Analyzing algorithms for Multi-Object Tracking

Oancea Sorina
AI Multimedia Lab
National University of Science and Technology POLITEHNICA Bucharest
Bucharest, Romania
sorina.oancea@upb.ro

Bogdan Ionescu
AI Multimedia Lab
National University of Science and Technology POLITEHNICA Bucharest
Bucharest, Romania
bogdan.ionescu@upb.ro

***Abstract*—This article covers the challenges of multi-object tracking in computer vision, which is crucial for applications such as surveillance or augmented reality. The research scope is to develop some adaptive methods for accurate object tracking in real-time, using deep learning techniques, particularly Convolutional Neural Networks(CNNs). The technology is advanced for the object detection through deep learning, however some challenges remain, like tracking small or fast-moving objects or handling environmental variations. This article aims to identify the limitations of the current tracking algorithms, exploring cutting-edge techniques, and proposing new methodologies to improve the efficiency and accuracy of the tracking process. Through simulation and real-world validation, the article will evaluate the performance of the old approaches, while leaving room for the exploration for the new solutions that could contribute to the development of more robust systems used for practical application across various fields. The algortihms analysed were SORT, DeepSORT, TrackFormer, TransTrack, CenterTrack and YOLO.**

***Index Terms*—DeepLearning, Multiple Object Tracking(MOT), Object detection, Convolutional Neural Networks(CNN), Transformer**

I. Introduction

With object tracking being a fundamental task in c omputer vision, applications span across areas such as surveillance, autonomous driving and robotics. The challenge of tracking multiple objects simultaneously, called "Multiple Object Tracking", involves associating objects across frames, managing blockages and predicting their trajectories. In recent years, the algorithms have advanced significantly due to improvements in deep learning and the integration of sophisticated tracking techniques. These methods have big advantages in challenging tracking situations, especially when there are obstacles, changes in how objects look, or when objects interact with each other. The article compares these newer techniques with real-time detection models like YOLO, to show what works well and what doesn't, especially when it comes to tracking multiple objects in real-world settings.

II. Algorithms Overview

A. SORT(Simple Online and Real-Time Tracking)

SORT is an efficient tracking algorithm that combines object detection with a Kalman filter to predict the future position of detected objects. The Kalman filter allows SORT to perform well in real-time scenarios, handling noisy detections and providing robust tracking even in low-complexity environments. One of its key features is the **Hungarian algorithm**, which is used for associating detections with previously predicted tracks. However, SORT has limitations, particularly in crowded scenes where object occlusions or overlaps occur.

B. DeepSort

An extension of the original SORT algorithm, DeepSORT incorporates deep learning to enhance object re-identification and association. By using a convolutional neural network (CNN) to extract feature embeddings from detected objects, DeepSORT can handle challenging scenarios where objects are occluded or appear in low-resolution frames. The combination of Kalman filtering with deep learning significantly improves the algorithm's robustness in environments with complex object affiliations.

C. TrackFormer

The algorithm uses a Transformer to address the problem of MOT (Multiple Object Tracking – the simultaneous tracking of multiple objects in video sequences, where each detected object is assigned a trajectory that 'tracks' the object across multiple frames). The Transformer is a class of neural network architectures used in natural language processing (NLP) and computer vision (CV). It requires training on large, grouped datasets, allowing the model to learn to track objects in various scenarios and lighting conditions. After training, the model can be used in real-time to track objects within video sequences. The performance is high when compared to traditional methods used for object tracking. The algorithm simultaneously manages the initialization of the object recognition process, the identification of the object, and the calculation of its trajectory using an encoder-decoder Transformer architecture, eliminating the need for additional association methods (including graphical optimization and object motion modeling). The results obtained by this algorithm are promising and will certainly inspire future research in the field of using Transformers for tracking and detection processes.

D. TransTrack

It represents a multiple object tracking (MOT) model that combines the Transformer architecture with traditional detection techniques, offering superior performance. A CNN backbone architecture is used to extract object features, followed by

1st Workshop on Artificial Intelligence for Multimedia, AI Multimedia Lab, CAMPUS Research Institute, POLITEHNICA Bucharest, Bucharest, Romania, 8 November, 2024

the application of a class of Transformer architectures required for inter-frame associations and detection. The model performs continuous updates of predictions as new information becomes available in the video sequence, ensuring the accuracy of the process of object tracking in real time. The algorithm is flexible and can be adapted to process a wide variety of scenarios and data types across different application domains, without the need for hardware modifications to handle these data.

