International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 01 | Jan - 2026 SJIF Rating: 8.586 ISSN: 2582-3930

Object Detection and Tracking Using YOLO & SORT

Akash A Navale
Information Science and Engineering
Jawaharlal Nehru New College of Engineering
Shimoga, India
navaleakash810@gmail.com

Adithya S S
Information Science and Engineering
Jawaharlal Nehru New College of Engineering
Shimoga, India
adithyass1899@gmail.com

Aakash S Ganiger
Information Science and Engineering
Jawaharlal Nehru New College of Engineering
Shimoga, India
akashganiger2004@gmail.com

Adarsh B K
Information Science and Engineering
Jawaharlal Nehru New College of Engineering
Shimoga, India
adarshgoudru7@gmail.com

Mrs. Rashmi R. B.E, M.Tech
Associate Professor,
Information Science and Engineering
Jawaharlal Nehru New College of Engineering
Shimoga, India
rashmiindnraj@jnnce.ac.in

***Abstract*— Object detection and tracking are fundamental tasks in computer vision that enable real-time analysis of visual data in dynamic environments. This research paper presents a comprehensive study of a YOLOv11 and DeepSORT-based object detection and tracking system designed for robust performance in real-world scenarios. The proposed system integrates YOLOv11, one of the latest deep learning detection models, with DeepSORT, a sophisticated multi-object tracking algorithm that combines Kalman Filter motion prediction and appearance-based feature embeddings. Enhanced architectural components including C3k2 blocks, SPPF, and C2PSA attention modules ensure high accuracy detection even under challenging conditions such as occlusion, varying illumination, and rapid motion. The system achieves above 25 FPS on mid-range computing hardware while maintaining high detection accuracy and stable tracking. Applications range from surveillance and traffic monitoring to autonomous vehicles and intelligent security systems. This paper provides detailed architectural design, mathematical formulations, implementation methodology, and comprehensive experimental results demonstrating the effectiveness of the proposed approach.**

I. INTRODUCTION

Object detection and tracking represent cornerstone technologies in the field of computer vision and artificial intelligence. These capabilities enable machines to automatically identify and locate objects such as humans, vehicles, animals, and other items in images or video streams, and subsequently track their movement across consecutive frames. Today, these technologies are integral to numerous real-world applications including CCTV surveillance systems, traffic monitoring, autonomous vehicles, industrial automation, and sports analysis. Object detection fundamentally answers the question of what and where objects are located within visual data, while tracking extends this capability by following these objects over time. [1], [4], [12]

In practical terms, object detection in a shopping mall security camera system helps identify a new person entering the frame, while tracking maintains awareness of that person's path and movements until they exit the viewing area. Such capabilities enable systems to understand and respond to dynamic scenes in real-time, providing actionable intelligence for safety, security, and operational efficiency. Modern deep learning approaches have revolutionized these fields, with algorithms like YOLO (You Only Look Once) setting new standards for real-time detection performance and SORT (Simple Online and Realtime Tracking) providing efficient tracking mechanisms. [10], [11], [5]

However, significant challenges persist in practical deployments. Occlusion, where one object obscures another, creates ambiguity in both detection and tracking. Changing lighting conditions, shadows, reflections, and dynamic backgrounds all introduce noise and complexity. Maintaining real-time processing speeds without sacrificing accuracy remains challenging, particularly for resource-constrained devices. Beyond safety and surveillance, object detection and tracking find applications in healthcare for patient monitoring, agriculture for livestock and crop management, retail for customer behavior analysis, and transportation for traffic management and accident prevention. [2], [3], [11]

To address these challenges, recent research has increasingly focused on integrating high-speed deep learning–based detectors with robust multi-object tracking frameworks. Advanced detection models such as the latest YOLO variants offer improved accuracy and inference speed, while tracking algorithms enhanced with motion prediction and appearance feature modeling help preserve object identities over time. By combining detection and tracking into a unified pipeline, modern systems are able to achieve reliable real-time performance even in complex, crowded, and dynamically changing environments. Such integrated approaches form the foundation for intelligent vision systems capable of operating effectively in real-world scenarios where accuracy, speed, and robustness are equally critical. [6], [12]

