

# Pedestrian Tracking in Autonomous Vehicles

Subjects: **Computer Science, Artificial Intelligence**

Contributor: Majdi Sukkar , Madhu Shukla , Dinesh Kumar , Vassilis C. Gerogiannis , Andreas Kanavos , Biswaranjan Acharya

Effective collision risk reduction in autonomous vehicles relies on robust and straightforward pedestrian tracking. Challenges posed by occlusion and switching scenarios significantly impede the reliability of pedestrian tracking.

pedestrian tracking

object detection

multi-object tracking (MOT)

autonomous vehicles

## 1. Introduction

In recent years, the significance of pedestrian tracking in autonomous vehicles has garnered considerable attention due to its pivotal role in ensuring pedestrian safety. Existing state-of-the-art approaches for pedestrian tracking in autonomous vehicles predominantly rely on object detection and tracking algorithms, including Faster R-CNN (Region-based Convolutional Neural Network), YOLO (You Only Look Once) and SORT algorithms [1]. However, these methods encounter issues in scenarios where pedestrians are occluded or only partially visible [2].

More specifically, the challenges in pedestrian tracking are multifaceted, ranging from crowded urban environments to unpredictable pedestrian behavior [3]. Existing algorithms often struggle to handle scenarios where individuals move behind obstacles, cross paths, or exhibit sudden changes in direction. Furthermore, adverse weather conditions, low lighting, and dynamic urban landscapes pose additional hurdles for accurate pedestrian tracking [4]. These issues underscore the need for advanced pedestrian tracking solutions that can adapt to diverse and complex real-world scenarios.

In response to these challenges, the research takes inspiration from recent advancements in deep learning and object detection. Deep learning techniques, with their ability to learn intricate patterns and representations from data, have shown promise in overcoming the limitations of traditional tracking methods. Leveraging the capabilities of the YOLOv8 algorithm for object detection, the approach aims to enhance the accuracy and robustness of pedestrian tracking in dynamic and challenging environments [5]. By addressing the limitations of existing algorithms, researchers aspire to contribute to the development of pedestrian tracking systems that are both reliable and adaptable.

The quest for improved pedestrian tracking is not solely confined to the domain of autonomous vehicles [6]. The relevant applications extend to various fields, such as surveillance, crowd management, and human-computer interaction. Accurate pedestrian tracking is crucial for ensuring public safety, optimizing traffic flow, and enhancing the overall efficiency of smart city initiatives [7][8]. The effectiveness of pedestrian tracking is also integral for a

myriad of applications in the context of urban environments, and in smart cities, where the integration of technology aims to enhance the quality of life, with pedestrian tracking playing a crucial role. Efficient tracking systems can contribute to optimized traffic management, improved public safety, and enhanced urban planning <sup>[9]</sup>. Beyond traffic applications, pedestrian tracking finds applications in surveillance, where monitoring and analyzing pedestrian movement are essential for security <sup>[10]</sup>. Additionally, in human–computer interaction scenarios, accurate tracking is pivotal for creating responsive and adaptive interfaces, offering a wide array of possibilities for innovative applications <sup>[11]</sup>. Therefore, the advancements in pedestrian tracking have far-reaching implications, influencing various aspects of the daily lives and the development of smart city ecosystems.

## **| 2. Pedestrian Tracking in Autonomous Vehicles**

Recent advancements in pedestrian tracking encompass a diverse array of methods and algorithms, broadly classified into the following categories:

