

The Visual Object Tracking Challenge Results VOT-ST2020, VOT-RT2020, VOT-LT2020

Matej Kristan, Jiri Matas, Aleš Leonardis, Michael Felsberg, Roman Pflugfelder, Joni-Kristian Kämäräinen, Luka Čehovin Zajc, Gustavo Fernandez, Alan Lukežič, Martin Danelljan, Ondrej Drbohlav, Linbo He, Yan Song, Jinyu Yang, Yushan Zhang, Goutam Bhat

Outline

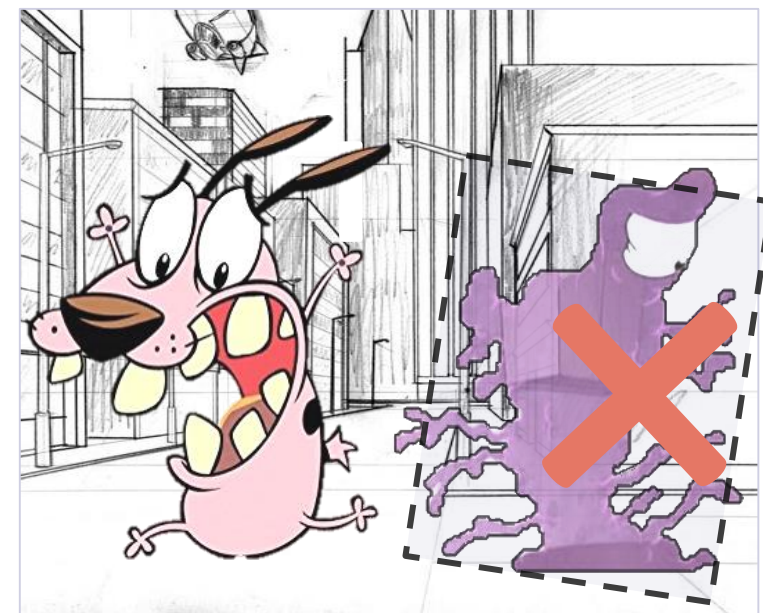
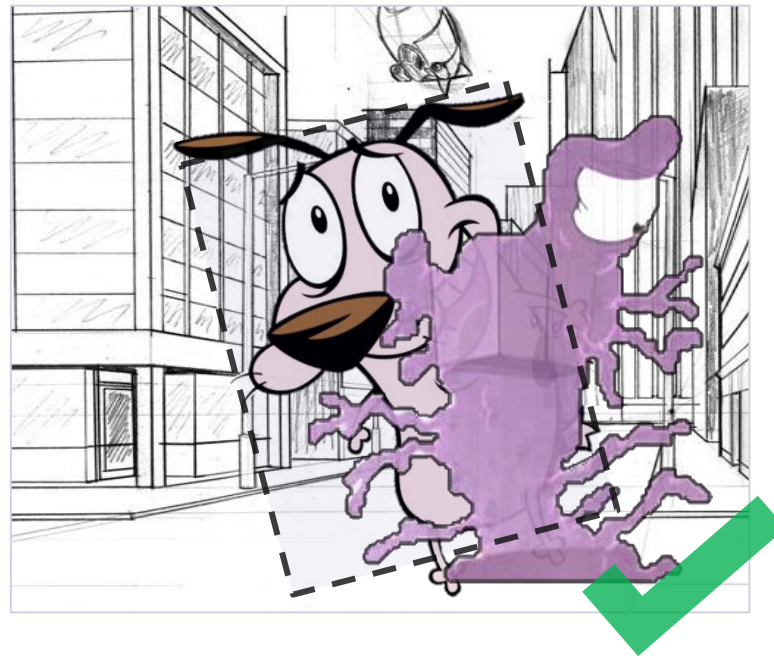
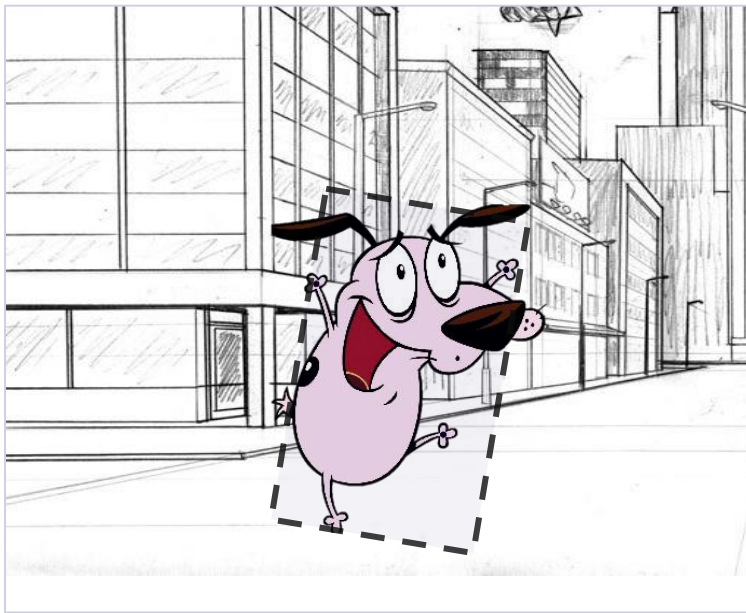
1. Scope of the VOT2020 ST/RT/LT challenges
2. Results overview (VOT2020 ST/RT/LT)
3. Winner announcement (VOT2020 ST/RT/LT)

The VOT 2020 workshop

VOT2020 ST/RT CHALLENGES: OVERVIEW

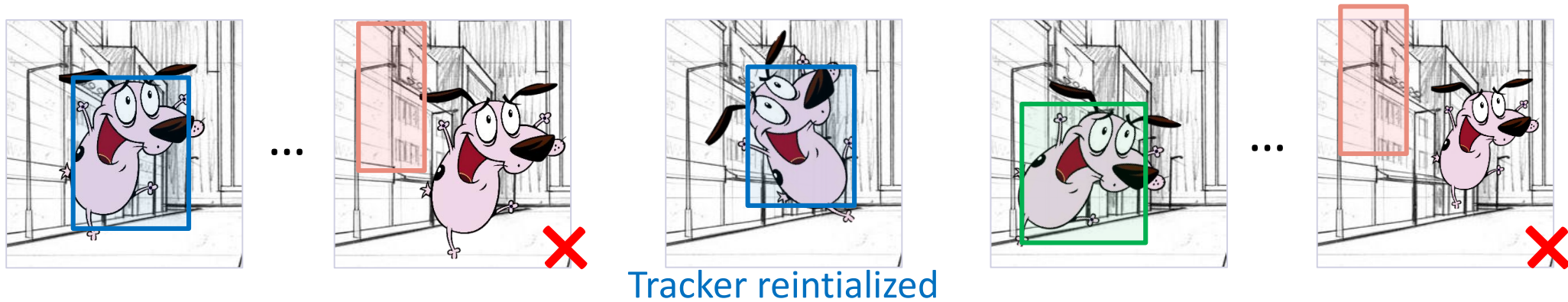
VOT-ST(2013-2019) challenge scope

- Mission: promote trackers that track for as long as possible
- Short-term, single-target, causal trackers
- Tracker reports the target **state location** (e.g., a bounding box)
- No redetection: drift is considered a failure and **tracker is reset**

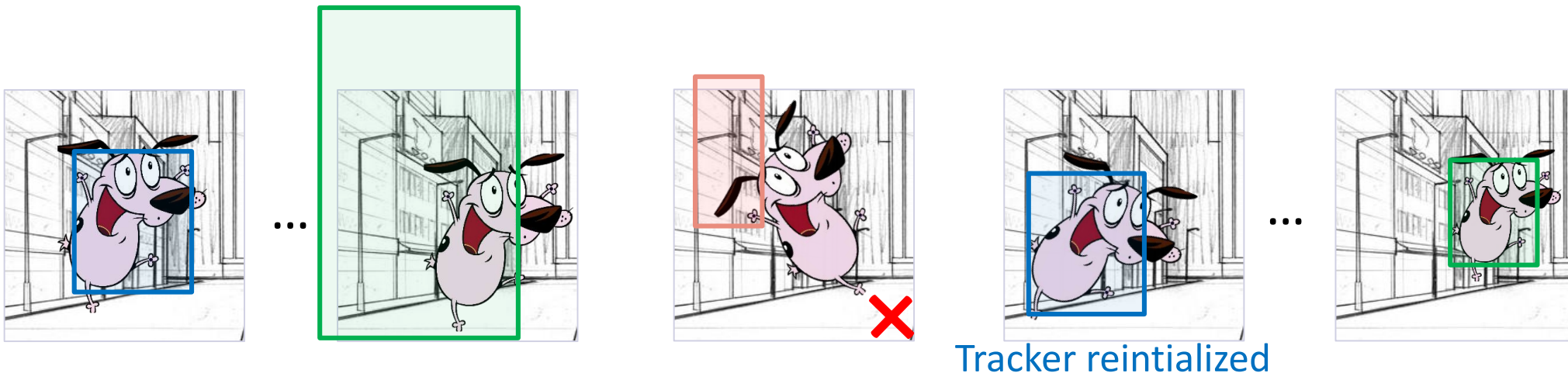


... but trackers are getting better

- A failure at some frame affects the next failure (a tuning opportunity)

