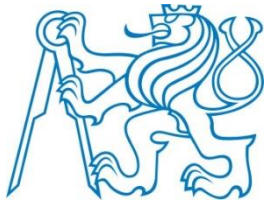
The Visual Object Tracking Challenge Results VOT2017

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

University *of Ljubljana*
Faculty *of Computer and Information Science*

UNIVERSITY OF BIRMINGHAM



LINKÖPING UNIVERSITY

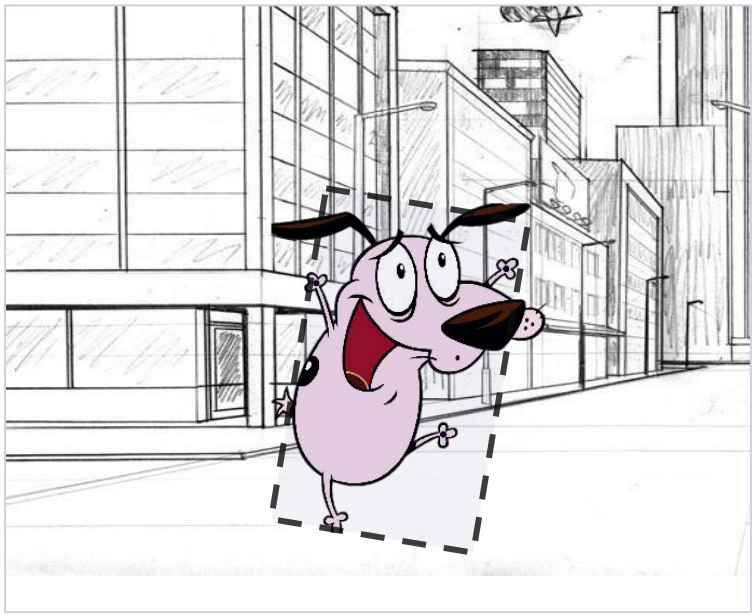
AUSTRIAN INSTITUTE OF TECHNOLOGY

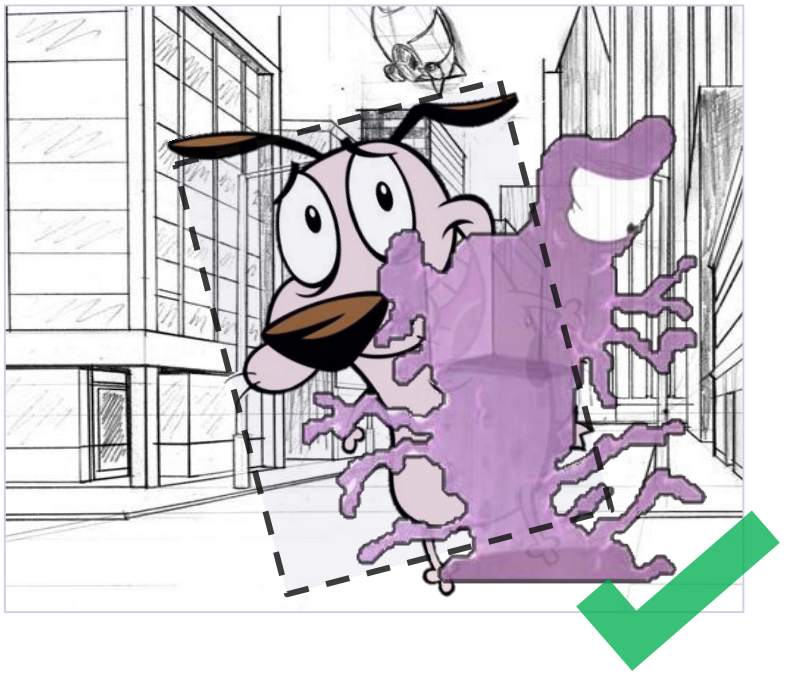
Outline

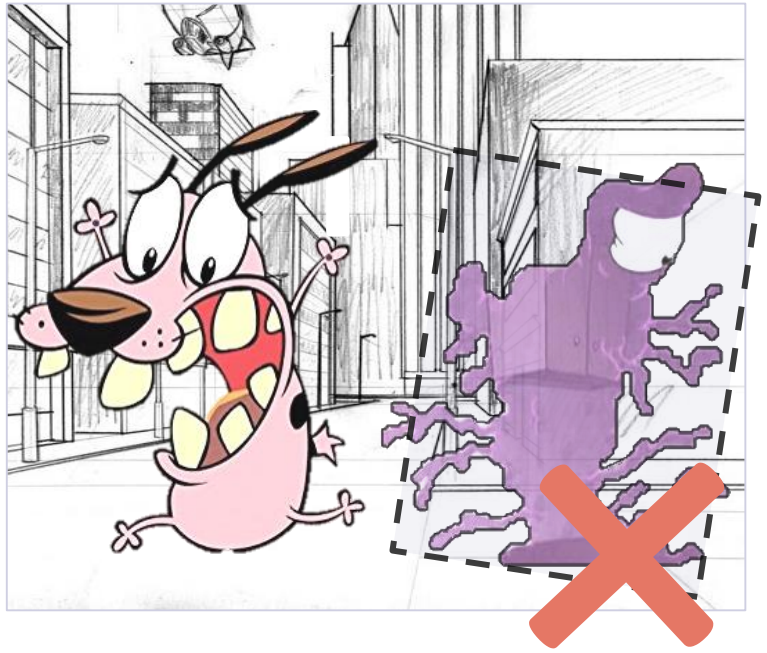
1. Scope of the VOT challenge
2. VOT2017 challenge overview and novelties introduced
3. VOT2017 results overview + winner announcement
4. Summary and outlook

Scope of the VOT2017 challenges

- Short-term, single-target, causal trackers
- No redetection: drift is considered a failure and tracker is reset







- Target state represented by a rotated bounding box

The VOT2017 evaluation system

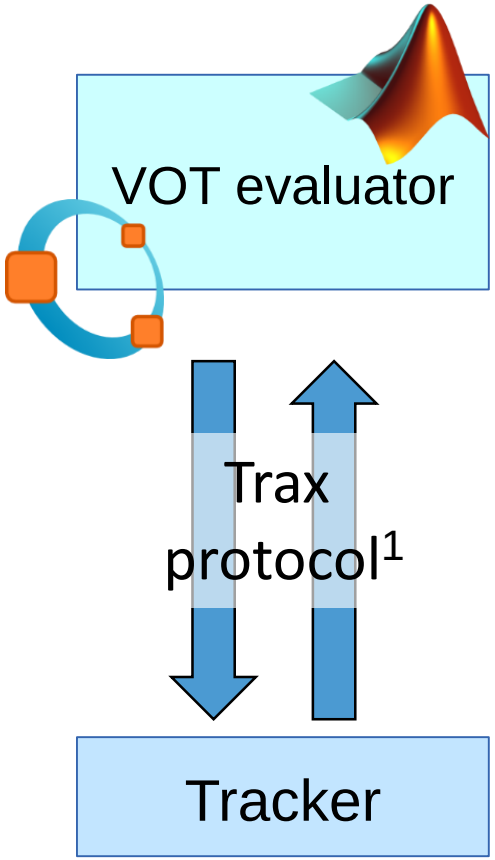
- A Matlab toolkit automatically performs a battery of standard experiments

  Currently the most advanced toolkit in visual tracking.
  https://github.com/vicoslab/vot-toolkit

- Plug and play!
  - Supports major programming languages and operating systems

- New version: Changes made to support new experiments

- Transition to the latest version of the toolkit was required (VOT2013 toolkit not supported)



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

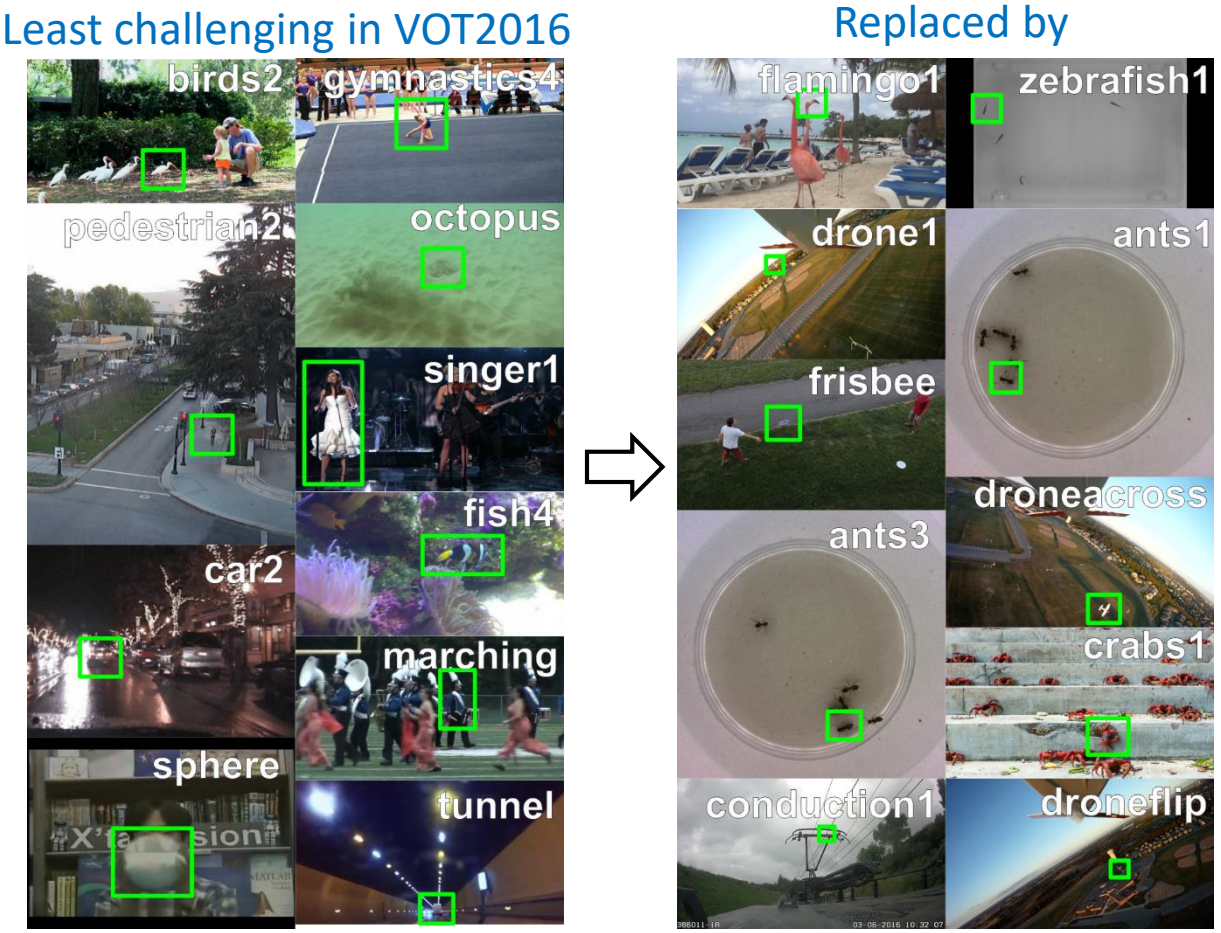
The VOT2017 dataset

- VOT approach: Keep it sufficiently small, well annotated and diverse
- Developed the VOT dataset construction methodology
- New this year two datasets: public and a sequestered

- Public: refreshed VOT2016 dataset
  (for tracking community)
- Sequestered: new sequences, similar to the VOT2017 public
  (for winner identification)

