International Journal of Creative and Open Research in Engineering and Management
ISSN: **3108-1754 (Online)**
Volume 02 Issue 06 June-2026 | Impact Factor: 3.5

Smart Crowd Management and RFID-Assisted Tracking System Using YOLOv8 and DeepSORT

Pradip Paithane
Department of Information Technology
Baramati, India
paithanepradip@gmail.com

Alisha Shikandar Shaikh
Department of Information Technology
Baramati, India
alishashaikh2210@example.com

Anushka Suresh Kharat
Department of Information Technology
Baramati, India
anushkakharat86@gmail.com

Jadhav Rutuja Ramesh
Department of Information Technology
Baramati, India
rutujadhav4804@gmail.com

Priyanka Ramesh Bade
Department of Information Technology
Baramati, India
badepriyanka2021@gmail.com

***Abstract*—Crowd surveillance is an essential part of crowd safety and management, especially in large gatherings where manual surveillance becomes ineffective. This paper aims to provide a thorough review of the existing literature on crowd detection and tracking systems, with a main emphasis on deep learning-based solutions. In this review, special attention is given to the You Only Look Once (YOLO) family of object detection models, especially YOLOv8, and their combination with multi-object tracking algorithms like DeepSORT. The review examines how recent advances in anchor-free detection, decoupled head models, and appearance-tracking methods have enhanced real-time performance in complex crowd settings. The literature has also identified some of the key challenges in crowd surveillance as extreme occlusion, lighting changes, identity switches, and poor cross-dataset generalization. Finally, this paper highlights some of the open research areas in crowd surveillance, including multimodal data fusion, multi-camera tracking, and adaptive AI models, which are necessary for developing efficient intelligent crowd surveillance systems.**

***Index Terms*—YOLOv8, DeepSORT, crowd detection, real-time surveillance, abnormal behavior detection, dense crowds, object tracking, computer vision, deep learning**

I. INTRODUCTION

Large public gatherings, such as religious events, sports, and political rallies, tend to have highly dense crowds in a closed environment. Occurrences such as the Kumbh Mela bring together millions of people, thus increasing the chances of dense crowds and abnormal crowd behavior. The safety of the public in such events can only be ensured by monitoring and detecting potential danger situations. However, manual surveillance by CCTV cameras is highly dependent on the human operator and tends to be less effective in high-density and complex environments [4], [12].

To address these problems, there has been a growing need for automated crowd surveillance systems based on computer vision and deep learning techniques in recent years. These systems are intended to automatically detect, track, and analyze pedestrian behavior in real-time, enabling decision-makers to act accordingly. Among the various object detection techniques, the You Only Look Once (YOLO) series of single-stage object detection algorithms have gained popularity due to their ability to achieve high detection accuracy with low latency, making them suitable for real-time surveillance applications [1], [19].

With continuous architectural improvements, recent YOLO versions have demonstrated enhanced performance capabilities in crowded environments. Various studies have indicated that recent YOLO versions have demonstrated enhanced feature learning capabilities, faster convergence rates, and enhanced robustness to challenging conditions such as illumination changes and overlapping persons [1], [5]. In particular, YOLOv8 introduces a new anchor-free detection mechanism and decoupled head architecture, which have been shown to improve detection accuracy and robustness in crowded environments [5], [6], [13].

Detection alone is not sufficient for efficient crowd monitoring, and it is equally important to maintain consistency in identity across video frames. For this, multi-object tracking techniques such as DeepSORT are commonly coupled with YOLO-based detectors. DeepSORT integrates motion prediction via a Kalman filter with appearance-based feature embeddings, which helps in efficient identity linking even in cases of partial occlusion and re-entry [8], [28]. A few recent works have demonstrated that the integration of YOLOv8 with DeepSORT helps in improving the stability of tracking and minimizing identity swaps in challenging pedestrian settings [5], [6], [8].

Apart from the approaches using deep learning, physics-inspired models like the Social Force Model and fluid dynamics-based methods have also been considered for crowd behavior analysis. These models offer macroscopic perspectives on the crowd behavior, in addition to data-driven crowd detection and tracking methods [27]. Nevertheless, these models are not flexible enough or in real-time as required in contemporary surveillance systems.

In addition to vision-based surveillance, Radio Frequency Identification (RFID) technology can enhance crowd management by enabling the identification and localization of registered individuals carrying RFID-enabled wristbands or tags. RFID systems provide unique identification without requiring direct visual contact, making them useful in crowded environments where camera-based tracking may be affected by severe occlusions. The integration of RFID with YOLOv8 and DeepSORT can improve person tracking accuracy, support missing-person identification, and facilitate real-time crowd monitoring during large-scale events such as the Kumbh Mela.

