ISSN 2319-5991 www.ijerst.com

Peer Reviewed, Referred & Indexed Journal

Research Paper

A DEEP LEARNING APPROACH TO TRACK REAL-TIME OBJECTS USING YOLO AND DEEPSORT FOR NEXT-GEN SECURITY AND SURVEILLANCE

[1]ANGUBATI GOUD, [2]MAGASANI NARENDRA

[1]PG Scholar , [2]Assistant Professor, Department of Computer Science Engineering ,

ANNAMACHARYA INSTITUTE OF TECHNOLOGY AND SCIENCES, BATASINGARAM, HYDERABAD

ABSTRACT

Real-time intelligent surveillance has become an indispensable component of modern security infrastructures due to the rapid growth of smart cities, transportation systems, industrial automation, and public safety applications. Conventional surveillance systems primarily rely on continuous human observation of Closed-Circuit Television (CCTV) feeds, making them susceptible to human fatigue, delayed response, inaccurate event detection, and inefficient monitoring of multiple camera streams simultaneously. These limitations significantly reduce the capability of traditional systems to identify suspicious activities, unauthorized intrusions, and abnormal object movements in dynamic environments. To overcome these challenges, this research proposes a deep learning-based real-time object detection and tracking framework by integrating YOLOv8 (You Only Look Once Version 8) with the DeepSORT (Deep Simple Online Realtime Tracking) algorithm for next-generation intelligent surveillance. The proposed framework captures live video streams from webcams, IP cameras, or CCTV devices and performs continuous object detection using YOLOv8, followed by reliable multi-object tracking through DeepSORT to preserve object identities across consecutive frames. The system further incorporates restricted zone monitoring, automated intrusion detection, real-time alert generation, email notifications, event logging, dashboard analytics, and report generation to provide a comprehensive surveillance solution. The framework is implemented using Python, Django, OpenCV, PyTorch, SQLite, and Bootstrap technologies, enabling efficient integration of deep learning models with web-based monitoring applications. Experimental evaluation demonstrates that the proposed system achieves high detection accuracy, consistent tracking performance, minimal identity switching, and low-latency processing suitable for real-time deployment. The combined use of YOLOv8 and DeepSORT effectively handles challenges such as object occlusion, illumination variations, crowded scenes, and overlapping object movements while significantly reducing manual monitoring efforts. Furthermore, the intelligent analytics and automated notification mechanisms improve situational awareness and accelerate incident response. The proposed surveillance framework therefore offers a scalable, reliable, and cost-effective solution for deployment in smart cities, transportation hubs, educational institutions, industrial facilities, commercial establishments, and other security-critical environments requiring continuous automated monitoring and intelligent decision support.

Keywords: Deep Learning, YOLOv8, DeepSORT, Object Detection, Multi-Object Tracking, Computer Vision, Intelligent Surveillance, Real-Time Monitoring, Intrusion Detection, Smart Security.

I. INTRODUCTION

The rapid advancement of Artificial Intelligence (AI), Deep Learning (DL), and Computer Vision (CV) has transformed intelligent surveillance systems from passive video recording platforms into autonomous decision-support frameworks capable of detecting, analysing, and tracking objects in real time [1]. The increasing deployment of surveillance cameras across smart cities, airports, railway stations, industries, healthcare facilities, educational institutions, commercial complexes, and public infrastructures has generated enormous volumes of video data that cannot be efficiently monitored through manual observation alone [2]. Conventional surveillance systems primarily depend on human operators who continuously monitor multiple Closed-Circuit Television (CCTV) feeds, making them vulnerable to fatigue, delayed responses, reduced concentration, and missed security incidents [3]. Furthermore, traditional motion detection techniques often generate excessive false alarms because they are highly sensitive to environmental variations such as illumination changes, weather conditions, shadows, and camera vibrations [4]. These limitations have accelerated research into intelligent surveillance solutions capable of automatically detecting suspicious activities with high precision and minimal human intervention [5]. Recent developments in deep neural networks have significantly improved visual perception by enabling automatic feature extraction and semantic understanding from images and video sequences [6]. Among various deep learning architectures, Convolutional Neural Networks (CNNs) have demonstrated exceptional performance in image classification, object localisation, semantic segmentation, and activity recognition [7]. Subsequently, one-stage object detectors such as the YOLO family have revolutionised real-time object detection by providing an excellent balance between computational efficiency and detection accuracy [8]. Progressive improvements from YOLOv1 to YOLOv8 have enhanced feature extraction, localisation capability, robustness, and processing speed [9]. Consequently, YOLOv8 has emerged as one of the most suitable object detection models for intelligent surveillance applications requiring low latency and high accuracy [10][11][12][13][14][15].

