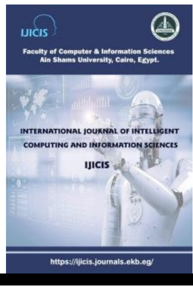
IJICIS, Vol.24, No.3, 11-24
DOI: 10.21608/ijicis.2024.309762.1347

International Journal of Intelligent Computing and Information Sciences

https://ijicis.journals.ekb.eg/



ENHANCEMENT ONLINE MULTI OBJECT TRACKING IN DYNAMIC ENVIRONMENT

Metwally Rashad

Department of Computer Science,
Faculty of Computers and Artificial Intelligence, Benha University,
Cairo, Egypt
metwally.rashad@fci.bu.edu.eg (M.R.)

Eman Mohamed*

Department of Computer Science,
Faculty of Computers and Artificial Intelligence, Benha University,
Cairo, Egypt
emanseleem9913@gmail.com

Ahmed A. El-Sawy

Department of Computer Science,
Faculty of Computers and Artificial Intelligence, Benha University,
Cairo, Egypt
ahmed.el_sawy@fci.bu.edu.eg (A.A.E.-S.)

Ahmed H. AbuEAtta

Department of Computer Science,
Faculty of Computers and Artificial Intelligence, Benha University,
Cairo, Egypt
ahmed.aboalatah@fci.bu.edu.eg (A.H.A.-A.)

Received 2024-08-04; Revised 2024-09-10; Accepted 2024-09-19

Abstract: *Object tracking is crucial for a wide range of computer vision applications, including autonomous navigation and surveillance systems. This paper introduces StrongSort, a novel object tracking algorithm designed to tackle the difficulties of achieving real-time accuracy in challenging environments. StrongSort leverages the capabilities of YOLOv8, a leading object detection model, to achieve this goal. The foundation of StrongSort lies in its integration with YOLOv8, which provides excellent object detection accuracy and speed. Leveraging YOLOv8's detection outputs, StrongSort utilizes a combination of object embeddings, motion prediction and a deep association mechanism to create a robust tracking framework. This enables StrongSort to handle occlusions, scale variations and abrupt object movements effectively. One of the key contributions of StrongSort is its ability to handle multiple object tracking making it suitable for multi-object tracking scenarios, such as autonomous vehicles navigating through urban environments or surveillance systems monitoring crowded areas. The algorithm employs a hierarchical approach that accurately associates detected objects across frames while maintaining low computational overhead. Experimental results show that StrongSort surpasses existing object tracking algorithms in key areas: accuracy, robustness, and speed. Furthermore, its efficiency enables real-time performance on standard hardware, making it a practical choice for a variety of applications.*

Keywords: *Multi-object tracking, You only look once (YOLO), YOLOv8, DeepSORT, StrongSort.*

***Corresponding Author**: Eman Mohamed

Computer Science Department, Faculty of Computers and Artificial Intelligence, Benha University, Cairo, Egypt.

Email address: emanseleem9913@gmail.com

1 Introduction

Accurately and reliably tracking objects within complex and dynamic scenes represents a fundamental challenge in computer vision with wide-ranging implications. The demand for real-time high-performance object tracking algorithms continues to grow driven by applications such as autonomous vehicles navigating busy urban environments and surveillance systems monitoring public spaces. This paper introduces "StrongSort" a novel object tracking algorithm that significantly advances the field by leveraging the capabilities of YOLOv8 a cutting-edge object detection model. Within the field of computer vision, achieving accurate and robust object tracking in complex and dynamic scenes is a fundamental challenge with broad applications. The demand for real-time high-performance object tracking algorithms is constantly increasing, driven by use cases ranging from autonomous vehicles navigating crowded urban environments to surveillance systems monitoring public spaces. This paper introduces "StrongSort" a novel object tracking algorithm that leverages .the capabilities of YOLOv8, a state-of-the-art object detection model, to achieve significant advancements in the field. Object tracking remains a critical component in computer vision tasks [2] bridging the gap between object detection and understanding object motion and behavior over time. YOLOv8 with its impressive object detection capabilities offers a robust starting point for object tracking. StrongSort capitalizes on this foundation and extends it to provide a comprehensive solution for object tracking in complex scenarios. Multiple Object Tracking is a subfield of computer vision that focuses on tracking multiple objects simultaneously in complex environments [3]. Unlike single-object tracking [12], MOT aims to maintain the identities and trajectories of numerous objects at the same time. This is crucial for applications like autonomous driving, surveillance and robotics [20], where understanding the interactions between multiple objects is essential. MOT algorithms face various challenges such as dealing with occlusions changes in object size, object interactions [15] and accurately matching objects across different frames. The primary objective of StrongSort is to address the intricacies of object tracking that arise in real-world scenarios. It tackles the challenges of handling scale variations, object occlusions, abrupt movements and the tracking of multiple objects simultaneously. The synergy between YOLOv8's detection accuracy and StrongSort's tracking capabilities yields a system that excels in both precision and speed making it suitable for a wide array of applications where real-time accurate tracking is essential. You Only Look Once is a groundbreaking technology in computer vision and object detection. YOLO’s key innovation is its ability to detect objects in real-time within a single pass [4], unlike traditional multi-stage methods (YOLO: Real-Time Object Detection, 2018). This efficiency stems from its deep neural network architecture, which simultaneously predicts bounding boxes and class probabilities for multiple objects within an image grid. YOLO’s speed and accuracy have made it a popular choice for various applications including autonomous vehicles, surveillance systems, augmented reality and robotics [18]. The ongoing development of YOLO with versions like YOLOv8[17] highlights its lasting impact on computer vision, where it continues to advance real-time object detection [5].StrongSort is a powerful object tracking algorithm that distinguishes itself through a multifaceted approach built upon the robust object detection capabilities of YOLOv8. This foundation enables StrongSort to excel in handling complex scenarios [6].The strength of StrongSort lies in its sophisticated framework, which combines advanced motion prediction, object embeddings and a deep association mechanism. This framework ensures accurate object identity maintenance across frames, enabling reliable tracking even amidst complex dynamics and occlusions. Moreover, StrongSort introduces a hierarchical approach to manage multiple objects, making it highly adaptable to scenarios requiring simultaneous tracking of numerous objects.

