VOT
visual object tracking

The 10th Visual Object Tracking Challenge Results VOT-ST2022, VOT-RT2022

Matej Kristan, Aleš Leonardis, Jiri Matas, Joni-Kristian Kämäräinen, Hyung Jin Chang, Roman Pflugfelder, Gustavo Fernandez, Luka Čehovin Zajc, Alan Lukežič, Michael Felsberg, Martin Danelljan, Ondrej Drbohlav, Johanna Björklund, Yushan Zhang, Zhongqun Zhang, Song Yan, Wenyan Yang, Dingding Cai, Christoph Mayer

University of Ljubljana
Faculty of Computer and Information Science

UNIVERSITY OF BIRMINGHAM

LINKÖPING UNIVERSITY

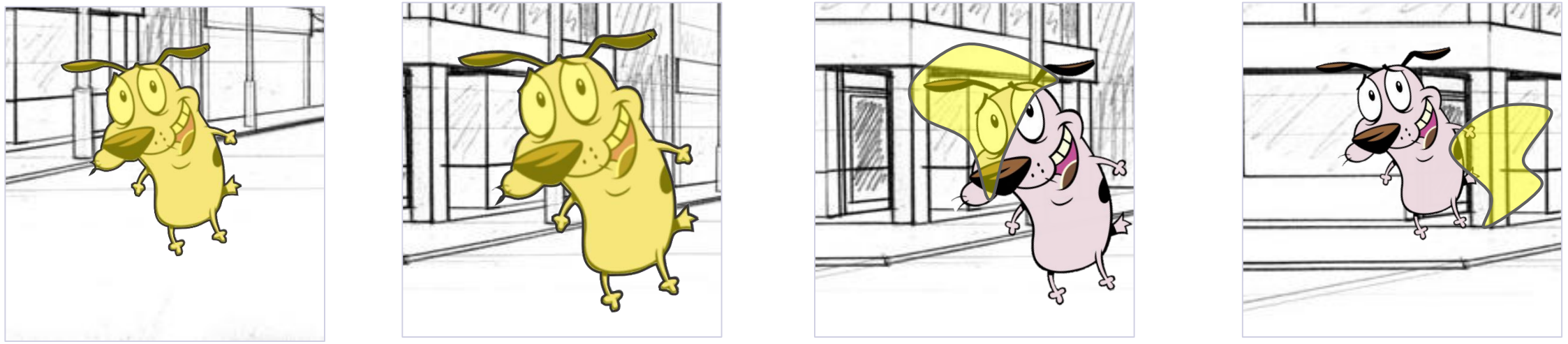
AUSTRIAN INSTITUTE OF TECHNOLOGY

Tampere University

ETH zürich

VOT-ST2022 challenge scope

- Short-term, single-target, causal trackers
- Initialize on the first frame and report position on the rest

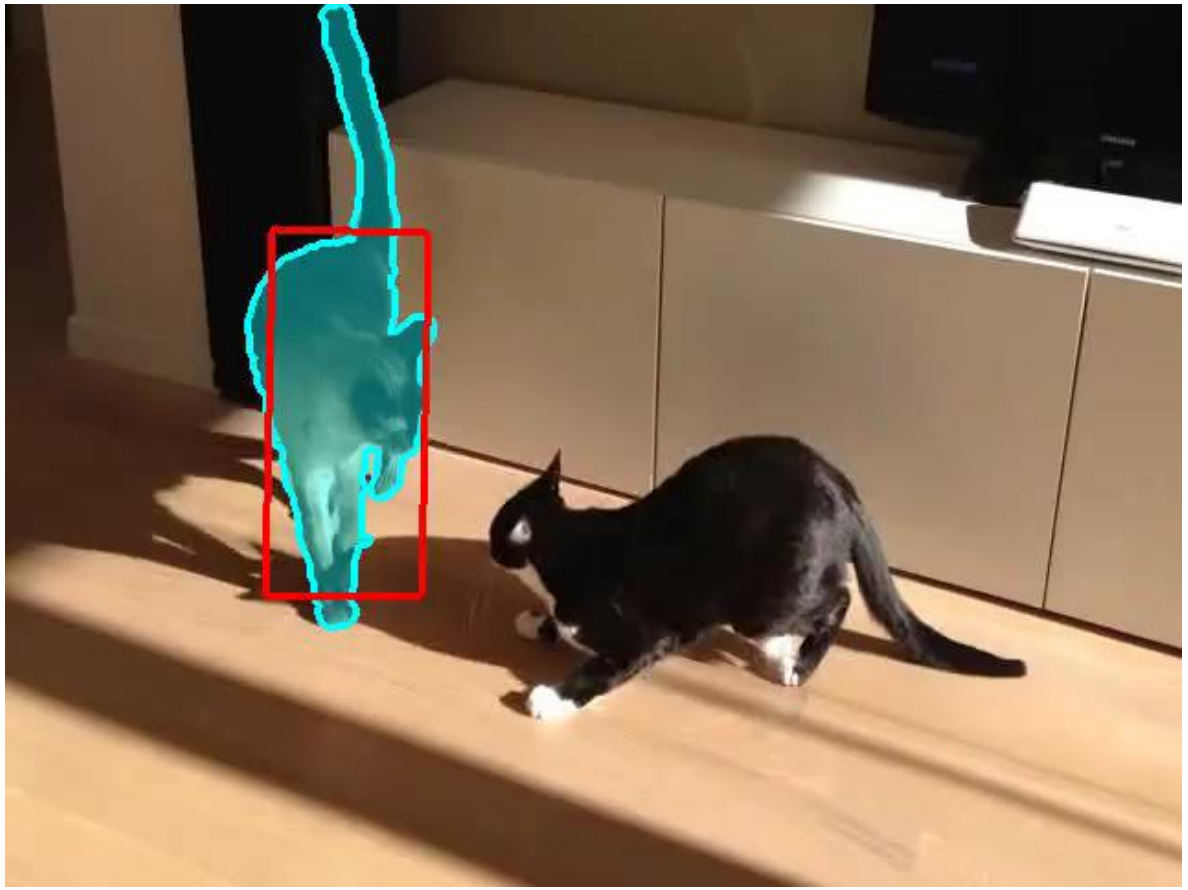


- Mission: promote trackers that track for as long as possible
- No redetection: drifting off the target is considered a failure

The VOT-ST2022 dataset

- Public dataset (60 sequences) +
  Sequestered dataset (60 sequences)
  Winner identified on *sequestered dataset*
- Both datasets refreshed
  - 2 challenging sequences added to each

- All frames manually segmented!
- Each frame annotated by 6 attributes



Red – VOT2019 annotation by a bounding box
Blue – VOT2022 annotation by a segmentation mask

