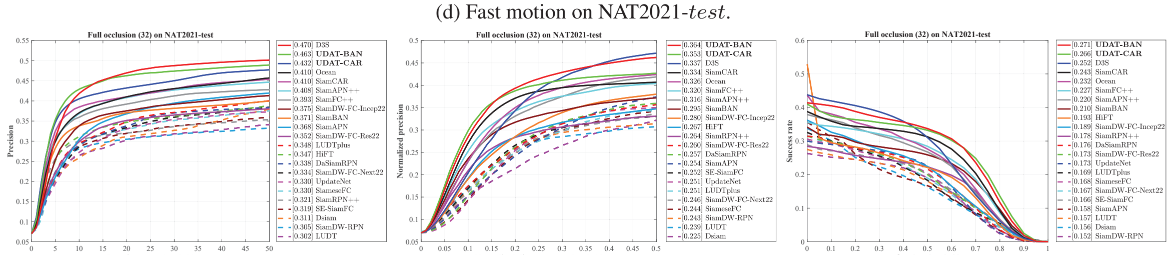
(b) Background clutter on NAT2021-*test*.



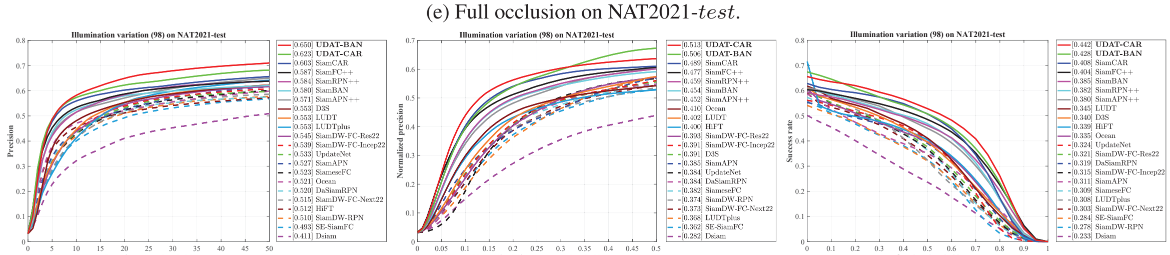
(c) Camera motion on NAT2021-*test*.



(d) Fast motion on NAT2021-*test*.

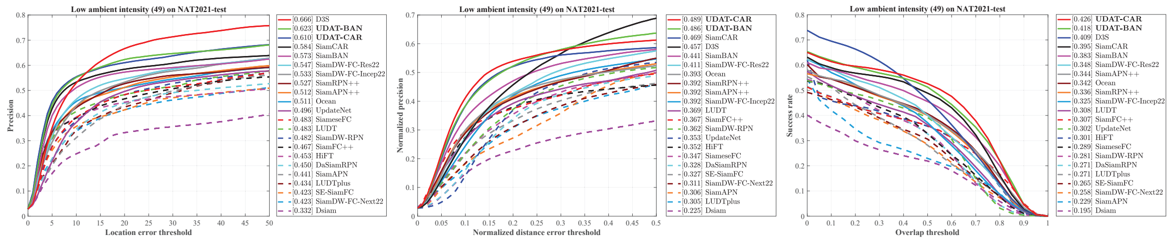


(e) Full occlusion on NAT2021-*test*.

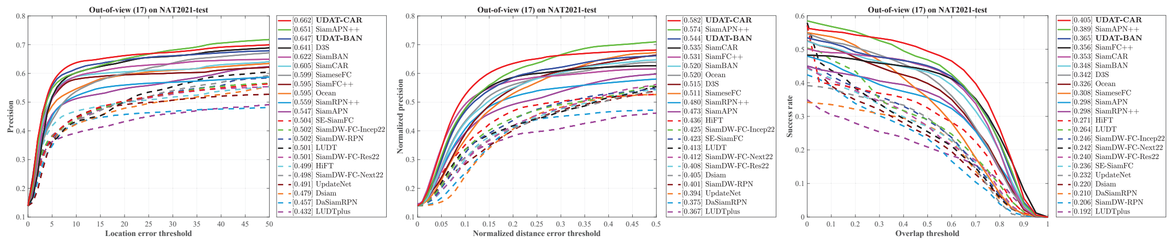


(f) Illumination variation on NAT2021-*test*.

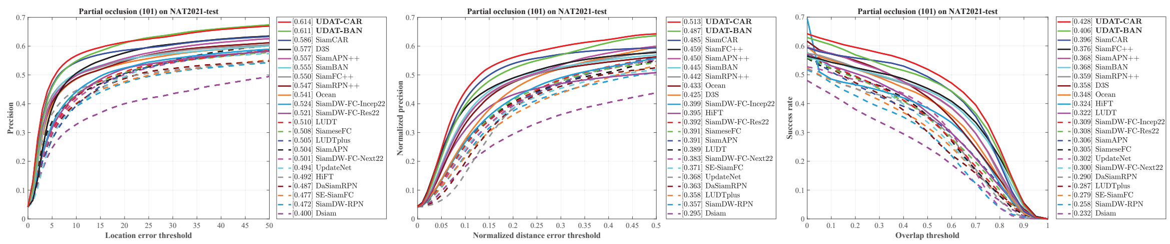
Figure 3. Detailed attribute-based performance on NAT2021-*test*.