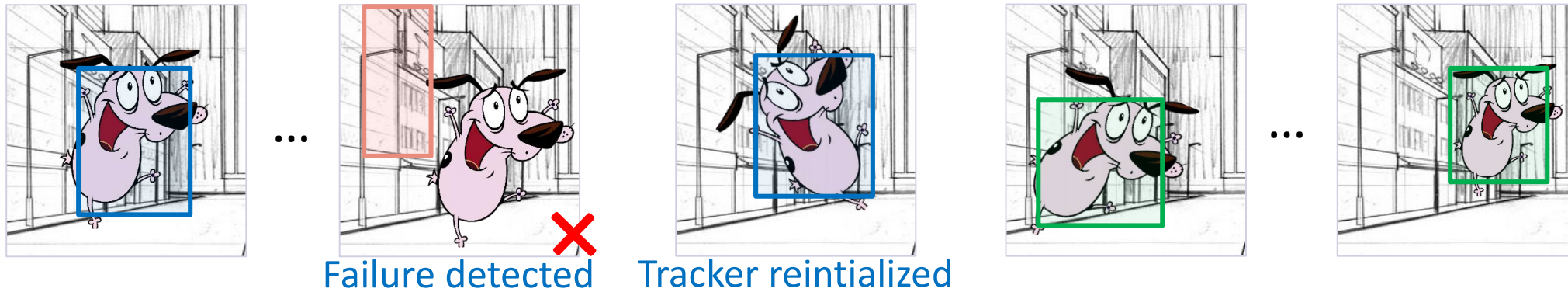


- Intentional bounding box over-inflation

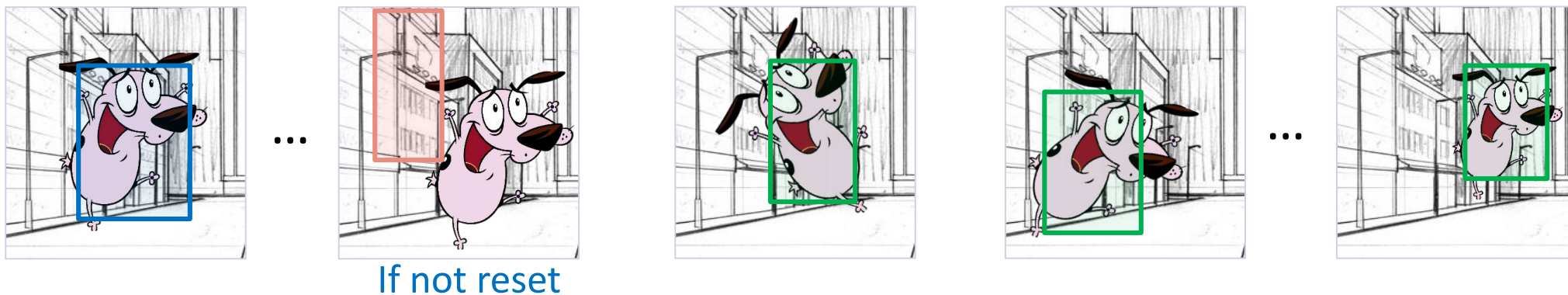


... but trackers are getting better

- Failure definition (0 overlap) penalizes even short-term failures

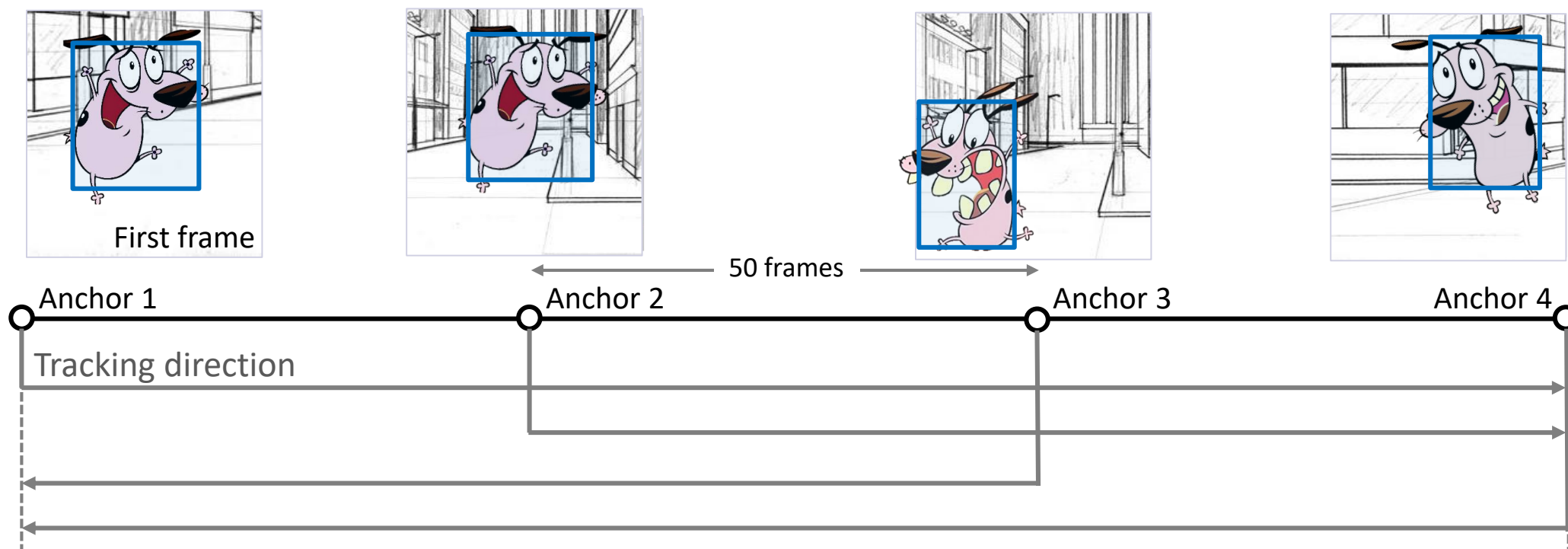


- A tracker might have recovered from a *short-term* failure



VOT-ST2020: VOT Anchor-based protocol

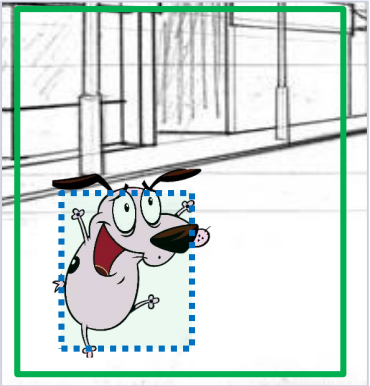
- Introduce initialization points (anchors) equal for all trackers
- Track in the direction of the largest number of tracking frames
- Each anchor produces one subsequence



- All trackers initialized on the same init-points

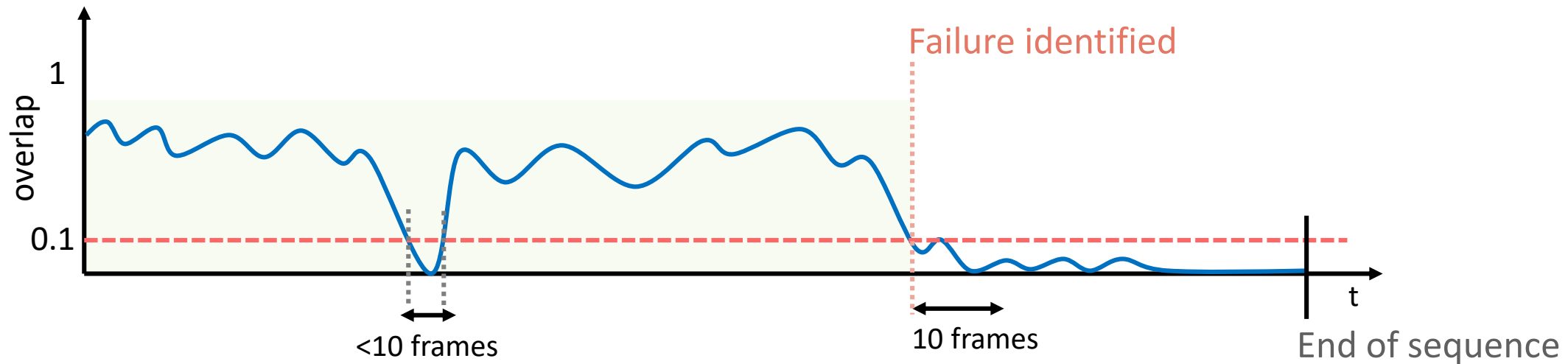
Accounting for short-term failure recovery

- Potential failure = overlap < $\theta_{\Phi} = 0.1$



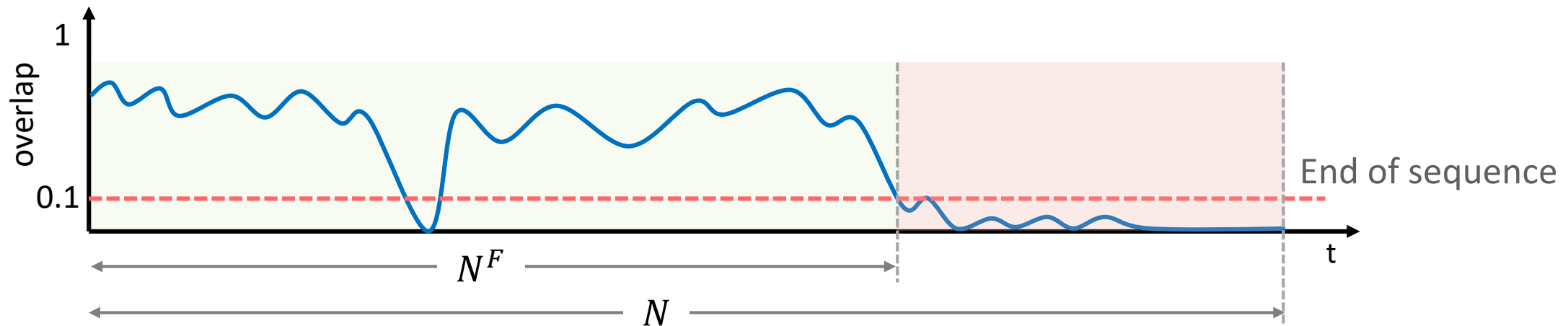
Prevent “gaming” where a tracker would predict the “entire image” as a bounding box to prevent reset identification

- Failure if the tracker does not recover within $\theta_N=10$ frames



VOT-ST2020 Performance measures

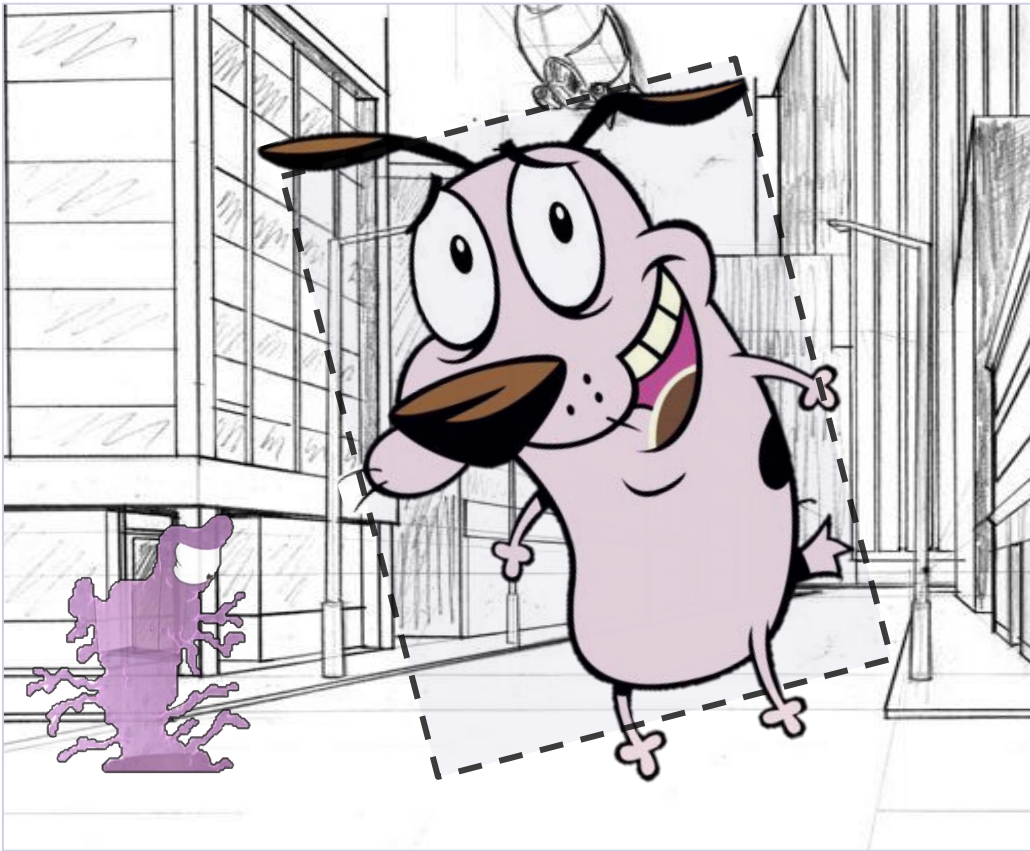
- Accuracy (A): average overlap on the successfully tracked period
- Robustness (R): Percentage of the tracked sub-sequence (N^F / N)