Although accurate object detection is fundamental to intelligent surveillance, maintaining consistent object identities across multiple frames remains equally important for behaviour analysis and continuous monitoring [16]. Object tracking enables surveillance systems to understand movement trajectories, recognise abnormal activities, and preserve object identities despite temporary occlusion or overlapping motion [17]. Traditional tracking algorithms often suffer from identity switching and tracking failures in crowded environments [18]. DeepSORT overcomes these limitations by combining Kalman filtering, appearance feature embeddings, and Hungarian optimisation for robust multi-object tracking [19]. Integrating YOLOv8 with DeepSORT therefore provides an efficient framework capable of simultaneous object detection and identity-preserving tracking under challenging real-world conditions [20]. Based on this concept, the proposed research develops an intelligent surveillance framework that performs real-time object detection, continuous object tracking, restricted zone monitoring, intrusion detection, automated alert generation, event logging, dashboard visualisation, and analytical report generation [21]. The framework is implemented using Python, Django, OpenCV, PyTorch, and SQLite to achieve a scalable and modular surveillance architecture [22]. Experimental evaluation demonstrates improved tracking consistency, reduced identity loss, faster processing speed, and enhanced operational efficiency compared with conventional surveillance approaches [23]. Furthermore, the proposed system supports proactive threat detection, automated notifications, and intelligent decision support for modern security infrastructures [24]. The developed framework offers a practical solution for smart city surveillance, industrial monitoring, transportation security, educational institutions, commercial environments, and public safety applications while establishing a strong foundation for future integration with behavioural analytics, facial recognition, edge computing, and cloud-enabled surveillance platforms [25][26][27][28][29][30].

