

Intelligent People Counting for Crowd Management during Ziyarat Al-Arbaeen in Karbala

Asst.Lect.Sinan Sameer Mahmood Al-Dabbagh

University of Baghdad / College of Science, Department of Remote
Sensing and GIS / Lab of Seismology and Spatial Data Analysis,
senan.samir@sc.uobaghdad.edu.iq

Asst.Lect. Raghda A. Ali

University of Baghdad / College of Science, Department of Computer
Science

raghda.Ali@sc.uobaghdad.edu.iq Hala Abdulsalam Jasim :

Asst.Lect. Hala Abdulsalam Jasim

University of Baghdad / College of Science, Department of Remote
Sensing and GIS / Lab of Seismology and Spatial Data Analysis,
hala.abd@sc.uobaghdad.edu.iq

Abstract

The ziyarat Al-Arbaeen in Karbala, Iraq, is one of the largest annual gatherings of people in the world, attracting millions of visitors. This poses significant challenges in crowd control, ensuring smooth flow of traffic, and ensuring the safety of visitors. This research aims to develop an intelligent people counting system using advanced computer vision techniques, relying on convolutional neural networks (CNNs), to accurately and effectively detect individuals in video clips, without the use of facial recognition or identification techniques. The system analyzes the input videos to calculate the total number of people appearing in each frame. The results are displayed as a representation of the total number, allowing for a more accurate understanding of the crowd distribution in the studied locations. The system currently does not include real-time data processing, but the possibility of developing this aspect and adding automatic alerts when crowd density increases is an important future endeavor that could enhance the system's ability to effectively manage crowds. This research represents a first step towards employing artificial intelligence techniques to support crowd control during major religious events and highlights the potential of computer vision as an effective tool to help ensure the safety and security of visitors.

Keywords: Crowd Management, People Counting, Ziyarat Al-Arbaeen, Artificial Intelligence, YOLOv5.

Introduction

Karbala is a holy city in Iraq, which experiences a special religious phenomenon once a year; ziyarat Al-Arbaeen. This occasion is the most important event that is spiritually significant to the Muslims because millions remember the killing of Imam Hussein (peace be upon him). It is ziyarat Al-Arbaeeni the biggest nonviolent human gathering in the globe and its attendance is over twenty million visitors of all over Iraq and the world who observe the occasion by walking, one hundreds of kilometers to the gathering place and more (Kadhim et al., 2025).

However, sheer presence of people in such a large number, notwithstanding its peaceful model and sheer popular organization that this kind of influx already has, presents enormous organizational problems concerning crowd safety and control of the crowd movement, at key points, like the entrances into the holy shrines, the roads leading to Karbala, and service stations (Lin et al., 2024; Collini et al., 2024; Ahmed et al., 2021). It has been demonstrated that the current systems that use manual inspection or physical counting of tourist numbers are no longer effective enough to keep up with the numbers that keep rising at the rate in which they are implied (Elsisi et al., 2021). The necessity to find urgent solutions in the form of smart solutions that will be able to deliver precise and real-time information needed to facilitate well-informed decisions on the management of a crowd, direction, and onward movement (Shambour & Gutub, 2022; Wen et al., 2021). In this regard, such technologies as artificial intelligence (AI) and in particular computer vision algorithms using convolutional neural networks (CNNs) can be regarded as one of the most promising activities in the field (Lin et al., 2022; Pervaiz et al., 2022). These technologies allow the computerized analysis of images and video snippets, the detection of visual trends, and the extraction of useful information being in the captured scene (Fan et al., 2022; Pervaiz, Jalal, & Kim, 2021). Among the most significant applications which may be

used to serve the ziyarat Al-Arbaeen, an automated counting of people with the help of video analysis can be mentioned, and it is the task of this research.