- **Multi-Object Tracking (MOT) Methods:**
  - These methods are designed to concurrently track multiple pedestrians within a scene. Traditional approaches often employ the Hungarian algorithm for association, linking detections across frames <sup>[5]</sup>.
  - Recent advancements leverage deep neural networks, such as TrackletNet <sup>[12]</sup> and DeepSORT, to learn features and association scores, enhancing tracking accuracy.
- **Re-identification-based Methods:**
  - This category integrates appearance-based and geometric feature-based techniques to identify and track pedestrians across different camera views. Notably, Siamese networks are employed to learn a similarity metric between pairs of images <sup>[13]</sup>.
  - Contemporary methods incorporate attention mechanisms to emphasize discriminative regions of pedestrians in order to improve tracking precision <sup>[14]</sup>.
- **Online Multi-Object Tracking:** Online tracking methods dynamically track multiple pedestrians in real-time. Techniques include deep learning-based re-identification <sup>[15]</sup> and online multiple hypothesis tracking <sup>[16]</sup>.
- **Multi-Cue Fusion:** Strategies in this category amalgamate various cues, such as color, shape, and motion, to enhance tracking robustness in complex scenes. For instance, the multi-cue multi-camera pedestrian tracking (MCMC-PT) method integrates color, shape, and motion cues from multiple cameras for comprehensive pedestrian tracking <sup>[15]</sup>.

This diverse landscape of pedestrian tracking methodologies underscores the ongoing efforts to address challenges in occlusion, partial visibility, and dynamic scenarios. Each approach brings its unique strengths,

contributing to the advancement of pedestrian tracking in autonomous vehicle applications.

In the research presented in [17], three frameworks for multi-object tracking were evaluated: tracking-by-detection (TBD), joint-detection-and-tracking (JDT), and a transformer-based tracking method. DeepSORT and StrongSORT were classified under the TBD framework. The front-end detector's performance significantly impacts tracking, and enhancing it is crucial. The transformer-based framework excels in MOTA (multiple object tracking accuracy) but has a large model size, while the JDT framework balances accuracy and real-time performance.

In [18], an occlusion handling strategy for a multi-pedestrian tracker is proposed, capable of retrieving targets without the need for re-identification models. The tracker can manage inactive tracks and cope with tracks leaving the camera's field of view, achieving state-of-the-art results on three popular benchmarks. However, the performed comparison does not include DeepSORT or StrongSORT algorithms.

The authors of [19] propose a framework model based on YOLOv5 and StrongSORT tracking algorithms for real-time monitoring and tracking of workers wearing safety helmets in construction scenarios. The use of deep learning-based object detection and tracking algorithms improves accuracy and efficiency in helmet-wearing detection. The study suggests that changing the box regression loss function from CIOU to Focal-EIOU can further improve detection performance. However, the study's limitation is its evaluation on a specific dataset, potentially limiting generalization to other datasets or scenarios.

The VOT2020 challenge assessed different tracking scenarios, introducing innovations like using segmentation masks instead of bounding boxes and new evaluation methods. Most trackers relied on deep learning, particularly Siamese networks, emphasizing the role of AI and deep learning in advancing object tracking for future improvements [20]. In [21], the utilization of the YOLOv5 model and the StrongSORT algorithm for ship detection, classification, and tracking in maritime surveillance systems is examined. The practical results demonstrate high accuracy in ship classification and the capability to track at a speed approaching real-time. The influence of StrongSORT contributes to enhancing tracking speed, confirming its effectiveness in maritime surveillance systems.

In [22], fine-tuning plays a crucial role in adapting the pre-trained YOLOv5 model for brain tumor detection, significantly enhancing the model's performance in identifying specific brain tumors within radiological images. In another study [23], a direct comparison with the performance of StrongSORT in its advanced version was conspicuously omitted. While utilizing the KC-YOLO approach based on YOLOv5 in their detection algorithm, the study did not surpass the capabilities of more recent versions like YOLOv8. Introducing newer updates could potentially enhance StrongSORT++, as illustrated in [4], showcasing its substantial outperformance across various metrics, including HOTA and IDF1, on the MOT17 and MOT20 datasets.