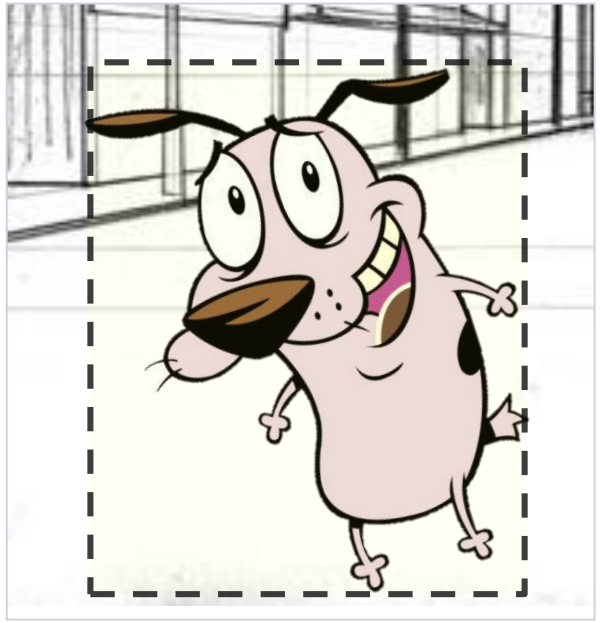
VOT-ST challenge variations

- Bounding boxes abandoned in VOT2020, but reintroduced this year due to pertaining significant research interest in the community
- Standard VOT anchor-based evaluation used (A, R, EAO)

VOT-STs2022

Segmentation mask

VOT-STb2022



Bounding box

Realtime constraint:

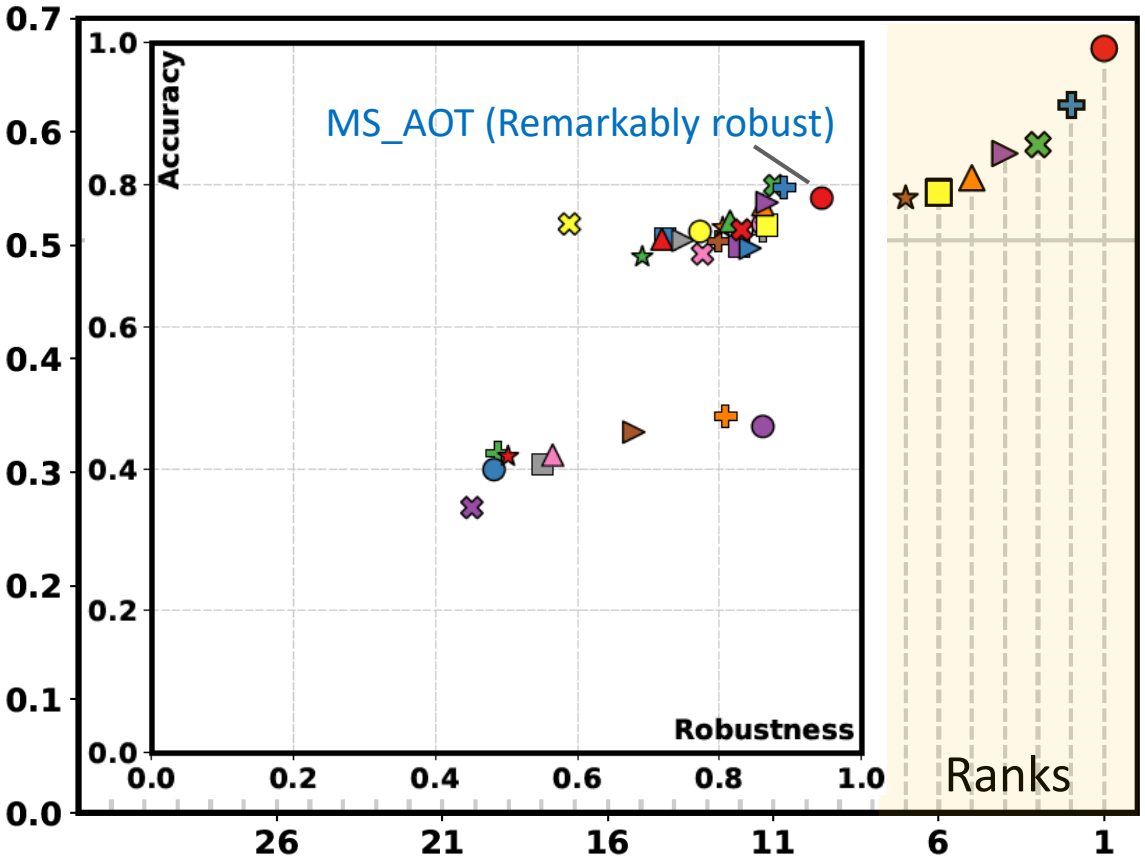
- *Process @20fps*
- *Winners identified on the public dataset*

Variants:

- VOT-RTs2022
- VOT-RTb2022

VOT-STs2022 results on public dataset (31 trackers)

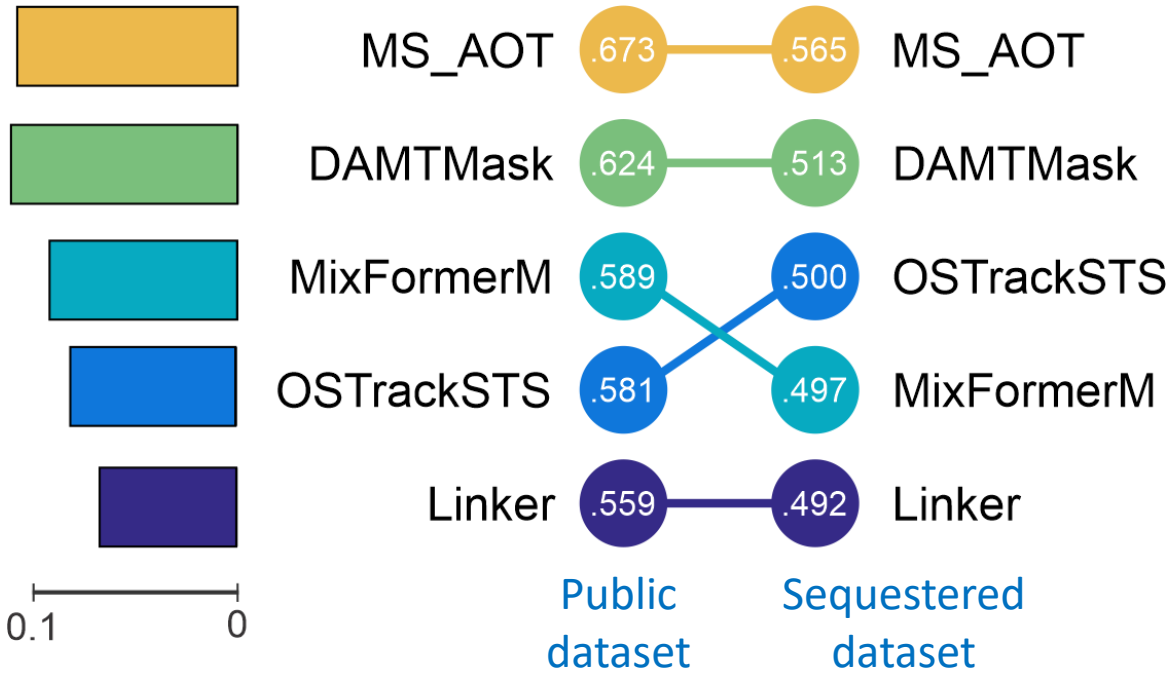
- Top trackers: (1) MS_AOT, (2) DAMTMask (3) MixFormerM, (4) OSTrackSTS, (5) Linker, (6) SRATransTS, (7) TransT_M, (8) DGformer, (9) TransLL, (10) LWL-B2S
- Core methodology:
  - 9 transformers, 1 deep DCF
  - Most use: Mixformer[1], TransT[2]
  - 7 two-stage:
    (i) box localization + (ii) segmentation
- Top performer (MS_AOT) stands out:
  - Single-stage, based on pure
    video object segmentation method[1]

