"Deep Learning Based on Crowd Monitoring using Yolo Algorithm"

Payal Balasaheb Lawand[1], Mrunali Randhir Khairnar[2], Kalyani Sunil Shelke[3], Vijaya Pundalik Nikam[4], Prof. Priyanka P. Kakade[5]

[1,2,3,4,5]Department of Computer Engineering, Brahma Valley College of Engineering and Research Institute, Nashik

Abstract:
Effective crowd management is essential for ensuring public safety in large gatherings. Traditional deep learning approaches for crowd analysis, including people counting, detection, and movement tracking, often require high computational resources, making them unsuitable for real-time applications on edge devices. This paper presents a Convolutional Neural Network (CNN)-based model designed to efficiently process crowd data while optimizing computational and memory demands. The proposed system enables real-time people detection, tracking, and movement estimation, allowing authorities to monitor and manage crowds proactively. By leveraging lightweight deep learning techniques, the model ensures high accuracy while maintaining efficiency, making it suitable for smart surveillance and public safety applications.

Keywords: Crowd Management, Real-Time Crowd Analysis, People Detection, Tracking, Convolutional Neural Network (CNN), Edge Computing, Deep Learning, Movement Estimation, Smart Surveillance, Public Safety.

INTRODUCTION

Managing large crowds efficiently is crucial for ensuring public safety, especially in high-density gatherings such as festivals, sporting events, and transportation hubs. Traditional methods rely on manual monitoring or high-resource deep learning models, which are often impractical for real-time applications on edge devices. These approaches struggle with computational inefficiencies, leading to delays in detecting potential risks and responding effectively.

To address these challenges, this paper presents a Convolutional Neural Network (CNN)-based crowd management system that enables real-time people detection, tracking, and movement estimation. The model is optimized for computational efficiency, making it suitable for deployment on edge devices. By leveraging deep learning techniques, the system enhances surveillance capabilities while minimizing resource consumption. This solution provides a scalable and cost-effective approach to crowd monitoring, improving response times and reducing the risk of accidents in large public gatherings.

LITERATURE SURVEY

[1] "Realtime Crowd Monitoring—Estimating Count, Speed, and Direction of People Using Hybridized YOLOv4", utilizes the YOLOv4 model to enhance real- time object detection and tracking. This research, published in IEEE Access Transactions (2022), demonstrates how YOLOv4 can effectively estimate the number of people in a crowd, analyze their movement speed, and predict their direction, providing a reliable solution for large-scale crowd management.

[2] "The Limo Powered Crowd Monitoring System: Deep Life Modeling for Dynamic Crowd With Edge-Based Information Cognition", focuses on using multi-sensor devices and edge cloud computing to improve dynamic crowd analysis. Published in the IEEE Sensors Journal (Volume: 22, Issue: 18, September 2022), this research highlights the advantages of edge-based information processing, reducing latency and enhancing real-time decision-making in dynamic crowd environments.

[3] "Crowd Monitoring and Classification", published in IEEE Access Transactions (2022), explores advanced techniques in Intelligent Systems and Computing for crowd analysis. This study emphasizes how intelligent computing methods can be applied to classify different crowd behaviors, ensuring efficient

monitoring and safety management in public spaces.
[4] "A Unified Multi-Scale Deep Convolutional Neural Network for Fast Object Detection in Computer Vision— ECCV (Lecture Notes in Computer Science)", published in IEEE Access Transactions (September 2022), presents a multi-scale deep learning approach for improving object detection speed and accuracy. This method enhances real- time performance in crowd surveillance systems by leveraging deep convolutional neural networks to quickly and efficiently detect individuals in crowded scenes.

METHODOLOGY

The proposed system follows a structured approach to enable real-time crowd detection, tracking, and movement estimation. It begins with data collection and preprocessing, where images and videos from public crowd datasets are gathered to train the model. To improve model accuracy and generalization, data augmentation techniques such as scaling, rotation, and noise addition are applied. The collected images are then resized, normalized, and formatted to ensure compatibility with deep learning models, enhancing the efficiency of feature extraction.

For model selection and implementation, the system employs YOLOv3 (You Only Look Once v3), a real- time object detection model known for its speed and accuracy. A lightweight Convolutional Neural Network (CNN) is used to extract spatial features, improving detection precision. To enhance real-time tracking, techniques such as Deep SORT and Kalman Filtering are implemented, allowing the system to track individual movements within a crowd. Additionally, optical flow analysis and Recurrent Neural Networks (RNNs) help predict movement direction and identify potential risks, such as overcrowding or sudden changes in crowd behavior.