The impact of fine-tuning is evident in a machine learning study [24], where deep learning models identified diseases in maize leaves. Fine-tuning significantly improved the performance of pre-trained models, resulting in disease classification accuracy rates exceeding 93%. VGG16, InceptionV3, and Xception achieved accuracy rates

surpassing 99%, demonstrating the effectiveness of transfer learning, as previously shown in [25], and the positive impact of fine-tuning on disease detection in maize leaves [24].

In **Table 1**, a comparison is presented between StrongSORT and other multi-object tracking methods with a focus on the advantages and disadvantages of its method.

**Table 1.** Comparison overview between StrongSORT and other multi-object trackers.

Method	Advantages	Disadvantages
StrongSORT [4]	<ul style="list-style-type: none"><li>Simple and efficient</li><li>Robustness to occlusions</li><li>Robustness to identity switches</li></ul>	<ul style="list-style-type: none"><li>Runs relatively slowly</li><li>Negligible extra computational cost</li></ul>
DeepSORT [26]	Simple and efficient	Limited robustness to occlusions and identity switches
TrackletNet [12]	Robust to occlusions and identity switches	Complex architecture and not real-time
DMAN [27]	Robust to occlusions and identity switches	Not real-time
SORT [28]	Simple and efficient	Limited robustness to occlusions and identity switches
ATOM [29]	Accurate tracking	Complex architecture and not real-time
IVDM [23]	Enhanced handling of occlusions and identity switches	Needs evaluation against real-time performance

In recognizing the identified limitations within StrongSORT, such as a slightly slower runtime in specific cases, it becomes imperative to strike a balance in utilizing its features while simultaneously addressing challenges in diverse implementation environments. Future research endeavors will delve into a meticulous examination of the algorithm's efficiency concerning occlusions and identity switches.

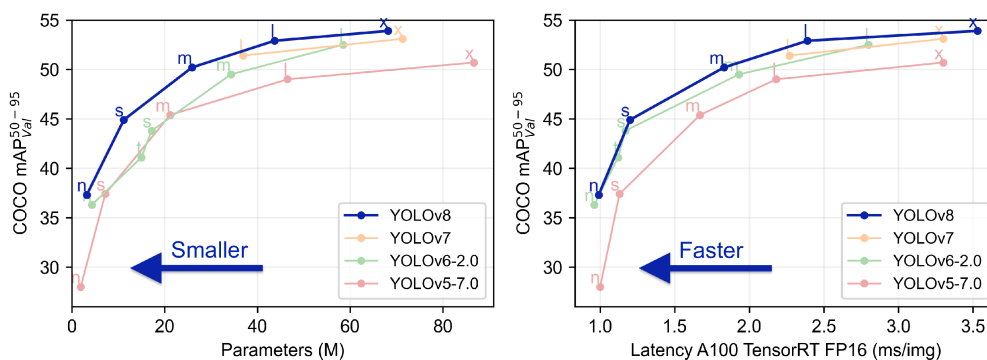
## 2.1. YOLOv8

YOLOv8 (You Only Look Once version 8) represents a significant advancement over its predecessors, YOLOv7 and YOLOv6, incorporating multiple features that enhance both speed and accuracy. A notable addition is the spatial pyramid pooling (SPP) module, enabling YOLOv8 to extract features at varying scales and resolutions [30]. This facilitates the precise detection of objects of different sizes. Another key feature is the cross stage partial

network (CSP) block, reducing the network's parameters without compromising accuracy, thereby improving both training times and overall performance [31].

Evaluation of prominent object detection benchmarks, including COCO and Pascal VOC datasets, showcases YOLOv8's exceptional capabilities. The COCO dataset, with 80 object categories and complex scenes, saw YOLOv8 achieving the highest-ever mean average precision (mAP) score of 55.3% among single-stage object detection algorithms. Additionally, it achieved a remarkable real-time speed of 58 frames per second on an NVIDIA GTX 1080 Ti GPU. In summary, YOLOv8 stands out as an impressive object detection algorithm, delivering high accuracy and real-time performance. Its ability to process the entire image simultaneously makes it well suited for applications such as autonomous driving and robotics. With its innovative features and outstanding performance, YOLOv8 is poised to remain a preferred choice for object detection tasks in the foreseeable future [32].