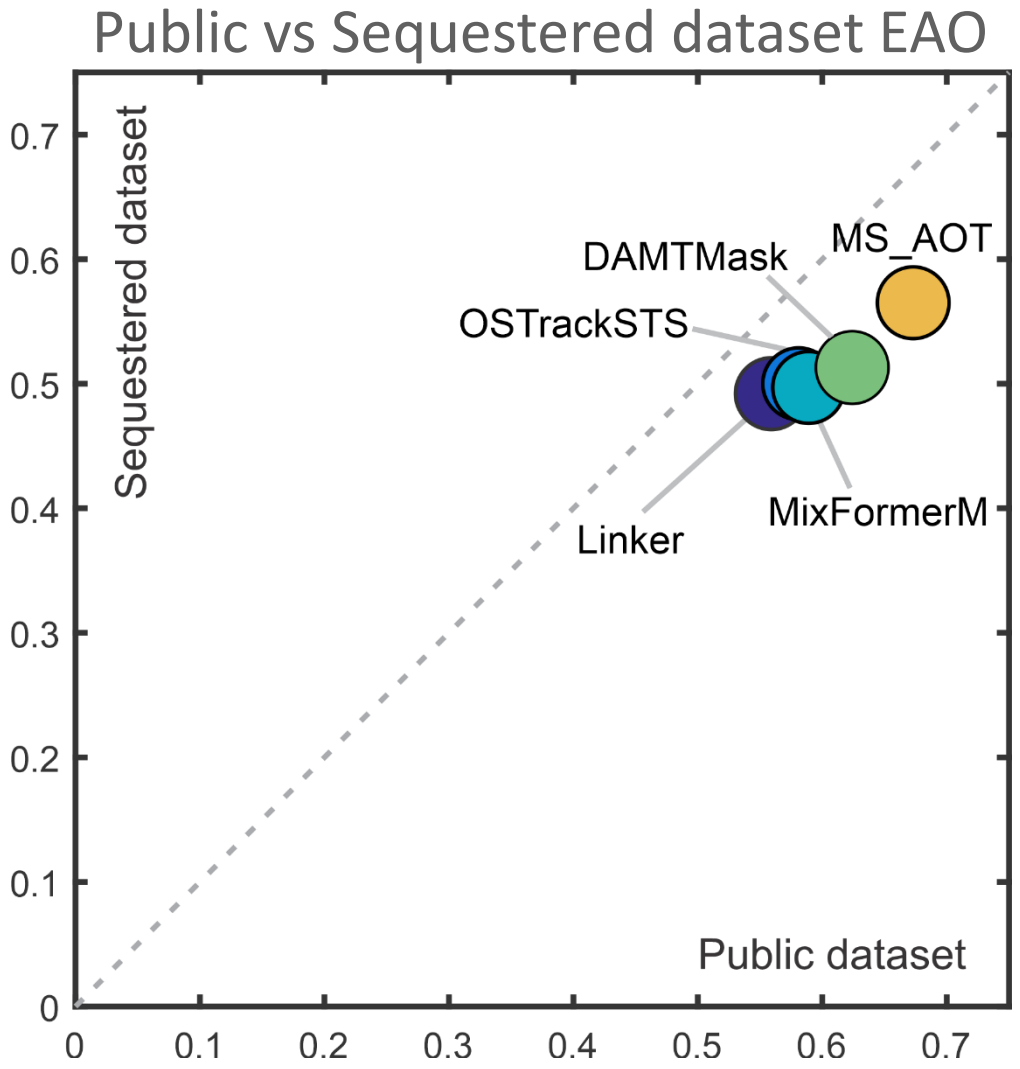


[1]Cui et al. CVPR2022, [2]Chen et al. CVPR2021, [3]Yang et al. Neurips 2021

VOT-STs2022 results on sequestered dataset

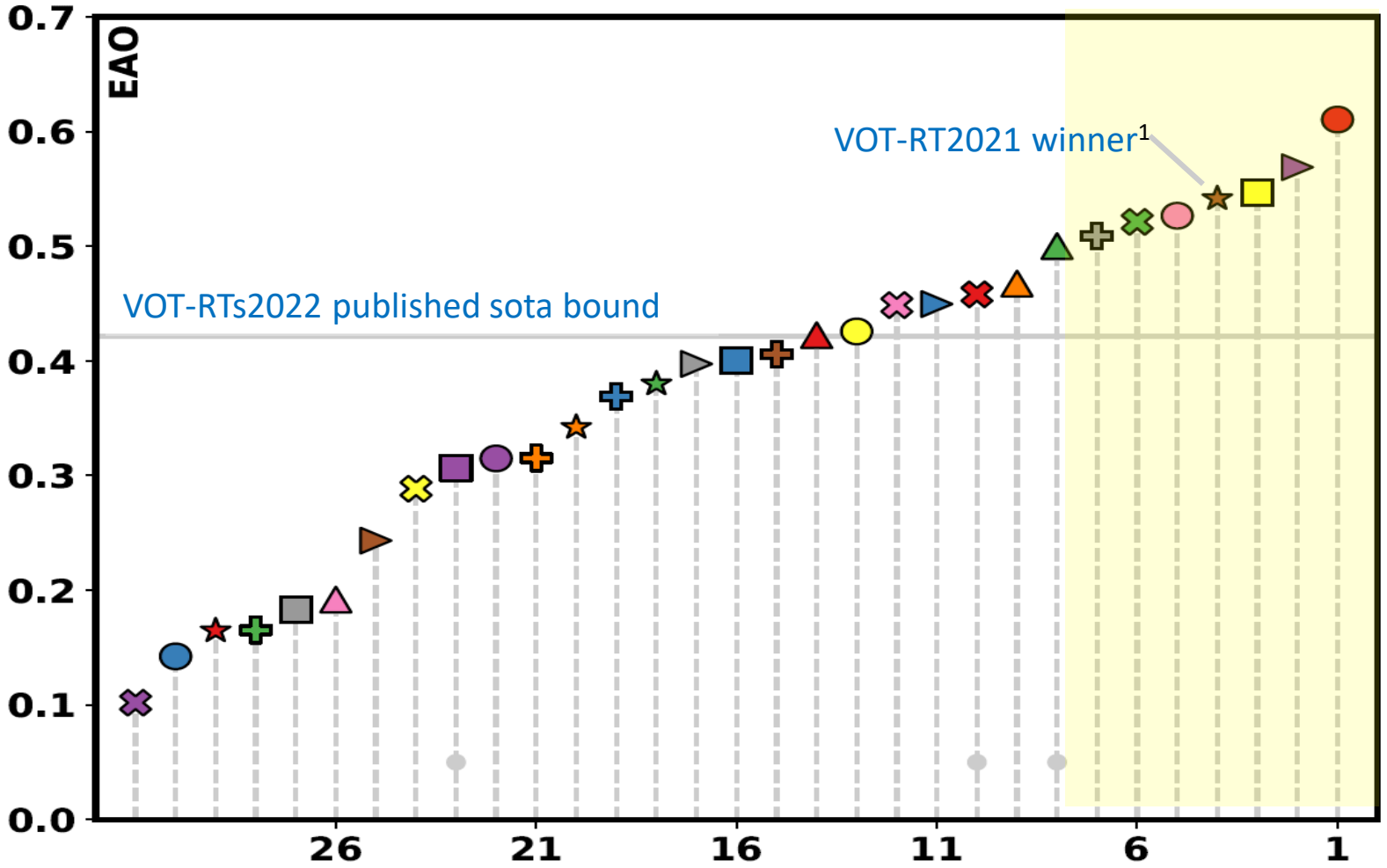
- Comparable results between public and sequestered set
  - Slight relative performance differences
  - Clearly stands out: MS_AOT