After detection and tracking, the system performs post-processing for crowd counting, where it analyzes the detected individuals to estimate the total crowd size. The system also incorporates anomaly detection to flag irregular crowd behavior, such as unexpected gathering patterns or potential safety hazards. This ensures that security personnel receive timely alerts for any potential risks, enabling proactive intervention.

To assess system reliability, evaluation metrics such as precision, recall, and F1-score are used to measure the accuracy of crowd detection. Additionally, latency is tested to ensure the system operates in real time without significant delays. The model's performance is also compared with existing crowd monitoring solutions to validate its effectiveness and efficiency in diverse environments.

For real-world deployment, the trained model is integrated into IoT-enabled smart cameras, allowing on-site data processing. To enhance efficiency, optimization techniques like quantization and pruning are applied, reducing computational demands without compromising accuracy. Frameworks such as TensorRT and TensorFlow Lite (TFLite) further accelerate the model’s performance, ensuring smooth operation on edge devices.

Finally, the system is integrated with surveillance networks for continuous monitoring. A real-time alert mechanism notifies security personnel of any unusual crowd activity, ensuring a swift response to potential emergencies. Additionally, a dashboard visualizes key crowd metrics, such as density, movement patterns, and risk levels, providing authorities with a clear overview for better decision-making. The system undergoes real-world testing in large public gatherings to evaluate its scalability, usability, and reliability, ensuring a robust and effective crowd management solution.

OBJECTIVE

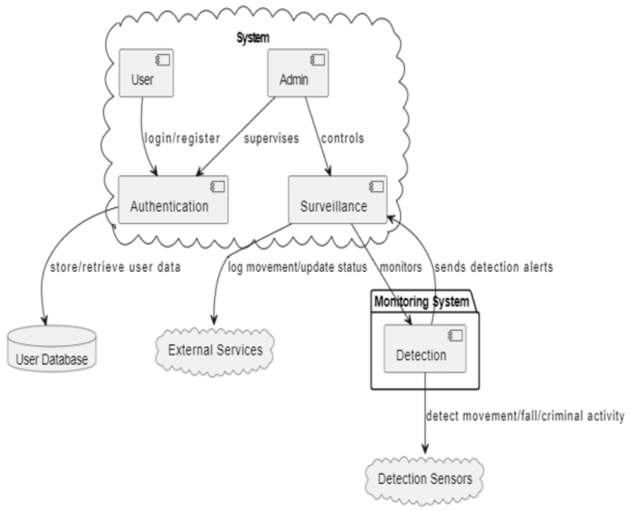
1. **Accurate Estimation of Crowd Size** The system aims to count the number of people in a specific area as accurately as possible. By using cameras and smart technology, it can analyze the crowd and estimate how many people are present at any given time. This helps in understanding how crowded a place is, whether it's an event, a public gathering, or a busy street.
2. **Better Planning and Management of Resources**
   Knowing the crowd size helps in distributing resources like security personnel, medical teams, and food supplies effectively. For example, if a concert or festival has more attendees than expected, organizers can quickly arrange for additional security and facilities to maintain order and ensure a smooth experience for everyone.
3. **Fire Detection for Safety** The system includes fire detection technology to identify potential fire hazards. If a fire is detected in a crowded area, the system can send an alert immediately. This allows authorities

to act fast, evacuate people, and prevent injuries or damage.

4. **Ensuring Safety in Large Gatherings** Large crowds can sometimes become dangerous, especially in enclosed spaces like stadiums or auditoriums. The system continuously monitors crowd density and movement to detect signs of overcrowding, stampedes, or other safety risks. This ensures that people remain safe and emergencies can be prevented before they happen.
5. **Quick and Effective Emergency Response** In case of an emergency, such as a fire, medical issue, or security threat, the system helps authorities respond faster. By tracking the crowd and identifying high-risk areas, emergency teams can be directed to the right locations without delay. This improves the chances of saving lives and reducing harm.

Overall, the system is designed to improve public safety, manage crowds efficiently, and provide a reliable way to handle emergencies in busy places.

