Improvising Object Tracking Algorithm SORT for Long-Term Trajectory Extraction

Rutul Patel*, Saumil Patel†, Kunj Kanzariya‡, and Ayush Patel§
*†‡§Computer Science and Engineering, Ahmedabad University, Ahmedabad, India
{rutul.p, saumil.p, kunj.k1, ayush.p3}@ahduni.edu.in

***Abstract*—This paper compares Simple Online and Realtime Tracking (SORT) and its advanced version StrongSORT multi-object tracking (MOT) approaches. Both approaches aim to address the inherent challenges in MOT, such as occlusions and the need for real-time tracking, by employing a pragmatic approach that combines speed with simplicity. While SORT lays the groundwork by using a straightforward tracking mechanism alongside object detection, Strong SORT builds upon this foundation with deep learning enhancements for both detection and appearance feature extraction. A key feature of Strong SORT is its sophisticated data association strategy, which merges motion information from a Kalman filter with appearance information to improve tracking accuracy and robustness. Our analysis, utilizing the MOT17 dataset, reveals that Strong SORT significantly outperforms SORT, particularly in complex scenarios characterized by occlusions and rapid object movements.**

***Index Terms*—component, formatting, style, styling, insert**

I. Introduction

In Multi-object Tracking (MOT) we aim to track, all the object present in the frames of the video. The MOT is usually divided into two part object detection followed by object tracking[1]. Object detection acts as the foundation for object tracking. This is an important process in various application in our day-to-day life including surveillance security to autonomous driving cars. Many studies have been proposed in the literature which aims to improve SORT for multi-object tracking. In[2] the authors have used deep learning to learn the features about the object. These improvements are aimed at addressing the problems such as increasing accuracy in identification of object, minimizing the switches in the identity when objects are re-identified, improving the tracking even when the object is obscured or occluded.

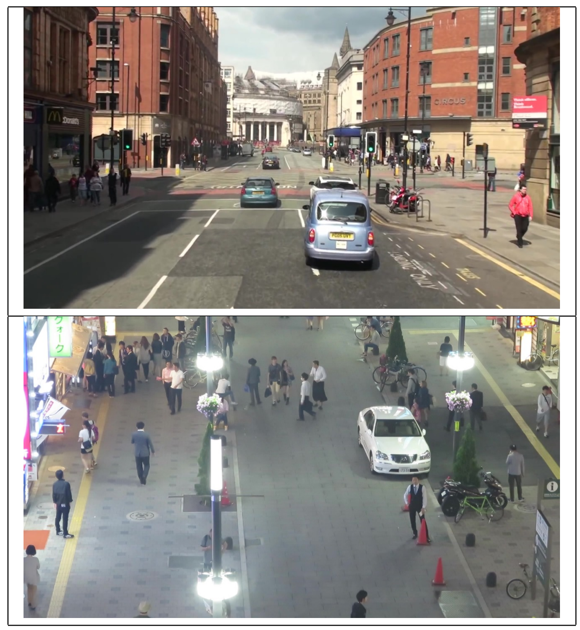
II. Data Description

A. Multiple Object Tracking Dataset - 17

The MOT17 dataset is a benchmark collection designed for evaluating multi-object tracking algorithms. It comprises several video sequences capturing diverse, real-world scenarios, each accompanied by precise annotations of pedestrian positions across frames. For the SORT algorithm, which focuses on tracking objects based on motion and appearance, MOT17 provides a valuable testing ground. The dataset includes both raw video and precomputed detection files, allowing assessment of tracking performance under varied conditions. With its mix of crowded scenes and varying lighting, MOT17 challenges SORT's capabilities in data association and maintaining object identities over time, making it an ideal resource for validating and enhancing tracking methodologies.

Identify applicable funding agency here. If none, delete this.

TABLE I
MOT17 Dataset Example.



III. Methodology

A. SORT

The Simple Online and Realtime Tracking (SORT) algorithm represents a pragmatic approach to multi-object tracking (MOT) that prioritizes speed and simplicity. At its core, SORT relies on two main steps: object detection in each frame, followed by a straightforward tracking mechanism that associates detected objects across frames using the Hungarian algorithm based on predicted object locations. Object detection can be performed by any state-of-the-art detection model, and tracking is facilitated by a Kalman filter, which predicts object positions in new frames based on their past position.

SORT's primary advantage lies in its computational efficiency, enabling real-time tracking in various applications. However, it faces challenges in scenarios with frequent occlusions or significant appearance changes, where its simple association mechanism might lead to identity switches or track loss. Despite these limitations, SORT serves as a foundational method for MOT, offering a balance between performance and speed that is crucial for many real-time applications.

