\Orchestrating a brighter world NEC

REDUCING FALSE POSITIVES IN OBJECT TRACKING WITH SIAMESE NETWORK

Takuya Ogawa[1], Takashi Shibata[1†], Shoji Yachida[1],
and Toshinori Hosoi[1]

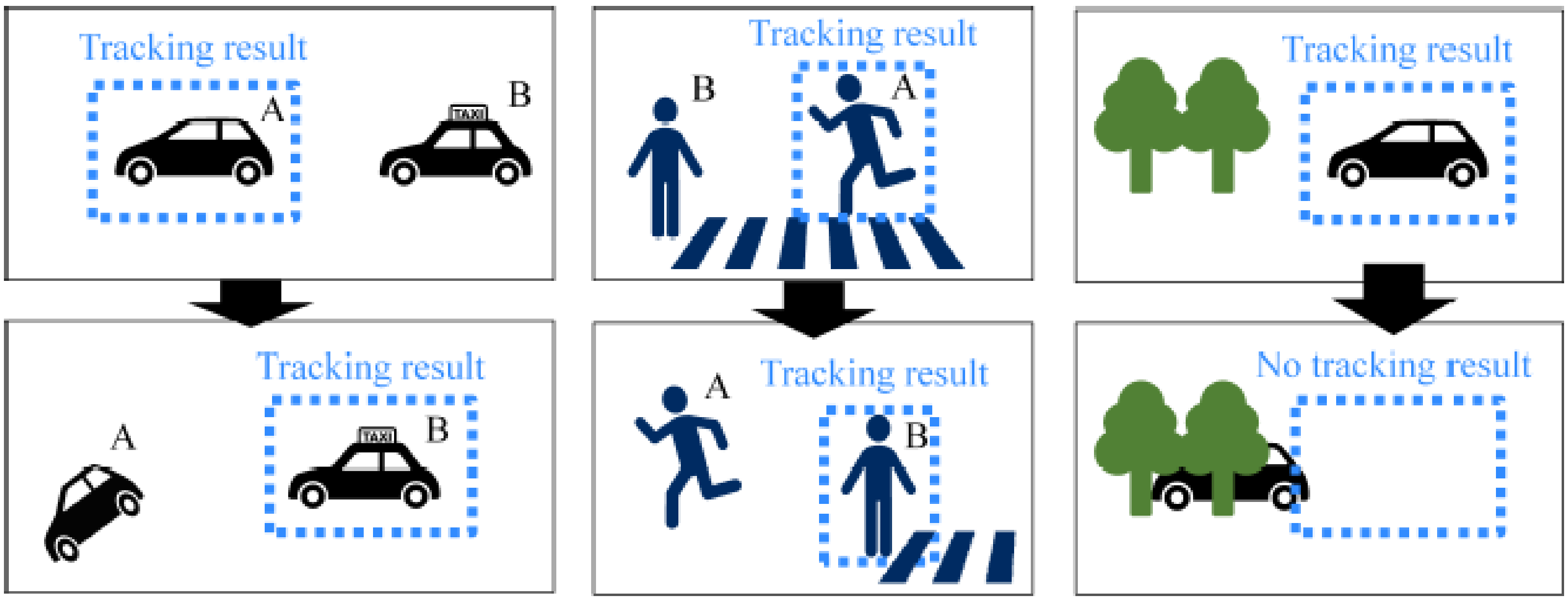
1: NEC, †: this author now belongs to NTT