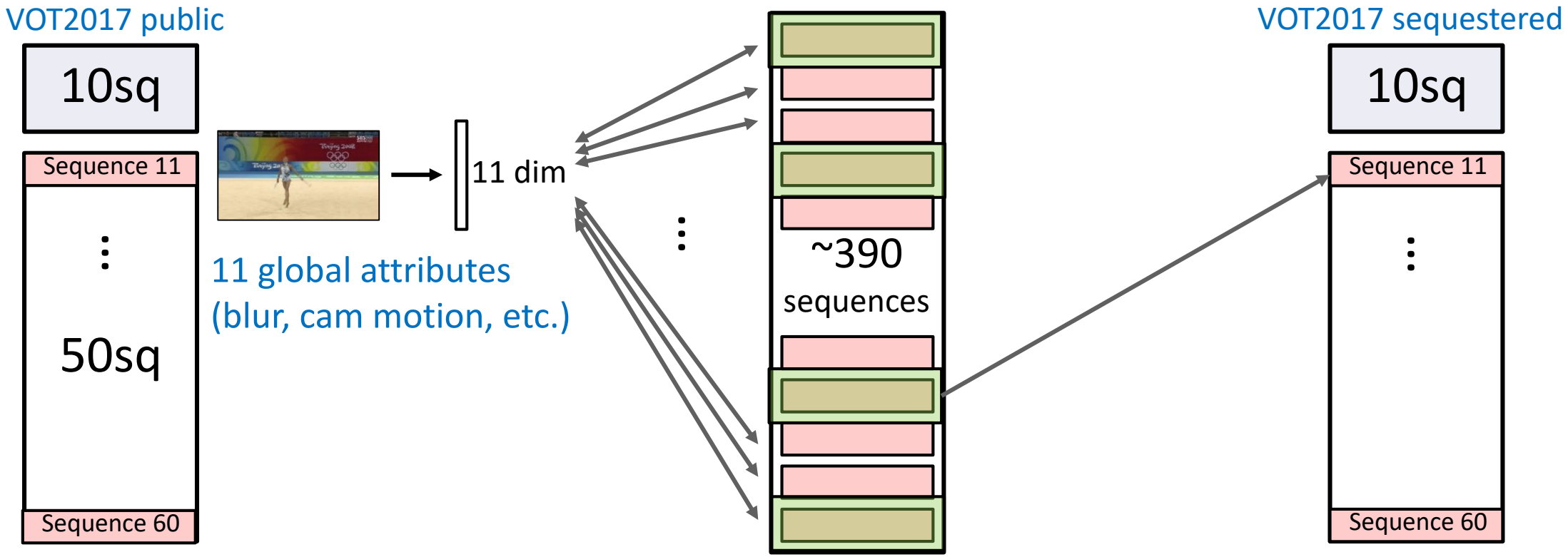
The VOT2017 public dataset

- VOT2016 dataset (60 sequences) not saturated, still challenging (refresh required only)
- Obtained 10 pairs of new sequences (not present in other benchmarks)
- Replaced 10 least challenging sequences in VOT2016 (~17%)
- VOT2016 difficulty level estimation methodology applied



The VOT2017 sequestered dataset

- Guideline: similar to the public dataset in attributes, but not easier
- 10 remaining new sequences entered the sequestered dataset
- Remaining 50 sequences selected from a pool of ~390 sequences



The VOT2017 dataset annotation

- Each image annotated by 6 attributes:

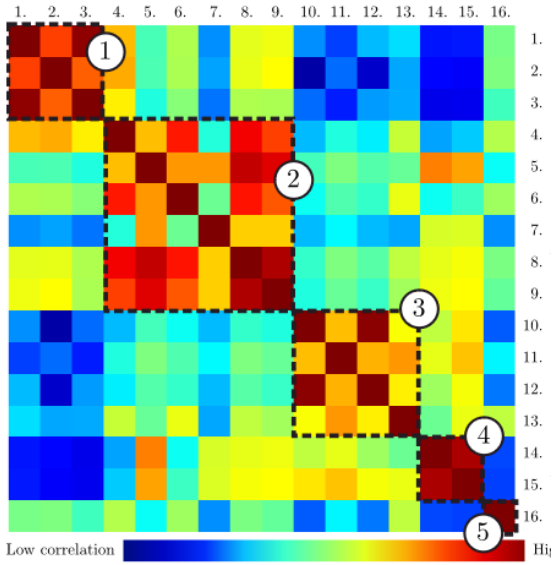
  Occlusion, Illumination change , Object motion, Object size change, Camera motion, Unassigned

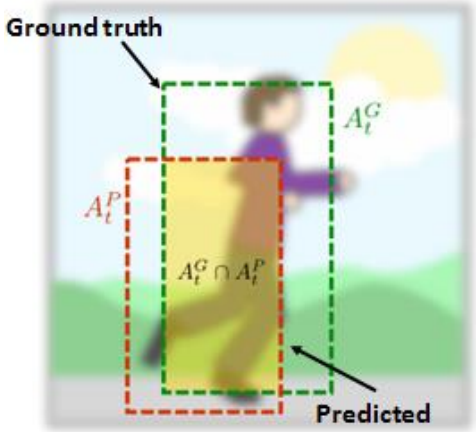
- Each image semi-automatically segmented
- A bounding box fitted automatically to segmentation mask (VOT2016 protocol)

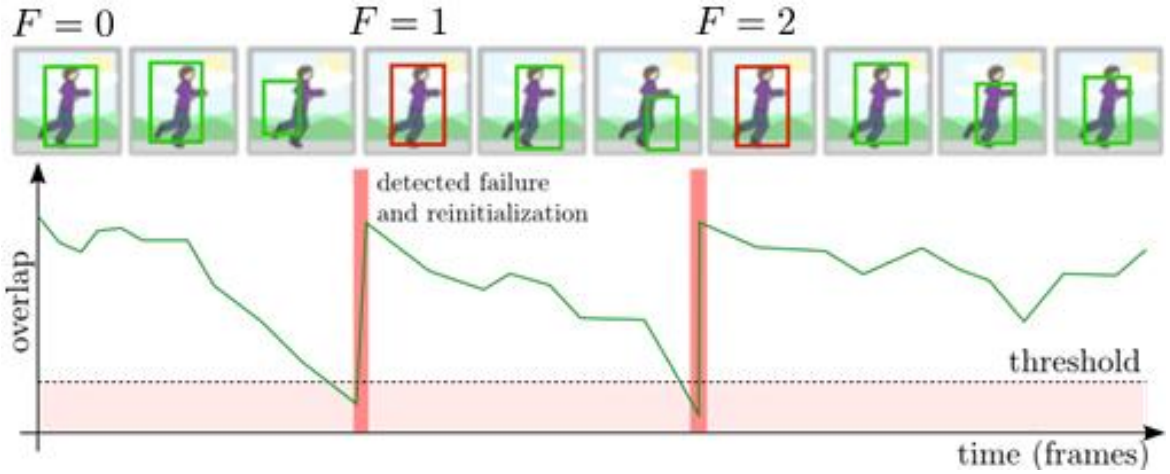
Evaluation methodology as in VOT2016

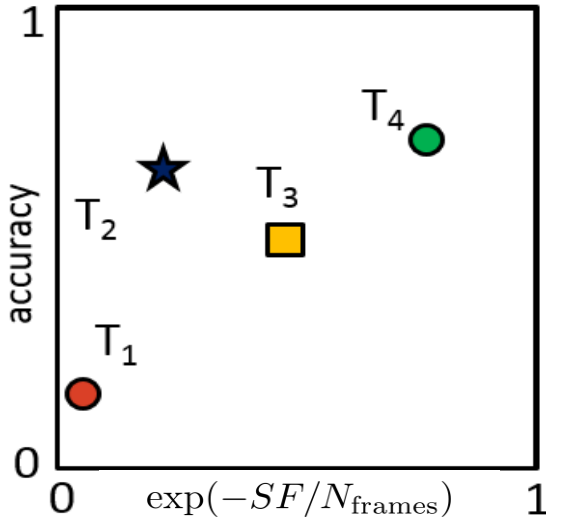
- Two weakly correlated measures[2] chosen according to[1]:
  - Robustness (number of times a is reinitialized)
  - Accuracy (average overlap while tracking)
  - + Combination of basic measures (EAO)

Performance measure correlation analysis[1]









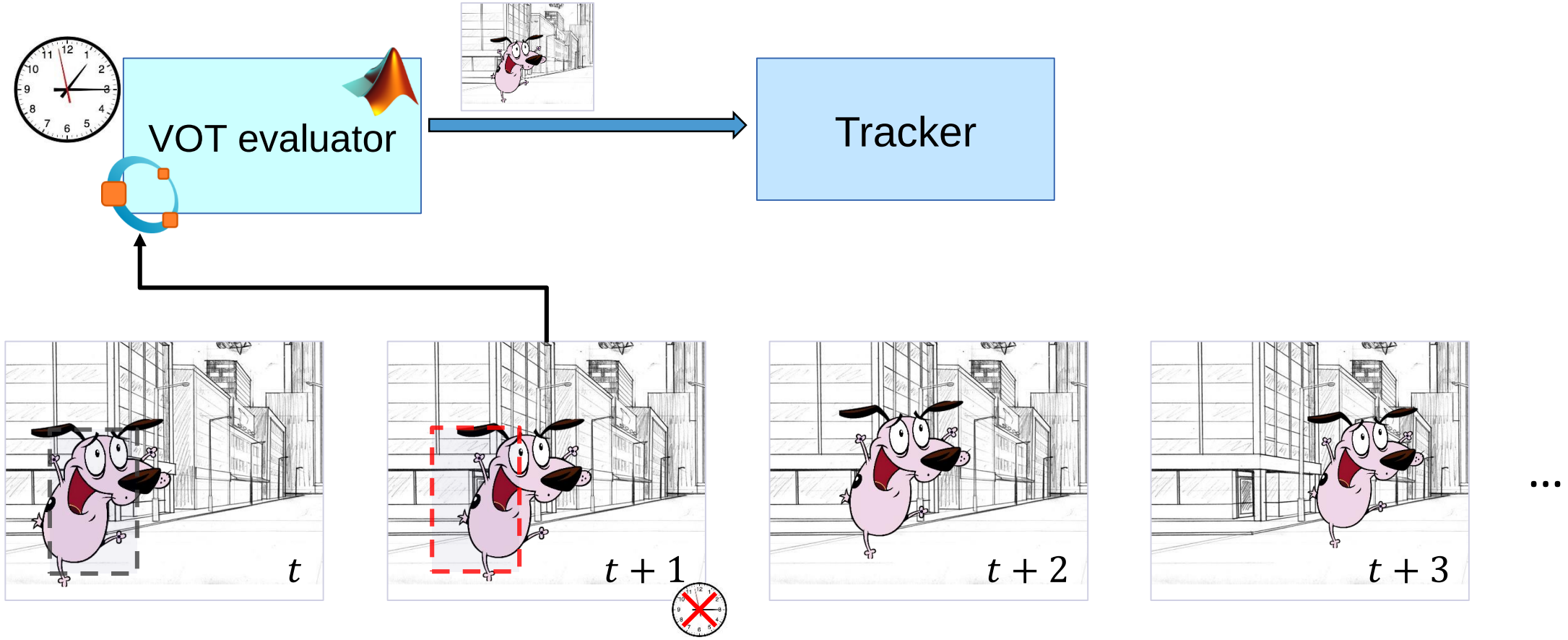
[1]Čehovin, Leonardis, Kristan. *Visual object tracking performance measures revisited*, IEEETIP 2016
[2]Kristan et al., A Novel Performance Evaluation Methodology for Single-Target Trackers, IEEETPAMI 2016