SYATEM ARCHITECTURE



Fig(a): System Architecture

The diagram represents a smart crowd monitoring and surveillance system that works in a structured and secure way. It begins with two main types of users: the User and the Admin. The user can log in or register to access the system, while the admin is responsible for supervising and controlling the entire system. Before accessing any features, both must go through an authentication process, which ensures that only authorized individuals can use the system.

The authentication module plays an important role in maintaining security. It verifies login details and connects with the user database to store and retrieve user information. Once the authentication is successful, users and admins can interact with the system. This ensures that all activities are properly tracked and managed in a secure manner.

After login, the system moves to the surveillance stage. This part of the system continuously monitors activities using cameras or other devices. The admin has control over this section and can oversee how the system is functioning. It keeps track of movements and updates the status in real time, helping in continuous observation of the environment.

The monitoring system is the core component where intelligent processing takes place. It includes a detection module that analyzes the data received from the surveillance system. This module uses advanced techniques to identify different activities such as normal movement, falls, or any suspicious or criminal behavior. It plays a key role in understanding what is happening in the monitored area.

Detection sensors are used to collect real-time data from the environment. These sensors, such as cameras or motion detectors, capture information and send it to the monitoring system. Based on this data, the system performs analysis and makes decisions. If any unusual activity is detected, the system immediately generates alerts.

Finally, the system can send alerts to the admin or concerned authorities for quick action. It may also use external services like cloud platforms or APIs to improve performance and data handling. Overall, this system acts as a smart and automated solution for monitoring crowds, improving safety, and responding quickly to potential risks.

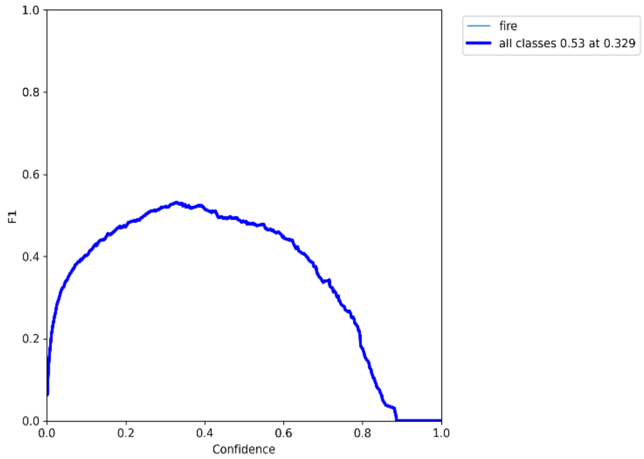
FUCTIONAL REQUIREMENTS

1. Real-Time Crowd Detection: The system must detect and count individuals in a crowd using a CNN model.
2. People Tracking: It should track the movement of each detected person across multiple frames.
3. Movement Estimation: The system should predict the direction and speed of crowd movement using algorithms like optical flow or RNNs.
4. Anomaly Detection: Identify sudden density changes or irregular movement patterns and flag them as potential risks.
5. Alerts and Notifications: Automatically send real-time alerts to security personnel when crowd density exceeds a safe threshold or anomalies are detected.
6. Data Visualization: Display live crowd metrics, movement trends, and risk zones through a user-friendly dashboard.
7. Data Logging: Store historical data for future analysis and performance evaluation.