Furthermore, the demand for scalable and adaptable object detection and tracking systems has grown with the widespread deployment of camera-based sensing in smart cities and intelligent infrastructure. Real-world environments are unpredictable, with varying camera angles, crowd densities, and motion patterns that challenge static or narrowly trained

models. An effective system must not only perform well in controlled settings but also generalize across diverse scenarios while maintaining consistent accuracy and low latency. This necessity has driven the development of architectures that emphasize efficient feature extraction, attention mechanisms, and optimized tracking strategies, enabling dependable operation across a broad range of applications and hardware platforms. [3], [6], [11]

In conclusion, object detection and tracking have become indispensable components of modern intelligent vision systems, enabling machines to perceive, interpret, and respond to dynamic visual environments. Advances in deep learning–based detection and robust tracking techniques have significantly improved the reliability and real-time performance of these systems in complex real-world conditions. By addressing challenges such as occlusion, illumination variation, and computational constraints, integrated detection–tracking frameworks continue to expand their applicability across surveillance, transportation, industrial, and analytical domains, reinforcing their critical role in the evolution of computer vision–driven technologies. [1], [4], [12]

II. Related Work

[1] Real-time object detection and tracking are critical components of many modern computer vision applications, including surveillance systems, traffic monitoring, and autonomous platforms. With rapid advances in deep learning and artificial intelligence, it has become possible to perform accurate real-time object detection and tracking using state-of-the-art algorithms. One such approach combines the YOLOv11 detection model with the DeepSORT tracking algorithm, forming a robust and efficient multi-object detection and tracking framework. YOLOv11 is a single-stage deep learning detector optimized for speed and accuracy, making it suitable for real-time video analysis in dynamic environments. The model incorporates advanced architectural components such as C3k2 blocks, SPPF layers, and attention mechanisms to improve feature extraction and detection performance. DeepSORT enhances traditional tracking methods by integrating Kalman Filter-based motion prediction with deep appearance feature embeddings, enabling reliable identity association across consecutive frames. The combination of detection accuracy and tracking stability allows the system to maintain object identities even under occlusion, illumination variation, and rapid movement. This integrated framework is capable of processing video streams in real-time while maintaining high detection precision and consistent tracking behavior. The ability to detect and track multiple objects simultaneously is especially valuable in scenarios where timely analysis and decision-making are required. By processing video input continuously, the system can generate detection and tracking outputs that support higher-level reasoning and automated responses. Such capabilities are essential in applications involving safety, monitoring, and intelligent automation. The balance between computational efficiency and robustness makes this approach suitable for deployment on mid-range computing platforms without requiring specialized hardware. The integration of YOLOv11 and DeepSORT demonstrates how modern deep learning techniques can be effectively combined to address long-standing challenges in object detection and tracking. In conclusion, real-time object detection and tracking using a YOLOv11 and DeepSORT-based framework provides a powerful solution for intelligent vision applications. Continued advancements in deep learning architectures and tracking algorithms are expected to further improve performance, scalability, and reliability in future systems across a wide range of real-world domains.