In this paper, a thorough review of the current state of art in crowd detection and tracking methods is presented, with a special emphasis on YOLOv8-based crowd detection models and their combination with DeepSORT crowd tracking frameworks. Through the analysis of recent literature, the most important developments, challenges, and open research issues in the area of crowd surveillance systems are outlined, including the handling of severe occlusions, identity association across cameras, and robustness to different environments.

II. EVOLUTION OF OBJECT DETECTION FOR CROWD ANALYSIS

Object detection is the building block of any intelligent video surveillance system, as it has a direct impact on the performance of the tracking and behavior analysis modules that follow. In the context of crowd surveillance, the problem at hand is to detect individuals under heavy occlusion, scale changes, and spatial overlap. The YOLO family of object detectors has shown a clear evolutionary path over the past decade, from anchor-based models to more flexible and efficient anchor-free models.

A. The Pre-YOLOv8 Era

The initial forms of YOLO created the foundation for real-time detection but had challenges in dense settings. YOLOv1 relied on a grid system that had difficulties with small objects, and YOLOv2 incorporated anchor boxes for enhanced localization. The emergence of YOLOv3 presented a major improvement with the addition of the Feature Pyramid Network (FPN), which enabled multi-scale detection, an important aspect of detecting both close and far-away people in a crowd.

In addition to the YOLO-based methods, part-aware and density-aware detection models have also played important roles in pedestrian detection research. Yang et al. [25] proposed a part-aware multi-scale fully convolutional network, which improves detection performance by modeling each body part separately, especially in the case of partial occlusion. Also, Tian et al. [26] presented PaDNet for pan-density crowd counting, which handles scale changes using density-adaptive learning strategies. These studies have paved the way for current multi-scale feature aggregation methods used in YOLO-based detectors.

The subsequent versions, YOLOv4 and YOLOv5, concentrated on optimizing speed and efficiency. YOLOv4 brought about the CSPDarknet backbone network and Mosaic data augmentation, while YOLOv5 gained popularity for its PyTorch ecosystem with a modular design apt for industrial use [19]. YOLOv7 advanced further with the E-ELAN architecture, which provided a better flow of gradients for the detection of small overlapping objects [14]. The evolution of these models up to YOLOv7 is presented in Table I, which emphasizes their applicability for crowd detection.

TABLE I
COMPARISON OF YOLO FAMILY (V1 TO V7) FOR CROWD DETECTION

Version	**Backbone**	**Key Innovation**	**Suitability**	**Advantages**	**Limitations**
YOLOv1 [28]	Custom CNN	Unified detection in a single pass (Grid System).	Very Low	• Extremely fast for real-time use. • Conceptually simple.	• Struggles with small objects. • Poor localization in dense groups.
YOLOv2 [28]	Darknet-19	Anchor Boxes & Batch Normalization.	Low-Mod	• Higher recall than v1. • Robust to input size variation.	• Misses many small/overlapping objects.
YOLOv3 [25]	Darknet-53	Feature Pyramid Network (FPN) for Multi-scale detection.	Moderate	• Detects objects at three scales. • Improved small-object accuracy.	• Slower inference. • ID switches under occlusion.
YOLOv4 [19]	CSPDarknet 53	CSPNet & PANet with advanced augmentation.	High	• GPU-optimized parallelism. • Mosaic augmentation improves robustness.	• Large model size. • Complex tuning.
YOLOv5 [13]	CSPDarknet (Mod)	PyTorch implementation with auto-anchor learning.	High	• Easy deployment. • Modular variants.	• Limited robustness in dense occlusion. • Accuracy plateau.
YOLOv7 [1], [14]	E-ELAN	Efficient Layer Aggregation with model scaling.	High	• Superior gradient flow. • Strong small-object detection.	• Higher computational demand than YOLOv5.

B. The Rise of YOLOv8

The development of YOLOv8 by Ultralytics is a significant paradigm shift in object detection architecture, as the earlier versions were not designed for high-density areas. As mentioned in the recent research articles by Oise et al. [5] and Ujlambkar et al. [6], YOLOv8 has a decoupled detection head where classification and bounding box regression are processed separately. This makes the optimization process more stable and faster, along with better detection accuracy, which is a critical requirement for people detection in a crowded religious gathering.

Anchor-Free Detection Mechanism: Unlike the earlier YOLO versions (v1-v7), YOLOv8 completely does away with the requirement for anchor boxes. According to Alkandary et al. [1], the anchor-free mechanism of YOLOv8 enables the model to learn dynamically about irregular object shapes and spatial distributions. This is highly beneficial for crowd surveillance, where overlapping human figures, different viewpoints, and partial occlusions are quite prevalent. The absence of anchor box dimension tuning requirements for crowd datasets makes YOLOv8 more flexible and robust.

