The 1st Visual Object Tracking Segmentation VOTS2023 Challenge Results

Matej Kristan, Jiri Matas, Martin Danelljan, Michael Felsberg, Hyung Jin Chang, Luka Čehovin Zajc, Alan Lukežič, Ondrej Drbohlav, Zhongqun Zhang, Khanh-Tung Tran, Xuan-Son Vu, Johanna Bjorklund, Christoph Mayer, Yushan Zhang, Lei Ke, Jie Zhao, Gustavo Fernandez

Intro/Motivation

- The VOT formed in 2013 to support general object trackers

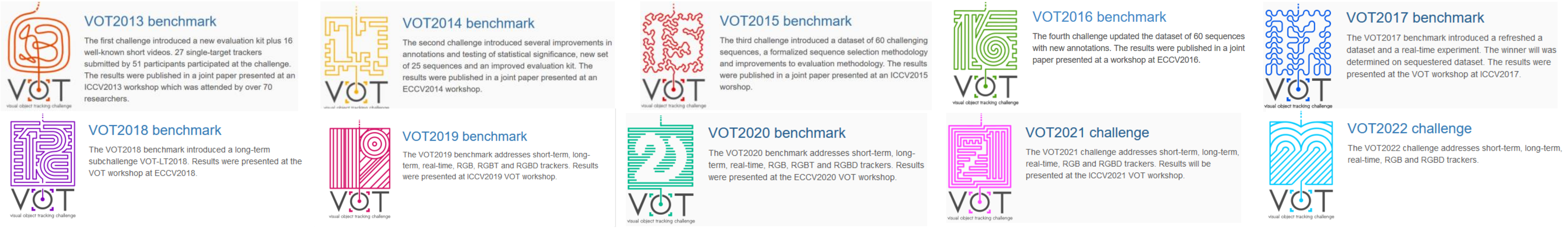
- "*Track any image region (including unknown instances or their parts) given a single training example.*"

- Research questions:
  Representations, Self-supervision, Robust localization...

- To better study these, the VOT challenges were restricted to single-target tracking, separating short-term and long-term tracking

Intro/Motivation

- Ten challenges organized to explore various research questions



VOT2013 benchmark

The first challenge introduced a new evaluation kit plus 16 well-known short videos. 27 single-target trackers submitted by 51 participants participated at the challenge. The results were published in a joint paper presented at an ICCV2013 workshop which was attended by over 70 researchers.

VOT2014 benchmark

The second challenge introduced several improvements in annotations and testing of statistical significance, new set of 25 sequences and an improved evaluation kit. The results were published in a joint paper presented at an ECCV2014 workshop.

VOT2015 benchmark

The third challenge introduced a dataset of 60 challenging sequences, a formalized sequence selection methodology and improvements to evaluation methodology. The results were published in a joint paper presented at an ICCV2015 worshop.

VOT2016 benchmark

The fourth challenge updated the dataset of 60 sequences with new annotations. The results were published in a joint paper presented at a workshop at ECCV2016.

VOT2017 benchmark

The VOT2017 benchmark introduced a refreshed a dataset and a real-time experiment. The winner will was determined on sequestered dataset. The results were presented at the VOT workshop at ICCV2017.

VOT2018 benchmark

The VOT2018 benchmark introduced a long-term subchallenge VOT-LT2018. Results were presented at the VOT workshop at ECCV2018.

VOT2019 benchmark

The VOT2019 benchmark addresses short-term, long-term, real-time, RGB, RGBT and RGBD trackers. Results were presented at ICCV2019 VOT workshop.

VOT2020 benchmark

The VOT2020 benchmark addresses short-term, long-term, real-time, RGB, RGBT and RGBD trackers. Results were presented at the ECCV2020 VOT workshop.

VOT2021 challenge

The VOT2021 challenge addresses short-term, long-term, real-time, RGB and RGBD trackers. Results will be presented at the ICCV2021 VOT workshop.

VOT2022 challenge

The VOT2022 challenge addresses short-term, long-term, real-time, RGB and RGBD trackers.

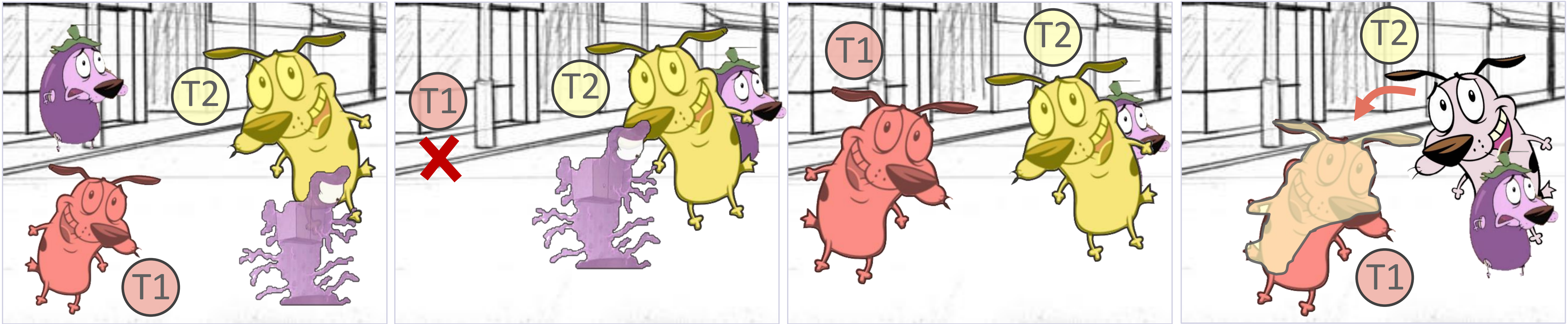
- The field has matured to a point where advancements expected by relaxing the restrictions