E-mail: takuya_ogawa@nec.com

Introduction: single object tracking

Practical long-term tracking is needed for real-time surveillance

Problems: drift and loss



Drift Drift Loss

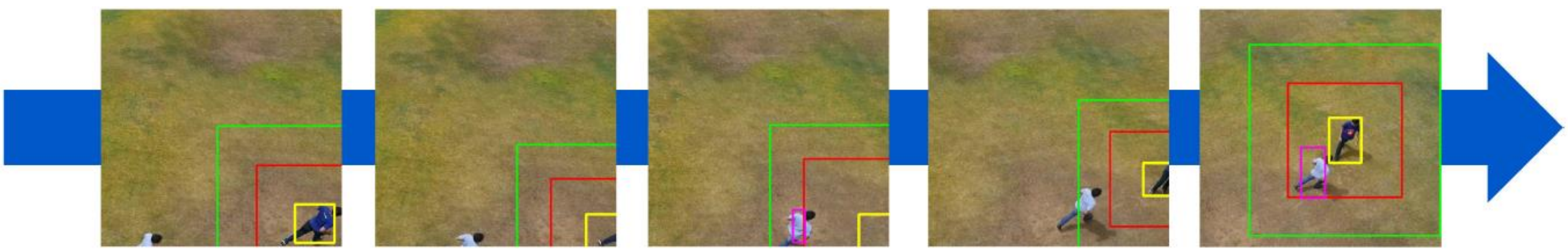
Goal

- Reduce the occurrence of drift and recover from the loss of the target

Concept of our method

Conventional [1] : Drift to similar object and never re-track the target

Proposed : Recognize the similar objects in advance and keep tracking a true target

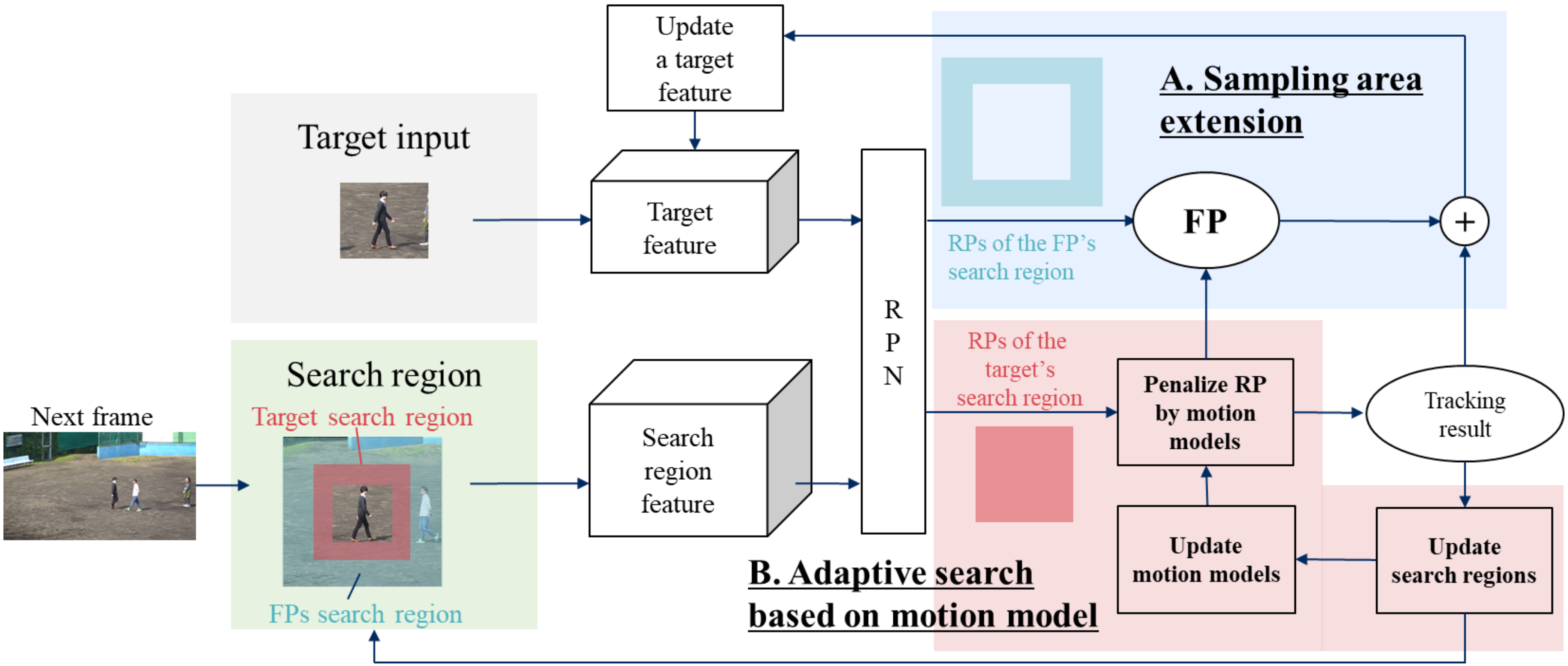


t_1 t_2 t_3 t_4 t_5

[1] Z. Zhu et al. "Distractor-aware siamese networks for visual object tracking," ECCV2018

