

# Developing Surveillance System using Deep Sort and YOLOV8

Isha Mohite<sup>1</sup>, Sakshi More<sup>2</sup>, Yash Thakre<sup>3</sup>, Kalyani Pawar<sup>4</sup>, Prof. Priti Malkhede

<sup>1,2,3,4</sup>Department of Artificial Intelligence and Data Science

P.E.S. Modern College of Engineering, Pune, India

<sup>5</sup>Assistant Professor, Department of Artificial Intelligence and Data Science

P.E.S. Modern College of Engineering, Pune, India

**Abstract**—This review paper explores recent advancements in AI-based smart surveillance systems designed for real-time crowd monitoring, missing person detection, and abnormal behavior analysis [18], [20]. With increasing public safety concerns in crowded environments such as airports, railway stations, malls, and smart cities, traditional manual surveillance methods have become ineffective and error-prone [2]. Modern surveillance systems leverage deep learning algorithms such as YOLO for object detection, Deep SORT for multi-object tracking, and CNN-VideoMAE models for behavioral analysis to automatically monitor live video feeds and generate real-time alerts [7], [9], [15]. This paper reviews existing methodologies, models, system architectures, and datasets used in smart surveillance, highlighting their strengths and limitations [18]. Key challenges such as real-time processing, scalability, data privacy, and environmental variations are also discussed [17], [19], [20]. The study provides insights into future research directions focusing on accuracy improvement, system scalability, and ethical surveillance deployment [19], [30].

**Index Terms**—Smart Surveillance System, Deep Learning, YOLO, Object Detection, Behavior Analysis, Crowd Monitoring

## I. INTRODUCTION

Surveillance systems play a crucial role in ensuring public safety and security, especially in crowded urban environments [18], [20]. Traditional CCTV-based monitoring relies heavily on manual observation, making it inefficient for real-time threat detection and crowd analysis [2]. Human operators often miss critical events due to fatigue, information overload, and limited attention span [2]. To overcome these limitations, Artificial Intelligence (AI) and Deep Learning (DL) technologies have been integrated into

surveillance systems [4], [18]. AI-based surveillance solutions can automatically detect individuals, track their movements, identify missing persons, and analyze human behavior patterns from live video streams [7], [9], [15]. Such systems significantly improve response time and monitoring accuracy [18]. This review paper focuses on smart surveillance systems that utilize object detection, multi-object tracking, face recognition, and abnormal behavior detection techniques [7], [9], [15]. The aim is to study existing literature, compare methodologies, and identify research gaps relevant to real-time intelligent surveillance systems [18], [20].

## II. LITERATURE REVIEW

### A. Traditional Surveillance Systems

Early surveillance systems relied on static CCTV cameras monitored manually by security personnel [2]. These systems lacked automation, real-time intelligence, and behavioral understanding [2]. They were primarily useful for post-event analysis rather than proactive security management [2]. Manual observation also introduced high chances of human error and delayed responses [2].

### B. Object Detection Techniques in Surveillance

Recent developments in object detection have significantly improved surveillance capabilities [7], [8], [28]. Deep learning-based models such as R-CNN, Fast R-CNN, SSD, and YOLO (You Only Look Once) are widely used for detecting humans and objects in video streams. YOLO-based architectures are particularly popular due to their real-time performance and high accuracy, making them suitable for crowd monitoring applications [7], [23]. YOLO detects

objects by predicting bounding boxes and class probabilities in a single forward pass, ensuring low latency [7].

C. Multi-Object Tracking Approaches

After detection, tracking individuals across frames is essential for understanding crowd movement [2], [9]. Techniques such as Kalman Filters, SORT, and Deep SORT are commonly used in surveillance systems [9]. Deep SORT improves tracking performance by integrating appearance descriptors and motion prediction, assigning unique IDs to individuals even in crowded and occluded environments [9]. This approach is highly effective for continuous person tracking in real-time scenarios [9].

D. Face Recognition in Surveillance Systems

Face recognition enables identification of missing persons or known individuals in surveillance video [16]. Modern systems utilize deep convolutional neural networks (CNNs) to extract facial embeddings and compare them with stored databases [4], [16]. Despite high accuracy, face recognition systems face challenges such as pose variation, lighting conditions, occlusion, and privacy concerns [16], [17]. Therefore, several studies emphasize secure storage and ethical handling of facial data [16].

E. Abnormal Behavior Detection

Behavioral analysis adds intelligence to surveillance systems by detecting suspicious activities such as running, fighting, loitering, or unauthorized access [12], [15]. Existing research employs CNN-based spatial feature extraction combined with VideoMAE networks to model temporal motion patterns [15], [26]. These models can successfully identify abnormal behavior based on movement trajectories and activity sequences [12], [27]. However, behavior detection accuracy is affected by crowd density, environmental factors, and data availability [12], [29].