[2] Object detection and tracking are essential components of many computer vision applications, and accurate real-time analysis of visual data can significantly improve system reliability and safety. To address this requirement, an object detection and tracking system based on a YOLOv11 and DeepSORT framework has been proposed. The proposed system processes video streams to detect and track multiple objects with high accuracy and real-time performance. YOLOv11 is employed as the primary detection model due to its optimized single-stage architecture and improved feature extraction capability. The model integrates advanced architectural components such as C3k2 blocks, SPPF layers, and spatial attention mechanisms to enhance detection robustness. DeepSORT is used as the tracking module to maintain object identities across consecutive frames. The tracking algorithm combines Kalman Filter-based motion prediction with deep appearance feature embeddings to achieve reliable data association. The motion model predicts object trajectories, while the appearance model reduces identity switches during occlusion and overlap. The proposed system demonstrates strong detection accuracy and stable tracking performance in dynamic environments. It is capable of handling challenges such as illumination variation, partial occlusion, and rapid object movement. Experimental evaluation shows that the system operates at real-time speeds while maintaining consistent tracking accuracy. The integrated framework supports continuous video analysis and generates reliable tracking outputs suitable for higher-level decision-making systems. The system performs efficiently on mid-range computing platforms without requiring specialized hardware. Compared to traditional detection-only or motion-only tracking approaches, the proposed system provides improved robustness and scalability. In addition to accuracy and efficiency, the system offers practical deployment advantages. It can be integrated into existing surveillance, traffic monitoring, and autonomous systems with minimal modification. The modular design allows easy adaptation to different environments and application requirements, making it suitable for a wide range of intelligent vision-based applications.

[3] The development of accurate and efficient deep learning models for real-time object detection and tracking is of great importance in a wide range of computer vision applications. In this context, a deep-learning-based system integrating YOLOv11 for object detection with DeepSORT for multi-object tracking has been proposed. The system is designed to improve the performance of existing detection and tracking approaches by enhancing detection accuracy, preserving object identities, and maintaining real-time processing speed while reducing computational overhead. YOLOv11 serves as the detection backbone and is optimized for fast single-stage

inference with improved feature extraction capability. The model incorporates architectural enhancements such as C3k2 blocks, SPPF layers, and spatial attention mechanisms to generate more discriminative feature representations. DeepSORT is employed to track detected objects across video frames by combining Kalman Filter-based motion estimation with deep appearance feature embeddings. The motion model predicts object trajectories, while the appearance model improves data association under occlusion and crowding. The proposed system demonstrates robust detection and tracking performance in dynamic environments. Experimental evaluation shows that the system achieves stable real-time performance while maintaining high tracking accuracy and reduced identity switches. The improved performance is attributed to the effective combination of YOLOv11's advanced detection architecture and DeepSORT's appearance-aware tracking strategy. The detection model extracts meaningful spatial features from video frames, while the tracking module refines object associations over time. This integrated approach enables accurate multi-object tracking even when objects undergo partial occlusion or rapid motion. The proposed system has significant implications for intelligent vision applications including surveillance, traffic monitoring, and autonomous systems. Its high accuracy, robustness, and computational efficiency make it suitable for deployment on mid-range hardware platforms. The ability to achieve reliable performance without excessive computational cost also reduces system complexity and deployment overhead in real-world applications.

[4] The proposed object detection and tracking system is a deep learning–based computer vision framework that integrates a YOLOv11 detection model with the DeepSORT multi-object tracking algorithm. The system is designed to improve real-time visual perception by detecting and tracking multiple objects in continuous video streams. YOLOv11 enables accurate and fast object detection, while DeepSORT maintains object identities across frames using motion prediction and appearance-based feature embeddings. The system aims to enhance situational awareness by providing reliable detection and tracking results in real time. Experimental evaluation demonstrates that the system achieves high detection accuracy and stable tracking performance while operating at real-time speeds. The system's modular architecture allows it to be efficiently deployed on standard computing platforms and easily integrated into existing intelligent vision infrastructures. The proposed system has potential applications in surveillance, traffic monitoring, intelligent transportation systems, and autonomous platforms. By combining high accuracy, real-time performance, and deployment flexibility, the proposed object detection and tracking system represents a meaningful contribution to intelligent vision-based systems with promising implications for real-world applications.

[5] This paper proposes a methodology to develop a reliable and computationally inexpensive real-time automatic accident detection system with minimal hardware requirements. The proposed detection stage uses Mini-YOLO, a deep learning model architecture trained using knowledge distillation, with reduced model size and computational overhead compared to its counterpart, YOLO (You-Only-Look-Once). Mini-YOLO achieves an average precision (AP) score of 34.2 on the MS-COCO dataset, outperforming other detection algorithms in runtime complexity. The proposed system achieves a staggering 28 frames per second on a low-end machine, making it an efficient solution for real-time accident detection. The use of knowledge distillation enables the transfer of knowledge from a larger network to a smaller network, reducing computational overhead without sacrificing accuracy. The proposed system's efficiency and reliability make it a promising solution for real-time accident detection with minimal hardware requirements.