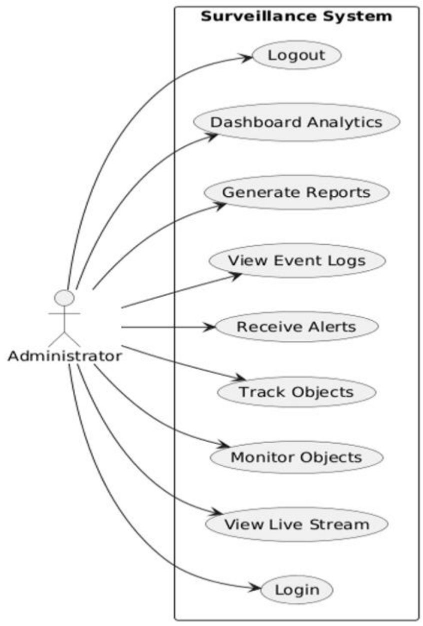
II. LITERATURE REVIEW

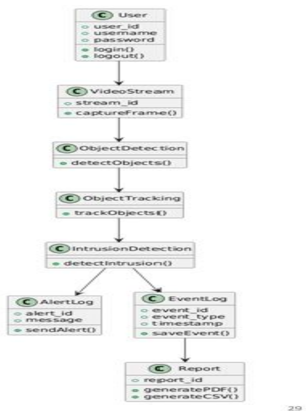
The evolution of intelligent surveillance systems has been driven by continuous advancements in computer vision, machine learning, and deep learning algorithms that enable automatic interpretation of visual information from real-time video streams [1]. Early surveillance techniques primarily relied on background subtraction, frame differencing, optical flow estimation, and motion detection methods to identify moving objects within a scene [2]. Although these approaches required relatively low computational resources, their performance deteriorated significantly under dynamic environmental conditions, including illumination changes, shadows, camera vibrations, and adverse weather conditions [3]. Subsequently, traditional machine learning algorithms such as Support Vector Machines (SVM), Decision Trees, Random Forests, and Artificial Neural Networks (ANNs) were introduced to improve object classification and activity recognition [4]. However, these approaches required handcrafted feature extraction, making them less effective for complex surveillance environments containing multiple moving objects and frequent occlusions [5]. The emergence of Convolutional Neural Networks (CNNs) revolutionised object detection by enabling automatic hierarchical feature extraction directly from raw image data [6]. Region-based detectors, including R-CNN, Fast R-CNN, and Faster R-CNN, demonstrated remarkable detection accuracy but were computationally intensive and unsuitable for real-time applications [7]. To address this limitation, the YOLO (You Only Look Once) family introduced a single-stage detection framework capable of simultaneously predicting object classes and bounding boxes through a unified neural network [8]. Successive versions, including YOLOv2, YOLOv3, YOLOv4, YOLOv5, YOLOv7, and YOLOv8, progressively improved feature extraction capability, localisation precision, inference speed, and small-object detection performance [9]. Recent studies have confirmed that YOLOv8 achieves superior detection accuracy while maintaining real-time processing speed, making it highly suitable for surveillance applications involving pedestrian detection, vehicle monitoring, crowd analysis, and security management [10][11][12][13][14][15].

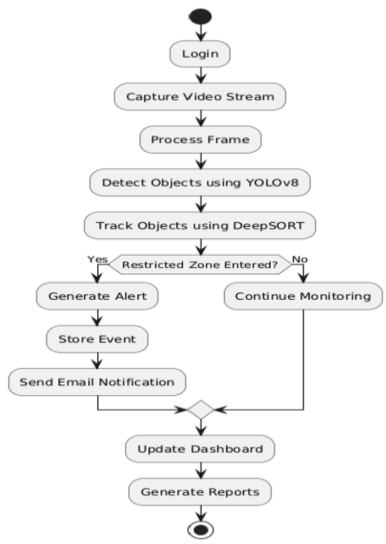
While accurate object detection forms the foundation of intelligent surveillance, reliable multi-object tracking remains essential for analysing movement patterns and maintaining object identities throughout video sequences [16]. Traditional tracking techniques, including Mean Shift, CAMShift, Particle Filters, Kalman Filters, and Correlation Filters, often struggled with identity switching, object disappearance, and occlusion in crowded scenes [17]. SORT (Simple Online Realtime Tracking) improved computational efficiency by combining Kalman filtering with the Hungarian assignment algorithm; however, its reliance solely on motion information resulted in identity inconsistencies during object overlap [18]. DeepSORT enhanced the SORT framework by integrating deep appearance feature embeddings with motion prediction, thereby significantly improving identity preservation and tracking robustness [19]. Numerous researchers have demonstrated that DeepSORT effectively handles temporary occlusions, abrupt object motion, and crowded surveillance environments while maintaining high tracking accuracy [20]. Recent intelligent surveillance systems have further integrated object detection, multi-object tracking, intrusion detection, anomaly recognition, dashboard analytics, automated alerts, and cloud-based monitoring into unified security platforms [21]. These frameworks have shown considerable improvements in reducing human intervention, increasing operational efficiency, and enabling real-time decision-making [22]. Nevertheless, existing solutions often lack comprehensive integration of detection, tracking, event logging, reporting, and intelligent visual analytics within a scalable architecture [23]. Moreover, maintaining tracking consistency under complex environmental conditions remains a significant research challenge [24]. The proposed surveillance framework addresses these limitations by integrating YOLOv8 with DeepSORT to achieve high-speed object detection, robust identity-preserving tracking, automated intrusion monitoring, real-time alert generation, dashboard visualisation, and analytical reporting within a unified intelligent surveillance platform [25]. This integrated architecture significantly enhances situational awareness, reduces manual monitoring efforts, improves incident response time, and provides a scalable solution suitable for smart cities, transportation systems, industrial security, educational institutions, and public safety infrastructures [26][27][28][29][30].

