



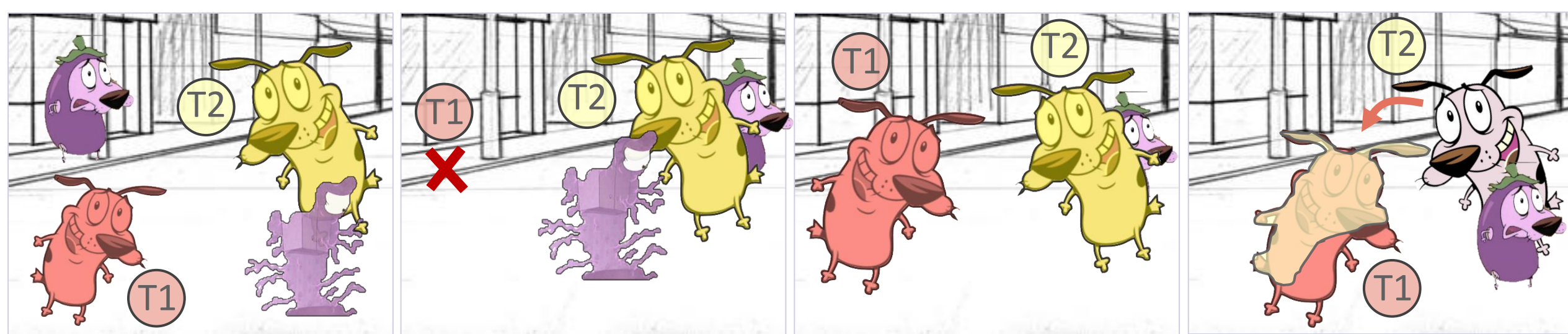
# The 4<sup>th</sup> VOTS2026

## Tracking by Segmentation Challenge Results

**Matej Kristan**, Jiří Matas, Pavel Tokmakov, Adam Harley, Alan Lukežič, Michael Felsberg, Luka Čehovin Zajc, Khanh-Tung Tran, Anurag Bagchi, Matjaž Zupančič Muc, Xuan-Son Vu, Johanna Björklund, Michal Neoral, Hyung Jin Chang, Gustavo Fernández

# The VOTS2026 challenge scope

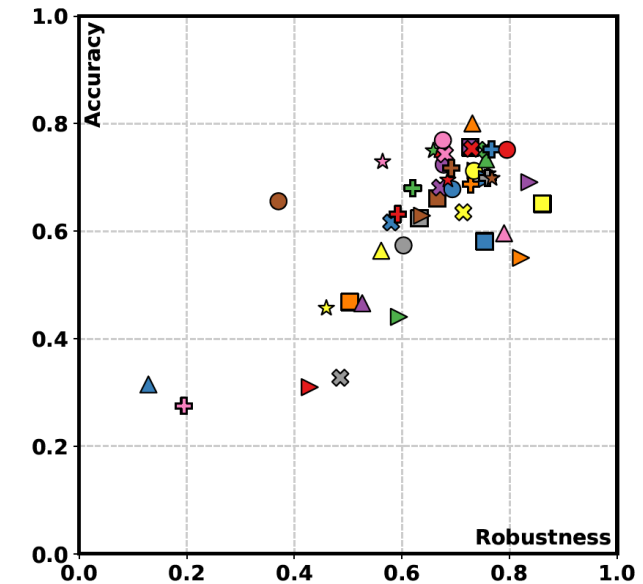
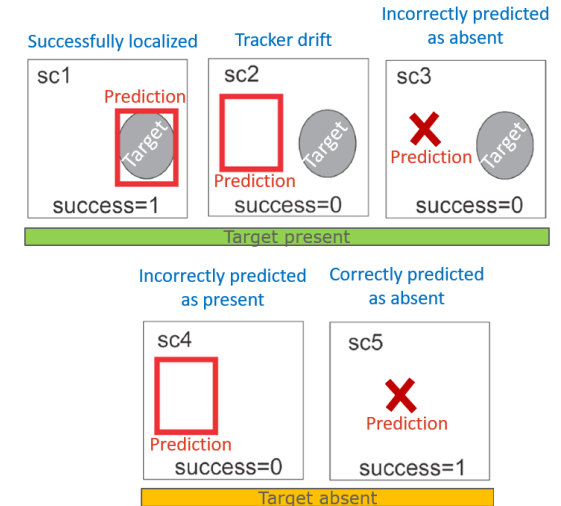
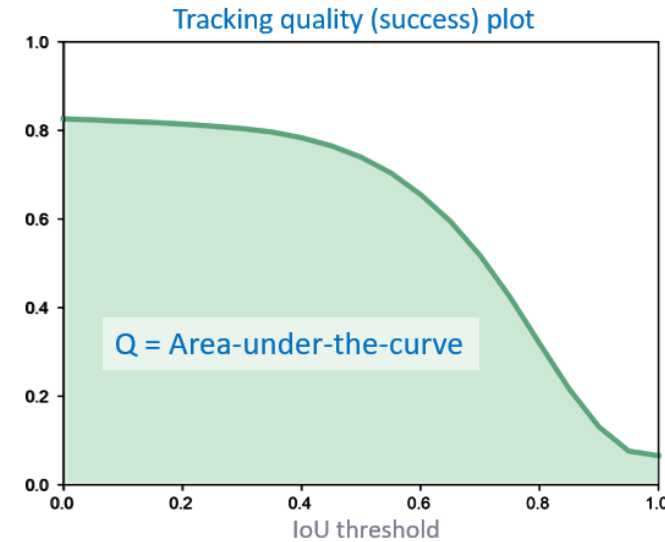
- General object tracking by segmentation
- Initialize on all targets in the first frame and report position in the rest