[6] This paper presents a deep learning–based methodology for real-time object detection and tracking that emphasizes reliability and computational efficiency for practical deployment. The proposed system integrates YOLOv11, an advanced single-stage object detection architecture, with the DeepSORT multi-object tracking algorithm to achieve accurate detection and stable tracking in dynamic environments. YOLOv11 is designed to provide high detection accuracy while maintaining low inference latency through optimized architectural components such as C3k2 blocks and SPPF layers. DeepSORT complements the detection stage by incorporating Kalman Filter–based motion prediction and appearance feature embeddings, enabling consistent object identity preservation across frames. The combined framework is capable of operating in real time on mid-range hardware, making it suitable for continuous video analysis. By jointly addressing detection accuracy and tracking robustness, the proposed system effectively handles challenges such as occlusion, rapid object motion, and illumination variation. The integration of efficient detection and appearance-aware tracking reduces computational overhead while maintaining reliable performance. The system's design demonstrates that high-speed object detection and tracking can be achieved without excessive hardware requirements. This approach provides a practical and scalable solution for real-time visual perception tasks. The efficiency and robustness of the proposed framework make it well suited for applications such as surveillance, traffic monitoring, and intelligent transportation systems, where real-time processing and accuracy are critical.

[7] A real-time human tracking system known as Pfinder was introduced to detect and track human body parts such as the head and hands using statistical models of color and shape. The system operates at interactive speeds on general-purpose hardware and demonstrates reliable performance across different environments and lighting conditions. Pfinder employs probabilistic modeling and Kalman filtering to maintain stable tracking while handling partial occlusion and motion variability. The approach avoids complex feature representations and instead relies on region-based blob tracking, which reduces computational complexity. Due to its efficiency and robustness, the system has been successfully applied to gesture recognition, human–computer interaction, and video surveillance applications.

[8] A simple yet highly efficient real-time 3D multi-object tracking system is presented to balance accuracy, computational efficiency, and system simplicity. The proposed

approach obtains 3D object detections from LiDAR point clouds and performs tracking using a combination of a 3D Kalman filter and the Hungarian algorithm for state estimation and data association. Unlike many complex state-of-the-art methods, the system relies on classical tracking components while extending the object state space to include 3D position, size, velocity, and orientation. Despite the simplicity of the tracking pipeline, the method achieves state-of-the-art performance on standard 3D multi-object tracking benchmarks such as KITTI and nuScenes. The system operates at real-time speeds exceeding 200 frames per second, making it one of the fastest reported 3D tracking solutions. In addition to the tracking framework, limitations in existing evaluation practices are addressed through the introduction of a standardized 3D multi-object tracking evaluation tool along with new integral performance metrics. These metrics enable fair and comprehensive comparison of 3D tracking systems across different operating thresholds. The results demonstrate that efficient system design and high-quality detections play a critical role in achieving reliable tracking performance. Overall, the work establishes a strong baseline for future research in both 2D and 3D multi-object tracking domains.

[9] A view-based approach for human action recognition was introduced using temporal templates constructed from motion information in video sequences. The method represents actions through motion-energy images and motion-history images, which jointly encode where motion occurs and how it evolves over time. These templates are matched against stored models of known actions, enabling direct recognition of motion patterns without requiring explicit three-dimensional reconstruction of the human body. The approach supports automatic temporal segmentation and exhibits invariance to linear variations in action speed. Real-time performance is achieved on standard computing platforms due to the simplicity of the representation and matching process. Experimental results on multiple human actions demonstrate that motion-based representations can effectively discriminate between different activities. However, the method assumes reliable motion extraction and may struggle in scenarios involving complex backgrounds or multiple interacting subjects, highlighting limitations when applied to highly cluttered real-world environments.