With the growing challenges created by the millions of visitors arriving during the ziyarat Al-Arbaeen, there is a necessity of the development of smart tools able to efficiently comprehend the number and dispersion of these masses in order to emphasize the utilization of spatial analysis and geographic information systems (GIS) methodologies in order to trace the movement of the visitors and propose dynamic pathways to ease the congestions, which will provide rich spatial information that can be used in counting the visitors (Jasim et al., 2025). In this work, a prototype of the intelligent tool is developed aiming towards counting the number of people in a video the system is not based on tracking and facial recognition and does not provide snapshots of individuals. As opposed, it processes every frame of the video and counts the number of people in it, presenting the number in an end picture along with nothing more but the number of people. It is in use today on pre-input videos; although, in the future, it can be extended onto the instant playback. The study will undertake a number of objectives that include:

- Development of intelligent and efficient system to count the people in crowded areas accurately and without much technicality involved.
- Delivering correct quantitative information that will serve to comprehend the dynamics of the crowd that gather as part of the ziyarat Al-Arbaeen.
- Providing the regulatory and field authorities with intelligent AI-based solutions that enhance the strategies of managing crowd.
- Paving the way towards the emerging real-time systems which might in the future be incorporated with field cameras to give early warning as and when required.

Accordingly, the research can be taken as more than just another technical project; it is an actual contribution to the application of AI to the benefit of society in sensitive, real-world, religious, and cultural scenarios, and yet it is simple to implement and easy to field.

Related Work

The methods to analyze the population density and crowd counting in crowded places have experienced great transformation because of the improvement in computer vision and deep learning. Some recent works have already proved the efficiency of such methods in enhancing the accuracy of monitoring and analysis, particularly in the complicated environment, like trains, religious crowds, and smart cities. The most noteworthy works referring to this topic and considering its approach to different technical and applied aspects are reviewed as follows (Gong et al., 2021). First, (Jin et al., 2024) provided the incorporation of convolutional neural networks (CNNs) along with YOLO algorithms to the people counting and detecting the population density activities in a detailed manner. The paper has argued why optimization and regularization algorithms are relevant in boosting model performance, as well as generalizing the model in averting overfitting, making them more applicable in other situations of density and crowding. The paper has shown that a combination of these methods helps not only the process of training to go faster but also make it more stable in such an environment as a public square and transport hub.

In contrast to this, the work of (Gao et al., 2024) was aimed at giving a holistic summary of methods in crowd counting in IoT settings with emphasis on the usefulness of using WiFi data and smart sensors to improve result accuracy in the estimate. The criteria to influence the crowd counting variables were categorized into three broad categories, namely, object detection, behavior monitoring and individual counting. Among the main challenges identified in the study, it was noted that dealing with many types of human activities, future crowding forecasting, and combining many machine learning methods under distributed settings were the most significant ones. A practical example of religious crowd control concern is the study of (Halboob et al., 2024), which devised a smart framework in managing crowds in

the case of the Umrah season. The issues that this framework tries to tackle include, counting, centering, density estimation, as well as behavior monitoring. A rules engine was introduced to identify abnormal behavior, and the findings of this revealed that the system could handle greater volumes of pilgrims without compromising the accuracy. The paper advised that in future, the rules engine could be developed with deep learning.

Regarding the smart cities, (Qaraqe et al., 2024) presented an integrated system titled as the PublicVision that is supposed to analyse the behaviour of crowds in real-time depending on the degree of violence and the number of people. The system has a sophisticated model of analyzing video based on Swin Transformer and employs VPN and IPSec as a method of providing secure data communication. Real-time data processing with the help of the NVIDIA Deep Stream SDK also became a possibility, creating an opportunity to use this technology in the future in the process of controlling crowds during some major event or protests.

The work by (Jin et al., 2024) brought to the fore the deployment of CNN and YOLO methods in the crowd counting and crowd detection tasks, especially the optimization and constraining tools, which made it easier to deal with the variety and imbalanced nature of data. In the analysis, the solutions proved the methods to be more viable in accommodating cases of crowding, altering lighting, and reflectivity.