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

# VOTS performance measures

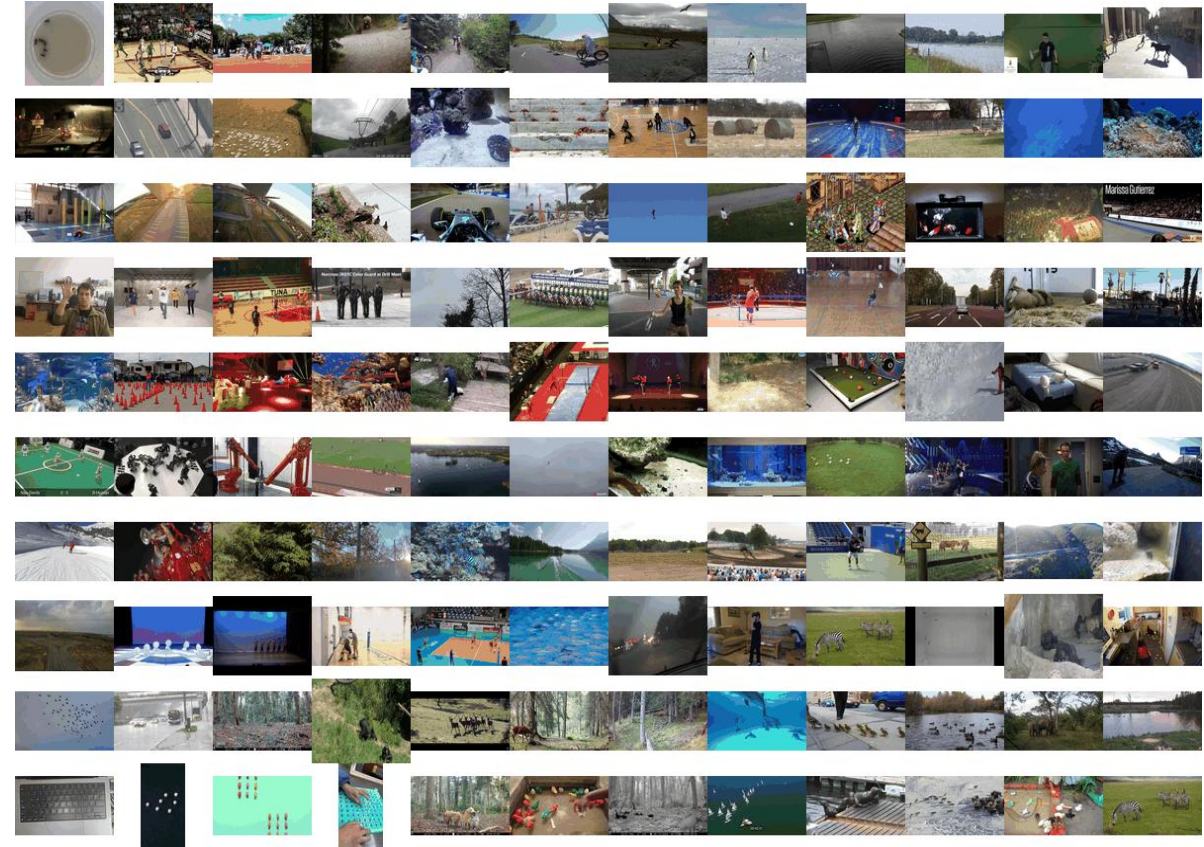
- Introduced in VOTS2023<sup>[1]</sup>
- Primary: **Tracking quality,  $Q$** 
  - (i) localization, (ii) re-detection,
  - (iii) target absence identification
- Auxiliary performance measures:
  - Accuracy/Robustness**
  - “Why did the tracker fail while target visible?”*  
(Not Reported Error, **NRE** ; Drift Rate Error, **DRE**)
  - “How well is target absence determined?”*  
(Absence Detection Quality, **ADQ**)