VOT-RTs2022 realtime challenge results

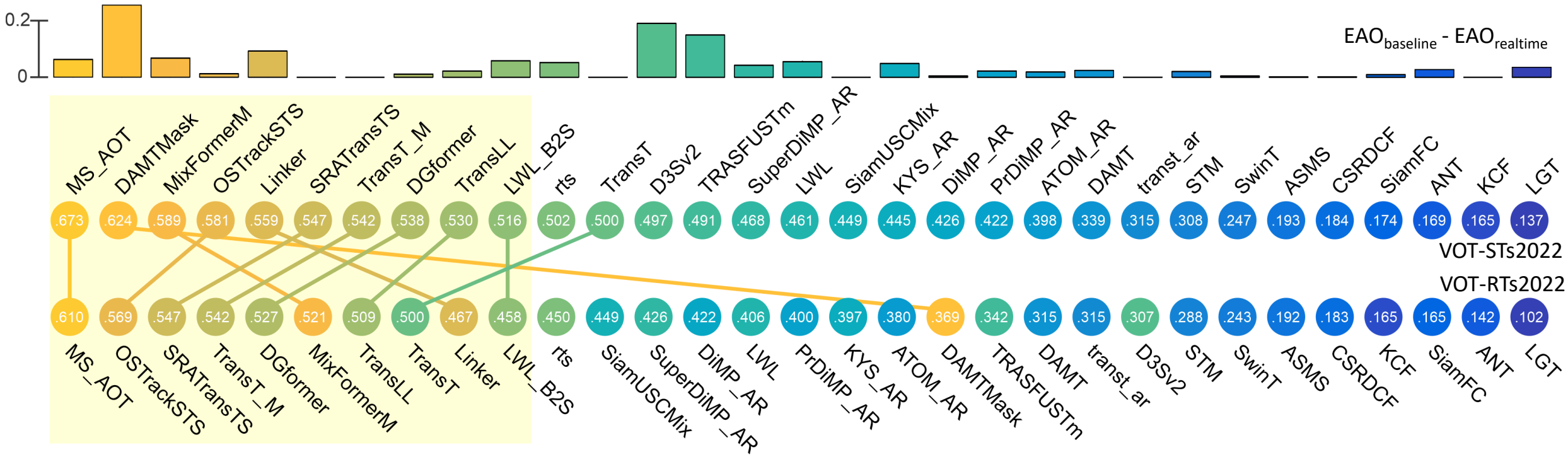
- Top 10: (1) MS_AOT, (2) OSTrackSTS, (3) SRATransTS, (4) TransT_M, (5) DGformer, (6) MixFormerM, (7) TransLL, (8) TransT, (9) Linker, (10) RTS



- 9 are transformers
- 3 outperform the VOT-RT2021 winner[1]
- Top: MS_AOT
- 45% of submissions outperform VOT-RT2022 sota bound

[1]TransT_M [Chen et al., Arxiv2022]

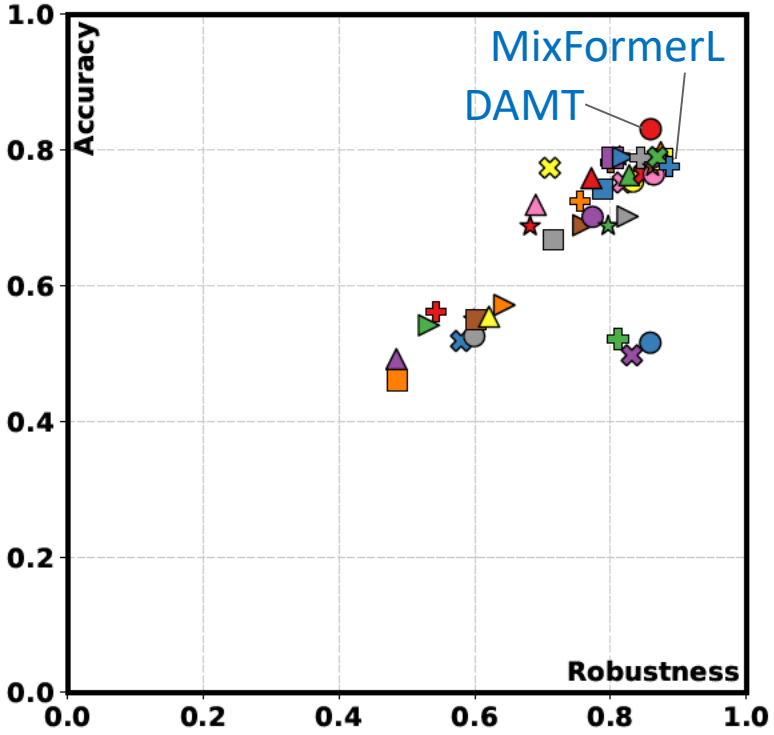
VOTs2022 Realtime vs Baseline results



- 9 top VOT-RTs2022 trackers among top 10 on VOT-STs2022 challenge!
- The top RT tracker MS_AOT is top in VOT-STs2022

VOT-STb2022 results on public dataset (41 trackers)

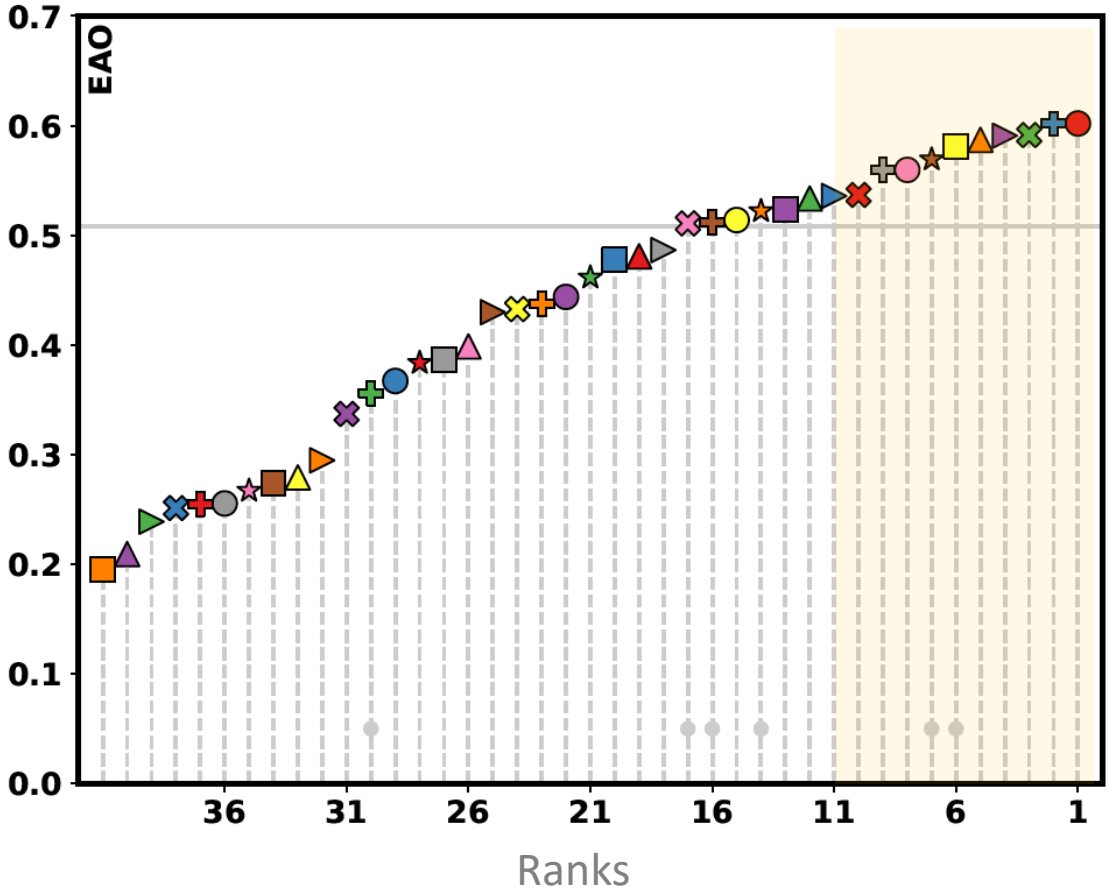
- Top trackers: (1) DAMT, (2) MixFormerL, (3) OSTrackSTB, (4) APMT_MR, (5) Mixformer, (6) APMT_RT, (7) ADOTstb, (8) SRATransT, (9) Linker_B, (10) TransT_M
- All top trackers are transformers