The VOT2017 main challenges

- VOT2017 baseline challenge
  - Experiment with resets upon drifts (VOT standard)
  - Experiments without resets (OTB standard)

- New challenge: VOT2017-realtime challenge
  - Since 2015 efforts made towards stressing the importance of speed
  - Introduction of EFO measure
  - But speed does not directly translate to performance in realtime scenario

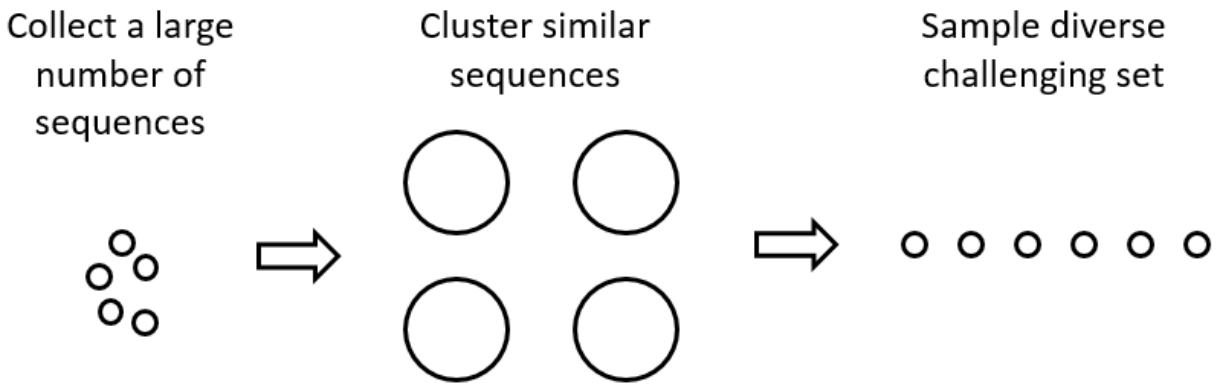
VOT2017-realtime challenge (NEW)



- Tracker required to produce the output faster than frame-rate
- The VOT reset-based methodology used

VOT-TIR2017 challenge

- Task: tracking with thermal videos
- VOT-TIR2016 has not been saturated, therefore in 2017 re-opened
- Follows VOT2013 selection and annotation approach:
  - Small (25 sequences) but diverse dataset according to VOT methodology



  - Derived from Linköping Thermal InfraRed (LTIR) dataset[1]
  - Annotation according to VOT methodology (illumination change becomes dynamics)

[1]A. Berg, J. Ahlberg, M. Felsberg, *A Thermal Object Tracking Benchmark*. AVSS 2015.

VOT2017 participation + winner requirements

- Different parameters don't make a different tracker
- Realtime experiment – running on a single machine

Results paper coauthorship:

- Sufficiently well performing tracker (NCC)
- Agree to publish the code online at VOT2017 page

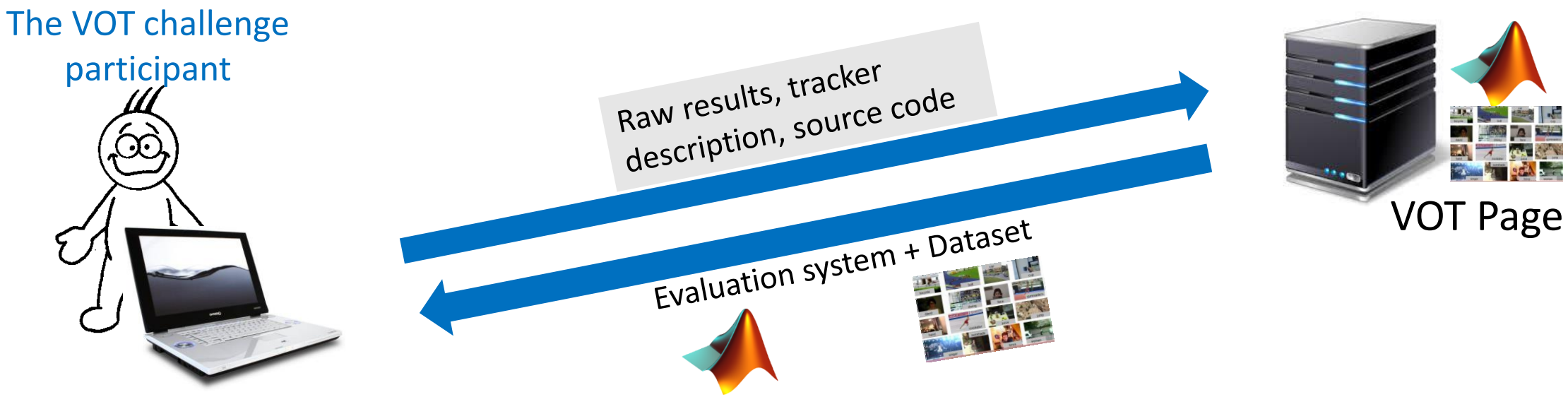
Additional constraint for winners:

- Best-performing tracker *not submitted by the organizers*
- Training on tracking benchmarks prohibited
- VOT2017: Top performance on VOT2017-sequestered dataset
- VOT-TIR2016: Significant improvement compared to 2016

The VOT 2017 workshop

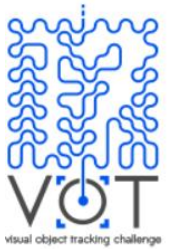
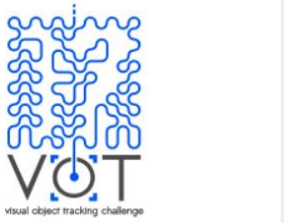
VOT2017 MAIN CHALLENGE PARTICIPATION AND SUBMITTED TRACKERS

VOT2017 participation



- Participants download the toolkit and the VOT2017 dataset
- Toolkit automatically performs all experiments
- Submission of raw results + tracker code required
- Top ten trackers re-run by VOT committee on the sequestered dataset

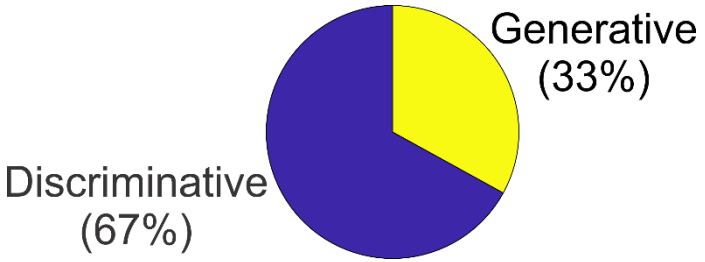
Evolution of VOT submitted trackers

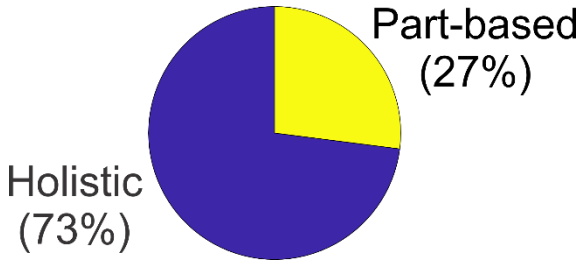
	2013	2014	2015	2016	2017
# Submissions	27	38	62	70	51
Submitted trackers design types	Many diverse trackers submitted	8 Discriminative (sSVM, DCF) 11 Generative 6 Part-based [many diverse]	16 DCF [many diverse] 3 CNN-based	14 CNN-based 27 use DCF [many diverse]	
Top performing	**Discriminative** **Generative** Part-based Holistic	**3 DCF**, Holistic 1 Part-based	**2 CNN-based** 1 sSVM-based 1 Part-based	**CF+CNN** **CNN**, **DCF**	

Kristan et al., "The Visual Object Tracking VOT2013 challenge results," ICCV Workshops 2013
Kristan et al., "The Visual Object Tracking VOT2014 challenge results," ECCV Workshops 2014
Kristan et al., "The Visual Object Tracking VOT2015 challenge results", ICCV Workshops 2015
Kristan et al., "The Visual Object Tracking VOT2016 challenge results", ECCV Workshops 2016
Kristan et al., "The Visual Object Tracking VOT2017 challenge results", ICCV Workshops 2017
Kristan et al., "A Novel Performance Evaluation Methodology for Single-Target Trackers", IEEE TPAMI 2016

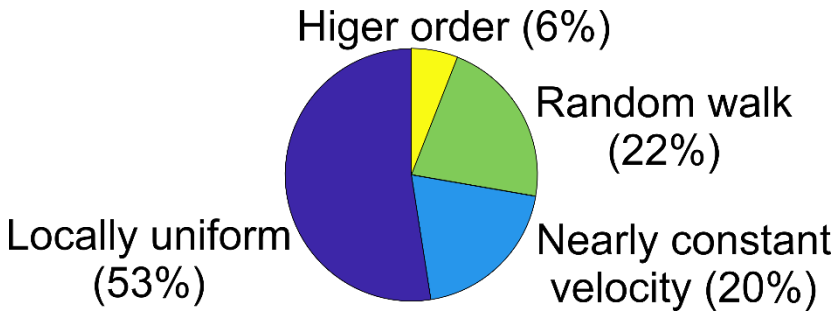
VOT2017: 51 trackers tested

Visual model:





Motion model:



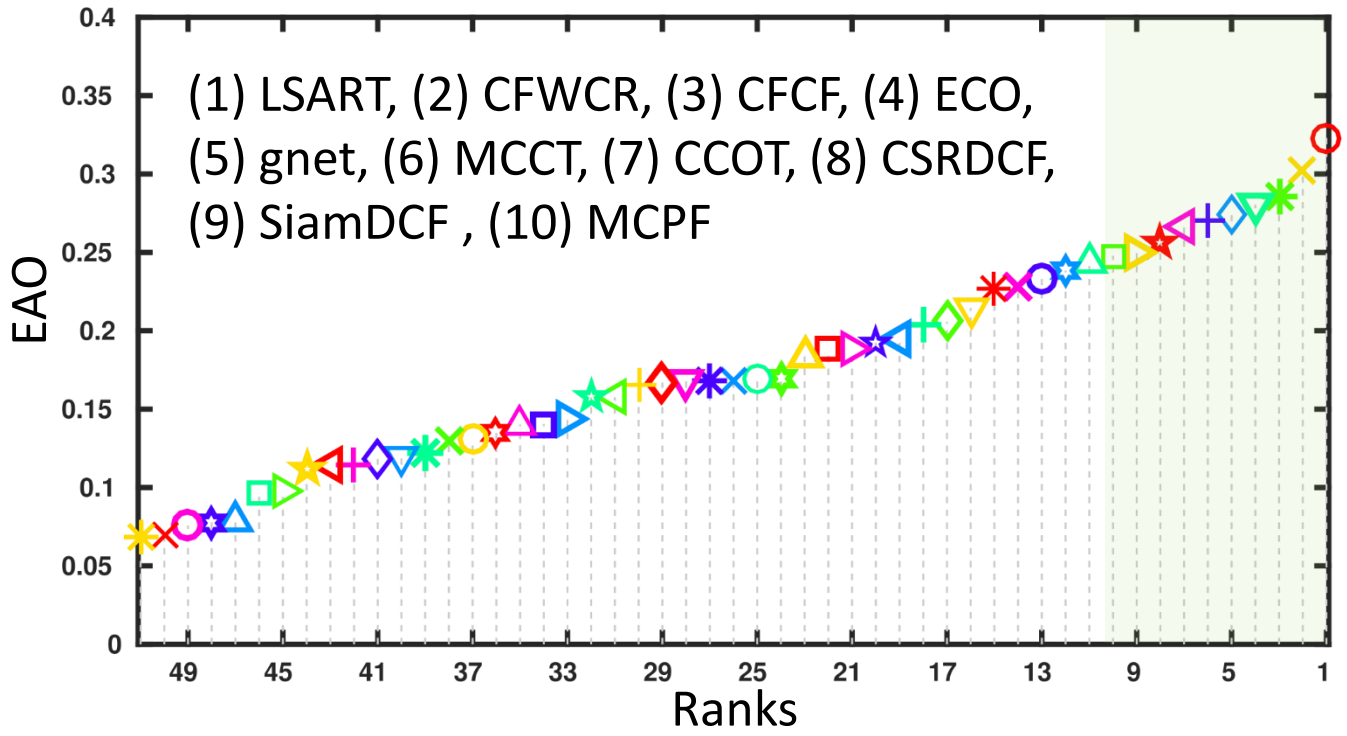
- Discriminative correlation filters (49%)
- Deep convolutional neural nets (31%)
- Mean-shift-based (10%)
- Structured SVM (4%)
- Combination of trackers (3%)

