



Long-term Visual Object Tracking

Dong Wang

wdice@dlut.edu.cn

Dalian University of Technology



Long-term Visual Object Tracking

- Long-term Tracking vs Short-term Tracking
- ‘Skimming-Perusal’ Long-term Tracking (ICCV19)
- Long-term Tracking with Meta-Updater (CVPR20)
- New Opportunities and Challenges



Long-term Visual Object Tracking

- Long-term Tracking vs Short-term Tracking
- ‘Skimming-Perusal’ Long-term Tracking (ICCV19)
- Long-term Tracking with Meta-Updater (CVPR20)
- New Opportunities and Challenges



Short-term Tracking

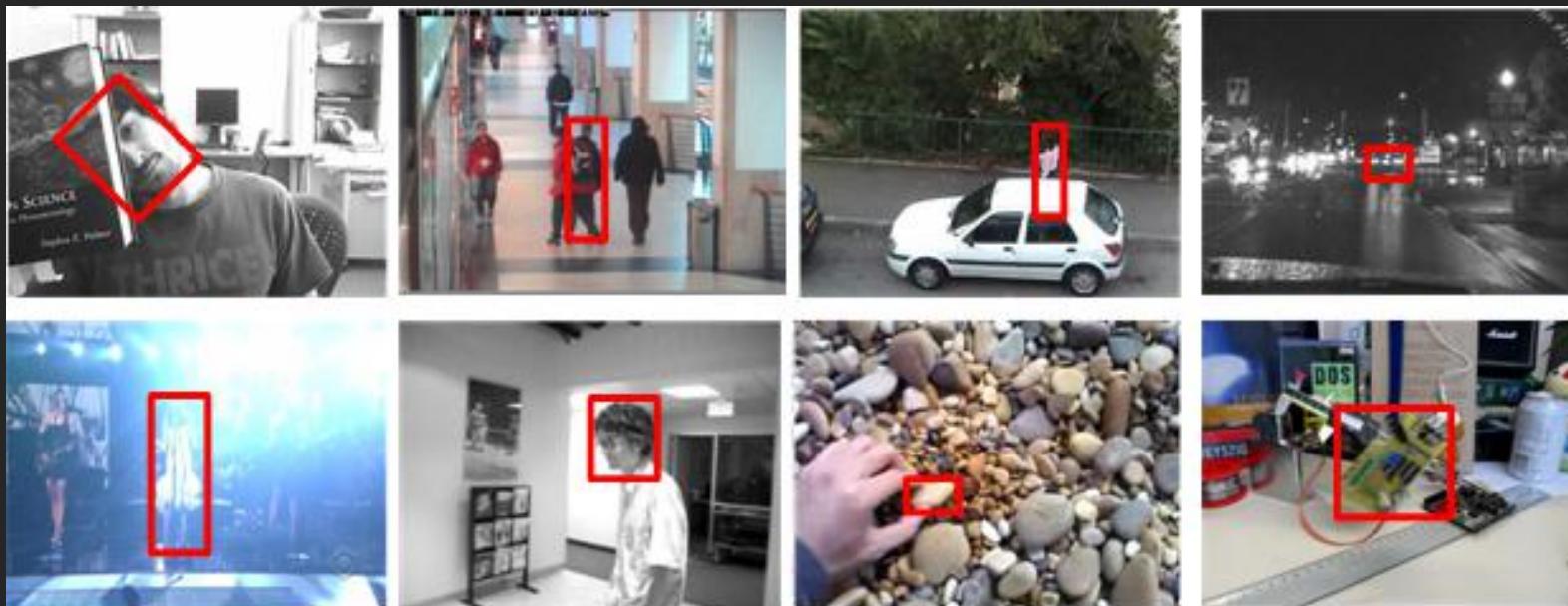
■ Goal

Track an arbitrary object in a video given its initial location

Single-object, Model-free

■ Challenges

Occlusion, Light Change, Background Clutter, etc.



Short-term Tracking

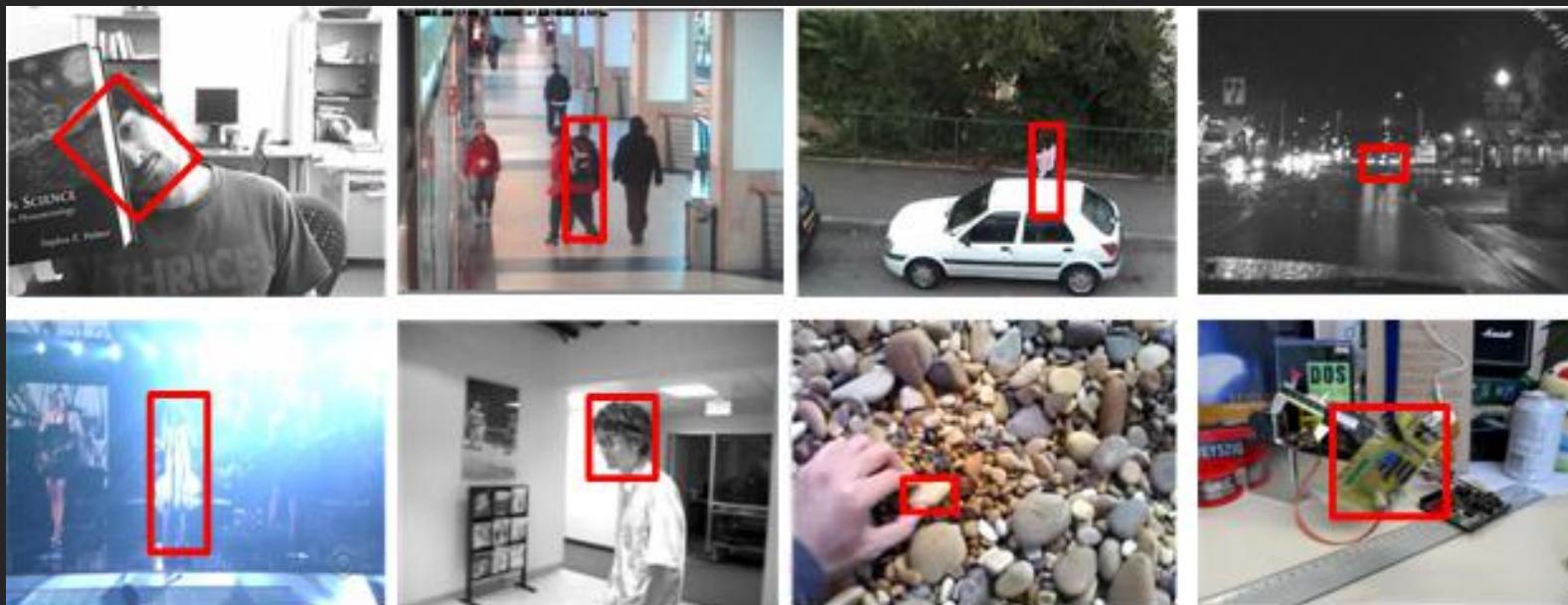
■ Goal

Track an arbitrary object in a video given its initial location

Single-object, Model-free

■ Challenges

Occlusion, Light Change, Background Clutter, etc.



◆ Subspace Learning (IVT)

◆ Online Classifier (MIL,
Structured SVM)

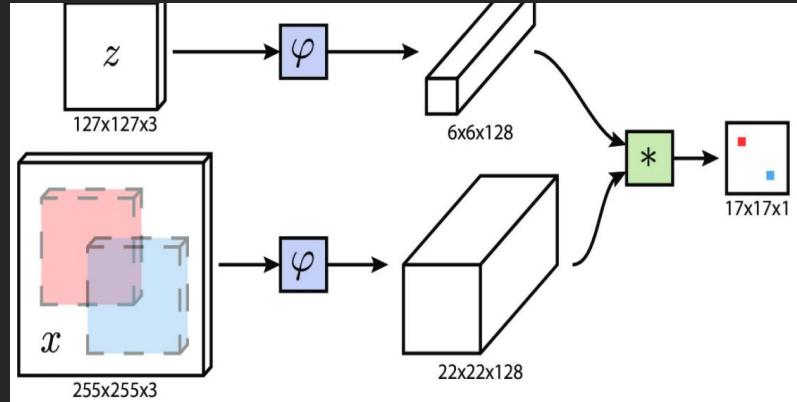
◆ Sparse Representation
(L1APG, SCM, ASLA,
MTT, OSPT, CT...)

◆ Correlation Filter
(KCF, SRDCF, BACF,
ECO, HCFT, DRT, UPDT)

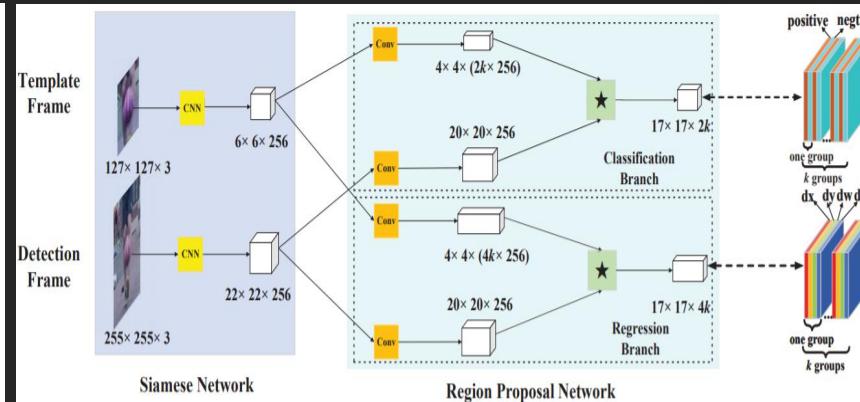
◆ Deep Learning
(MDNet, FCNT, SiamFC
SiamRPN, DaSiamRPN)

Short-term Tracking: One-shot vs Online

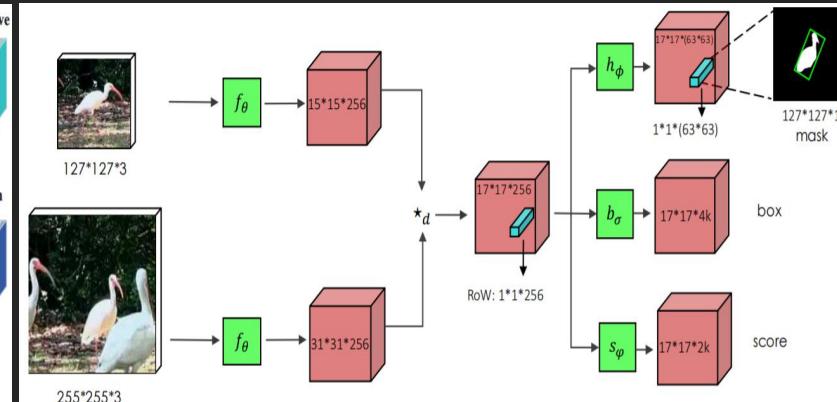
SiamFC (ECCVW16)



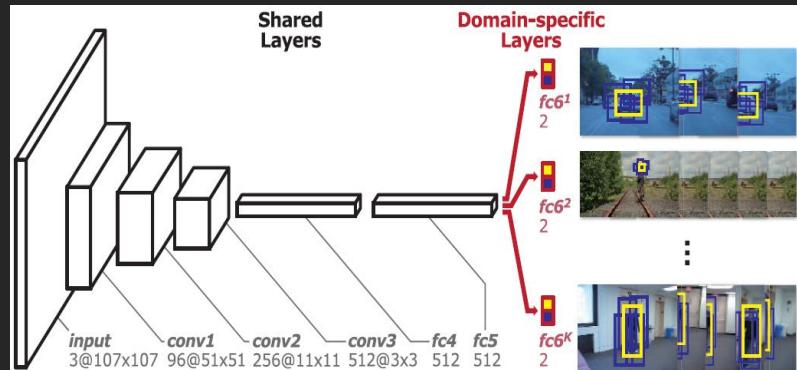
SiamRPN (CVPR18)



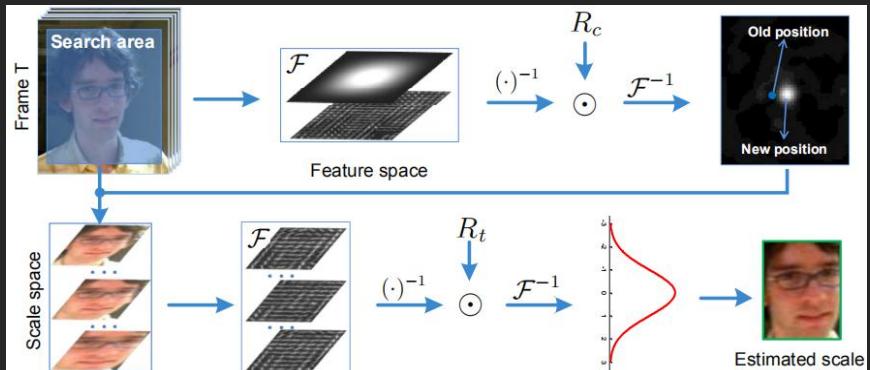
SiamMask (CVPR19)



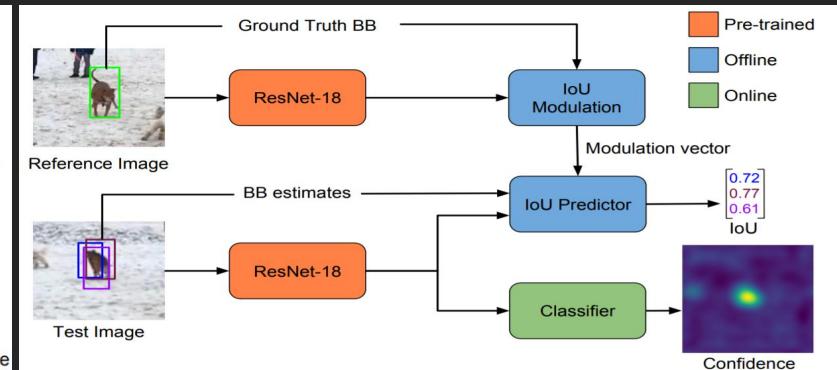
MDNet (CVPR16)



ECO (ICCV17)

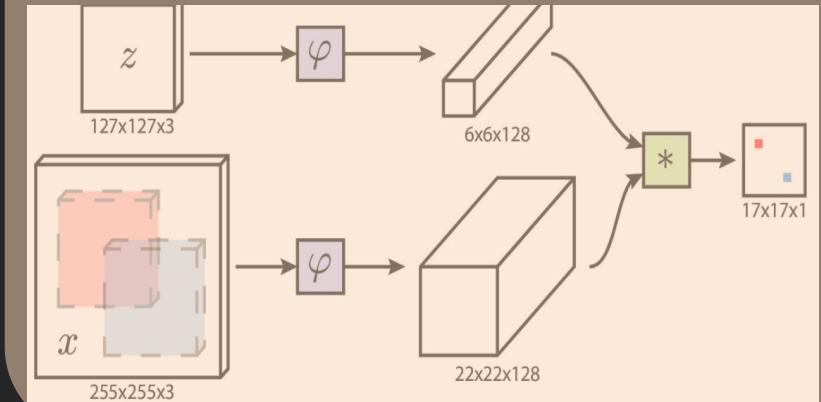


ATOM (CVPR19)