Visual object tracking segmentation challenge

VOTS

The VOTS2023 challenge scope

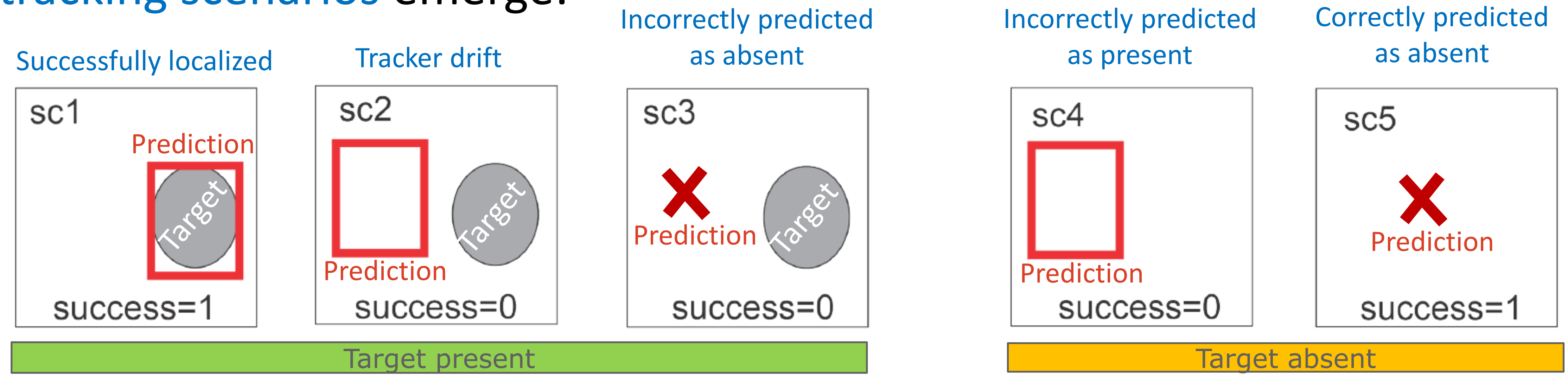
- General object Short/Long-term, Single/Multi-target segmentation trackers
- Initialize on all targets in the first frame and report position in the rest



- Determine the target absence and redetect when it reappears
- Drifting off the target to background or another object is considered failure

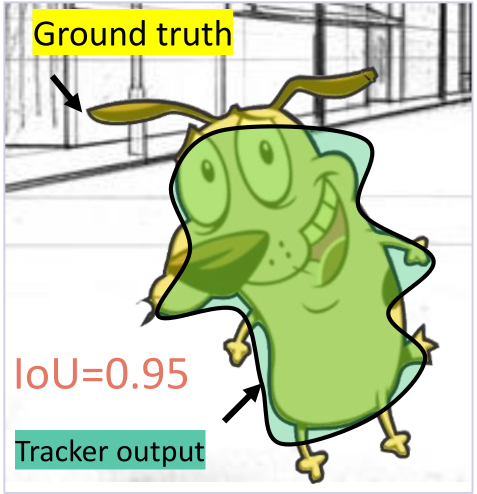
Per-target performance measures

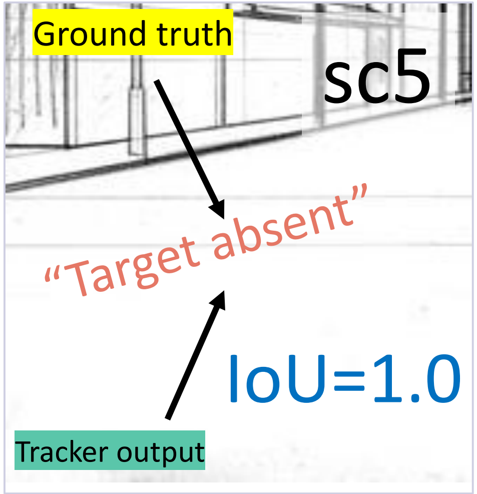
5 tracking scenarios emerge:



- IoU as a standard measure of agreement between prediction and GT
- Require IoU value definition for sc5