2 Related Work

2.1 DeepSORT Tracker

DeepSORT utilizes a two-branch framework for object tracking, comprising an appearance branch and a motion branch [4]. The appearance branch leverages deep appearance descriptors extracted from a simple Convolutional Neural Network pretrained on the person re-identification dataset. This approach allows DeepSORT to effectively learn and recognize object appearances for robust tracking.These descriptors are used to extract appearance features from the detections in each frame. Additionally [8], a feature bank mechanism is employed to store the features of the preceding 100 frames for each tracklet. As new detections are introduced, we compute the minimum cosine distance between the feature bank B_i corresponding to the i-th tracklet and the feature f_j associated with the j-th detection.

$$d(i,j) = min\left\{1 - f_j^T f_k^{(i)} \middle| f_k^{(i)} \in B_i\right\}. \quad (1)$$

DeepSORT: Tackling the Challenges of Multiple Object Tracking:

The core challenge in Multiple Object Tracking is the simultaneous tracking and association of multiple object identities within a scene. DeepSORT tackles this challenge by leveraging deep learning techniques to enhance both object association and tracking across video frames [13]. This approach enables DeepSORT to effectively maintain object identities and trajectories even in complex scenarios with multiple moving objects. Here are some key aspects of DeepSORT's role in MOT:

- Feature Embeddings: DeepSORT using deep convolutional neural networks (CNNs) to extract appearance features or embeddings for each detected object. These embeddings encode unique characteristics of each object, making it easier to distinguish between them, even in crowded scenes.
- Association: to track objects as they move through different video frames. This algorithm hinges on the comparison of appearance embeddings, which like unique fingerprints for each object. By measuring the similarity between these embeddings [1], DeepSORT can accurately connect the dots and maintain consistent object identities even when objects are briefly hidden or move out of sight [3]. This ability to bridge gaps in object visibility is paramount in multi-object tracking scenarios ensuring that objects are not misidentified or lost track of amidst complex and dynamic scenes.
- Scalability: to effectively track a fluctuating number of objects within a scene. This inherent adaptability is crucial in dynamic environments where the object count can change rapidly. For instance, in applications like monitoring pedestrian flow or tracking vehicles in traffic, DeepSORT effortlessly adjusts to accommodate the ebb and flow of objects ensuring accurate and reliable tracking even amidst fluctuating object densities [4].
- Real-time tracking is a cornerstone of its effectiveness for MOT applications. It achieves this by carefully balancing the need for accurate tracking with the constraints of computational efficiency [5]. This makes it a powerful tool for systems that demand both timely and accurate tracking results [6]. Whether it's a self-driving car navigating through traffic or a security system monitoring a crowded area DeepSORT's real-time performance ensures that critical information is available the moment it's needed.

2.2 StrongSORT

NAS Kalman Traditional Kalman filters while effective in many tracking scenarios face limitations when dealing with real-world complexities. Specifically, they are sensitive to low-quality detections which are common in noisy environments and they lack a mechanism to explicitly account for varying scales of detection noise [7]. This can lead to inaccurate state estimations and hinder overall tracking performance.To overcome these limitations, we leverage the NSA Kalman algorithm a more robust variation of the traditional Kalman filter as implemented in GIAOTracker. This algorithm explicitly addresses the challenges posed by low-quality detections and varying noise scales resulting in more accurate and reliable tracking results.

Motion Cost Enhancement We provide an improved method that differs from Strongsort's which mainly uses the appearance feature distance as a matching cost during the first association stage with in motion distance only acting as a gating mechanism[26]. We deal with the assignment issue by combining details about motion and appearance [3]. To make this improvement [20] a cost matrix with the notation. This enhancement involves the calculation of a cost matrix, denoted as C, which is derived from a weighted sum of appearance cost ($\mathrm{A_a}$) and motion cost ($\mathrm{A_m}$)

$$\mathrm{C} = \lambda A_a + (1 - \lambda) Am \quad (2)$$

Where the weight factor λ is set to 0.98.

Vanilla Matching Interestingly our analysis reveals that the matching cascade algorithm a key component of SortSORT can become a performance bottleneck as the tracker's capabilities improve [27]. This seemingly counterintuitive observation stems from the fact that stronger trackers are inherently more robust to challenging data associations [12]. Consequently, the additional prior constraints imposed by the matching cascade become overly restrictive hindering overall matching accuracy [32].

Our solution to this is to use a more direct global linear assignment strategy in place of the matching cascade which is a straightforward but efficient solution. This streamlining removes needless obstacles from the tracker's capabilities enabling it to fully utilize its expanded powers [28].