[10] An enhanced front-vehicle detection approach was presented for large vehicle fleet management using a single front-mounted camera and deep learning techniques. The study employed YOLOv4 supplemented with a fence-based geometric constraint to improve detection accuracy and reduce false detections under varying lighting and road conditions. By avoiding hardware upgrades and focusing on software optimization, the method achieved high accuracy and stable real-time performance. The results demonstrated that single-stage deep learning detectors can effectively support collision avoidance and driving safety in resource-constrained fleet environments.

[11] An object detection and tracking framework for video surveillance applications was presented using deep learning and artificial intelligence techniques. The approach employed CNN-based models for single-object detection and YOLOv3 for multi-object detection on urban vehicle, KITTI, and COCO datasets. Detected objects were tracked across video frames using the Simple Online Real-Time Tracking (SORT) algorithm, enabling efficient multi-object tracking in traffic surveillance scenarios. Experimental results demonstrated real-time performance with high accuracy, precision, and mean average precision under varying illumination and traffic conditions, validating the suitability of YOLO-based detection combined with SORT for intelligent surveillance systems.

[12] A comprehensive survey of deep learning–based visual tracking methods was presented, covering the progression from traditional hand-crafted feature trackers to modern end-to-end deep learning approaches. The study categorizes tracking algorithms based on network architecture, learning strategy, and tracking paradigm, including discriminative, generative, and hybrid models. Performance comparisons across widely used benchmarks highlight how deep trackers handle challenges such as occlusion, scale variation, fast motion, and background clutter. The survey also discusses training strategies, dataset bias, and evaluation metrics that influence tracker performance. By identifying limitations related to computational complexity and generalization, the work outlines future research directions for developing more robust and efficient visual tracking systems suitable for real-world applications.

III. METHODOLOGY

Visual Data Acquisition: A live camera feed or pre-recorded video is used as input to capture continuous frames containing multiple dynamic objects such as pedestrians and vehicles. These frames form the raw visual data for detection and tracking.

Processing Stage: Each video frame is resized and normalized before being processed by the YOLOv11 object detection model. YOLOv11 detects objects in real time by predicting bounding boxes, class labels, and confidence scores. The detection results are passed to the DeepSORT tracking module for further analysis.

Feature Extraction and Tracking: YOLOv11 provides object localization, while DeepSORT extracts appearance features using a deep neural network. Object motion is predicted using a Kalman Filter, and detections are associated with existing tracks using the Hungarian Algorithm. This ensures consistent object tracking under occlusion and motion variations.

Application Interface and Output: The final output is displayed through a visual interface showing bounding boxes, class labels, and unique tracking IDs. Performance metrics such as FPS and object count are also overlaid, supporting real-time monitoring and analysis.