Enhanced Feature Aggregation and Multi-Scale Detection: The YOLOv8 model adopts a Cross Stage Partial (CSP) backbone network with a PANet neck, which assists in efficient multi-scale feature fusion. According to Ujlambkar et al. [6], this is required for crowd density mapping because it allows the detector to accurately identify individuals at different ranges from the camera, ranging from close-up shots of pedes-trians to small distant figures in wide-area surveillance videos. In addition, Gu¨ndu¨z and Is¸ık [19] indicate that this enhanced feature extraction process results in superior performance on conventional crowd datasets compared to the earlier YOLO versions.

Superior Performance in Comparative Studies: A full-fledged comparative analysis carried out by Alkandary et al. [1], comparing the different versions of YOLO, from v3 to v10, confirmed that YOLOv8 is indeed superior to its earlier versions. According to the analysis, YOLOv8 has a more optimal trade-off between the accuracy of detection, resistance to occlusion, and the speed of processing. This was further supported by Sheng et al. [8] in the context of pedestrian tracking, where the enhanced accuracy of YOLOv8 was found to have a direct positive impact on identity linking when combined with DeepSORT.

Suitability for Real-Time Anomaly Detection: The real-time processing abilities of YOLOv8 have also been further confirmed in applications related to anomaly detection. Oise et al. [5] demonstrated the effectiveness of YOLOv8-DeepSORT with attention and adaptive optimization, confirming that the model is capable of maintaining high frame rates when processing high-density video streams. This confirms YOLOv8 as not only a detector but also the core component of intelligent crowd surveillance systems that can notify authorities of abnormal crowd behaviors in real-time.

Khan et al. [7] focused on violence detection in industrial crowd surveillance settings using deep spatio-temporal learning. This study emphasizes the significance of modeling temporal features for anomaly detection, indicating that spatial detection alone (YOLO, for instance) may not be sufficient in modeling behavioral anomalies.

Alsubai et al. [15] have discussed AI-based crowd density analysis for sustainable smart cities. Their work highlights the need for distributed processing frameworks in massive gatherings.

Li et al. [17] have described the use of explainable AI in crowd density analysis models. Their approach using attention maps is a crucial requirement for public safety applications, where transparency is paramount.

Mohanty et al. [18] have designed PublicVision, which concentrates on secure surveillance systems. Their work highlights privacy-preserving systems, which is a significant ethical consideration that is usually neglected in detection-tracking studies.

III. DEEPSORT FOR MULTI-OBJECT TRACKING

Deep Simple Online and Real-Time Tracking (DeepSORT) is a paradigm shift in multi-object tracking (MOT) as it extends the SORT algorithm to overcome its major drawback of identity swap during occlusion. Although originally proposed by Wojke et al. [28], DeepSORT combines appearance descriptors with motion prediction, making it the de facto method for real-time tracking in surveillance systems. The ability of DeepSORT to preserve identity across consecutive frames has been proven in several works concentrating on high-density scenarios [27].

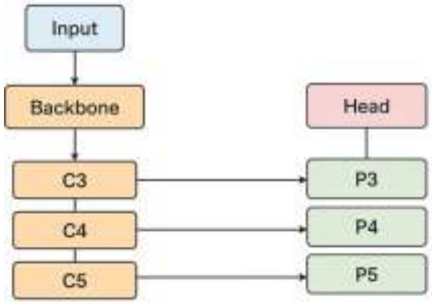

Fig. 1. YOLO V8 Architecture.

Lin et al. [23] introduced a unified framework that combines detection and re-identification. This is achieved by integrating detection and re-identification into a single deep learning model, which decreases the inconsistency of features between modules. Compared to the joint embedding method, DeepSORT has a faster training process.

Lee et al. [11] generalized multi-object tracking to group detection and clustering, focusing more on behavioral analysis than mere trajectory continuity. The results show that trajectory clustering can be used for early warnings of abnormal crowd formation behavior.

Du et al. [21] further improved DeepSORT with StrongSORT, adding appearance-linking strategies and motion-appearance fusion, which greatly reduced fragmentation in long-term tracking.

A. Architecture and Mechanism

The strength of DeepSORT lies in its two-branch structure, which combines motion and appearance features in an effective manner:

- **Motion Modeling Branch:** Using a Kalman filter, this branch predicts the future position of bounding boxes based on past motion cues. Although the SORT tracker solely depended on this feature, resulting in a failure rate in occluded scenarios, DeepSORT makes use of this feature as a spatial constraint to search for associations.
- **Appearance Modeling Branch:** This is the most important part of the DeepSORT algorithm for crowd monitoring. This branch makes use of a Convolutional Neural Network (CNN) that is trained on large-scale person re-identification datasets to learn a discriminative visual embedding for detected persons. Using the cosine distance between the feature vectors, the algorithm is

capable of re-identifying a person even if it is temporarily hidden by other persons or obstacles.