2.3 StrongSORT with YOLOX

StrongSort marks a significant advancement in object tracking intricately linked to the progress made in object detection by YOLOX. Capitalizing on YOLOX's exceptional accuracy and efficiency in pinpointing objects, StrongSort introduces a powerful new tracking algorithm. This algorithm leverages YOLOX's precise object localization and identification abilities to enhance the robustness and precision of object tracking especially in challenging environments. By seamlessly incorporating YOLOX's detection results as a cornerstone [35], StrongSort elevates the reliability and accuracy of object tracking in complex scenarios.

2.4 ByteTrack

Traditional Multi-Object Tracking methods often prioritize high-scoring detection boxes, which can lead to overlooking objects with lower scores such as those that are partially hidden or occluded [8]. This over-reliance on high scores often results in missed objects and incomplete tracking trajectories.

To address this issue a new approach proposes a more inclusive strategy [9]. This method considers a broader range of detection boxes including those with lower scores by leveraging their similarities with existing tracklets. This inclusive approach helps to recover true objects that might have been missed and effectively filters out background noise leading to more complete and accurate tracking results. This innovative approach has been successfully applied to enhance the performance of various state-of-the-art trackers [10].

ByteTrack employs a lightweight neural network architecture for tracking objects within video sequences in Figure 1. It utilizes a Siamese network design that processes pairs of input frames and produces a similarity score determining whether the objects in those frames correspond. ByteTrack a high-performance object tracking algorithm leverages a simple yet remarkably effective data augmentation strategy to excel in challenging tracking scenarios. This approach allows ByteTrack to outperform many state-of-the-art object tracking algorithms achieving superior results with fewer parameters and faster processing times. This efficiency makes ByteTrack a compelling choice for real-time applications where speed and accuracy are paramount.

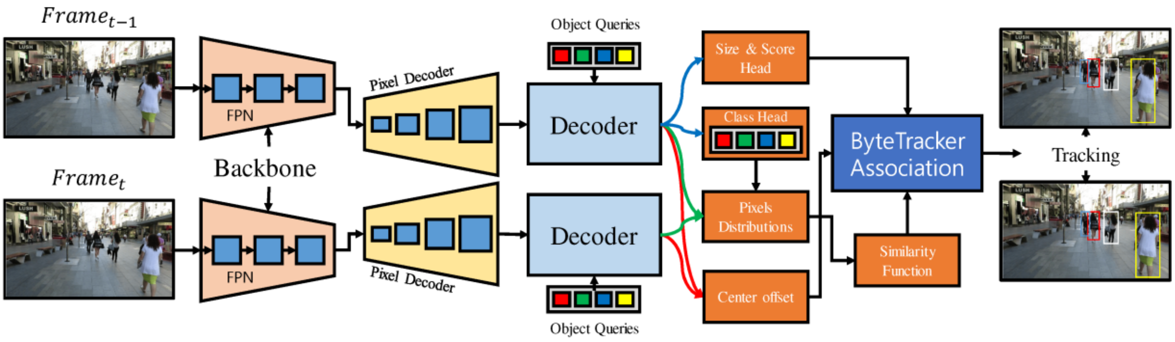


Figure 1. ByteTrack

2.5 OC-SORT Tracker

Kalman filter (KF) techniques are widely used in multi-object tracking but often rely on the assumption of linear object motion. While this assumption is generally valid for short periods of occlusion, it can lead to significant inaccuracies over extended durations. Moreover, when measurements are unavailable traditional KF methods rely heavily on prior state estimations resulting in error accumulation during occlusion and substantial variance in motion direction.

This study challenges the notion that complex KF variations are necessary for optimal performance. We demonstrate that a basic Kalman filter when coupled with careful management of occlusion-induced noise accumulation can achieve state-of-the-art tracking results [17]. Our approach deviates from conventional estimation-centric methods by incorporating object observations (measurements from object detectors) to construct a "virtual trajectory" spanning occlusion periods [39]. This virtual trajectory aids in correcting errors in filter parameters leading to improved accuracy.

It was introduced as Observation-Centric SORT (OC-SORT) preserving its simplicity real-time capability and online nature while enhancing robustness during occlusion and accommodating non-linear motion. When supplied with off-the-shelf detections [18], OC-SORT achieves impressive speeds of over 700 FPS on a single CPU and attains state-of-the-art results across various datasets, including

MOT17, MOT20, head tracking and notably DanceTrack where object motion exhibits pronounced non-linearity [35].

3 Methodology

Detection by YOLOV8, YOLO models have significantly advanced the field of Computer Vision achieving state-of-the-art performance in object detection. The latest iteration, YOLOv8 as shown in Figure 2, exemplifies the model family's key strength: predicting all objects within an image in a single forward pass hence the name "You Only Look Once”.
The main distinction introduced by the YOLO models was the framing of the task at hand [19]. The authors of the paper reframed the object detection task as a regression problem (predict the bounding box coordinates) instead of classification [24].
YOLO models are also faster to train and have the ability to produce high accuracy with smaller model sizes [21]. They can be trained on single GPUs making them more accessible to developers like us. YOLOv8 is the latest iteration of these YOLO models (as of early 2023). It has undergone a few major changes from its ancestors such as anchor-free detection the introduction of C3 convolutions and mosaic augmentation [34].