In their article that concentrated on the accuracy of people detection and counting using the YOLOv8 algorithm in still images and moving videos and confronted the issues of overlap and the resulting change of shape and size due to change of angle and perspective, (Shyaa & Hashim, 2024) aimed at improving the effectiveness of people detection and counting in moving videos and still images. The results of the study have revealed that YOLOv8 garnered an excellent level of recognition and counting accuracy of up to 100%, opening out wide opportunities of its use

in security surveillance, crowd management, and people flow analysis in events. YOLOv8 has also been suggested to work together with intelligent systems like the Internet of Things by enhancing security and studying the flow of people day in and day out.

The article by (Siva et al., 2025) discusses the realization of the intelligent surveillance system utilizing the YOLOv8 algorithm to provide proper, real-time identification of people and threats in a sequence of live videos. The system incorporates a mixing of artificial intelligence together with a deep learning algorithm, to assess the crowd together with identifying any anomalies, as well as decrease false alarms and optimize performance across low-power equipment. The results showed a detection precision of up to 95.4 percent, an anomaly recognition precision of 92.7 percent and thus the system could be used in crowd control, traffic management and law enforcement and can be extended into smart infrastructure in the future.

Finally, in their work (V. M. R. et al., 2025) proposed the use of a YOLO-based architecture being used to detect and count the passengers in the metro cars. Although both speed and accuracy was high, the research demonstrated problems that may be encountered in crowded environments including visual obscuration, varying light, and movement blur in detecting passengers. The paper suggested enhancement of performance by turning to newer versions of YOLO, namely, YOLOv5 and YOLOv8, as well as implementing compensatory methods that would allow enhancing stability and accuracy, including visual simulation and multi-frame tracking.

Methodology

This paper describes the development of a video-processing intelligent system based on the modern deep learning pipeline, the YOLOv5 model. The system relies on the artificial intelligence tools to interpret visual images and describe their semantics. AI is a broad-based specialty whose main aim is to equip computers with the ability to act in a manner that is closely similar to the analysis and judgement of human beings. Deep learning is a branch of AI and therefore uses neural networks as a part of teaching machines using data. YOLOv5 is a fine-tuned deep learning platform specifically designed to execute image interpretation quickly and productively. Its prominent benefit is the fact that it is able to process an entire image without having to go through the process of segmentation or re-entry of images. To achieve this, the model divides the input into a conventional grid of small squares and examines cells individually, to determine whether a specific object exists or not. Upon the identification of such an object, the system will also predict its categorization and outline its location through a rectangle box. Redundant detections are then suppressed and overlap is eliminated through the process of non-maximum suppression resulting only in predictions which have the highest confidence.

1. Stages of the Proposed System:

The system consists of Several stages. The first stage is video input using a digital camera or a pre-recorded video to explore a specific area. Then, a frame extraction step is needed so that the video will be divided into a series of frames using the OpenCV library, with each frame analyzed separately. After framing the video, YOLOv5 will be used to detect objects in the viewing area within a square. Each object will be classified into one of the types (Cars, Animals, Human, etc.), and then counting the

number of detected squares representing people in each frame, as a result, the number of recorded people will be displayed for each image. The flowchart of the presented system is illustrated in Fig.1.