For data association, DeepSORT employs a matching cascade strategy followed by the Hungarian algorithm. This approach prioritizes recently observed tracklets, which is particularly effective in real-time applications where minimizing latency is crucial.

B. Enhancements and Variants in Crowd Research

Although the traditional DeepSORT offers a good baseline, some recent studies have found flaws in the algorithm when dealing with very dense scenarios, thus giving rise to the concept of ”Improved” variants and successors.

Improved DeepSORT for Pedestrian Tracking: In the context of crowd control, Sheng et al. [8] presented a multi-object pedestrian tracking approach using YOLOv8 and Improved DeepSORT. Their contribution focused on optimizing the association score tailored for human motion, and they were able to show a substantial decrease in ID switches during the tracking process in occluded settings.

StrongSORT and Advanced Association: Du et al. [21] presented StrongSORT, which essentially overcame the ”fragmentation” problems that come with the original DeepSORT algorithm. StrongSORT integrates an appearance-linking module and relies on the affinity of joint motion and appearance for track association. A comparison study by Alkandary et al. [1] showed that, although the traditional DeepSORT is still a very competitive approach, StrongSORT performs better in terms of tracking accuracy in very cluttered environments.

Attention and Adaptive Optimization: Oise et al. [5] further optimized the YOLOv8-DeepSORT system by using attention mechanisms and adaptive optimization techniques. Their ”YOLOv8-DeepSORT” system demonstrated that improving the feature extraction procedure in the tracker leads to better performance in variable lighting conditions and fast crowd dynamics, which are typical of events such as the Kumbh Mela.

C. Suitability for Crowd Surveillance

The literature clearly verifies that the combination of DeepSORT with new detectors produces a synergy effect for mass gathering monitoring.

- **Occlusion Handling:** Ujlambkar et al. [6] used DeepSORT for crowd density estimation, highlighting that the key to successful crowd density estimation is the use of the appearance model to ensure accurate counting when people move behind buildings or when they are densely packed.
- **Scalability:** Athilakshmi et al. [20] showed that the combination of DeepSORT with new detectors such as YOLONAS improves the real-time human tracking function, which indicates that DeepSORT is flexible enough to work with new architectures of detectors.

In conclusion, DeepSORT and its variants have the required infrastructure to convert discrete detection results into continuous meaningful paths. This is particularly important for crowd safety monitoring, where it is important to analyze not only the presence of people in the scene but also the flow of people over time [14], [20].

IV. APPLICATION TO MASS GATHERINGS: KUMBH MELA AND HAJJ

Religious mass gatherings are some of the most difficult real-world scenarios for the application of intelligent crowd surveillance systems, especially due to the high density of the crowd, diversity of crowd motion, and environmental factors. Of these, the Hajj and Kumbh Mela are often cited in the literature as benchmark scenarios for assessing the scalability and efficacy of computer vision solutions.

A. The Hajj Benchmark

The Hajj, which is a pilgrimage to Saudi Arabia, is the most similar real-world scenario to the Kumbh Mela in terms of crowd density and safety requirements. There have been a few instances in the literature where the Hajj scenario has been used to assess the efficacy of state-of-the-art crowd surveillance solutions. Albattah et al. [24] used CNN-based solutions for crowd density estimation and management during the Hajj, where they pointed out the importance of using effective detection models that can deal with extreme densities and dynamic crowd behavior in religious gatherings.

B. The Kumbh Mela Context

In the context of the Kumbh Mela, which is known to be one of the largest religious events in the world, scalability and real-time processing capabilities are of utmost importance. Khamitkar [4] highlights the importance of intelligent and AI-powered infrastructure that is capable of processing large amounts of video data over a vast geographical region. A YOLOv8-DeepSORT based surveillance system is most appropriate for dealing with the operational requirements of the Kumbh Mela for the following reasons:

- **Real-Time Processing:** YOLOv8 has the capability of performing high-speed processing while retaining high levels of detection accuracy [1], which makes it possible to process multiple CCTV feeds simultaneously. This is particularly important for dealing with the large perimeter and congregation areas that are present at the Kumbh Mela.
- **Occlusion Handling:** The dynamic nature of the Mela, which is characterized by high levels of chaos, leads to frequent and prolonged periods of occlusion due to the presence of large human crowds and temporary structures. DeepSORT’s appearance-based re-identification system has the capability of maintaining continuous tracking of individuals even during periods of temporary occlusion, thus preventing fragmentation of identity [6].
- **Zone-Based Monitoring:** As suggested by Siva et al. [3] and Alasmari et al. [12], dividing the area of surveillance into several zones and using threshold alert systems helps in decentralized crowd management. This way of zone-wise surveillance helps in identifying the early signs

of local congestion conditions, which in turn prevents stampede and congestion-related accidents.