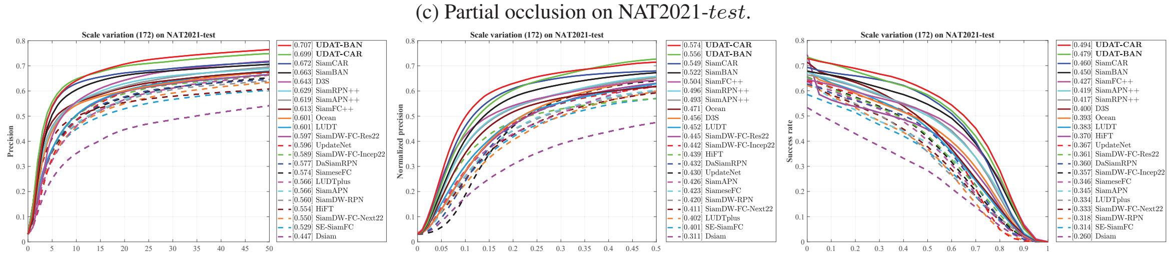
(a) Low ambient intensity on NAT2021-*test*.



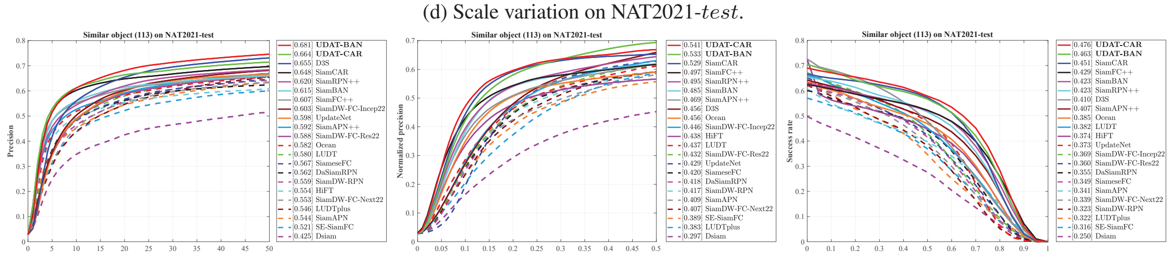
(b) Out-of-view on NAT2021-*test*.



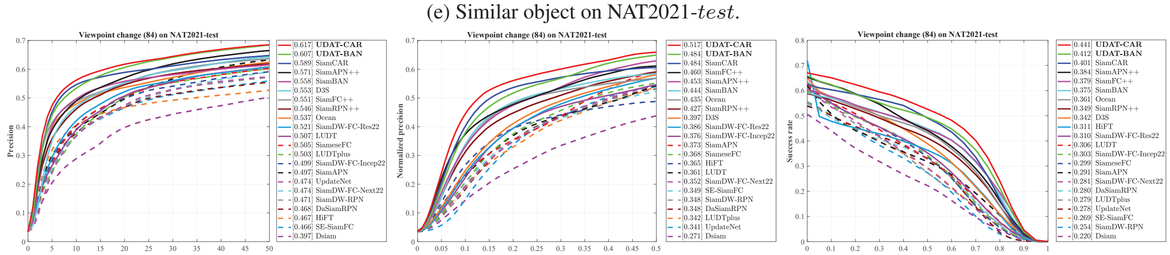
(c) Partial occlusion on NAT2021-*test*.



(d) Scale variation on NAT2021-*test*.

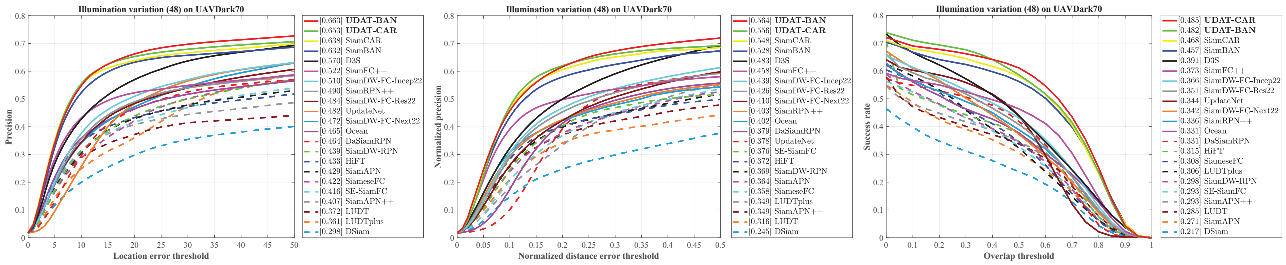


(e) Similar object on NAT2021-*test*.