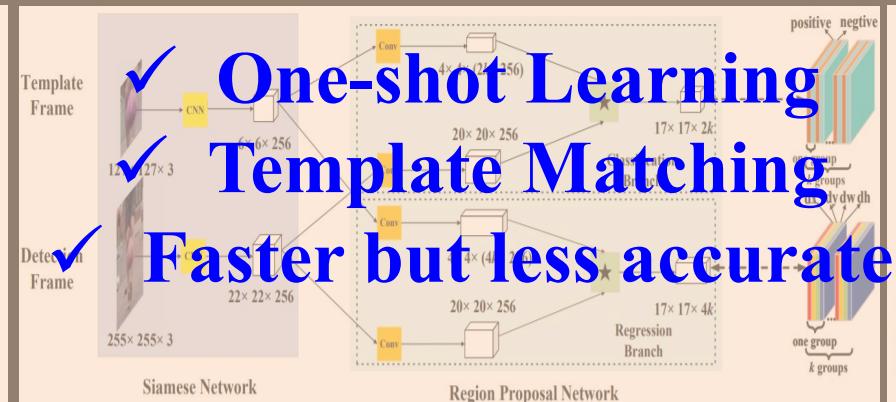


Short-term Tracking: One-shot vs Online

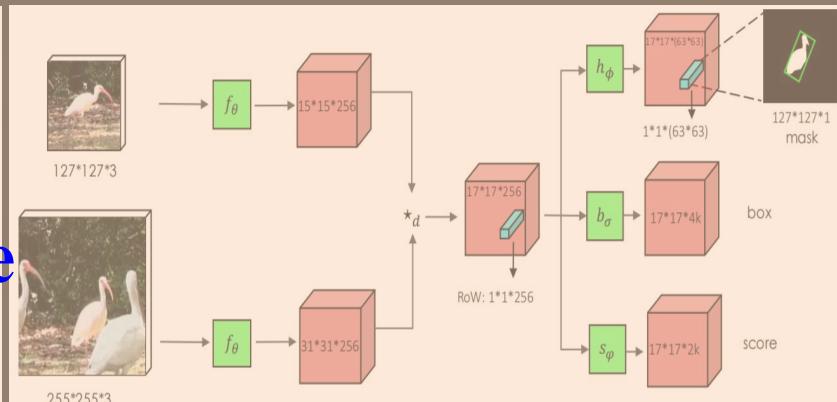
SiamFC (ECCVW16)



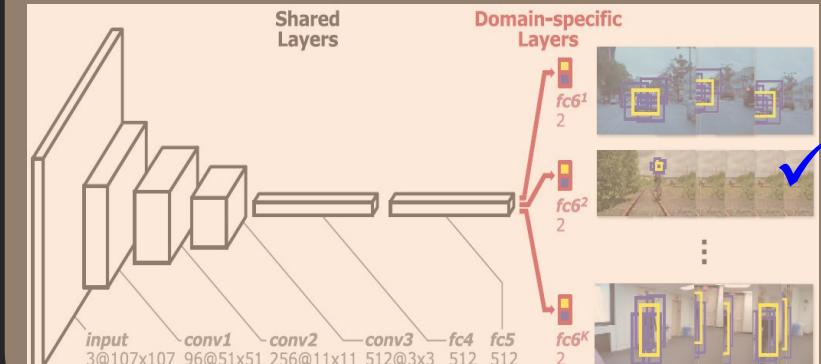
SiamRPN (CVPR18)



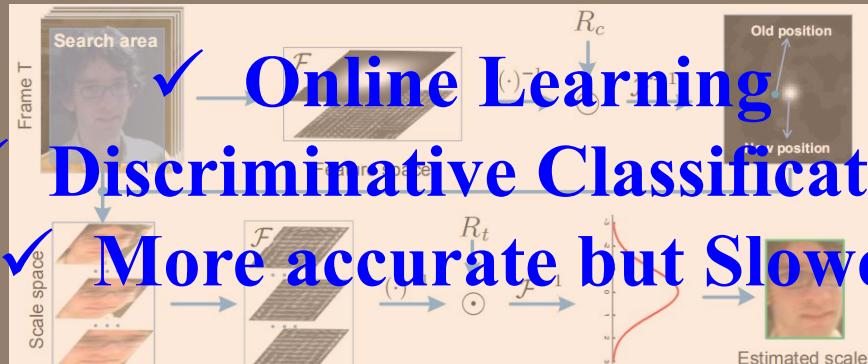
SiamMask (CVPR19)



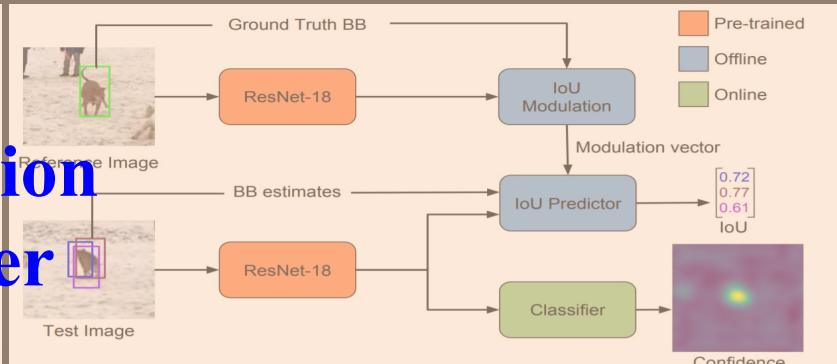
MDNet (CVPR16)



ECO (ICCV17)

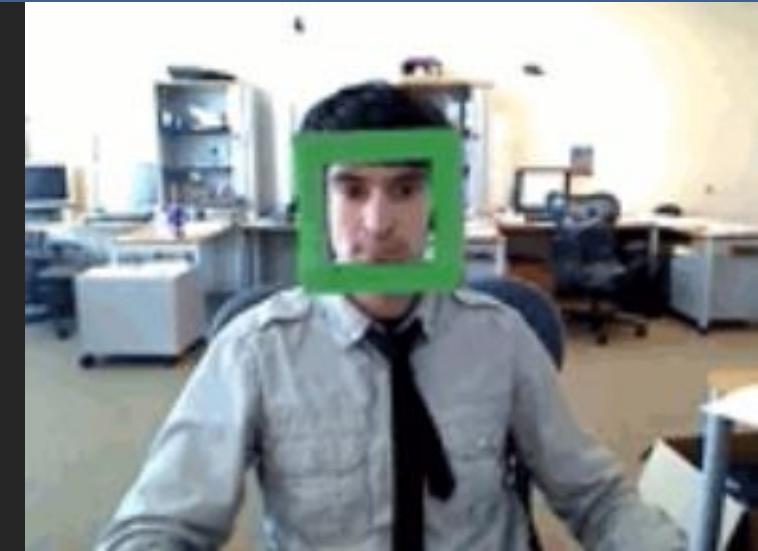


ATOM (CVPR19)



To Freshman

- 王蒙蒙-基于计算机视觉的目标跟踪算法概览
<https://www.bilibili.com/video/av61881441?from=search&seid=7456371587432981285>
 - VALSE合肥 20190412-Workshop 9: 在线视觉跟踪
https://www.iqiyi.com/v_19rsl7fz98.html
 - 踪迹谁知 - 视觉目标跟踪
https://www.iqiyi.com/v_19rsjugqrc.html
 - 将门创投 | 中科院自动化所在读博士生朱政: 高性能单目标跟踪算法的探索
<https://www.bilibili.com/video/av33603018?from=search&seid=6288243859705072663>
 - 【极市】张志鹏-CVPR2019:基于siamese网络的单目标跟踪
https://www.bilibili.com/video/av52113951/?spm_id_from=333.788.videocard.19
 - 【计算机视觉】Siammask目标跟踪与分割统一的算法——作者中科院自动化所王强博士
<https://www.bilibili.com/video/av61878688?from=search&seid=17780490606506838259>
 - 【极市】朱政-基于孪生网络结构的SiamRPN系列目标跟踪算法
<https://www.bilibili.com/video/av67335639/>
 - 【CCF-CV视界无限】CCF-CV视界无限第4期视觉目标跟踪全部PPT
<https://pan.baidu.com/s/13Rwf2N8SyVd3GoQXo1dt3g&shfl=sharepset>
- 活动总结和专家观点参见: <https://mp.weixin.qq.com/s/qJkcurgaac57INOKYoODMg>



Long-term Visual Object Tracking

■ Long-term Video Sequences

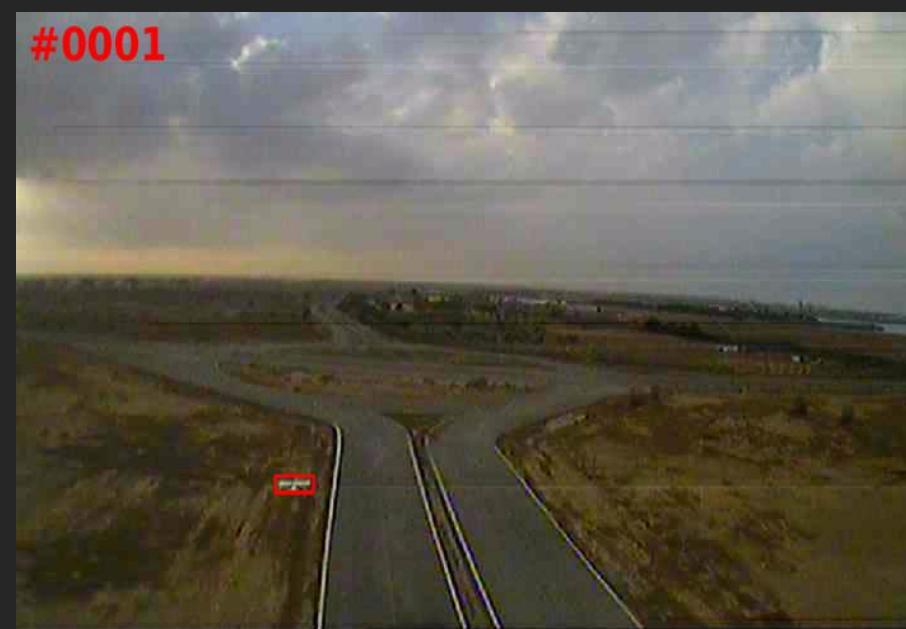
■ Frequent Disappearance and Reappearance

□ Both tracking and re-detection Ability

	LTB35 [18]	OxUvA [30]	OTB100 [33]	VOT2017 [13]
Avg frames	4196	4235	590	350
Max frames	29700	37440	3870	1494
Min frames	1389	900	72	36
absent labels	12%	52%	0%	0%
Avg absent labels	503	2.2	0	0

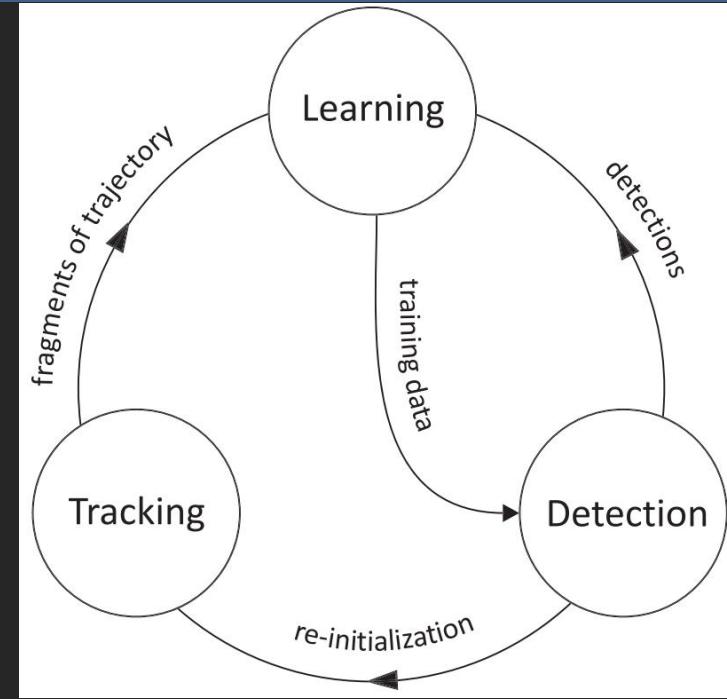


Long-term Visual Object Tracking



Existing Long-term Trackers

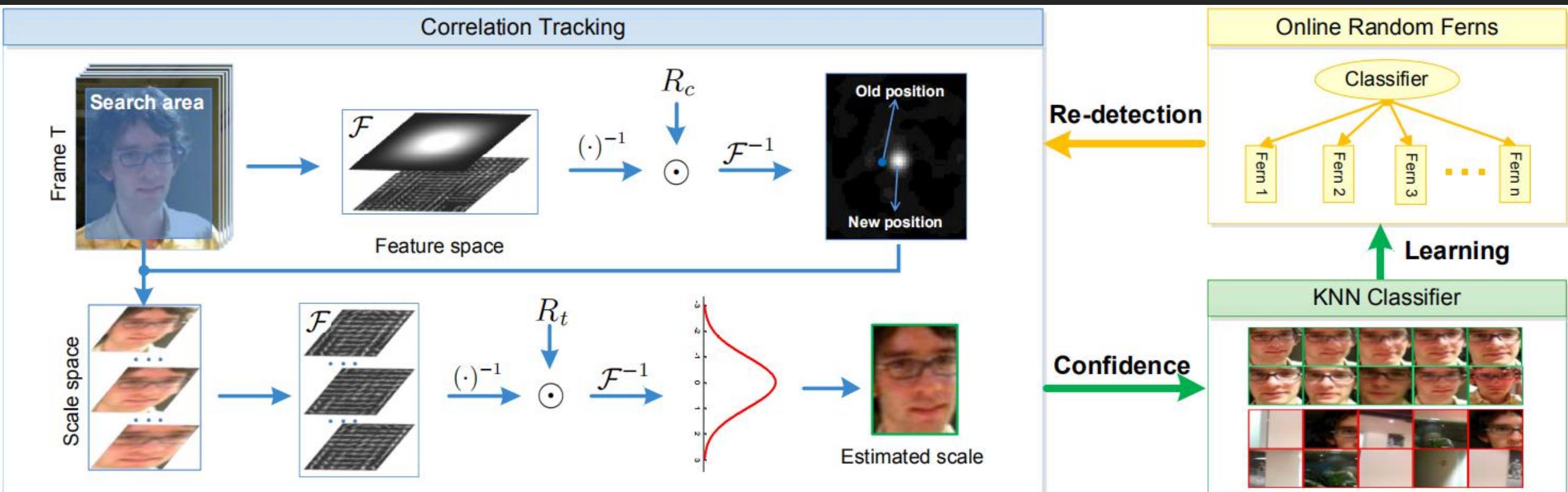
- Optical-flow-based Local Matcher
- Global Re-detection with An Ensemble of Online Weak Classifiers



- TLD: Zdenek Kalal, Krystian Mikolajczyk, Jiri Matas: Tracking-Learning-Detection. IEEE TPAMI. 34(7): 1409-1422 (2012)
<https://github.com/wangdongdut/Long-term-Visual-Tracking>

Existing Long-term Trackers

- Local Tracking with DSST
- Global Re-detection with Online Random Ferns



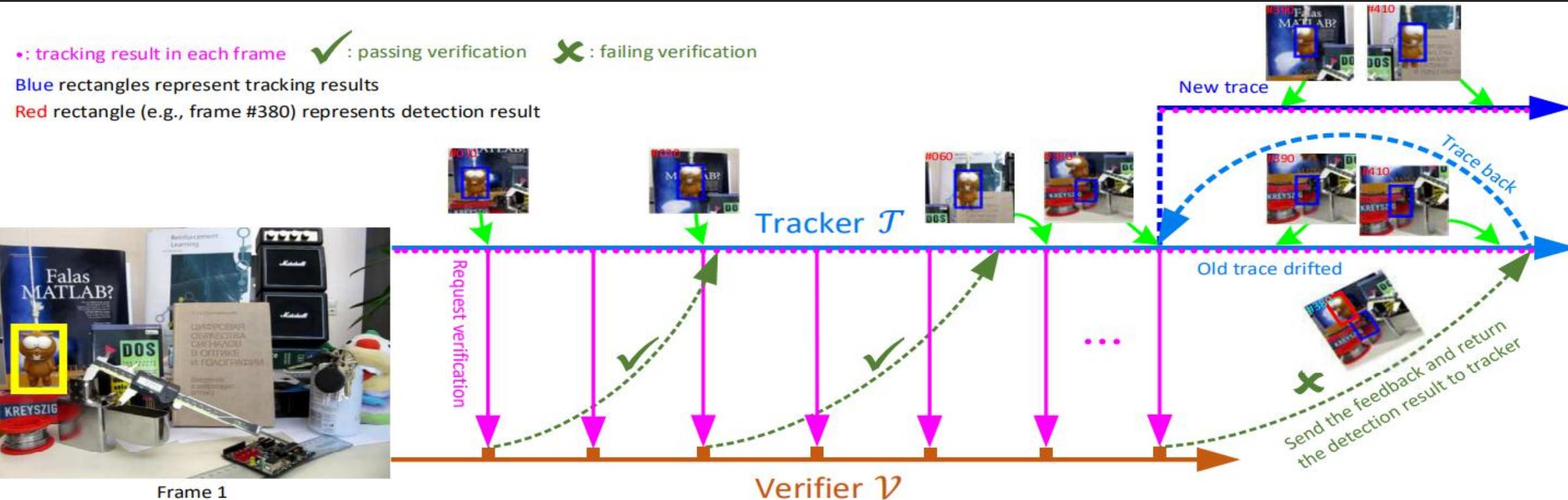
■ LCT: Chao Ma, Xiaokang Yang, Chongyang Zhang, Ming-Hsuan Yang. Long-term correlation tracking. CVPR, 2015.