Thus, the findings from both Hajj and Kumbh Mela-related researches validate that the integration of YOLO detection with effective multi-object tracking is a feasible and scalable solution for intelligent surveillance systems in religious mass gatherings.

C. *Role of RFID in Crowd Management*

Radio Frequency Identification (RFID) technology has emerged as an effective solution for crowd management in large public gatherings. RFID tags embedded in wristbands, identity cards, or wearable devices allow the unique identification of individuals through radio signals. Unlike computer vision systems, RFID does not require line-of-sight visibility and can function even in highly congested environments.

In a Kumbh Mela scenario, RFID-enabled wristbands can be distributed to pilgrims during registration. RFID readers installed at entry gates, exit points, checkpoints, and critical zones continuously collect location information of tagged individuals. This information can be integrated with YOLOv8-based crowd detection and DeepSORT-based tracking systems to provide enhanced situational awareness.

The combined RFID and computer vision framework offers several advantages, including improved tracking accuracy, identification of missing persons, reduced identity-switching problems, crowd flow analysis, and emergency response support. Furthermore, RFID-generated location data can serve as an additional verification layer when visual tracking becomes unreliable due to severe occlusions or poor lighting conditions.

V. PERFORMANCE ANALYSIS AND BENCHMARKING

A comparison of benchmarking studies shows that YOLOv8-based systems always provide better precision-recall tradeoff in dense crowd settings. Alkandary et al. [1] showed that YOLOv8 surpasses YOLOv5 and YOLOv7 in terms of mAP50-95 while having comparable inference speed. Gunduz and Isik [19] achieved mAP50-95 of 53.2

Regarding tracking performance, Sheng et al. [8] achieved MOTA of 77.5

Nonetheless, cross-dataset generalization is still a challenge. Alasmari et al. [12] pointed out that there is no common crowd surveillance benchmark, leading to variable evaluation metrics among studies. Physics-inspired models reviewed by Zhang et al. [27] are helpful for macroscopic analysis but are not applicable for real-time analysis.

These results suggest that, although YOLOv8-DeepSORT integration performs well in real-time tracking, there is still a need for further research in domain adaptation, benchmark standardization, and multi-camera association.

VI. SYSTEM DESIGN AND UML DIAGRAMS

This section presents the Unified Modeling Language (UML) diagrams that describe the overall design and workflow of the proposed intelligent crowd surveillance system. The diagrams illustrate the interaction between users and the system, the sequence of operations, the flow of activities, the structural organization of system components, and the deployment architecture. These UML models provide a clear understanding of the proposed framework before discussing the underlying detection and tracking techniques.

A. *Use Case Diagram*

The use case diagram illustrates the interaction between the primary actors, such as the Administrator, Security Officer, and Surveillance System. It represents the major functionalities of the proposed system, including crowd monitoring, person detection, object tracking, anomaly detection, alert generation, and report management.

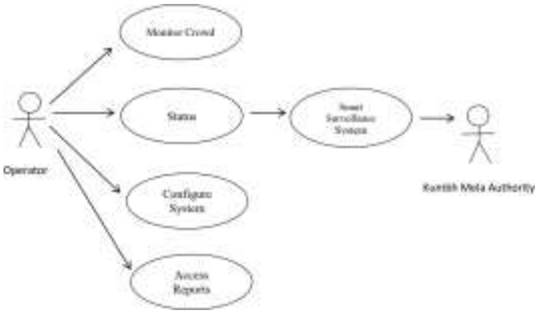

Fig. 2. Use Case Diagram of the Smart Crowd Surveillance System

B. *Activity Diagram*

The activity diagram describes the workflow of the proposed surveillance system. It begins with capturing video streams from CCTV cameras, followed by frame preprocessing, person detection using YOLOv8, object tracking using DeepSORT, anomaly detection, alert generation, and continuous monitoring.

C. *Sequence Diagram*

The sequence diagram illustrates the interaction between different system components during the surveillance process. It shows how video frames are processed, objects are detected and tracked, anomalies are identified, and alerts are sent to the administrator.