<sup>1</sup>Kristan et al., VOTS2023

# VOTS2026 dataset refreshed from VOTS2025

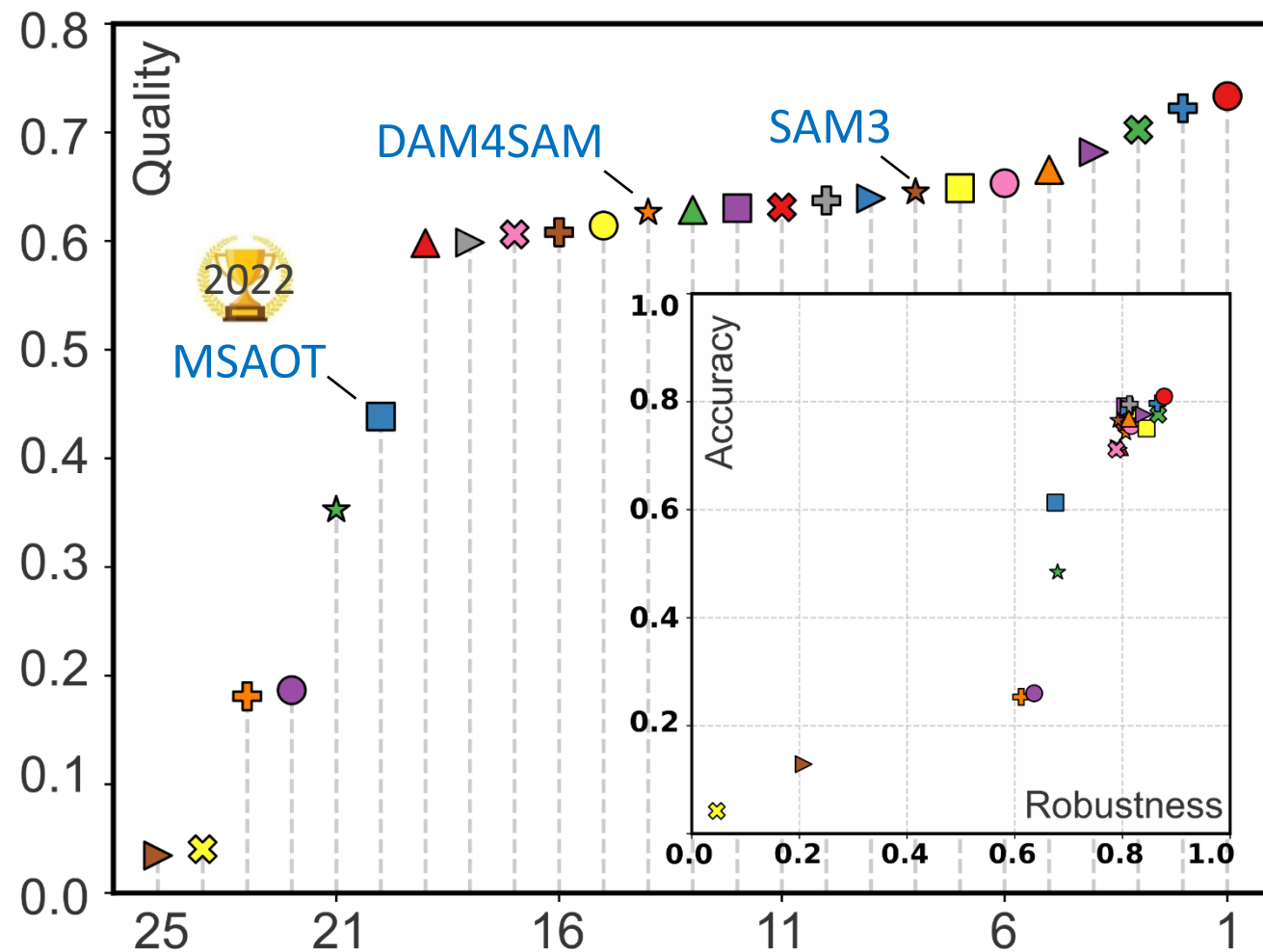
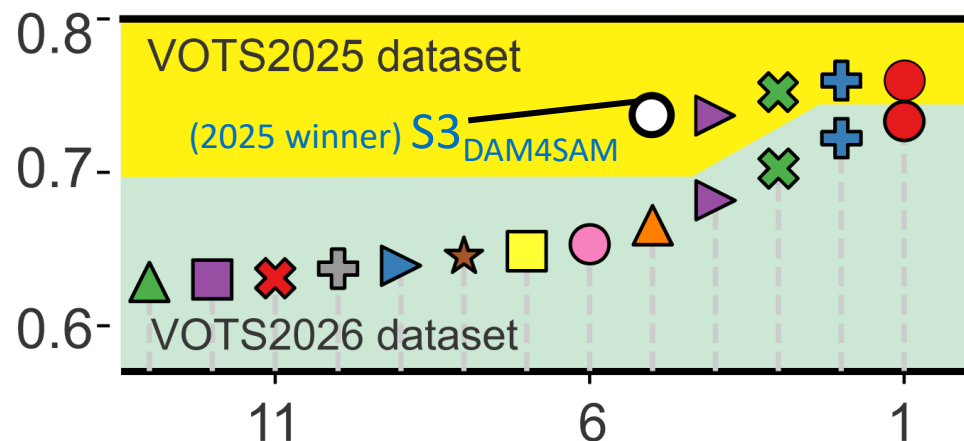
- 50 sequences removed, 26 added
- Stats: **120 sequences**; **271 targets**, of these **158** leave the FOV (med 2, max 23)
- Sequence properties:
  - Min/Max = **63/10.7k frames** ;
  - 2.3 targets** per sequence ;
  - Median **target absence**: **20 frames**
- To prevent overfitting:
  - Sequences + initialization frames GT **publicly available**.
  - **GT of test frames sequestered**, evaluation carried out on a **dedicated server**.
  - **During the challenge**, results reported for 2025 dataset only.