Proposed Frameworks

2 frameworks for avoiding drift/loss of the target



FPC = false positive candidate, RP=region proposal

A. Sampling area extension

Online learning of similar objects before approaching

Extend a sampling area for the false positive candidates only

- Similar object = false positive candidate (FPC)
- Able to recognize and learn FPC in advance

Conditions for sampling high quality FPC

- Reliability L_{fpc} near the boundary

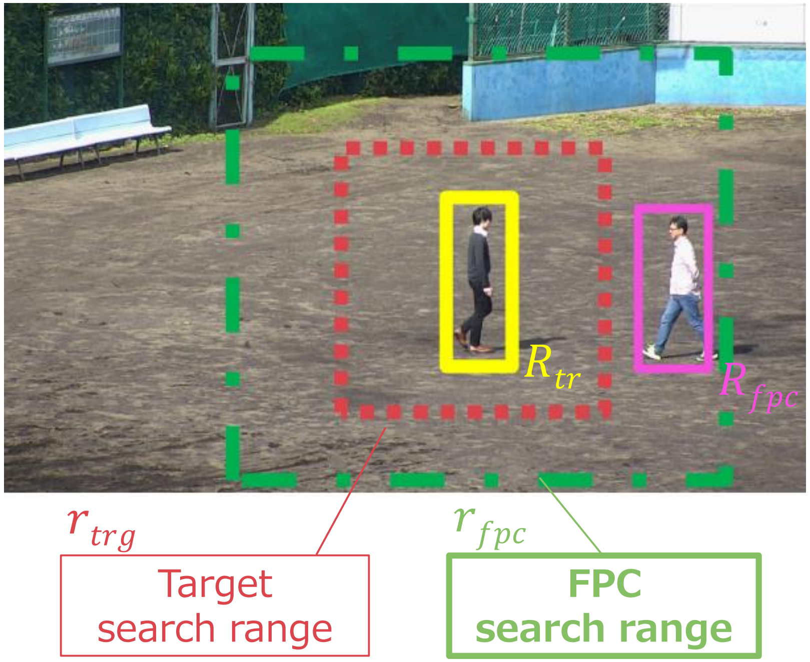
$$0.5 \leq Lfp_c \leq 0.9$$

- Low IoU with the target

$$IoU(R_{tr}, Rfp_c) < 0.1$$

Online learning

- Same as conventional method [1]
  - Feature difference between the target and FPC



$$T_{base} = \sqrt{w^2 + h^2}$$

$$r_{trg} = k * Tbase \qquad 1 \leq k < b$$

$$r_{fpc} = b * Tbase$$

[1] Z. Zhu et al. “Distractor-aware siamese networks for visual object tracking,” ECCV2018

B. Adaptive search based on motion models

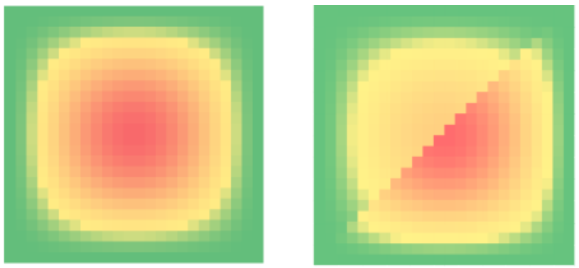
Weighting inside a search region by its motion to reduce drifting

Adopt motion models (Stay/Move)