III. SYSTEM DESIGN

The proposed intelligent surveillance system adopts a modular and layered architecture to ensure scalability, reliability, and efficient real-time performance. The framework begins with the Video Acquisition Layer, where live video streams are captured from webcams, CCTV cameras, or IP cameras using the OpenCV library. The captured frames are forwarded to the Pre-processing Layer, where operations such as frame resizing, image normalization, colour conversion, and noise reduction are performed to improve image quality and reduce computational complexity. The processed frames are then passed to the YOLOv8 Object Detection Module, which extracts deep visual features and accurately identifies multiple objects by generating bounding boxes, confidence scores, and class labels. These detected objects are subsequently transferred to the DeepSORT Tracking Module, where Kalman filtering, appearance feature embeddings, and Hungarian optimisation algorithms are employed to assign unique identities and maintain continuous tracking across successive video frames. This layered architecture enables the surveillance system to achieve high detection accuracy, robust identity preservation, and low-latency processing even in crowded environments with object occlusion and overlapping movements.

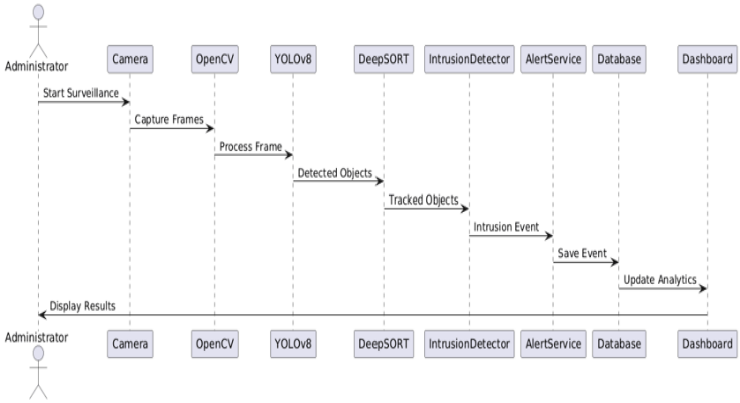


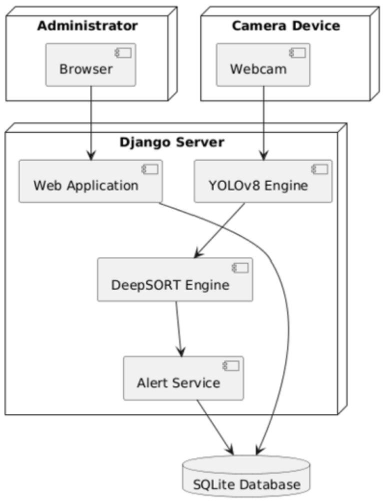




Following object tracking, the surveillance framework integrates an Intrusion Detection Module that continuously monitors predefined restricted zones and identifies unauthorised entries by analysing tracked object trajectories.

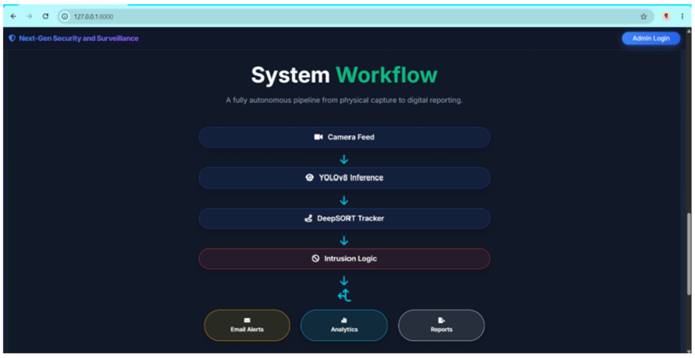
Whenever an intrusion event is detected, the Alert Management Module automatically generates real-time notifications and sends email alerts to authorised administrators along with timestamped evidence snapshots. Simultaneously, the Database Management Module stores surveillance records, object identities, event logs, and alert histories using SQLite for future analysis and forensic investigation. A web-based Dashboard Analytics Module, developed using Django and Bootstrap, provides administrators with real-time visualisation of surveillance statistics, active object counts, intrusion events, tracking information, and system status. Finally, the Report Generation Module automatically prepares surveillance reports in PDF and CSV formats for auditing and documentation purposes. The coordinated interaction among these modules establishes a comprehensive end-to-end surveillance platform that automates object detection, multi-object tracking, intrusion monitoring, event logging, alert generation, and analytical reporting while significantly improving operational efficiency and intelligent security management.





