SAM2RL: Towards Reinforcement Learning Memory Control in Segment Anything Model 2

Alen Adamyan[1,2,†], **Tomáš Čížek**[1,†], **Matej Straka**[1], **Klara Janouskova**[2], **Martin Schmid**[1,3]
alen.am0161@gmail.com, {cizek,schmidm}@kam.mff.cuni.cz, matej.straka825@student.cuni.cz, klara.janouskova@fel.cvut.cz

[1]**Department of Applied Mathematics, Charles University, Prague, Czech Republic**
[2]**Visual Recognition Group, Czech Technical University in Prague**
[3]**EquiLibre Technologies, Inc.**

[†] Equal contribution.

Abstract

Segment Anything Model 2 (SAM 2) has demonstrated strong performance in object segmentation tasks and has become the state-of-the-art for visual object tracking. The model stores information from previous frames in a memory bank, enabling temporal consistency across video sequences. Recent methods augment SAM 2 with hand-crafted update rules to better handle distractors, occlusions, and object motion. We propose a fundamentally different approach using reinforcement learning for optimizing memory updates in SAM 2 by framing memory control as a sequential decision-making problem. In an overfitting setup with a separate agent per video, our method achieves a relative improvement over SAM 2 that exceeds by more than three times the gains of existing heuristics. These results reveal the untapped potential of the memory bank and highlight reinforcement learning as a powerful alternative to hand-crafted update rules for memory control in visual object tracking.

1 Introduction

Reinforcement learning (RL) is well-suited for sequential decision-making tasks in dynamic environments by learning from delayed rewards to optimize long-term returns. *Visual Object Tracking* (VOT) poses precisely this challenge, requiring a tracker to localize a target in a continuous video stream while adapting to appearance changes, occlusions, and distractions over time. These challenges align naturally with the strengths of RL in learning adaptive policies under uncertainty.

The current state-of-the-art in visual object tracking, *Segment Anything Model 2* (SAM 2) (Ravi et al., 2024), integrates a promptable segmentation framework with a memory bank. The memory bank always retains memories from the most recent frames to maintain temporal consistency. However, this fixed update rule overlooks each frame's relevance to the tracking quality. Building upon SAM 2, several recent studies have proposed heuristic enhancements to this memory bank update mechanism, leading to notable performance improvements. These heuristics target specific scenarios where SAM 2 underperforms, such as motion-aware selection (Yang et al., 2024; 2025) and distractor-aware updates (Videnovic et al., 2024).

Rather than hand-crafting these update rules, we propose to learn them directly by framing the SAM 2 memory updates as a reinforcement learning task. We estimate the upper-bound performance achievable with learned memory updates by overfitting an RL agent (SAM2RL) to a set of

videos individually. In this overfitting setup, SAM2RL achieves more than three times the performance gains of prior methods over the original SAM 2, highlighting the potential of RL to improve SAM 2 memory bank updates. This improvement could set a new state-of-the-art for VOT without modifying any trainable parameters of the original SAM 2.

The paper is organized as follows. In Section 2, we define the visual tracking task and review relevant prior work. In Section 3, we describe the SAM 2 architecture and formulate memory bank updates as a reinforcement learning task. In Section 4, we overfit our policy on a set of videos and compare its performance against prior methods. Finally, Section 5 concludes with a discussion of our findings and outlines directions for future work.

2 Related Work

Vision Object Tracking (VOT) is the task of locating and following a specific object over time in a video sequence, given its initial position in the first frame (Marvasti-Zadeh et al., 2022). Traditionally, this task has been approached using bounding boxes to represent the target object, offering a simple yet effective way to estimate object location and scale. In recent years, the field has shifted toward more precise localization methods, with tracking-by-segmentation becoming the dominant approach. Common challenges of VOT include occlusion, motion blur, and appearance changes.