**Added from:** AnimalTrack: 6 ; ChimpACT: 2 ; DiDi: 1 ;  
GMOT40: 5 ; MammAlps: 2 ; MOSSEv2: 7 ; SA-FARI: 3

# VOTS2026 challenge results: 25 trackers tested

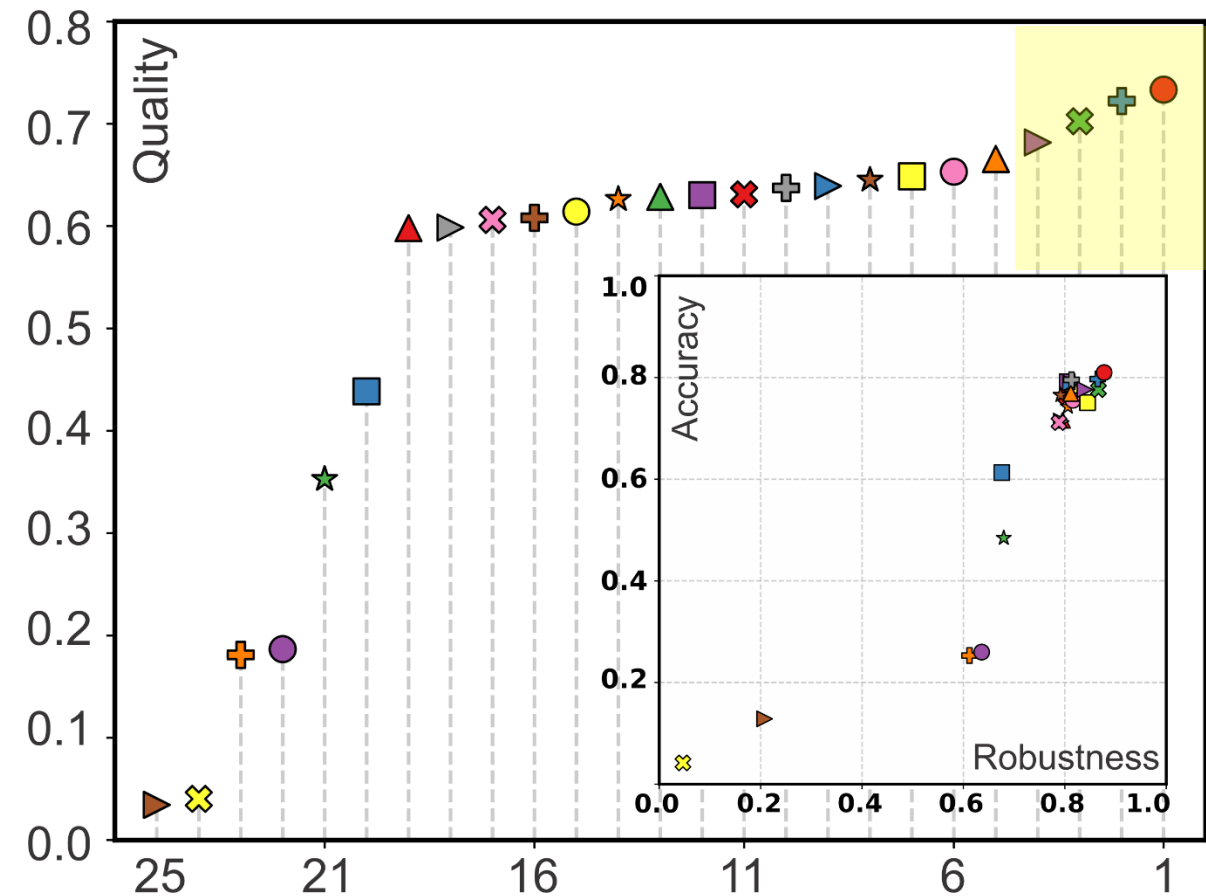
- Baseline 1: VOT2022 winner MSAOT<sup>[1]</sup>  
(76% outperform it)
- Baseline 2: DAM4SAM<sup>[2]</sup>  
(52% outperform it)
- Baseline 3: SAM3<sup>[3]</sup>  
(28% outperform it)
- On **VOTS2025** dataset:



<sup>1</sup>Yang, et al. NeurIPS 2021 ; <sup>2</sup>Videnvić, et al. CVPR 2025, <sup>3</sup>Carion, et al. ICRL 2026

# VOTS2026 challenge top 4 trackers

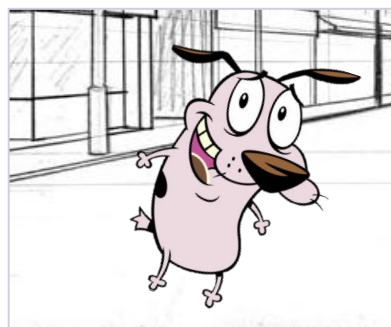
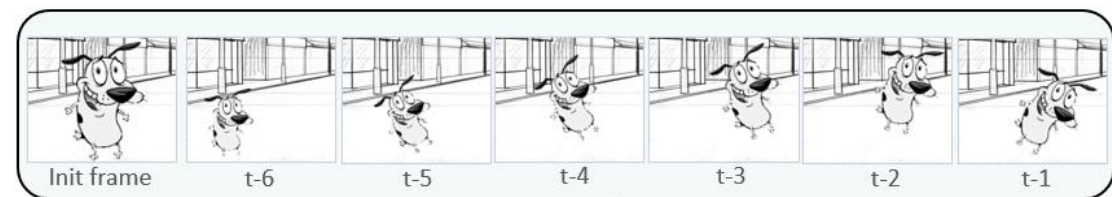
- Top 4 trackers: VLM-SAFEDAM, SAFEDAM, ACTrack, SAM3Dual
- All memory-based VOS methods
- VLM-SAFEDAM ● / SAFEDAM +:
  - SAM3 + DAM4SAM + VLM + distractors
- ACTrack ✕:
  - Agentic tracker (tool use)  
SAM3 + VLM prompting
- SAM3Dual ▶:
  - SAM3 + with DAM4SAM-like memory ; multi-memory



# VLM-SAFEDAM, SAFEDAM

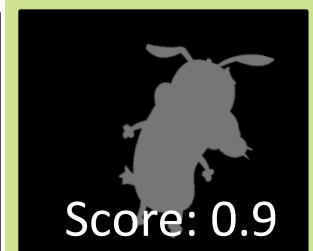
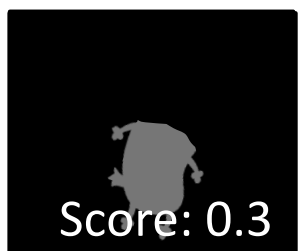
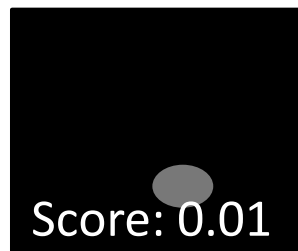
SAM2/3 [ICLR2025/ICLR2026]

