VOT
visual object tracking

The Visual Object Tracking Challenge Results VOT2018

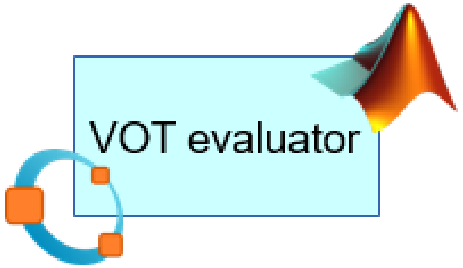
Matej Kristan, Aleš Leonardis, Jiri Matas, Michael Felsberg, Roman Pflugfelder, Luka Čehovin, Gustavo Fernandez, Alan Lukežič, Tomaš Vojir, Goutam Bhat, Abdelrahman Eldesokey, et al.

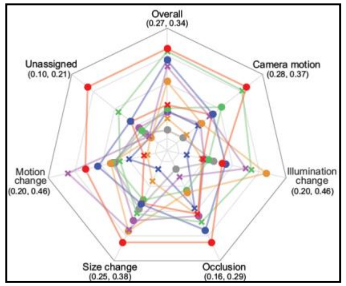
University of Ljubljana Faculty of Computer and Information Science
UNIVERSITY OF BIRMINGHAM
LINKÖPING UNIVERSITY
AUSTRIAN INSTITUTE OF TECHNOLOGY

The emergence of VOT initiative

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

- The VOT initiative (February 2013)
- Goal: Establish evaluation standards -> development of trackers
- Four pillars of VOT:
  - Evaluation system
  - Datasets
  - Evaluation methodology
  - Community building (VOT challenges)





The VOT challenge evolution

	Perf. Measures	Dataset size	Target box	
VOT2013	ranks, A, R	16, manual select.	manual	VOT-ST
VOT2014	ranks, A, R, EFO	25, manual select.	manual	VOT-ST
VOT2015	EAO, A, R, EFO	60, fully auto	manual	VOT-ST, VOT-TIR
VOT2016	EAO, A, R, EFO	60, fully auto	auto	VOT-ST, VOT-TIR,
VOT2017	EAO, A, R, $EAO_{realtime}$	60x2, fully auto	auto	VOT-ST, VOT-RT, VOT-TIR

- Carefully developed datasets, annotation, measures, toolkits, subchallenges
- VOT2018:
  - Short-term tracking challenge
  - Short-term tracking real-time challenge
  - Long-term tracking challenge

Outline

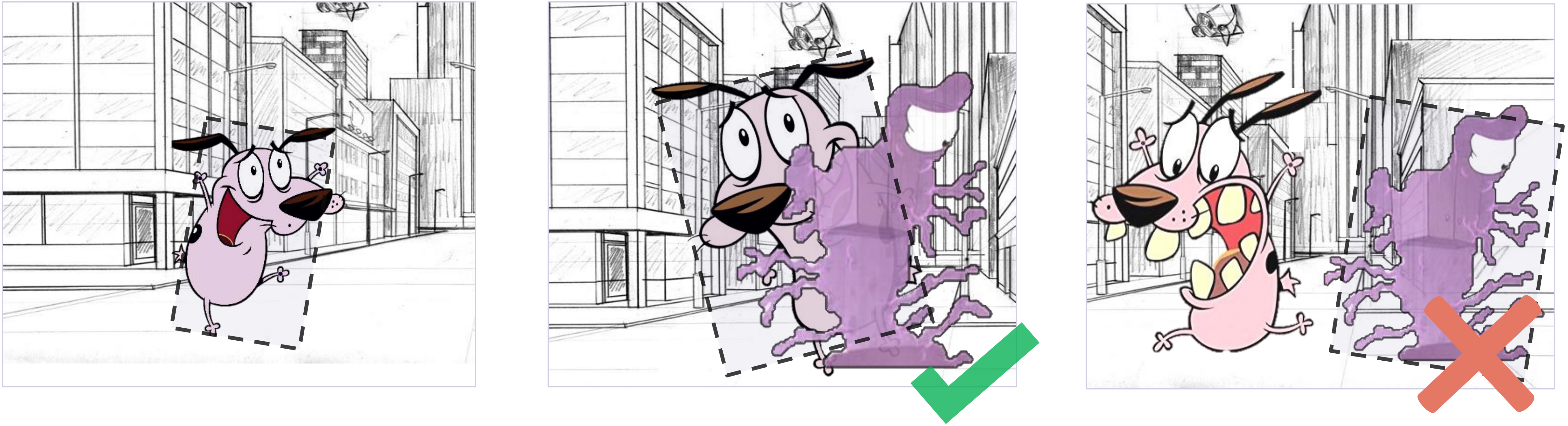
1. Scope of the VOT2018 (sub) challenges

2. VOT2018 results overview

3. Winner announcement

VOT2018 short-term (ST) challenge

- Short-term, single-target, causal trackers
- Tracker reports the target state as a rotated bounding box



- No redetection: drift is considered a failure and tracker is reset

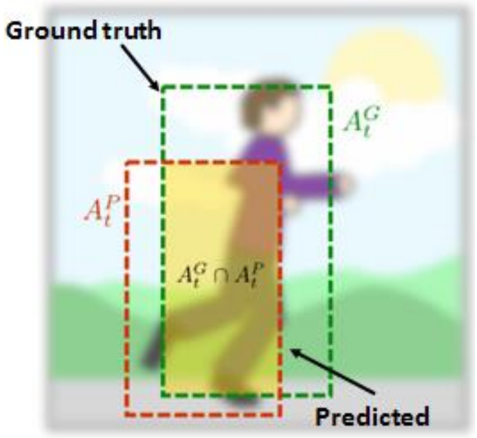
The VOT2018 ST dataset

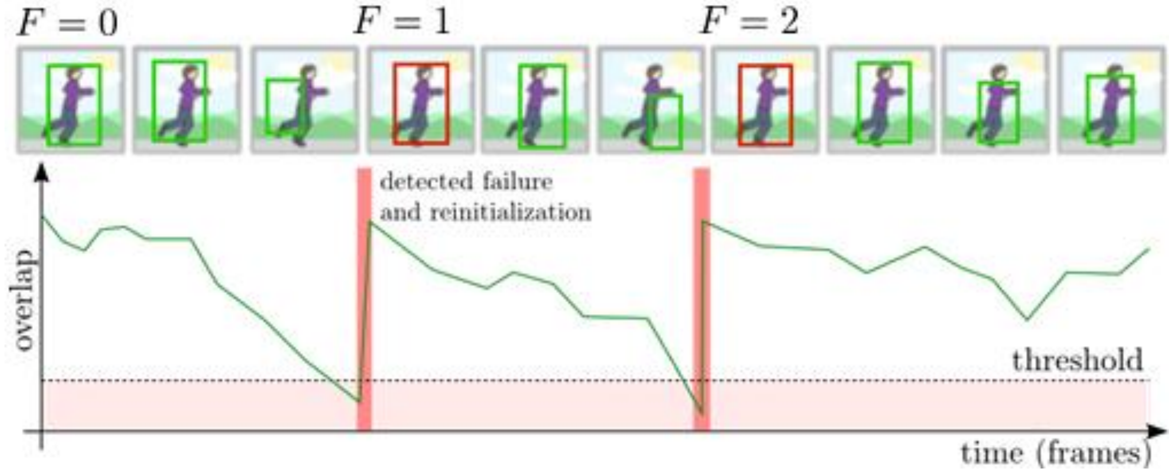
- VOT2017 dataset did not saturate → same dataset used in VOT2018
- Public dataset (60 sequences) + Sequestered dataset (60 sequences)
- Each image annotated by 6 attributes:

  Occlusion, Illumination change , Object motion, Object size change, Camera motion, Unassigned
- Rotated bounding box automatically computed from pre-segmented image

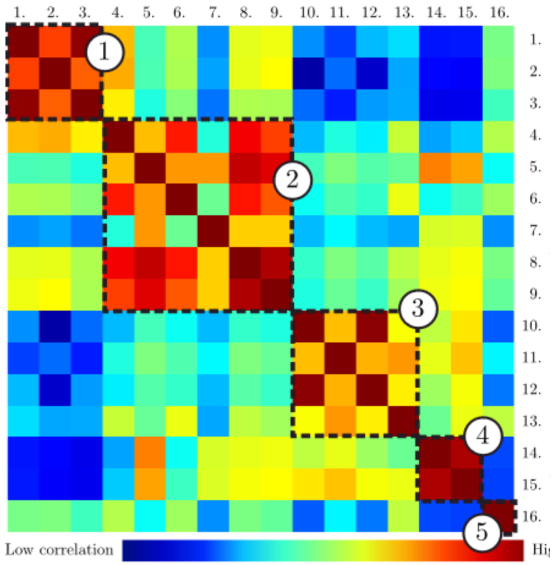
The VOT2018 ST evaluation methodology

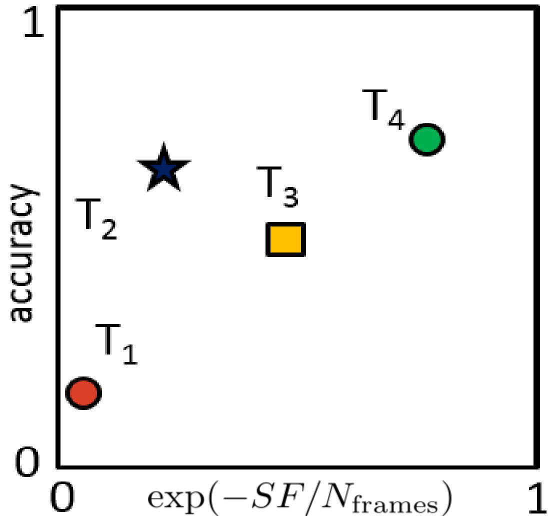
- Two weakly correlated measures[2] chosen according to[1]:
  - Robustness (number of times a is reinitialized)
  - Accuracy (average overlap while tracking)
  - + Combination of basic measures (EAO)





Performance measure correlation analysis[1]