- Adjust a position reliability W by its motion

$$W = w_{2D} * R(\theta) * \begin{pmatrix} 1 & 1 & 1 \\ 1 & c & d \\ 1 & d & d \end{pmatrix} \qquad c = \begin{cases} 1+\beta & (stay) \\ 1 & (move) \end{cases} \qquad d = \begin{cases} 1 & (stay) \\ \gamma & (move) \end{cases}$$

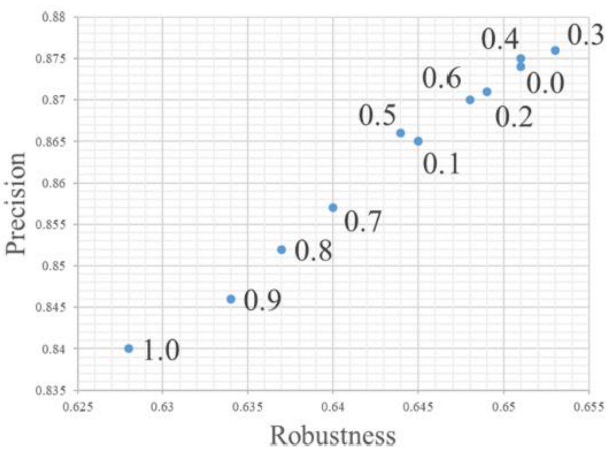
$R(\theta)$: *Rotation Matrix*, w_{2D}: *2D cosine window*



Stay Move

- Determine motion constraint value γ by experiment

$$\beta = 1.0, \quad \gamma = 0.3$$



Update a reliability L of each candidate in the same way as [1]

- Introduce size penalty

$$p_{new} = e^{k*max(\frac{r}{r'},\frac{r'}{r})*max(\frac{s}{s'},\frac{s'}{s})} * W$$

- Taking visual reliability α into account the reliability L

$$L = \alpha * p_{new}$$

[1] Z. Zhu et al. "Distractor-aware siamese networks for visual object tracking," ECCV2018

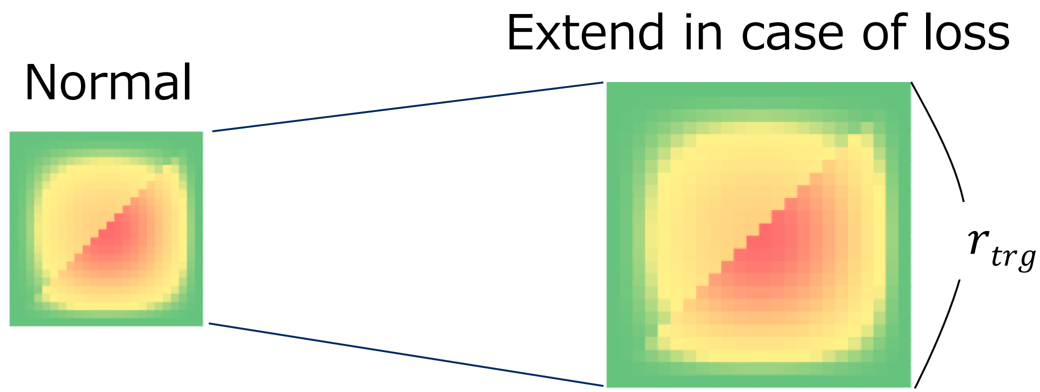
B. Adaptive search based on motion models

Extend the search region in case of losing the target

- Detect loss by target reliability L_{trg}

$$L_{trg} \leq 0.6$$

- Extend the search region according to the num. of consecutive loss



$$r_{trg} = (k + \tau * nlo_{ss} * a) * Tba_{se}$$

τ : an adjusting parameter of the expansion amount
n_{loss} : the frame number of the latest consecutive loss
a : a target acceleration just before loss

Evaluation I : Ablation Study

Both frameworks improved robustness and the combination is better

Metric

- Long-term tracking benchmark dataset: UAV123 [2]
  - Long-term tracking dataset includes absence of the target (causes drift and loss)
- One-pass evaluation (OPE)

A	B	*Robustness*	*Precision*
–	–	0.489	0.725
✓	–	0.570	0.776
–	✓	0.564	0.758
✓	✓	**0.595**	**0.791**