$\text{IoU}_{sc5}=1.0$

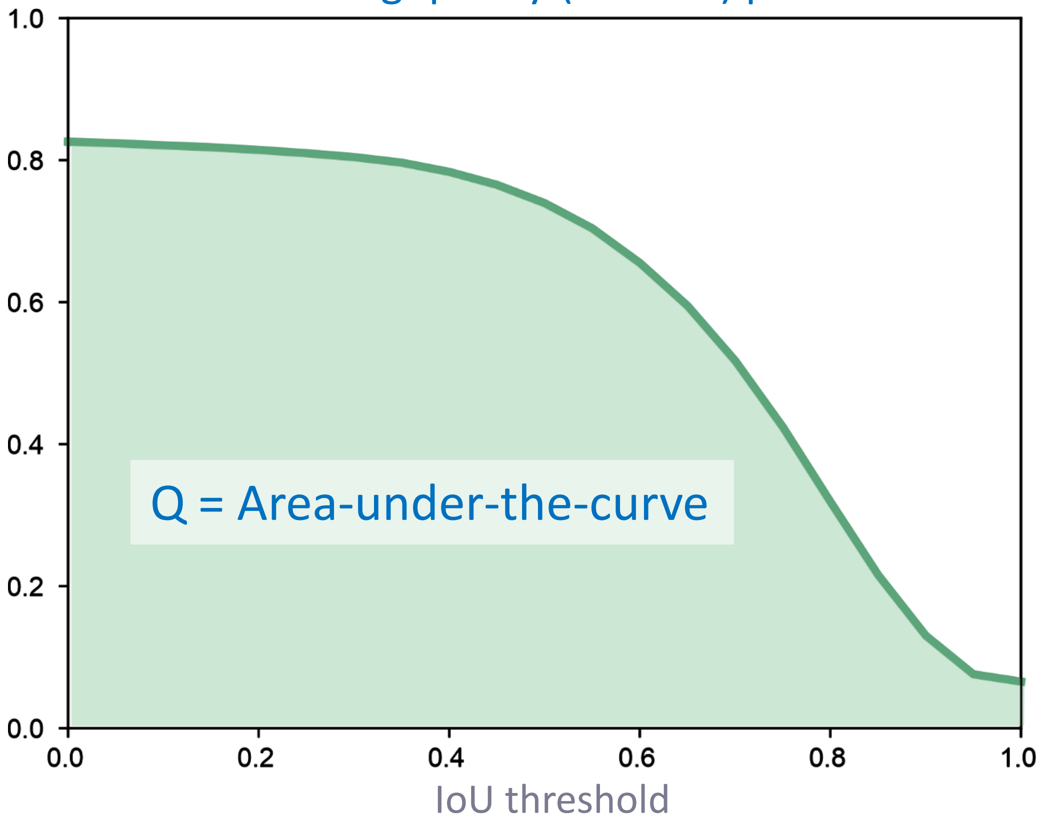




Primary performance measure

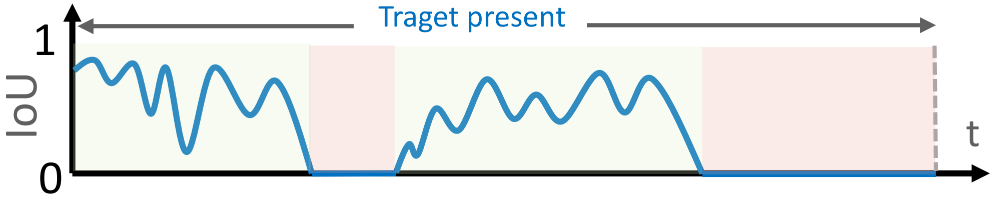
- Performance summarized by the classical success w.r.t IoU plot (i.e., tracking quality plot)
- Success plot calculated individually for each target in each sequence and then averaged
- Primary measure: *Tracking quality Q* (area-under-the-curve)

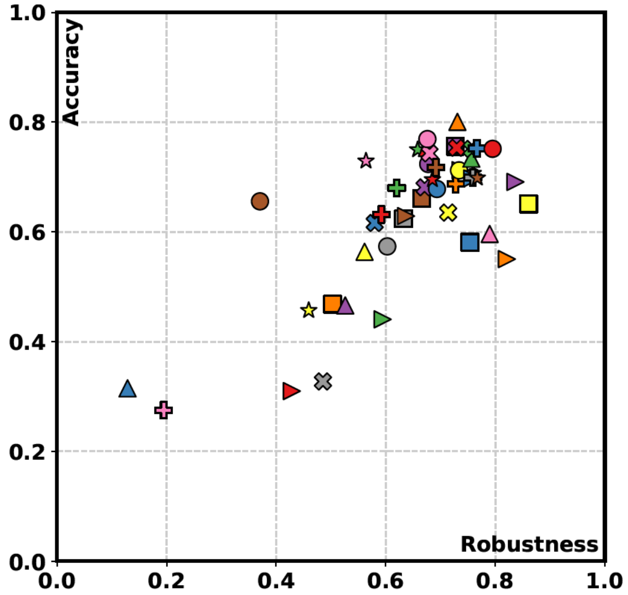
Tracking quality (success) plot



Auxiliary performance measures

- Accuracy/Robustness (@IoU=0.0 when target present)





“Why did the tracker fail while target visible?”

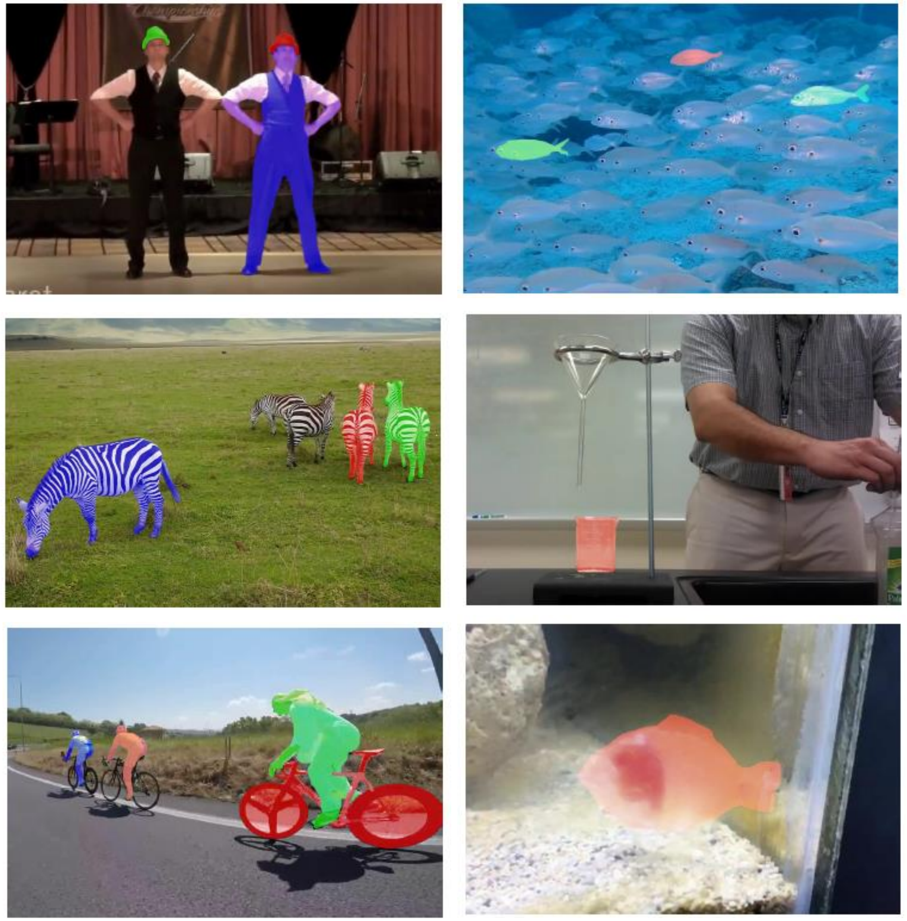
- NotReportedError (NRE): % frames incorrectly predicted target absent
- DriftRateError (DRE): % frames tracker drifted while predicting target present

“How well is target absence determined?”