Segment Anything Model 2 (SAM 2) (Ravi et al., 2024) extends the promptable transformer architecture of *Segment Anything Model* (SAM) (Kirillov et al., 2023) with a streaming memory mechanism, enabling unified image and video segmentation. Leveraging large-scale pretraining, it enables real-time zero-shot tracking across diverse domains, such as self-driving, video editing, or wildlife monitoring. SAM 2's fixed-length memory bank always stores only the most recent frames, suffering from redundancy and irrelevant content. To address these limitations, several recent extensions propose improved memory update mechanisms using hand-crafted update rules.

SAM2Long (Ding et al., 2024) mitigates error accumulation in long videos by using multiple parallel memory banks, improving occlusion handling and object reappearance at increased computational cost. *DAM4SAM* (Videnovic et al., 2024) enhances distractor handling with distractor-aware update rules. *SAMURAI* (Yang et al., 2024) incorporates motion-aware memory selection via a *Kalman filter* (Kalman, 1960) for refined mask selection, while *MoSAM* (Yang et al., 2025) introduces motion-guided prompting and enhances the memory bank with spatial-temporal memory selection for accurate predictions.

3 Method

In this section, we formulate memory selection in SAM 2 as a reinforcement learning task, where an agent learns a memory update policy to optimize tracking performance. We first give a brief overview of the SAM 2 architecture, and then describe how memory selection is framed as a reinforcement learning task.

SAM 2 Architecture. SAM 2 achieves temporally consistent segmentation of a target object by combining per-frame image embeddings with a memory-attention mechanism. As illustrated in Figure 1, each incoming frame is first processed by the *image encoder* to produce a feature embedding. This embedding is then conditioned on *memories* inside the *memory bank* using the *memory attention* and also on *tracking prompts*, such as the input mask of the first frame. The resulting conditioned embedding is passed into the *mask decoder*, which generates the final segmentation mask. The *memory encoder* then transforms the predicted mask and the corresponding image embedding into a *memory feature*. This memory feature is added to the memory bank with fixed capacity $N = 7$, which always retains the initial frame along with the most recent $N - 1$ memories.

Memory Selection via Reinforcement Learning. We frame the SAM 2 memory selection problem as a reinforcement learning task. The tracking of a single video object forms an episode in the