The VOT 2017 workshop

VOT2017 MAIN CHALLENGE RESULTS

VOT2017 results on public dataset

- **Tracking approach:** All apply DCF on top of multi-dim features
- **Features:** All CNN (+handcrafted), except CSRDCF (HoG+Colornames)
  - Most CNN trained for detection, except CFCF and SiamDCF (trained for localization)
- **Design:** most are holistic, except LSART (channels specialize to different parts on target)



VOT2017 results on public dataset

- Top trackers are among the most robust trackers

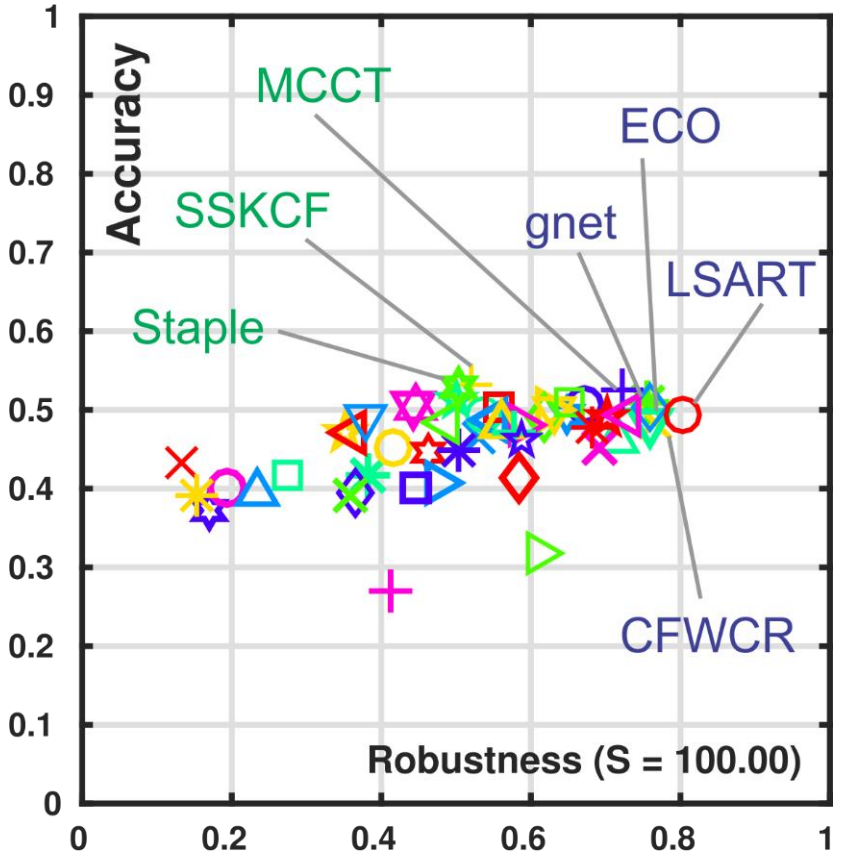
  (1) LSART, (2) CFWCR, (3) ECO, (4) gnet

- Top in accuracy:

  (1) SSKCF, (2) Staple, (3) MCCT

- Per-attribute analysis:

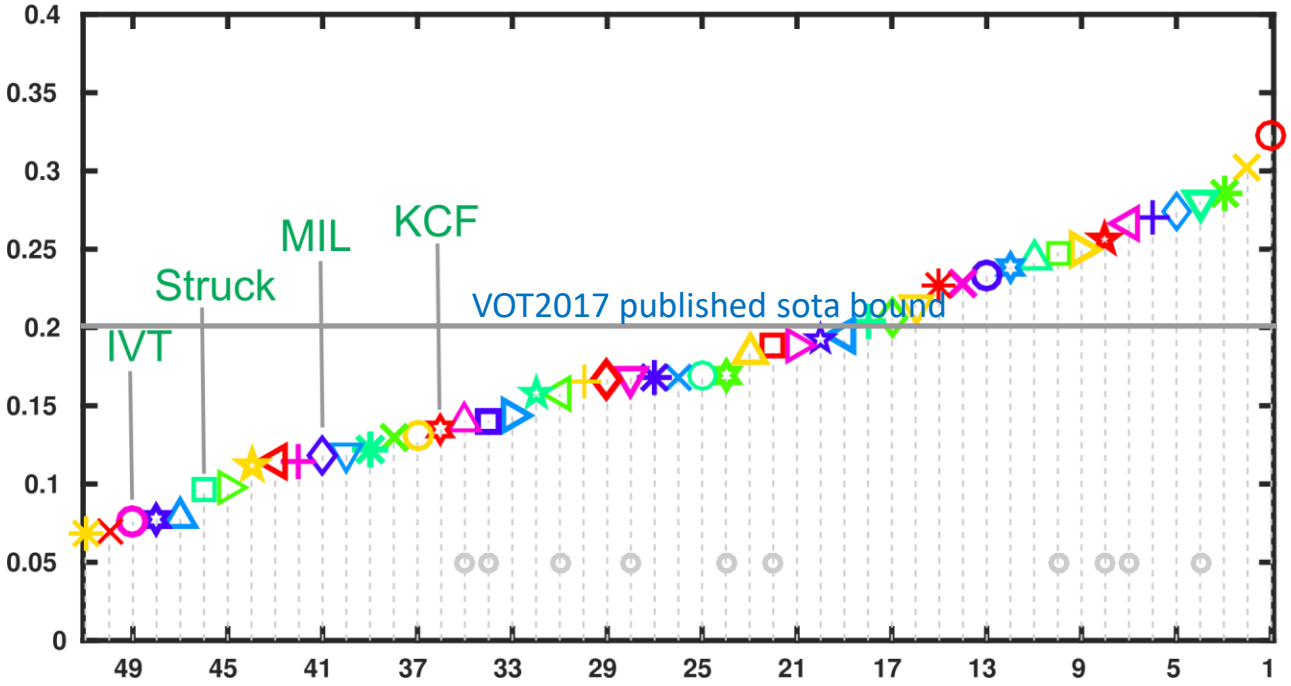
	cam. mot.	ill. ch.	mot. ch.	occl.	scal. ch.
Accuracy	0.48	0.46	0.45 ③	0.39 ①	0.41 ②
Robustness	0.84	1.16 ②	0.97 ③	1.19 ①	0.69



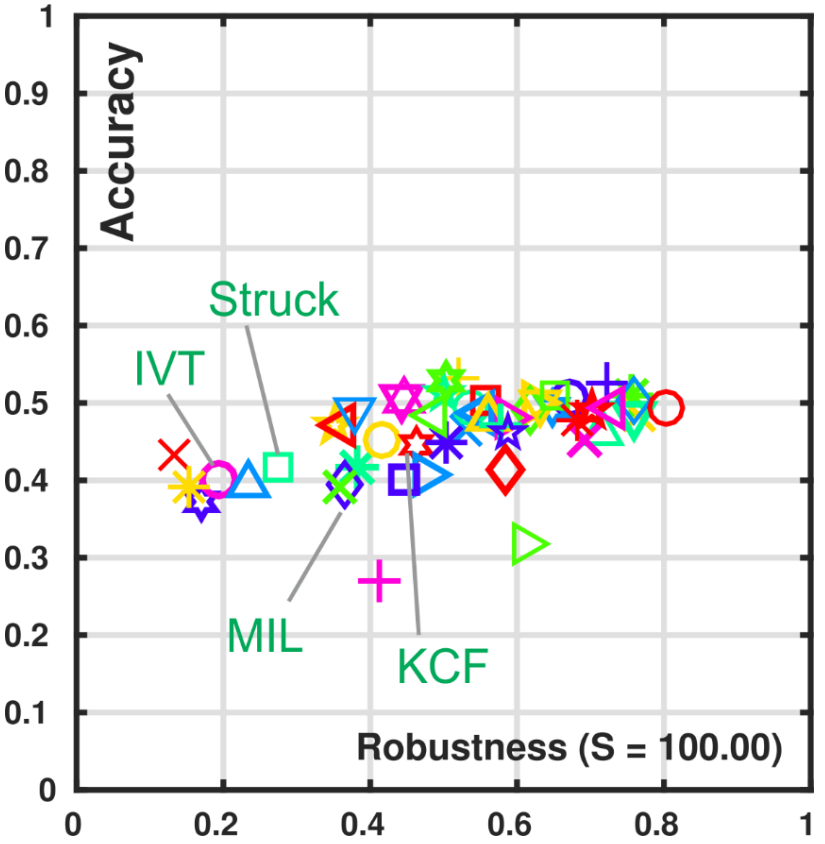
- Most failures due to: Occlusion
- Mostly affects accuracy: Occlusion + Scale change

VOT2017 results on public dataset

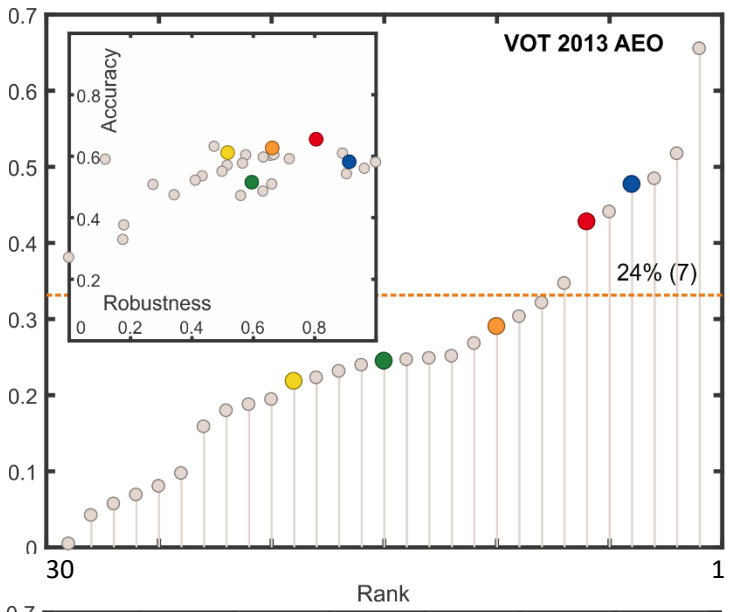
- Baselines ranked at the very tail of the benchmark
- 10 trackers published at major CV venues (2016>)
  - Their average performance: VOT2017 sota bound
  - Over 35% submissions exceed this bound