- AbsenceDetectionQuality (ADQ): % frames target correctly predicted absent

VOTS2023 dataset

- Source: LaGOT[1], UTB180[2], TOTB[3], VOT-LT2021,VOT-LT2022, VOT-ST2022
- Selection criteria:
  - Sequences challenging for modern architectures
  - Properties: (i) visually-similar objects, (ii) substantial appearance changes, (iii) cluttered background, (iv) entering-exiting field-of-view
  - Diverse object and scene types (Air, Ground, Underwater)
  - Opaque as well as transparent objects
- Annotation: Segmentation masks
  - Include parts of objects as targets



[1] Mayer et al. ArXiv 2023; [2]Alawode et al. ACCV2022; [3]Fan et al. ICCV2021

VOTS2023 dataset

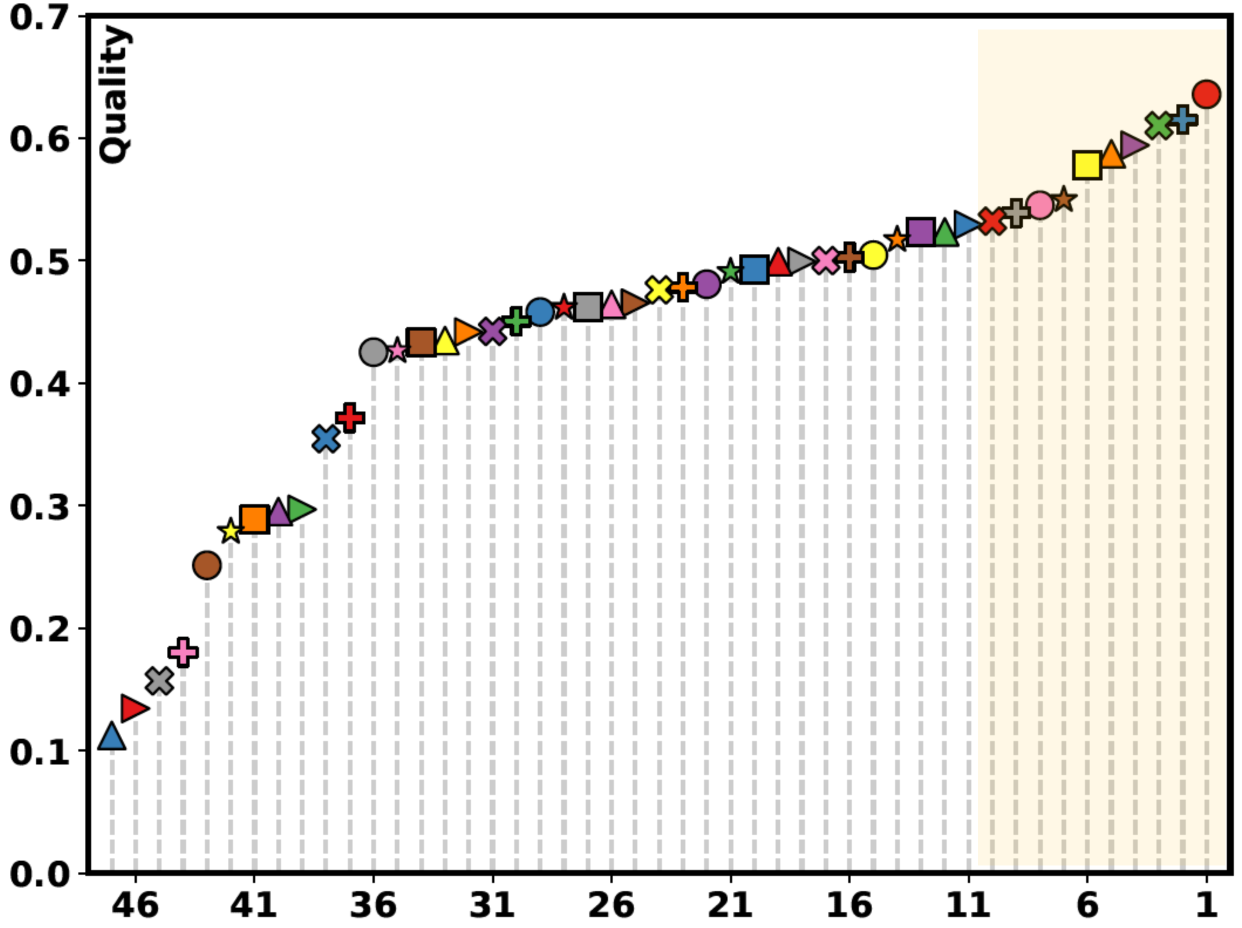
- Stats: 144 sequences ; 341 targets ; 168 targets leave the FOV at least once
- Sequence properties:
  - min/max = 63/10.7k frames
  - On average 2.37 targets per sequence annotated
  - Median target absence: 18 frames

- To prevent overfitting:
  - Sequences + initialization frames GT publicly available.
  - GT of test frames sequestered, evaluation carried out on a dedicated server.

VOTS2023 challenge results: 47 trackers tested

- Top trackers: DMAOT, HQTrack, MVOSTracker, Dynamic$_{\text{DEAOT}}$, seqtrack, DMNet, aot, MCMOT, rts_rts50_002, VAPT

- Dominant design choices:
    - Transformer-based
    - Single-stage ST1/LT0 trackers
    - Same architecture used for frame-to-frame localization and re-detection



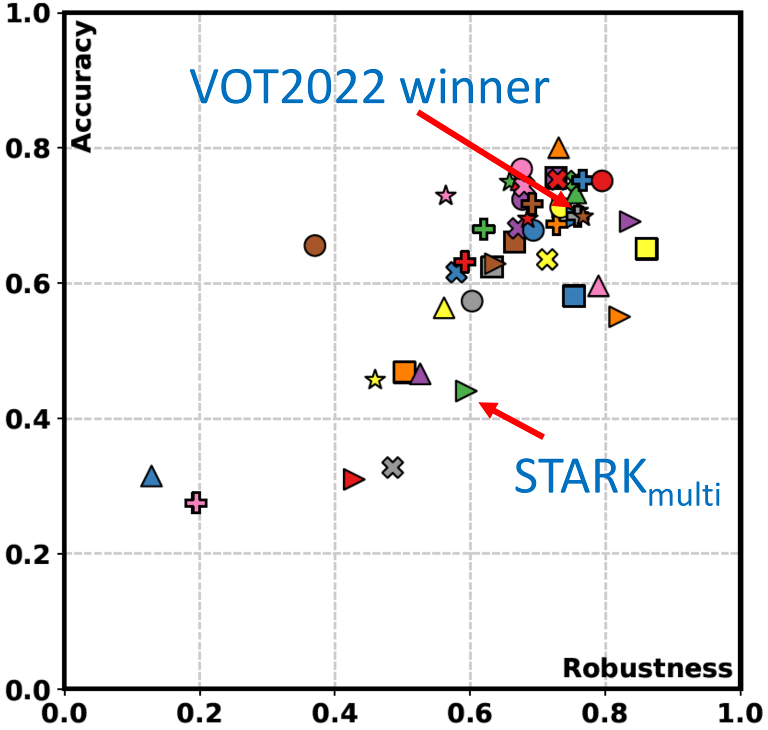
VOTS2023 challenge quality of submissions