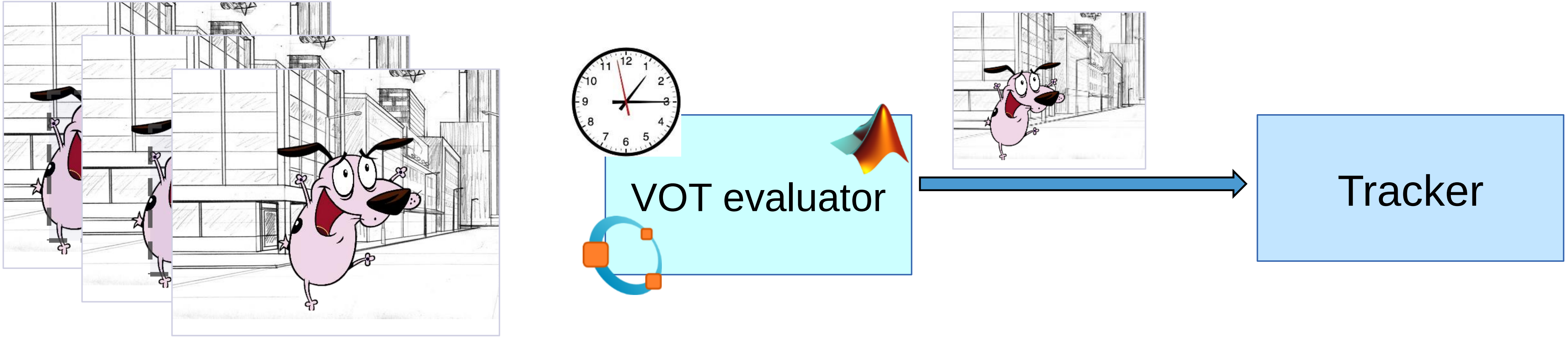


[1]Čehovin, Leonardis, Kristan. *Visual object tracking performance measures revisited*, IEEETIP 2016

[2]Kristan et al., A Novel Performance Evaluation Methodology for Single-Target Trackers, IEEETPAMI 2016

The VOT2018 ST real-time challenge

- Required to process sequences at ~20 fps



- The VOT2018 ST public dataset used for this

- Same performance evaluation protocol and measures as VOT2018 ST

The VOT 2018 workshop

VOT2018 LONG-TERM TRACKING CHALLENGE

Long-term tracking (LTT)

- Required long-term tracker properties:
  - Determine when the target has been lost (or disappeared)
  - Re-detect the target after losing the target

Short-term vs long-term spectrum[1]

ST/LT levels	Position reported	Determines target lost?	Target re-detection
ST_0: Basic ST	each frame	no	no
ST_1: Basic ST with conservative updating	each frame	not explicitly, selective update of visual model	no
LT_0: Pseudo LT	only when visible	yes	no
LT_1: Re-detecting LT	only when visible	yes	yes

- ST_0 (e.g., KCF[2], MS[3])
- ST_1 (e.g., MDNet[4], ECO[5]) -> easily converted to LT_0
- LT_1 (e.g., TLD[5])

[1]Lukežič, Čehovin, Vojir, Matas, Kristan, *Now you see me: evaluating performance in long-term visual tracking*, arXiv2018
[2]Enriques et al. PAMI 2015 ; [2]Comaniciu et al. PAMI 2002; [3]Nam et al. *CVPR2016*;
[4]Danelljan et al. *CVPR2017*; [5]Kalal et al. PAMI 2011

VOT2018 LT tracking dataset

- VOT approach: Keep it sufficiently small, well annotated and diverse
- Most challenging element: lots of target disappearances
- The dataset that meets these requirements: LTB35[1]

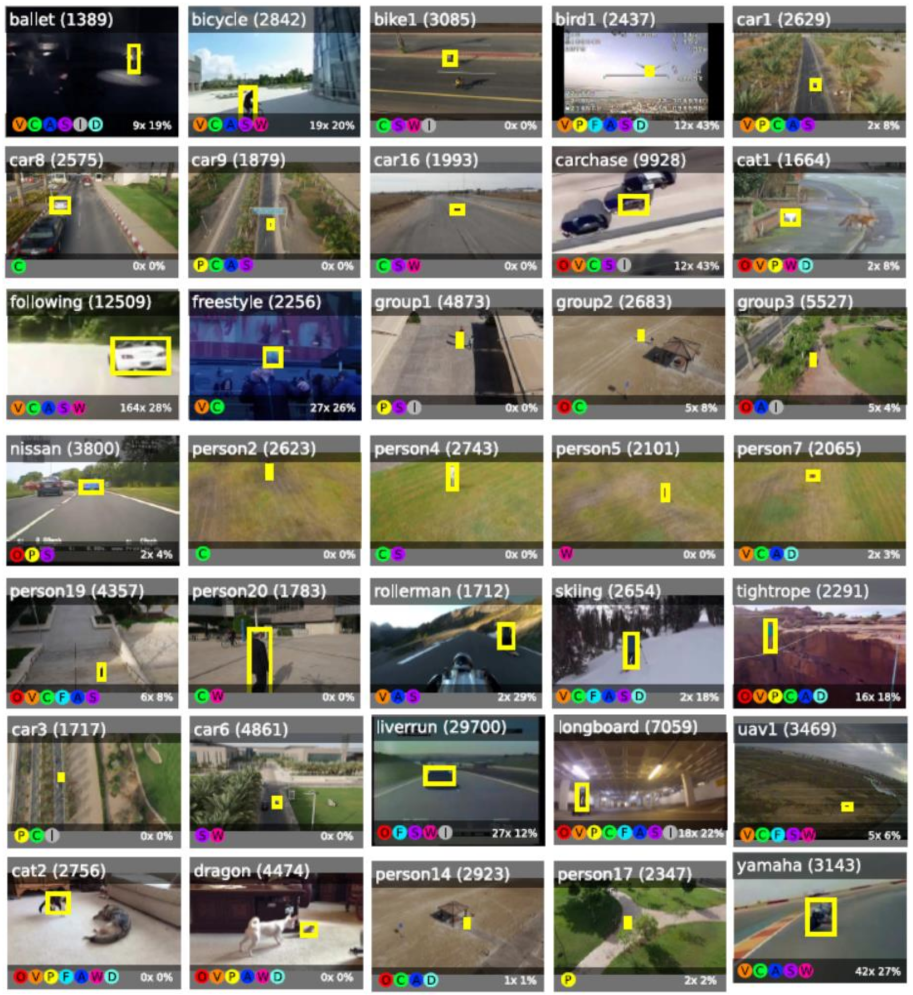
LTB35 =
20 (from UAV20L[2]– small objects & many disappearances)
+3 (from[3] – challenging long sequences)
+6(new from Youtube – many disappearances)
+6 (new generated from[4]

[1]Lukežič, et al., Now you see me: evaluating performance in long-term visual tracking, Arxiv2018
[2]Mueller et al., A benchmark and simulator for uav tracking, ECCV2016
[3]Kalal et al., Tracking-Learning-Detection, TPAMI2010
[4]Čehovin et al., Beyond Standard Benchmarks: Parameterizing Performance Evaluation in Visual Object Tracking, ICCV2017

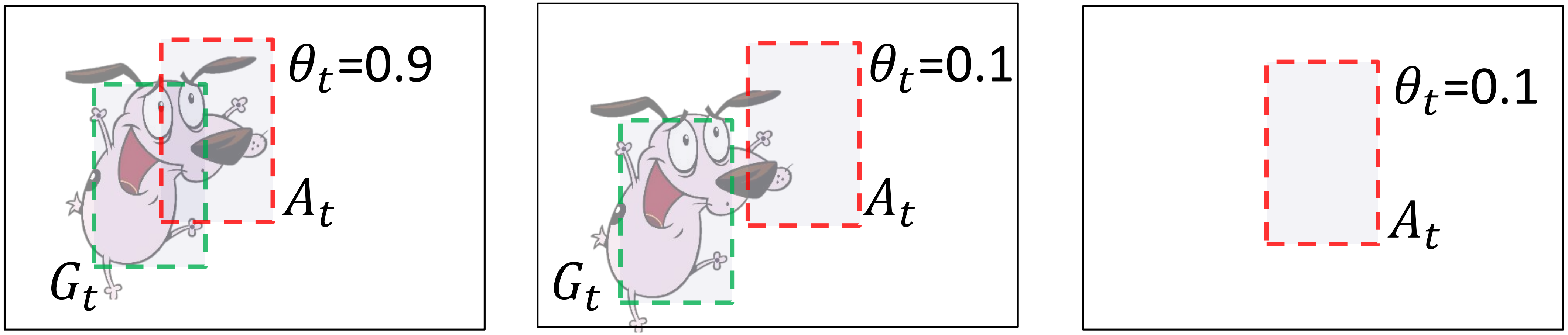
VOT2018 LT tracking dataset

- 35 sequences (146,847 frames)
- Axis aligned bounding box annotations

  (persons, car, motorcycle, bicycle, boat, animals, etc.)
- Resolution: 290x217 - 1280x720
- Average per sequence disappearance: 12
- Average target absence period: 40 frames
- Nine per-sequence attributes:
  (1) full occlusion, (2) out-of-view motion, (3) partial occlusion, (4) camera motion, (5) fast motion, (6) scale change, (7) aspect ratio change, (8) viewpoint change, (9) similar objects



LT performance measure design[1]

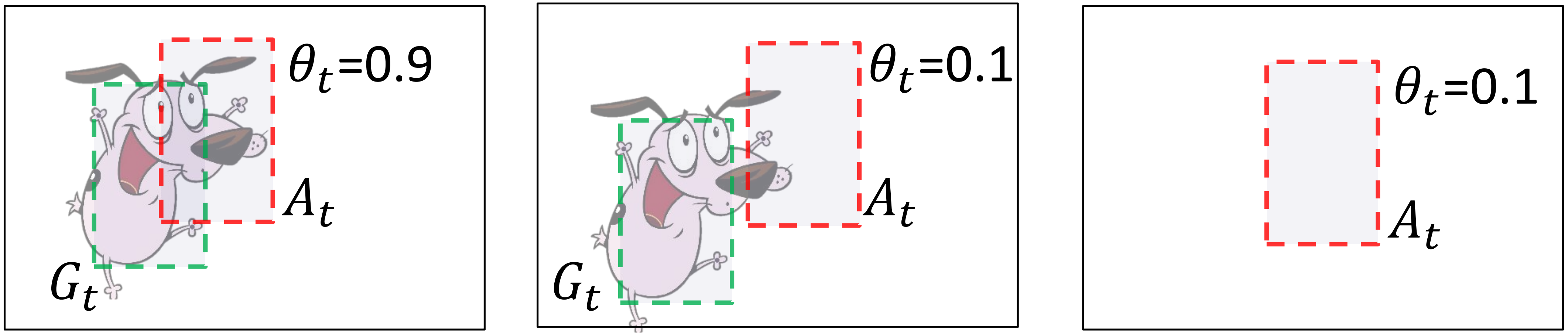


- Requirements: (i) localization accuracy, (ii) target absence prediction accuracy, (iii) re-detection accuracy
- Precision (Pr) … % of all predictions A_t that agree with GT G_t
- Recall (Re) … % of all GT boxes that that agree with predictions A_t
- F-measure … a standard Pr/Re tradeoff

$$F = 2PrRe/(Pr + Re)$$

[1]Lukežič, et al., Now you see me: evaluating performance in long-term visual tracking, Arxiv2018