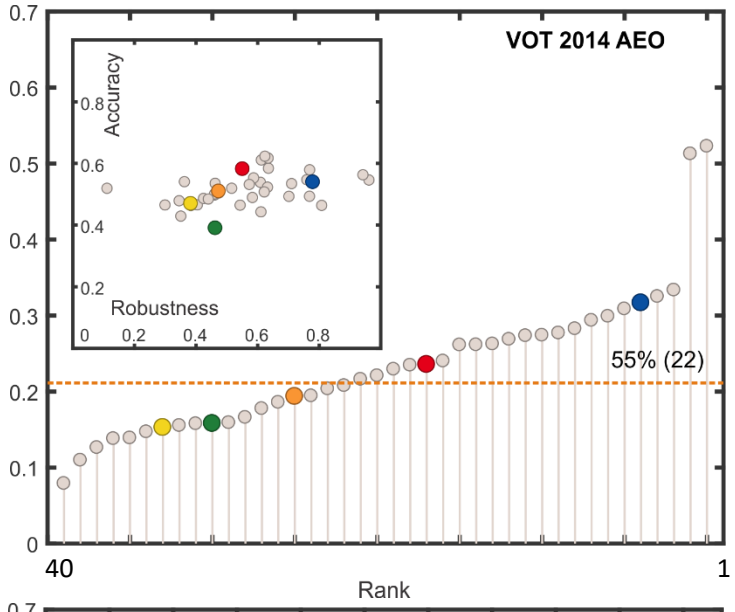


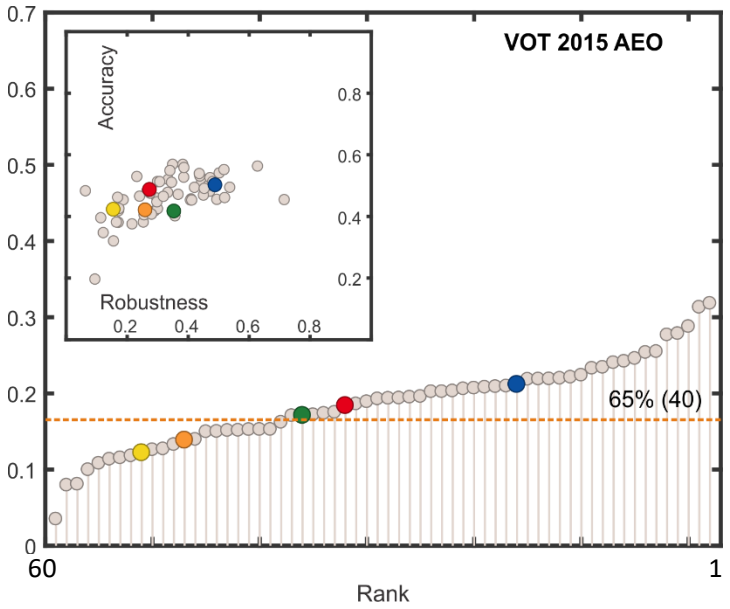
Most submissions are highly advanced trackers!

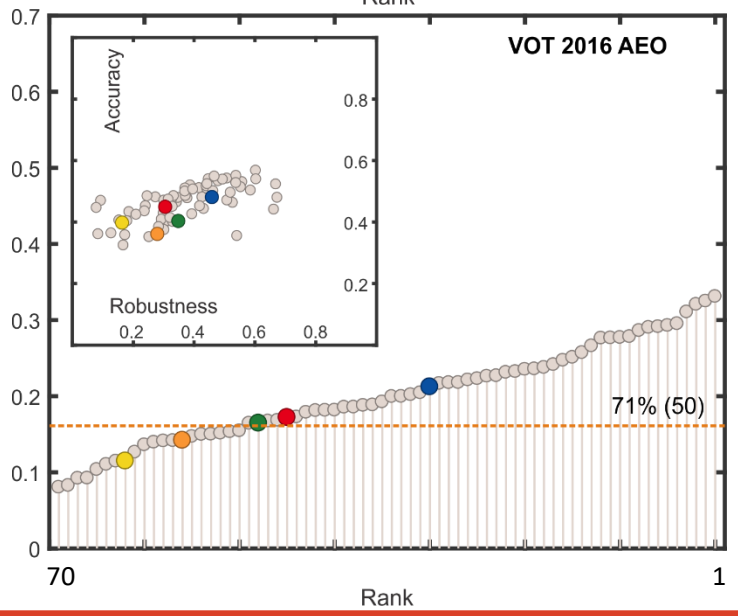


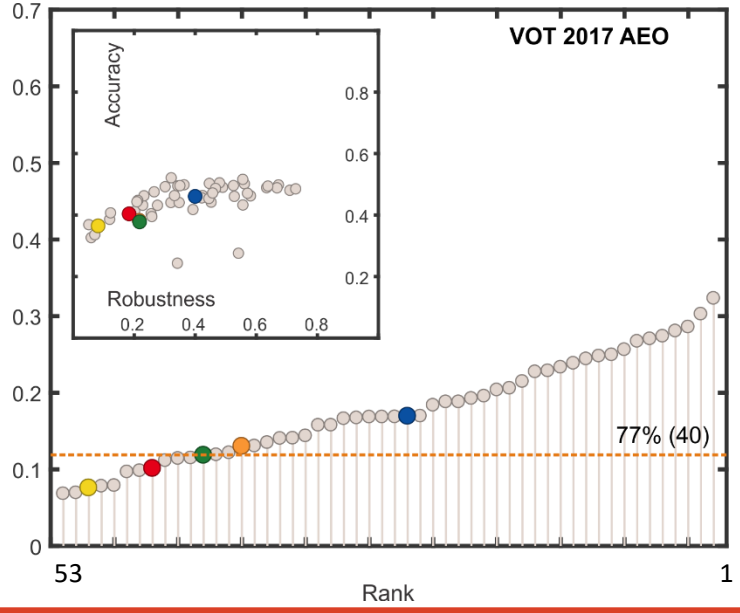
VOT submitted trackers quality over the years











The submitted trackers are:

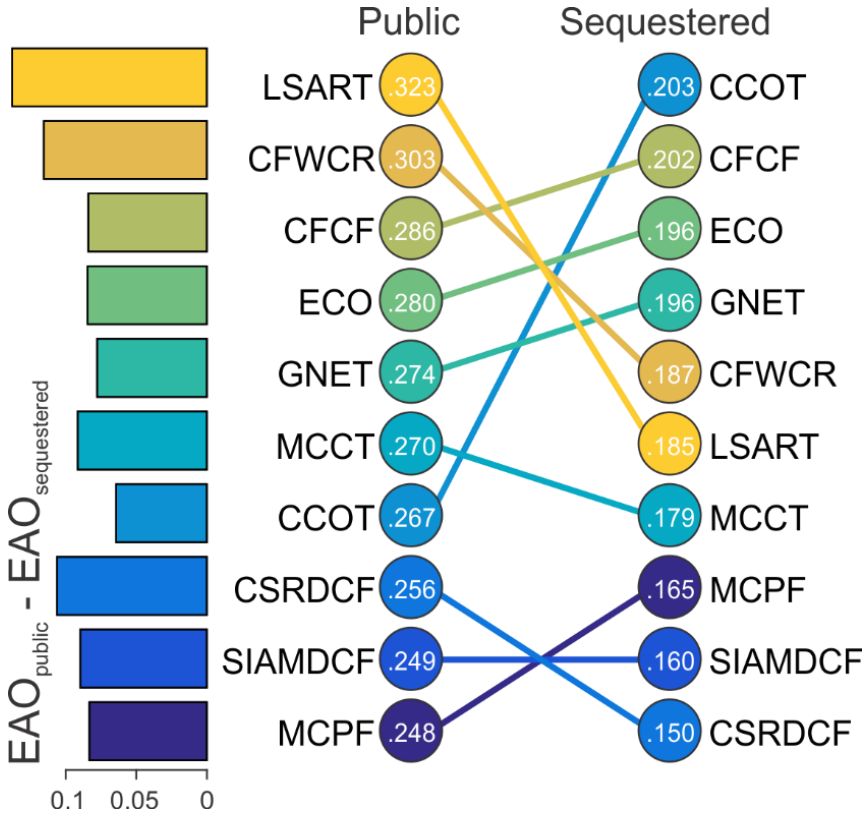
... getting more accurate

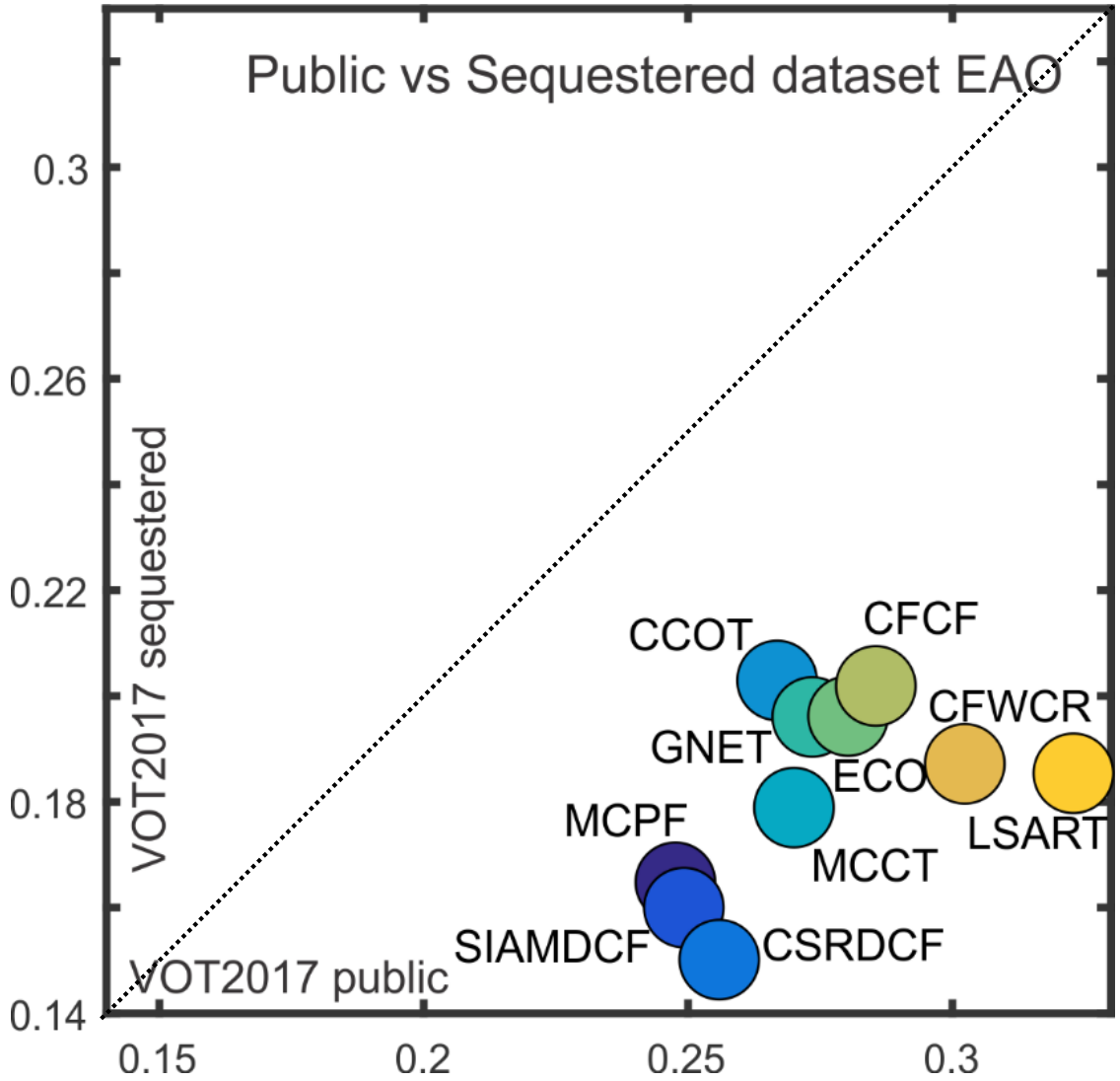
... getting more robust

... increasingly exceeding baselines

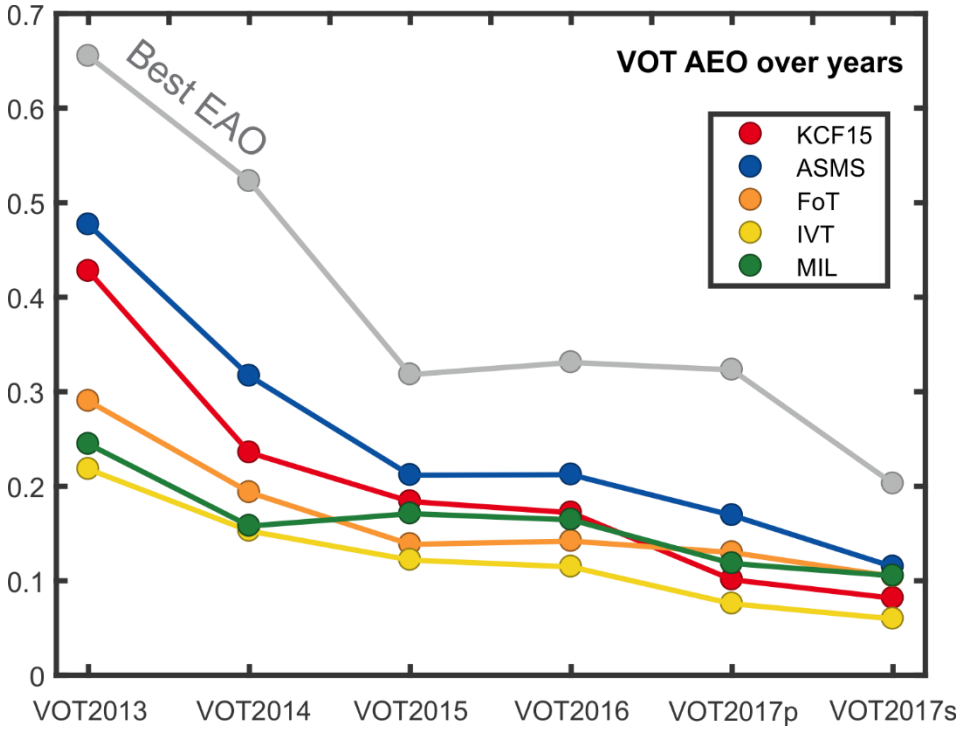
VOT2017 results on sequestered dataset

- Fairly stable ranks
- Greatest change: LSART
- Smallest change: CCOT





The VOT datasets tracking difficulty



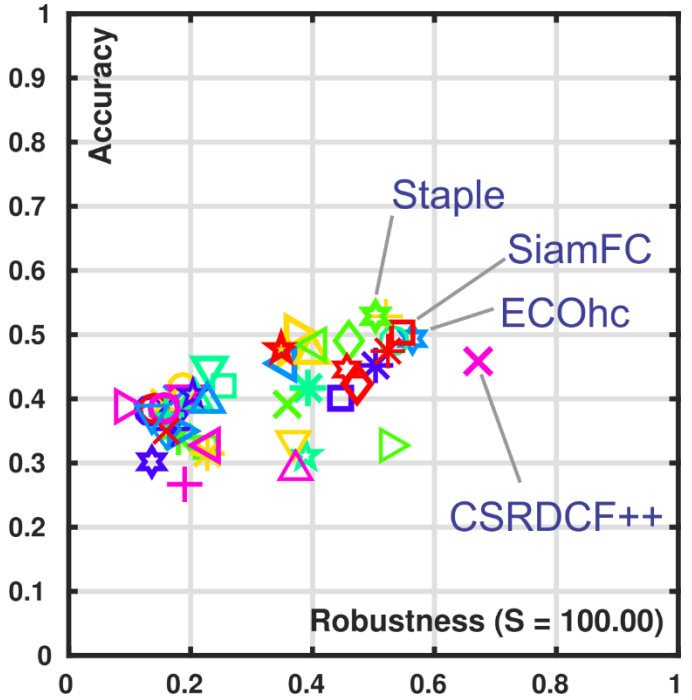
Dataset increasingly more challenging
Top performance stabilized on public dataset

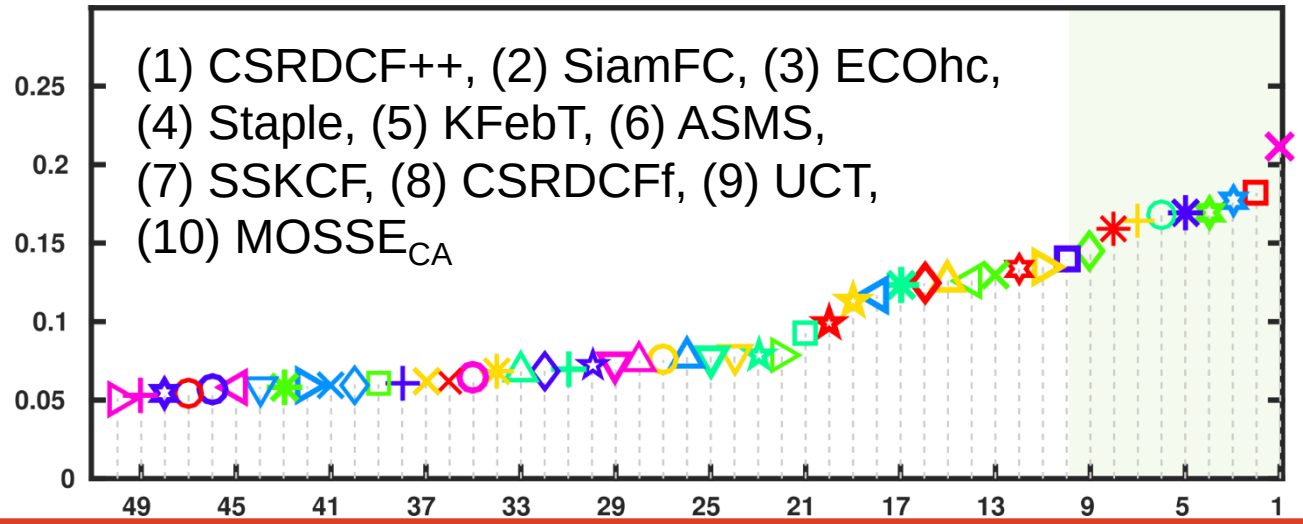
The VOT 2017 workshop

VOT-REALTIME RESULTS

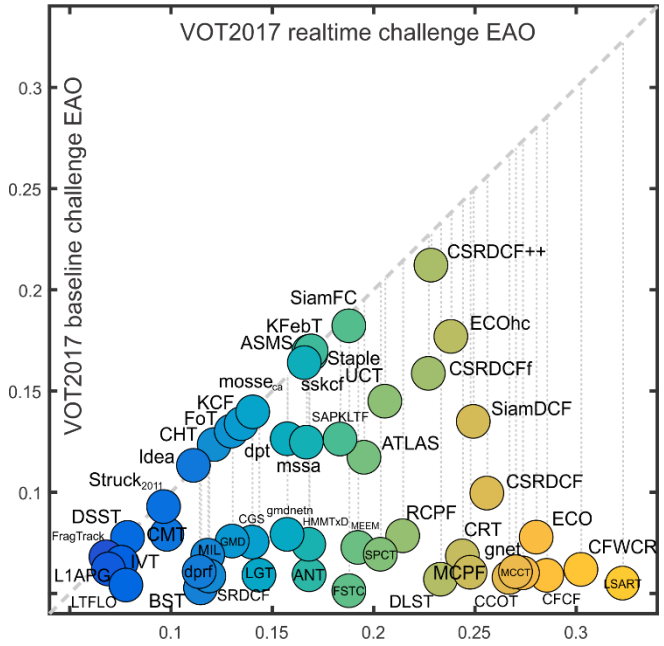
VOT2017-realtime results on public dataset

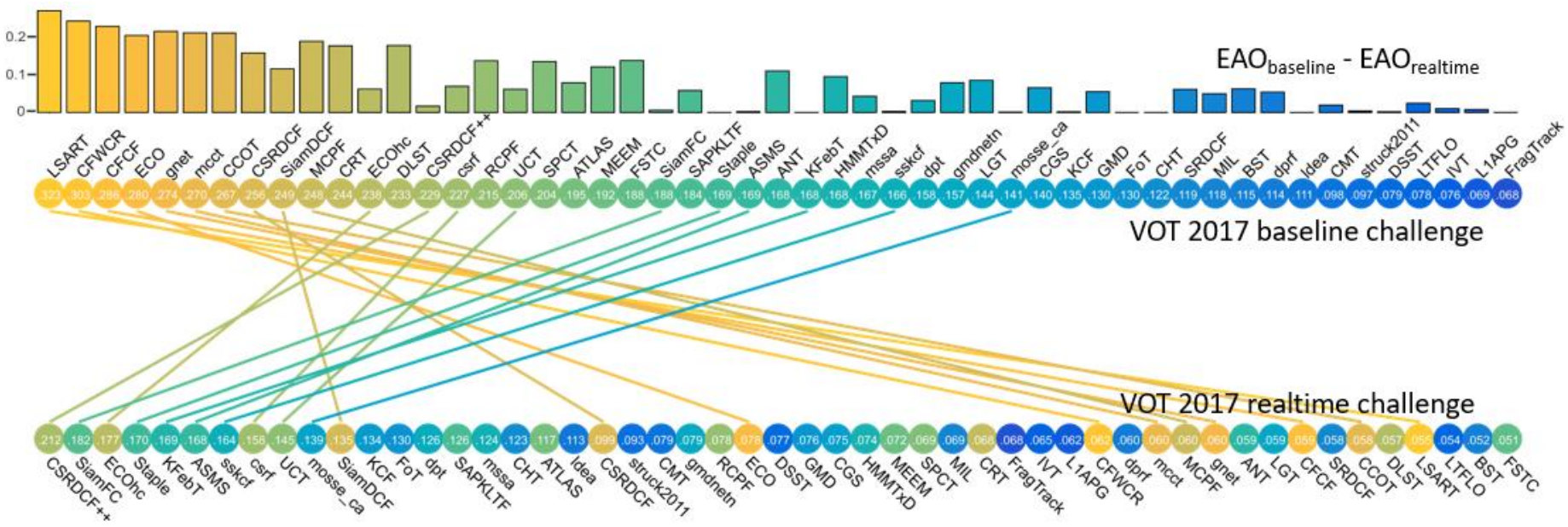
- **Tracking approach:** All apply DCF except from ASMS (scale-adaptive MS), SiamFC and UCT (CNN correlation)
- **Features:** Handcrafted, some apply CNN
  - SiamFC pretrained for template correlation
  - SiamDCF: pretrained for CF localization
  - UCT: CNN trained online
- Implementation/hardware:
  - Most: Matlab, C++ feature extraction
  - CSRDCF++: C++/CPU
  - ECOhc: Matlab, C++/CPU
  - SiamFC: Matlab, C++/GPU





Realtime vs Baseline experiment





- 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 2017 workshop

VOT-TIR CHALLENGE AND RESULTS

(Main organizer: Michael Felsberg)