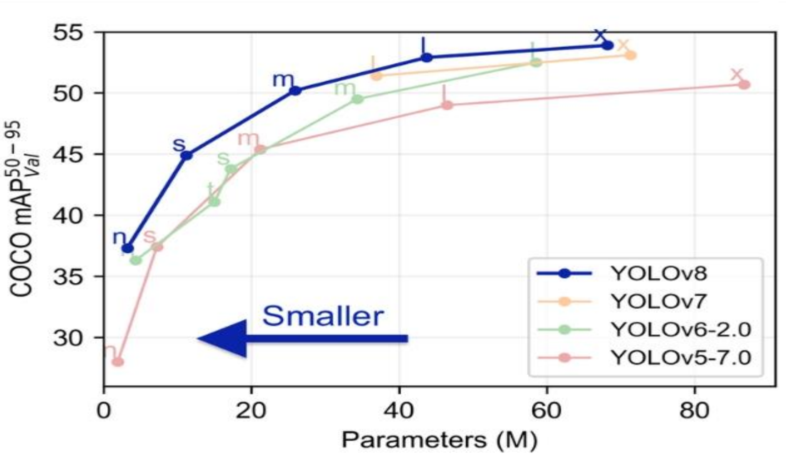


Figure 2:Comparing different YOLO version

YOLOv8 has more parameters than its predecessors, such as YOLOv5 but fewer parameters than YOLOv6. It offers about 33% more mAP for n-size models and generally a greater mAP across the board [33].
DeepSORT with YOLOV8 When YOLOv8 and DeepSORT are integrated, a potent system for real-time object tracking is produced. YOLOv8 a fast and accurate object detection system recognizes things in video frames. Then, DeepSORT takes over giving each observed object a unique ID and making predictions about its future positions based on its appearance and movement patterns Figure3. This combination makes it possible to follow things continuously even when they move rapidly or are momentarily obscured its improved counterpart in scenarios involving extended occlusions or disappearances Table 1.

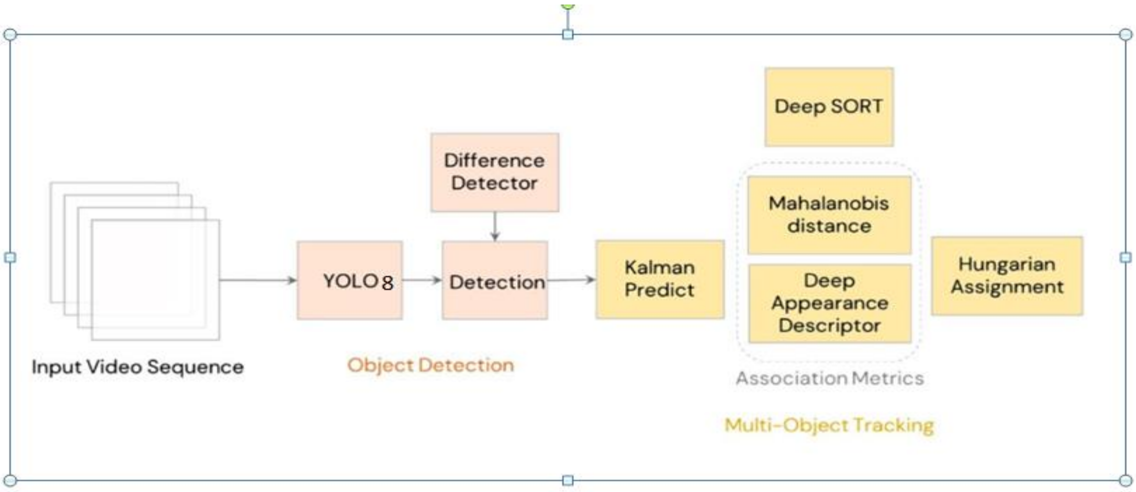


Figure 3:DeepSORT with YOLOV8

Table 1: Comparison with state-of-the-art MOT methods on the MOT17 test setDeepsort with yolov8. The best results for each metric are bolded and highlighted in red.

Method	Ref.	HOTA (↑)	IDF1 (↑)	MOTA (↑)	AssA (↑)	DetA (↑)	IDs (↓)
DeepSORT	ICIP2017	61.2	74.5	78.0	59.7	63.1	1,821
DeepSORT+YOLOV8	OUR	61.7	74.6	78.1	59.9	63.0	1,618

StrongSORT with YOLOV8, Building on the DeepSORT architecture [16] StrongSORT is an improved object tracking method designed to address typical tracking issues including missing associations and missing detections [1]. It does this by adding motion prediction models and stronger matching algorithms. StrongSORT resolves these problems to deliver more accurate and dependable object trajectories throughout time [22], especially in intricate situations involving occlusions or transient object disappearances from the camera's field of view. This makes it an effective tool for uses in video analysis, autonomous driving and surveillance [23] Table 2.

This paper embarks on revisiting the classic DeepSORT tracker, enhancing it significantly across multiple dimensions encompassing object detection, feature embedding, and trajectory association [22]. The resulting tracker named StrongSORT offers a formidable and equitable benchmark for the MOT community in Figure 4.

Moreover, this paper introduces two lightweight and adaptable algorithms designed to tackle two inherent challenges in MOT: missing association and missing detection. In contrast to the prevalent approach of associating short tracklets into complete trajectories with high computational complexity [25].