<https://github.com/wangdongdut/Long-term-Visual-Tracking>

DUT, IIAU-LAB

Existing Long-term Trackers

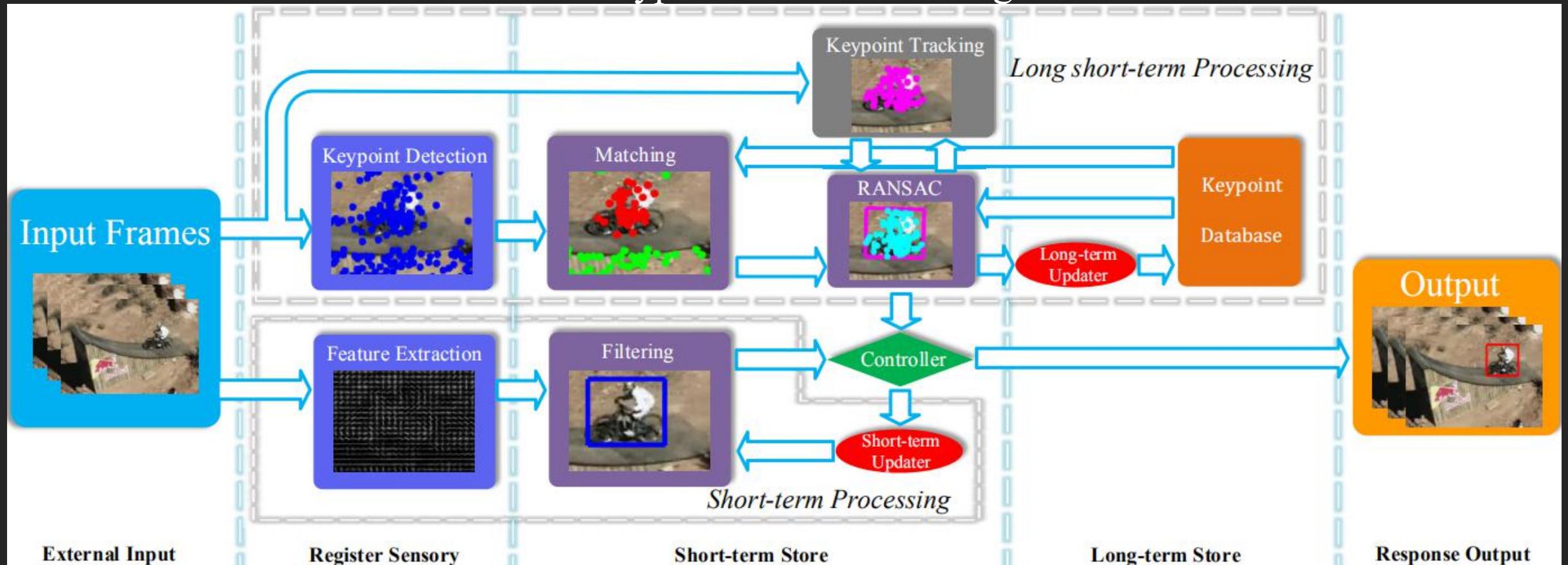
- Local Tracking with DSST
- Global Re-detection with Siamese Network



- PTAV: Heng Fan, Haibin Ling. Parallel Tracking and Verifying: A Framework for Real-Time and High Accuracy Visual Tracking. ICCV, 2017.

Existing Long-term Trackers

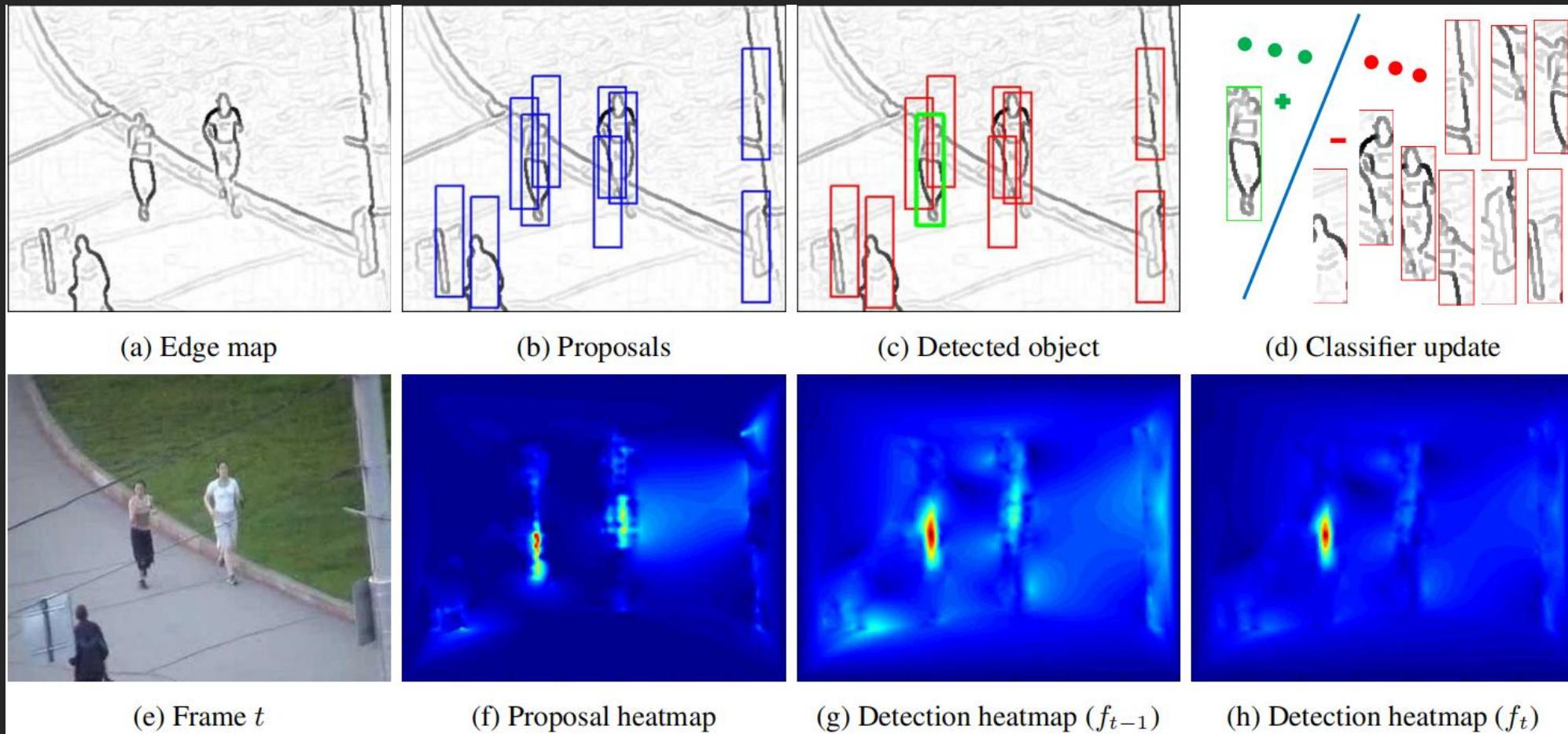
- Local Tracking with An Integrated Correlation Filter
- Global Re-detection with Keypoint-based Matching



- **MUSTer**: Zhibin Hong, Zhe Chen, Chaohui Wang, Xue Mei, Danil V. Prokhorov, Dacheng Tao. MUSTer: A cognitive psychology inspired approach to object tracking. CVPR, 2015.

Existing Long-term Trackers

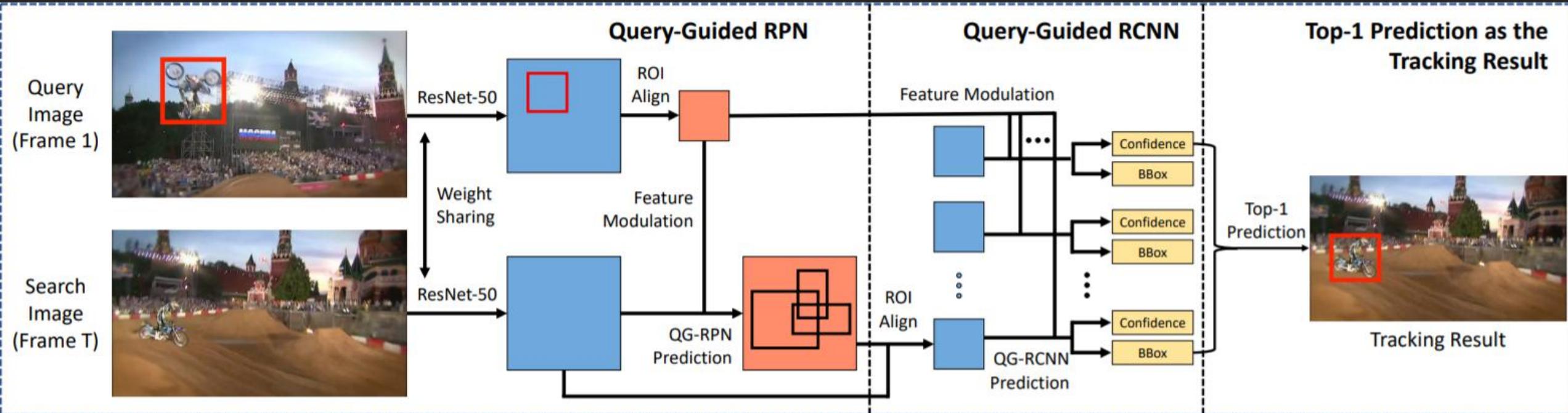
- Global re-detection with Instance-specific Edgebox proposals



- EBT: Gao Zhu, Fatih Porikli, Hongdong Li. Beyond Local Search: Tracking Objects Everywhere with Instance-Specific Proposals. CVPR, 2016.

Existing Long-term Trackers

- Global candidate proposal with Instance-specific (Query-guided) RPN
- Global detection with Instance-specific (Query-guided) RCNN



NO motion model, NO online learning, NO punishment on position or scale changes,
NO scale smoothing and NO trajectory refinement

- **GlobalTrack:** Lianghua Huang, Xin Zhao, Kaiqi Huang. GlobalTrack: A Simple and Strong Baseline for Long-term Tracking. AAAI, 2020. [<https://github.com/huanglianghua/GlobalTrack>]

Long-term Tracking Dataset

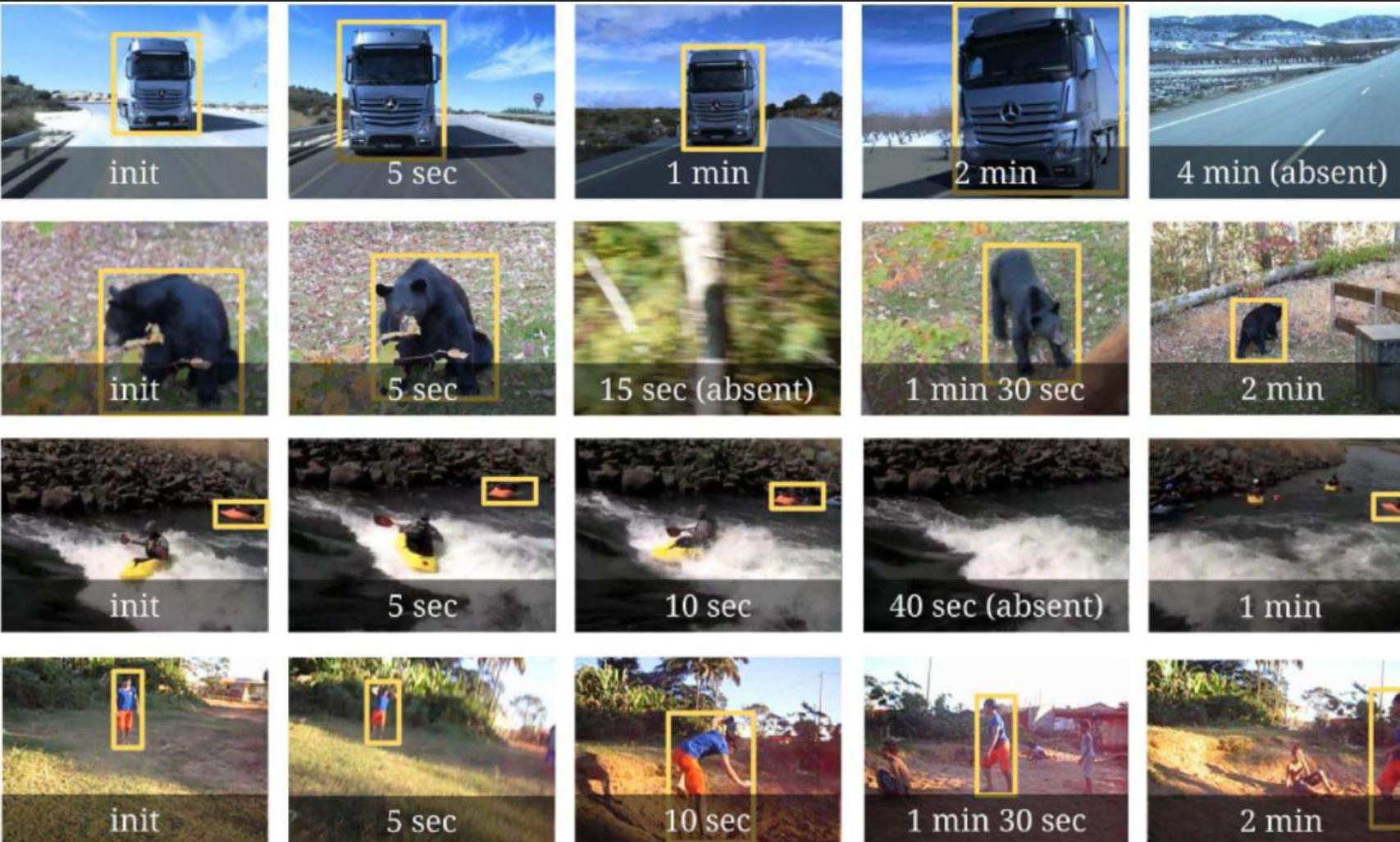
■ VOT 2018/2019/2020 Long-term Challenge



- The VOT-2018 LTB35 dataset is presented in Visual Object Tracking (VOT) challenge 2018 for evaluating long-term trackers, which includes 35 challenging videos of various objects (e.g., persons, car, motorcycles, bicycles and animals) with a total frame length of 146847 frames. Each contains on average 12 long-term target disappearances, and each disappearance lasts on average 40 frames.
- VOT-2019 extends to 50 videos.
- VOT-2020 is same as VOT-2019

Long-term Tracking Dataset

■ OxUvA Long-term Benchmark



- The OxUvA long-term dataset consists of 366 object tracks in 337 videos, which are carefully selected from the YTBB dataset and sparsely labeled at a frequency of 1Hz
- OxUvA dev (with 200 tracks) and OxUvA test (with 166 tracks) sets.

