

# Learning Dynamic Memory Networks for Object Tracking

Tianyu Yang<sup>[0000–0002–9674–5220]</sup> and Antoni B. Chan<sup>[0000–0002–2886–2513]</sup>

Department of Computer Science, City University of Hong Kong, Hong Kong, China  
ti anyyang8-c@my. ci tyu. edu. hk, abchan@ci tyu. edu. hk

**Abstract.** Template-matching methods for visual tracking have gained popularity recently due to their comparable performance and fast speed. However, they lack effective ways to adapt to changes in the target object’s appearance, making their tracking accuracy still far from state-of-the-art. In this paper, we propose a dynamic memory network to adapt the template to the target’s appearance variations during tracking. An LSTM is used as a memory controller, where the input is the search feature map and the outputs are the control signals for the reading and writing process of the memory block. As the location of the target is at first unknown in the search feature map, an attention mechanism is applied to concentrate the LSTM input on the potential target. To prevent aggressive model adaptivity, we apply gated residual template learning to control the amount of retrieved memory that is used to combine with the initial template. Unlike tracking-by-detection methods where the object’s information is maintained by the weight parameters of neural networks, which requires expensive online fine-tuning to be adaptable, our tracker runs completely feed-forward and adapts to the target’s appearance changes by updating the external memory. Moreover, unlike other tracking methods where the model capacity is fixed after offline training – the capacity of our tracker can be easily enlarged as the memory requirements of a task increase, which is favorable for memorizing long-term object information. Extensive experiments on OTB and VOT demonstrates that our tracker MemTrack performs favorably against state-of-the-art tracking methods while retaining real-time speed of 50 fps.

**Keywords:** Addressable Memory, Gated Residual Template Learning

## 1 Introduction

Along with the success of convolution neural networks in object recognition and detection, an increasing number of trackers [4, 13, 22, 26, 31] have adopted deep learning models for visual object tracking. Among them are two dominant tracking strategies. One is the tracking-by-detection scheme that online trains an object appearance classifier [22, 26] to distinguish the target from the background. The model is first learned using the initial frame, and then fine-tuned using the training samples generated in the subsequent frames based on the newly predicted bounding box. The other scheme is template matching, which adopts either the target patch in the first frame [4, 29] or the previous frame [14]