D. *Class Diagram*

The class diagram represents the structural organization of the proposed system. It includes the major classes such as Camera, VideoProcessor, YOLODetector, DeepSORTTracker, AlertManager, Database, and Administrator along with their relationships and responsibilities.

E. *Deployment Diagram*

The deployment diagram illustrates the physical architecture of the proposed system. It shows the deployment of CCTV cameras, processing server, deep learning models, database server, and administrator workstation connected through the network for real-time crowd surveillance.

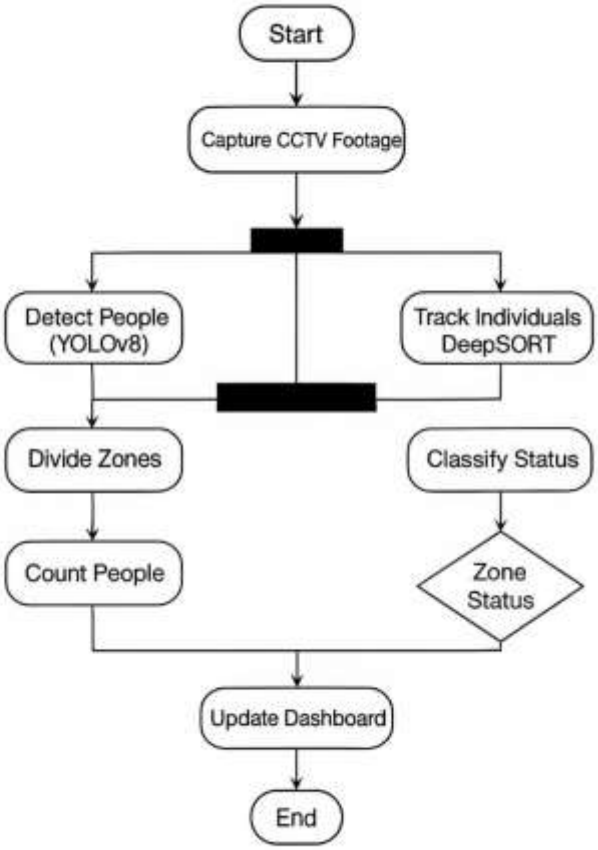

Fig. 3. Activity Diagram of the Smart Crowd Surveillance System

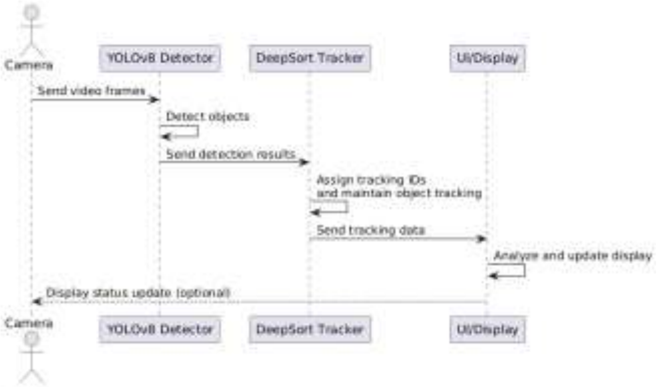

Fig. 4. Sequence Diagram of the Smart Crowd Surveillance System

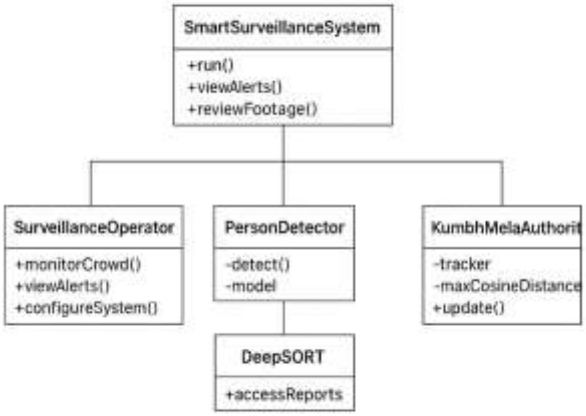

Fig. 5. Class Diagram of the Smart Crowd Surveillance System

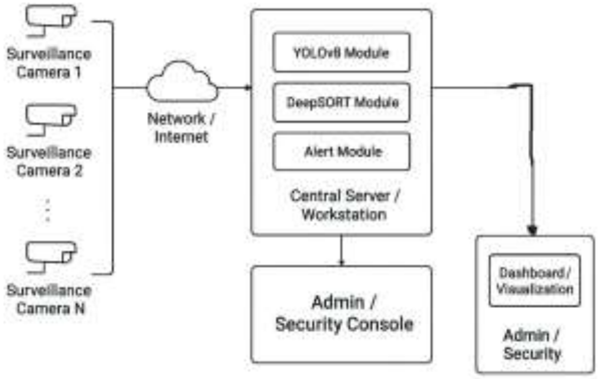

Fig. 6. Deployment Diagram of the Smart Crowd Surveillance System

VII. CONCLUSION

This paper has provided a comprehensive review of the recent advancements in intelligent crowd surveillance, with a special focus on the deep learning-based detection and tracking approaches. The YOLO family of models has been reviewed, with a special emphasis on the role of architectural improvements, specifically the anchor-free paradigm and decoupled head of YOLOv8, in enhancing the performance of real-time processing in dense crowd surveillance.