Long-term Tracking Dataset

■ TLP

- 50 HD videos from real-world scenarios, with an average of 13500 frames per sequence.

Abhinav Moudgil, Vineet Gandhi. Long-term Visual Object Tracking Benchmark. ACCV, 2018.
[<https://amoudgl.github.io/tlp/>]

Long-term Tracking Dataset

■ LaSOT

- 1400 challenging sequences (*1120 for training and 280 for testing*) with 70 tracking categories, with an average of 2500 frames per sequence.



Heng Fan, Liting Lin, Fan Yang, Peng Chu, Ge Deng, Sijia Yu, Hexin Bai, Yong Xu, Chunyuan Liao, Haibin Ling, LaSOT: A High-quality Benchmark for Large-scale Single Object Tracking. CVPR, 2019. [<https://cis.temple.edu/lasot/>]

Long-term Visual Object Tracking

- Long-term Tracking vs Short-term Tracking
- ‘Skimming-Perusal’ Long-term Tracking (ICCV19)
- Long-term Tracking with Meta-Updater (CVPR20)
- New Opportunities and Challenges



‘Skimming-Perusal’ Tracking

‘Skimming-Perusal’ Tracking:

A Framework for Real-Time and Robust Long-term Tracking

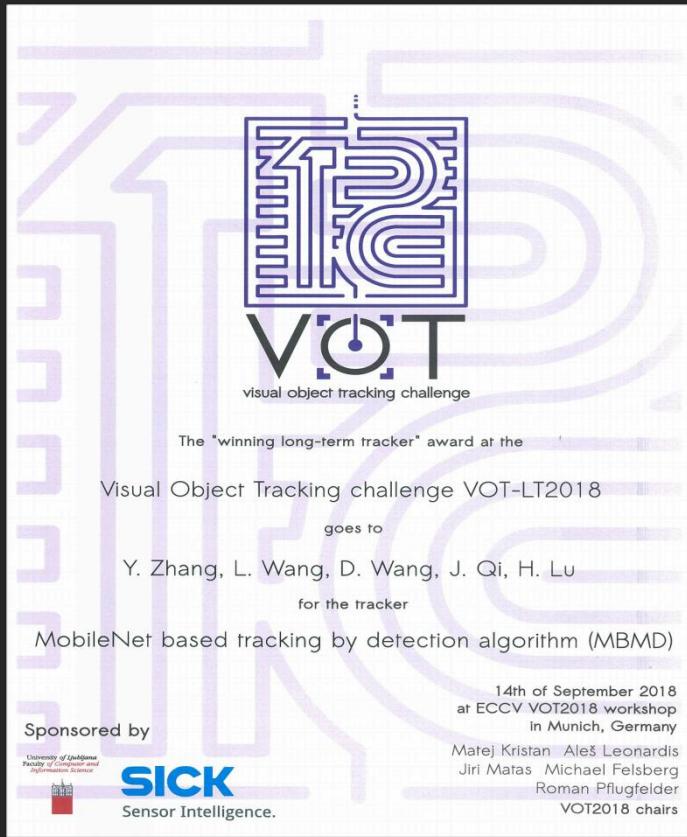
(SPLT)

Bin Yan, Haojie Zhao, Dong Wang*, Huchuan Lu, Xiaoyun Yang, ‘Skimming-Perusal’ Tracking:
A Framework for Real-Time and Robust Long-term Tracking, ICCV 2019.

➤ Code: <https://github.com/iiau-tracker/SPLT>

VOT18 Long-term Winner

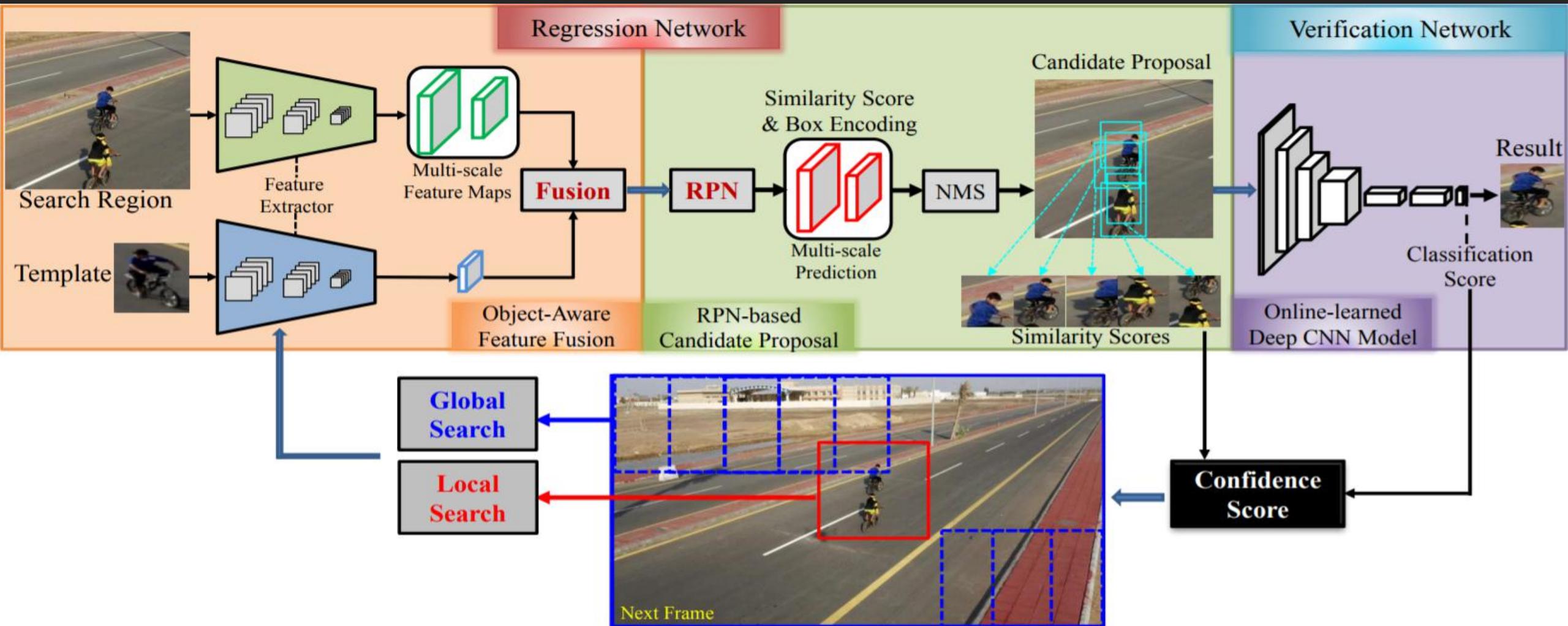
The winner of the VOT18 Long-term Challenge



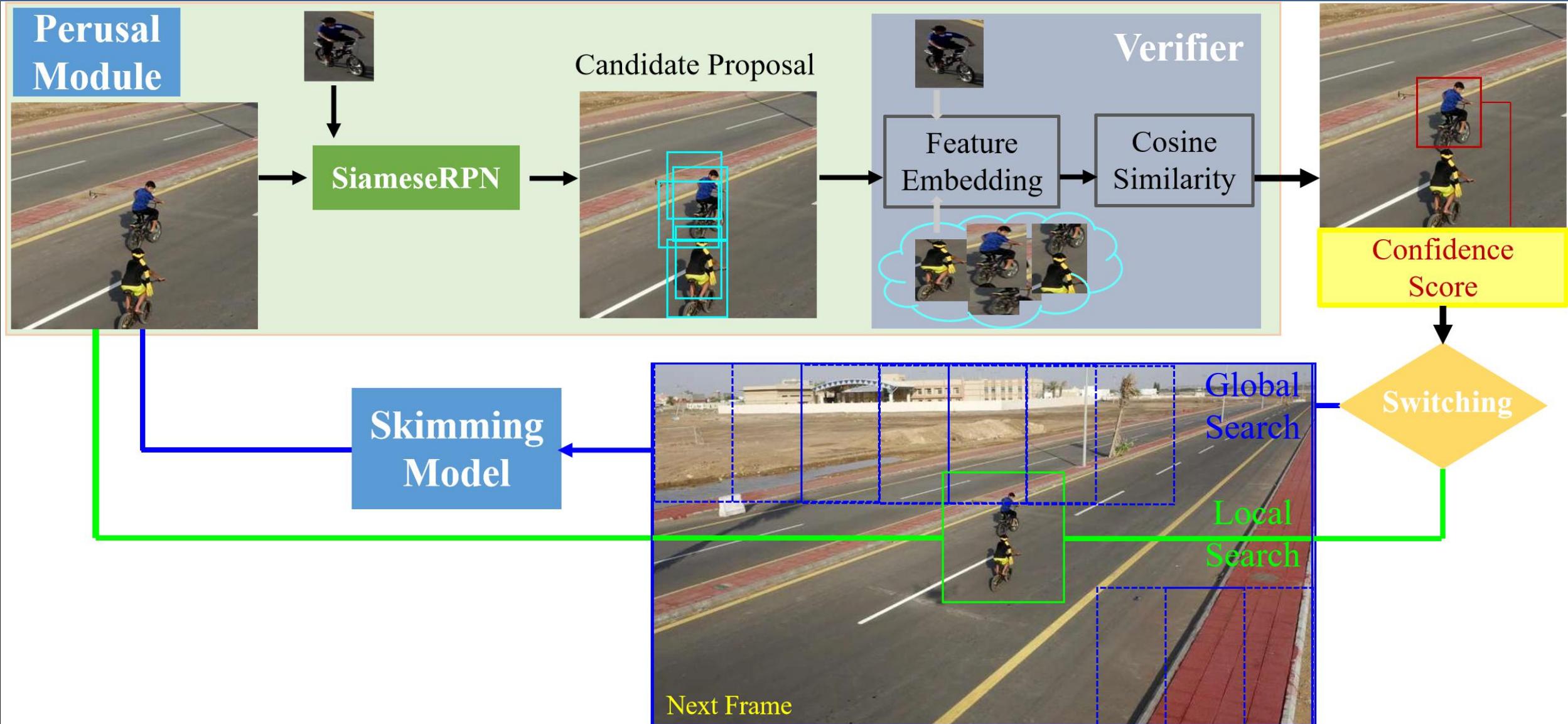
	Tracker	F-score	Pr	Re	ST/LT	Frames (Success)
1.	MBMD	0.610 ①	0.634 ②	0.588 ①	LT ₁	1 (100%)
2.	DaSiam_LT	0.607 ②	0.627 ③	0.588 ②	LT ₁	- (0%)
3.	MMLT	0.546 ③	0.574	0.521 ③	LT ₁	0 (100%)
4.	LTSINT	0.536	0.566	0.510	LT ₁	2 (100%)
5.	SYT	0.509	0.520	0.499	LT ₁	0 (43%)
6.	PTAVplus	0.481	0.595	0.404	LT ₁	0 (11%)
7.	FuCoLoT	0.480	0.539	0.432	LT ₁	78 (97%)
8.	SiamVGG	0.459	0.552	0.393	ST ₀ → LT ₀	- (0%)
9.	SLT	0.456	0.502	0.417	ST ₁ → LT ₀	0 (100%)
10.	SiamFC	0.433	0.636 ①	0.328	ST ₀ → LT ₀	- (0%)
11.	SiamFCDet	0.401	0.488	0.341	LT ₁	0 (83%)
12.	HMMTxD	0.335	0.330	0.339	LT ₁	3 (91%)
13.	SAPKLT	0.323	0.348	0.300	LT ₀	- (0%)
14.	ASMS	0.306	0.373	0.259	ST ₀ → LT ₀	- (0%)
15.	FoT	0.119	0.298	0.074	ST ₀ → LT ₀	0 (6%)

- Code: <https://github.com/xiaobai1217/MBMD> ➤ <https://zhuanlan.zhihu.com/p/44737353>
➤ VOT18: <https://www.votchallenge.net/vot2018/> ➤ <https://arxiv.org/abs/1809.04320>