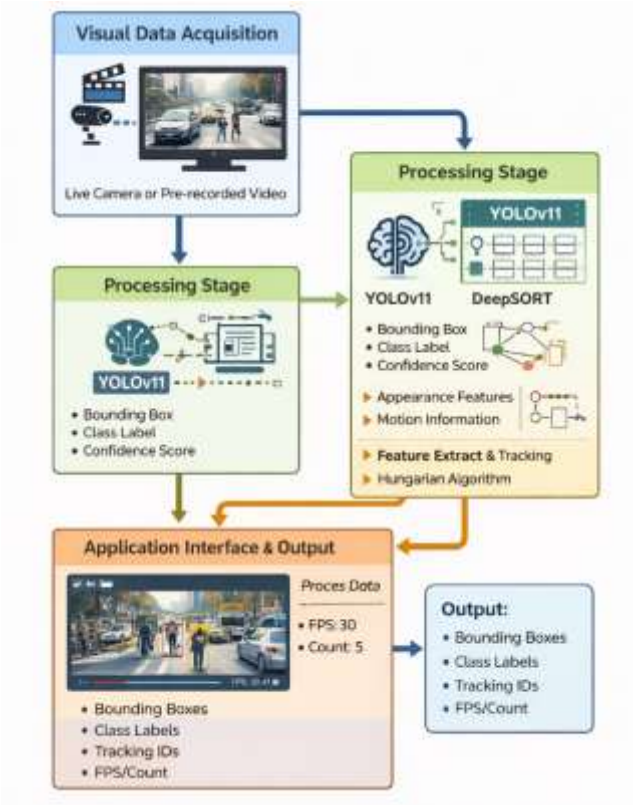

Fig. 1. Object Detection and Tracking System Workflow

The Detection, Tracking, and Visualization Modules work together seamlessly to enable accurate real-time object detection and multi-object tracking in dynamic environments. The Visualization Module serves as the primary interface of the system, displaying real-time video output with annotated bounding boxes, object class labels, confidence scores, and unique tracking IDs. It provides an intuitive view of how objects are detected and tracked across frames, along with performance indicators such as frame rate and object count.

The Detection Module, powered by YOLOv11, acts as the perceptual component of the system. It processes incoming video frames and identifies multiple objects simultaneously with high accuracy and low latency. By extracting spatial features and predicting object locations and classes in a single forward pass, the detection module enables fast and reliable perception of the scene.

At the core of the system lies the Tracking and Processing Module, implemented using DeepSORT. This module integrates motion prediction through Kalman filtering with appearance-based feature extraction to maintain consistent object identities over time. It associates detections across frames using optimal assignment strategies and ensures stable tracking even under occlusion, overlapping objects, and rapid movement. From a system perspective, the processing module fuses detection outputs into meaningful temporal information, transforming raw visual data into structured tracking results suitable for real-time analysis and intelligent video applications.

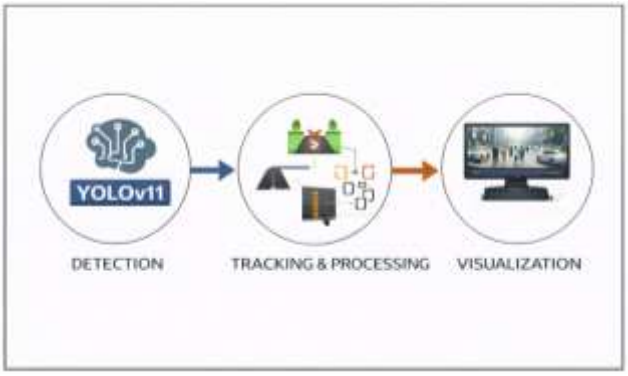

Fig 2. Architecture of the YOLOv11 and DeepSORT-Based Object Detection and Tracking System

IV. Proposed Model

The proposed model focuses on real-time object detection and multi-object tracking using deep learning and computer vision techniques. The system is designed to process video data efficiently, detect multiple objects accurately, and maintain their identities across consecutive frames. The model integrates YOLOv11 for object detection and DeepSORT for object tracking, forming a unified detection–tracking framework suitable for dynamic environments.

The process begins with video data acquisition, where live camera feeds or pre-recorded video sequences are used as input. These video streams consist of continuous frames containing multiple objects such as pedestrians and vehicles. The collected video frames are prepared for processing through resizing and normalization to ensure compatibility with the deep learning models.

Once the data is prepared, object detection is performed using the YOLOv11 model. YOLOv11 detects objects in each frame by predicting bounding boxes, class labels, and confidence scores in a single forward pass. The detected objects are filtered using Non-Maximum Suppression to remove redundant detections and retain the most accurate bounding boxes. This stage provides precise spatial information required for tracking.

The detected objects are then passed to the DeepSORT tracking module. DeepSORT combines motion and appearance information to track objects over time. A Kalman Filter is used to predict object positions in subsequent frames, while a deep neural network extracts appearance features for each object. The Hungarian Algorithm is applied to associate new detections with existing tracks based on motion consistency and appearance similarity. This enables reliable identity preservation even during occlusion and overlapping scenarios.

Finally, the output of the system is visualized through an interface that displays bounding boxes, object class labels, confidence values, and unique tracking IDs. Additional performance metrics such as frame rate and object count are also displayed. The proposed model ensures accurate, efficient, and real-time object detection and tracking, making it suitable for applications such as surveillance, traffic monitoring, and intelligent video analysis.

