

# Online Multi-Object Tracking with Dual Matching Attention Networks

Ji Zhu<sup>1,2</sup>, Hua Yang<sup>1</sup>, Nian Liu<sup>3</sup>, Minyoung Kim<sup>4</sup>,  
Wenjun Zhang<sup>1</sup>, and Ming-Hsuan Yang<sup>5,6</sup>

<sup>1</sup>Shanghai Jiao Tong University <sup>2</sup>Visbody Inc

<sup>3</sup>Northwestern Polytechnical University <sup>4</sup>Massachusetts Institute of Technology

<sup>5</sup>University of California, Merced <sup>6</sup>Google Inc

Fjizhu1023, liuni an228@gmail.com minykim@mit.edu

Fhyang, zhangwenjun@sjtu.edu.cn mhyang@ucmerced.edu

**Abstract.** In this paper, we propose an online Multi-Object Tracking (MOT) approach which integrates the merits of single object tracking and data association methods in a unified framework to handle noisy detections and frequent interactions between targets. Specifically, for applying single object tracking in MOT, we introduce a cost-sensitive tracking loss based on the state-of-the-art visual tracker, which encourages the model to focus on hard negative distractors during online learning. For data association, we propose Dual Matching Attention Networks (DMAN) with both spatial and temporal attention mechanisms. The spatial attention module generates dual attention maps which enable the network to focus on the matching patterns of the input image pair, while the temporal attention module adaptively allocates different levels of attention to different samples in the tracklet to suppress noisy observations. Experimental results on the MOT benchmark datasets show that the proposed algorithm performs favorably against both online and offline trackers in terms of identity-preserving metrics.

**Keywords:** Multi-object tracking · Cost-sensitive tracking loss · Dual matching attention network.

## 1 Introduction

Multi-Object Tracking (MOT) aims to estimate trajectories of multiple objects by finding target locations and maintaining target identities across frames. In general, existing MOT methods can be categorized into offline and online methods. Offline MOT methods use both past and future frames to generate trajectories while online MOT methods only exploit the information available up to the current frame. Although offline methods have some advantages in handling ambiguous tracking results, they are not applicable to real-time vision tasks.

Recent MOT methods mainly adopt the tracking-by-detection strategy and handle the task by linking detections across frames using data association algorithms. However, these approaches heavily rely on the quality of detection

---

Corresponding author.



