LT performance measure design



- Agreement = sufficient overlap:

Detection “certainty” threshold

$$\Omega(A_t, G_t) \geq \tau_\Omega \longrightarrow \Omega(A_t(\tau_\theta), G_t) \geq \tau_\Omega$$

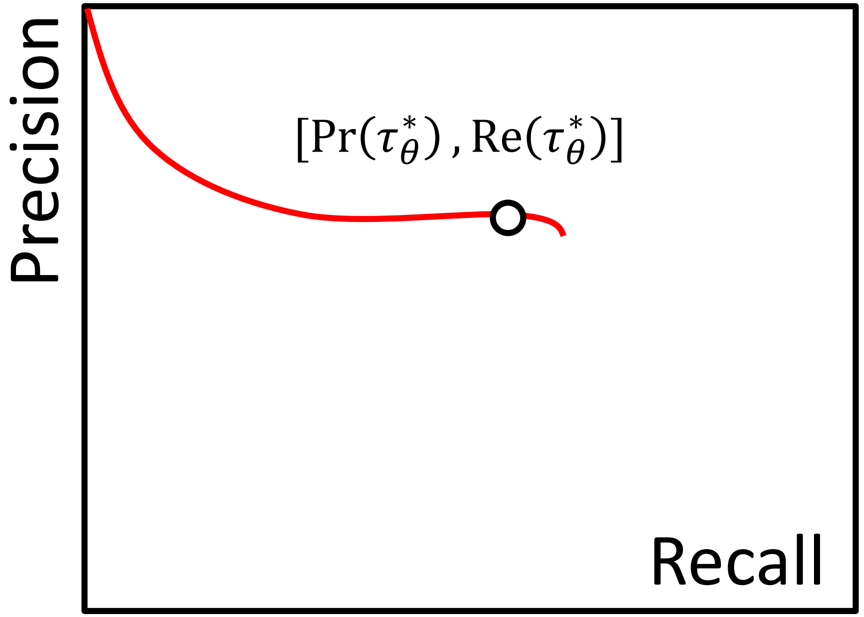
- Precision and Recall depend on two thresholds: $\mathrm{Pr}(\tau_\theta, \tau_\Omega)$, $\mathrm{Re}(\tau_\theta, \tau_\Omega)$
- The overlap threshold is avoided by integrating it out

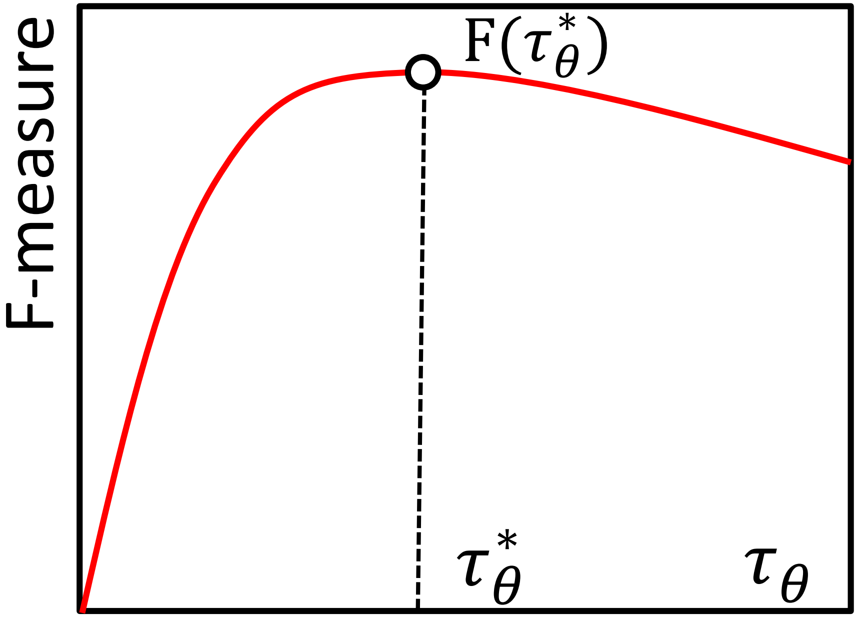
$$Pr(\tau_\theta) = \int_0^1 Pr(\tau_\theta, \tau_\Omega) d\tau_\Omega = \frac{1}{N_p} \sum_{t \in \{t : A_t(\tau_\theta) \neq \emptyset\}} \Omega(A_t(\tau_\theta), G_t),$$

$$Re(\tau_\theta) = \int_0^1 Re(\tau_\theta, \tau_\Omega) d\tau_\Omega = \frac{1}{N_g} \sum_{t \in \{t : G_t \neq \emptyset\}} \Omega(A_t(\tau_\theta), G_t)$$

Primary LT performance measures

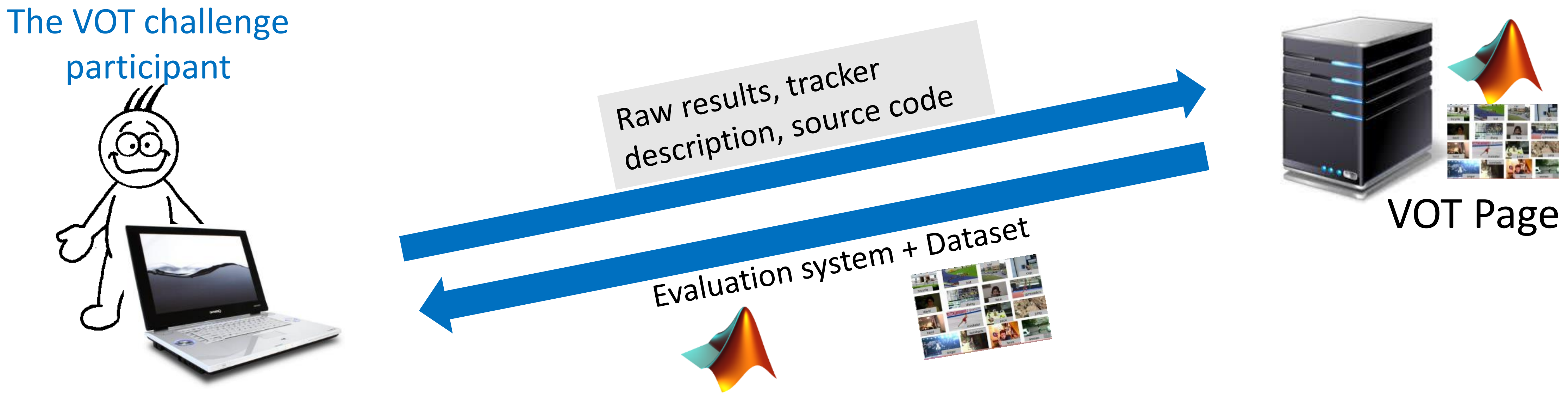
- Primary measures are $\mathrm{Pr}(\tau_\theta^*)$, $\mathrm{Re}(\tau_\theta^*)$ and $\mathrm{F}(\tau_\theta^*)$ evaluated at detection certainty threshold that maximizes the tracker F-measure





- Primary scores thus fully avoid manually setting the thresholds
- In short-term setup, $\mathrm{F}(\tau_\theta^*)$ reduces to a standard ST measure!

VOT2018 participation



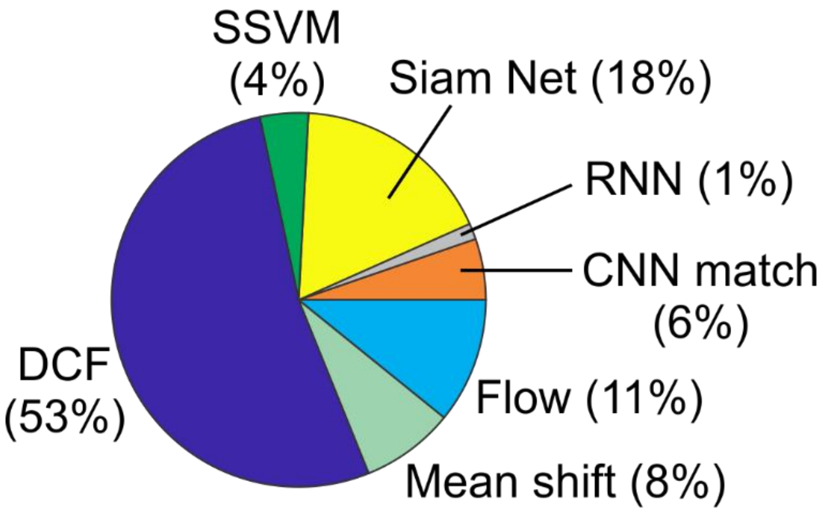
- Participants download the toolkit and the VOT2018 datasets
- Toolkit automatically performs all experiments
- Submission of raw results + tracker code required
- Top ten trackers re-run by VOT committee on the sequestered dataset

The VOT 2018 workshop

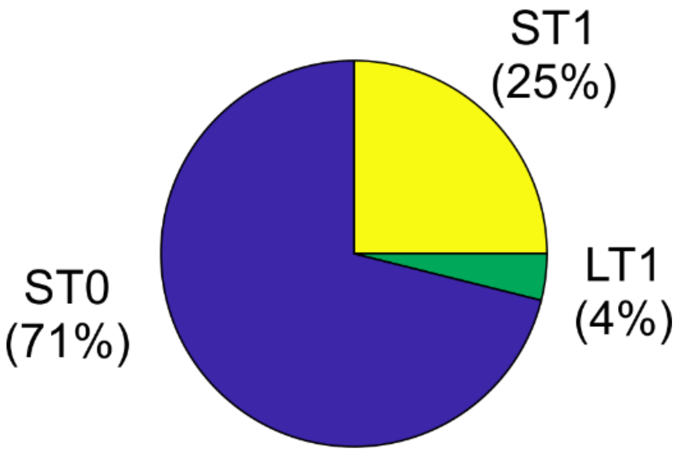
VOT2018 ST CHALLENGE RESULTS

VOT2018: 72 trackers tested

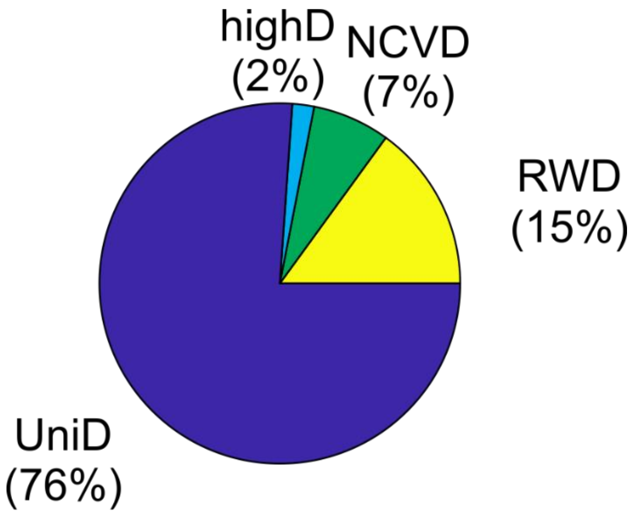
Tracking approach:



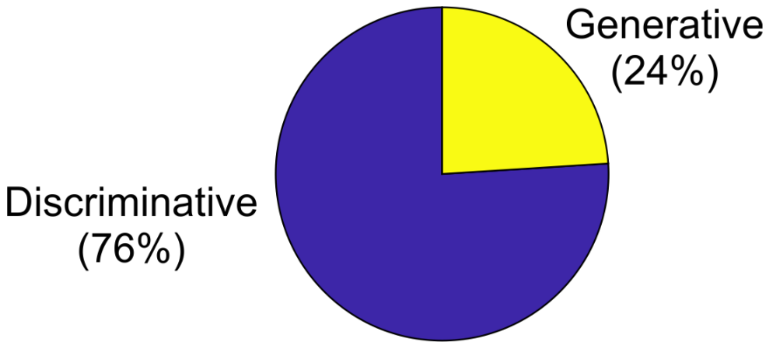
ST/LT category:



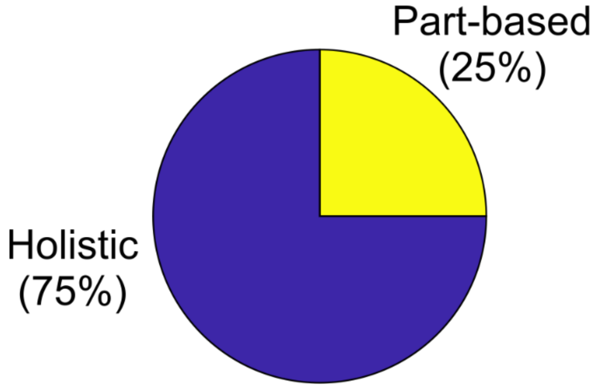
Motion model:



Generative/discriminative:



Holistic/parts:



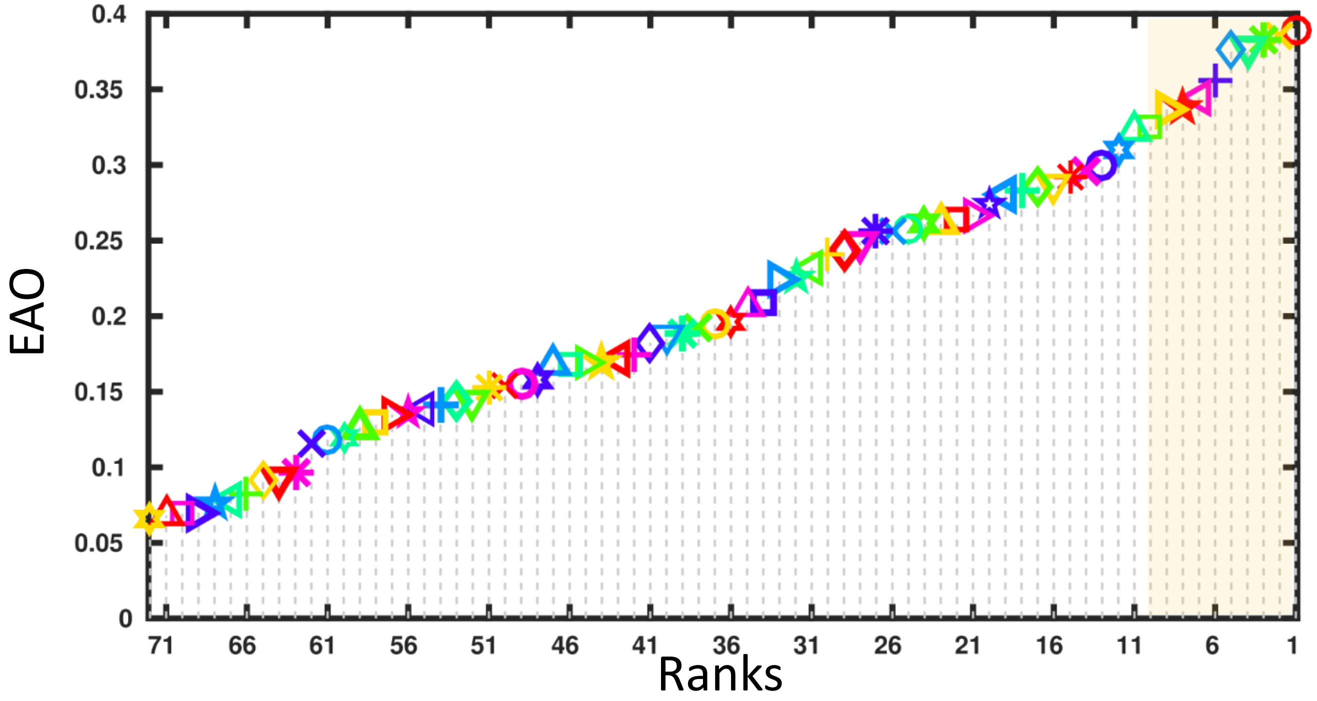
VOT2018 ST results on public dataset

- **Top trackers:** (1) LADCF, (2) MFT, (3) SiamRPN, (4) UPDT, (5) RCO, (6) DRT, (7) DeepSTRCF1, (8) SA_Siam_R, (9) CPT , (10) DLSTpp

- **Tracking approach in top 10:**

  8 DCF, except SiamPRN and SA_Siam_R (based on Siamese nets)

- **Features:**
  - All CNN (+handcrafted)
    Top two apply Resnet50
  - Most CNN trained for detection
  - Some trained for localization, (e.g. MFT)

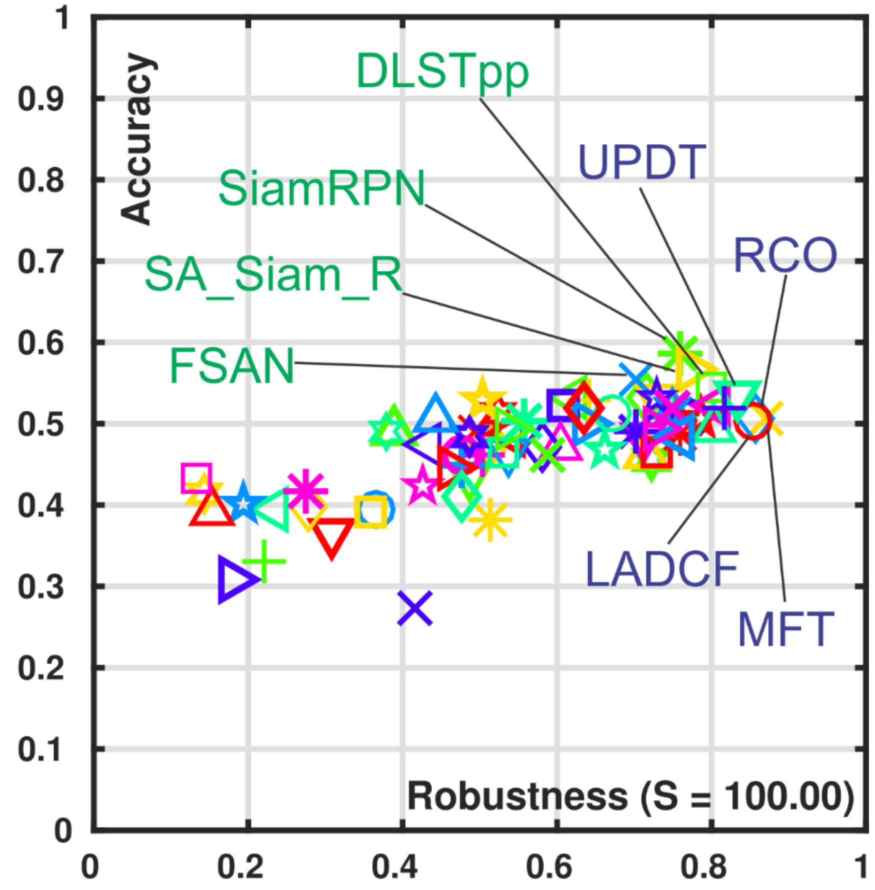


VOT2018 results on public dataset

- Top trackers are among the most robust trackers
  (1) MFT, (2) LADCF, (3) RCO, (4) UPDT
- Top in accuracy:
  (1) SiamRPN, (2) SA_Siam_R, (3) FSAN, (4) DLSTpp
- Per-attribute analysis:

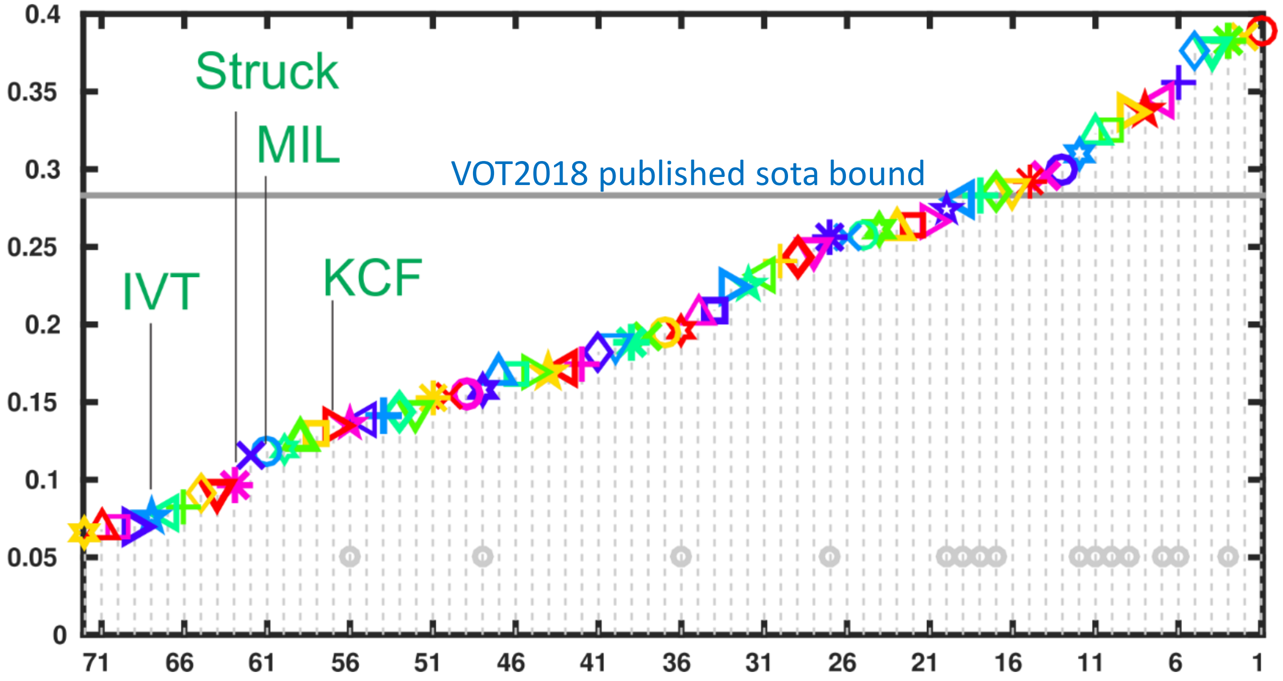
	CM	IC	MC	OC	SC
Accuracy	0.49	0.47	0.47 ③	0.40 ①	0.43 ②
Robustness	0.74	1.05 ②	0.87 ③	1.19 ①	0.61