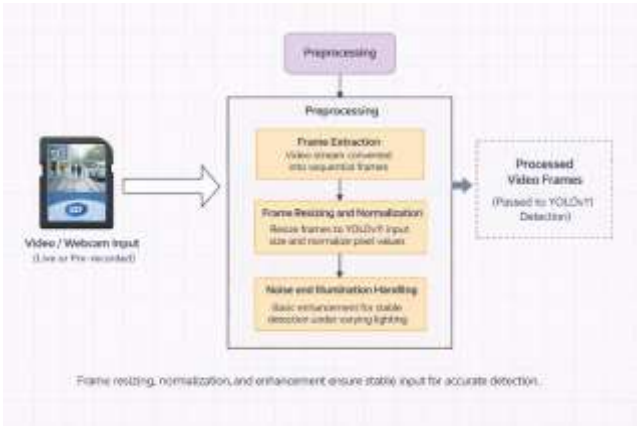

Fig. 3. Preprocessing Pipeline for Video-Based Object Detection and Tracking

V. RESULTS

The primary outcome of the proposed YOLOv11 and DeepSORT-based system is its ability to perform accurate real-time object detection and consistent multi-object tracking in dynamic video environments. The system successfully detects multiple object classes such as pedestrians and vehicles and assigns unique tracking IDs to each object, maintaining identity continuity across consecutive frames. YOLOv11 demonstrates reliable detection performance by accurately localizing objects with bounding boxes and class labels, even in scenes containing motion and moderate occlusion.

The integration of DeepSORT significantly enhances tracking performance by preserving object identities over time. By combining Kalman Filter-based motion prediction with appearance feature extraction, the system effectively reduces identity switches and maintains stable tracking when objects overlap or temporarily leave the frame. Experimental results show that tracked objects retain consistent IDs throughout their movement, validating the robustness of the tracking module.

The system operates at real-time speeds on mid-range hardware, achieving smooth video processing with a stable frame rate. Performance indicators such as frames per second (FPS) and object count are displayed alongside detection results, demonstrating the system's efficiency and responsiveness. The visual output confirms accurate synchronization between detection and tracking, with bounding boxes, class labels, and tracking IDs correctly updated frame by frame.

Overall, the results indicate that the proposed system achieves a strong balance between detection accuracy, tracking stability, and computational efficiency. The YOLOv11 and DeepSORT integration proves effective for real-time video-based object detection and tracking, making the system suitable for applications such as surveillance, traffic monitoring, and intelligent video analytics.