B. *Strong SORT*

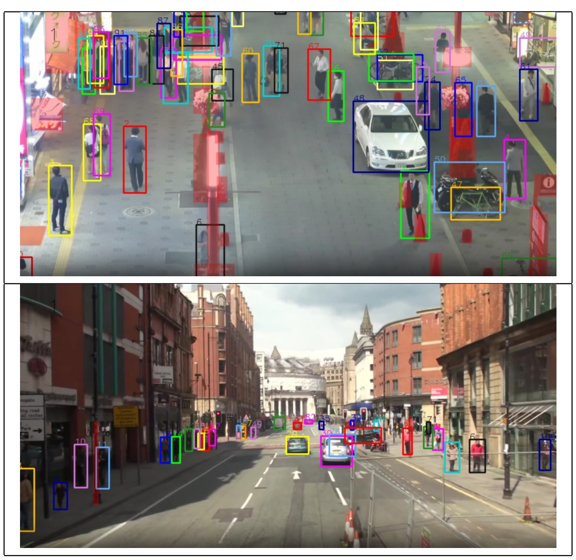
Building on the foundations of SORT, Strong SORT introduces enhancements that significantly improve tracking accuracy and robustness, addressing some of the limitations of the original algorithm. Strong SORT incorporates advanced deep learning models for both object detection and appearance feature extraction, enabling more precise object identification and tracking even in challenging conditions.

A key feature of Strong SORT is its sophisticated data association strategy, which combines motion information from a Kalman filter with appearance information obtained from deep learning-based feature extractors. This dual approach reduces identity switches and improves tracking continuity, especially in scenes with occlusions, interactions among objects, or significant variations in object appearance.

Moreover, Strong SORT introduces mechanisms for managing track lifecycles, including the initiation of new tracks for newly detected objects, the continuation of existing tracks, and the termination of tracks that no longer correspond to observed objects. These mechanisms ensure that the tracker maintains an accurate representation of the scene, enhancing its utility in applications requiring precise and reliable MOT.

In essence, Strong SORT represents an evolution of SORT, leveraging the latest advancements in computer vision and deep learning to offer superior performance in MOT tasks. Its development underscores the ongoing efforts within the research community to create more effective and efficient tracking algorithms.

TABLE II
RESULTS OF TRACKING ON MOT17.



IV. RESULTS AND INFERENCES

In our comparative study between the SORT and strongSORT tracking algorithms using the MOT17 dataset, specifically on the MOT-17-SDP subset, we observed that strongSORT outperformed SORT in terms of Multi-Object Tracking Accuracy (MOTA), achieving scores of 34.152 and 23.07, respectively. Despite the relatively low MOTA scores, it's important to note that both algorithms demonstrated proficient tracking capabilities in practical video frame tests. The discrepancy in the quantitative evaluation, particularly the lower accuracy, might stem from potential errors in our calculation methodology or the inherent challenges presented by the dataset, such as occlusions and rapid movement.

TABLE III
COMPARISON OF MOTA VALUES FOR DIFFERENT TRACKING METHODS

Tracking Method	**MOTA**
SORT (SDP)	23
Strong SORT (FRCNN)	34.15
SORT (DPM)	11

These findings suggest that while both SORT and strongSORT are effective for real-time tracking, strongSORT offers a notable advantage in handling complex scenarios, as evidenced by its superior MOTA score. Future investigations will aim to refine the evaluation process to ensure accuracy and further explore the algorithms' performance nuances.

V. ACKNOWLEDGEMENT

We acknowledge Dr. Mehul Raval and his team for their support and guidance throughout this study.

REFERENCES

[1] Du, Yunhao, Zhicheng Zhao, Yang Song, Yanyun Zhao, Fei Su, Tao Gong, and Hongying Meng. "Strongsort: Make Deepsort Great Again." *arXiv.org*, February 22, 2023. https://arxiv.org/abs/2202.13514.

[2] Maggiolino, Gerard Ahmad, Adnan Cao, Jinkun Kitani, Kris. (2023). Deep OC-Sort: Multi-Pedestrian Tracking by Adaptive Re-Identification. 3025-3029. 10.1109/ICIP49359.2023.10222576.

[3] P. Dendorfer et al., "MOTChallenge: A Benchmark for Single-Camera Multiple Target Tracking." *arXiv*, 2020. doi: 10.48550/ARXIV.2010.07548.