Same (top) EAO:
DAMT & MixFormerL

More accurate

More robust



VOT-STb2022 results on sequestered dataset

- Two ways to evaluate performance: (i) mask GT and (ii) bounding box GT

Public dataset ranks (in EAO)

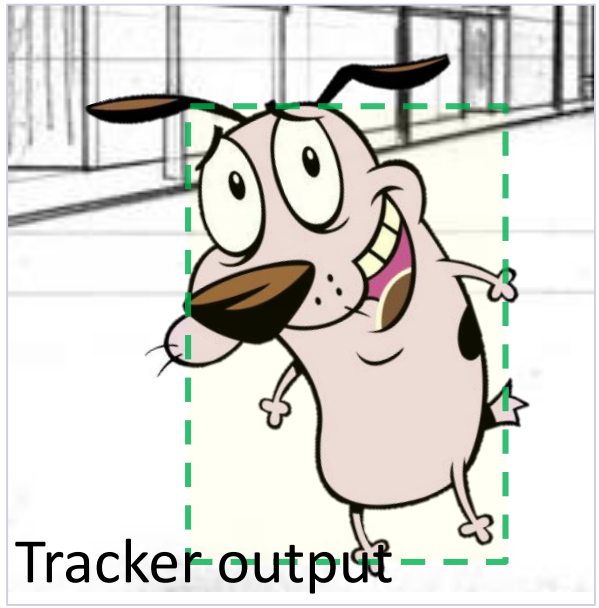
MixFormerL
DAMT
OSTrackSTB
APMT_MR
ADOTstb

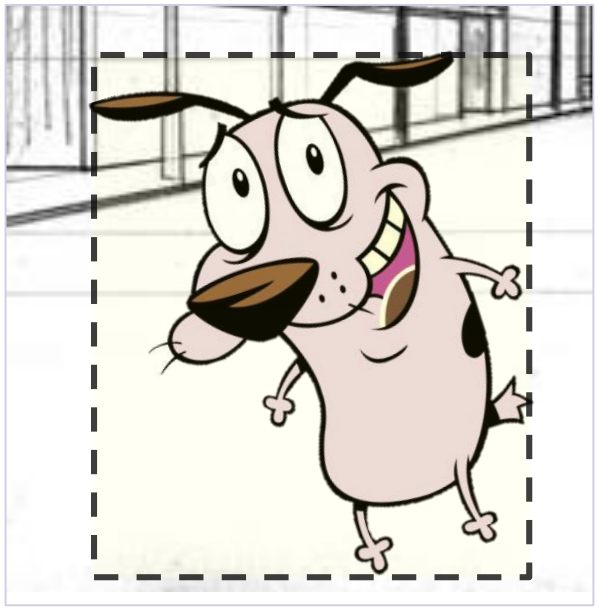
Sequestered (bounding box GT)

	Tracker	EAO	A	R
1.	OSTrackSTB	0.523	0.800	0.881
2.	APMT_MR	0.508	0.800	0.862
3.	MixFormerL	0.500	0.837	0.825
4.	ADOTstb	0.499	0.812	0.840
5.	DAMT	0.479	0.804	0.826

Sequestered (mask GT)

	Tracker	EAO	A	R
1.	APMT_MR	0.322	0.528	0.845
2.	OSTrackSTB	0.309	0.517	0.839
3.	MixFormerL	0.306	0.542	0.803
4.	ADOTstb	0.301	0.532	0.806
5.	DAMT	0.289	0.503	0.807