The integration of YOLOv8 with DeepSORT has been utilized to address the challenges of identity preservation. It is evident from the existing literature that the integration of appearance-based feature embedding and motion prediction has led to the reduction of identity switches and enhanced the stability of tracking in partial occlusion conditions. These factors make the YOLOv8-DeepSORT system amenable for real-time crowd surveillance in large-scale public events.

However, there are still some challenges that need to be addressed, such as severe occlusion, identity association across cameras, and illumination variations. Future research should aim at multimodal data fusion, multi-camera tracking, and adaptive learning. Furthermore, the application of predictive analytics and explainable AI will play a pivotal role in enhancing the situational awareness and trustworthiness of automated crowd surveillance systems for events such as Kumbh Mela and Hajj. Future research should focus on multimodal data fusion by integrating computer vision techniques with RFID, IoT sensors, and multi-camera networks. Such hybrid systems can improve tracking reliability, enhance crowd safety, and support real-time decision-making in large-scale public gatherings.

REFERENCES

[1] K. Alkandary, A. S. Yildiz, and H. Meng, “A comparative study of YOLO series (v3–v10) with DeepSORT and StrongSORT: A real-time tracking performance study,” *Electronics*, vol. 14, no. 5, p. 876, Feb. 2025, doi: 10.3390/electronics14050876.

[2] R. Nasir, Z. Jalil, M. Nasir, and T. Alsubait, “An enhanced framework for real-time dense crowd abnormal behavior detection using YOLOv8,” *Artificial Intelligence Review*, vol. 58, no. 4, 2025.

[3] P. Siva *et al.*, “Smart surveillance systems using YOLOv8: A scalable approach for crowd and threat detection,” *International Journal of Research in Applied Engineering and Technology (IJRAET)*, vol. 14, no. 1, 2025.

[4] N. S. Khamitkar, “The role of artificial intelligence in enhancing the Maha Kumbh Mela 2025: A technological revolution in pilgrimage management,” *International Journal of Multidisciplinary Research*, 2025.

[5] G. P. Oise *et al.*, “YOLOv8-DeepSORT: A high-performance framework for real-time multi-object tracking with attention and adaptive optimization,” *Journal of Science Research and Reviews*, vol. 2, no. 2, pp. 92–100, May 2025, doi: 10.70882/josrar.2025.v2i2.50.

[6] D. Ujlambkar *et al.*, “Crowd density mapping and anomaly detection using YOLOv8 and DeepSORT,” *International Journal of Scientific Research in Engineering and Management (IJSREM)*, vol. 9, no. 6, pp. 1–19, Jun. 2025, doi: 10.55041/IJSREM49725.

[7] H. Khan, X. Yuan, L. Qingge, and K. Roy, “Violence detection from industrial surveillance videos using deep learning,” *IEEE Access*, vol. 13, pp. 15363–15375, Jan. 2025, doi: 10.1109/ACCESS.2025.3531213.

[8] W. Sheng, J. Shen, Q. Huang, Z. Liu, and Z. Ding, “Multi-object pedestrian tracking method based on YOLOv8 and improved DeepSORT,” *Mathematical Biosciences and Engineering*, vol. 21, no. 2, pp. 1791–1805, Jan. 2024, doi: 10.3934/mbe.2024077.

[9] Y. Chen *et al.*, “Learning discriminative features for crowd counting,” *IEEE Transactions on Image Processing*, vol. 33, pp. 3749–3763, Jun. 2024, doi: 10.1109/TIP.2024.3408609.

[10] A. Aziz and A. Bajpai, “Attire-based anomaly detection in restricted areas using YOLOv8,” *arXiv preprint*, arXiv:2404.00645, 2024.

[11] H. Lee *et al.*, “Real-time human group detection and clustering in crowded environments using enhanced multi-object tracking,” *IEEE Access*, vol. 12, pp. 184027–184039, Dec. 2024, doi: 10.1109/ACCESS.2024.3503661.

[12] A. M. Alasmari, N. S. Farooqi, and Y. A. Alotaibi, “Recent trends in crowd management using deep learning techniques: A systematic literature review,” *Journal of Umm Al-Qura University for Engineering and Architecture*, vol. 15, pp. 355–383, Jun. 2024, doi: 10.1007/s43995-024-00071-3.