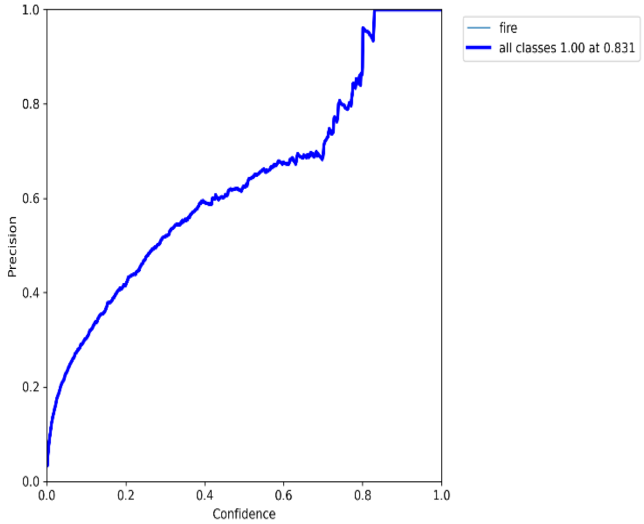
NON FUCTIONAL REQUIREMENTS

1. Real-Time Processing: Ensure low latency for quick detection, tracking, and alert generation.
2. Scalability: Support multiple cameras and large crowds without compromising performance.
3. Accuracy: Maintain high detection accuracy for crowd counting and movement estimation.
4. Efficiency: Optimize computational and memory demands to run smoothly on edge devices.
5. Security: Implement data encryption for secure transmission and prevent unauthorized access.
6. Reliability: Ensure continuous system operation with minimal downtime, especially during public events.
7. User-Friendly Interface: Provide a simple, intuitive dashboard for real-time monitoring.
8. Maintainability: Ensure the system is modular, allowing for easy updates and enhancements.
9. Energy Efficiency: Optimize power consumption to extend edge device battery life.

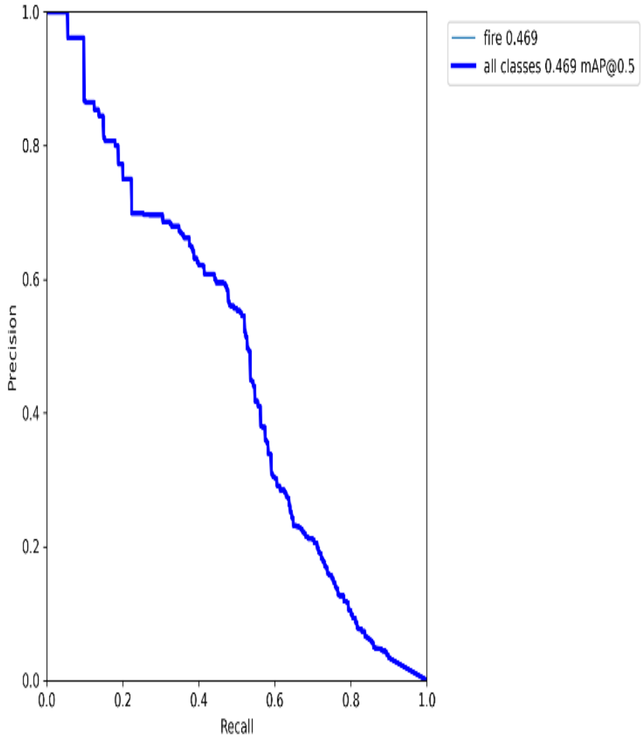
RESULT



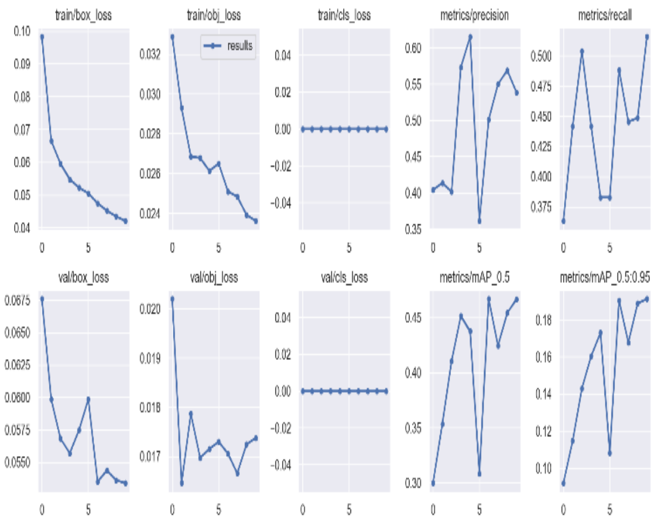
This graph shows how the F1 score (a balance between precision and recall) changes as the model's confidence threshold increases. The blue line represents "all classes" with an overall F1 of 0.53 at a confidence of 0.329. The light blue line (labeled "fire") seems to represent performance for a specific class called "fire".The curve starts low near 0 confidence, rises to a peak around 0.3–0.4 confidence, and then steadily drops as confidence increases further. This means the model gives its best balanced performance (highest F1) at a moderate confidence level around 0.33. After that, being too strict with confidence hurts the overall score.



This graph displays Precision on the y-axis against Confidence on the x-axis. The blue line (all classes) reaches a perfect precision of 1.00 at confidence 0.831.It starts near 0 when confidence is low and keeps climbing higher as the confidence threshold increases. This is expected behavior — when the model is only allowed to make predictions it is very sure about (high confidence), its precision improves a lot because it makes fewer wrong predictions. At very high confidence (~0.83), the precision becomes 100% for all classes combined.