**Figure 1** illustrates the performance of YOLOv8 in real-world scenarios [33].



**Figure 1.** YOLOv8 performance.

## 2.2. DeepSORT

DeepSORT is an advanced real-time multiple objects tracking algorithm that leverages deep learning-based feature extraction coupled with the Hungarian algorithm for assignment. The code structure of DeepSORT comprises the following key components [34]:

- **Feature Extraction:** Responsible for extracting features from input video frames, including bounding boxes and corresponding features.
- **Detection and Tracking:** Detects objects in each video frame and associates them with their tracks using the Hungarian algorithm.
- **Kalman Filter:** Predicts the location of each object in the next video frame based on its previous location and velocity.
- **Appearance Model:** Stores and updates the appearance features of each object over time, facilitating re-identification and appearance updates.

- Re-identification: Matches the appearance of an object in one video frame with its appearance in a previous frame.
- Output: Generates the final output—a set of object tracks for each video frame.

These components collaboratively form a comprehensive tracking algorithm. DeepSORT can undergo training using a substantial dataset of video frames and corresponding object bounding boxes. The training process of DeepSORT fine-tunes the various algorithm's components, such as the appearance model and feature extraction, aiming to enhance the algorithm's overall performance. Despite its strengths, DeepSORT also faces unique challenges, including:

- Blurred Objects: Tracking difficulties due to image artifacts caused by blurred objects [\[35\]](#).
- Intra-object Changes: Challenges in handling changes in the shape or size of objects [\[36\]](#).
- Non-rigid Objects: Difficulty in tracking objects appearing for short durations [\[37\]](#).
- Transparent Objects: Challenges in detecting objects made of transparent materials.
- Non-linear Motion: Difficulty in tracking irregularly moving objects [\[38\]](#).
- Fast Motion: Challenges posed by quickly moving objects.
- Similar Objects: Difficulty in differentiating objects with similar appearances [\[39\]](#).
- Occlusion: Tracking challenges when objects overlap or obstruct each other [\[15\]](#).
- Scale Variation: Difficulty when objects appear at different scales in the image [\[40\]](#).

To address these challenges, the StrongSORT algorithm has been proposed aiming to offer solutions which enhance tracking robustness and efficiency [\[4\]](#).

## 2.3. StrongSORT

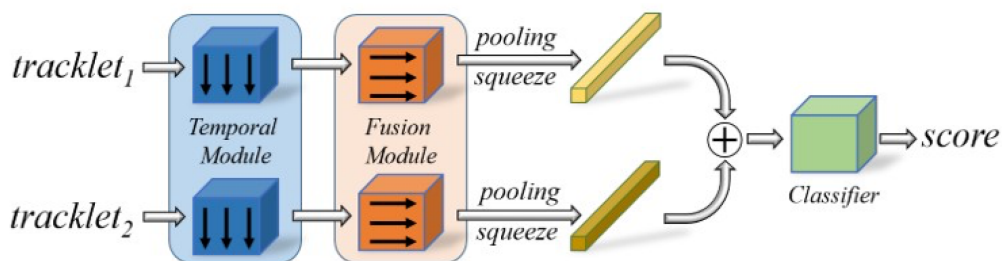
The StrongSORT algorithm enhances the original DeepSORT by introducing the AFLink algorithm for temporal matching to person tracking and the GSI algorithm for temporal interpolation of matched individuals. New configuration options have been incorporated to facilitate these improvements [\[4\]](#):

- AFLink: A flag indicating whether the AFLink algorithm should be utilized for temporal matching.
- Path\_AFLink: The path to the AFLink algorithm model to be employed.