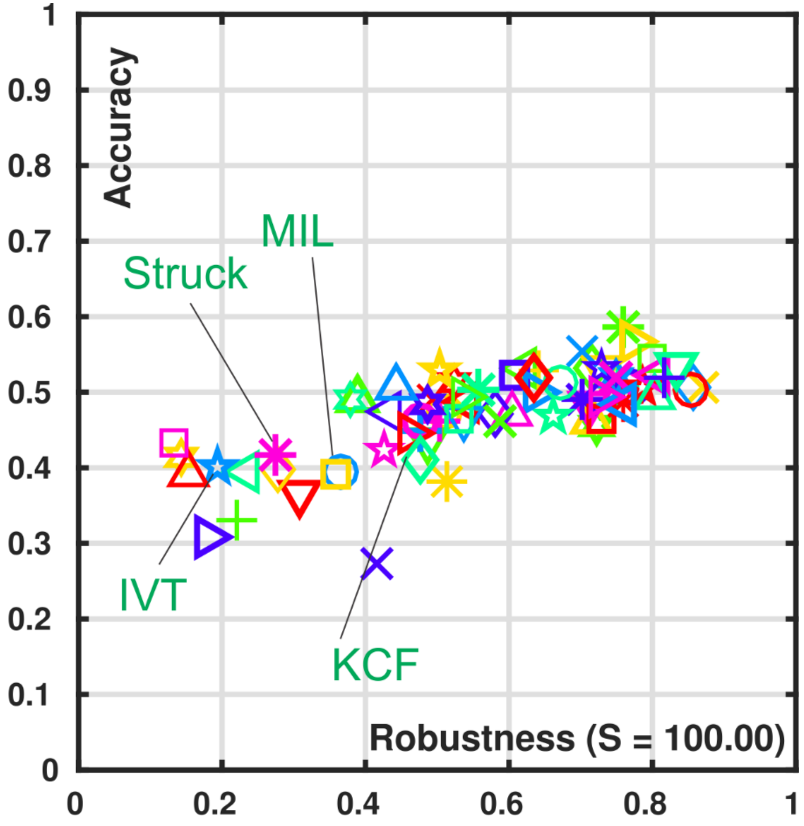
- Most failures due to: Occlusion
- Mostly affects accuracy: Occlusion + Scale change



VOT2018 results on public dataset

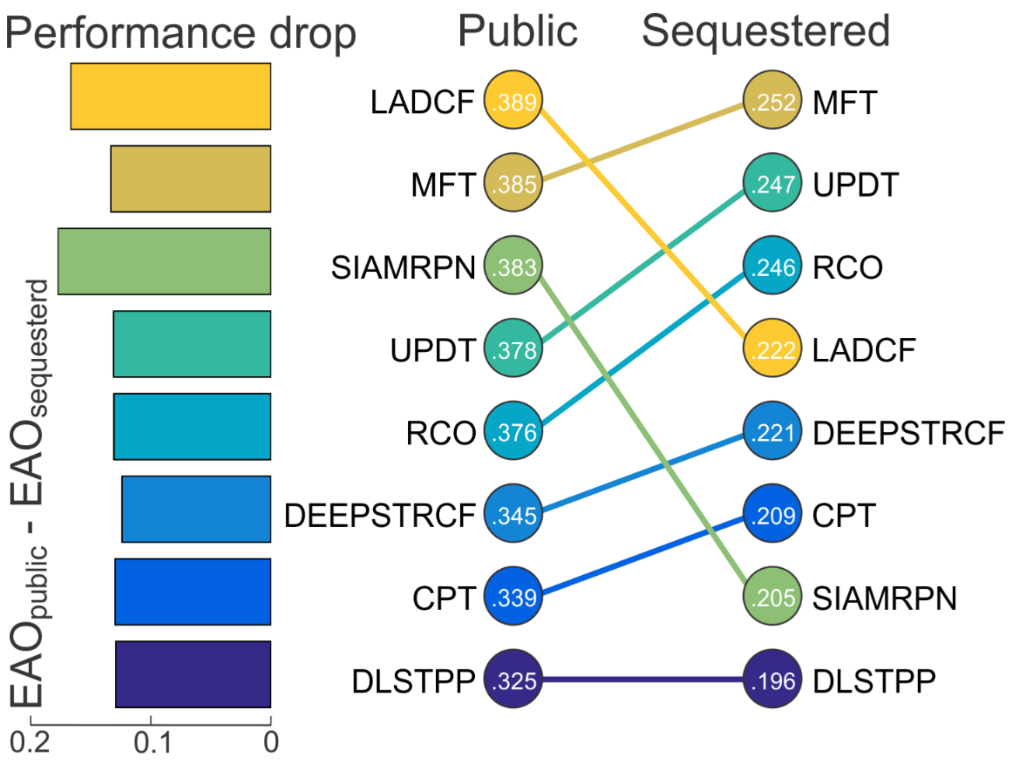
- Baselines ranked at the very tail of the benchmark
- 19 trackers published at major CV venues (2017≤)
  - Their average performance: VOT2018 sota bound
  - Over 26% submissions exceed this bound