This is a Precision-Recall curve, which is commonly used to evaluate object detection or classification models. The blue line shows "all classes" with an mAP (mean Average Precision) of 0.469 at IoU=0.5. The light blue "fire" line shows the performance for the fire class (0.469).The curve starts with high precision at low recall (left side) and gradually drops as recall increases. This tells us the trade-off: if you want the model to find more objects (higher recall), you have to accept more mistakes (lower precision). The area under this curve (mAP) being 0.469 means the overall performance is moderate.



This image is a collection of multiple small graphs showing how the model trained over epochs (probably 10 epochs here).The top-left graphs show training losses (box loss, obj loss, cls loss) steadily decreasing — this is good, meaning the model is learning.The bottom-left graphs show validation losses also decreasing or stabilizing.The right-side graphs show metrics like precision, recall, mAP@0.5, and mAP@0.5:0.95 improving over training time, though they fluctuate a bit.Overall, it looks like a successful training run where losses went down and performance metrics went up, which indicates the model is improving and not overfitting badly.

CONCLUSION

The proposed CNN-based crowd management system provides an efficient, real-time solution for monitoring and analyzing large public gatherings. By integrating deep learning techniques with real-time tracking and movement estimation, the system enhances safety, prevents overcrowding, and minimizes security risks. Optimized for edge devices, it ensures computational efficiency without compromising performance. The automated alert mechanism enables timely interventions, improving crowd control and event management. With high accuracy, scalability, and real-time responsiveness, this solution addresses the limitations of traditional surveillance systems, making it a valuable tool for smart city applications, public events, and emergency response planning. Future enhancements may include improved anomaly detection and integration with advanced AI-driven decision-making systems.

In simple terms, this project builds a smart and intelligent system that can monitor large crowds automatically using AI and deep learning. Instead of relying only on human security or basic CCTV cameras, this system can understand what is happening in real time. It can track how people are moving, identify crowded areas, and quickly detect if something unusual or risky is happening.

One of the biggest advantages of this system is safety. When too many people gather in one place, it can become dangerous and lead to accidents like stampedes. This system helps prevent such situations by sending alerts in advance, so authorities can take action on time. It also helps in managing big events like festivals, concerts, or public gatherings more efficiently.

Another important benefit is that the system is designed to work on edge devices, which means it does not require very powerful computers all the time. This makes it faster, cost-effective, and easier to use in real-world situations. It can process data quickly and provide instant results without delays.

The system is also scalable, which means it can be used in small areas as well as large smart city environments. It improves overall surveillance by reducing human effort and increasing accuracy. Compared to traditional systems, it is more reliable because it can continuously monitor and analyze situations without getting tired or missing important details.

In the future, this system can be made even smarter by adding advanced features like better detection of suspicious behavior, predicting possible risks before they happen, and integrating with other smart technologies. Overall, this project is a powerful step toward safer public spaces and smarter crowd management.

FUTURE SCOPE

In the future, this crowd management system can become much more advanced and intelligent with the help of new technologies. One important improvement is advanced behavior detection. The system can be trained to understand different types of human activities such as running, fighting, panic situations, or people moving in unusual patterns. This will help in identifying dangerous situations much earlier and reducing the chances of accidents like stampedes or riots.

Another major improvement is prediction and forecasting. By using past data and AI models, the system can predict crowd growth, peak times, and possible risk areas. For example, it can tell in advance if a place is likely to become overcrowded in the next few minutes or hours. This will help event organizers and authorities plan better and take preventive actions before problems occur.

The system can also be enhanced by integrating it with drones and aerial surveillance. Drones can capture live video from the top view, which gives a better understanding of crowd density and movement in large areas. This is especially useful in festivals, protests, or emergency situations where ground cameras may not cover everything.

In addition, the system can beconnected with IoT devices and sensors like smart cameras, motion sensors, and wearable devices. These sensors can provide more accurate and real-time data, improving the overall performance of the system. For example, sensors can detect temperature, noise levels, or sudden movements, which can indicate unusual crowd behavior.

Another future scope is the development of mobile applications and alert systems. Authorities can receive instant notifications on their phones about overcrowding or suspicious activities. In some cases, alerts can also be sent to the public to guide them to safer routes or less crowded areas.

The system can also be integrated with smart city infrastructure, such as traffic management systems and emergency response services like police, ambulance, and fire departments. This will help in better coordination during large events or emergencies, ensuring quick response and improved safety.