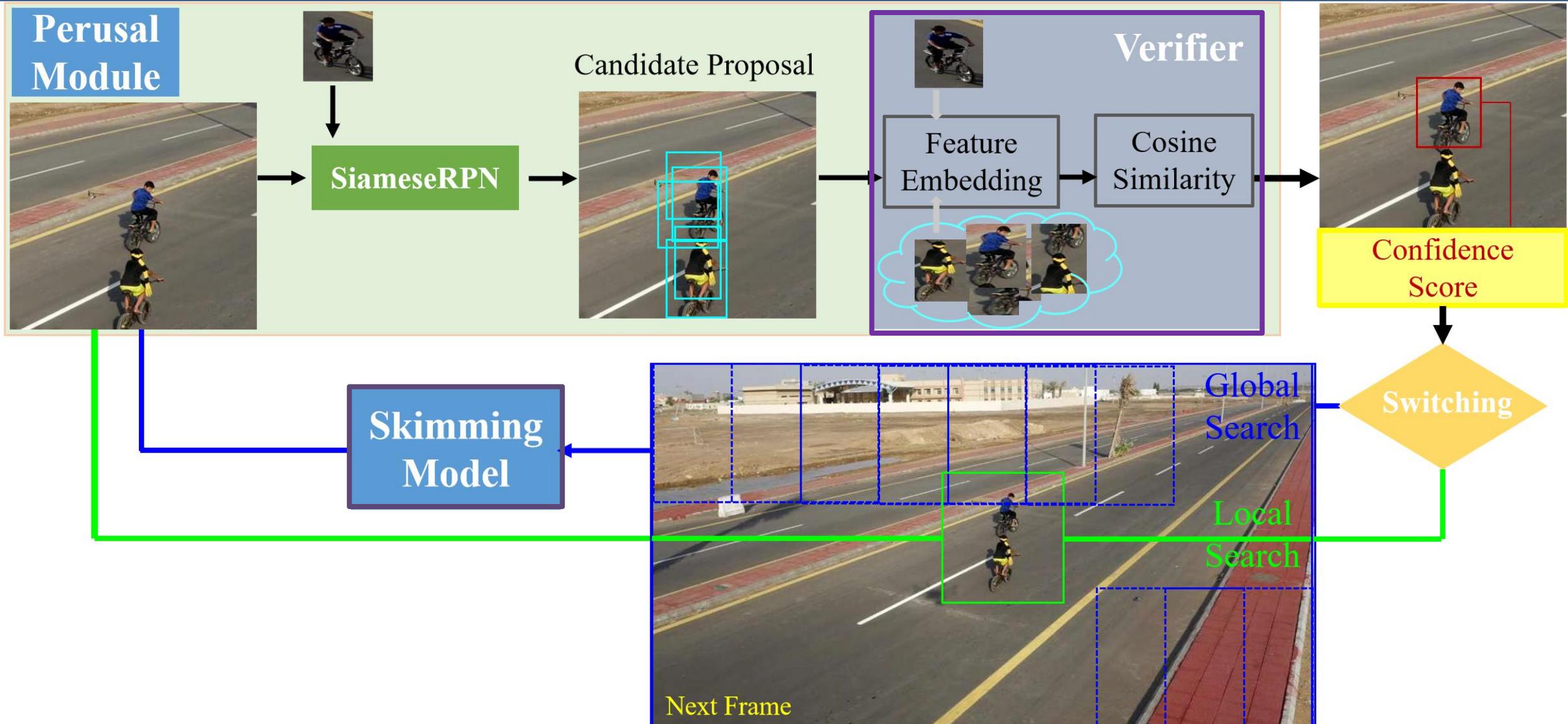
VOT18 Long-term Winner



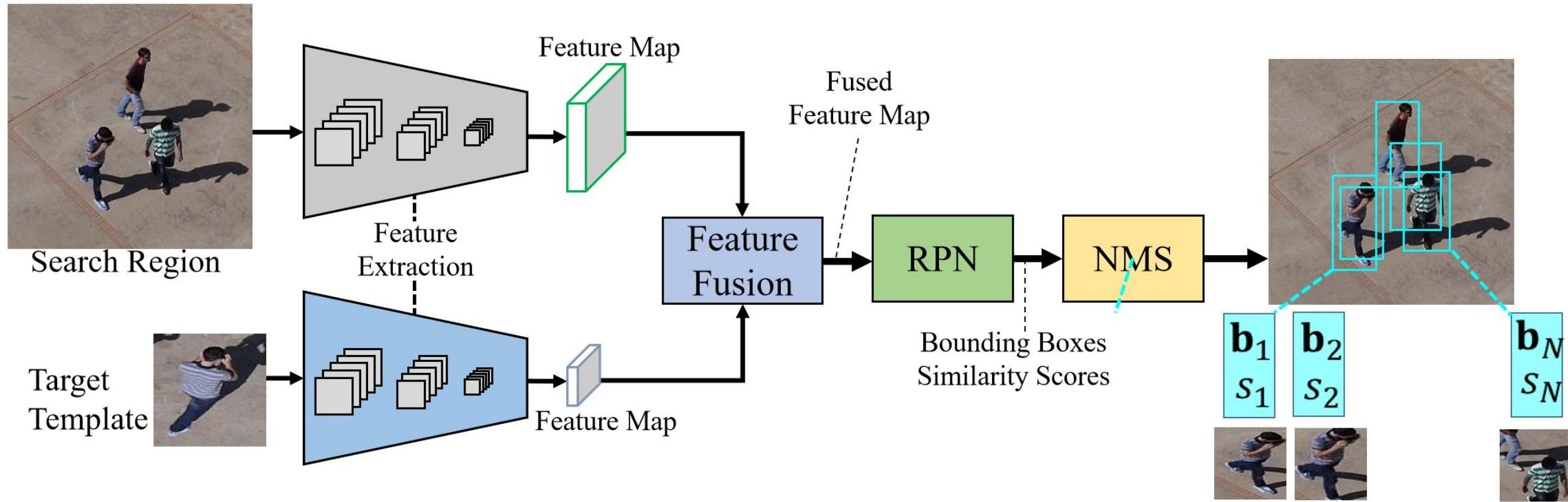
‘Skimming-Perusal’ Tracking



‘Skimming-Perusal’ Tracking



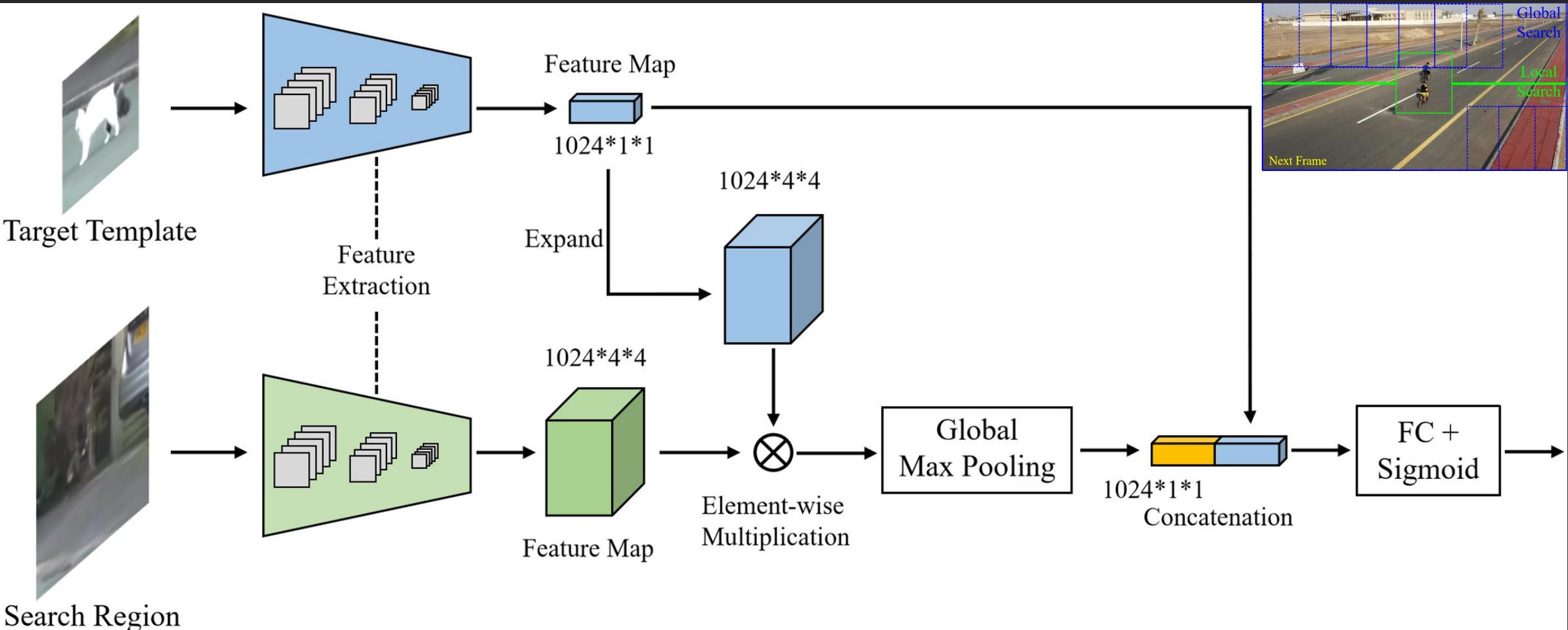
SPLT-SiameseRPN



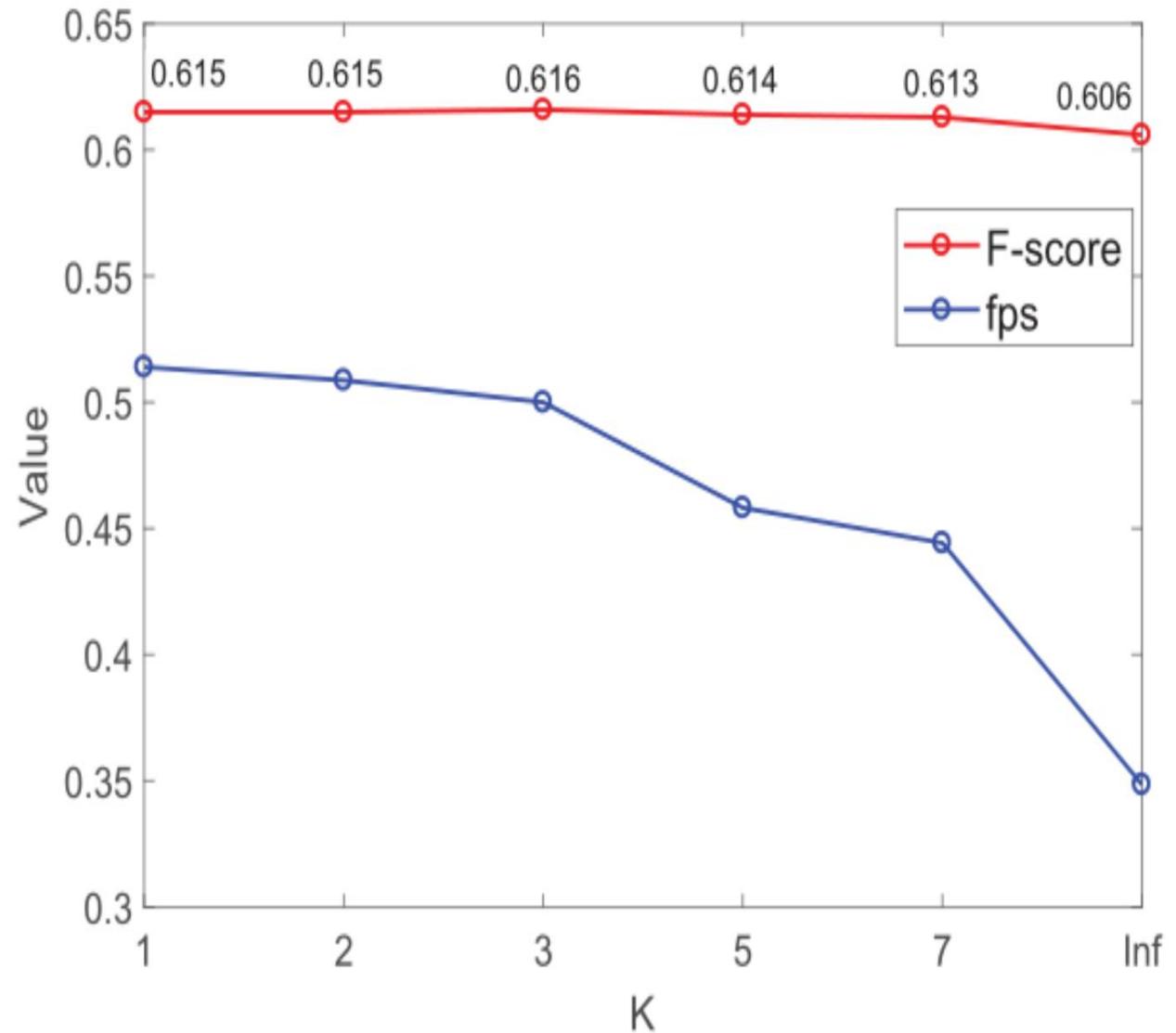
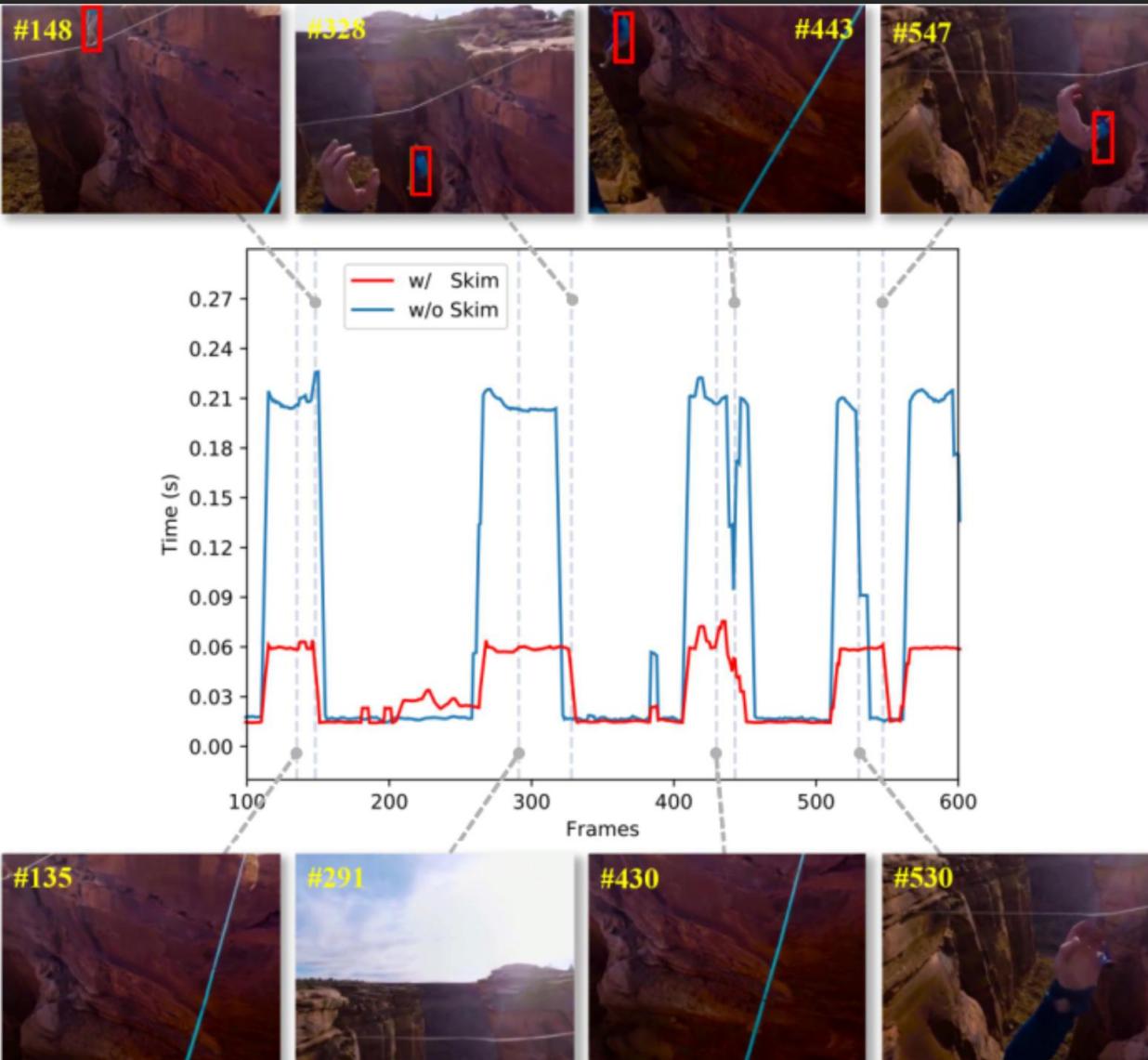
Related Codes:

- <https://github.com/xiaobai1217/MBMD> [Ours]
- <https://github.com/STVIR/pysot> [SenseTime]
- <https://github.com/researchmm/SiamDW> [MSRA]

SPLT-Skimming



SPLT-Skimming



VOT 2018 Long-term Challenge

Tracker	F-score	Pr	Re	Frames (Success)
Ours	0.616	0.633	0.600	1 (100%)
MBMD	0.610	0.634	0.588	1 (100%)
DaSiam_LT	0.607	0.627	0.588	- (0%)
MMLT	0.546	0.574	0.521	0 (100%)
LTSINT	0.536	0.566	0.510	2 (100%)
SYT	0.509	0.520	0.499	0 (43%)
PTAVplus	0.481	0.595	0.404	0 (11%)
FuCoLoT	0.480	0.539	0.432	78 (97%)
SiamVGG	0.459	0.552	0.393	- (0%)
SLT	0.456	0.502	0.417	0 (100%)
SiamFC	0.433	0.636	0.328	- (0%)
SiamFCDet	0.401	0.488	0.341	0 (83%)
HMMTxD	0.335	0.330	0.339	3 (91%)
SAPKLT	0.323	0.348	0.300	- (0%)
ASMS	0.306	0.373	0.259	- (0%)
FoT	0.119	0.298	0.074	- (6%)

25.7fps
4.4fps

✓ A PC machine with an
Intel i7 CPU (32G RAM)
and a NVIDIA GTX1080Ti
GPU (11G memory)