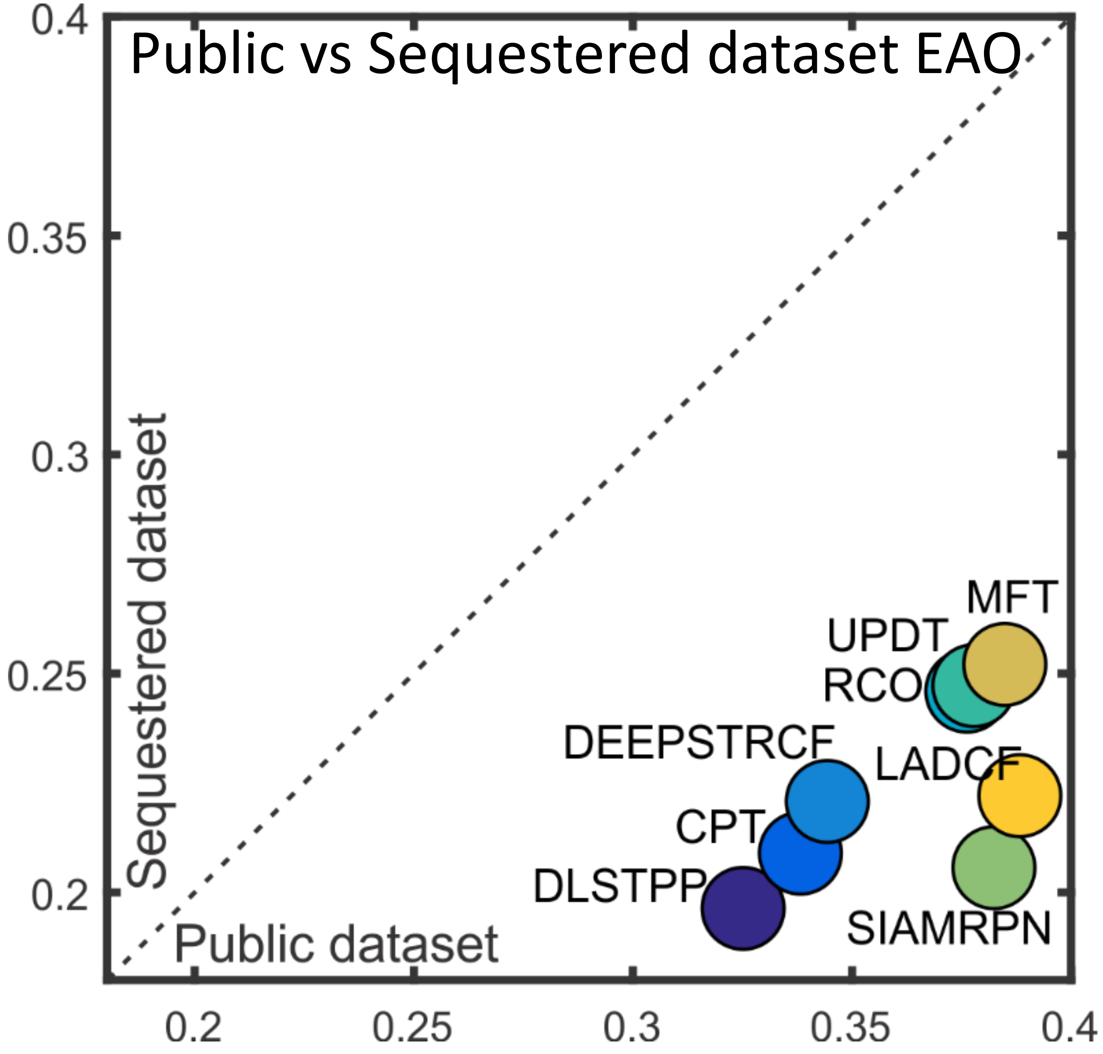




VOT2018 results on sequestered dataset

- Fairly stable ranks
- Greatest change: SiamRPN
- Smallest rank change: MFT

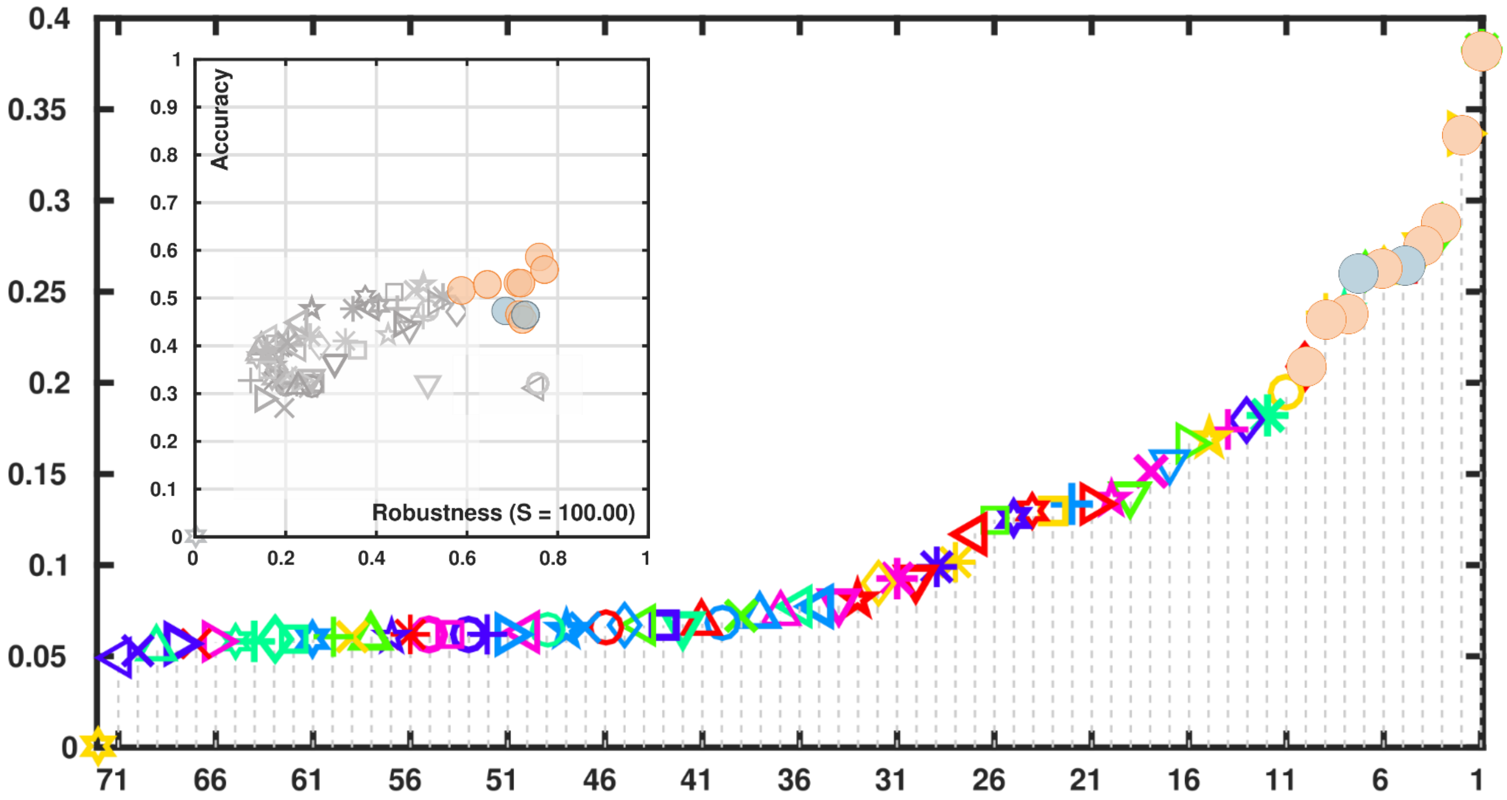




Matej Kristan (matej.kristan@fri.uni-lj.si)

VOT2018 ST realtime results

- Top 10: (1) SiamRPN, (2) SA_Siam_R, (3) SA_Siam_P, (4) SiamVGG, (5) CSRTPP, (6) LWDNTm, (7) LWDNTthi, (8) CSTEM, (9) MBSiam, (10) UpdateNet



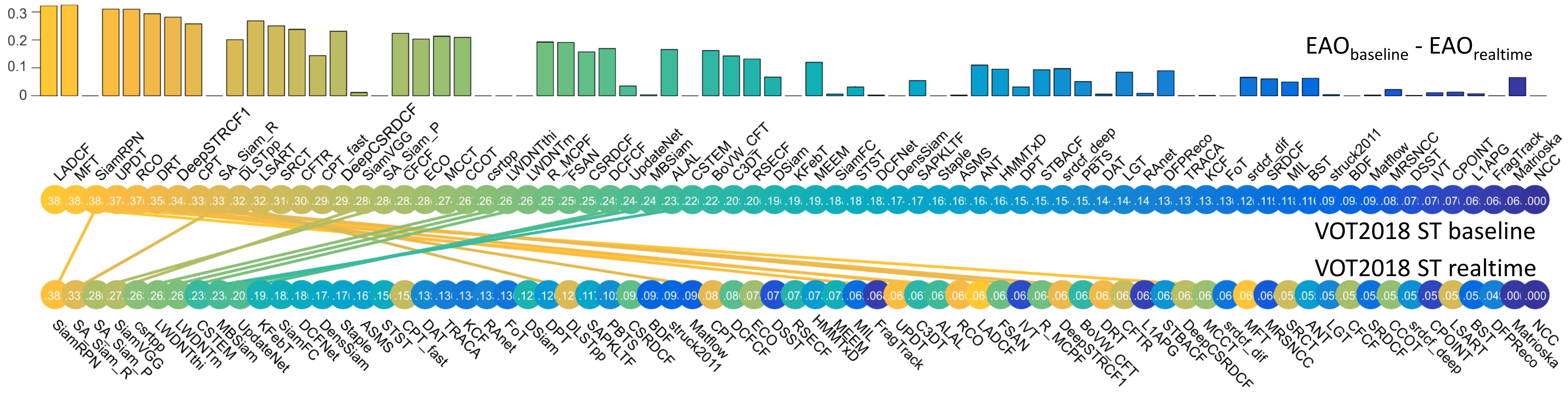
Two classes:

Approach: Siamese (e.g., SiamFc[1])
Features: CNN
Hardware: GPU

Approach: CSRDCF[2]
Features: HOG+CN
Hardware: CPU

[1]Bertinetto, et al., VOT2016, [2]Lukežič, et. al., CVPR2017

VOT2018 ST Realtime vs Baseline results



- Most of the top baseline performers drop with real-time constraint
- The drop is smaller for real-time trackers on baseline challenge
- Some achieve top real-time performance AND perform well on the baseline

The VOT 2018 workshop

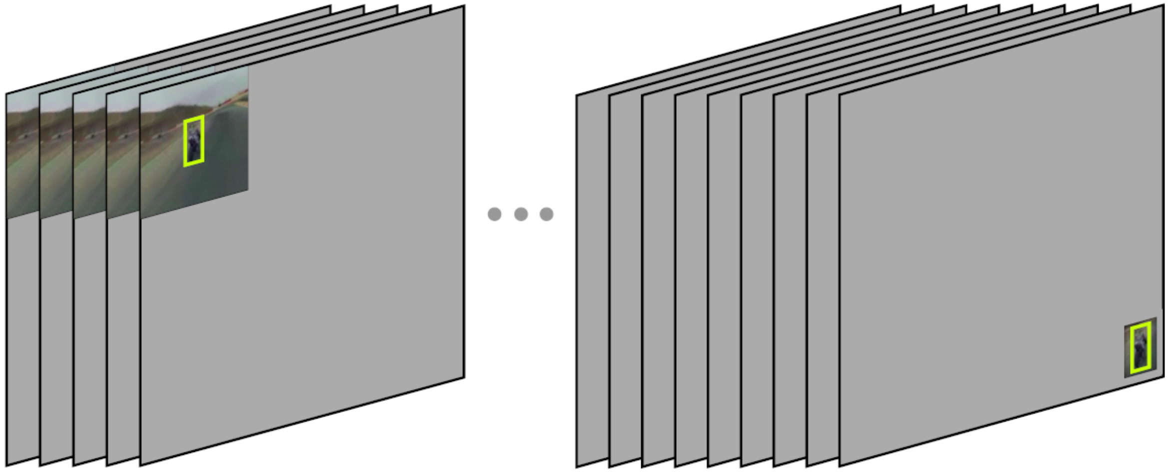
VOT LONG-TERM CHALLENGE RESULTS

VOT2018 long-term challenge: trackers tested

- 15 trackers tested
- 5 trackers from ST_0 class (elevated to LT0)
- 10 trackers from LT_1 class
- LT trackers composed of: a short-term component and a detector

Image-wide re-detection test

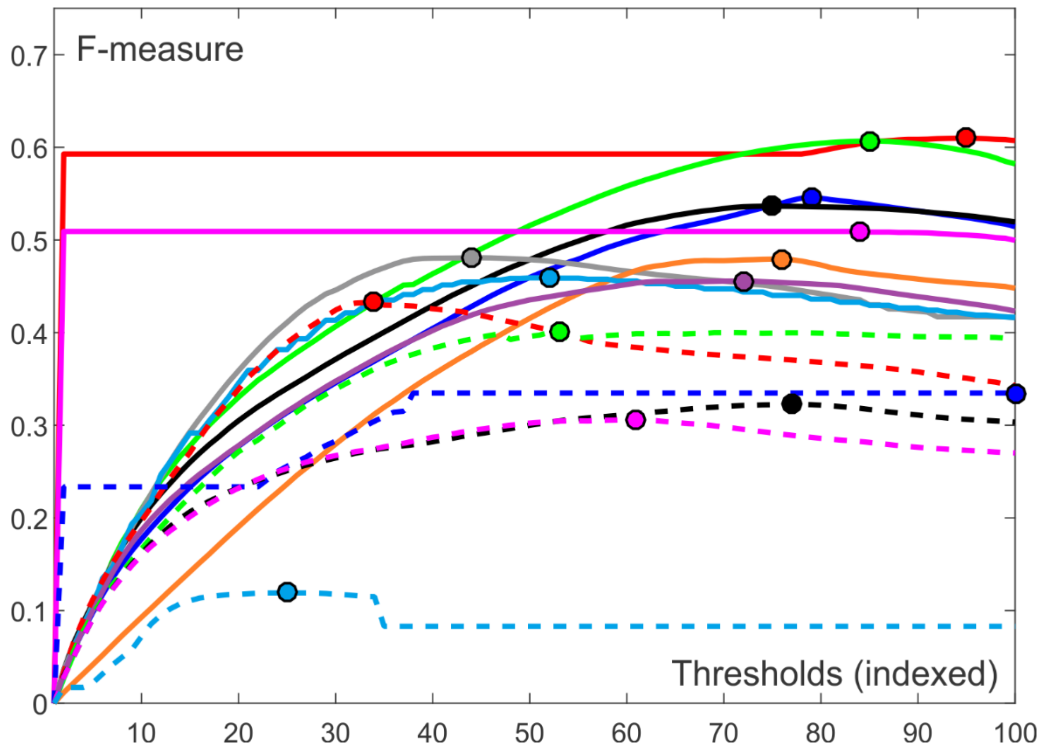
- Designed to test a crucial ingredient: *Image-wide re-detection*
- Created artificial sequences: *Only target position changes*