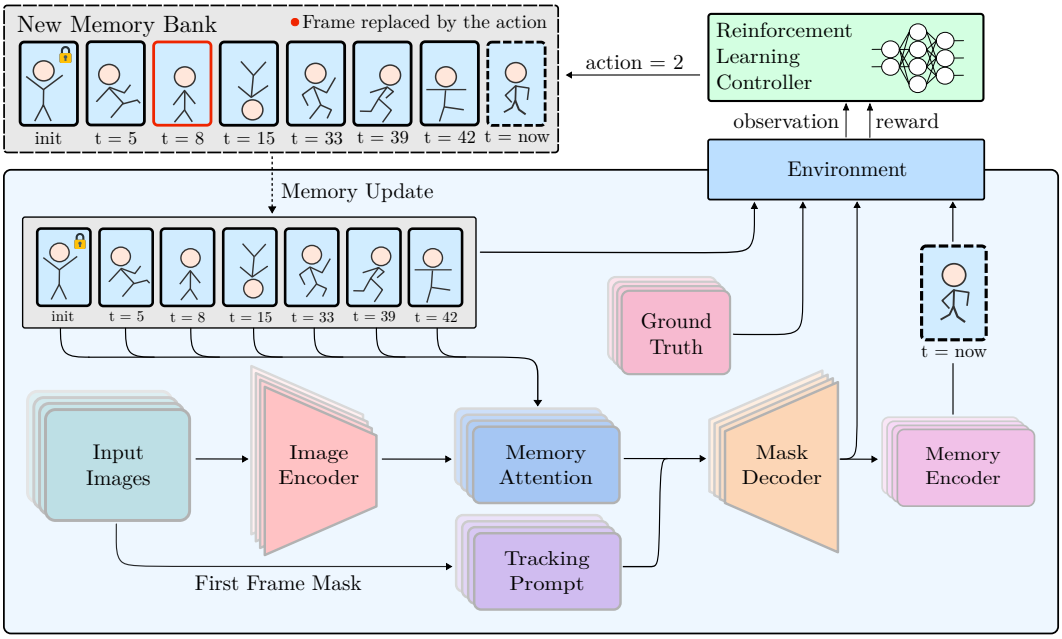

Figure 1: In SAM2RL, the image encoder processes a new frame, which is then conditioned by the memory bank of past predictions and tracking prompts. The mask decoder uses this representation to generate a segmentation mask. This mask is used to provide a reward to the RL memory controller and also by the memory encoder to create a new memory for the current frame. The controller then observes memories and decides which memory frames to retain.

RL environment operated by an agent who performs memory bank updates. At each timestep, the incoming input image is first processed by the frozen SAM 2 pipeline, producing the predicted mask and the encoded memory feature. The agent then receives *observation* consisting of the representation of the stored memories in the memory bank and the memory feature of the current frame. Based on this observation, the agent selects an *action* that either discards the incoming memory feature or inserts it into the memory bank by replacing one of the $N-1$ stored memories. The objective is to learn a memory update policy that optimizes the *tracking quality* Q, which is the primary tracking measure of the VOTS challenge (Kristan et al., 2023). The tracking quality is measured as the mean *intersection-over-union* (IoU) on frames with a visible object, with a penalty for incorrect prediction when the object is not present. Formally, the tracking quality Q for a video of length T with a predicted mask $\hat{M}_t$ and a ground truth mask M_t at timestep t is defined as

$$Q = \frac{1}{T}\sum_{t=1}^{T} q_t, \quad \text{where} \quad q_t = \begin{cases} 1, & \hat{M}_t = \emptyset \text{ and } M_t = \emptyset, \\ \mathrm{IoU}(\hat{M}_t, M_t), & \text{otherwise.} \end{cases} \tag{1}$$

To guide the learning process towards maximizing the tracking quality Q, we set the *reward* at timestep t to q_t. The final episode *return* to maximize is thus equal to $\sum_{t=1}^{T} \gamma^{t-1} q_t$ with $\gamma \in (0, 1]$.

4 Evaluation

The memory bank is a relatively small component of the overall SAM 2 architecture. Therefore, it remains unclear how much of the tracking performance, even on the training dataset, can be improved by optimizing only the memory update policy. In this section, we explore the power of the memory bank and its update policy by overfitting a sequence of videos individually using a reinforcement learning algorithm, as described in Section 3. We compare the obtained results with those achieved by previous methods using hand-crafted update rules to show how much of the capacity is effectively utilized by the current methods.

Table 1: Comparison of SAM 2-based methods and the proposed SAM2RL with an overfitted memory bank policy, evaluated on 64 videos from the SA-V training set using standard VOTS metrics.

Method	**Quality [%]**	**Accuracy [%]**	**Robustness [%]**
SAM 2	71.95	73.01	88.84
SAMURAI	-4.70 67.25	-0.30 72.71	+1.83 90.67
DAM4SAM	+1.60 73.55	-0.05 72.96	+0.36 89.20
SAM2RL (ours)	**+4.91 76.86**	**+2.52 75.53**	**+2.33 91.17**