- Overall A/R: weighted average over all sequences
- EAO measure – combines the per-subsequence results

VOT-ST2020 target position reporting

VOT-ST2019 position = rot. bound. box



VOT-ST2020 position = seg. mask

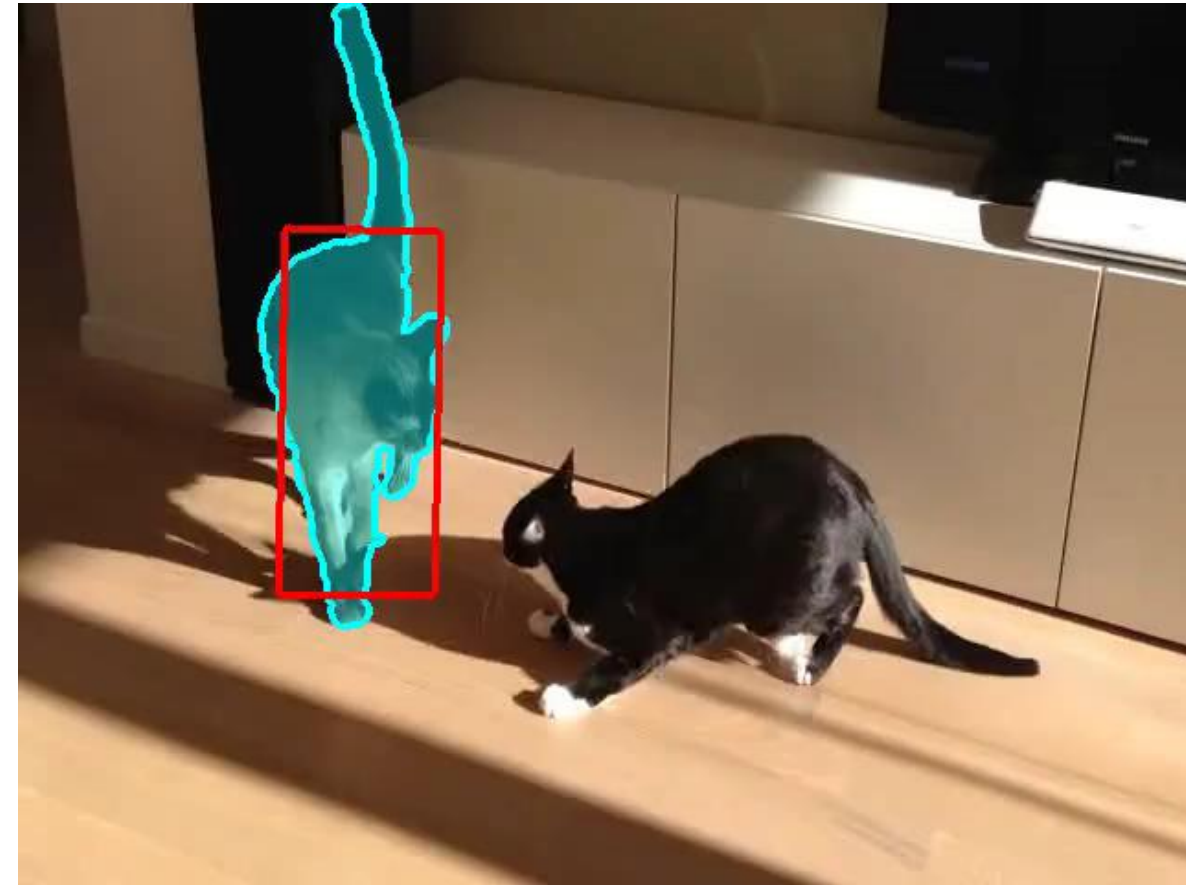


- Trackers may report bounding box or a mask
- Evaluated against a ground truth segmentation mask

The VOT-ST2020 dataset

- Public dataset (60 sequences) + Sequestered dataset (60 sequences)
Winner identified on *sequestered dataset*
- Both datasets refreshed
 - A challenging sequence added to each
- All frames manually segmented!
- Bounding boxes not provided (obsolete)

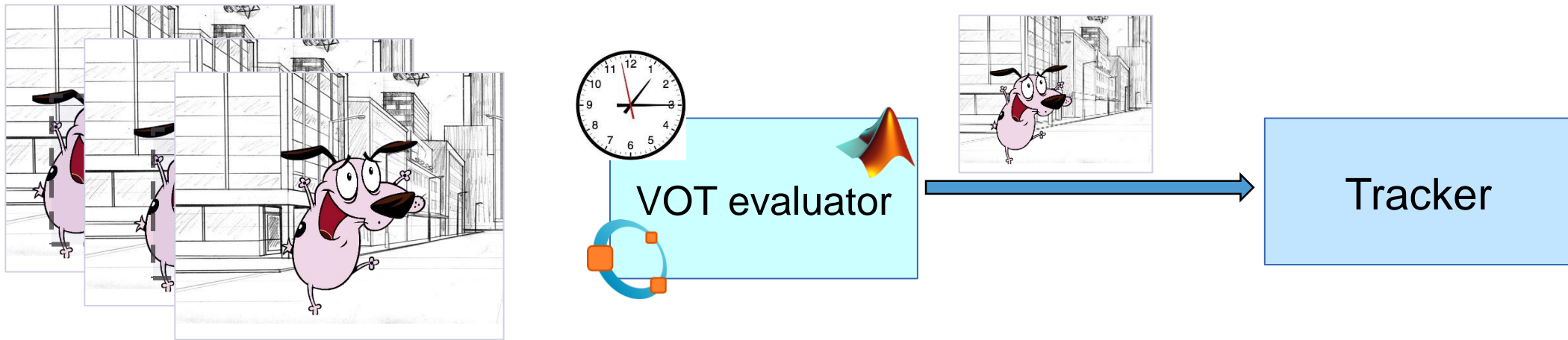
- Each frame annotated by 6 attributes:
Occlusion, Illumination change, Object motion, Object size change, Camera motion, Unassigned



Red – VOT2019 annotation by a bounding box
Blue – VOT2020 annotation by a segmentation mask

The VOT2020 ST real-time challenge (VOT-RT2020)

- Introduced in VOT2017
- Same performance evaluation protocol and measures as VOT-ST2020
- Required to process sequences at ~ 20 fps

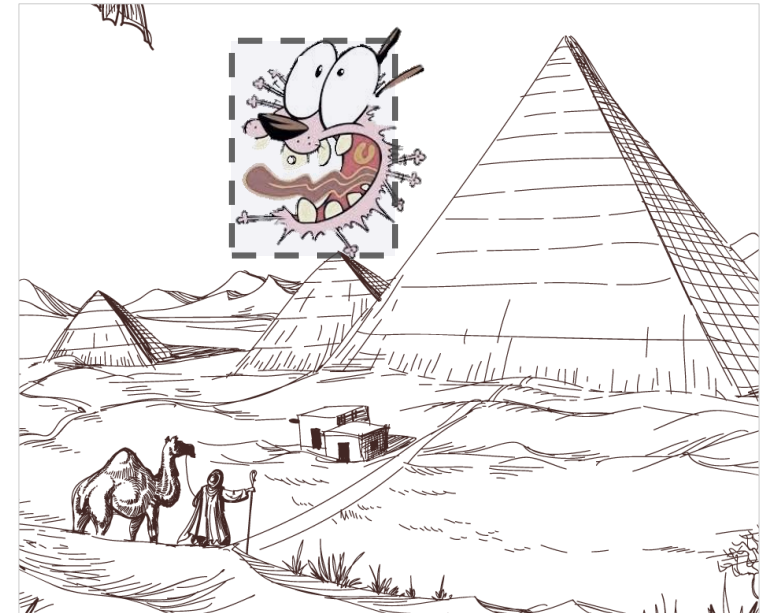


- Winner identification protocol:
Top EAO on the VOT-ST2020 public dataset

The VOT 2020 workshop

VOT2020 LT CHALLENGE: OVERVIEW

VOT2020 long-term challenge (VOT-LT2020)

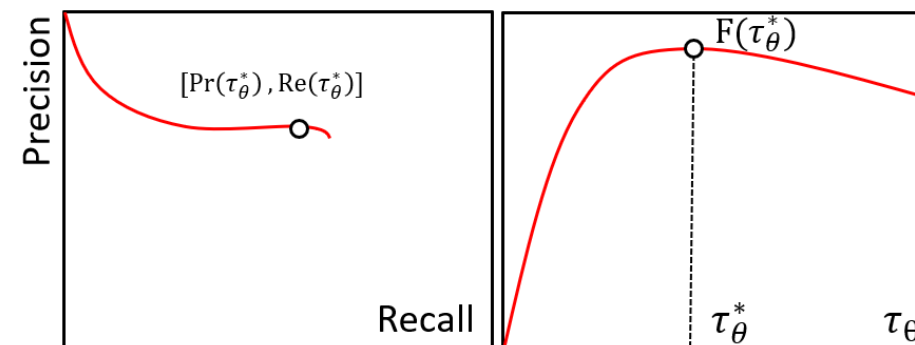
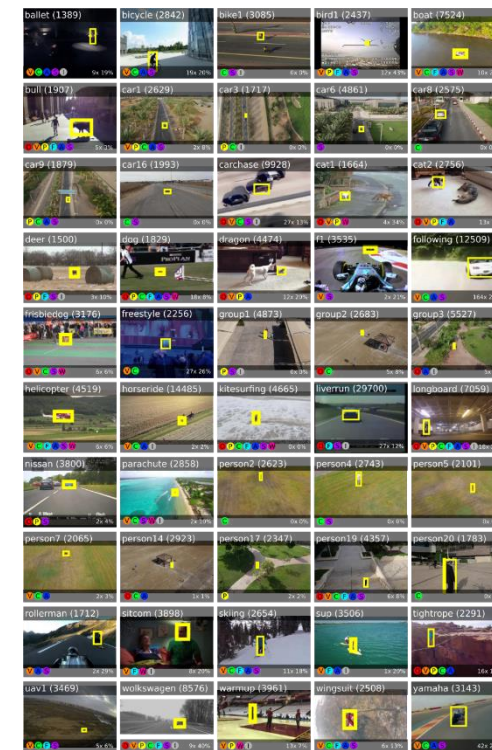