- GSI: A flag indicating whether the GSI algorithm should be employed for temporal interpolation.
- Interval: The temporal interval to be applied in the GSI algorithm.
- Tau: The temporal interval to be used in the GSI algorithm.

The original DeepSORT application is updated through the integration of the AFLink and GSI algorithms, providing enhanced capabilities for temporal matching and interpolation of tracked individuals.

**Figure 2** presents the framework of the AFLink model. It adopts the spatio-temporal information of two tracklets as input and predicts their connectivity [\[4\]](#).



**Figure 2.** Framework of the AFLink model.

## References

1. Razzok, M.; Badri, A.; Mourabit, I.E.; Ruichek, Y.; Sahel, A. Pedestrian Detection and Tracking System Based on Deep-SORT, YOLOv5, and New Data Association Metrics. *Information* 2023, 14, 218.
2. Bhola, G.; Kathuria, A.; Kumar, D.; Das, C. Real-time Pedestrian Tracking based on Deep Features. In *Proceedings of the 4th International Conference on Intelligent Computing and Control Systems (ICICCS)*, Madurai, India, 13–15 May 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 1101–1106.
3. Li, R.; Zu, Y. Research on Pedestrian Detection Based on the Multi-Scale and Feature-Enhancement Model. *Information* 2023, 14, 123.
4. Du, Y.; Zhao, Z.; Song, Y.; Zhao, Y.; Su, F.; Gong, T.; Meng, H. StrongSORT: Make DeepSORT Great Again. *IEEE Trans. Multimed.* 2023, 25, 8725–8737.
5. Dendorfer, P.; Rezatofighi, H.; Milan, A.; Shi, J.; Cremers, D.; Reid, I.D.; Roth, S.; Schindler, K.; Leal-Taixé, L. MOT20: A Benchmark for Multi Object Tracking in Crowded Scenes. *arXiv* 2020, arXiv:2003.09003.

6. Xiao, C.; Luo, Z. Improving Multiple Pedestrian Tracking in Crowded Scenes with Hierarchical Association. *Entropy* 2023, 25, 380.
7. Myagmar-Ochir, Y.; Kim, W. A Survey of Video Surveillance Systems in Smart City. *Electronics* 2023, 12, 3567.
8. Tao, M.; Li, X.; Xie, R.; Ding, K. Pedestrian Identification and Tracking within Adaptive Collaboration Edge Computing. In *Proceedings of the 26th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, Rio de Janeiro, Brazil, 24–26 May 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 1124–1129.
9. Son, T.H.; Weedon, Z.; Yigitcanlar, T.; Sanchez, T.; Corchado, J.M.; Mehmood, R. Algorithmic Urban Planning for Smart and Sustainable Development: Systematic Review of the Literature. *Sustain. Cities Soc.* 2023, 94, 104562.
10. AL-Dosari, K.; Hunaiti, Z.; Balachandran, W. Systematic Review on Civilian Drones in Safety and Security Applications. *Drones* 2023, 7, 210.
11. Vasiljevas, M.; Damaševičius, R.; Maskeliūnas, R. A Human-Adaptive Model for User Performance and Fatigue Evaluation during Gaze-Tracking Tasks. *Electronics* 2023, 12, 1130.
12. Wang, G.; Wang, Y.; Zhang, H.; Gu, R.; Hwang, J. Exploit the Connectivity: Multi-Object Tracking with TrackletNet. In *Proceedings of the 27th ACM International Conference on Multimedia (MM)*, Nice, France, 21–25 October 2019; pp. 482–490.
13. Li, R.; Zhang, B.; Kang, D.; Teng, Z. Deep Attention Network for Person Re-Identification with Multi-loss. *Comput. Electr. Eng.* 2019, 79, 106455.
14. Jiao, S.; Wang, J.; Hu, G.; Pan, Z.; Du, L.; Zhang, J. Joint Attention Mechanism for Person Re-Identification. *IEEE Access* 2019, 7, 90497–90506.
15. Guo, W.; Jin, Y.; Shan, B.; Ding, X.; Wang, M. Multi-Cue Multi-Hypothesis Tracking with Re-Identification for Multi-Object Tracking. *Multimed. Syst.* 2022, 28, 925–937.
16. Kang, W.; Xie, C.; Yao, J.; Xuan, L.; Liu, G. Online Multiple Object Tracking with Recurrent Neural Networks and Appearance Model. In *Proceedings of the 13th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, Chengdu, China, 17–19 October 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 34–38.
17. Guo, S.; Wang, S.; Yang, Z.; Wang, L.; Zhang, H.; Guo, P.; Gao, Y.; Guo, J. A Review of Deep Learning-Based Visual Multi-Object Tracking Algorithms for Autonomous Driving. *Appl. Sci.* 2022, 12, 10741.
18. Stadler, D.; Beyerer, J. Improving Multiple Pedestrian Tracking by Track Management and Occlusion Handling. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Virtual, 19–25 June 2021; pp. 10958–10967.