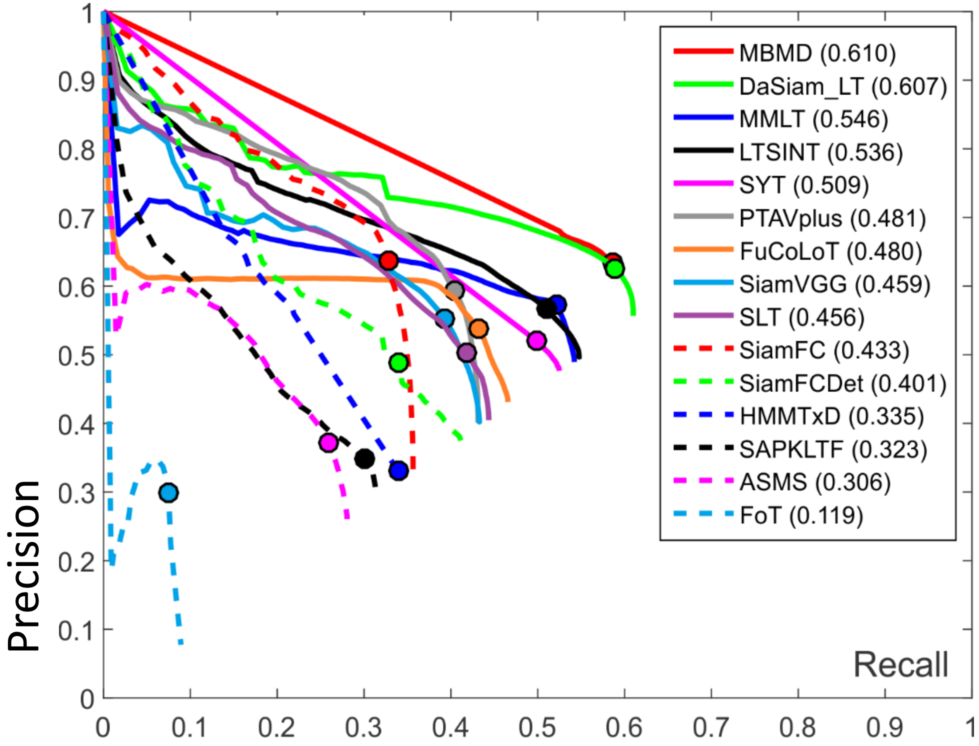


	Tracker	Frames (Success)
1.	MBMD	1 (100%)
2.	DaSiam_LT	- (0%)
3.	MMLT	0 (100%)
4.	LTSINT	2 (100%)
5.	SYT	0 (43%)
6.	PTAVplus	0 (11%)
7.	FuCoLoT	78 (97%)
8.	SiamVGG	- (0%)
9.	SLT	0 (100%)
10.	SiamFC	- (0%)
11.	SiamFCDet	0 (83%)
12.	HMMTxD	3 (91%)
13.	SAPKLTF	- (0%)
14.	ASMS	- (0%)
15.	FoT	0 (6%)

Image-wide detection: 8 trackers
(nearly instant detection, with exceptions)
7 trackers id not pass the LT1 test: Local detection at most

Overall results – the benchmark



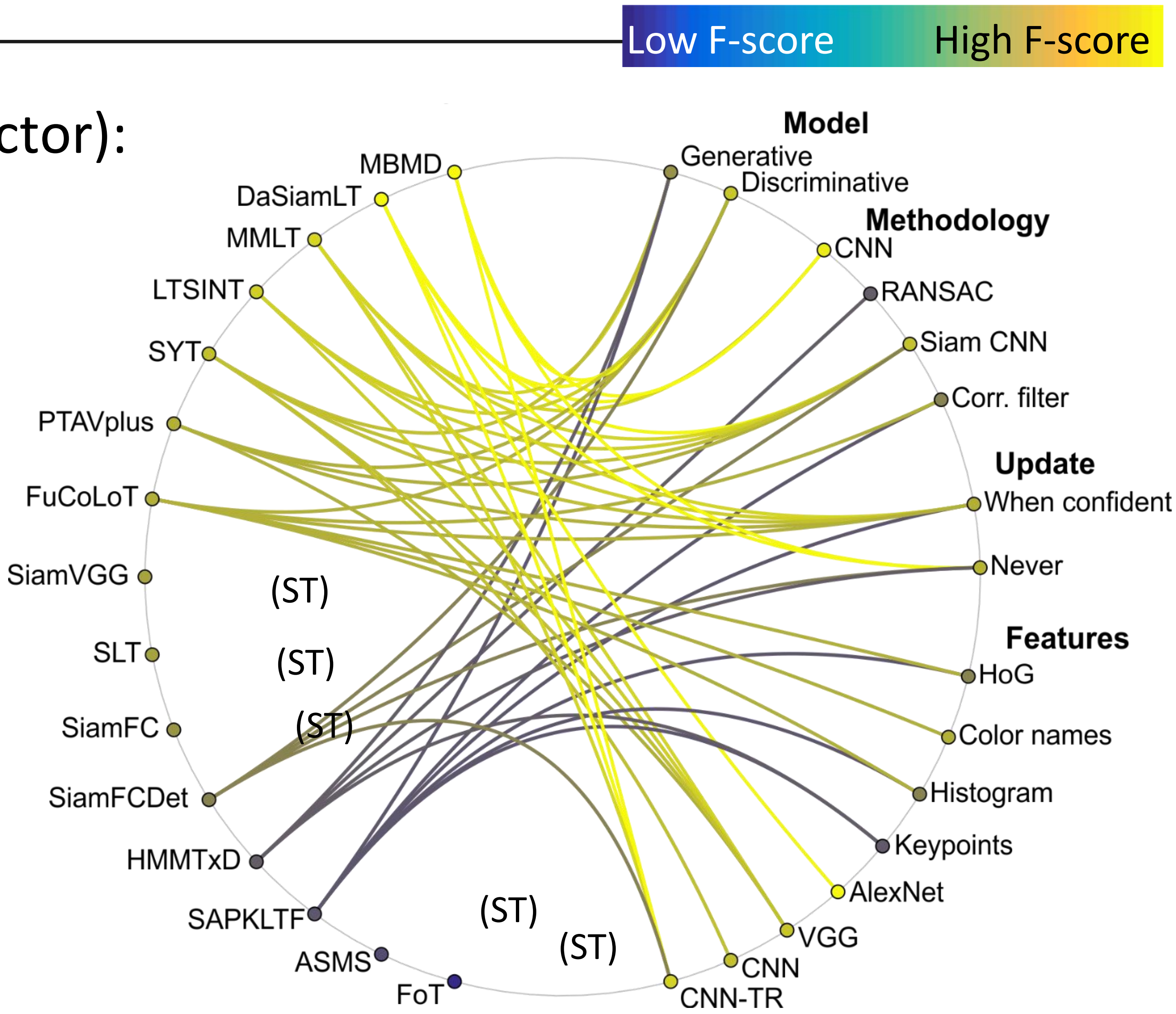


- Top 10: 9 CNN-based (mostly Siamese architectures), 2 apply DCF
- Best: MBMD (Based on MDNet[1]) – bounding box regression + classification
- All top trackers are LT_1 with image-wide detection except DaSiam_LT

[1]Nam, Han, Learning, Multi-Domain Convolutional Neural Networks for Visual Tracking, CVPR2016

Architecture analysis:

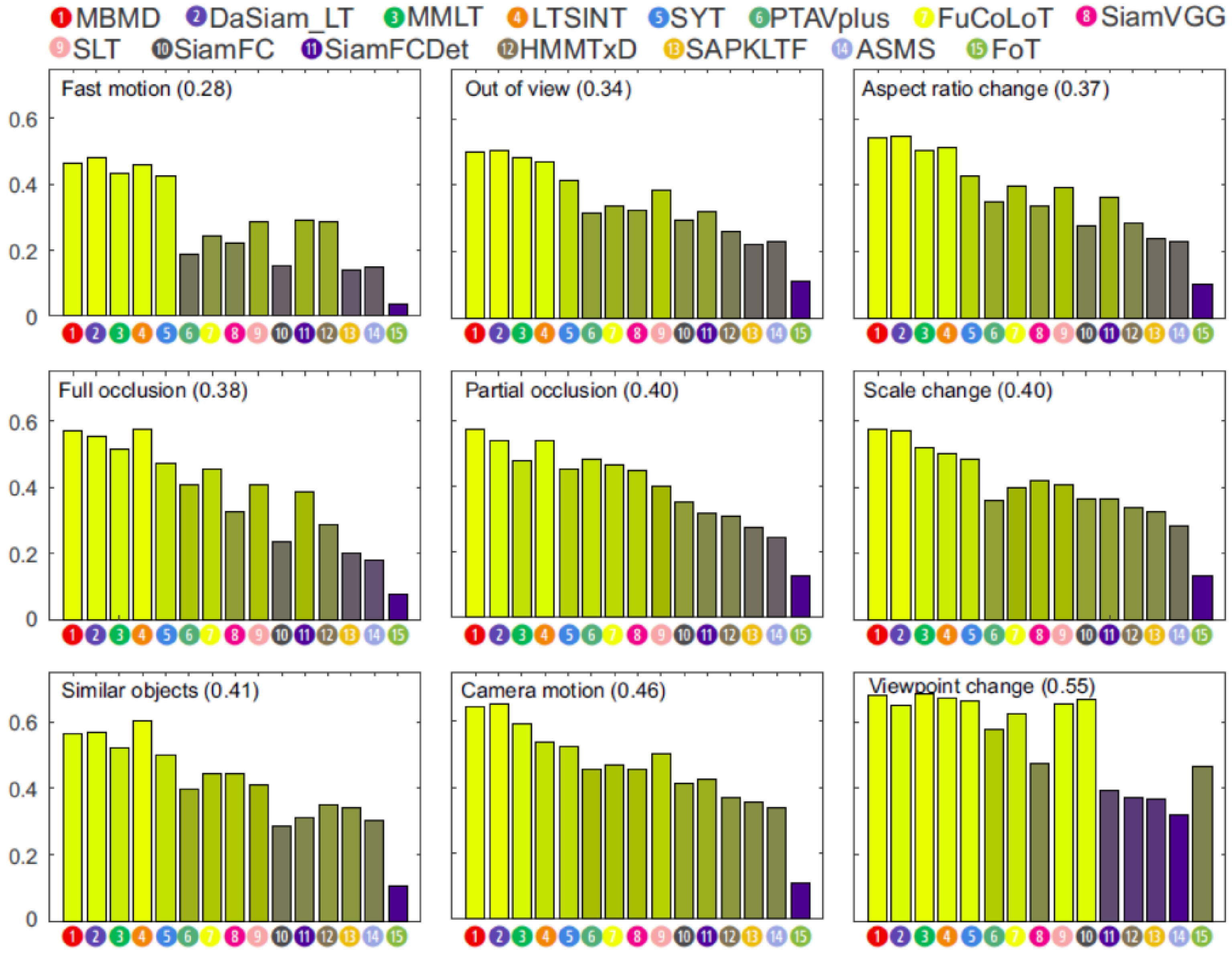
- Long-term component (Detector):
  - CNNs consistently deliver top performance
  - DCFs show great promise (when properly learned)
  - Deep features dominate
- Short-term component:
  - Follow trend observed in ST tracking benchmarks



