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
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)

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) OStTrackSTS, (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
- Top: MS_AOT
- 45% of submissions outperform VOT-RT2022 sota bound

1TransT_M [Chen et al., Arxiv2022]
• 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) OTrackSTB, (4) APMT_MR, (5) Mixformer, (6) APMT_RT, (7) ADOTstb, (8) SRATransT, (9) Linker_B, (10) TransT_M

- All top trackers are transformers

Graph showing accuracy and robustness with DAMT and MixFormerL highlighted as having the same top EAO, with DAMT being more accurate and MixFormerL more robust.
VOT-STb2022 results on sequestered dataset

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

<table>
<thead>
<tr>
<th>Public dataset ranks</th>
<th>Sequestered (bounding box GT)</th>
<th>Sequestered (mask GT)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(in EAO)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MixFormerL</td>
<td>1. OSTrackSTB 0.523 0.800 0.881</td>
<td>1. APMT_MG 0.322 0.528 0.845</td>
</tr>
<tr>
<td>DAMT</td>
<td>2. APMT_MR 0.508 0.800 0.862</td>
<td>2. OSTrackSTB 0.309 0.517 0.839</td>
</tr>
<tr>
<td>OSTrackSTB</td>
<td>3. MixFormerL 0.500 0.837 0.825</td>
<td>3. MixFormerL 0.306 0.542 0.803</td>
</tr>
<tr>
<td>APMT_MR</td>
<td>4. ADOTstb 0.499 0.812 0.840</td>
<td>4. ADOTstb 0.301 0.532 0.806</td>
</tr>
<tr>
<td>ADOTstb</td>
<td>5. DAMT 0.479 0.804 0.826</td>
<td>5. DAMT 0.289 0.503 0.807</td>
</tr>
</tbody>
</table>

Tracker output

GT(ii)

GT(i)
VOT-RTb2022  realtime challenge results

• Top 10: (1) OTrackSTB, (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
• Top: OTrackSTB
• 54% of submissions outperform VOT-RT2022 sota bound

1TransT_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)

<table>
<thead>
<tr>
<th>Tracker</th>
<th>EAO</th>
<th>A</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>MixFormer</td>
<td>0.602</td>
<td>0.831</td>
<td>0.859</td>
</tr>
<tr>
<td>DAMT</td>
<td>0.602</td>
<td>0.776</td>
<td>0.887</td>
</tr>
<tr>
<td>OTrackSTb</td>
<td>0.591</td>
<td>0.790</td>
<td>0.869</td>
</tr>
<tr>
<td>APMT_MR</td>
<td>0.591</td>
<td>0.787</td>
<td>0.877</td>
</tr>
<tr>
<td>MixFormerL</td>
<td>0.587</td>
<td>0.797</td>
<td>0.874</td>
</tr>
<tr>
<td>APMT_RT</td>
<td>0.581</td>
<td>0.787</td>
<td>0.877</td>
</tr>
<tr>
<td>ADOTstb</td>
<td>0.569</td>
<td>0.775</td>
<td>0.862</td>
</tr>
<tr>
<td>SRATransT</td>
<td>0.560</td>
<td>0.764</td>
<td>0.864</td>
</tr>
<tr>
<td>Linker_B</td>
<td>0.560</td>
<td>0.789</td>
<td>0.844</td>
</tr>
<tr>
<td>TransT_M</td>
<td>0.537</td>
<td>0.765</td>
<td>0.849</td>
</tr>
<tr>
<td>vitrack</td>
<td>0.536</td>
<td>0.789</td>
<td>0.818</td>
</tr>
<tr>
<td>SuperFus</td>
<td>0.534</td>
<td>0.763</td>
<td>0.828</td>
</tr>
<tr>
<td>SwinTrack</td>
<td>0.524</td>
<td>0.788</td>
<td>0.803</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Tracker</th>
<th>EAO</th>
<th>A</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>MS_AOT</td>
<td>0.673</td>
<td>0.781</td>
<td>0.944</td>
</tr>
<tr>
<td>DAMTMask</td>
<td>0.624</td>
<td>0.796</td>
<td>0.891</td>
</tr>
<tr>
<td>MixFormerM</td>
<td>0.589</td>
<td>0.799</td>
<td>0.878</td>
</tr>
<tr>
<td>OTrackSTS</td>
<td>0.581</td>
<td>0.775</td>
<td>0.867</td>
</tr>
<tr>
<td>Linker</td>
<td>0.559</td>
<td>0.772</td>
<td>0.861</td>
</tr>
<tr>
<td>SRATransTS</td>
<td>0.547</td>
<td>0.743</td>
<td>0.866</td>
</tr>
<tr>
<td>TransT_M</td>
<td>0.542</td>
<td>0.743</td>
<td>0.865</td>
</tr>
<tr>
<td>DGformer</td>
<td>0.538</td>
<td>0.744</td>
<td>0.861</td>
</tr>
<tr>
<td>TransLL</td>
<td>0.530</td>
<td>0.735</td>
<td>0.861</td>
</tr>
<tr>
<td>LWL_B2S</td>
<td>0.516</td>
<td>0.736</td>
<td>0.831</td>
</tr>
<tr>
<td>rts</td>
<td>0.502</td>
<td>0.710</td>
<td>0.843</td>
</tr>
<tr>
<td>TransT</td>
<td>0.500</td>
<td>0.749</td>
<td>0.815</td>
</tr>
<tr>
<td>D3Sv2</td>
<td>0.497</td>
<td>0.713</td>
<td>0.827</td>
</tr>
</tbody>
</table>
Box trackers vs Segmentation trackers