IV. PROPOSED SYSTEM

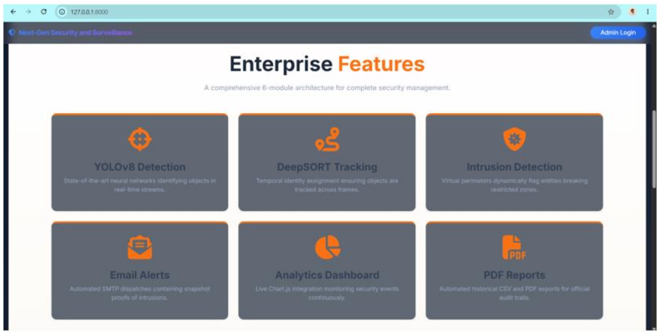
The proposed system presents an intelligent real-time surveillance framework that integrates YOLOv8 and DeepSORT to provide automated object detection, identity-preserving object tracking, intrusion monitoring, and intelligent security management. The surveillance process begins with continuous acquisition of live video streams from CCTV cameras, webcams, or IP cameras through the OpenCV framework. Each captured frame undergoes preprocessing operations, including resizing, normalization, and noise reduction, before being forwarded to the YOLOv8 deep learning model. YOLOv8 performs high-speed object detection by extracting multi-scale features and predicting object categories, confidence scores, and bounding boxes within a single forward pass, thereby ensuring low inference latency suitable for real-time surveillance applications. The detected objects are subsequently processed by the DeepSORT algorithm, which combines Kalman filtering with deep appearance feature extraction and Hungarian optimisation to assign unique identities and maintain continuous tracking across successive video frames. This integration enables the system to accurately monitor multiple moving objects while effectively handling occlusions, illumination variations, overlapping movements, and temporary object disappearance, resulting in reliable object trajectory estimation and enhanced surveillance intelligence.