Attribute analysis

Most challenging:

- Fast motion
- Out of view
- Aspect ratio change

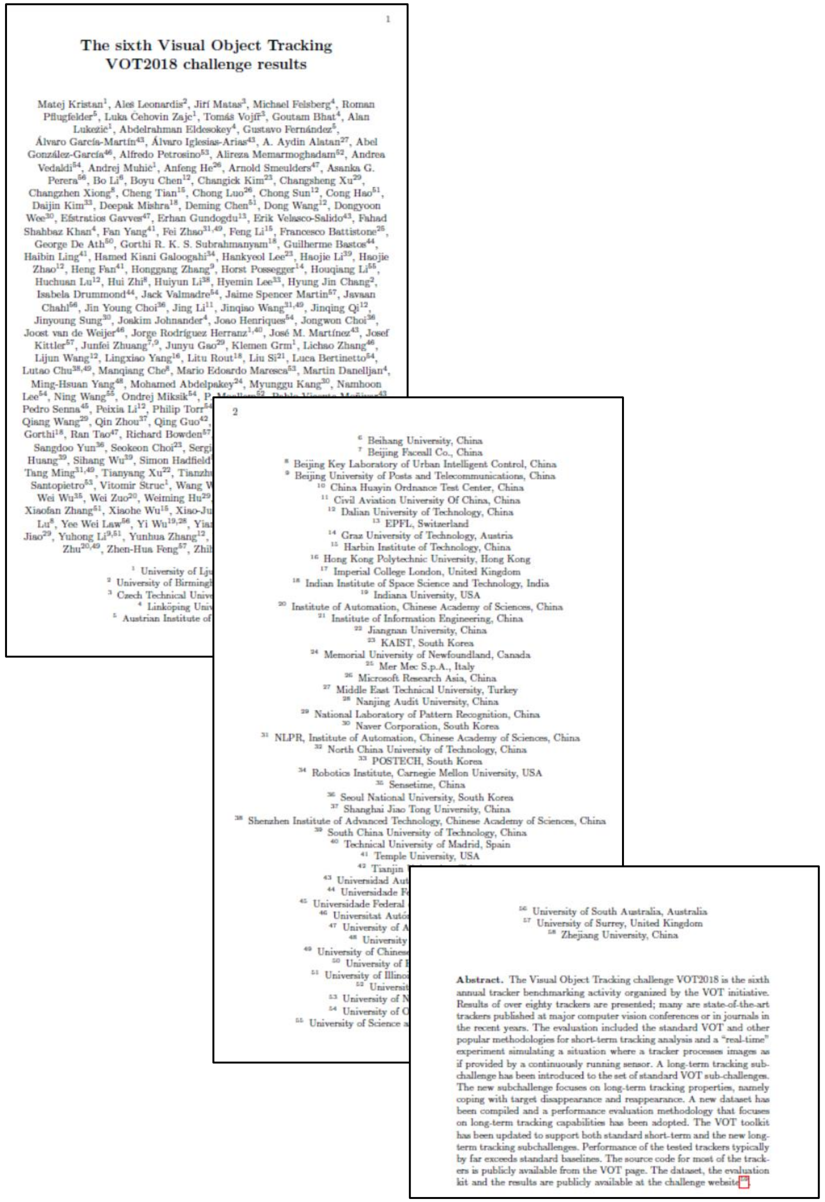


VOT2018 challenges Summary

- VOT2018 ST baseline:
  - DCF the dominant methodology, Deep features the dominant features
  - Wider use of deep features trained for localization
- VOT2018 ST realtime:
  - Best performance: fully convolutional deep approaches and DCFs
  - Some of the fastest trackers also rank among best on baseline
- VOT2018 LT:
  - Explicit object detection integrated
  - Nearly all top-performers CNN-based, only one purely DCF
  - Top two performers do not update the detector

VOT2018 online resources

- Results in 155 coauthor paper (58 institutions)
- Available at http://www.votchallenge.net/vot2018
  - Presentations + papers + Dataset + Evaluation kit
  - Guidelines on how to evaluate your trackers on VOT2018 and produce graphs for your papers
    (directly comparable to 72 ST and ~10 LT trackers!)
- VOT is open source:
  - All toolkits and protocols on Github
  - Trackers code available:
    VOT2018 (72 ST, 15 LT), VOT2017 (36), VOT2016 (39)



VOT2018 awards:

Winners of the VOT2018 short-term challenge:

MFT by: S. Bai, Z. He, J. Zhuang

“Multi-solution Fusion for Visual Tracking ”

VOT2018 awards:

Winners of the VOT2018 ST realtime challenge:

SiamRPN by: Q. Wang, Z. Zhu, B. Li, W. Wu, W. Hu, W. Zou

"High Performance Visual Tracking with Siamese Region Proposal Network"

(talk after the Poster session!)

visual object tracking challenge

University of Ljubljana
Faculty of Computer and Information Science

SICK
Sensor Intelligence.

VOT2018 awards:

Winners of the VOT2018 long-term challenge:

MBMD by: Y. Zhang, L. Wang, D. Wang, J. Qi, H. Lu

"MobileNet based tracking by detection algorithm"

(talk after the Poster session!)

University of Ljubljana
Faculty of Computer and Information Science

SICK
Sensor Intelligence.

visual object tracking challenge

Thanks

- The VOT2018 committee


M. Kristan


J. Matas


A. Leonardis


M. Felsberg


R. Pflugfelder


G. Fernandez


L. Čehovin


T. Vojir


A. Lukežič


G. Bhat


A. Eldesokey

- Everyone who participated or contributed

Matej Kristan1, Ales Leonardis2, Jir Matas3, Michael Felsberg4, Roman Pfugfelder5, Luka Cehovin Zajc1, Tomas Voj~r3, Goutam Bhat4, Alan Lukezic1, Abdelrahman Eldesokey4, Gustavo Fernandez5, Alvaro Garca-Martn43, Alvaro Iglesias-Arias43, A. Aydin Alatan27, Abel Gonzalez-Garca46, Alfredo Petrosino53, Alireza Memarmoghadam52, Andrea Vedaldi54, Andrej Muhic1, Anfeng He26, Arnold Smeulders47, Asanka G. Perera56, Bo Li6, Boyu Chen12, Changick Kim23, Changsheng Xu29, Changzhen Xiong8, Cheng Tian15, Chong Luo26, Chong Sun12, Cong Hao51, Daijin Kim33, Deepak Mishra18, Deming Chen51, Dong Wang12, Dongyoon Wee30, Efstratios Gavves47, Erhan Gundogdu13, Erik Velasco-Salido43, Fahad Shahbaz Khan4, Fan Yang41, Fei Zhao31;49, Feng Li15, Francesco Battistone25,George De Ath50, Gorthi R. K. S. Subrahmanyam18, Guilherme Bastos44, Haibin Ling41, Hamed Kiani Galoogahi34, Hankyeol Lee23, Haojie Li39, Haojie Zhao12, Heng Fan41, Honggang Zhang9, Horst Possegger14, Houqiang Li55,Huchuan Lu12, Hui Zhi8, Huiyun Li38, Hyemin Lee33, Hyung Jin Chang2, Isabela Drummond44, Jack Valmadre54, Jaime Spencer Martin57, Javaan Chahl56, Jin Young Choi36, Jing Li11, Jinqiao Wang31;49, Jinqing Qi12, Jinyoung Sung30, Joakim Johnander4, Joao Henriques54, Jongwon Choi36, Joost van de Weijer46, Jorge Rodrguez Herranz1;40, Jose M. Martnez43, Josef Kittler57, Junfei Zhuang7;9, Junyu Gao29, Klemen Grm1, Lichao Zhang46, Lijun Wang12, Lingxiao Yang16, Litu Rout18, Liu Si21, Luca Bertinetto54, Lutao Chu38;49, Manqiang Che8, Mario Edoardo Maresca53, Martin Danelljan4, Ming-Hsuan Yang48, Mohamed Abdelpakey24, Myunggu Kang30, Namhoon Lee54, Ning Wang55, Ondrej Miksik54, P. Moallem52, Pablo Vicente-Mo~nivar43, Pedro Senna45, Peixia Li12, Philip Torr54, Priya Mariam Raju18, Qian Ruihe21, Qiang Wang29, Qin Zhou37, Qing Guo42, Rafael Martn-Nieto43, Rama Krishna Gorthi18, Ran Tao47, Richard Bowden57, Richard Everson50, Runling Wang32, Sangdoo Yun36, Seokeon Choi23, Sergio Vivas43, Shuai Bai7;9, Shuangping Huang39, Sihang Wu39, Simon Had eld57, Siwen Wang12, Stuart Golodetz54, Tang Ming31;49, Tianyang Xu22, Tianzhu Zhang29, Tobias Fischer17, Vincenzo Santopietro53, Vitomir Struc1, Wang Wei10, Wangmeng Zuo15, Wei Feng42, Wei Wu35, Wei Zuo20, Weiming Hu29, Wengang Zhou55, Wenjun Zeng26, Xiaofan Zhang51, Xiaohe Wu15, Xiao-Jun Wu22, Xinmei Tian55, Yan Li8, Yan Lu8, Yee Wei Law56, Yi Wu19;28, Yiannis Demiris17, Yicai Yang39, Yifan Jiao29, Yuhong Li9;51, Yunhua Zhang12, Yuxuan Sun12, Zheng Zhang58, Zheng Zhu20;49, Zhen-Hua Feng57, Zhihui Wang12, and Zhiqun He7;9

- VOT2018 sponsors:

University of Ljubljana
Faculty of Computer and Information Science

SICK
Sensor Intelligence.