A single FIFO memory



SAM3

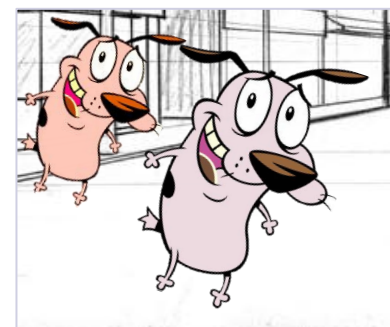
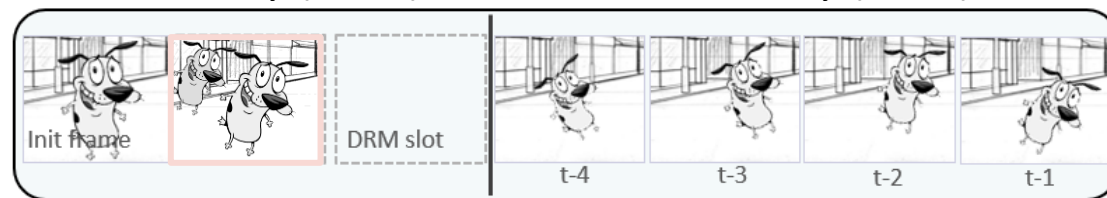
FIFO



DAM4SAM [CVPR2025]

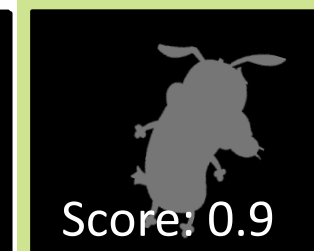
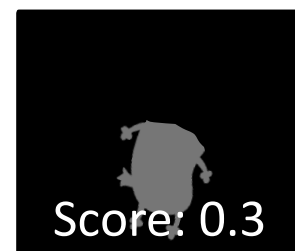
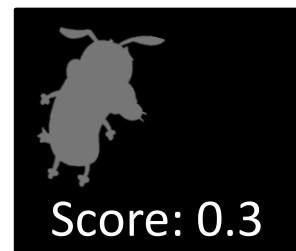
Distractor resolving  
memory (DRM)

Recent appearance  
memory (RAM)



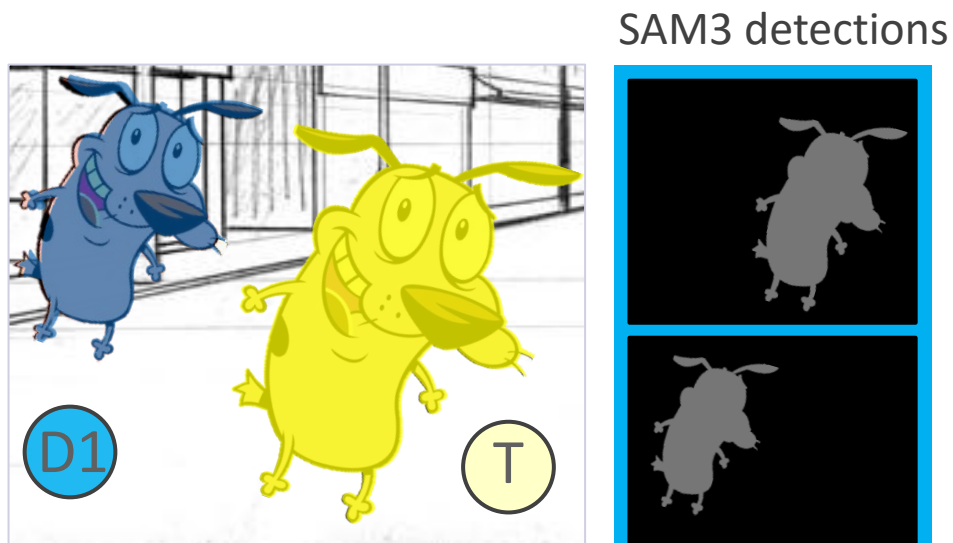
SAM2

FIFO  
update  
DRM

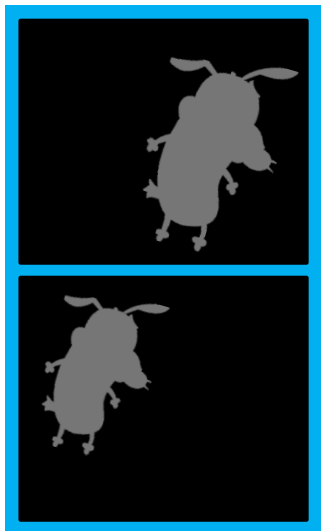


# VLM-SAFEDAM, SAFEDAM

(1) Distractor presence detected:



SAM3 detections

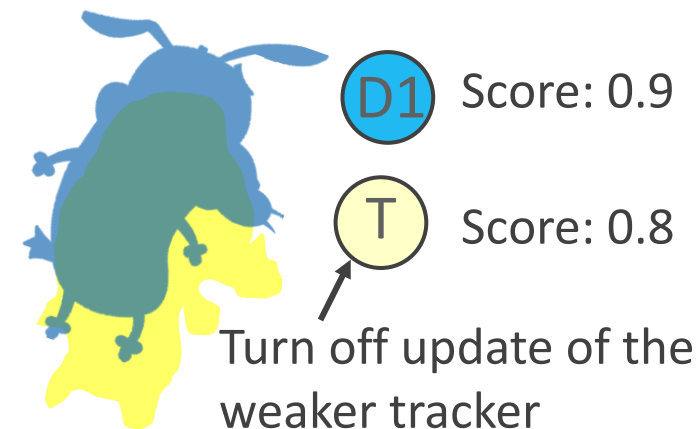
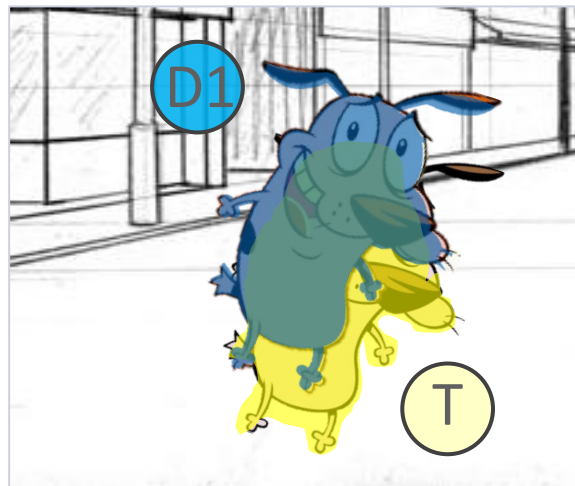


Predicted SAM3 masks:

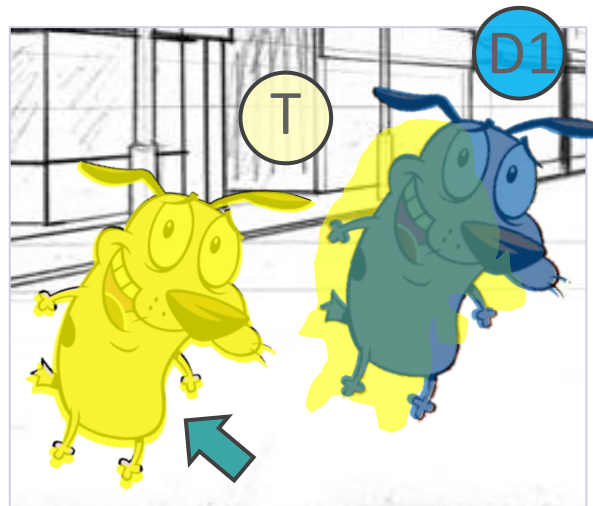


*Run SAM3 detector & init a new tracker on the distractor.*

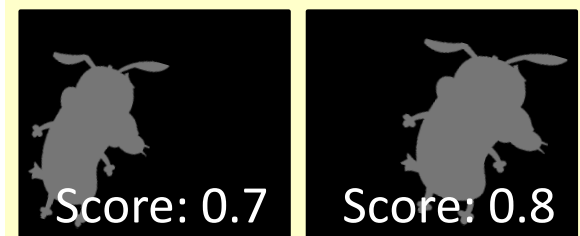
(2) If sufficient overlap between T & D1:



(3) Select alternative mask if strong enough:



Predicted SAM3 masks (T):



# VLM-SAFEDAM, SAFEDAM

- 1<sup>st</sup> frame: Qwen3.6-27B-FP8  
check for numbers/letters
  - NO → SAFEDAM;
  - YES → SAFEDAM+VLM