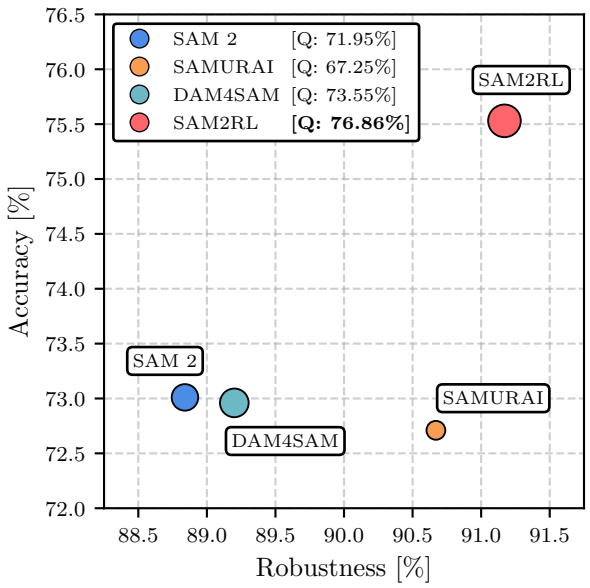

Figure 2: Accuracy vs. robustness plot comparing the performance of SAM 2-based models and SAM2RL, with marker size indicating the tracking quality.

Training Setup. Our experiments are designed to intentionally overfit individual videos, without any expectation that the learned policy will generalize beyond them. To this end, we represent the agent's observation using a simple binary vector that uniquely encodes both the indices of the image frames stored in the memory bank and the current timestep. The policy network is then a two-layer multilayer perceptron (MLP), with an input size equal to the video length, a hidden layer of 1024 units, and a softmax output over the N possible actions. We train the network using *Proximal Policy Optimization* (PPO) (Schulman et al., 2017) separately on 64 random videos from the *SA-V training dataset* (Ravi et al., 2024), with an average length of 80 frames recorded at 6 frames per second. All evaluated models employ the largest available *Hiera-L* image encoder (Ryali et al., 2023) with input resolution 256×256 and SAM 2.1 checkpoint. We always track only the first object in a video.

We evaluate the environment and train the policy network for each video using a single NVIDIA A100 40GB GPU. The environment simulation runs at approximately 750 frames per second. In each PPO iteration, we collect a dataset of 16k samples and train the policy for 2 epochs. The training process runs for 150 iterations and takes approximately 1 hour on average per video.

Quantitative Results. We compare the tracking performance of the proposed SAM2RL with SAM 2 and its two best-performing publicly available variants, SAMURAI (Yang et al., 2024) and DAM4SAM (Videnovic et al., 2024). We evaluate the tracking performance in Table 1 using three standard metrics from the VOTS challenge (Kristan et al., 2023). The tracking quality Q