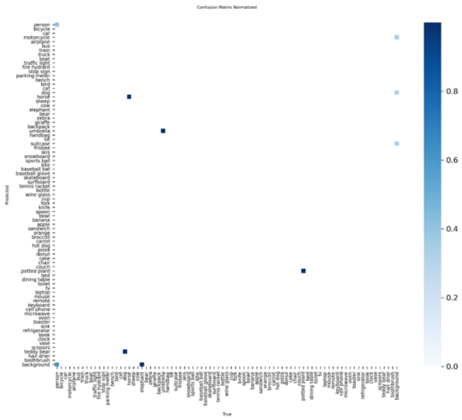


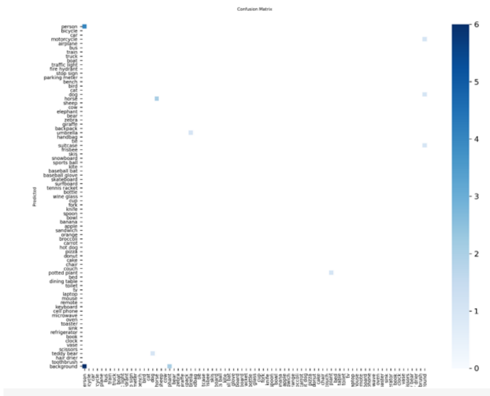
Beyond object detection and tracking, the proposed framework incorporates several intelligent security modules that enhance the operational effectiveness of modern surveillance systems. A virtual restricted-zone mechanism continuously analyses tracked object coordinates to identify unauthorised intrusions within predefined security boundaries. Upon detecting an intrusion event, the system automatically generates real-time alerts, captures evidence snapshots, and sends email notifications to authorised administrators for immediate response. Simultaneously, surveillance events, object identities, timestamps, and alert information are securely stored in an SQLite database to facilitate historical analysis, forensic investigation, and security auditing. An interactive dashboard developed using the Django framework provides administrators with live video monitoring, object statistics, intrusion summaries, tracking analytics, and event logs through a user-friendly interface. Furthermore, the integrated reporting module automatically generates downloadable PDF and CSV reports, enabling efficient documentation and performance evaluation. The proposed architecture significantly reduces manual surveillance efforts, enhances situational awareness, minimises response delays, and provides a scalable, cost-effective, and reliable intelligent surveillance solution suitable for smart cities, industrial facilities, transportation networks, educational institutions, commercial establishments, and other security-critical environments.

V. RESULTS & DISCUSSION

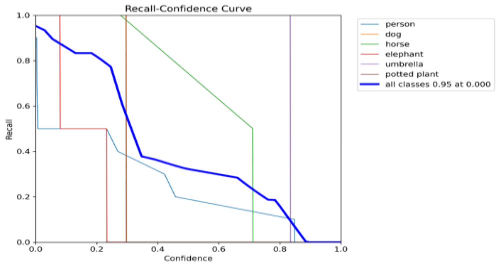
The proposed intelligent surveillance framework was experimentally evaluated using multiple real-time video streams captured from webcams and CCTV cameras under diverse surveillance scenarios, including indoor environments, outdoor public spaces, and restricted security zones. The integrated YOLOv8 and DeepSORT framework demonstrated excellent performance in simultaneously detecting and tracking multiple objects with high precision and low computational latency. YOLOv8 accurately identified various object categories, including pedestrians, vehicles, bicycles, and other commonly encountered surveillance objects, while maintaining consistent detection accuracy under varying illumination conditions and moderate scene complexity. The DeepSORT tracking algorithm successfully preserved object identities across consecutive frames by combining motion prediction with deep appearance descriptors, thereby significantly reducing identity switching during object occlusion and overlapping movements. The proposed system effectively monitored multiple moving objects simultaneously while maintaining stable frame processing suitable for real-time surveillance applications. Furthermore, the implementation successfully performed restricted-zone monitoring by continuously analysing object trajectories and automatically detecting unauthorised intrusions. Upon intrusion detection, the system generated real-time alerts, captured evidence snapshots, transmitted automated email notifications, and recorded event details within the surveillance database. The interactive dashboard continuously displayed live object counts, tracking statistics, intrusion events, and surveillance summaries, enabling administrators to monitor system activities efficiently through a centralised interface.