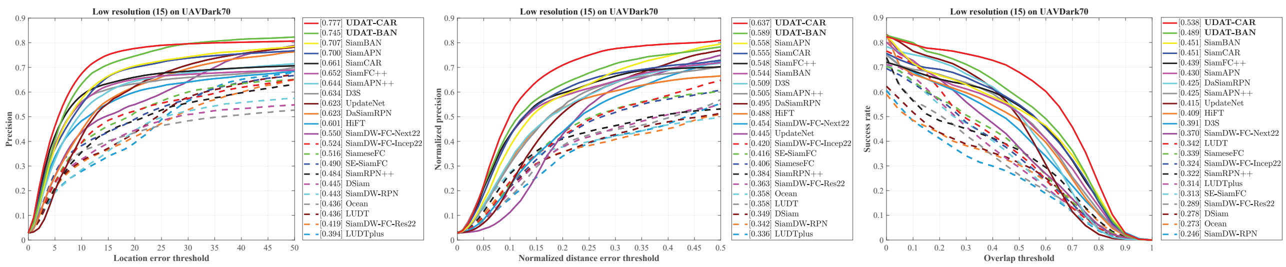


(f) Viewpoint change on NAT2021-*test*.

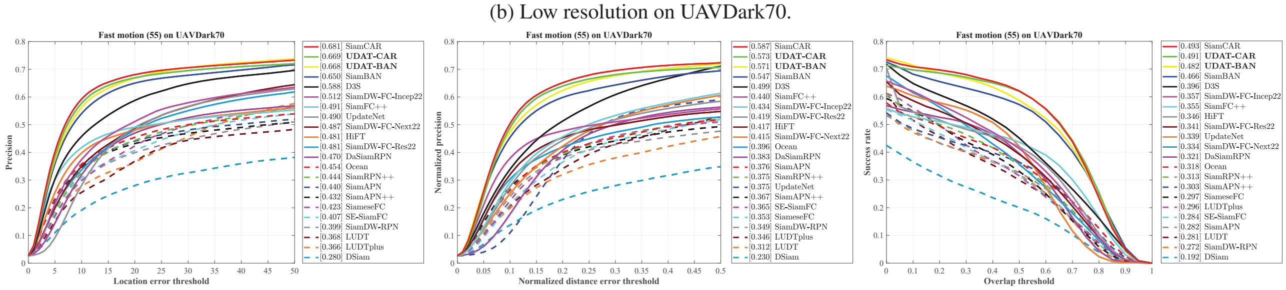
Figure 4. Detailed attribute-based performance on NAT2021-*test*.



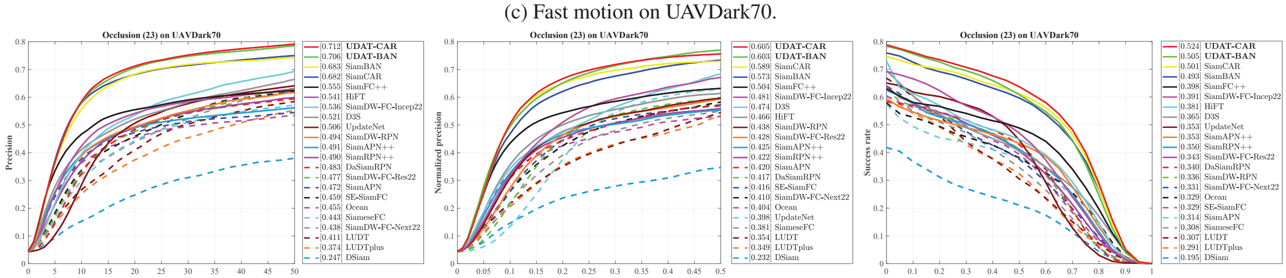
(a) Illumination variation on UAVDark70.



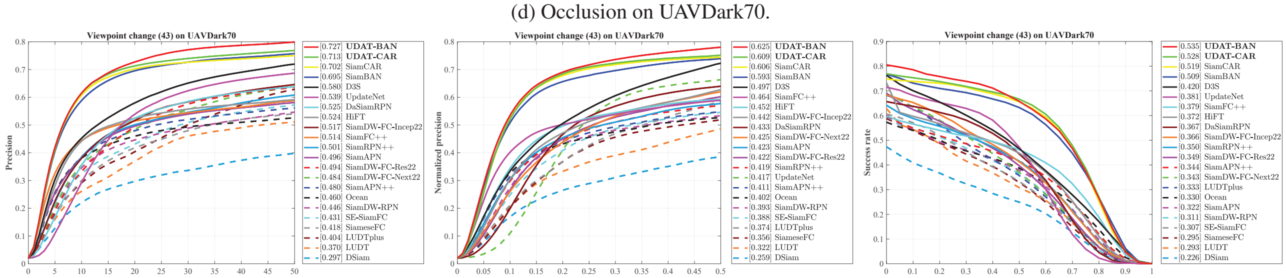
(b) Low resolution on UAVDark70.



(c) Fast motion on UAVDark70.



(d) Occlusion on UAVDark70.



(e) Viewpoint change on UAVDark70.

Figure 5. Detailed attribute-based performance on UAVDark70.