19. Li, F.; Chen, Y.; Hu, M.; Luo, M.; Wang, G. Helmet-Wearing Tracking Detection Based on StrongSORT. *Sensors* 2023, 23, 1682.
20. Kristan, M.; Leonardis, A.; Matas, J.; Felsberg, M.; Pflugfelder, R.P.; Kämäräinen, J.; Danelljan, M.; Zajc, L.C.; Lukežić, A.; Drbohlav, O.; et al. The 8th Visual Object Tracking VOT2020 Challenge Results. In *Proceedings of the Workshops on Computer Vision (ECCV)*, Glasgow, UK, 23–28 August 2020; *Lecture Notes in Computer Science*. Springer: Berlin/Heidelberg, Germany, 2020; Volume 12539, pp. 547–601.
21. Pham, Q.H.; Doan, V.S.; Pham, M.N.; Duong, Q.D. Real-Time Multi-vessel Classification and Tracking Based on StrongSORT-YOLOv5. In *Proceedings of the International Conference on Intelligent Systems & Networks (ICISN)*; *Lecture Notes in Networks and Systems*. Springer: Berlin/Heidelberg, Germany, 2023; Volume 752, pp. 122–129.
22. Shelatkar, T.; Bansal, U. Diagnosis of Brain Tumor Using Light Weight Deep Learning Model with Fine Tuning Approach. In *Proceedings of the International Conference on Machine Intelligence and Signal Processing (MISP)*; *Lecture Notes in Electrical Engineering*. Springer: Berlin/Heidelberg, Germany, 2022; Volume 998, pp. 105–114.
23. Li, J.; Wu, W.; Zhang, D.; Fan, D.; Jiang, J.; Lu, Y.; Gao, E.; Yue, T. Multi-Pedestrian Tracking Based on KC-YOLO Detection and Identity Validity Discrimination Module. *Appl. Sci.* 2023, 13, 12228.
24. Subramanian, M.; Shanmugavadivel, K.; Nandhini, P.S. On Fine-Tuning Deep Learning Models Using Transfer Learning and Hyper-Parameters Optimization for Disease Identification in Maize Leaves. *Neural Comput. Appl.* 2022, 34, 13951–13968.
25. Sukkar, M.; Kumar, D.; Sindha, J. Improve Detection and Tracking of Pedestrian Subclasses by Pre-Trained Models. *J. Adv. Eng. Comput.* 2022, 6, 215.
26. Kapania, S.; Saini, D.; Goyal, S.; Thakur, N.; Jain, R.; Nagrath, P. Multi Object Tracking with UAVs using Deep SORT and YOLOv3 RetinaNet Detection Framework. In *Proceedings of the 1st ACM Workshop on Autonomous and Intelligent Mobile Systems*, Linz, Austria, 8–10 July 2020; pp. 1–6.
27. Zhu, J.; Yang, H.; Liu, N.; Kim, M.; Zhang, W.; Yang, M. Online Multi-Object Tracking with Dual Matching Attention Networks. In *Proceedings of the 15th European Conference on Computer Vision (ECCV)*, Munich, Germany, 8–14 September 2018; *Lecture Notes in Computer Science*. Springer: Berlin/Heidelberg, Germany, 2018; Volume 11209, pp. 379–396.
28. Bewley, A.; Ge, Z.; Ott, L.; Ramos, F.T.; Upcroft, B. Simple Online and Realtime Tracking. In *Proceedings of the International Conference on Image Processing (ICIP)*, Phoenix, AZ, USA, 25–28 September 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 3464–3468.
29. Danelljan, M.; Bhat, G.; Khan, F.S.; Felsberg, M. ATOM: Accurate Tracking by Overlap Maximization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern*