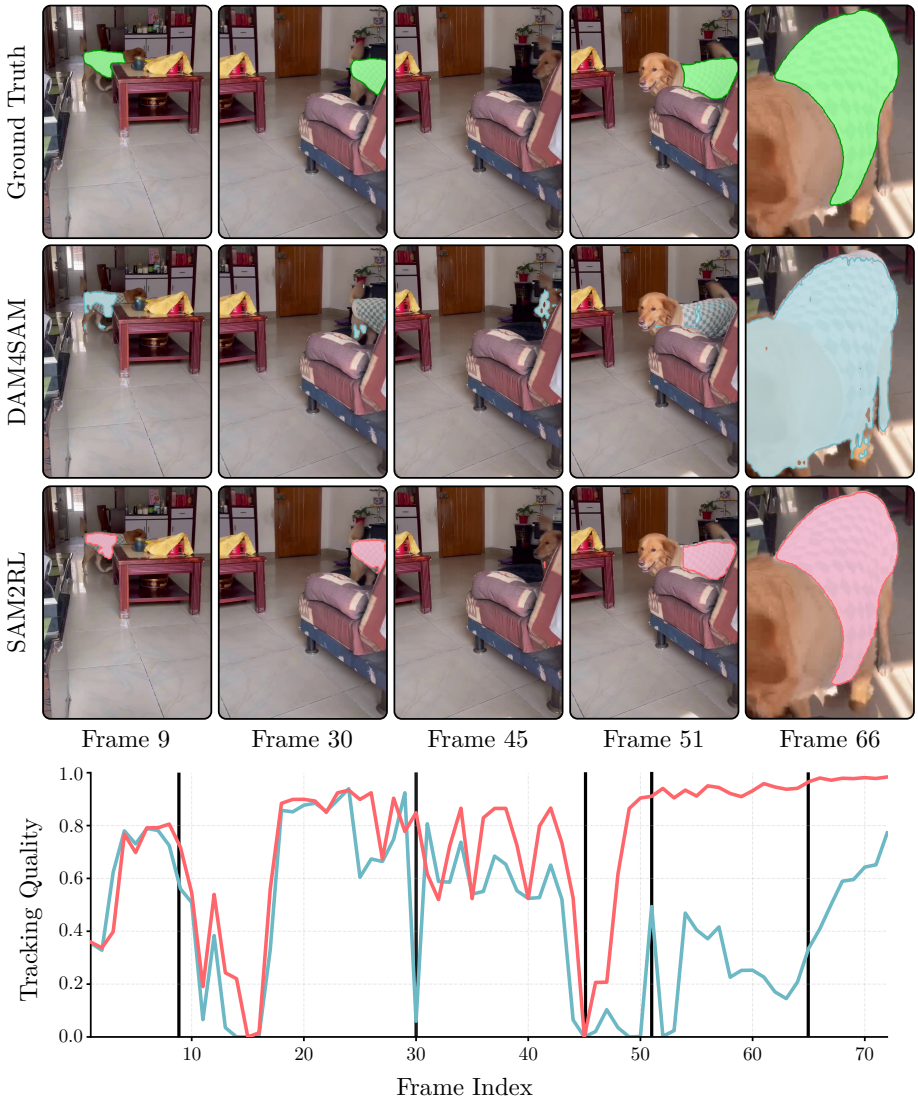

Figure 3: Visualization of predicted masks and the tracking quality of SAM2RL and DAM4SAM on a single selected video. SAM2RL produces masks that more closely match the ground truth, and successfully recovers the object after its disappearance, while DAM4SAM often fails to detect it.

from Equation (1) is the primary performance measure optimized by the RL policy. The *accuracy* and *robustness* are two additional metrics to evaluate mask quality on frames with a visible object only. The accuracy is equal to the mean IoU in frames with nonzero IoU, and the robustness is the proportion of frames with nonzero IoU in frames with a visible object. The accuracy and robustness thus describe the trade-off between the mask precision and the chance of detecting an object with at least a minimal overlap. We visualize all the metric values in Figure 2.

SAM2RL achieves the highest performance in all three VOTS metrics, surpassing SAM 2 and both of its enhanced variants. In particular, SAM2RL improves the tracking quality by +4.91%, accuracy by +2.52%, and robustness by +2.33% compared to SAM 2. In contrast, SAMURAI, in our training setup, underperforms relative to SAM 2 in tracking quality metric by -4.70%, which we attribute to its tendency to predict a mask even when the object is not visible, trading off for a modest gain in robustness by +1.83%. DAM4SAM shows a moderate improvement over SAM 2 with a +1.60% increase in tracking quality, but it still falls short compared to SAM2RL. These results indicate

that optimizing the memory bank's update policy can lead to substantial improvements in tracking performance, exceeding those achieved by current methods.

Qualitative Results. Figure 3 presents a qualitative comparison of the predicted masks generated by SAM2RL and DAM4SAM in a selected video from the SA-V dataset. DAM4SAM is used as a baseline, as it outperforms both SAM 2 and SAMURAI on this specific video. SAM2RL produces masks that match the ground truth more accurately than those from DAM4SAM. In particular, after the tracked object disappears completely in frame 45, SAM2RL successfully recovers object tracking, whereas DAM4SAM either fails to detect the object completely or produces inaccurate masks. On the other hand, we observe that although SAM2RL is deliberately overfitted on this video, it still fails to achieve optimal performance, suggesting the limit of the memory bank capacity.