- VOT-STs2022 winner MS_AOT run on public STb2022 dataset
  - Initialize by AlphaRef\(^1\); Output is bounding box fitted 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 AlphaRef¹; Output is bounding box fitted to mask prediction ¹[Yan et al., CVPR2021]

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

Δ_EAO = 0.039
Announcing 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
Announcing 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**


• **Everyone who participated or contributed**

Kang Ben, Goutam Bhat, Hong Cha, Chen, Jiayi Chen, Shengyong Chen, Xin Xingyan, Xin, Xiu Yi Chen, Yu-Hsi Chen, Zhiming Chen, Yangming Chen, Jing, Chen, Xi, Yu, Hsi, Chen, 12, Zhixing Chen, 16, Yangming Cheng, 16, Anq, Angelo Ciaramel, 16, Yutao Cui, 16, Benjamin D’zubr, 1, Mohana Murali, 16, Qili Deng, 16, Debajoti Dhar, 1, Shangze Di, 14, Emanuel Di Nardo, 46, 17, Daniel K. Du, 16, Matteo Dunnhofe, 51, Heng Fan, 48, Zhenhua Feng, 50, Zhihong Fu, 16, Shang Gao, 41, Rama Krishna Gorthi, 22, Eric Granger, 27, Q.H. Gu, 15, Himanshu Gupta, 19, Jianfeng He, 49, Keji He, 13, Yan Huang, 13, Deepak Jangid, 19, Rongrong Jia, 53, Cheng Jia, 30, Yingjie Jiang, 26, Felix Jaremo, 24, Zhe Kang, 26, Madhu Kiran, 27, Josef Krtitko, 50, Simiao Lai, 17, Xiangyu Lan, 33, Dongwook Lee, 34, Hyunjong Lee, 34, Seohyung Lee, 34, Hui Li, 17, Ming Li, 17, Wangkai Li, 49, Xi Li, 55, Xianxian Li, 57, Xiaolin Liu, 16, Zhe Li, 41, Liteng Lin, 37, Halbin Ling, 40, Bo Liu, 25, Chang Liu, 18, Si Liu, 23, Huchuan Liu, 18, Rafael M. O. Cruz, 27, Bingpeng Ma, 44, Chao Ma, 36, Jie Ma, 21, Yinchao Ma, 49, Niki Martine, 51, Alireza Memarmoghdam, 45, Christian Micheloni, 51, Payman Moallem, 45, Le Thanh Nguyen-Meidine, 27, Siyang Pan, 35, ChangBeom Park, 34, Danda Paude, 10, Matthieu Paul, 10, Houwen Peng, 28, Andreas Robinson, 4, Lito Rut, 39, Shiguang Shan, 24, Kristian Simonato, 51, Tianhui Song, 30, Xiaoning Song, 26, Chao Sun, 55, Jingna Sun, 16, Shengyong Tang, 26, Radu Timotheo, 10, Chi-Yi Tsai, 42, Luc Van Gool, 10, Om Prakash Verma, 19, Dong Wang, 18, Fei Wang, 49, Liang Wang, 13, Liangliang Wang, 16, Lijun Wang, 18, Limin Wang, 30, Qiang Wang, 35, Gangshan Wu, 30, Jinlin Wu, 13, Xiaojun Wu, 26, Fei Xie, 38, Tianyang Xu, 26, Wei Xu, 16, Yong Xu, 37, Yuanyou Xu, 55, Wanli Xue, 13, Zicheng Xun, 14, Bin Yan, 18, Dawei Yang, 49, Jinyu Yang, 41, Wankou Yang, 38, Xiaoyun Yang, 33, Yi Yang, 55, Yichun Yang, 30, Zongxin Yang, 55, Botao Ye, 24, Fisher Yu, 10, Hongyuan Yu, 13, Jiayin Yu, 13, Qianjin Yu, 49, Wei Chen, 13, Kang Ze, 26, Jiang Zhi, 38, Chengwei Zhang, 17, Chunhu Zhang, 36, Kaihua Zhang, 29, Tianzhu Zhang, 49, Wenkang Zhang, 38, Zhibin Zhang, 43, Zhipeng Zhang, 31, Jie Zhao, 18, Shaochuan Zhao, 26, Feng Zheng, 41, Haixia Zheng, 54, Min Zheng, 16, Bineng Zhong, 20, Jiawen Zhu, 18, Xuefeng Zhu, 26, and Yueting Zhuang, 55

• **VOT2022 sponsor:**

[Image of sponsor logo]

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