Table 2: Comparison with Strongsort with YoloX by more epochs and Strongsort with YOLOv8, MOT methods on the MOT17 test set. The best results for each metric in Deep sort and strongsort are bolded and highlighted in red.

Method	Ref.	HOTA (↑)	IDF1 (↑)	MOTA (↑)	AssA (↑)	DetA (↑)	IDs (↓)
SORT	ICIP2016	34.0	39.8	43.1	31.8	37.0	143.3

MTDF	TMM2019	37.7	45.2	49.6	34.5	42.0	5,567
DeepMOT	CVPR2020	42.4	53.8	53.7	42.7	42.5	1,947
ISEHDADH	TMM2019	-	-	54.5	-	-	3,010
Tracktor++	ICCV2019	44.8	55.1	56.3	45.1	44.9	1,987
TubeTK	CVPR2020	48.0	58.6	63.0	45.1	51.4	4,137
CRF-MOT	TMM2022	-	60.4	58.9	-	-	2,544
CenterTrack	ECCV2020	52.2	64.7	67.8	51.0	53.8	3,039
TransTrack	arxiv2020	54.1	63.5	75.2	47.9	61.6	3,603
PermaTrack	ICCV2021	55.5	68.9	73.8	53.1	58.5	3,699
CSTrack	TIP2022	59.3	72.6	74.9	57.9	61.1	3,567
FairMOT	IJCV2021	59.3	72.3	73.7	58.0	60.9	3,303
CrowdTrack	AVSS2021	60.3	73.6	75.6	59.3	61.5	2,544
CorrTracker	CVPR2021	60.7	73.6	76.5	58.9	62.9	3,369
RelationTrack	TMM2022	61.0	74.7	73.8	61.5	60.6	1,374
OC-SORT	arxiv2022	61.7	76.2	76.0	62.0	61.6	2,199
ByteTrack	ECCV2022	62.8	77.2	78.9	62.2	63.8	2,310
DeepSORT	ICIP2017	61.2	74.5	78.0	59.7	63.1	1,821
DeepSORT+YOLOV8	**our**	61.5	74.6	78.1	59.9	63.0	1,718
StrongSORT + YOLOX	arxiv2022	62.9	78.5	78.3	63.7	63.6	1,446
StrongSORT + YOLOV8	**ours**	**68.43**	**81.88**	**85.35**	**66.33**	**65.55**	**625**

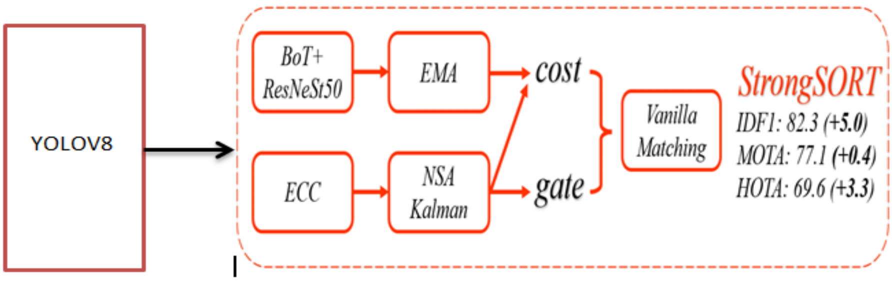


Figure 4:StrongSORT with YOLOV8

In combining the object detection prowess of YOLOv8 with the robust tracking capabilities of StrongSORT creates a powerful synergy for monitoring objects in video streams [2]. The state-of-the-art detection model YOLOv8 quickly and accurately recognizes a variety of items. Then, StrongSORT takes over using deep learning methods to guarantee consistent tracking between frames [36] even in cases when objects are temporarily occluded or vanish. In domains where keeping precise item IDs in dynamic situations is essential such as sports analytics autonomous driving and surveillance this seamless integration is especially helpful [14]. Through the integration of YOLOv8 and StrongSORT developers can construct advanced systems that are proficient in evaluating intricate visual data [38].

4 Experimental Results

4.1 Dataset

The MOT17 dataset is part of the Multiple Object Tracking Benchmark (MOT Challenge), which provides a standardized evaluation framework for multi-object tracking algorithms [37]. MOT17 specifically refers to the 2017 version of this dataset. Here are some key details about the MOT17 dataset:

- **Data Content**: MOT17 contains a diverse set of video sequences captured in various scenarios such as pedestrian tracking in crowded streets and tracking of objects in surveillance videos [29]. It consists of 7 sequences and 5,316 frames for training and 7 sequences and 5919 frames for testing.

- **Annotations**: Each video sequence in the MOT17 dataset comes with ground truth annotations [30]. These annotations include information about the location and identity of each object (e.g., pedestrians or vehicles) in each frame of the video Figure 5.

- **Challenges**: MOT17 sequences are designed to challenge tracking algorithms with factors such as occlusions, scale variations and object interactions [31]. These challenges make it a valuable resource for evaluating the robustness of multi-object tracking algorithms.