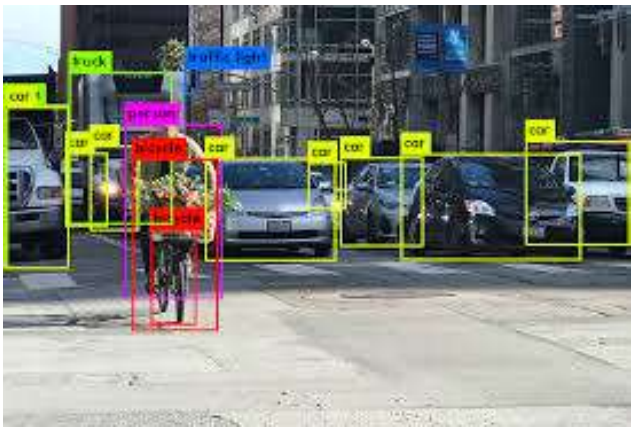

Fig. 4 Expected Result

Fig 5. Result of Detecting Vehicle on the Road by Web Cam

Fig. 6 Multiple Object Output of YOLOv11 and DeepSORT on Live Webcam Feed

Fig. 7 Sample Detection and Tracking Output on Video Sequence

VI. CONCLUSION

In recent years, the rapid growth of camera-based systems and intelligent visual analytics has increased the demand for accurate and efficient object detection and tracking solutions. Many real-world applications such as video surveillance, traffic monitoring, and smart city systems require continuous monitoring of dynamic environments. Traditional computer vision techniques often struggle to maintain accuracy and real-time performance under challenges such as occlusion, varying illumination, and complex backgrounds.
To address these limitations, deep learning–based object detection and tracking methods have gained significant attention. One of the most effective approaches in this domain is the integration of real-time object detectors with robust multi-object tracking algorithms. YOLOv11 is a state-of-the-art single-stage object detection model capable of detecting multiple objects simultaneously with high speed and accuracy. By processing video frames in a single forward pass, YOLOv11 enables real-time detection of objects such as pedestrians and vehicles. However, object detection alone is insufficient for understanding object movement and behavior over time.
To overcome this limitation, the proposed system integrates YOLOv11 with DeepSORT, a multi-object tracking algorithm that combines motion estimation and appearance-based feature extraction. DeepSORT uses a Kalman Filter to predict object trajectories and a deep neural network to extract appearance embeddings, allowing consistent object identity assignment across consecutive frames. This integration enables reliable tracking even in scenarios involving occlusion and overlapping objects.
The proposed system processes video input from live webcams or pre-recorded sequences, detects objects using YOLOv11, and tracks them using DeepSORT in real time. The system demonstrates stable performance across different environments while maintaining low computational overhead. Despite challenges such as dataset diversity and complex scene dynamics, the YOLOv11 and DeepSORT-based framework provides an effective solution for real-time object detection and tracking. Overall, this approach offers a scalable and practical foundation for intelligent video analysis applications, contributing to advancements in computer vision–based monitoring systems.

REFERENCES

[1] F. Porikli and A. Yilmaz, “Object Detection and Tracking,” in Handbook of Pattern Recognition and Computer Vision, 4th ed., 2012.
[2] R. Khanam and M. Hussain, “YOLOv11: An Overview of the Key Architectural Enhancements,” University of Huddersfield, Oct. 24, 2024.
[3] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, “Simple Online and Realtime Tracking (SORT),” 2016.
[4] Y. Sun, Z. Sun, and W. Chen, “The evolution of object detection methods,” Engineering Applications of Artificial Intelligence, vol. 133, p. 108458, Apr. 2024.
[5] D. M. Gavrila, “Pedestrian detection: A review of the state of the art,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 1, pp. 1–14, 2000.
[6] Y. Ming and Y. Zhang, “ADT: Object tracking algorithm based on adaptive detection,” IEEE Access, vol. 8, pp. 169407–169416, 2020, doi: 10.1109/ACCESS.2020.3023872.
[7] C. Wren, A. Azarbayejani, T. Darrell, and A. Pentland, “Pfinder: Real-time tracking of the human body,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pp. 780–785, 1997.
[8] X. Weng, J. Wang, D. Held, and K. M. Kitani, “3D multi-object tracking: A baseline and new evaluation metrics,” in Proc. European Conference on Computer Vision (ECCV), 2020, pp. 210–227, doi: 10.1007/978-3-030-58571-6_13.
[9] J. Davis and A. Bobick, “The representation and recognition of action using temporal templates,” in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1997.
[10] C.-Y. Mu, P. Kung, C.-F. Chen, and S.-C. Chuang, “Enhancing front-vehicle detection in large vehicle fleet management,” Remote Sensing, vol. 14, no. 7, p. 1544, Mar. 2022, doi: 10.3390/rs14071544.
[11] Mohana and H. V. R. Aradhya, “Object detection and tracking using deep learning and artificial intelligence for video surveillance applications,” International Journal of Advanced Computer Science and Applications (IJACSA), vol. 10, no. 12, pp. 517–530, 2019, doi: 10.14569/IJACSA.2019.0101269.
[12] S. M. Marvasti-Zadeh, L. Cheng, H. Ghanei-Yakhdan, and S. Kasaei, “Deep learning for visual tracking: A comprehensive survey,” arXiv preprint, arXiv:1912.00535, Jan. 2021.