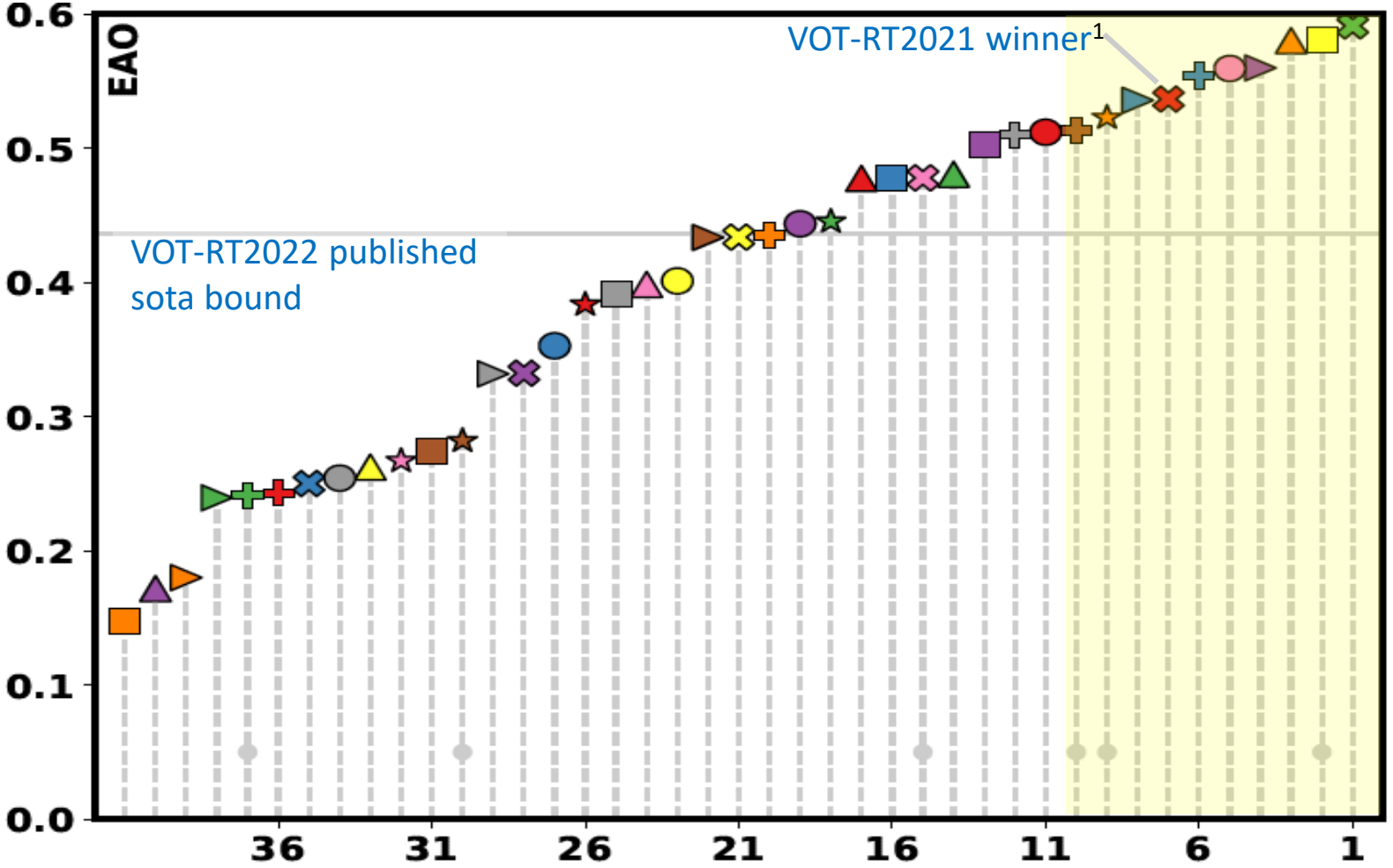


GT(ii)

GT(i)

VOT-RTb2022 realtime challenge results

- Top 10: (1) OSTrackSTB, (2) APMT_RT, (3)MixFormer, (4) APT_MR, (5) SRATransT, (6) DAMT, (7) TransT_M, (8) vittrack, (9) SBT, (10) TransT



- All 10 transformers
- 6 outperform the VOT-RT2021 winner[1]
- Top: OSTrackSTB
- 54% of submissions outperform VOT-RT2022 sota bound

[1]TransT_M [Chen et al., Arxiv2022]

Box trackers vs Segmentation trackers

- VOT-STb2022 top 10 performers:
  - 7 perform well in VOT-STs2022
  - 3 of top 4 VOT-STb2022 are among top VOT-STs2022 (3rd, 2nd, 4th)

VOT-STb2022

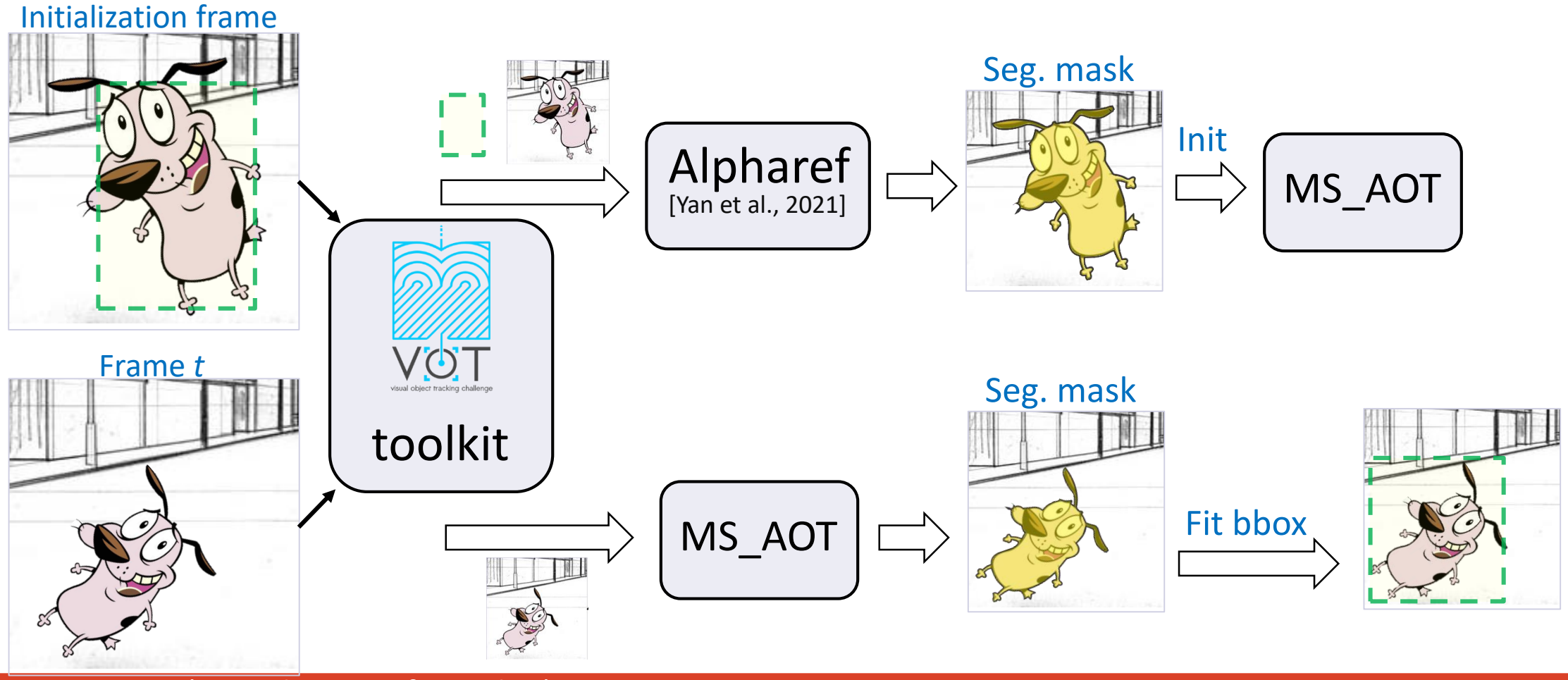
Tracker	EAO	A	R
MixFormerL	0.602①	0.831①	0.859
DAMT	0.602②	0.776	0.887①
OSTrackSTB	0.591③	0.790	0.869
APMT_MR	0.591	0.787	0.877③
MixFormer	0.587	0.797②	0.874
APMT_RT	0.581	0.787	0.877②
ADOTstb	0.569	0.775	0.862
SRATransT	0.560	0.764	0.864
Linker_B	0.560	0.789	0.844
TransT_M	0.537	0.765	0.849
vittrack	0.536	0.789	0.818
SuperFus	0.534	0.763	0.828
SwinTrack	0.524	0.788	0.803

VOT-STs2022

Tracker	EAO	A	R
MS_AOT	0.673①	0.781③	0.944①
DAMTMask	0.624②	0.796②	0.891②
MixFormerM	0.589③	0.799①	0.878③
OSTrackSTS	0.581	0.775	0.867
Linker	0.559	0.772	0.861
SRATransTS	0.547	0.743	0.866
TransT_M	0.542	0.743	0.865
DGformer	0.538	0.744	0.861
TransLL	0.530	0.735	0.861
LWL_B2S	0.516	0.736	0.831
rts	0.502	0.710	0.843
TransT	0.500	0.749	0.815
D3Sv2	0.497	0.713	0.827