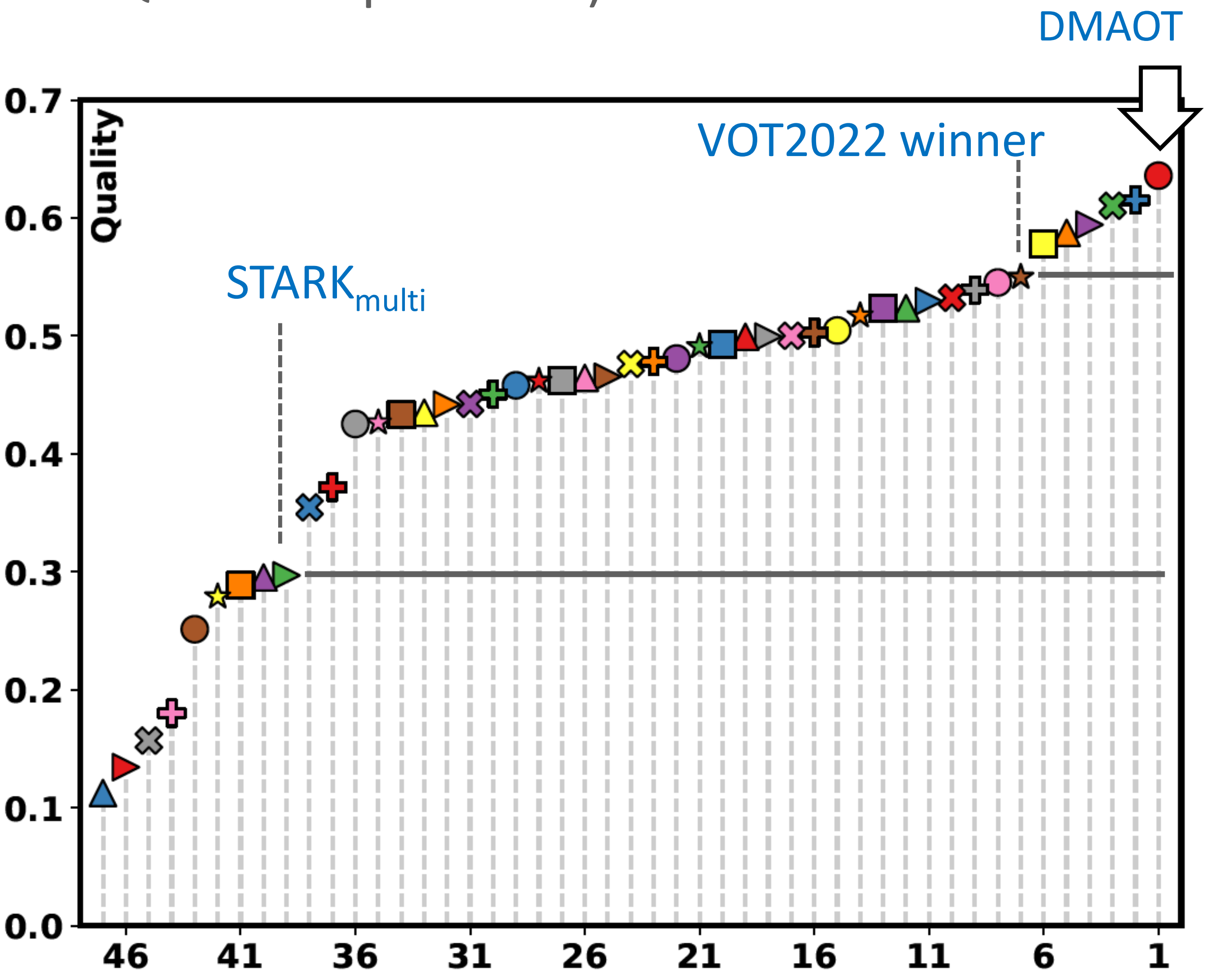
- Baseline 1: Independent STARKs[1] (47% in Q w.r.t. top tracker)

  80% of submissions outperform it

- Baseline 2: VOT2022 winner AOT[2]