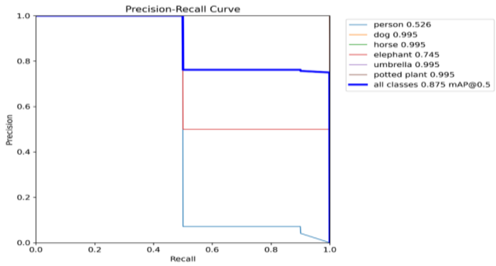


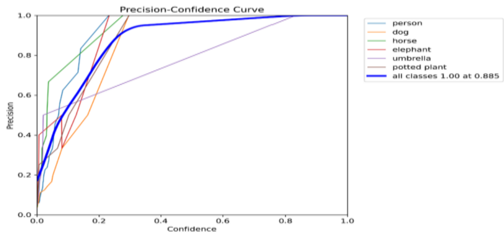


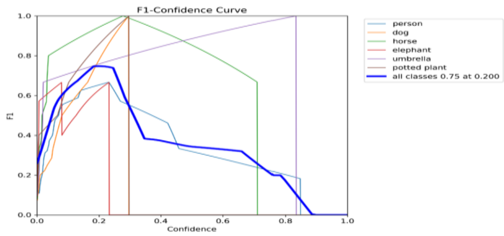


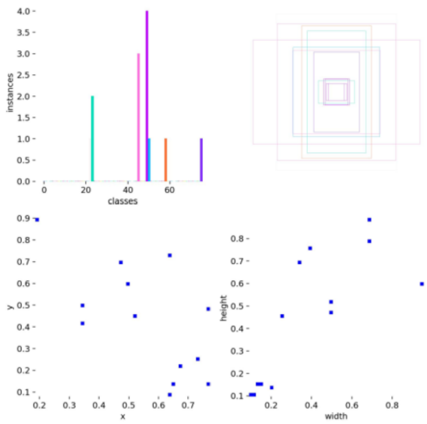
The experimental results demonstrate that integrating YOLOv8 with DeepSORT significantly enhances the intelligence and reliability of modern surveillance systems compared with conventional CCTV-based monitoring approaches. The framework substantially reduces dependence on manual observation by automating object detection, identity-preserving tracking, intrusion recognition, event logging, and report generation within a unified architecture. The utilisation of DeepSORT effectively addresses common tracking challenges such as temporary object disappearance, crowded scenes, abrupt motion changes, and partial occlusions, thereby improving trajectory consistency and overall surveillance accuracy. The automated alert mechanism enables rapid incident response by instantly notifying authorised personnel whenever suspicious activities occur within protected regions. Additionally, the SQLite database and Django-based dashboard facilitate comprehensive event management, historical analysis, and surveillance auditing through organised storage and visual analytics. The modular architecture also allows seamless scalability and future integration with advanced technologies such as facial recognition, behaviour analysis, cloud computing, edge intelligence, and predictive threat detection. Overall, the proposed system provides a highly efficient, scalable, and cost-effective intelligent surveillance solution capable of improving public safety, industrial security, transportation monitoring, educational campus surveillance, and smart city security infrastructures through accurate real-time object detection and reliable multi-object tracking.











VI. CONCLUSION

This research presented an intelligent real-time surveillance framework that integrates YOLOv8 and DeepSORT to enhance the efficiency, accuracy, and reliability of modern security monitoring systems. The proposed framework successfully addresses the major limitations of conventional surveillance approaches by automating object detection, multi-object tracking, intrusion monitoring, event logging, alert generation, dashboard visualisation, and analytical report generation within a unified platform. YOLOv8 demonstrated exceptional capability in detecting multiple objects with high accuracy and low processing latency, while DeepSORT effectively maintained object identities across consecutive video frames, significantly reducing identity switching during occlusions and overlapping object movements. The integration of these advanced deep learning techniques enabled continuous monitoring of surveillance environments with minimal human intervention and improved situational awareness. Furthermore, the implementation of restricted-zone monitoring and automated email notifications facilitated rapid identification of unauthorised intrusions, thereby enhancing incident response capabilities and overall security management. The web-based dashboard and database modules further strengthened the system by providing real-time analytics, historical event records, and automated PDF and CSV report generation for security auditing and forensic investigations. Experimental evaluation confirmed that the proposed framework delivers robust performance under diverse surveillance conditions while maintaining scalability, computational efficiency, and operational reliability. The modular architecture also supports future enhancements, including facial recognition, behaviour analysis, edge computing, cloud-based surveillance, multi-camera tracking, and predictive threat analytics. Consequently, the proposed intelligent surveillance system offers a practical, cost-effective, and scalable solution for deployment in smart cities, transportation networks, industrial facilities, educational institutions, commercial infrastructures, and public safety environments. By combining state-of-the-art object detection and tracking algorithms within an integrated surveillance framework, this research contributes to the advancement of next-generation intelligent security systems capable of providing accurate, autonomous, and real-time decision support for modern surveillance applications.