VOT 2018 Long-term Challenge

ballet

Long-term Visual Object Tracking

- Long-term Tracking vs Short-term Tracking
- ‘Skimming-Perusal’ Long-term Tracking (ICCV19)
- Long-term Tracking with Meta-Updater (CVPR20)
- New Opportunities and Challenges



Long-Term Tracking with Meta-Updater

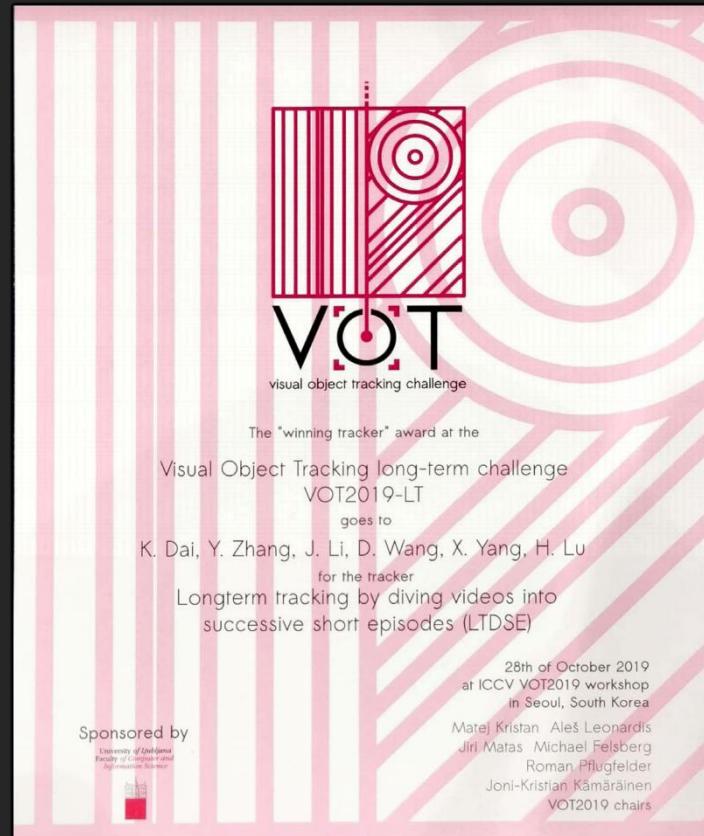
High-Performance Long-Term Tracking with Meta-Updater (LTMU)

Kenan Dai, Yunhua Zhang, Dong Wang*, Jianhua Li, Huchuan Lu, Xiaoyun Yang. High-Performance Long-Term Tracking with Meta-Updater, CVPR, 2020.

➤ Code: <https://github.com/Daikenan/LTMU>

Long-Term Tracking with Meta-Updater

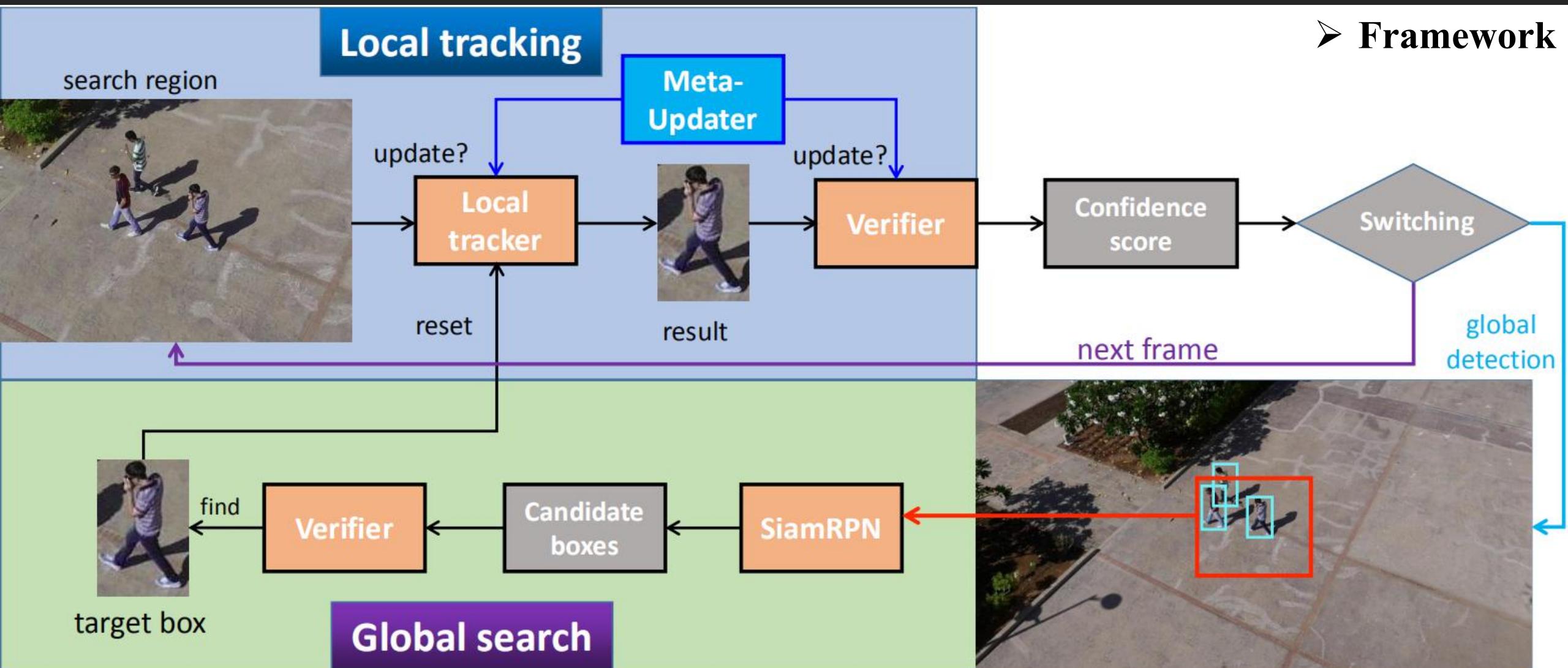
The winner of the VOT19 Long-term Challenge



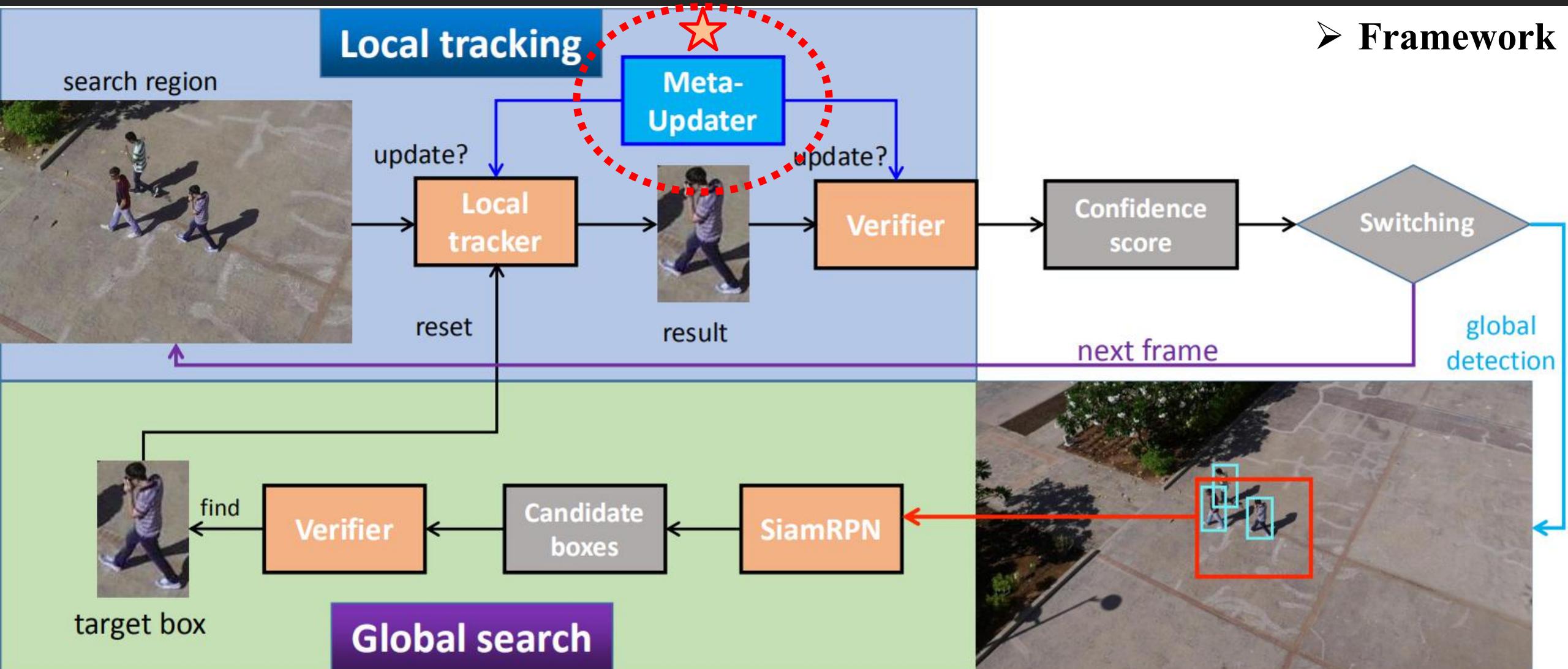
Tracker	F-score	Pr	Re	ST/LT
1. LT_DSE	0.695 ①	0.715 ③	0.677 ①	LT ₁
2. CLGS	0.674 ②	0.739 ②	0.619 ③	LT ₁
3. SiamDW_LT	0.665 ③	0.697	0.636 ②	LT ₁
4. mbdet	0.567	0.609	0.530	LT ₁
5. SiamRPNsLT	0.556	0.749 ①	0.443	LT ₁
6. Siamfcos-LT	0.520	0.493	0.549	LT ₁
7. CooSiam	0.508	0.482	0.537	LT ₁
8. ASINT	0.505	0.517	0.494	LT ₁
9. FuCoLoT	0.411	0.507	0.346	LT ₁

- Code: https://github.com/Daikenan/LT_DSE
- VOT19: <https://www.votchallenge.net/vot2019/>