- Required long-term tracker properties:
 - Determine whether the target has been lost (or disappeared)
 - Re-detect the target when it reappears
- Tracker output at each frame: bounding box + certainty score

VOT2020 long-term challenge (VOT-LT2020)

- VOT-LT2019 dataset used
 - Average sequence length >4k frames
 - Average per sequence disappearance: 10
 - Average target absence period: 50 frames
- Initialized at first frame, no reset at target loss
- Tracking properties measured:
Localization, Loss/Presence **detection**
- Tracking Precision, Recall & F-score¹:

$$\Pr(\tau_\theta), Re(\tau_\theta), F(\tau_\theta)$$



¹Lukežič, et al., Now you see me: evaluating performance in long-term visual tracking, TCyb2020

The new VOT toolkit

- Rewritten in Python, runs the VOT2020 experiments

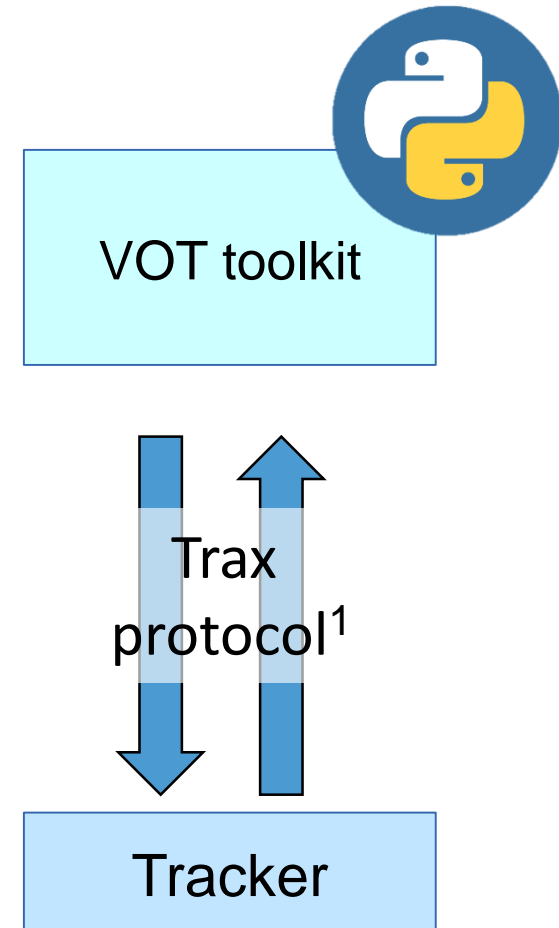
Previous Matlab toolkits¹ now obsolete.

- Download from the VOT homepage

https://www.votchallenge.net/howto/tutorial_python.html

- Plug and play!

- Supports major [programming languages](#) and operating systems



¹Luka Čehovin, TraX: The visual Tracking eXchange Protocol and Library, Neurocomputing, 2017

The VOT 2020 workshop

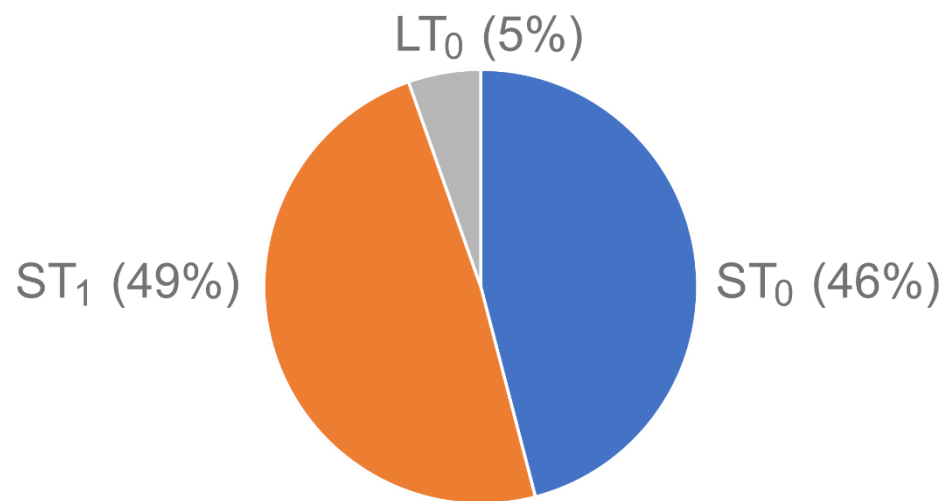
VOT-ST2020 & VOT-RT2020 CHALLENGE RESULTS

VOT-ST2020, VOT-RT2020: 37 trackers tested

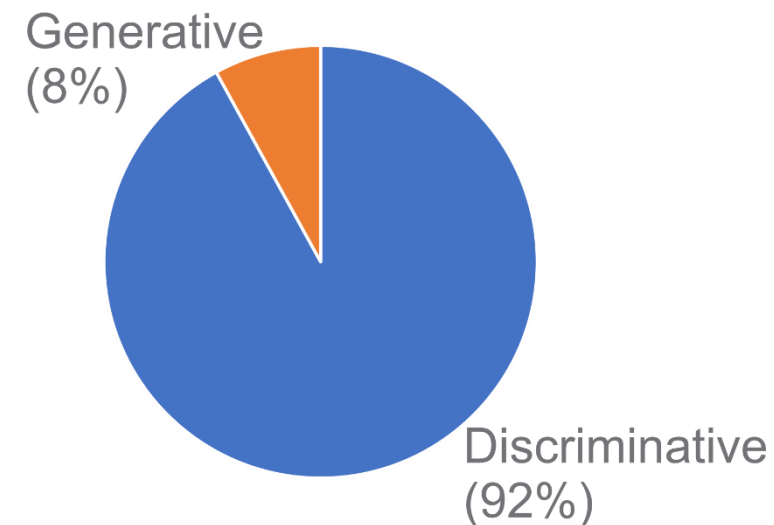
Tracking approach:

DCF: 68%
Siamese: 46%
Neither: 13%
Deep features: 86%

ST/LT category:



Target model:



Strong response in the VOT community:

57% of trackers reported a [segmentation mask](#)

VOT-ST2020 results on public dataset

- Top trackers: (1) RPT, (2) OceanPlus, (3) AlphaRef, (4) AFOD, (5) LWTL, (6) fastOcean, (7) TRASTmask, (8) DET50, (9) D3S, (10) Ocean

- All top trackers are deep trackers:

All use pretrained ResNet50

9 deep DCF (ATOM¹ DiMP²),

4 Siamese (e.g., SiamRPN³)

¹Danelljan et al. CVPR2019, ²Bhat et al. ICCV2019

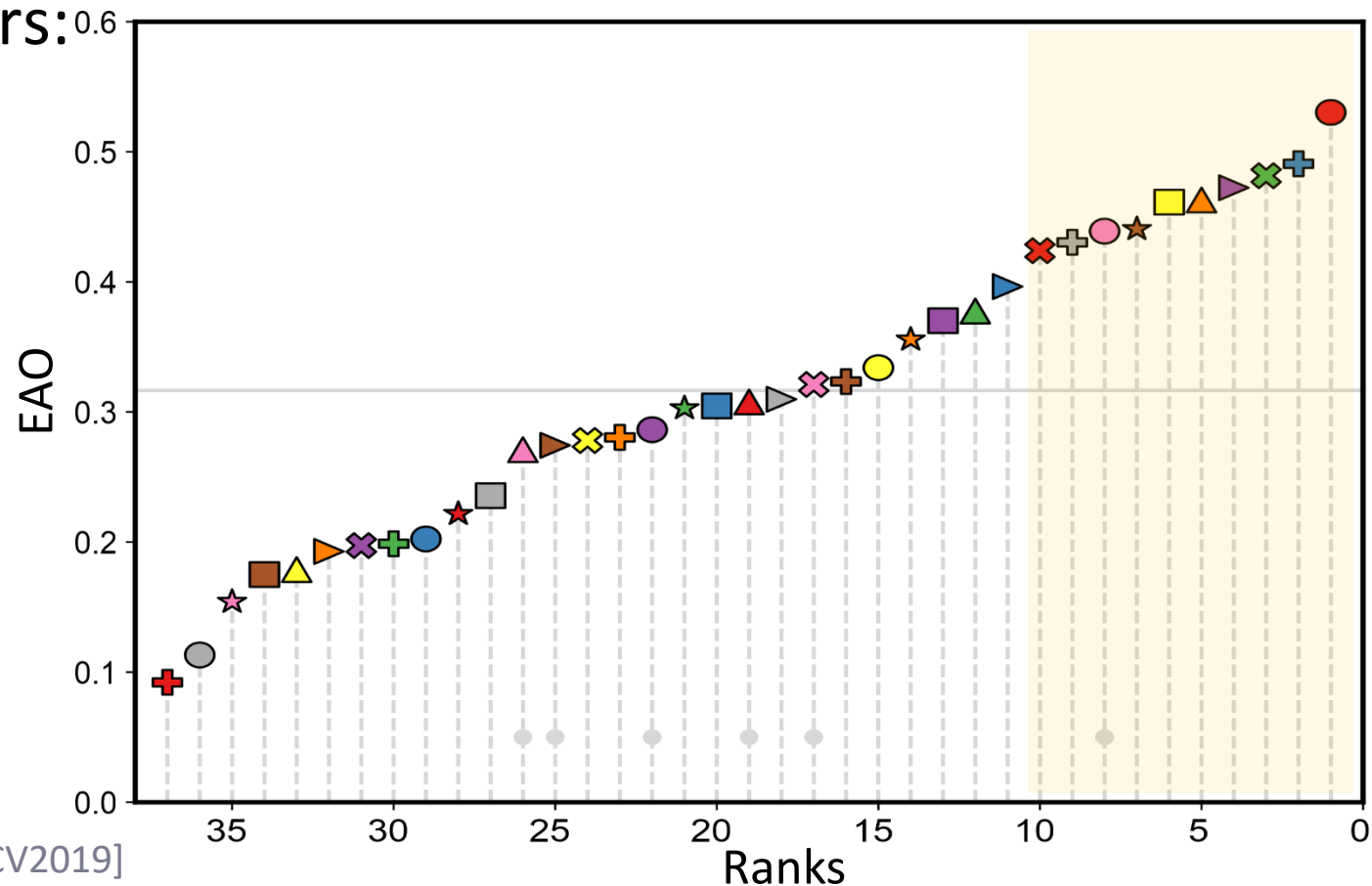
³Li et al. CVPR2019

- Localization:

All provide segmentation mask

Many apply region proposals^{4,5}

⁴Fcos [Tian et al. ICCV2019], ⁵RepPoints [Yang et al., ICCV2019]



VOT-ST2020 results on public dataset

- Top trackers are among the most robust trackers

(1) RPT, (2) OceanPlus, (3) fastOcean

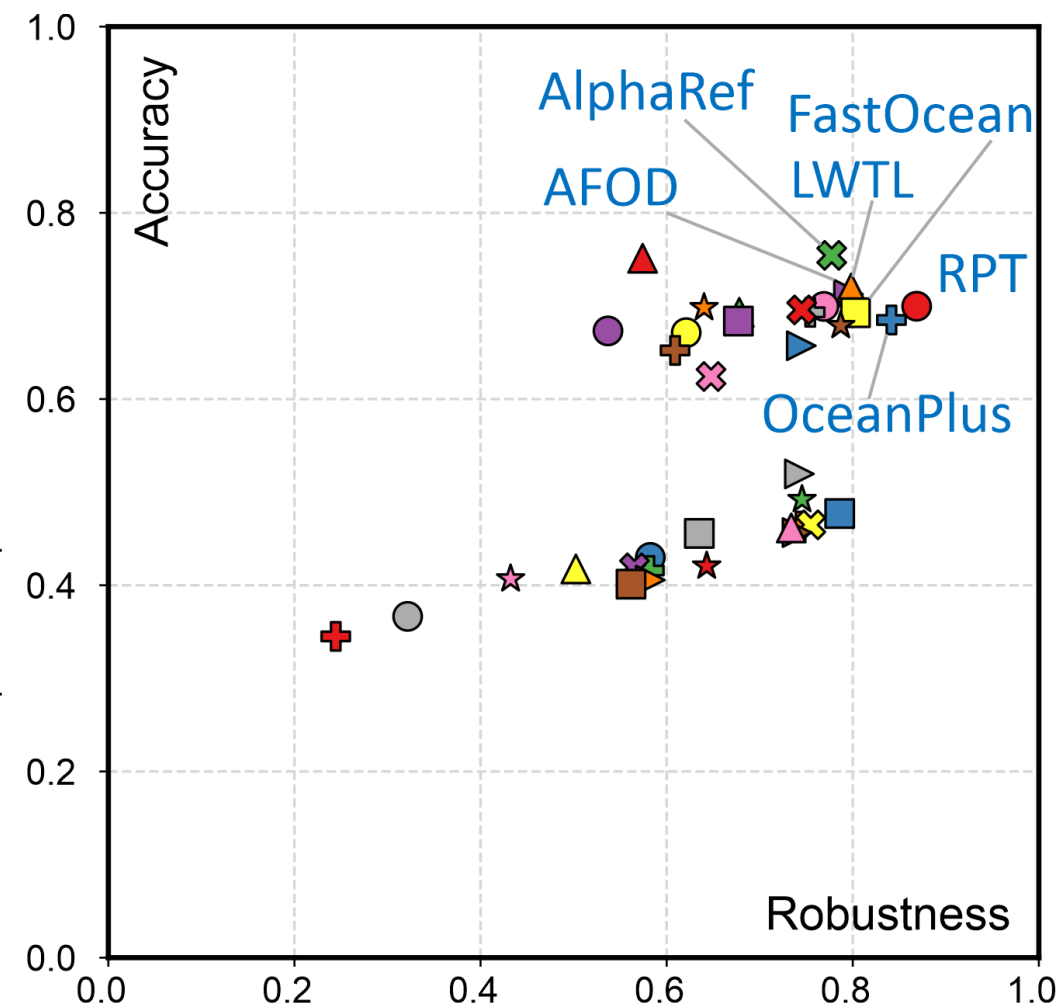
- Top in accuracy:

(1) AlphaRef, (2) STM, (3) LWTL

- Per-attribute analysis:

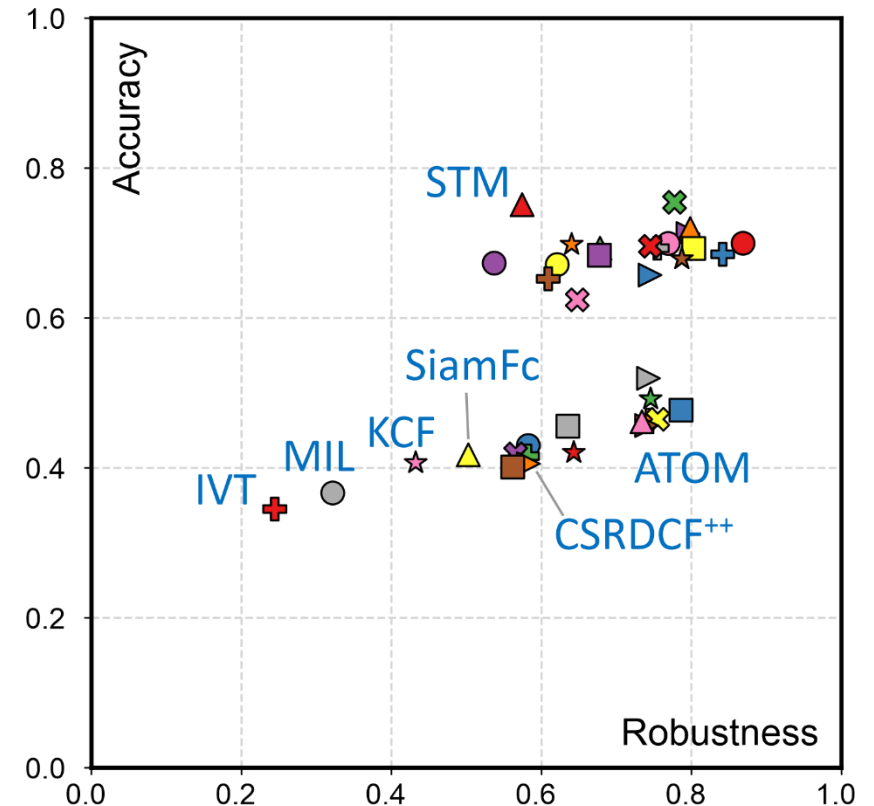
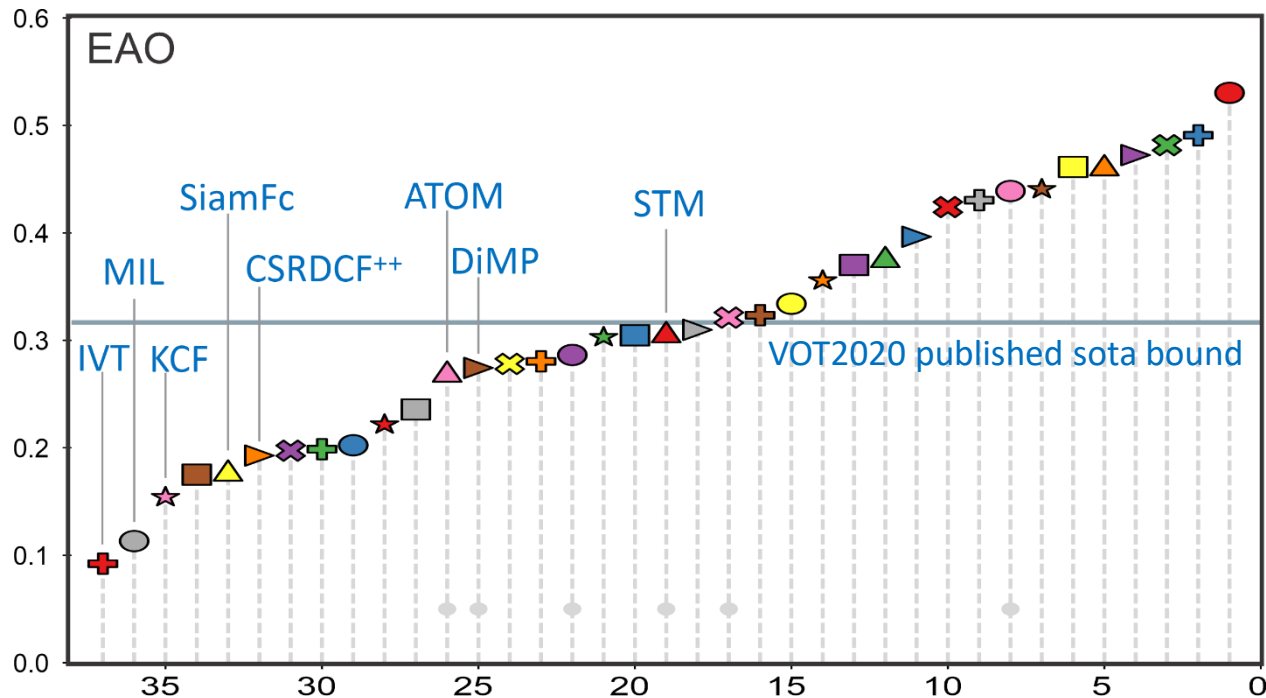
	CM	IC	OC	SC	MC
A	0.53 ③	0.54	0.45 ①	0.54	0.51 ②
R	0.70	0.77	0.60 ①	0.69	0.63 ②

- Most failures due to: **Occlusion**
- Mostly affects accuracy: **Occlusion**

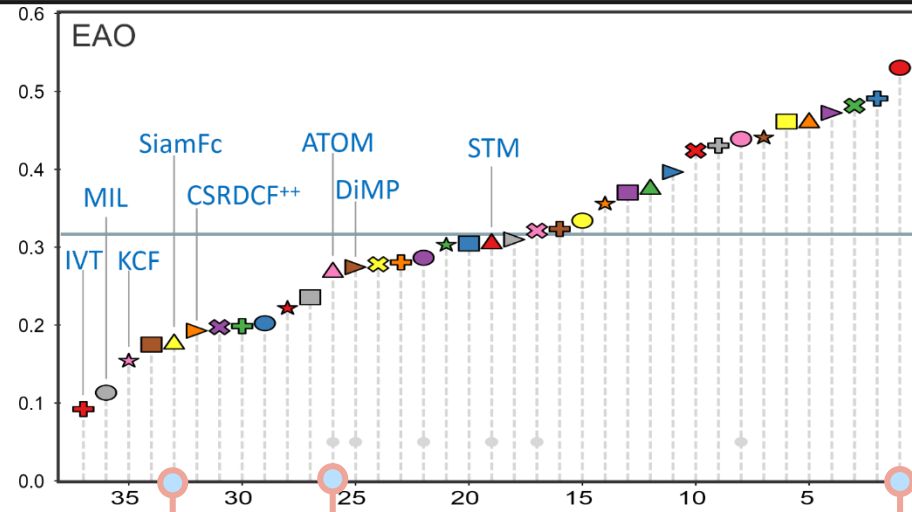


Tracking keeps advancing

- Baselines & 2019 SOTA ranked at the tail of the benchmark
 - 46% submissions above the VOT2020 sota bound
- Added by VOT2020: Video object segmentation “tracker” STM [Oh, et al. ICCV2019]
 - Outperforms 48% trackers in EAO & 2nd in accuracy