- Recognition (CVPR) Computer Vision Foundation; IEEE: Piscataway, NJ, USA, 2019; pp. 4660–4669.
30. Guo, M.; Xue, D.; Li, P.; Xu, H. Vehicle Pedestrian Detection Method Based on Spatial Pyramid Pooling and Attention Mechanism. *Information* 2020, 11, 583.
  31. Hussain, M. YOLO-v1 to YOLO-v8, the Rise of YOLO and Its Complementary Nature toward Digital Manufacturing and Industrial Defect Detection. *Machines* 2023, 11, 677.
  32. Sirisha, U.; Praveen, S.P.; Srinivasu, P.N.; Barsocchi, P.; Bhoi, A.K. Statistical Analysis of Design Aspects of Various YOLO-Based Deep Learning Models for Object Detection. *Int. J. Comput. Intell. Syst.* 2023, 16, 126.
  33. Ultralytics YOLOv8. Available online: <https://github.com/ultralytics/ultralytics> (accessed on 28 January 2024).
  34. Wojke, N.; Bewley, A.; Paulus, D. Simple Online and Realtime Tracking with a Deep Association Metric. In *Proceedings of the International Conference on Image Processing (ICIP)*, Beijing, China, 17–20 September 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 3645–3649.
  35. Guo, Q.; Feng, W.; Gao, R.; Liu, Y.; Wang, S. Exploring the Effects of Blur and Deblurring to Visual Object Tracking. *IEEE Trans. Image Process.* 2021, 30, 1812–1824.
  36. Janai, J.; Güney, F.; Behl, A.; Geiger, A. Computer Vision for Autonomous Vehicles: Problems, Datasets and State of the Art. *Found. Trends Comput. Graph. Vis.* 2020, 12, 1–308.
  37. Meimetis, D.; Daramouskas, I.; Perikos, I.; Hatzilygeroudis, I. Real-Time Multiple Object Tracking Using Deep Learning Methods. *Neural Comput. Appl.* 2023, 35, 89–118.
  38. Wang, Z.; Zheng, L.; Liu, Y.; Li, Y.; Wang, S. Towards Real-Time Multi-Object Tracking. In *Proceedings of the 16th European Conference on Computer Vision (ECCV)*, Glasgow, UK, 23–28 August 2020; *Lecture Notes in Computer Science*. Springer: Berlin/Heidelberg, Germany, 2020; Volume 12356, pp. 107–122.
  39. Ciaparrone, G.; Sánchez, F.L.; Tabik, S.; Troiano, L.; Tagliaferri, R.; Herrera, F. Deep Learning in Video Multi-Object Tracking: A Survey. *Neurocomputing* 2020, 381, 61–88.
  40. Song, S.; Li, Y.; Huang, Q.; Li, G. A New Real-Time Detection and Tracking Method in Videos for Small Target Traffic Signs. *Appl. Sci.* 2021, 11, 3061.

---

Retrieved from <https://encyclopedia.pub/entry/history/show/125566>