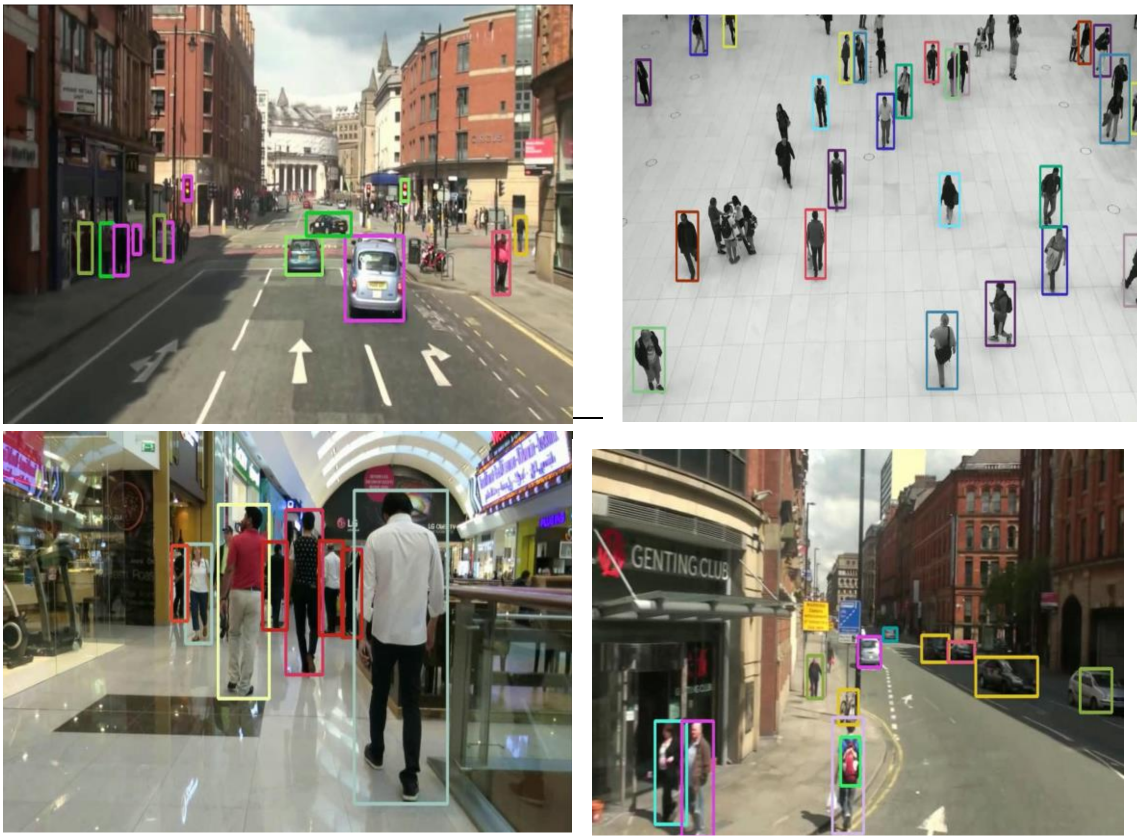


Figure 5: Sample tracking results visualization of StrongSORT on the test sets of MOT17. The box color corresponds to the ID

4.2 Metrics

A variety tracking performance is evaluated using a variety of indicators such as MOTA, IDs, IDF1, HOTA, AssA, and DetA

MOTA: MOTA primarily considers False Positives (FP), False Negatives (FN) and Identity Switches (IDs and prioritizes detection performance more highly**.**

$$\mathrm{MOTA}=1+\frac{\sum_t (FN_t+FP_t+IDSW_t)}{\sum_t GT_t} \qquad (3)$$

Where:

- FN_t is the number of false negatives (missed targets) at time t.
- FP_t is the number of false positives (false alarms) at time t
- $IDSW_t$ is the number of identity switches at time t
- GT_t is the number of ground truth objects at time t

IDF1 (Identity F1 Score): shows the identification recall and identity precision harmonic mean.

$$\mathbf{IDF1}=\frac{2.\mathrm{IDTP}}{2.\mathrm{IDTP}+\mathrm{FN}+\mathrm{FP}} \qquad (4)$$

Where:

- $IDTP$ is the number of correctly identified tracks
- FP is the number of false positives in track identities
- FN is the number of false negatives in track identities.

HOTA: offers a comprehensive evaluation by explicitly combining detection score (DetA) and association score (AssA). It strikes a balance between accurate detection and association making it a unified metric. Implementation HOTA is a more recent metric that aims to address some of the limitations of MOTA. It considers both the accuracy of object detection and the accuracy of object association

$$\mathrm{HOTA}=\sqrt{\frac{\sum_{t-1}^{T}\sum_{i-1}^{N_t}(TP(t,i))^{\alpha}}{\sum_{t-1}^{T}(N_t-M_t)+\sum_{i-1}^{N_t}(1-TP(t,i))^{\alpha}}} \qquad (5)$$

Where:

- T is the number of frames
- N_t is the number of ground truth objects in frame t
- M_t is the number of detected objects in frame t
- $TP(t,i)$ is 1 if the i-th ground truth object is correctly associated with a detection in frame t and 0 otherwise
- α is a parameter that controls the relative importance of detection accuracy and association accuracy. A common value is $\alpha = 2$.

AssA evaluates the tracker's ability to link detections to already-existing tracks. It is a part of the HOTA measurement.

DetA evaluates object detection precision without regard to association. It is a part of the HOTA measurement.

4.3 Detection

YOLOv8 stands out as an object detector by effectively balancing time efficiency and accuracy. It surpasses previous YOLO versions in accuracy while remaining competitive with other state-of-the-art object detection models. This balance makes YOLOv8 a compelling choice for real-world applications where both speed and precision are crucial.