VOT-TIR2017: 10 trackers tested

VOT-TIR2016 re-opened:

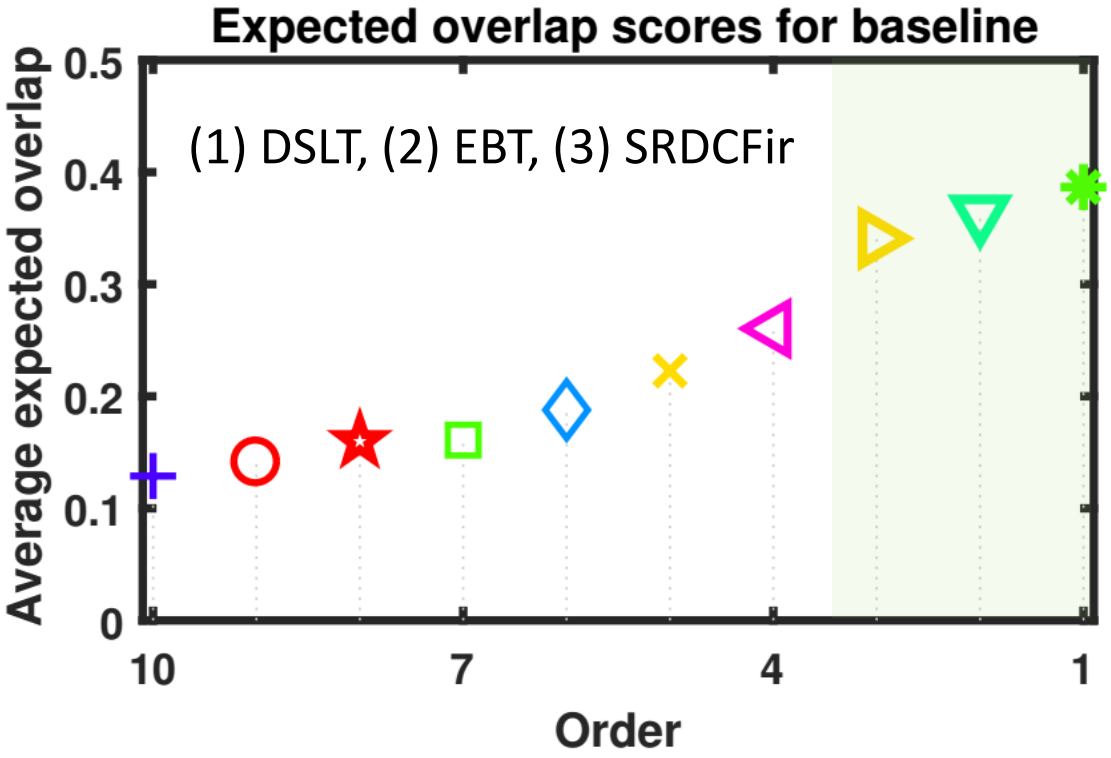
- Winner VOT-TIR2016 (EBT) not significantly better than top-performer 2015 (SRDCFir)
- Both added by the VOT-TIR2017 committee
- Also by committee: ECO

7+1 new submissions:

- Discriminative correlation filters (4)
- Deep convolutional neural nets (2)
- Mean-shift-based (1)
- (Structured) SVM (2)
- Combination of trackers (2)

VOT-TIR2017 results

- **Requirement:** Winner needs to perform better than EBT/SRDCFir
- **Tracking approach:** two main paradigms
  - EBT: object proposals (SVM)
  - SRDCF: DCF-based
  - DSLT: Struck-based (SVM)
- **Features:**
  - EBT: edges
  - SRDCFir: HOG, intensity, motion mask
  - DSLT: HOG and optical flow



VOT-TIR2017 results

- Top trackers are among the most robust trackers

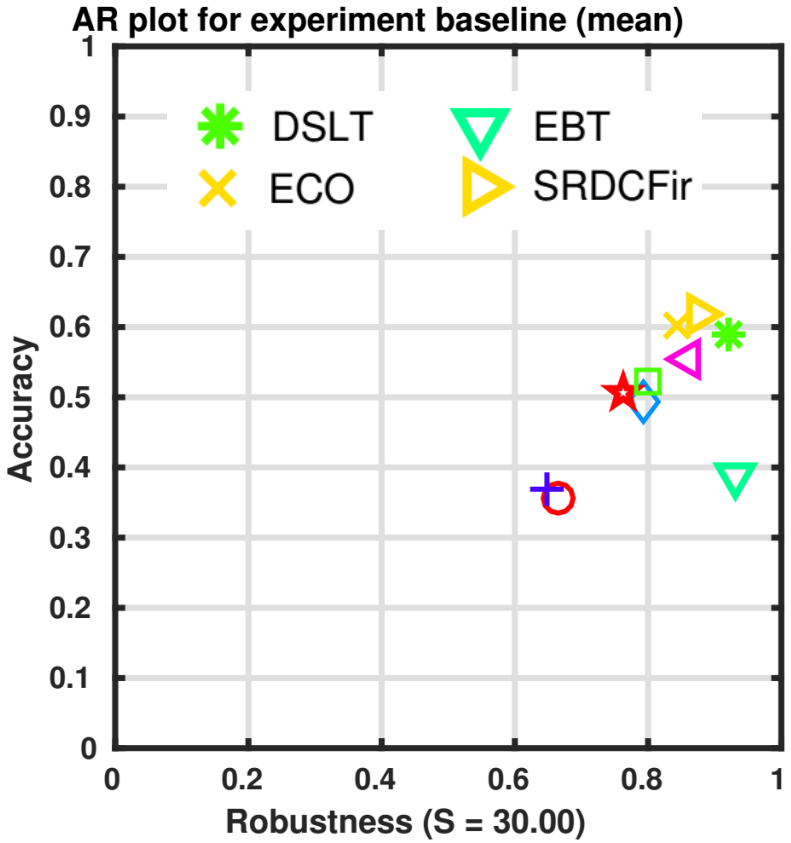
  (1) EBT, (2) DSLT, (3) SRDCFir

- Top in accuracy:

  (1) SRDCFir, (2) ECO, (3) DSLT

- Per-attribute analysis:

	cam. mot.	dyn. ch.	empty	mot. ch.	occl.	size ch.
Accuracy	0.51 ③	0.49 ①	0.54	0.49 ②	0.59	0.52
Robustness	0.72 ③	1.23 ①	0.82 ②	0.31	0.58	0.42



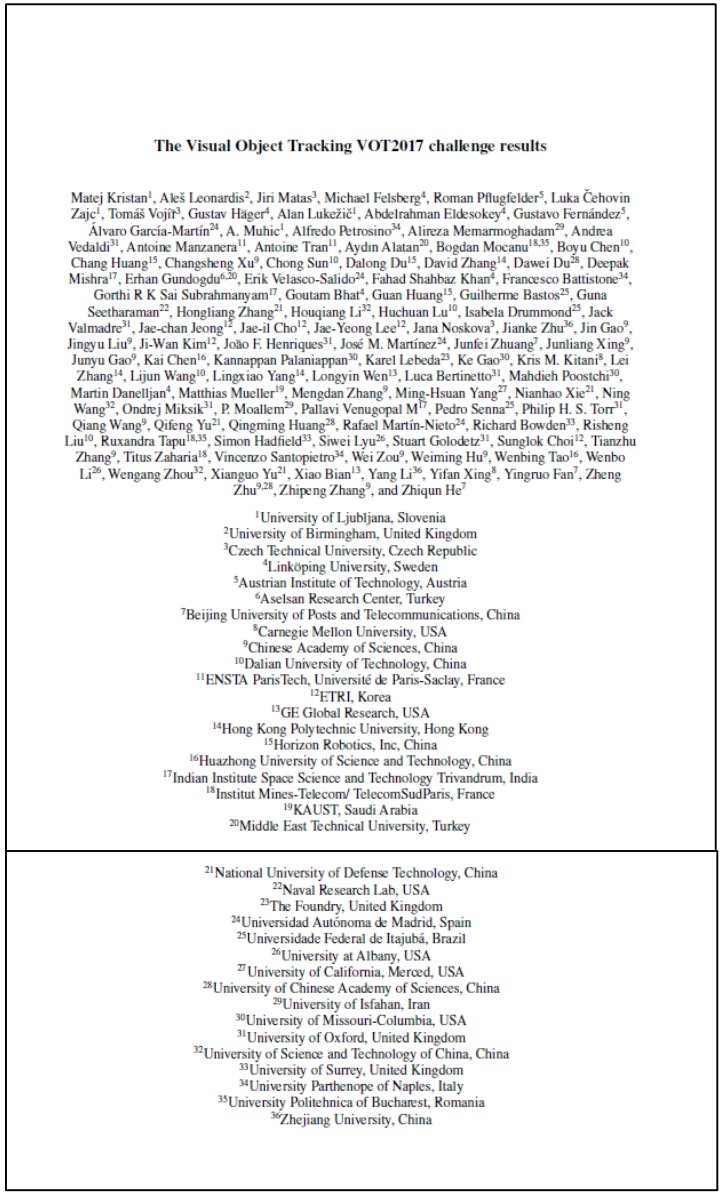
- Most failures due to: Dynamics change
- Mostly affects accuracy: Dynamics change

VOT2017 challenges Summary

- VOT2017 baseline:
  - Emergence of deep features trained for localization (seem to perform very well)
  - Correlation approaches (wide sense) perform best
  - Sequestered dataset more challenging than public
- VOT2017-realtime:
  - Best performance: fully convolutional deep approaches and discriminative CFs
  - Some of the fastest trackers also rank among best on baseline
- VOT-TIR2017:
  - Deep features are not (yet) significant for achieving sota performance
  - Motion features seem to be essential for robust tracking
  - Best performance by SVM-based approaches

VOT2017 online resources

- Results in 104 coauthor paper
- Available at http://www.votchallenge.net/vot2017
- Presentations + papers + Dataset + Evaluation kit
- Guidelines on how to evaluate your trackers on VOT2017 and produce graphs for your papers
  (directly comparable to 51 trackers!)
- VOT is open source:
  - All toolkits and protocols on Github
  - VOT2017: ~36 tracker source code
  - VOT2016: 39 tracker source code