Long-Term Tracking with Meta-Updater

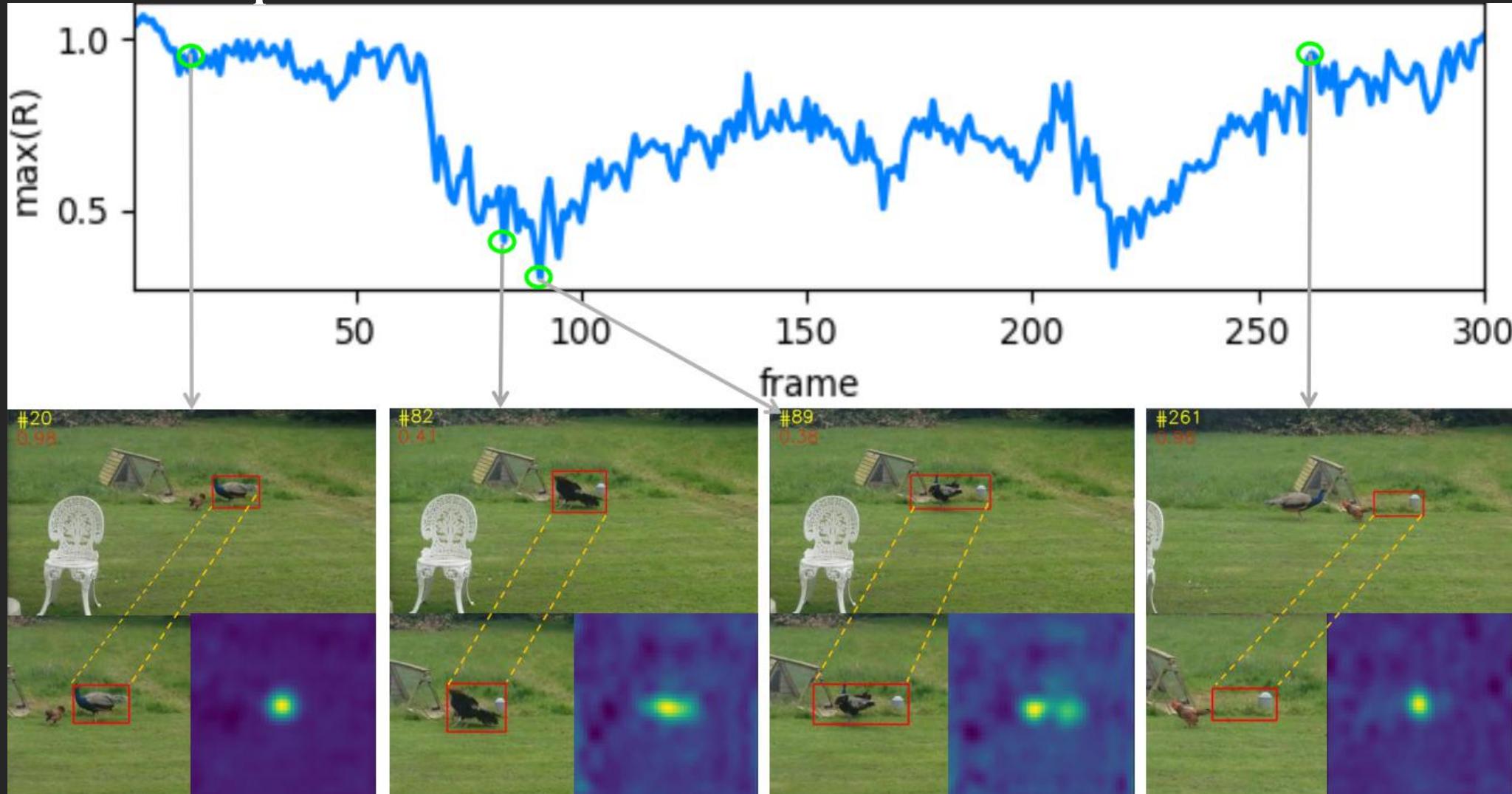


Long-Term Tracking with Meta-Updater

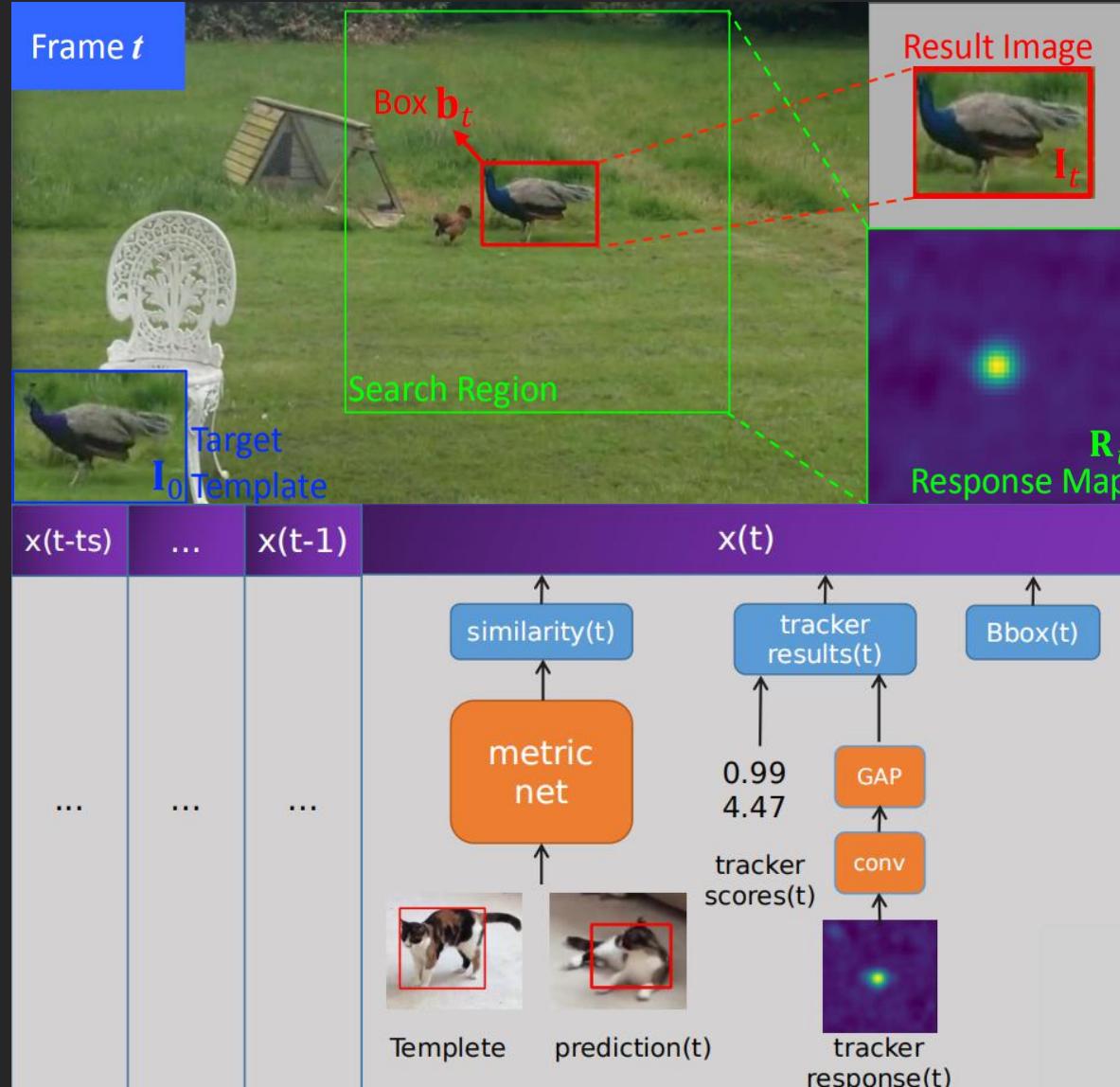


Long-Term Tracking with Meta-Updater

➤ Max Response?



Long-Term Tracking with Meta-Updater



➤ Sequential Information

- Geometric Cue:

$$\mathbf{b}_t = [x_t, y_t, w_t, h_t]$$

- Discriminative Cue:

- ✓ Score:

$$s_t^C = \max(\mathbf{R}_t)$$

- ✓ Map:

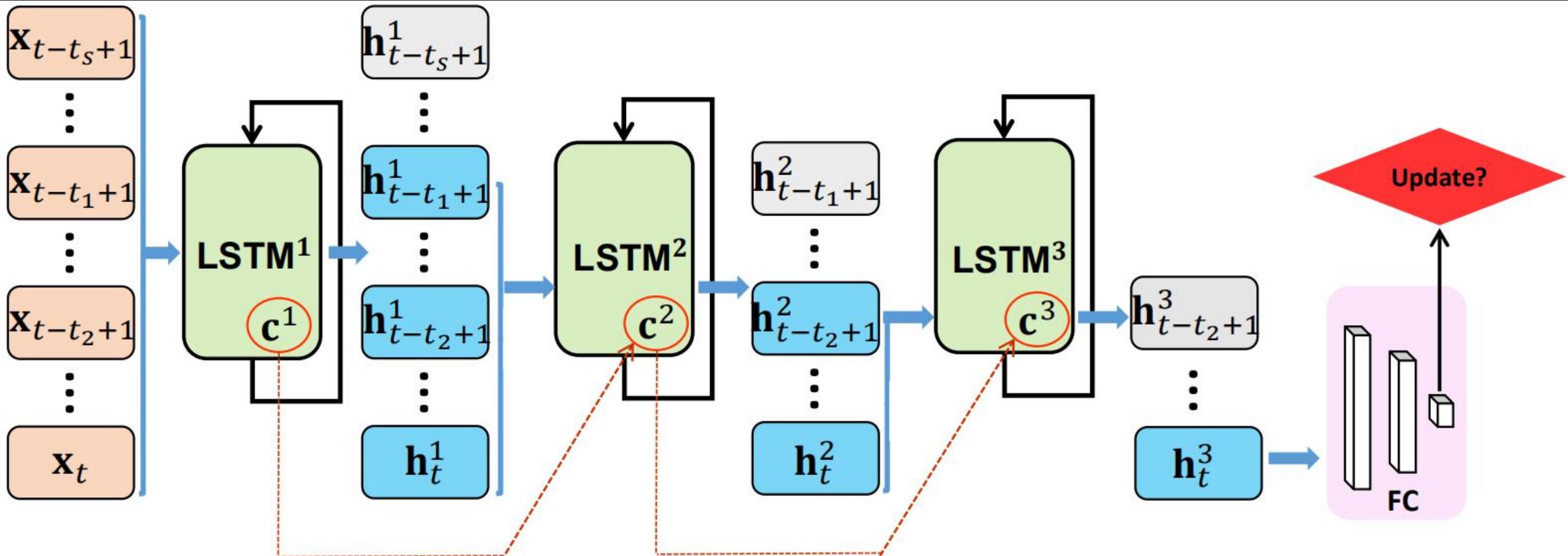
$$\mathbf{v}_t^R = f^R(\mathbf{R}_t; \mathbf{W}^R)$$

- Appearance Cue:

$$s_t^A = \|f^A(\mathbf{I}_t, \mathbf{W}^A) - f^A(\mathbf{I}_0, \mathbf{W}^A)\|_2$$

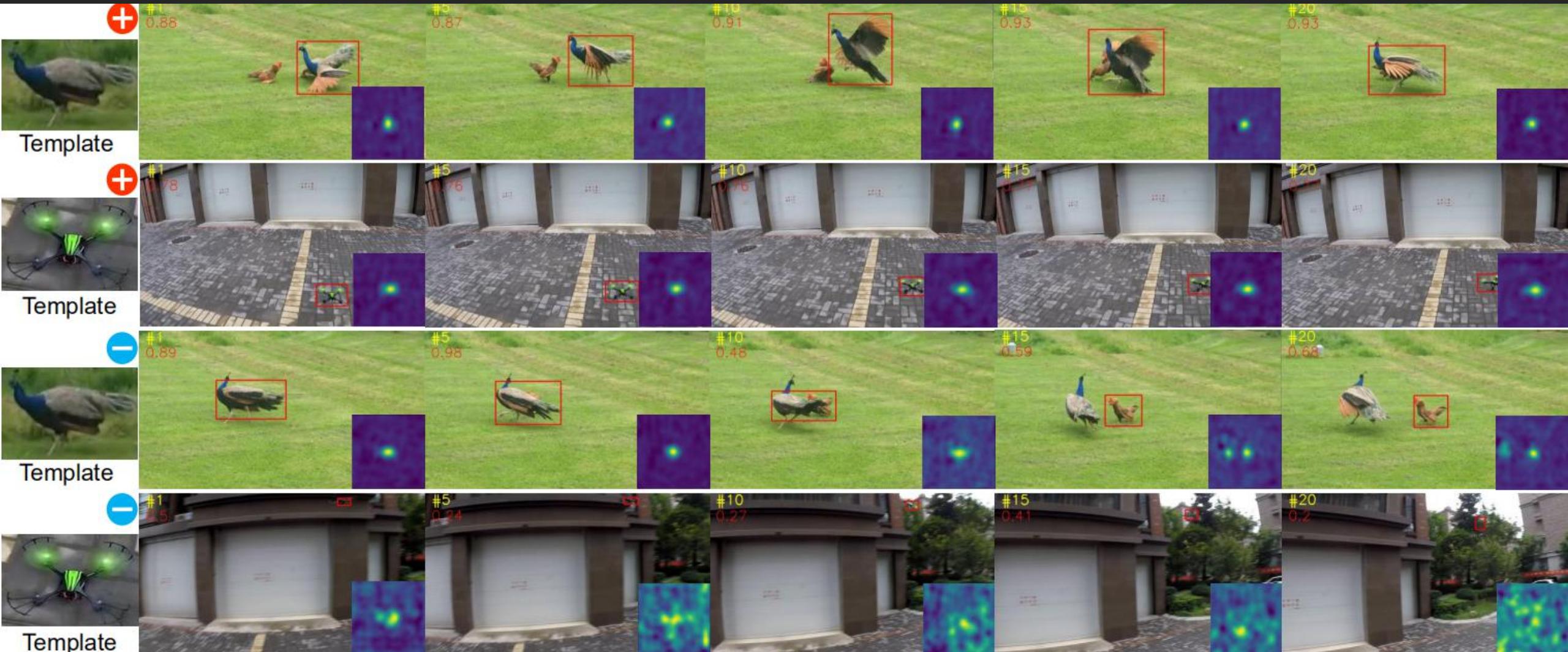
Long-Term Tracking with Meta-Updater

- Three-stage Cascaded LSTMs



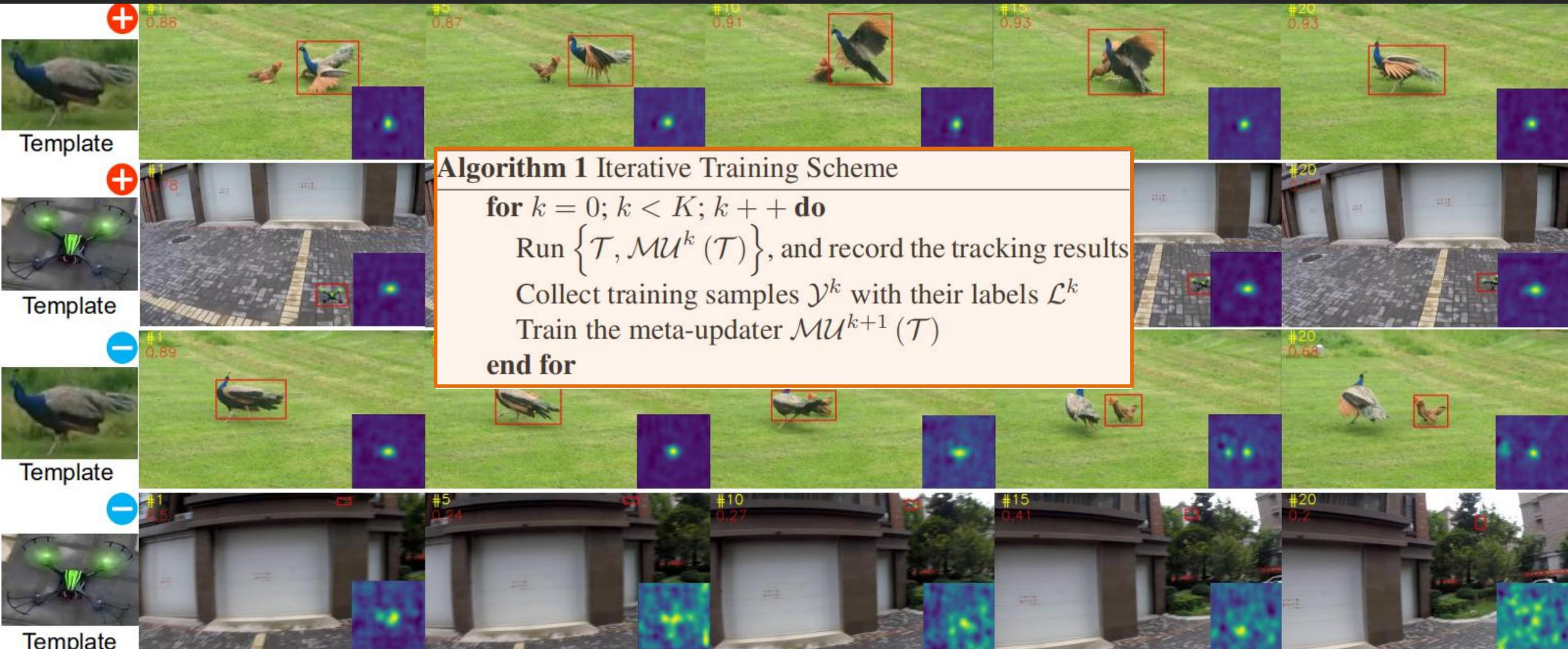
Long-Term Tracking with Meta-Updater

► Training



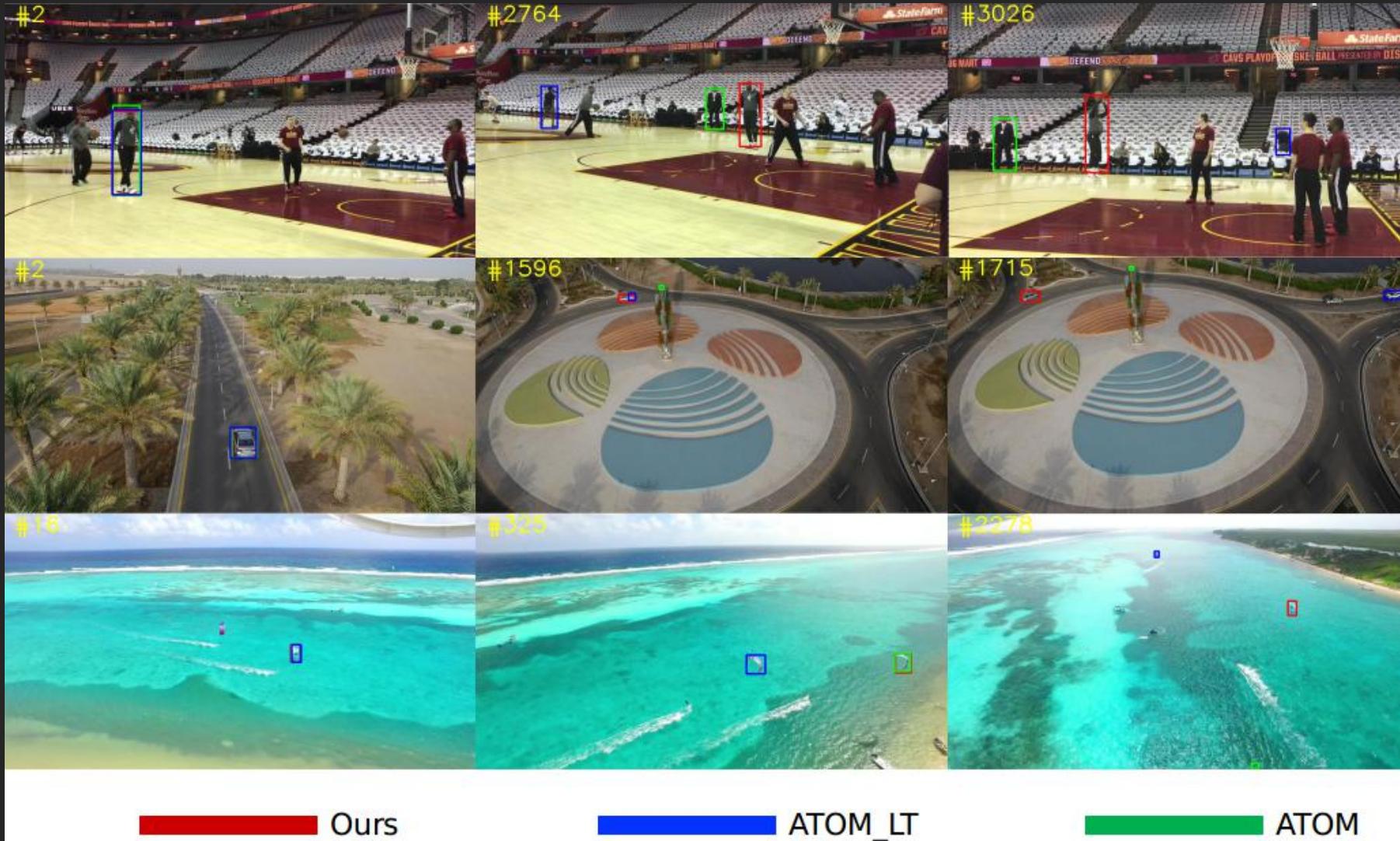
Long-Term Tracking with Meta-Updater

► Training



Long-Term Tracking with Meta-Updater

➤ Visual Results



Ours

ATOM_LT

ATOM

Long-Term Tracking with Meta-Updater

➤ VOT2018LT

Tracker	F-score	Pr	Re
LTMU(Ours)	0.690	0.710	0.672
SiamRPN++	0.629	0.649	0.609
SPLT	0.616	0.633	0.600
MBMD	0.610	0.634	0.588
DaSiam_LT	0.607	0.627	0.588
MMLT	0.546	0.574	0.521
LTSINT	0.536	0.566	0.510
SYT	0.509	0.520	0.499
PTAVplus	0.481	0.595	0.404
FuCoLoT	0.480	0.539	0.432
SiamVGG	0.459	0.552	0.393
SLT	0.456	0.502	0.417
SiamFC	0.433	0.636	0.328
SiamFCDet	0.401	0.488	0.341
HMMTxD	0.335	0.330	0.339
SAPKLT	0.323	0.348	0.300
ASMS	0.306	0.373	0.259