E. CenterTrack

CenterTrack is an efficient tracking algorithm that represents an advanced method for object tracking, focusing on the use of information about the central position of objects, which serves as a stable and significant feature that allows objects to be identified. The algorithm uses a convolutional neural network (CNN) to detect objects and estimate their central position in each frame of a video sequence. The position is then used to track the objects throughout the entire video, no matter the changes in positioning or the level of image illumination. The architecture of the algorithm is highly efficient, making it suitable for implementation on various hardware platforms.

F. YOLO(You Only Look Once)

YOLO is a real-time object detection algorithm developed by Joseph Redmon. It is recognized for its accuracy and speed of recognition, being able to detect objects in a single pass of image or video sequence processing. Object detection becomes a unique regression problem, processing the entire input image and simultaneously predicting bounding boxes and class probabilities for the objects. The YOLO algorithm is extremely fast and capable of processing data in real time. The architecture is simple, making it easy to implement and train. The algorithm performs detection and classification in a single pass. However, a notable disadvantage is that the early versions of YOLO struggled with detecting small objects in large images. In this algorithm, speed is prioritized, which can lead to trade-offs in data processing.

III. Conclusion about the studied algorithms

The algorithms discussed above(FairMOT, Deep SORT, TrackFormer, TransTrack, CenterTrack, and YOLO) presents distinct advantages for Multiple Object Tracking and Object Detection, yet come with their own set of limitations. FairMOT excels with its integrated detection and re-identification system, offering high accuracy and efficiency, especially in complex scenes. Deep SORT is build using traditional tracking methods by introducing deep learning for re-identification. TrackFormer makes use of Transformer-based architectures to enhance long-range object tracking and association, delivering high performance but requiring substantial hardware resources. TransTrack, similar with TrackFormer, uses Transformers to refine object association across frames, proving effective in dynamic scenarios but relying on large datasets for training. CenterTrack simplifies the tracking process by focusing on the central position of objects, having real-time performance with impressive results on standard datasets. Lastly, YOLO remains one of the fastest and most efficient algorithms for object detection, with real-time capabilities, though earlier versions struggled with small object detection. Overall, each algorithm has its strengths and challenges, and the choice of model depends on the specific application requirements, including the trade-offs between accuracy, computational resources, and real-time performance.

References

[1] Yifu Zhang, Chunyu Wang, Xinggang Wang, Wenjun Zeng, Wenyu Liu, FairMOT: On the Fairness of Detection and Re-Identification in Multiple Object Tracking, arXiv, 2021.

[2] Jonathan Tan, Lyuyu Shen, Marcelo Ang Jr., FairMOT-X: Real-time One-shot Methods For Multi-class Multi-object Tracking, 17th International Conference on Control, Automation, Robotics and Vision (ICARCV), December 11-13, Singapore, 2022.

[3] Yuhao Wang, Hangzhang Cheng, Xintong Zhou, Wei Luo, Haopeng Zhang, MOVING SHIP DETECTION AND MOVEMENT PREDICTION IN REMOTE SENSING VIDEOS, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLIII-B2-2020, 2020, XXIV ISPRS Congress (2020 edition).

[4] Abdul Majid, Qinbo, Saba Brahmani, Deep SORT Related Studies, International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 2456-3307.

[5] Tim Meinhardt, Alexander Kirillov, Laura Leal-Taixe, Christoph Feichtenhofer, FairMOT: TrackFormer: Multi-Object Tracking with Transformers, CVPR, 2022.

[6] Zhou, X., Koltun, V., Krähenbühl, P. (2020). Tracking Objects as Points. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, JM. (eds) Computer Vision – ECCV 2020. ECCV 2020.

[7] Mohammed Razzok, Abdelmajid Badri, Ilham El Mourabit, Yassine Ruichek, Aïcha Sahel, Pedestrian Detection and Tracking System Based on Deep-SORT, YOLOv5, and New Data Association Metrics, Information 2023.