The Visual Object Tracking VOT2017 challenge results

Matej Kristan[1], Aleš Leonardis[2], Jiri Matas[3], Michael Felsberg[4], Roman Pflugfelder[5], Luka Čehovin Zajc[1], Tomáš Vojíř[3], Gustav Häger[4], Alan Lukežič[1], Abdelrahman Eldesokey[4], Gustavo Fernández[5], Álvaro García-Martín[24], A. Muhic[1], Alfredo Petrosino[34], Alireza Memarmoghadam[29], Andrea Vedaldi[31], Antoine Manzanera[11], Antoine Tran[11], Aydın Alatan[20], Bogdan Mocanu[18,35], Boyu Chen[10], Chang Huang[15], Changsheng Xu[9], Chong Sun[10], Dalong Du[15], David Zhang[14], Dawei Du[28], Deepak Mishra[17], Erhan Gundogdu[6,20], Erik Velasco-Salido[24], Fahad Shahbaz Khan[4], Francesco Battistone[34], Gorthi R K Sai Subrahmanyam[17], Goutam Bhat[4], Guan Huang[15], Guilherme Bastos[25], Guna Seetharaman[22], Hongliang Zhang[21], Houqiang Li[32], Huchuan Lu[10], Isabela Drummond[25], Jack Valmadre[31], Jae-chan Jeong[12], Jae-il Cho[12], Jae-Yeong Lee[12], Jana Noskova[3], Jianke Zhu[36], Jin Gao[9], Jingyu Liu[9], Ji-Wan Kim[12], João F. Henriques[31], José M. Martínez[24], Junfei Zhuang[7], Junliang Xing[9], Junyu Gao[9], Kai Chen[16], Kannappan Palaniappan[30], Karel Lebeda[23], Ke Gao[30], Kris M. Kitani[8], Lei Zhang[14], Lijun Wang[10], Lingxiao Yang[14], Longyin Wen[13], Luca Bertinetto[31], Mahdieh Poostchi[30], Martin Danelljan[4], Matthias Mueller[19], Mengdan Zhang[9], Ming-Hsuan Yang[27], Nianhao Xie[21], Ning Wang[32], Ondrej Miksik[31], P. Moallem[29], Pallavi Venugopal M[17], Pedro Senna[25], Philip H. S. Torr[31], Qiang Wang[9], Qifeng Yu[21], Qingming Huang[28], Rafael Martín-Nieto[24], Richard Bowden[33], Risheng Liu[10], Ruxandra Tapu[18,35], Simon Hadfield[33], Siwei Lyu[26], Stuart Golodetz[31], Sunglok Choi[12], Tianzhu Zhang[9], Titus Zaharia[18], Vincenzo Santopietro[34], Wei Zou[9], Weiming Hu[9], Wenbing Tao[16], Wenbo Li[26], Wengang Zhou[32], Xianguo Yu[21], Xiao Bian[13], Yang Li[36], Yifan Xing[8], Yingruo Fan[7], Zheng Zhu[9,28], Zhipeng Zhang[9], and Zhiqun He[7]

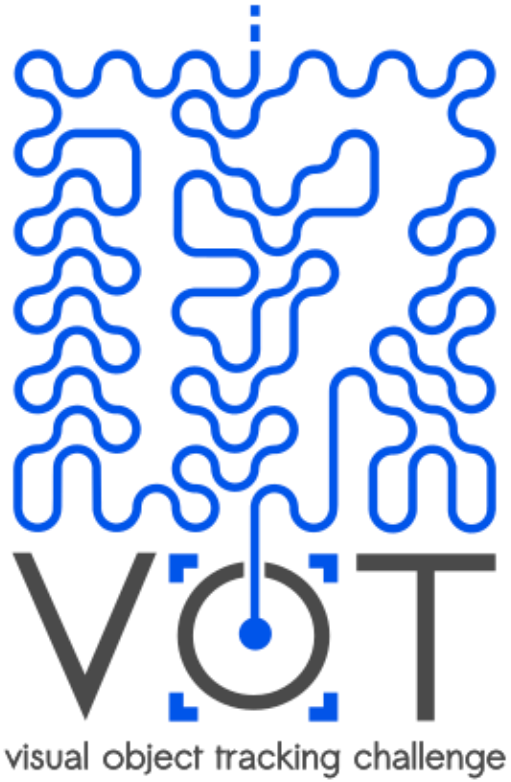
[1]University of Ljubljana, Slovenia
[2]University of Birmingham, United Kingdom
[3]Czech Technical University, Czech Republic
[4]Linköping University, Sweden
[5]Austrian Institute of Technology, Austria
[6]Aselsan Research Center, Turkey
[7]Beijing University of Posts and Telecommunications, China
[8]Carnegie Mellon University, USA
[9]Chinese Academy of Sciences, China
[10]Dalian University of Technology, China
[11]ENSTA ParisTech, Université de Paris-Saclay, France
[12]ETRI, Korea
[13]GE Global Research, USA
[14]Hong Kong Polytechnic University, Hong Kong
[15]Horizon Robotics, Inc, China
[16]Huazhong University of Science and Technology, China
[17]Indian Institute Space Science and Technology Trivandrum, India
[18]Institut Mines-Telecom/ TelecomSudParis, France
[19]KAUST, Saudi Arabia
[20]Middle East Technical University, Turkey
[21]National University of Defense Technology, China
[22]Naval Research Lab, USA
[23]The Foundry, United Kingdom
[24]Universidad Autónoma de Madrid, Spain
[25]Universidade Federal de Itajubá, Brazil
[26]University at Albany, USA
[27]University of California, Merced, USA
[28]University of Chinese Academy of Sciences, China
[29]University of Isfahan, Iran
[30]University of Missouri-Columbia, USA
[31]University of Oxford, United Kingdom
[32]University of Science and Technology of China, China
[33]University of Surrey, United Kingdom
[34]University Parthenope of Naples, Italy
[35]University Politehnica of Bucharest, Romania
[36]Zhejiang University, China

VOT2017 awards:

Winners of the VOT2017 main challenge:

CFCF by: Erhan Gundogdu, Aydin Alatan

“Convolutional Features for Correlation Filters”
Presentation next



University *of Ljubljana*
Faculty *of Computer and Information Science*

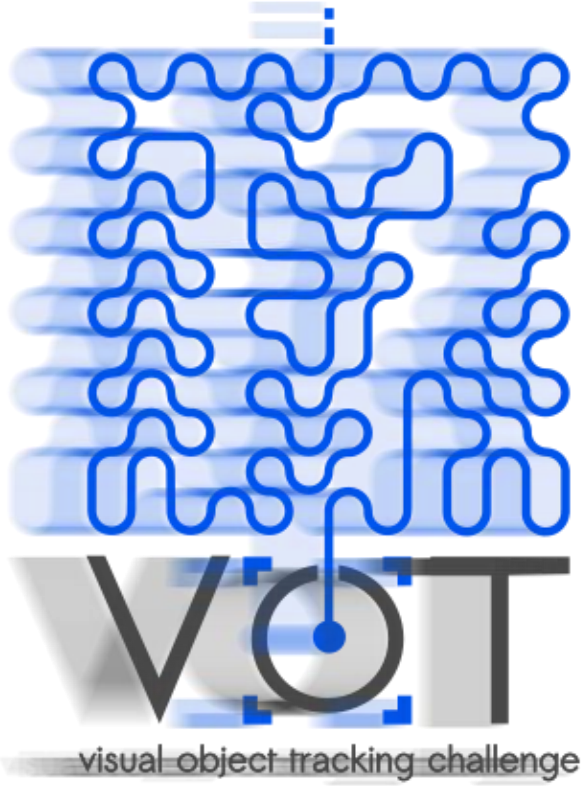
SICK
Sensor Intelligence.

VOT2017 awards:

Winners of the VOT2017 realtime challenge:

SiamFC by: Luca Bertinetto, Joao Henriques, Jack Valmadre, Andrea Vedaldi, Philip Torr

“Fully Convolutional Siamese Network”



University *of Ljubljana*
Faculty *of Computer and Information Science*

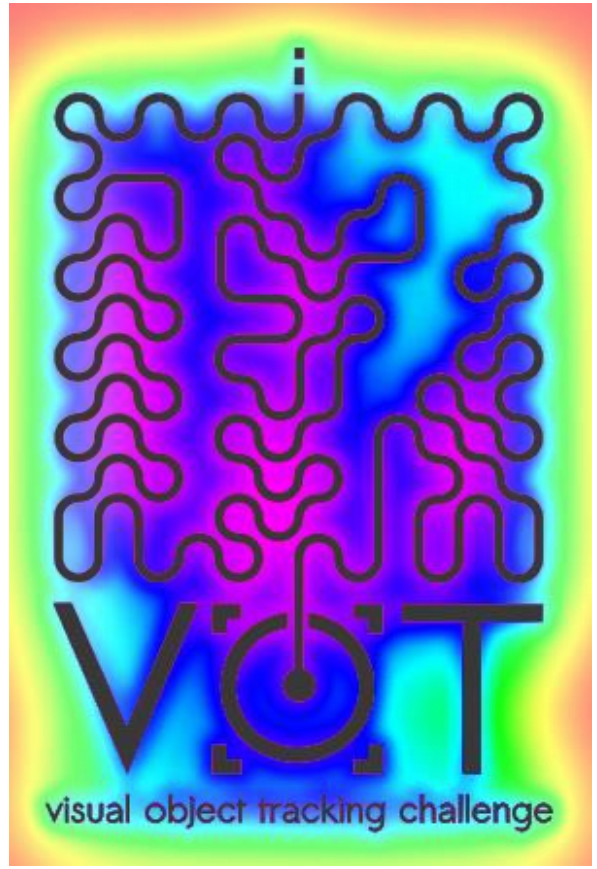
SICK
Sensor Intelligence.

VOT2017 awards:

Winners of the VOT-TIR2017 challenge:

DSLT by: Xianguo Yu, Qifeng Yu, Hongliang Zhang, Nianhao Xie

“Dense Structural Learning based Tracker”



University of Ljubljana
Faculty of Computer and Information Science

SICK
Sensor Intelligence.

Thanks

- The VOT2017 committee


M. Kristan


J. Matas


A. Leonardis


M. Felsberg


R. Pflugfelder


G. Fernandez


L. Čehovin


T. Vojir


A. Lukežič


G. Häger


A. Eldesokey

- Everyone who participated or contributed

Abdelrahman Eldesokey4, A´ lvaro Garc´ıa-Mart´ın24, A. Muhic1, Alfredo Petrosino34, Alireza Memarmoghadam29, Andrea Vedaldi31, Antoine Manzanera11, Antoine Tran11, Aydın Alatan20, Bogdan Mocanu18,35, Boyu Chen10, Chang Huang15, Changsheng Xu9, Chong Sun10, Dalong Du15, David Zhang14, Dawei Du28, Deepak Mishra17, Erhan Gundogdu6,20, Erik Velasco-Salido24, Fahad Shahbaz Khan4, Francesco Battistone34, Gorthi R K Sai Subrahmanyam17, Goutam Bhat4, Guan Huang15, Guilherme Bastos25, Guna Seetharaman22, Hongliang Zhang21, Houqiang Li32, Huchuan Lu10, Isabela Drummond25, Jack Valmadre31, Jae-chan Jeong12, Jae-il Cho12, Jae-Yeong Lee12, Jana Noskova3, Jianke Zhu36, Jin Gao9, Jingyu Liu9, Ji-Wan Kim12, Jo˜ao F. Henriques31, Jos´e M. Mart´ınez24, Junfei Zhuang7, Junliang Xing9, Junyu Gao9, Kai Chen16, Kannappan Palaniappan30, Karel Lebeda23, Ke Gao30, Kris M. Kitani8, Lei Zhang14, Lijun Wang10, Lingxiao Yang14, Longyin Wen13, Luca Bertinetto31, Mahdieh Poostchi30, Martin Danelljan4, Matthias Mueller19, Mengdan Zhang9, Ming-Hsuan Yang27, Nianhao Xie21, Ning Wang32, Ondrej Miksik31, P. Moallem29, Pallavi Venugopal M17, Pedro Senna25, Philip H. S. Torr31, Qiang Wang9, Qifeng Yu21, Qingming Huang28, Rafael Mart´ın-Nieto24, Richard Bowden33, Risheng Liu10, Ruxandra Tapu18,35, Simon Hadfield33, Siwei Lyu26, Stuart Golodetz31, Sunglok Choi12, Tianzhu Zhang9, Titus Zaharia18, Vincenzo Santopietro34, Wei Zou9, Weiming Hu9, Wenbing Tao16, Wenbo Li26, Wengang Zhou32, Xianguo Yu21, Xiao Bian13, Yang Li36, Yifan Xing8, Yingruo Fan7, Zheng Zhu9,28, Zhipeng Zhang9, and Zhiqun He7

VOT2017 sponsors:

University *of Ljubljana*
Faculty *of Computer and Information Science*

SICK
Sensor Intelligence.