Moreover, future improvements can focus on fully automated decision-making systems. With advanced AI, the system can not only detect problems but also suggest or take actions automatically, such as redirecting people, controlling entry points, or activating alarms.

Finally, improvements in accuracy,speed, and scalability will make the system more reliable and usable in different environments, from small indoor spaces to large outdoor events and entire smart cities.

Overall, the future scope of this project is very wide. With continuous advancements in AI and technology, this system can become a complete smart solution for managing crowds safely, efficiently, and intelligently in real-world situations.

REFERENCES:

1. F. Al-Shargie, T. B. Tang, and N. Nordin, “Mental stress assessment using simultaneous EEG and fNIRS,” *IEEE Access*, vol. 6, pp. 24096–24106, 2018.
2. H. Lu, S. Li, D. Zhu, and X. Wang, “Stress detection using speech features and machine learning,” *IEEE Transactions on Affective Computing*, vol. 12, no. 3, pp. 670–682, 2021.
3. B. Ivanova and K. Shoylekova, “Trends and Challenges in Surveillance – A Systematic Review of Camera Systems Implementing Artificial Intelligence,” 2024.
4. C. C. Corrigan and C. Lütge, “AI-powered public surveillance systems: why we (might) need them and how we want them,” 2022.
5. R. Amini and Z. Zilic, “Systematic Review of IoT-Based Solutions for User Tracking: Towards Smarter Lifestyle, Wellness and Health Management,” 2024.
6. H. Sharma and N. Kanwal, “Video surveillance in smart cities: current status, challenges & future directions,” 2024.
7. Jannat, A. Ilyas, and T. Saeed, “Exploration of Solutions for Smart Cities: Challenges in Privacy and Security,” 2021.
8. S. Ebadinezhad, “The Role of IoT in Enhancing Public Safety in Smart Cities,” 2024.
9. Y. G. D. Thomas and K. Jacob, “Overview of autonomous surveillance technologies and their potential applications in urban settings,” 2023.
10. W. Liu, M. Anguelov, D. Erhan, et al., “SSD: Single Shot MultiBox Detector,” *ECCV*, 2016.

11. J. Redmon and A. Farhadi, “YOLOv3: An Incremental Improvement,” 2018.
12. K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” *CVPR*, 2016.
13. Krizhevsky, I. Sutskever, and G. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” *NIPS*, 2012.
14. S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards Real-Time Object Detection,” *IEEE TPAMI*, 2017.
15. Z. Zhang, “Deep learning-based crowd density estimation: A survey,” *Pattern Recognition Letters*, 2020.
16. L. Zhang, M. Li, and X. Wang, “Crowd counting via multi-column CNN,” *CVPR*, 2016.
17. V. A. Sindagi and V. M. Patel, “A survey of recent advances in CNN-based crowd counting,” *Signal Processing*, 2018.
18. D. Helbing and A. Johansson, “Pedestrian, crowd and evacuation dynamics,” *Encyclopedia of Complexity and Systems Science*, 2017.
19. G. Olmschenk, “Real-time crowd analysis using deep learning,” 2021.
20. S. S. Channappayya, “Smart surveillance systems using AI and IoT,” *IEEE IoT Journal*, 2022.
21. X. Wang, “Intelligent multi-camera video surveillance: A review,” *Pattern Recognition Letters*, 2013.
22. M. Valera and S. Velastin, “Intelligent distributed surveillance systems: A review,” *IEE Proceedings*, 2005.
23. Hampapur, L. Brown, J. Connell, et al., “Smart video surveillance: Exploring the concept,” *IEEE Computer*, 2005.
24. N. Dalal and B. Triggs, “Histograms of Oriented Gradients for Human Detection,” *CVPR*, 2005.
25. P. Dollar, C. Wojek, B. Schiele, and P. Perona, “Pedestrian detection: A benchmark,” *CVPR*, 2009.
26. M. Khan, M. Ullah, and S. Ahmad, “Deep learning-based crowd behavior analysis,” 2020.
27. S. Ali and M. Shah, “A Lagrangian particle dynamics approach for crowd flow segmentation,” *CVPR*, 2007.
28. T. Bouwmans, “Traditional and recent approaches in background modeling for foreground detection,” *Computer Science Review*, 2014.
29. Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, 2015.
30. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press, 2016.