  13% (top 6 trackers) outperfrom it



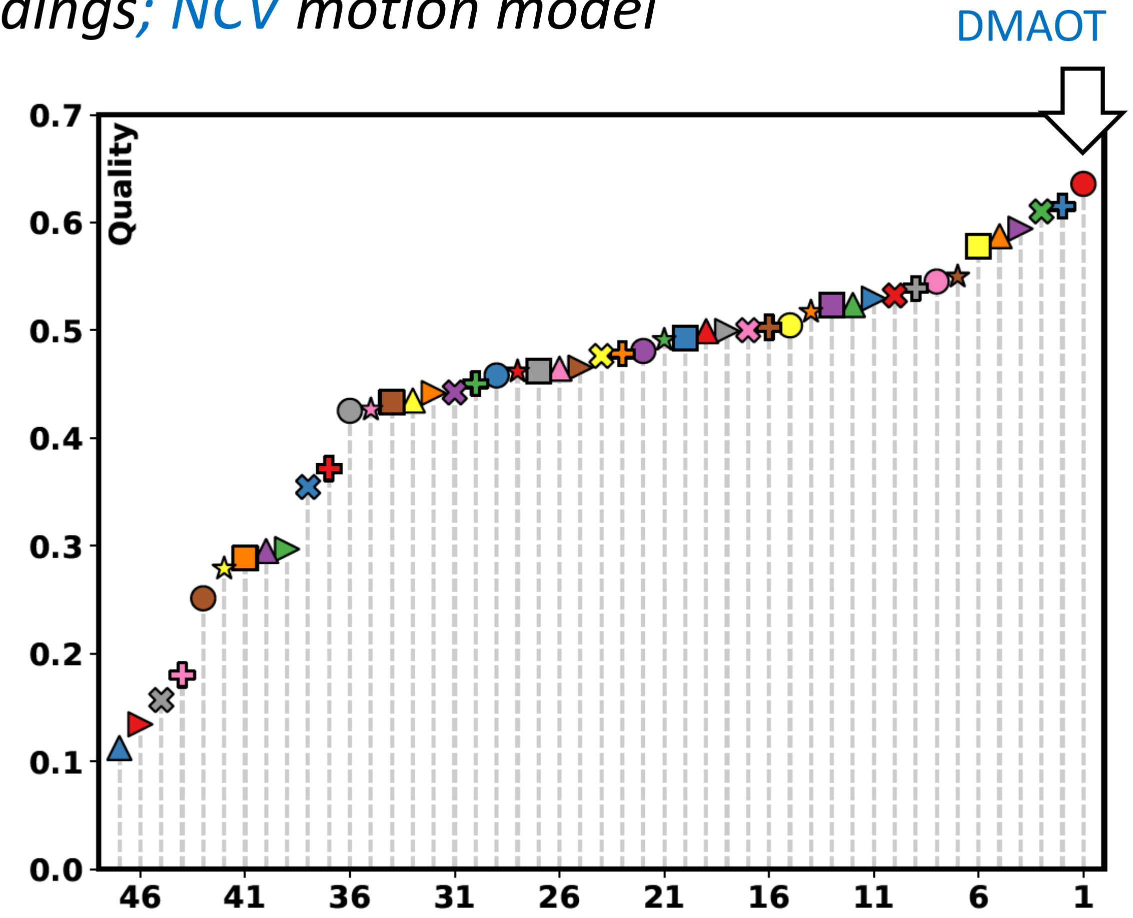


[1] Yan, et al. ICCV2021; [2] Yang, et al. NeurIPS 2021

VOTS2023 challenge results

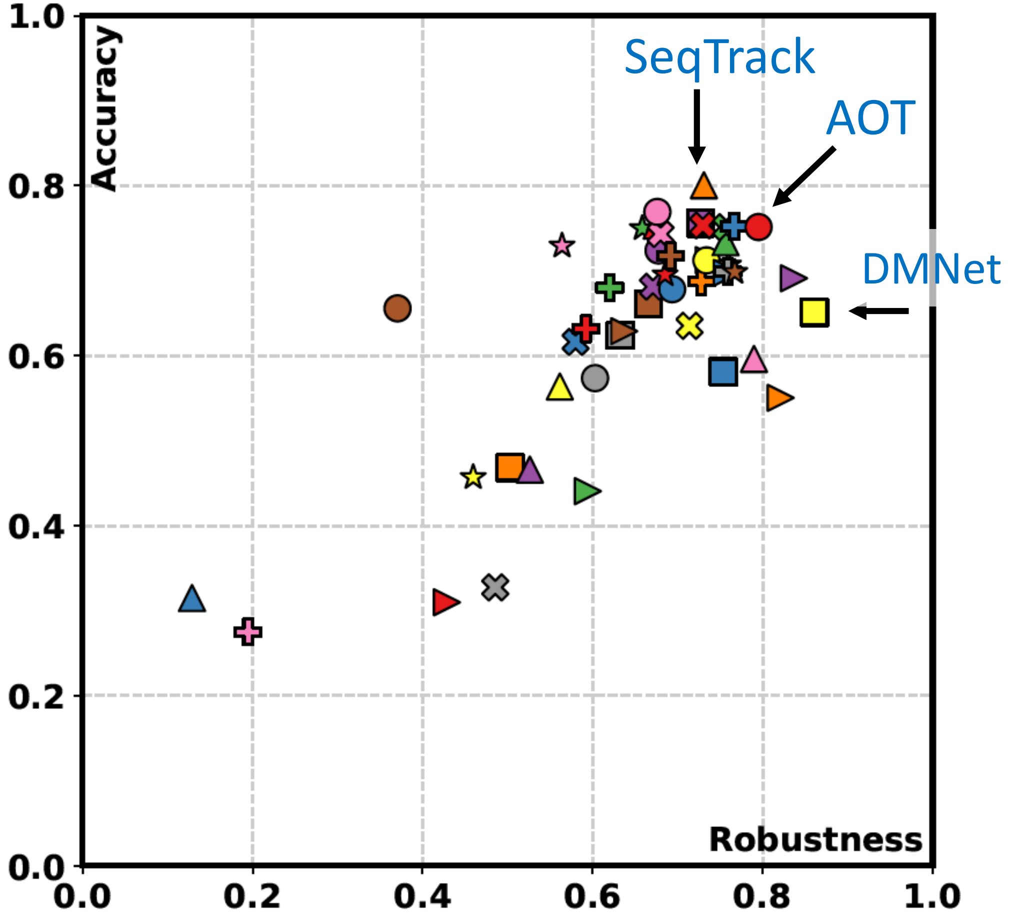
- Top performer DMAOT: Extends the VOT2022 winner AOT
  - *Swin transformer backbone ; Separates long-term and short-term target templates; gated propagation module for visual embeddings; NCV motion model*

- Very good Acc=0.751 & Rob=0.795 (localizes the target 80% of the time)
- Very low drifting (DRE=7%),
- Low false absence prediction (NRE=14%)
- Good target absence prediction: in ADQ=73% cases