4.4 Tracking:

DeepSORT an enhanced version of the SORT algorithm leverages the power of deep learning-based object detectors like YOLOv8. YOLOv8 excels at real-time object detection accurately identifying and bounding objects of interest within a scene. These detections are then fed into DeepSORT which uses a combination of appearance features and motion cues to estimate the likelihood of detections belonging to the same object across consecutive frames. This association process enables DeepSORT to track objects over time.

By combining the strengths of YOLOv8's detection capabilities and DeepSORT's tracking prowess, this integrated approach proves highly effective for applications such as surveillance, tracking and identification, where both rapid detection and persistent tracking are essential.

DeepSORT ranks first on MOT17 for metrics HOTA, IDF1, AssA and DetA and ranks second for MOTA and IDs. In particular, it yields an accurate association and outperforms the second-performance tracker by a large merge.

StrongSORT: achieves greater tracking accuracy and reliability by utilizing these improved features and association strategies. This makes it especially useful for applications where accurate and consistent tracking of multiple objects is crucial, like sports analytics, autonomous driving and surveillance.

For StrongSORT the matching distance threshold is established at 0.45. The warp mode for ECC is set to MOTION EUCLIDEAN. We employ a momentum term of 0.9 in EMA (Exponential Moving Average). The weight factor for appearance cost, denoted as alpha is set to 0.98.

StrongSORT ranks first on MOT17 for metrics HOTA, IDF1, AssA and DetA and ranks second for MOTA and IDs. In particular, it yields an accurate association and outperforms the second-performance tracker by a large merge.

StrongSORT is a highly effective algorithm for object tracking, known for its ability to maintain accurate object identities even in challenging scenarios. This robust performance stems from its innovative combination of a powerful appearance embedding model with a Kalman filter framework. This dual approach allows StrongSORT to leverage both visual information and motion dynamics, leading to more reliable and consistent tracking results. This enhanced tracking accuracy makes StrongSORT a valuable tool for various applications, including surveillance, autonomous driving, and video analysis

5 Conclusion and Future Work

In this paper, the accuracy became higher in deepsort and then improved when using strongsort. when using Deepsort with YoloV8,It became higher in HOTA(61.5), IDF1(74.6),MOTA(78.1), AssA(59.9), DetA(63.0), IDs(1,718).

This work presented a powerful integration of "StrongSORT" and "YOLOV8" for object tracking. By addressing limitations of its predecessor DeepSORT, "StrongSORT" achieves state-of-the-art tracking performance. When combined with the object detection capabilities of "YOLOV8" the resulting system offers robust and accurate real-time object tracking. when using strongsort ,It became higher in HOTA(68.432), IDF1(81.886),MOTA(85.356), AssA(66.332), DetA(65.559), IDs(625).

Even if StrongSORT's object tracking much outperforms DeepSORT's, there is still room for improvement [32]. In order to better handle abrupt changes in object appearance brought on by variables like viewpoint fluctuation or illumination variations, one approach would be to investigate more complex appearance models perhaps integrating transformer networks or attention mechanisms. To capture more complicated motion patterns and more precisely forecast object trajectories, the motion prediction models can be further improved by including extended short-term memory networks or other sequence learning approaches. To further improve tracking robustness in difficult situations, additional robust association metrics that incorporate appearance, motion and even contextual information should be investigated. To enchecment the strong sort to improve Tracking performance by using a variety of indicators MOTA, IDs, IDF1, HOTA, AssA and DetA.

References

1. Hu, H. Y. Monocular quasi-dense 3d object tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence2022
2. Y. Du *et al*., "StrongSORT: Make DeepSORT Great Again," in *IEEE Transactions on Multimedia*, vol. 25, pp. 8725-8737, 2023
3. . Han, S. H. (n.d.). Motion-aware multi-object tracking. Neurocomputing(2022)
4. Cao, J., Weng, X., Khirodkar, R., Pang, J., Kitani,K.: Observation-centric sort: Rethinking sort for robust multi-object tracking. arXiv preprint arXiv:2203.14360 (2022) .
5. Khurana, T., Dave, A., Ramanan, D.: Detecting invisible people. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 3174–3184 (2021).
6. Liang, C., Zhang, Z., Zhou, X., Li, B., Zhu, S., Hu,W.: Rethinking the competition between detection and reid in multiobject tracking. IEEE Transactions on Image Processing 31, 3182–3196 (2022) .
7. Stadler, D., Beyerer, J.: Modelling ambiguous assignments for multi-person tracking in crowds. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 133–14(2022).
8. Wu, J., Cao, J., Song, L., Wang, Y., Yang, M., Yuan, J.: Track to detect and segment: An online multi-object tracker. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 12352– 12361 (2021).
9. Sun, P., Cao, J., Jiang, Y., Yuan, Z., Bai, S., Kitani, K,Luo, P.: Dancetrack: Multi-object tracking in uniform appearance and diverse motion. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 20993–21002 (2022)
10. Wang, G., Wang, Y., Gu, R., Hu, W., Hwang, J.N.Split and connect: A universal tracklet booster for multiobject tracking. IEEE Transactions on Multimedia pp. 1–1(2022).
11. Wang, Q., Zheng, Y., Pan, P., Xu, Y.: Multiple object tracking with correlation learning. In:
12. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3876–3886 (2021)
13. Yu, E., Li, Z., Han, S., Wang, H.: Relationtrack: Relationaware multiple object tracking with