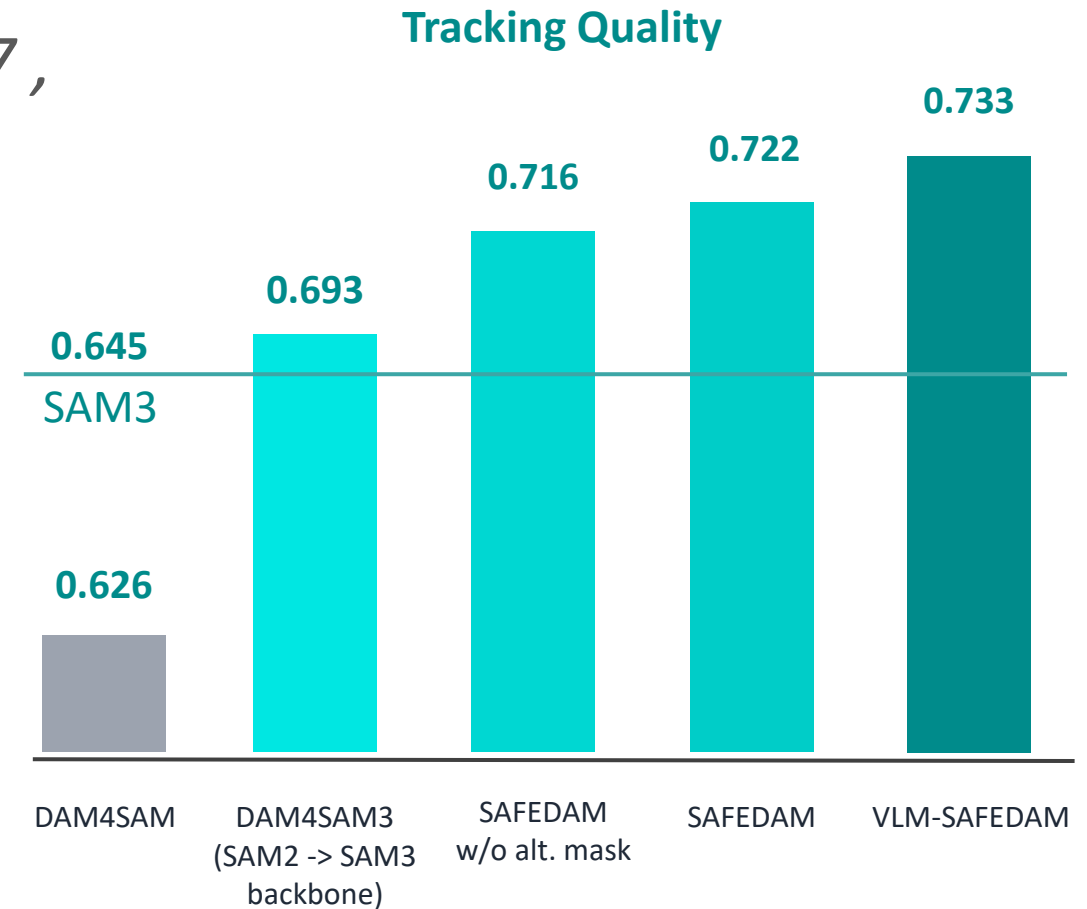


If 1<sup>st</sup> frame check passed (~2% of sequences on VOTS2026):

- Periodically run VLM on target region to verify it's the same target
- If verification failed or target mask empty:
  - Run VLM globally ; detect target by a bounding box
  - Convert box to mask by SAM3 prompting + VLM re-verification
  - Update the SAFEDAM memory if VLM verification is passed

# VLM-SAFEDAM on VOTS2026

- *Q-score* = 0.733 (~4% improvement over ACTrack)
- *VOTS aux*:  $A=0.81$ ,  $R=0.878$ ,  $DRE=0.057$ ,  $NRE=0.066$ ,  $ADQ=0.805$
- Good segmentation accuracy 81%, on-target in ~87% of frames
- Localization failure ~13% due to:
  - Falsely reported target as absent: 7%
  - Drifting: 6%
- Accurately predicts target absence in 81% cases



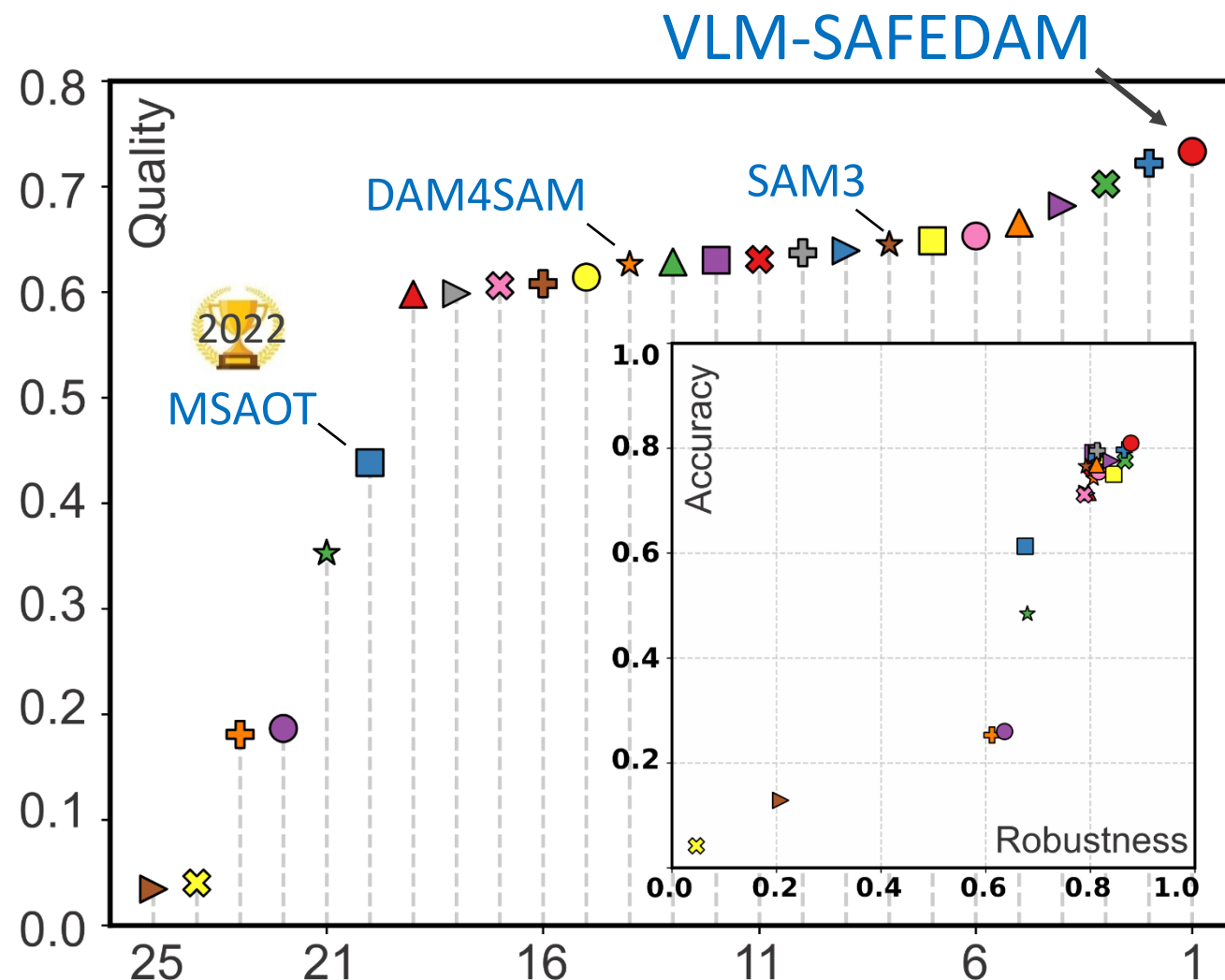


VOTS2026 challenge  
Winners:

VLM-SAFEDAM by:  
Dayu Wu, Yucheng Bai, Lv Wu, Han  
Fang

“VLM-DAM4SAM3 with anti-overlap mechanism”

(Full presentation available on VOTS2026 program page)



# Summary

---

- Strongest submissions all SAM3-based
- The gains come from: distractor modeling, identity verification, memory management
- VLMs-based & Agentic trackers emerging
- Benchmark opened for post-challenge evaluation  
Available: VOTS2026 35 trackers ; VOTS2025 134 trackers



# Thanks

- Everyone who participated or contributed

Matej Kristan<sup>1</sup>, Jiří Matas<sup>2</sup>, Pavel Tokmakov<sup>3</sup>, Adam Harley<sup>4</sup>, Michael Felsberg<sup>5</sup>, Luka Čehovin Zajc<sup>1</sup>, Alan Lukežič<sup>1</sup>, Khanh-Tung Tran<sup>6</sup>, Anurag Bagchi<sup>7</sup>, Matjaž Zupančič<sup>1</sup>, Muc<sup>1</sup>, Xuan-Son Vu<sup>8,9</sup>, Johanna Bjorklund<sup>6</sup>, Michal Neoral<sup>2</sup>, Hyung Jin Chang<sup>10</sup>, Gustavo Fernandez<sup>11</sup>, Mohamad Alansari<sup>25</sup>, Hasan AlMarzouqi<sup>25</sup>, Jeffrey Appiagyei<sup>37</sup>, Michael Arens<sup>18</sup>, Suleyman Aslan<sup>15</sup>, Minasadat Attari<sup>37</sup>, Gorkay Aydemir<sup>15</sup>, Volkan Aydingul<sup>17</sup>, Yucheng Bai<sup>30</sup>, Bishnu Bhusal<sup>37</sup>, Burak Can Biner<sup>15</sup>, Wenrui Cai<sup>13</sup>, Marco Cocco<sup>39</sup>, Zhenglin Du<sup>41</sup>, Matteo Dunnhofer<sup>39,42</sup>, Ahmet Rasim Emirdagi<sup>15</sup>, Han Fang<sup>30</sup>, Karthikeya Gadiparthi<sup>27</sup>, Alexandru-Marian Galea<sup>34</sup>, Rama Krishna Sai Gorthi<sup>23</sup>, Ziqi Gu<sup>26</sup>, Xiantao Hu<sup>26</sup>, XingSen Huang<sup>19</sup>, Sajid Javed<sup>25</sup>, Alisher Kendzhaev<sup>22</sup>, Chaehyun Kim<sup>14</sup>, JeongRae Kim<sup>14</sup>, Wooseok Kim<sup>14</sup>, Josef Kittler<sup>38</sup>, Yunus Bilge Kurt<sup>15</sup>, Hyeongyeong Lee<sup>14</sup>, Daoheng Li<sup>26</sup>, Hao Li<sup>16</sup>, Kunyu Li<sup>40</sup>, Xin Li<sup>28</sup>, Yuanjia Li<sup>24</sup>, Yuzhe Li<sup>13</sup>, Zhengyang Li<sup>41</sup>, Zixu Li<sup>26</sup>, Changwon Lim<sup>14</sup>, Qingjie Liu<sup>13</sup>, Deshui Miao<sup>20</sup>, Yonathan Michael<sup>25</sup>, Christian Micheloni<sup>39</sup>, Juan Mogollon<sup>37</sup>, Zahrasadat Mostafavi Ghahfarokhi<sup>37</sup>, Muzammal Naseer<sup>25</sup>, Kannappan Palaniappan<sup>37</sup>, Jianjun Qian<sup>26</sup>, Nasrin Rahimi<sup>15</sup>, Gani Rahmon<sup>37</sup>, Aleksandr Romanov<sup>22</sup>, Norbert Scherer-Negenborn<sup>18</sup>, Liangtao Shi<sup>21</sup>, Arunkumar Sivapuram<sup>33</sup>, Roman Solovyev<sup>12</sup>, Elham Soltani Kazemi<sup>37</sup>, Zhangyong Tang<sup>24</sup>, Imad Eddine Toubal<sup>37</sup>, Gavriil Tyukhtyaev<sup>22</sup>, Stephane Vujasinovic<sup>18</sup>, Fei Wang<sup>29</sup>, He Wang<sup>24</sup>, Yu-Xiong Wang<sup>36</sup>, Yi Wen<sup>41</sup>, Naoufel Werghi<sup>25</sup>, Dayu Wu<sup>30</sup>, Lv Wu<sup>30</sup>, Xiao-Jun Wu<sup>24</sup>, Jinxia Xie<sup>26</sup>, Tianyang Xu<sup>24</sup>, Ming-Hsuan Yang<sup>35</sup>, Misra Yavuz<sup>15</sup>, M. Akin Yilmaz<sup>15</sup>, Jiahao Zeng<sup>32</sup>, Haijun Zhang<sup>20</sup>, Lichao Zhang<sup>31</sup>, Zheyu Zhang<sup>36</sup>, Xu Zhao<sup>26</sup>, and Xue-Feng Zhu<sup>24</sup>

- VOTS2026 sponsors:



UNIVERSITY OF  
BIRMINGHAM