VOTS2023 challenge results

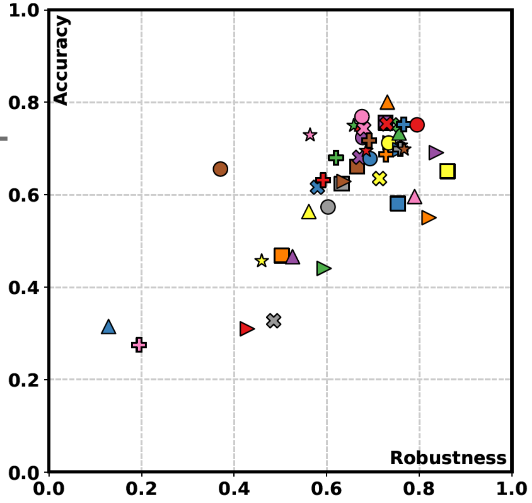
- The top-performer in Q (DMAOT) strikes a good balance in Acc/Rob
- Top robustness: DMNet (Rob=0.86) vs (DMAOT Rob = 0.795)
  - Reason might be the use of optimal transport formulation in segmentation/localization
- Top accuracy: SeqTrack
  - Bounding box tracker with SAM[1] segmentation
  - Care taken when to accept the SAM[1] result



[1] Kirillov, et al., Segment Anything, 2023

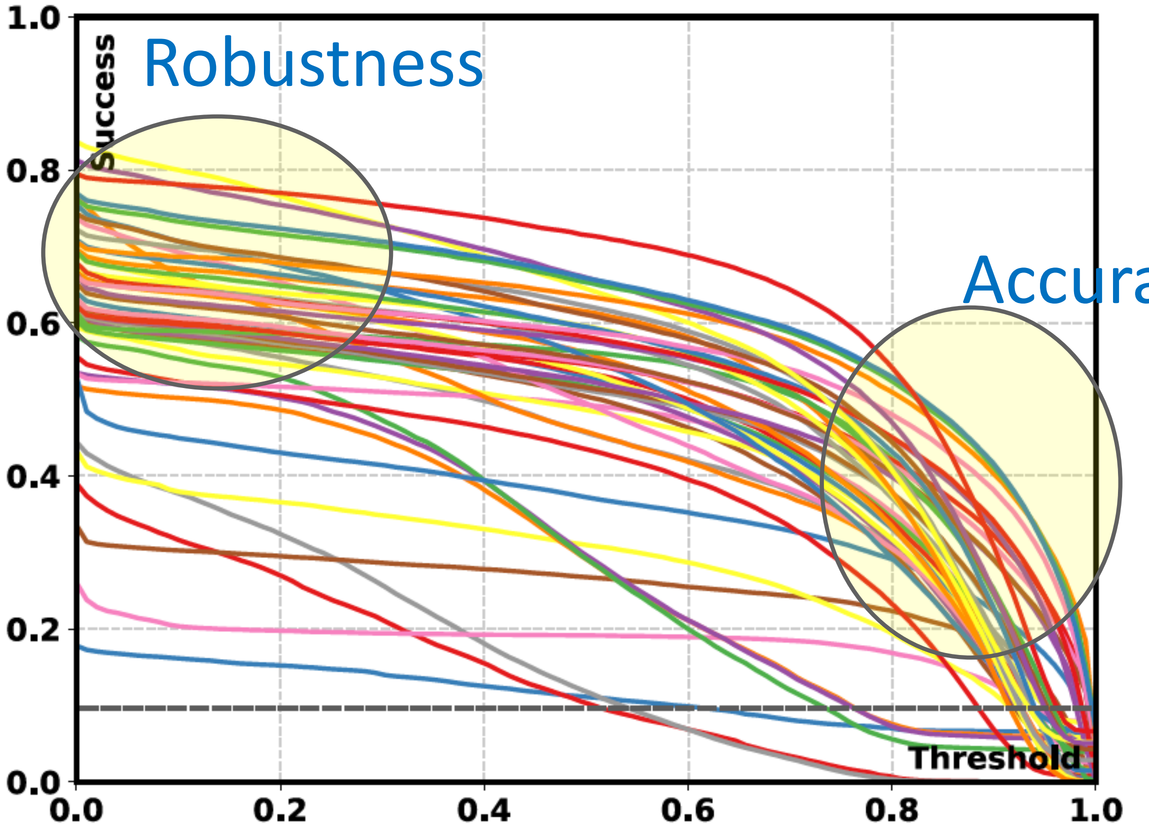
VOTS2023 challenge results

- Q @ low thresholds indicates robustness
- Q @ mid-to-high thresholds indicates accuracy



Clusters:

- @ low thresholds
  - Transformer feature extraction backbones
- @ medium-to-high thresholds
  - Careful use of SAM[1] for segmenting targets (mask or box refinement)



[1] Kirillov, et al., Segment Anything, 2023

VOTS2023 challenge
Winners:

DMAOT by: Yangming Cheng, Zongxin Yang, Yuanyou Xu, Xiaodi Li, Jiahao Li, Yi Yang, Yueting Zhuang

"Decoupled Memory AOT"

VOTS2023 challenge
Spotlight:

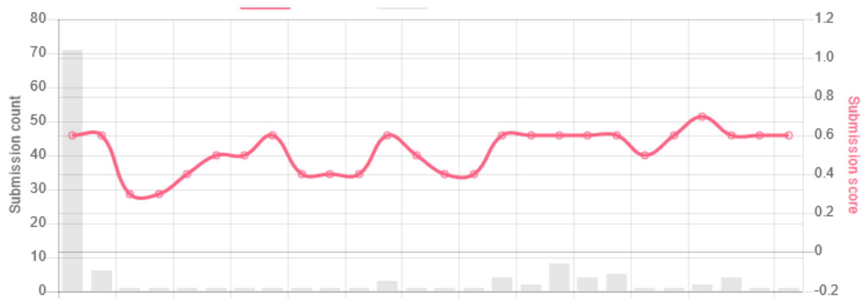
DMNet by: Yinchao Ma, Wangkai Li, Dawei Yang, Rui Sun, Qianjin Yu, Fei Wang, Tianzhu Zhang