III. COMPARISON OF EXISTING SYSTEMS

Techniq	Purpose	Advantages	Limitations
YOLO	Object detection	Real-time, fast	Sensitive to occlusion
Deep SORT	Tracking	Robust ID assignment	Dependent on

			detection quality
Face CNNs	Recognition	High accuracy	Privacy issues
CNN-Video MAE	Behavior analysis	Captures temporal patterns	Requires large datasets

IV. CHALLENGES IN SMART SURVEILLANCE SYSTEMS

High computational cost due to multi-camera feeds [19]. Handling large-scale deployments efficiently [19], [30]. Real-Time Processing and Latency Constraints [19]. Scalability Constraints (Multi-Camera Feeds) [19], [30]. Privacy, Security, and Legal Constraints [17], [20].

V. RESEARCH GAPS AND FUTURE DIRECTIONS

Need for lightweight models for low-resource environments [19], [23]. Improved detection under occlusion and dense crowds [7], [23]. Privacy-preserving surveillance techniques [17]. Integration with smart city and IoT infrastructures [20], [30]. Edge-based surveillance for reduced latency [19]. Future research should focus on efficient model optimization, ethical AI frameworks, and scalable real-time deployment strategies [19], [30].

VI. PUBLICATION PRINCIPLES

This review paper adheres to the publication principles of the International Journal of Innovative Research in Technology (IJIRT). The work focuses on reviewing and analyzing existing research contributions in the field of AI-based smart surveillance systems rather than presenting experimental or proprietary results.

The paper includes adequate discussion of previously published techniques related to object detection, multi-object tracking, face recognition, and abnormal behavior analysis. All concepts discussed have been interpreted and rewritten in original academic language to avoid duplication of previously published text. The review emphasizes clarity, technical accuracy, and reproducibility of ideas presented in existing literature.

Sufficient technical descriptions have been provided to allow readers to understand the reviewed

methodologies and their applicability to real-time surveillance scenarios. Ethical and legal considerations related to surveillance systems, including privacy protection and responsible deployment, have also been acknowledged as part of responsible scholarly communication.

## VII. CONCLUSION

This review paper presented a comprehensive study of AI-based smart surveillance systems designed for real-time crowd monitoring, missing person detection, and abnormal behavior identification [18], [20]. The integration of deep learning techniques such as YOLO for object detection, Deep SORT for tracking, and CNN-VideoMAE models for behavioral analysis has significantly improved the accuracy and efficiency of modern surveillance systems [7], [9], [15].

The review highlights that while current systems demonstrate promising performance, challenges related to real-time processing, scalability, environmental variability, and data privacy continue to limit large-scale deployment [17], [19], [20]. Addressing these challenges through optimized models, edge computing, and privacy-preserving techniques is essential for future development [19], [30].

Overall, intelligent surveillance systems have the potential to enhance public safety and operational efficiency when designed and deployed responsibly [20]. Continued research in this domain will contribute to more reliable, ethical, and scalable surveillance solutions for smart cities and public infrastructures [19], [30].

## VIII. APPENDIX

### Appendix A: Acronyms and Abbreviations

AI – Artificial Intelligence  
 CNN – Convolutional Neural Network  
 VideoMAE – Transformer-based video classification  
 YOLO – You Only Look Once  
 GPU – Graphics Processing Unit  
 RTSP – Real-Time Streaming Protocol  
 DL – Deep Learning

### Appendix B: Key Technologies Reviewed

YOLO-based Object Detection Models  
 Deep SORT Multi-Object Tracking Algorithm  
 CNN and VideoMAE-based Behavioral Analysis  
 Face Recognition using Deep Feature Embeddings

## ACKNOWLEDGMENT

The authors would like to express sincere gratitude to their faculty members and project guides for their valuable guidance and continuous support throughout the preparation of this review paper. The authors also acknowledge the availability of open-source research publications and toolkits that have contributed to understanding recent advancements in intelligent surveillance systems.