Box trackers vs Segmentation trackers

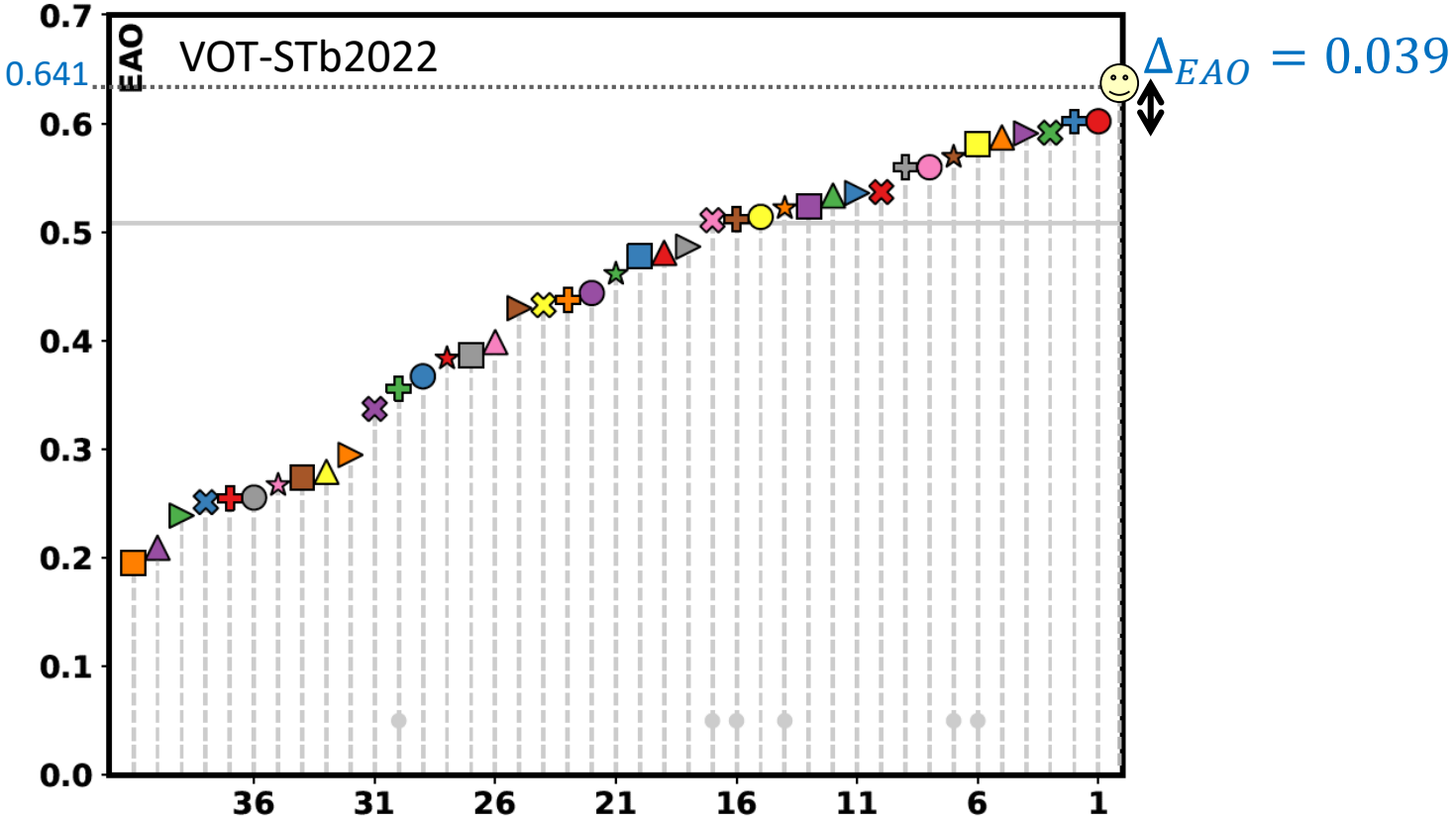
- VOT-STs2022 winner MS_AOT run on public STb2022 dataset
  - Initialize by AlphaRef[1] ; Output is bounding box fited to mask prediction [1][Yan et al., CVPR2021]

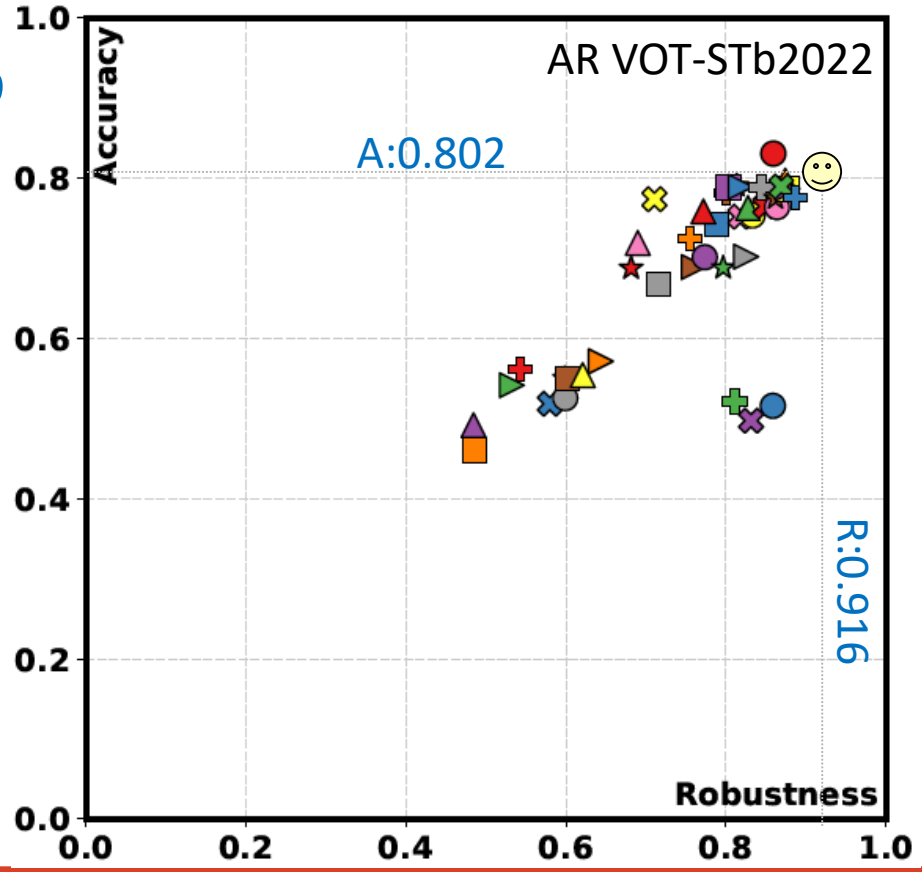


Box trackers vs Segmentation trackers

- VOT-STs2022 winner MS_AOT run on public STb2022 dataset
  - Initialize by AlphaRef[1] ; Output is bounding box fited to mask prediction [1][Yan et al., CVPR2021]

  (MS_AOT) EAO: 0.641, A:0.802, R:0.916



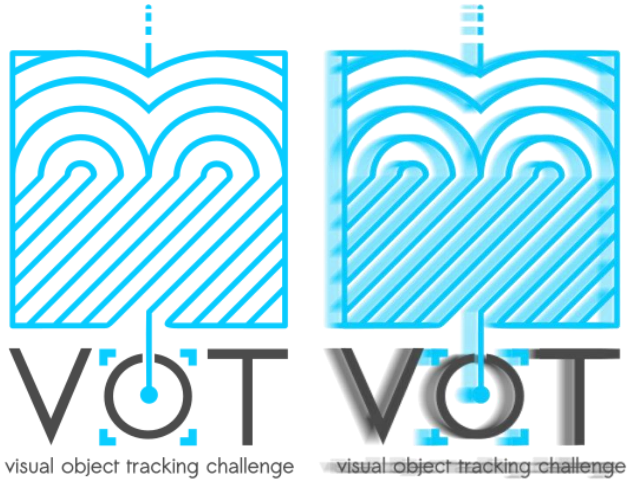


Annuncing VOT sequestered analysis service

- Since introduction of the VOT sequestered evaluation (2017) it is clear that trackers prone to overfitting on public datasets
- Decision: Create VOT-ST sequestered analysis service (VOT-SAS)
- Partnership with: WARA-ML
  - WASP: Wallenberg AI, Autonomous Systems and Software Program
  - Financed by the Knut and Alice Wallenberg foundation
  - The development team: Yushan Zhang, Nithesh Chandher Karthikeyan, Johanna Björklun, Michael Felsberg
  - Implement & maintain the VOT-SAS