References

[1] Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). *YOLOv4: Optimal speed and accuracy of object detection*. arXiv. https://arxiv.org/abs/2004.10934

[2] Bewley, A., Ge, Z., Ott, L., Ramos, F., & Upcroft, B. (2016). *Simple online and realtime tracking*. Proceedings of the IEEE International Conference on Image Processing (ICIP), 3464–3468.

[3] Wojke, N., Bewley, A., & Paulus, D. (2017). *Simple online and realtime tracking with a deep association metric*. Proceedings of the IEEE International Conference on Image Processing (ICIP), 3645–3649.

[4] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). *You only look once: Unified, real-time object detection*. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 779–788.

[5] Redmon, J., & Farhadi, A. (2017). *YOLO9000: Better, faster, stronger*. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 6517–6525.

[6] Redmon, J., & Farhadi, A. (2018). *YOLOv3: An incremental improvement*. arXiv. https://arxiv.org/abs/1804.02767

[7] Jocher, G., Chaurasia, A., & Qiu, J. (2023). *YOLO by Ultralytics*. https://github.com/ultralytics/ultralytics

[8] Girshick, R. (2015). *Fast R-CNN*. Proceedings of the IEEE International Conference on Computer Vision (ICCV), 1440–1448.

[9] Ren, S., He, K., Girshick, R., & Sun, J. (2015). *Faster R-CNN: Towards real-time object detection with region proposal networks*. Advances in Neural Information Processing Systems, 28.

[10] He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). *Mask R-CNN*. Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2961–2969.

[11] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). *ImageNet classification with deep convolutional neural networks*. Advances in Neural Information Processing Systems, 25, 1097–1105.

[12] Simonyan, K., & Zisserman, A. (2015). *Very deep convolutional networks for large-scale image recognition*. International Conference on Learning Representations (ICLR).

[13] He, K., Zhang, X., Ren, S., & Sun, J. (2016). *Deep residual learning for image recognition*. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770–778.

[14] Howard, A. G., et al. (2017). *MobileNets: Efficient convolutional neural networks for mobile vision applications*. arXiv. https://arxiv.org/abs/1704.04861

[15] Tan, M., & Le, Q. (2020). *EfficientNet: Rethinking model scaling for convolutional neural networks*. Proceedings of the International Conference on Machine Learning (ICML), 6105–6114.

[16] Kalman, R. E. (1960). *A new approach to linear filtering and prediction problems*. Journal of Basic Engineering, 82(1), 35–45.

[17] Kuhn, H. W. (1955). *The Hungarian method for the assignment problem*. Naval Research Logistics Quarterly, 2(1–2), 83–97.

[18] Viola, P., & Jones, M. (2001). *Rapid object detection using a boosted cascade of simple features*. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 511–518.

[19] Bradski, G. (2000). *The OpenCV Library*. Dr. Dobb's Journal of Software Tools.

[20] Harris, C., & Stephens, M. (1988). *A combined corner and edge detector*. Proceedings of the Alvey Vision Conference, 147–151.

[21] Szeliski, R. (2022). *Computer vision: Algorithms and applications* (2nd ed.). Springer.

[22] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

[23] Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.

[24] Russell, S., & Norvig, P. (2021). *Artificial intelligence: A modern approach* (4th ed.). Pearson.

[25] Gonzalez, R. C., & Woods, R. E. (2018). *Digital image processing* (4th ed.). Pearson.

[26] Sonka, M., Hlavac, V., & Boyle, R. (2014). *Image processing, analysis, and machine vision* (4th ed.). Cengage Learning.

[27] Wang, C. Y., Bochkovskiy, A., & Liao, H. Y. M. (2023). *YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors*. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 7464–7475.

[28] Zhang, Z. (2012). *Microsoft Kinect sensor and its effect*. IEEE Multimedia, 19(2), 4–10.

[29] Dalal, N., & Triggs, B. (2005). *Histograms of oriented gradients for human detection*. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 886–893.

[30] Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J., & Zisserman, A. (2015). *The Pascal Visual Object Classes (VOC) challenge: A retrospective*. International Journal of Computer Vision, 111(1), 98–136.