Score improvement

- By combining proposed 2 framework:
  - Improve **21.7% for robustness** and 9.1% for precision

Speed improvement

- **3 times faster (95.7 FPS)** than a video capture speed (30 FPS)
  - Optimization by our adaptive extension of the search region

Method	*SiamRPN* [5]	*DaSiamRPN* [1]	*gSiam (Proposed)*
FPS	**169.6**	29.1	95.7

[1] Z. Zhu et al. "Distractor-aware siamese networks for visual object tracking," ECCV2018
[2] M. Mueller et al. "A benchmark and simulator for uav tracking," ECCV2016
[5] B. Li et al. "High performance visual tracking with siamese region proposal network," CVPR2018

Evaluation II : Comparison with exiting methods

Our method improved robustness/precision in all attributes

Dataset variation

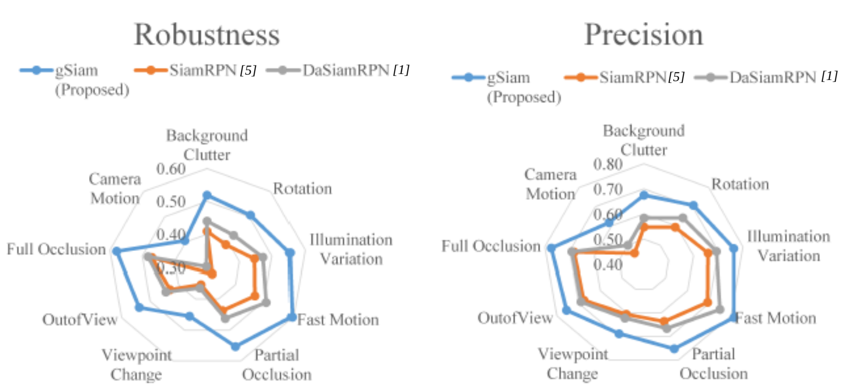
- 3 long-term tracking benchmarks

Dataset	*UAV123* [2]	*LaSOT* [3]	*VOT2018-LT* [4]
No. of seq.	123	280	35
No. of classes	9	70	11
Min len. of seq.	109	1000	1389
Max len. of seq.	3085	9999	29700
Med. len. of seq.	877	2102	2683
Min size of image	(720,480)	(202,224)	(240,320)
Max size of image	(1280,720)	(1280,720)	(1280,720)
Med. size of image	(1280,720)	(1280,720)	(1280,720)
Absence(No. of seq, ratio)	(33, 7.1%)	(89, 2.6%)	(24, 15.1%)

Evaluation results

- the robustness and precision improved in all attributes

Method (Robustness/Precision)	*UAV123* [2]	*LaSOT* [3]	*VOT2018-LT* [4]
gSiam (Proposed)	**0.595** **0.791**	**0.418** **0.447**	**0.484** **0.554**
SiamRPN [5]	0.466 0.696	0.349 0.253	0.109 0.113
DaSiamRPN [1]	<u>0.489</u> <u>0.725</u>	<u>0.388</u> <u>0.291</u>	<u>0.251</u> <u>0.498</u>



The results of attribution evaluation of [2]

We developed a more suitable method for long-term tracking than exiting methods

[1] Z. Zhu et al. "Distractor-aware siamese networks for visual object tracking," ECCV2018
[2] M. Mueller et al. "A benchmark and simulator for uav tracking," ECCV2016
[3] H. Fan et al. "Lasot: A high-quality benchmark for large-scale single object tracking," CVPR2019
[4] M. Kristan, "The sixth visual object tracking vot2018 challenge results," ECCV2018
[5] B. Li et al. "High performance visual tracking with siamese region proposal network," CVPR2018

Evaluation III : Comparison with SOTA methods

Our method is effective even compared with SOTA methods

Prerequisites

- Using deeper feature extraction model than Evaluation II (AlexNet -> ResNet-50)

Evaluation results

- Outperformed SOTA methods on all datasets for the robustness
- More effective for dataset with more absences of the target