Annuncing VOT sequestered analysis service

- Login using the EduGain service
- Submit a CONDA environment to the server
  - Prepare code with the toolkit, CUDA 11.x, minimum GPU requirement, etc.
- Tracker run, results to the leaderboard
- Detailed guidelines with demo code available soon
- Still in the process of securing hardware resources
- Plans to make beta test available by December
- Hope to offer leaderboard results for your ICCV2023 submissions



VOT-STs2022 &
VOT-RTs2022 Winners:

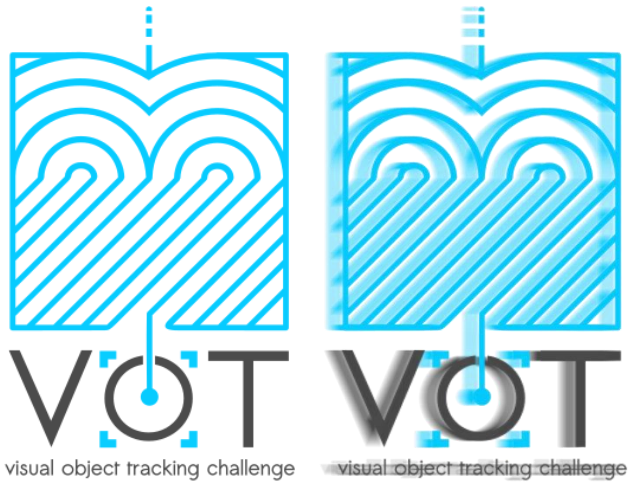
MS_AOT by: Zongxin Yang, Yangming Cheng, Yuanyou Xu, Chao Sun, Yi Yang, Yueting Zhuang

"Associating Objects with Multi-scale Transformers for Video Object Segmentation"

VOT-STb2022 (mask GT)
Winners:

APMT_MR by: Yinchao Ma, Dawei Yang, Qianjin Yu, Jianfeng He, Fei Wang, Wangkai Li, Tianzhu Zhang

"Adaptive Part Mining Tracker with Multi-Region"



VOT-STb2022 (box GT) &
VOT-RTb Winners:

OSTrackSTB by: Botao Ye, Hong Chang, Bingpeng Ma, Shiguang Shan, Xilin Chen

"One-stream tracker with online template updating"

Winners talks in Session III
(16:35-18:00, GMT+2 time zone)

Summary

- Transformers became the dominant methodology of top trackers
- Continued trend that top ST trackers are top on RT as well
- Observed emergence of remarkably robust segmentation trackers
- Invest more research into purely segmentation trackers
- Plans to publish the first sequestered ST tracker evaluation service VOT-SAS

Thanks

- The VOT2022 committee


M. Kristan


J. Matas


A. Leonardis


J. K. Kamarainen


H. J. Chang


R. Pflugfelder


G. Fernandez


L. Čehovin


A. Lukežič


M. Felsberg

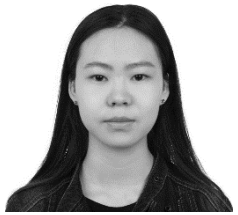

M. Danelljan


O. Drbohlav


Z. Zhang


Y. Song


Wenyan Yang


Z. Yushan


C. Mayer


Dingding Cai


Johanna Björklund

- Everyone who participated or contributed

Kang Ben18, Goutam Bhat10, Hong Chang24, Guangqi Chen16, Jiaye Chen26, Shengyong Chen43, Xilin Chen24, Xin Chen18, Xiuyi Chen13, Yiwei Chen35, Yu-Hsi Chen12, Zhixing Chen16, Yangming Cheng55, Angelo Ciaramella47, Yutao Cui30, Benjamin Dˇzubur1, Mohana Murali Dasari22, Qili Deng16, Debajyoti Dhar39, Shangzhe Di14, Emanuel Di Nardo46,47, Daniel K. Du16, Matteo Dunnhofer51, Heng Fan48, Zhenhua Feng50, Zhihong Fu16, Shang Gao41, Rama Krishna Gorthi22, Eric Granger27, Q.H. Gu15, Himanshu Gupta19, Jianfeng He49, Keji He13, Yan Huang13, Deepak Jangid19, Rongrong Ji53, Cheng Jiang30, Yingjie Jiang26, Felix J¨aremo Lawin4, Ze Kang26, Madhu Kiran27, Josef Kittler50, Simiao Lai18, Xiangyuan Lan32, Dongwook Lee34, Hyunjeong Lee34, Seohyung Lee34, Hui Li26, Ming Li17, Wangkai Li49, Xi Li55, Xianxian Li20, Xiao Li16, Zhe Li41, Liting Lin37, Haibin Ling40, Bo Liu25, Chang Liu18, Si Liu23, Huchuan Lu18, Rafael M. O. Cruz27, Bingpeng Ma44, Chao Ma36, Jie Ma21, Yinchao Ma49, Niki Martinel51, Alireza Memarmoghadam45, Christian Micheloni51, Payman Moallem45, Le Thanh Nguyen-Meidine27, Siyang Pan35, ChangBeom Park34, Danda Paudel10, Matthieu Paul10, Houwen Peng28, Andreas Robinson4, Litu Rout39, Shiguang Shan24, Kristian Simonato51, Tianhui Song30, Xiaoning Song26, Chao Sun55, Jingna Sun16, Zhangyong Tang26, Radu Timofte10,52, Chi-Yi Tsai42, Luc Van Gool10, Om Prakash Verma19, Dong Wang18, Fei Wang49, Liang Wang13, Liangliang Wang16, Lijun Wang18, Limin Wang30, Qiang Wang35, Gangshan Wu30, Jinlin Wu13, Xiaojun Wu26, Fei Xie38, Tianyang Xu26, Wei Xu16, Yong Xu37, Yuanyou Xu55, Wanli Xue43, Zizheng Xun14, Bin Yan18, Dawei Yang49, Jinyu Yang41, Wankou Yang38, Xiaoyun Yang33, Yi Yang55, Yichun Yang30, Zongxin Yang55, Botao Ye24, Fisher Yu10, Hongyuan Yu13, Jiaqian Yu35, Qianjin Yu49, Weichen Yu13, Kang Ze26, Jiang Zhai38, Chengwei Zhang17, Chunhu Zhang36, Kaihua Zhang29, Tianzhu Zhang49, Wenkang Zhang38, Zhibin Zhang43, Zhipeng Zhang31, Jie Zhao18, Shaochuan Zhao26, Feng Zheng41, Haixia Zheng54, Min Zheng16, Bineng Zhong20, Jiawen Zhu18, Xuefeng Zhu26, and Yueting Zhuang55

- VOT2022 sponsor:

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