decoupled representation. IEEE Transactions on Multimedia pp. 1–1 (2022).
14. Zhang, Y., Wang, C., Wang, X., Zeng, W., Liu, W.Fairmot: On the fairness of detection and re-identification in multiple object tracking. International Journal of Computer Vision 129(11), 3069–3087 (2022).
15. Yang, F., Chang, X., Sakti, S., Wu, Y., Nakamura, S.Remot: A model-agnostic refinement for multiple object tracking. Image and Vision Computing 106, 104091(2021).
16. Gao, J., Nevatia, R.: Revisiting temporal modeling for video-based person reid. arXiv preprint arXiv:1805.02104 (2018)
17. . Fern´andez Llorca, D., Hern´andez Mart´ınez, A.,Garc´ıa Daza, I.: Vision-based vehicle speed estimation:A survey. IET Intelligent Transport Systems 15(8),987–1005 (2021).
18. Gao, T., Pan, H., Wang, Z., Gao, H.: A crf-based framework for tracklet inactivation in online multi-object tracking. IEEE Transactions on Multimedia 24, 995–1007 (2021).
19. Ge, Z., Liu, S., Wang, F., Li, Z., Sun, J.: Yolox: Exceeding yolo series in 2021. arXiv preprint arXiv:2107.08430 (2021).
20. Luo, H., Jiang, W., Gu, Y., Liu, F., Liao, X., Lai, S., Gu, J.: A strong baseline and batch normalization neck for deep person re-identification. IEEE Transactions on Multimedia 22(10), 2597–2609 (2019).
21. Yu, F., Li, W., Li, Q., Liu, Y., Shi, X., Yan, J.: Poi:Multiple object tracking with high performance detection and appearance feature. In: European Conference on Computer Vision. pp. 36–42. Springer (2016).
22. Tokmakov, P., Li, J., Burgard, W., Gaidon, A.: Learning to track with object permanence. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 10860–10869 (2021).
23. Zhu, Y., Zhou, K., Wang, M., Zhao, Y., Zhao, Z.: A comprehensive solution for detecting events in complex surveillance videos. Multimedia Tools and Applications 78(1), 817–838 (2019).
24. Shao, S., Zhao, Z., Li, B., Xiao, T., Yu, G., Zhang, X,Sun, J.: Crowdhuman: A benchmark for detecting humanin a crowd. arXiv preprint arXiv:1805.00123 (2018)
25. Yang, F., Chang, X., Sakti, S., Wu, Y., Nakamura, S.: Remot: A model-agnostic refinement for multiple object tracking. Image and Vision Computing 106, 104091 (2021)
26. Bergmann, P., Meinhardt, T., Leal-Taixe, L.: Tracking without bells and whistles. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 941–951 (2019).
27. Bergmann, P., Meinhardt, T., Leal-Taixe, L.: Tracking without bells and whistles. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 941–951 (2019).
28. Wojke, N., Bewley, A., Paulus, D.: Simple online and realtime tracking with a deep association metric. In:2017 IEEE international conference on image processing (ICIP). pp. 3645–3649. IEEE (2017).
29. Redmon, J., Farhadi, A.: Yolov3: An incremental improvement.arXiv preprint arXiv:1804.02767 (2018)
30. Fu, Z., Angelini, F., Chambers, J., Naqvi, S.M.: Multilevel cooperative fusion of gm-phd filters for online multiple human tracking. IEEE Transactions on Multimedia 21(9), 2277–2291 (2019).
31. Wang, G., Wang, Y., Zhang, H., Gu, R., Hwang, J.N.: Exploit the connectivity: Multi-object tracking with trackletnet. In: Proceedings of the 27th ACM International Conference on Multimedia. pp. 482–490 (2019).

32. Ristani, E., Solera, F., Zou, R., Cucchiara, R., Tomasi,C.: Performance measures and a data set for multitarget,multi-camera tracking. In: European conference on computer vision. pp. 17–35. Springer (2016).
33. Ristani, E., Solera, F., Zou, R., Cucchiara, R., Tomasi,C.: Performance measures and a data set for multitarget,multi-camera tracking. In: European conference on computer vision. pp. 17–35. Springer (2016).
34. Milan, A., Leal-Taix´e, L., Reid, I., Roth, S., Schindler, K.: Mot16: A benchmark for multi-object tracking. arXiv preprint arXiv:1603.00831 (2016).\
35. Dai, P., Wang, X., Zhang, W., Chen, J.: Instance segmentation enabled hybrid data association and discriminative hashing for online multi-object tracking. IEEE Transactions on Multimedia 21(7), 1709–1723 (2018).
36. Naiel, M.A., Ahmad, M.O., Swamy, M., Lim, J., Yang,M.H.: Online multi-object tracking via robust collaborative model and sample selection. Computer Vision and Image Understanding 154, 94–107 (2017).
37. Zhang, S., Benenson, R., Schiele, B.: Citypersons: Adiverse dataset for pedestrian detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3213–3221 (2017).
38. Wang, B., Wang, G., Chan, K.L., Wang, L.: Tracklet association by online target-specific metric learning and coherent dynamics estimation. IEEE transactions on pattern analysis and machine intelligence 39(3), 589–602 (2016).
39. Ren, S., He, K., Girshick, R., Sun, J.: Faster r-cnn: Towards real-time object detection with region proposal networks. Advances in neural information processing systems 28 (2015).