Progress of the last 4 years

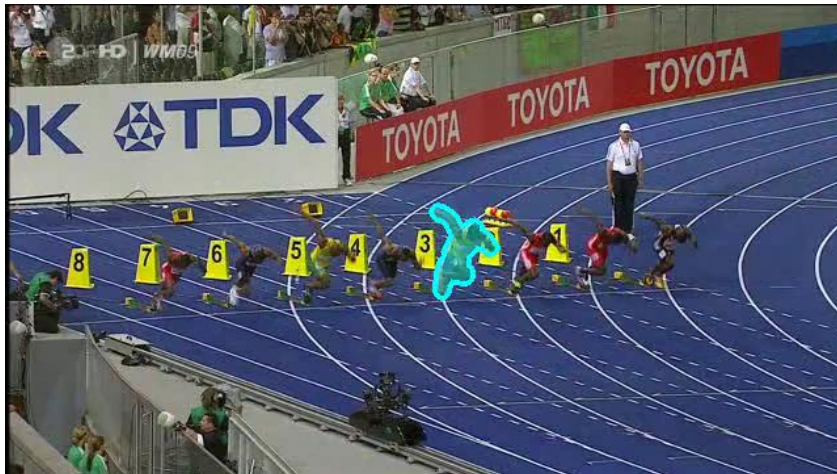


„Although tracking itself is by and large a solved problem...“
-- Jianbo Shi & Carlo Tomasi
CVPR1994 --

SiamFc [Bertinetto et al. VOTW2016]

ATOM [Danelljan et al. CVPR2019]

RPT [Ma et al. VOT2020]



Winner of VOT2017 RT

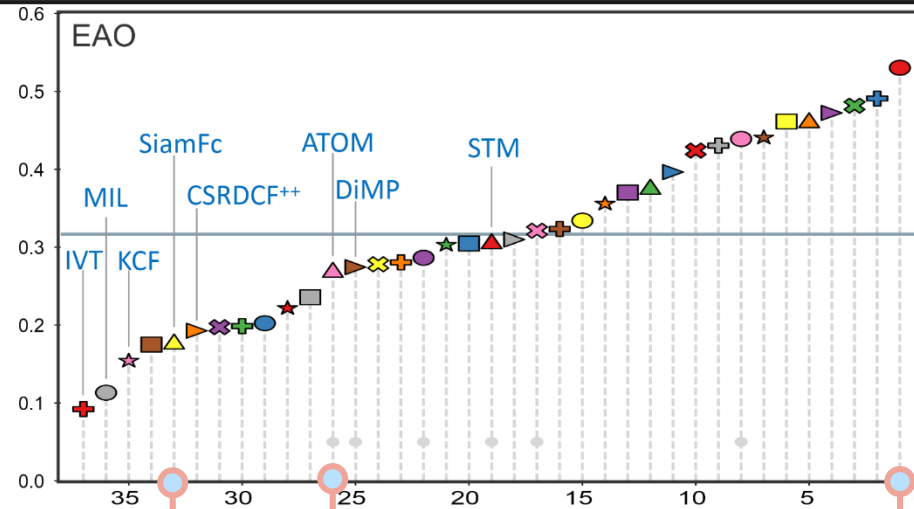


Widely used in the top-performers in 2019



Top performer on VOT2020 public set

Not there yet...



~~„Although tracking itself is by and large a solved problem...“
-- Jianbo Shi & Carlo Tomasi
CVPR1994 --~~

SiamFc [Bertinetto et al. VOTW2016]



Winner of VOT2017 RT

ATOM [Danelljan et al. CVPR2019]



Part of most top-performers in 2019

RPT [Ma et al. CVPR2019]

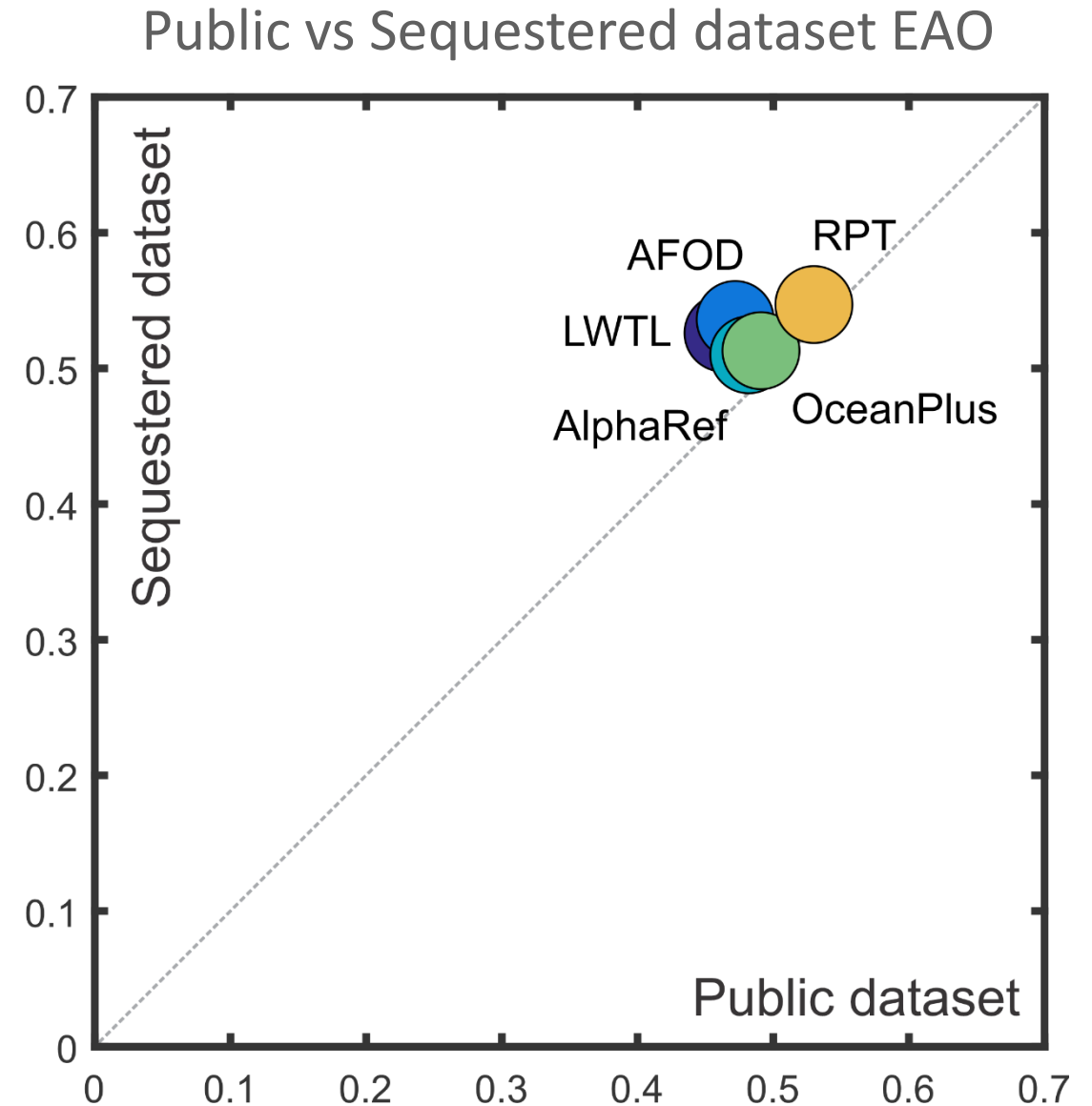
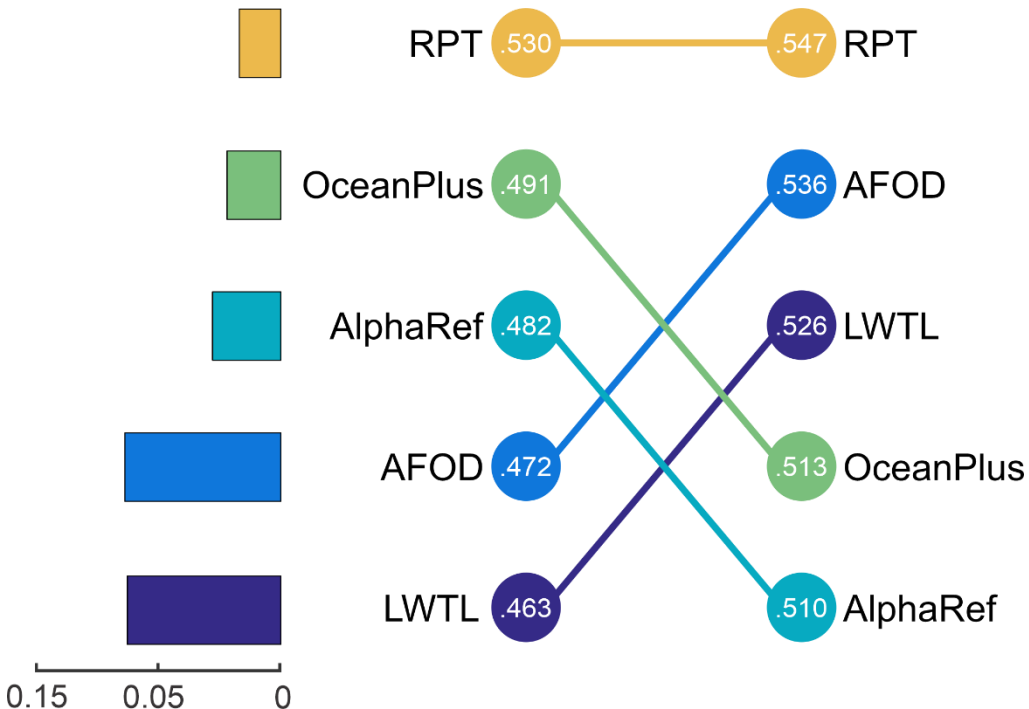


