Context-Aware Token Pruning and Discriminative Selective Attention for Transformer Tracking

Janani Kugarajeevan[1,1], Thanikasalam Kokul[1,*], Amirthalingam Ramanan[1], Subha Fernando[1]

[a]*Department of Computational Mathematics, University of Moratuwa, Sri Lanka*
[b]*Department of Interdisciplinary Studies, University of Jaffna, Sri Lanka*
[c]*Department of Computer Science, University of Jaffna, Sri Lanka*

Abstract

One-stream Transformer-based trackers have demonstrated remarkable performance by concatenating template and search region tokens, thereby enabling joint attention across all tokens. However, enabling a large number of background search tokens to attend to the target template tokens weakens the tracker's discriminative capability. Several token pruning methods have been proposed to mitigate background interference; however, they often remove tokens near the target, leading to the loss of essential contextual information and degraded tracking performance. Moreover, the presence of distractors within the search tokens further reduces the tracker's ability to accurately identify the target. To address these limitations, we propose CP-DATrack, a novel tracking framework designed to suppress interference from background and distractor tokens while enhancing computational efficiency. First, a learnable module is integrated between two designated encoder layers to estimate the probability of each search token being associated with the target. Based on these estimates, less-informative background tokens are pruned from the search region while preserving the contextual cues surrounding the target. To further suppress background interference, a discriminative selective attention mechanism is employed that fully blocks search-to-template attention in the early layers. In the subsequent encoder layers, high-probability

*Corresponding author

Email addresses: jananitha@univ.jfn.ac.lk (Janani Kugarajeevan), kokul@univ.jfn.ac.lk (Thanikasalam Kokul), a.ramanan@univ.jfn.ac.lk (Amirthalingam Ramanan), subhaf@uom.lk (Subha Fernando)

target tokens are selectively extracted from a localized region to attend to the template tokens, thereby reducing the influence of background and distractor tokens. The proposed CPDATrack achieves state-of-the-art performance on multiple benchmarks, particularly excelling in one-shot tracking on GOT-10k with an average overlap of 75.1%. The code, models, and results of this work are available at `https://github.com/JananiKugaa/CPDATrack.git`.

Keywords:
Visual Object Tracking, Token Pruning, Selective attention, Transformer Tracking

---

Journal of Visual Communication and Image Representation 119 (2026) 104884