Method (Robustness/Precision)	*UAV123* [2]	*LaSOT* [3]	*VOT2018-LT* [4]
gSiamR (Proposed)	**0.602**	**0.501**	**0.547**
	0.808	0.490	0.606
SiamRPN++ [6]	0.600	0.495	0.424
	0.775	**0.493**	**0.647**
SiamMask [7]	**0.602**	0.467	0.426
	0.775	0.469	0.636

[2] M. Mueller et al. "A benchmark and simulator for uav tracking," ECCV2016
[3] H. Fan et al. "Lasot: A high-quality benchmark for large-scale single object tracking," CVPR2019
[4] M. Kristan, "The sixth visual object tracking vot2018 challenge results," ECCV2018
[6] B. Li et al. "Siamrpn++:Evolution of siamese visual tracking with very deep networks," CVPR2019
[7] Q. Wang et al. "Fast online object tracking and segmentation: A unifying approach," CVPR2019

Appendix: Evaluation on short-term tracking dataset

Our method is also effective even on short-term tracking dataset

Evaluated by short-term tracking benchmark dataset

- OTB2015 [8]

Evaluation result

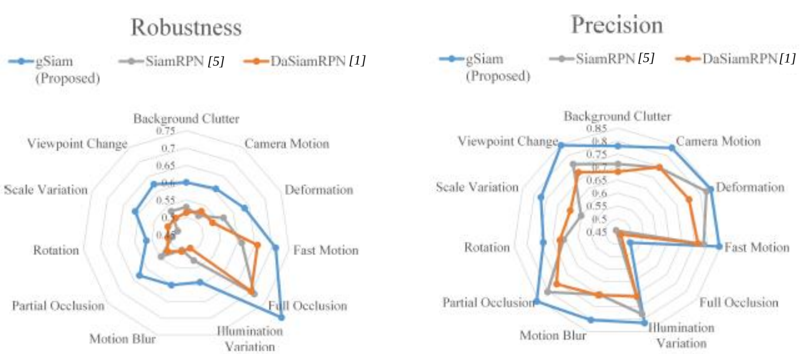
- the robustness and precision improved in all attributes even on short-term scenes

Comparison with other methods

Method	*SiamRPN [5]*	*DaSiamRPN [1]*	*gSiam*
Robustness/Precision	0.539	0.529	**0.628**
	0.785	0.763	**0.846**
	SiamMask [7]	*SiamRPN++ [6]*	*gSiamR*
Robustness/Precision	0.634	0.645	**0.654**
	0.833	0.846	**0.878**

Ablation study

III-B	*III-C*	*Robustness*	*Precision*
–	–	0.529	0.763
✓	–	0.585	0.795
–	✓	0.625	0.828
✓	✓	**0.628**	**0.846**



The result of attribution evaluation

Our method is effective in every situation of object tracking

[1] Z. Zhu et al. "Distractor-aware siamese networks for visual object tracking," ECCV2018
[5] B. Li et al. "High performance visual tracking with siamese region proposal network," CVPR2018
[6] B. Li et al. "Siamrpn++:Evolution of siamese visual tracking with very deep networks," CVPR2019
[7] Q. Wang et al. "Fast online object tracking and segmentation: A unifying approach," CVPR2019
[8] Y. Wu et al. "Object tracking benchmark," TPAMI2015

References

[1] Z. Zhu et al. “Distractor-aware siamese networks for visual object tracking,” ECCV2018

[2] M. Mueller et al. “A benchmark and simulator for uav tracking,” ECCV2016

[3] H. Fan et al. “Lasot: A high-quality benchmark for large-scale single object tracking,” CVPR2019

[4] M. Kristan, “The sixth visual object tracking vot2018 challenge results,” ECCV2018

[5] B. Li et al. “High performance visual tracking with siamese region proposal network,” CVPR2018

[6] B. Li et al. “Siamrpn++:Evolution of siamese visual tracking with very deep networks,” CVPR2019

[7] Q. Wang et al. “Fast online object tracking and segmentation: A unifying approach,” CVPR2019

[8] Y. Wu et al. “Object tracking benchmark,” TPAMI2015