Top performer on VOT2020 public set

VOT-ST2020 results on sequestered dataset

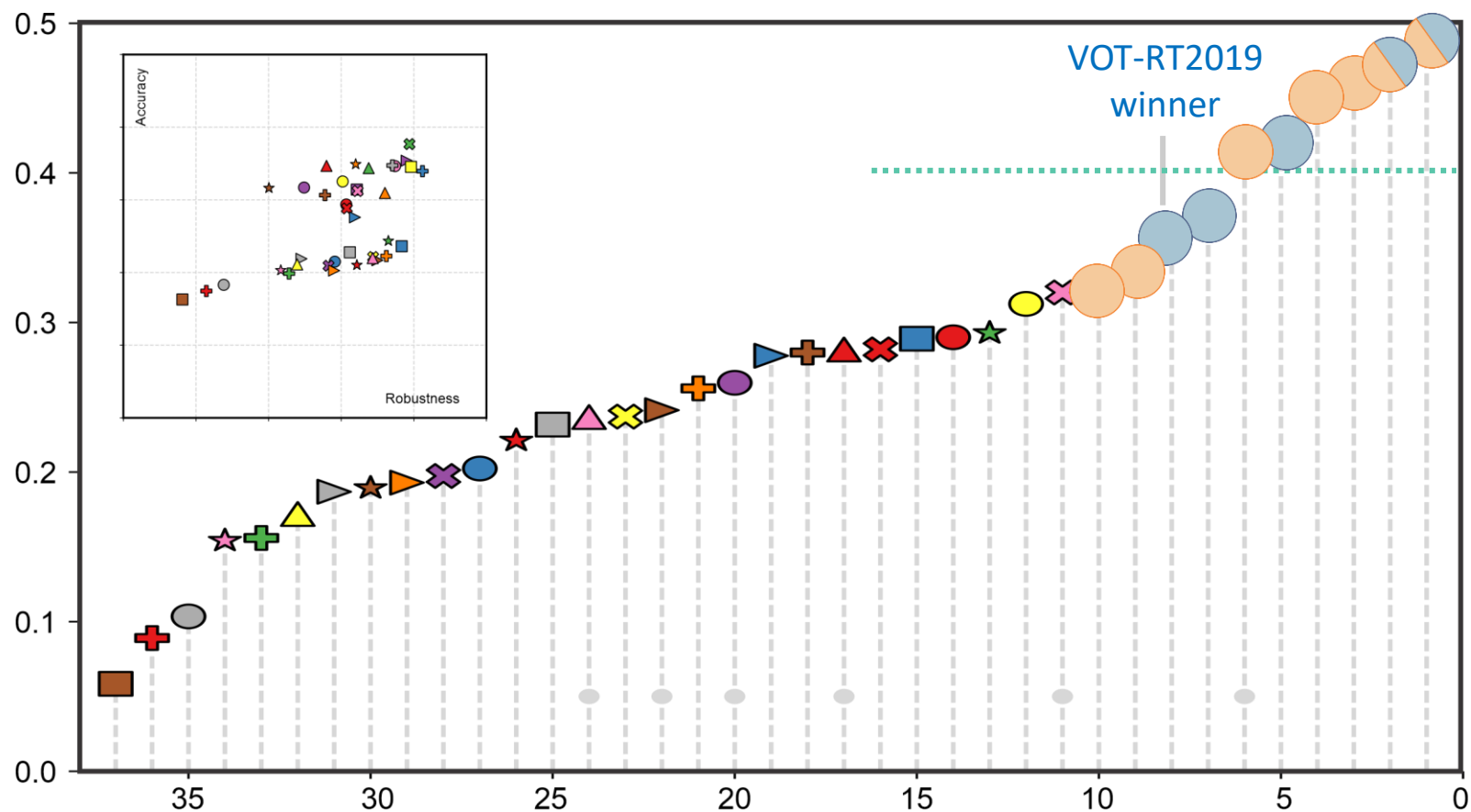
- Comparable results between public and sequestered set

- Biggest rank *ascent*: AFOD, LWTL
- Clearly stands out: RPT



VOT realtime challenge (VOT-RT2020) results

- **Top 10:** (1) AlphaRef, (2) OceanPlus, (3) AFOD, (4) fastOcean, (5) Ocean, (6) D3S, (7) AFAT, (8) SiamMargin, (9) LWTL, (10) TRASTmask



Two classes:



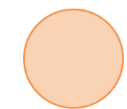
Approach:

Siamese correlation

Bounding box regression

(e.g., SiamRPN¹, SiamMask²)

GPU-based



Approach:

deep DCF correlation

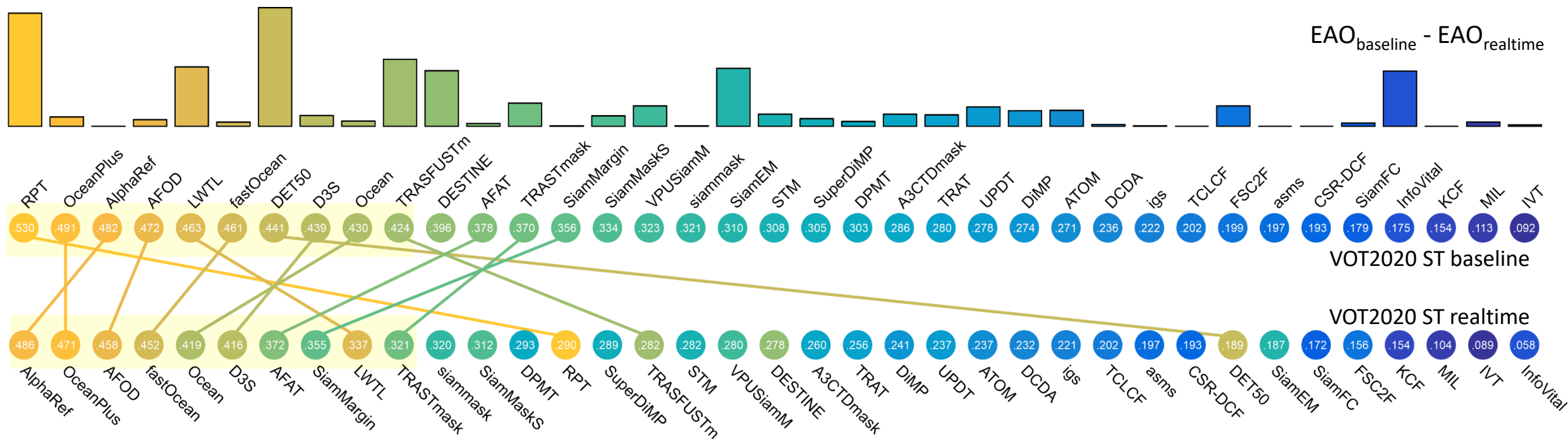
Bounding box regression

(e.g., ATOM³)

GPU-based

¹Li et al. CVPR2018, ²Wang et al. CVPR2019, ³Danelljan et al. CVPR2019

VOT2020 Realtime vs Baseline results



- 8 out of 10 top RT trackers among 10 top on VOT-ST2020 challenge!
 - AlphaRef, OceanPlus and AFOD ranked 3rd, 2nd and 4th on the VOT-ST2020
- Emergence of deep tracking architectures that no longer sacrifice speed for performance

The VOT 2020 workshop

VOT-LT2020 CHALLENGE RESULTS

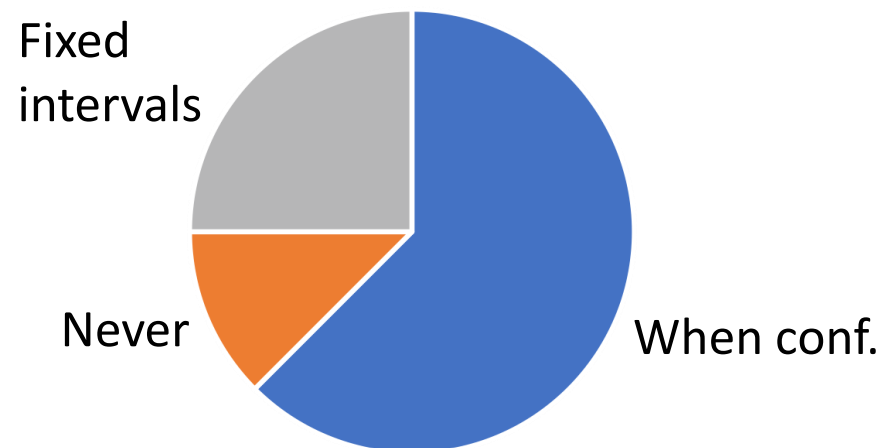
VOT-LT2020 challenge overview

- 8 trackers tested
- All trackers were from LT_1 class:
Explicit target absence detection and re-detection implemented

Features:

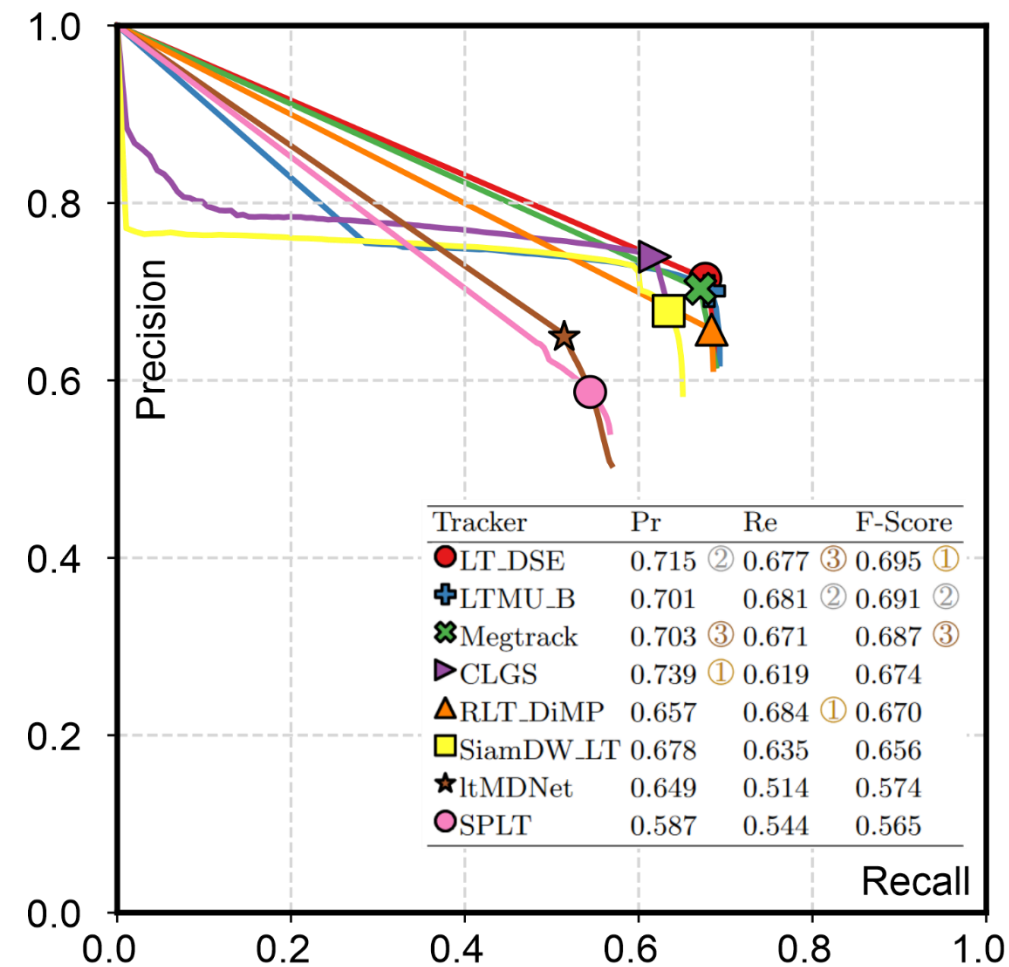


Model update:



VOT-LT2020 challenge results

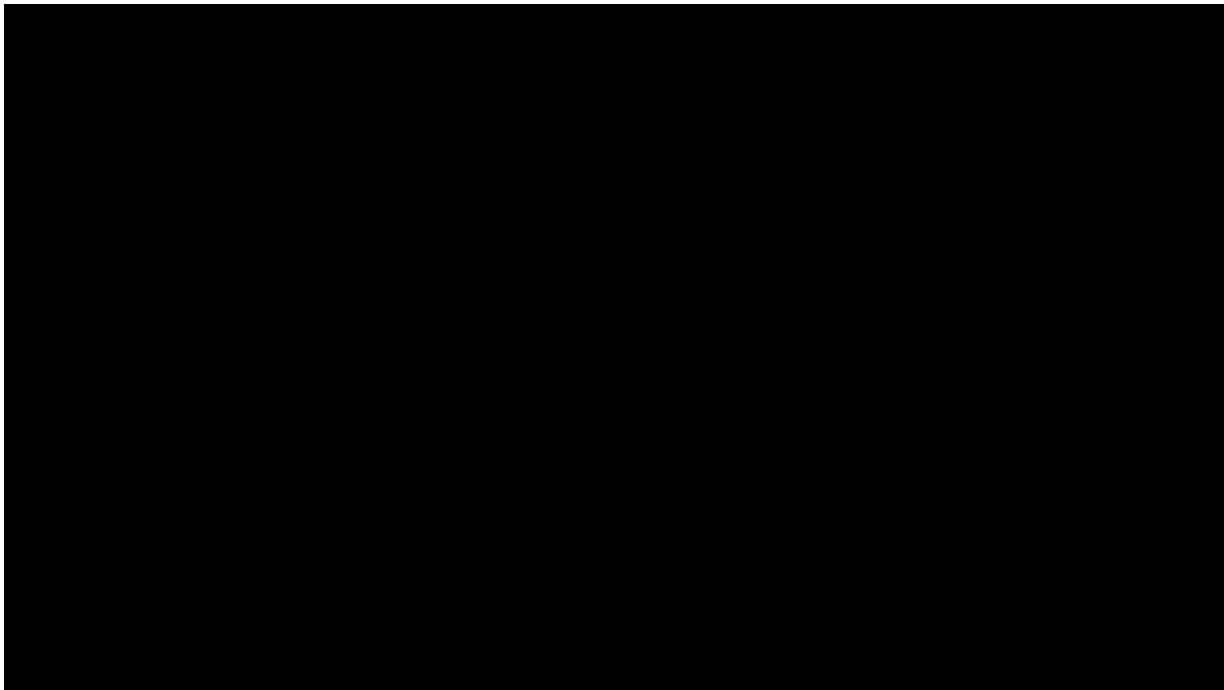
- Top 3 trackers: LTDSE, LTMUB, Megtrack
Apply a Short-term tracker + Detector
- Top-performer: LTDSE (winner of VOT-LT2019)
Contributed by the VOT committee
ST: ATOM¹ + SiamMask³
Absence det.: MDNet² (winner of VOT-ST2015)
Re-Det: RPN from MBMD⁴ (winner of VOT-LT2018)
- Second-best (~0.6% worse): LTMUB
A **simplified version of LTDSE** with an LSTM meta-updater



¹Danelljan et al. CVPR2019, ²Nam et al CVPR2016, ³Wang et al. CVPR2019, ⁴Zhang et al VOT2018

Advances since inception of VOT-LT (2018)

FCLT [Lukežič et al. ACCV2018]



A state-of-the-art LT tracker in 2017

LTMUB [Dai et al. VOT2020]



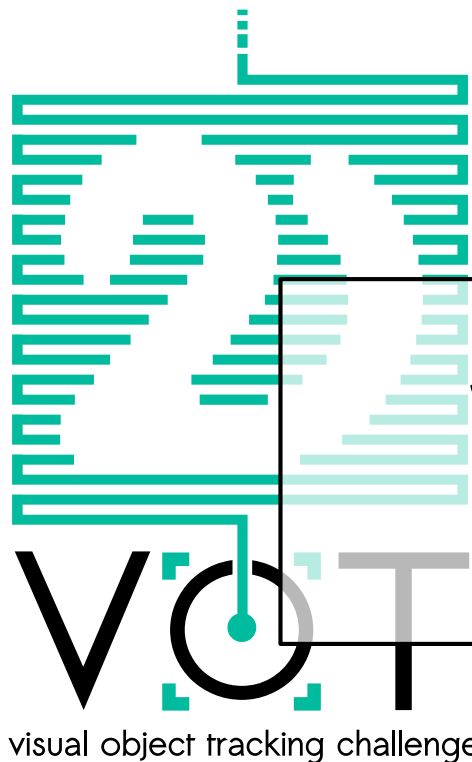
Top LT tracker in VOT2020

VOT2020 ST/RT/LT challenges summary

- VOT-ST2020:
 - Deep DCF and Siamese correlation the dominant methodology
 - All top performers provide (high-quality) target segmentation
 - Substantial performance improvements over last 2 years
- VOT-RT2020:
 - Emergence of deep tracking architectures that no longer sacrifice speed for performance
- VOT-LT2020:
 - Deep LT architectures stabilized (deep ST component, deep detector)
 - More work required on target appearance & detector learning

The VOT 2020 workshop

VOT2020 ST/RT/LT WINNER ANNOUNCEMENTS



Winners talks in the Live Session II
(18:00-20:00, UTC+1 time zone)



VOT-ST2020 Winners:

RPT by: [Haitao Zhang](#), [Linyuan Wang](#), [Ziang Ma](#), [Wei Lu](#), [Jun Yin](#) and [Miao Cheng](#)

“Learning Point Set Representation for Siamese Visual Tracking”

VOT-RT2020 Winners:

AlphaRef by: [Bin Yan](#), [Dong Wang](#), [Huchuan Lu](#) and [Xiaoyun Yang](#)

“Alpha-Refine”

VOT-LT2020 Winners:

LTMUB by: [Kenan Dai](#), [Dong Wang](#), [Jianhua Li](#), [Huchuan Lu](#) and [Xiaoyun Yang](#)

“A Baseline Long-Term Tracker with Meta-Updater”

Thanks

- The VOT2020 committee



M. Kristan



J. Matas



A. Leonardis



M. Felsberg



R. Pflugfelder



J. K. Kamarainen



G. Fernandez



L. Čehovin



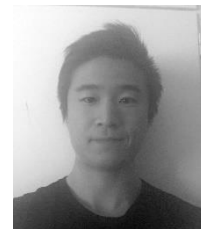
A. Lukežič



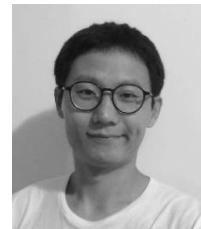
M. Danelljan



O. Drbohlav



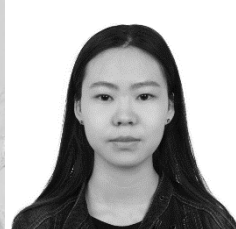
H. Linbo



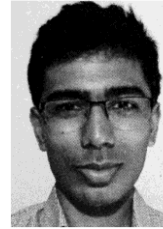
Y. Song



Y. Jinyu



Z. Yushan



G. Bhat

- Everyone who participated or contributed

Alexander Hauptmann11, Alireza Memarmoghadam14, Alvaro Garcia10, Andreas Robinson4, Anton Varfolomeiev28, Awet Hailelassie Gebrehiswot10, Bedirhan Uzun15, Bin Yan13, Bing Li21, Chen Qian32, Chi-Yi Tsai38, Christian Micheloni44, Dong Wang13, Fei Wang32, Fei Xie36, Felix Jaremo Lawin4, Fredrik Gustafsson45, Gian Luca Foresti44, Goutam Bhat8, Guangqi Chen32, Haibin Ling37, Haitao Zhang47, Hakan Cevikalp15, Haojie Zhao13, Haoran Bai34, Hari Chandana Kuchibhotla20, Hasan Saribas16, Heng Fan37, Hossein Ghanei-Yakhdan46, Houqiang Li42, Houwen Peng26, Huchuan Lu13, Hui Li22, Javad Khaghani39, Jesus Bescos10, Jianhua Li13, Jianlong Fu26, Jiaqian Yu31, Jingtao Xu31, Josef Kittler43, Jun Yin47, Junhyun Lee24, Kaicheng Yu19, Kaiwen Liu21, Kang Yang27, Kenan Dai13, Li Cheng39, Li Zhang41, Lijun Wang13, Linyuan Wang47, Luc Van Gool8, Luca Bertinetto17, Martin Danelljan8, Matteo Dunnhofer44, Miao Cheng47, Mohana Murali Dasari20, Ning Wang27, Ning Wang42, Pengyu Zhang13, Philip H.S. Torr41, Qiang Wang29, Radu Timofte8, Rama Krishna Sai Gorthi20, Seokeon Choi23, Seyed Mojtaba Marvasti-Zadeh39, Shaochuan Zhao22, Shohreh Kasaei33, Shoumeng Qiu35, Shuhao Chen13, Thomas Schon45, Tianyang Xu43, Wei Lu47, Weiming Hu21;29, Wengang Zhou42, Xi Qiu25, Xiao Ke18, Xiao-Jun Wu22, Xiaolin Zhang35, Xiaoyun Yang12;30, Xuefeng Zhu22, Yingjie Jiang22, Yingming Wang13, Yiwei Chen31, Yu Ye18, Yuezhou Li18, Yuncon Yao36, Yunsung Lee24, Yuzhang Gu35, Zezhou Wang13, Zhangyong Tang22, Zhen-Hua Feng43, Zhijun Mai40, Zhipeng Zhang21, Zhirong Wu26, and Ziang Ma47

- VOT2020 sponsor:



University of Ljubljana
Faculty of Computer and
Information Science

Photos & sound in c-break videos:
Pixabay, Matt Hardy, Nonima