## REFERENCES

- [1] P. Viola and M. Jones, "Rapid Object Detection Using a Boosted Cascade of Simple Features," IEEE Conference on Computer Vision and Pattern Recognition, 2001.
- [2] A. Yilmaz, O. Javed, and M. Shah, "Object Tracking: A Survey," ACM Computing Surveys, vol. 38, no. 4, 2006.
- [3] R. Mehran, A. Oyama, and M. Shah, "Abnormal Crowd Behavior Detection Using Social Force Model," IEEE Conference on Computer Vision and Pattern Recognition, 2009.
- [4] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," Advances in Neural Information Processing Systems, 2012.
- [5] H. Wang and C. Schmid, "Action Recognition with Improved Trajectories," IEEE International Conference on Computer Vision, 2013.
- [6] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," International Conference on Learning Representations, 2015.
- [7] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," IEEE Conference on Computer Vision and Pattern Recognition, 2016.
- [8] W. Liu, D. Anguelov, D. Erhan, et al., "SSD: Single Shot MultiBox Detector," European Conference on Computer Vision, 2016.
- [9] N. Wojke, A. Bewley, and D. Paulus, "Simple Online and Realtime Tracking with a Deep Association Metric," IEEE International Conference on Image Processing, 2017.
- [10] M. Ravanbakhsh, M. Nabi, E. Ricci, and N. Sebe, "Abnormal Event Detection in Videos Using

- Generative Adversarial Nets,” IEEE International Conference on Image Processing, 2017.
- [11] L. Chen, Y. Wei, and Z. Guo, “Real-Time Intelligent Video Surveillance System Based on Deep Learning,” *International Journal of Computer Applications*, vol. 182, no. 2, pp. 32–38, 2018.
- [12] S. Sultani, C. Chen, and M. Shah, “Real-World Anomaly Detection in Surveillance Videos,” *IEEE Conference on Computer Vision and Pattern Recognition*, 2018.
- [13] D. Chen, H. Li, Z. Xiao, and W. Chen, “Crowd Behavior Analysis Based on Deep Learning for Video Surveillance,” *Journal of Visual Communication and Image Representation*, vol. 59, pp. 275–284, 2019.
- [14] T. H. Vu, A. Osokin, and I. Laptev, “Context-Aware CNN for Multi-Person Tracking,” *IEEE Conference on Computer Vision and Pattern Recognition*, 2019.
- [15] Z. Zhang, Y. Liu, and Y. Wang, “Human Action Recognition Using Convolutional Neural Networks and Long Short-Term Memory,” *IEEE Access*, vol. 8, pp. 209620–209631, 2020.
- [16] A. K. Jain, P. Flynn, and A. A. Ross, *Handbook of Biometrics*, Springer, 2020.
- [17] A. Z. Koubaa and H. Bennaceur, “Privacy-Aware Intelligent Video Surveillance Systems: A Review,” *Journal of Network and Computer Applications*, vol. 124, pp. 93–107, 2021.
- [18] Y. Liu, X. Wang, and S. Chen, “Deep Learning-Based Intelligent Surveillance Systems: A Review,” *IEEE Access*, vol. 9, pp. 112345–112360, 2021.
- [19] M. Dos Santos, J. Silva, and R. Oliveira, “Real-Time Smart Surveillance Using Edge AI and Deep Learning,” *Sensors*, vol. 22, no. 18, pp. 1–19, 2022.
- [20] T. Nguyen, A. Pham, and Q. Le, “AI-Based Video Surveillance for Smart Cities: Recent Advances and Future Challenges,” *Journal of Intelligent Systems*, vol. 33, no. 1, pp. 45–62, 2024.
- [21] J. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, “YOLOv4: Optimal Speed and Accuracy of Object Detection,” *arXiv preprint arXiv:2004.10934*, 2020.
- [22] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, “YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors,” *arXiv preprint arXiv:2207.02696*, 2022.
- [23] G. Jocher, A. Chaurasia, and J. Qiu, “YOLOv8: Next-Generation Ultralytics Real-Time Object Detector,” *Ultralytics Technical Report*, 2023.
- [24] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri, “Learning Spatiotemporal Features with 3D Convolutional Networks,” *IEEE International Conference on Computer Vision (ICCV)*, 2015.
- [25] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
- [26] W. Luo, W. Liu, and S. Gao, “Remembering History with Convolutional LSTM for Anomaly Detection,” *IEEE International Conference on Multimedia and Expo (ICME)*, 2017.
- [27] Y. Liu, X. Wang, L. Zhou, and Y. Li, “Future Frame Prediction for Anomaly Detection – A New Baseline,” *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.
- [28] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017.
- [29] B. Kiran, D. Thomas, and C. Parakkal, “An Overview of Deep Learning-Based Methods for Unsupervised and Semi-Supervised Anomaly Detection in Videos,” *Journal of Imaging*, vol. 4, no. 2, 2018.
- [30] Z. Fang, J. Li, and Y. Tian, “Edge-AI-Enabled Real-Time Smart Surveillance: Architecture, Applications, and Challenges,” *IEEE Internet of Things Journal*, 2023.