[13] S. D. Cheke and M. U. Karande, “Crowd detection using YOLOv8,” *Journal of Emerging Technologies and Innovative Research (JETIR)*, vol. 11, no. 8, Aug. 2024.

[14] D. Kabangira and Y. Oh, “Detecting and tracking people in crowded environments with YOLOv7 and DeepSORT,” *Nanotechnology Perceptions*, vol. 20, no. 7, pp. 149–157, 2024.

[15] S. Alsubai *et al.*, “Design of artificial intelligence-driven crowd density analysis for sustainable smart cities,” *IEEE Access*, vol. 12, pp. 121983–121993, Apr. 2024, doi: 10.1109/ACCESS.2024.3390049.

[16] S. Khan *et al.*, “Enhancing crowd safety at Hajj: Real-time detection of abnormal behavior using YOLOv9,” *IEEE Access*, vol. 12, pp. 124567–124580, May 2024, doi: 10.1109/ACCESS.2024.3401256.

[17] J. Li *et al.*, “Integrating explainable artificial intelligence with advanced deep learning models for crowd density estimation,” *IEEE Access*, vol. 12, pp. 125223–125234, May 2024, doi: 10.1109/ACCESS.2024.3408902.

[18] S. N. Mohanty *et al.*, “PublicVision: A secure smart surveillance system for crowd behavior recognition,” *IEEE Access*, vol. 12, pp. 119842–119854, Apr. 2024, doi: 10.1109/ACCESS.2024.3385751.

[19] M. S. Gunduz and G. Isik, “A new YOLO-based method for real-time crowd detection from video and performance analysis of YOLO models,” *Journal of Real-Time Image Processing*, vol. 20, no. 5, Jan. 2023, doi: 10.1007/s11554-023-01276-w.

[20] R. Athilakshmi *et al.*, “Enhancing real-time human tracking using YOLONAS-DeepSORT fusion models,” in *Proc. 7th Int. Conf. Electronics, Communication and Aerospace Technology (ICECA)*, Chennai, India, 2023, pp. 1118–1125, doi: 10.1109/ICECA58529.2023.10394864.

[21] Y. Du *et al.*, “StrongSORT: Make DeepSORT great again,” *IEEE Transactions on Multimedia*, vol. 25, pp. 8725–8737, 2023, doi: 10.1109/TMM.2023.3240881.

[22] C. Jadhav, R. Ramteke, and R. K. Somkunwar, “Smart crowd moni-toring and suspicious behavior detection using deep learning,” *Revue d’Intelligence Artificielle*, vol. 37, no. 4, pp. 955–962, Aug. 2023.

[23] H.-W. Lin *et al.*, “Real-time multiple pedestrian tracking with joint detection and embedding deep learning model for embedded systems,” *IEEE Access*, vol. 10, pp. 51458–51471, May 2022, doi: 10.1109/ACCESS.2022.3173408.

[24] W. Albattah *et al.*, “Hajj crowd management using CNN-based approach,” *Computers, Materials and Continua*, vol. 66, no. 2, 2021.

[25] P. Yang *et al.*, “A part-aware multi-scale fully convolutional network for pedestrian detection,” *IEEE Transactions on Intelligent Trans-portation Systems*, vol. 22, no. 2, pp. 1125–1137, Feb. 2021, doi: 10.1109/TITS.2019.2963700.

[26] Y. Tian *et al.*, “PaDNet: Pan-density crowd counting,” *IEEE Transactions on Image Processing*, vol. 29, pp. 2714–2727, Jan. 2020, doi: 10.1109/TIP.2019.2952083.

[27] P. Paithane and S. Lokhande, “IoT based smart surveillance system for crowd monitoring,” *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, vol. 9, no. 4, pp. 2236–2241, 2020.

[28] X. Zhang, Q. Yu, and H. Yu, “Physics inspired methods for crowd video surveillance and analysis: A survey,” *IEEE Access*, vol. 6, pp. 66816–66830, Oct. 2018, doi: 10.1109/ACCESS.2018.2878733.

[29] N. Wojke, A. Bewley, and D. Paulus, “Simple online and real-time tracking with a deep association metric,” in *Proc. IEEE Int. Conf. Image Processing (ICIP)*, 2017, pp. 3645–3649, doi: 10.1109/ICIP.2017.8296962.

[30] M. Du, C. Jing, and M. Du, “Tag location method integrating GNSS and RFID technology,” *Journal of Global Positioning Systems*, vol. 14, no. 2, pp. 1–10, 2016.