ICCV2019
VOT2018LT
Winner

➤ OxUvALT

Tracker	MaxGM	TPR	TNR
LTMU(Ours)	0.751	0.749	0.754
SPLT	0.622	0.498	0.776
GlobalTrack	0.603	0.574	0.633
MBMD	0.544	0.609	0.485
SiamFC+R	0.454	0.427	0.481
TLD	0.431	0.208	0.895
LCT	0.396	0.292	0.537
MDNet	0.343	0.472	0
SINT	0.326	0.426	0
ECO-HC	0.314	0.395	0
SiamFC	0.313	0.391	0
EBT	0.283	0.321	0
BACF	0.281	0.316	0
Staple	0.261	0.273	0

→ AAAI
2020

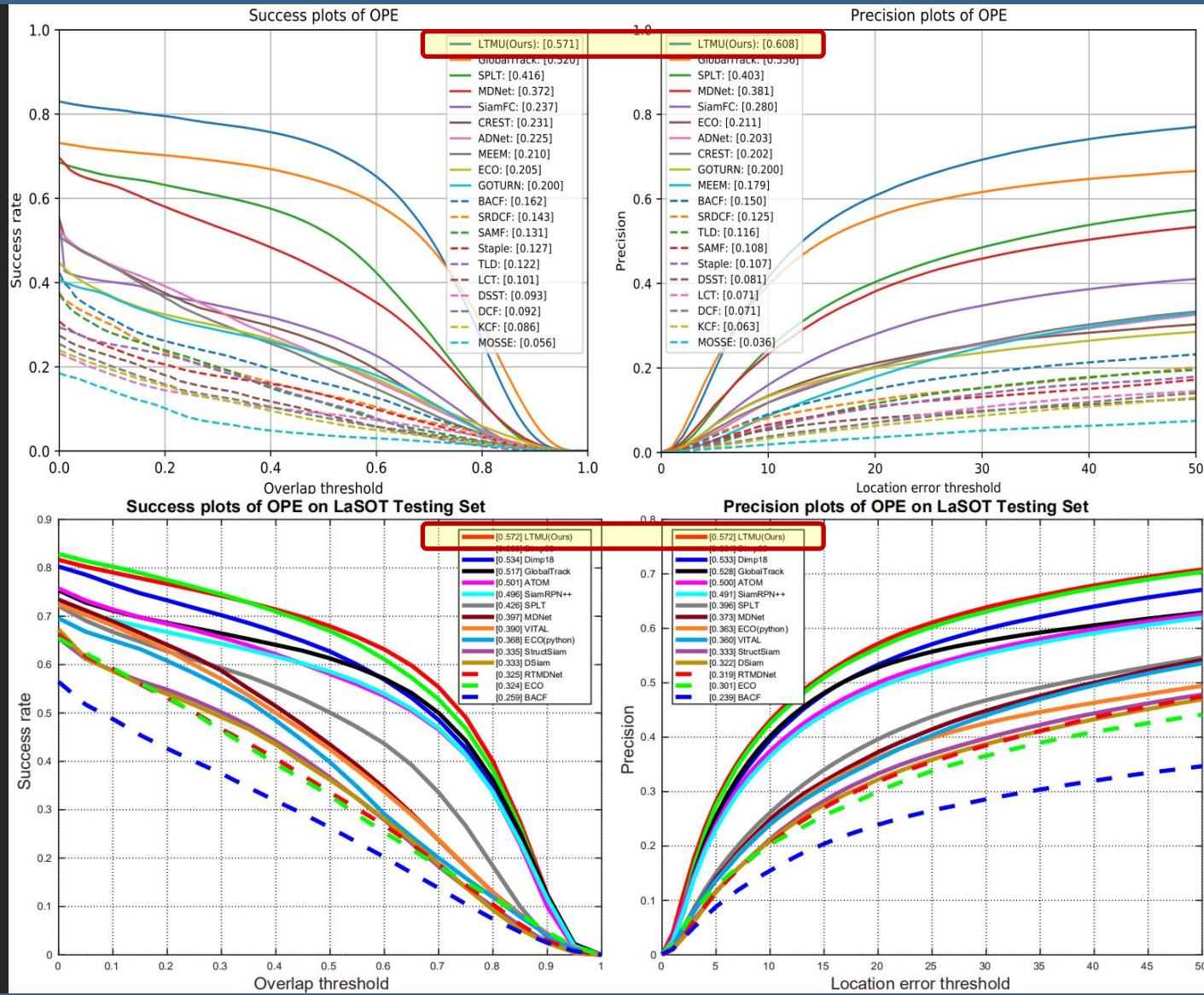
Long-Term Tracking with Meta-Updater

➤ TLP

➤ VOT2019LT

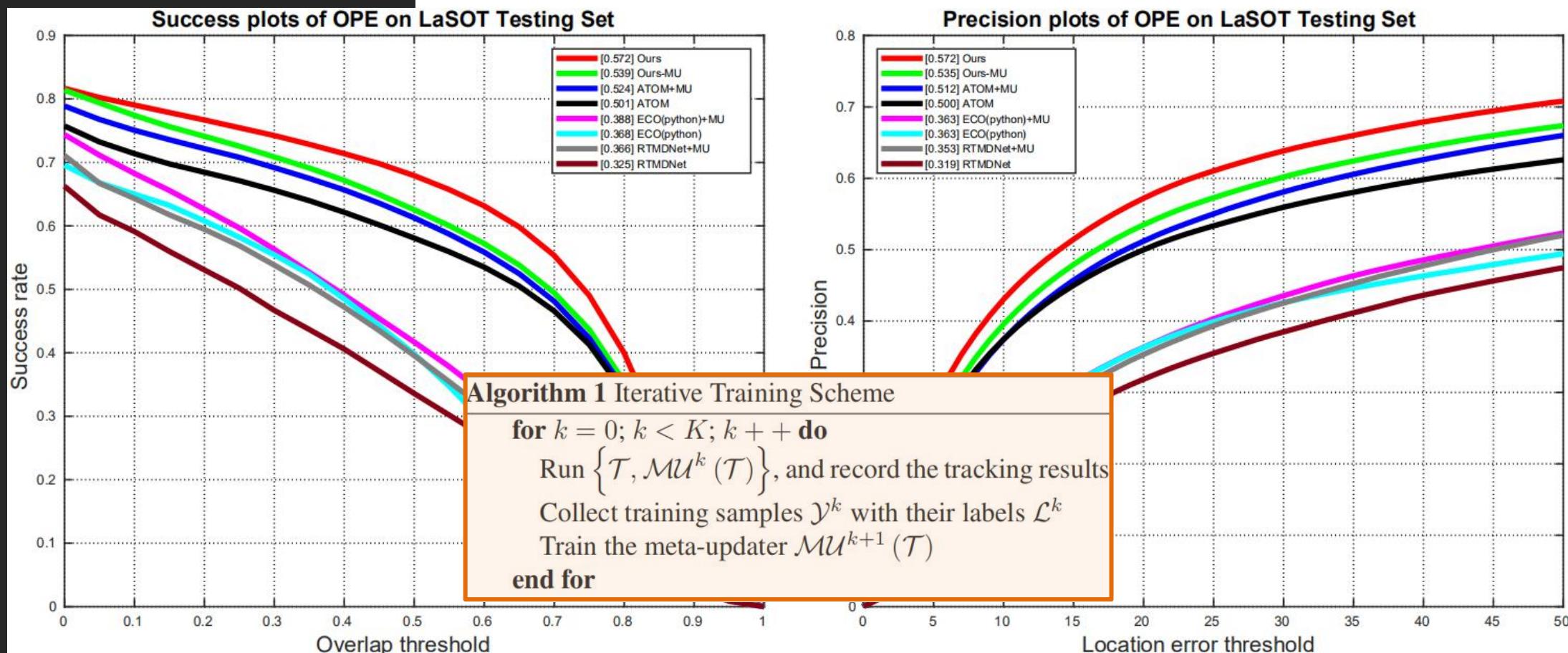
Tracker	F-score	Pr	Re
LTMU(Ours)	0.697	0.721	0.674
CLGS	0.674	0.739	0.619
SiamDW_LT	0.665	0.697	0.636
mbdet	0.567	0.609	0.530
SiamRPNsLT	0.556	0.749	0.443
Siamfcos-LT	0.520	0.493	0.549
CooSiam	0.508	0.482	0.537
ASINT	0.505	0.517	0.494
FuCoLoT	0.411	0.507	0.346

➤ LaSOT



Long-Term Tracking with Meta-Updater

➤ Generalization Ability



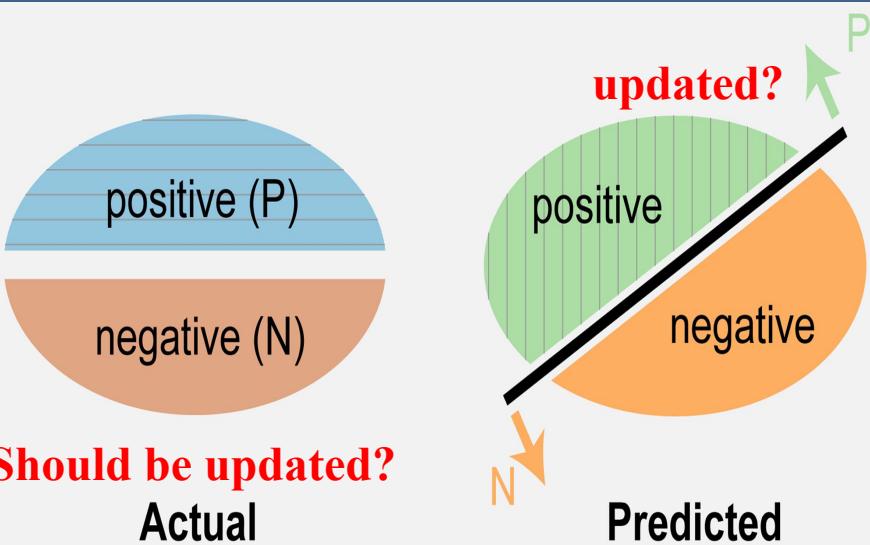
Long-Term Tracking with Meta-Updater

➤ Why it works?

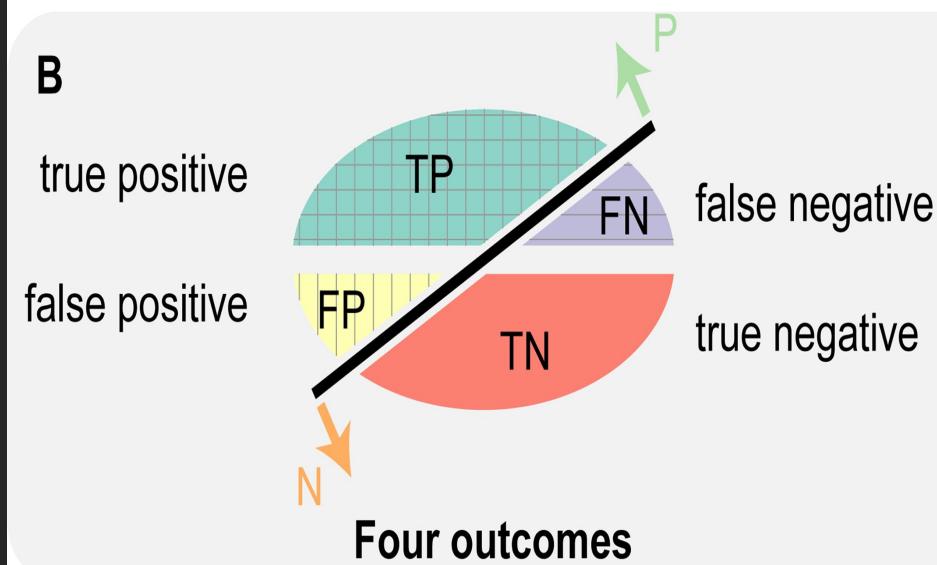
Tracker	Pr	Re	TNR
RTMDNet	0.599	0.993	0.402
RTMDNet+MU	0.909	0.902	0.898
ECO	0.583	1.000	0.000
ECO+MU	0.852	0.895	0.803
ATOM	0.765	0.997	0.310
ATOM+MU	0.931	0.886	0.845
Ours-MU	0.867	0.994	0.479
Ours	0.952	0.874	0.862

- $\text{Pr} = \text{TP}/(\text{TP}+\text{FP})$
- $\text{Re} = \text{TP}/(\text{TP}+\text{FN})$
- $\text{TNR} = \text{TN}/(\text{TN}+\text{FP})$

A



B



Long-term Visual Object Tracking

- Long-term Tracking vs Short-term Tracking
- ‘Skimming-Perusal’ Long-term Tracking (ICCV19)
- Long-term Tracking with Meta-Updater (CVPR20)
- New Opportunities and Challenges



New Opportunities and Challenges

➤ Dataset

- OTB2015 (100)
- VOT2019 (60)
- TC128 (128)
- PTB (100)
- ALOV300+ (314)
- UAV123 (123)
- NUS-PRO (365)
- NFS (100)
- TrackingNet (30,643)
- GOT-10K (10,000)
- LaSOT (1,400)

...

More than 50,000 Sequences!

Short-term
Tracking

Long-term
Tracking

- LTDT (6)
- NUS-PRO LT (5)
- UAV20L (20)
- VOT2018LT (35)
- VOT2019LT (50)
- VOT2020LT (50)
- TLP (50)
- OxUvA (337)

...

Just about 500 Sequences!

New Opportunities and Challenges

➤ Framework and Key Component

- A New Framework? Local tracking-Global Redetection?
- Efficient Global Search? EfficientDet?
- Learning-based Switching?
- Temporal Information Mining?
- Off-line Training?
- More Training Data?
-

New Opportunities and Challenges

➤ Machine Learning Algorithms

- One-shot Learning v.s. Online Learning?
- Auto-learning (NAS)?
- Reinforcement Learning?
- Transfer Learning?
- Positive Unlabeled learning?
- Learning with Noisy Samples?
-

Thank You



(Q&A)