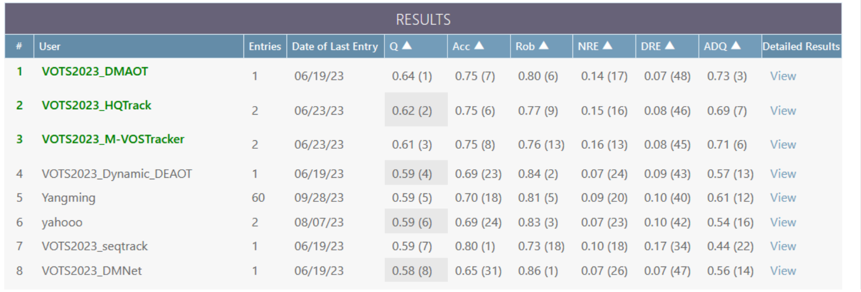
"Dynamic Matching Network"

Winners & Spotlight talks in Session II
@10:45

Summary

- New challenge: General Short/long-term, Single/Multi-target segmentation
- New performance measures, dataset, toolkit and eval server
- Evaluation server open for post-challenge evaluation





RESULTS

#	User	Entries	Date of Last Entry	Q ▲	Acc ▲	Rob ▲	NRE ▲	DRE ▲	ADQ ▲	Detailed Results
1	VOTS2023_DMAOT	1	06/19/23	0.64 (1)	0.75 (7)	0.80 (6)	0.14 (17)	0.07 (48)	0.73 (3)	View
2	VOTS2023_HQTrack	2	06/23/23	0.62 (2)	0.75 (6)	0.77 (9)	0.15 (16)	0.08 (46)	0.69 (7)	View
3	VOTS2023_M-VOSTracker	2	06/23/23	0.61 (3)	0.75 (8)	0.76 (13)	0.16 (13)	0.08 (45)	0.71 (6)	View
4	VOTS2023_Dynamic_DEAOT	1	06/19/23	0.59 (4)	0.69 (23)	0.84 (2)	0.07 (24)	0.09 (43)	0.57 (13)	View
5	Yangming	60	09/28/23	0.59 (5)	0.70 (18)	0.81 (5)	0.09 (20)	0.10 (40)	0.61 (12)	View
6	yahooo	2	08/07/23	0.59 (6)	0.69 (24)	0.83 (3)	0.07 (23)	0.10 (42)	0.54 (16)	View
7	VOTS2023_seqtrack	1	06/19/23	0.59 (7)	0.80 (1)	0.73 (18)	0.10 (18)	0.17 (34)	0.44 (22)	View
8	VOTS2023_DMNet	1	06/19/23	0.58 (8)	0.65 (31)	0.86 (1)	0.07 (26)	0.07 (47)	0.56 (14)	View

VOTS benchmark

- Similarly to VOT2022, evidence indicates remarkable robustness of segmentation trackers vs bbox trackers
- Encourage evaluation of bounding box trackers on the VOTS2023 benchmark (Robustness)

Thanks

- The VOTS2023 committee

M. Kristan

J. Matas

M. Danelljan

M. Felsberg

H. J. Chang

L. Čehovin Z.

A. Lukežič

O. Drbohlav

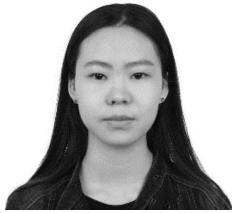
Z. Zhang

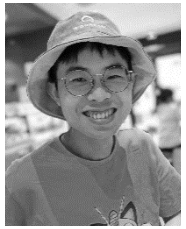
T. Tran

Xuan-Son Vu

Johanna Björklund

C. Mayer


Z. Yushan


Lei Ke

G. Fernandez

- Everyone who participated or contributed

Noor Al-Shakarji38, Dong An20, Michael Arens15, Stefan Becker15, Goutam Bhat3, Sebastian Bullinger15, Antoni B. Chan11, Shijie Chang13, Hanyuan Chen14, Xin Chen13, Yan Chen19, Zhenyu Chen13, Yangming Cheng42, Yutao Cui29, Chunyuan Deng16, Jiahua Dong32, Matteo Dunnhofer41, Wei Feng34, Jianlong Fu27, Jie Gao19, Ruize Han34, Zeqi Hao13, Jun-Yan He14, Keji He20, Zhenyu He18, Xiantao Hu17, Kaer Huang25, Yuqing Huang18, Yi Jiang9, Ben Kang13, Jin-Peng Lan14, Hyungjun Lee30, Chenyang Li14, Jiahao Li42, Ning Li17, Wangkai Li39, Xiaodi Li42, Xin Li31, Pengyu Liu13, Yue Liu23, Huchuan Lu13, Bin Luo14, Ping Luo33, Yinchao Ma39, Deshui Miao18, Christian Micheloni41, Kannappan Palaniappan38, Hancheol Park30, Matthieu Paul3, HouWen Peng26, Zekun Qian34, Gani Rahmon38, Norbert Scherer-Negenborn15, Pengcheng Shao23, Wooksu Shin30, Elham Soltani Kazemi38, Tianhui Song29, Rainer Stiefelhagen24, Rui Sun39, Chuanming Tang37, Zhangyong Tang23, Imad Eddine Toubal38, Jack Valmadre35, Joost van de Weijer12, Luc Van Gool3, Jash Vira35, St`ephane Vujasinovi´c15, Cheng Wan16, Jia Wan8, Dong Wang13, Fei Wang39, Feifan Wang34, He Wang23, Limin Wang29, Song Wang40, Yaowei Wang31, Zhepeng Wang25, Gangshan Wu29, Jiannan Wu33, Qiangqiang Wu11, Xiaojun Wu23, Anqi Xiao20, Jinxia Xie17, Chenlong Xu17, Min Xu10, Tianyang Xu23, Yuanyou Xu42, Bin Yan13, Dawei Yang39, Ming-Hsuan Yang36, Tianyu Yang22, Yi Yang42, Zongxin Yang42, Xuanwu Yin28, Fisher Yu3, Hongyuan Yu28, Qianjin Yu39, Weichen Yu10, YongSheng Yuan13, Zehuan Yuan9, Jianlin Zhang37, Lu Zhang13, Tianzhu Zhang39, Guodongfang Zhao21, Shaochuan Zhao23, Yaozong Zheng17,19, Bineng Zhong17, Jiawen Zhu13, Xuefeng Zhu23, Yueting Zhuang42, ChengAo Zong13, and Kunlong Zuo28

- VOTS2023 sponsors:

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

eyedea

UNIVERSITY OF
BIRMINGHAM