5 Conclusion

We introduced the first reinforcement learning approach to optimize memory update policies within SAM 2 for visual object tracking. Although the memory bank constitutes only a small component of the overall SAM 2 architecture, the significant performance gains of SAM2RL in an overfitting setup demonstrate that optimizing the update policy can greatly enhance the tracking quality. This highlights an opportunity to improve the current hand-crafted update rules using reinforcement learning-based update policies. We plan to achieve these improvements by scaling our approach through the integration of richer observations and suitable network architectures, enabling us to learn policies that would generalize beyond the training data, and thus push the current state-of-the-art in video object tracking.

Acknowledgments

Tomáš Čížek and Martin Schmid were supported by the grant no. 25-18031S of the Czech Science Foundation (GAČR). Martin Schmid was also supported by Center for Foundations of Modern Computer Science (Charles Univ. project UNCE 24/SCI/008). Klara Janouskova received support from Toyota Motor Europe, and Alen Adamyan from CEDMO 2.0 NPO (MPO 60273/24/21300/21000). Computational resources were provided by the e-INFRA CZ project (ID:90254), supported by the Ministry of Education, Youth and Sports of the Czech Republic.

References

Shuangrui Ding, Rui Qian, Xiaoyi Dong, Pan Zhang, Yuhang Zang, Yuhang Cao, Yuwei Guo, Dahua Lin, and Jiaqi Wang. Sam2Long: Enhancing Sam 2 for Long Video Segmentation with a Training-Free Memory Tree. *arXiv preprint arXiv:2410.16268*, 2024.

Rudolph Emil Kalman. A New Approach to Linear Filtering and Prediction Problems, 1960.

Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment Anything. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 4015–4026, 2023.

Matej Kristan, Jiří Matas, Martin Danelljan, Michael Felsberg, Hyung Jin Chang, Luka Čehovin Zajc, Alan Lukežič, Ondrej Drbohlav, Zhongqun Zhang, Khanh-Tung Tran, et al. The First Visual Object Tracking Segmentation VOTS2023 Challenge Results. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 1796–1818, 2023.

Seyed Mojtaba Marvasti-Zadeh, Li Cheng, Hossein Ghanei-Yakhdan, and Shohreh Kasaei. Deep Learning for Visual Tracking: A Comprehensive Survey. *IEEE Transactions on Intelligent Transportation Systems*, 23(5):3943–3968, 2022. DOI: 10.1109/TITS.2020.3046478.

Nikhila Ravi, Valentin Gabeur, Yuan-Ting Hu, Ronghang Hu, Chaitanya Ryali, Tengyu Ma, Haitham Khedr, Roman Rädle, Chloe Rolland, Laura Gustafson, et al. Sam 2: Segment Anything in Images and Videos. *arXiv preprint arXiv:2408.00714*, 2024.

Chaitanya Ryali, Yuan-Ting Hu, Daniel Bolya, Chen Wei, Haoqi Fan, Po-Yao Huang, Vaibhav Aggarwal, Arkabandhu Chowdhury, Omid Poursaeed, Judy Hoffman, Jitendra Malik, Yanghao Li, and Christoph Feichtenhofer. Hiera: A Hierarchical Vision Transformer without the Bells-and-Whistles. *arXiv preprint arXiv:2306.00989*, 2023.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal Policy Optimization Algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

Jovana Videnovic, Alan Lukezic, and Matej Kristan. A Distractor-Aware Memory for Visual Object Tracking with SAM2. *arXiv preprint arXiv:2411.17576*, 2024.

Cheng-Yen Yang, Hsiang-Wei Huang, Wenhao Chai, Zhongyu Jiang, and Jenq-Neng Hwang. Samurai: Adapting Segment Anything Model for Zero-Shot Visual Tracking with Motion-Aware Memory. *arXiv preprint arXiv:2411.11922*, 2024.

Qiushi Yang, Yuan Yao, Miaomiao Cui, and Liefeng Bo. MoSAM: Motion-Guided Segment Anything Model with Spatial-Temporal Memory Selection. *arXiv preprint arXiv:2505.00739*, 2025.