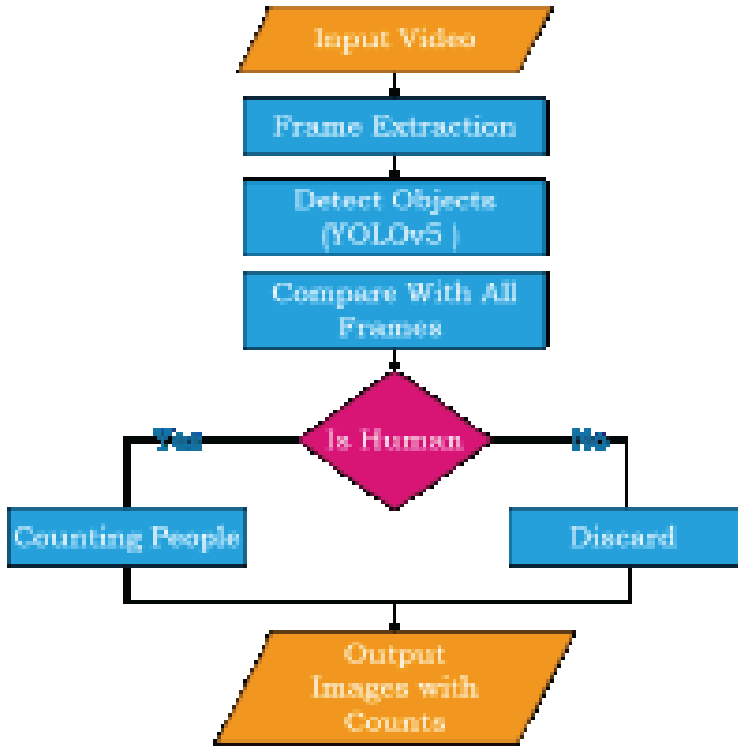


Figure 1: System Layout

2.Programming and Implementation Environment:

The system was developed and implemented using Python 3.9, due to its robust libraries and extensive support community, which contributes to accelerating development and integration with AI models. A set of libraries and software tools was selected to meet the requirements for video processing and running deep learning models efficiently. The libraries used in this work.

- OpenCV: To process video files and extract frames from them.
- PyTorch: To run and customize the pre-trained YOLOv5 model.
- NumPy and Matplotlib: To support numerical operations and visualize results.

The environment was chosen to provide a balance between performance and available resources a Laptop with an Intel Core i7 processor and 16 GB RAM, give the ability to execute the system efficiently in real time without the need for advanced hardware or dedicated graphics cards (GPU), reflecting its practical applicability in low-cost systems or when moving to real-world operating environments.

3.The YOLOv5 Model:

The YOLOv5 model is one of the latest and fastest computer vision models for object detection in images and videos. This model is characterized by its high speed and good accuracy in identifying objects, making it ideally suited for real-time applications such as counting people in crowded environments.

4.YOLOv5 Working Principle:

YOLOv5 is based on the basic idea of processing the entire image at once (One-Pass Detection), rather than segmenting the image or processing it in separate stages. The model divides the image into a grid and estimates the probability of an object being present within each grid, specifying the object's class (e.g., person, car, etc.) and its precise location within the image (Yue & Zhang, 2025).

The model is built from three main components:

1. Backbone: Used to extract basic features from the image, such as edges and general shapes. In YOLOv5, a CSPNet is used as the backbone to improve efficiency and reduce computational cost.
2. Neck: Used to combine features extracted from multiple levels, allowing the model to understand the overall context of objects in the image, whether small or large. FPN or PANet are typically used at this stage.
3. Head: Performs the actual prediction of bounding boxes and object classes. In our case, the model is trained to recognize only the “person” class.

5.Integration of YOLOv5 into the Proposed System Architecture:

A pre-trained YOLOv5 model was used with minor modifications to determine the number of people in each video frame captured by fixed surveillance cameras and drones. Rather than retraining the model from scratch, we relied on existing models available in the PyTorch library, leveraging its streamlined Python interface. The process begins by extracting frames from video clips using the OpenCV library, where each clip is converted into a series of still images. The images are then processed and passed to the YOLOv5 model to detect objects within them. The results are subsequently filtered to extract only objects classified as people, and their count is calculated based on the number of bounding boxes matching this category. The final result is displayed on the image, showing the number of detected people without showing the bounding boxes. The resulting images are then saved with the required data for documentation and analysis purposes. Fig. 2 illustrates the YOLOv5 structure and how each stage works:

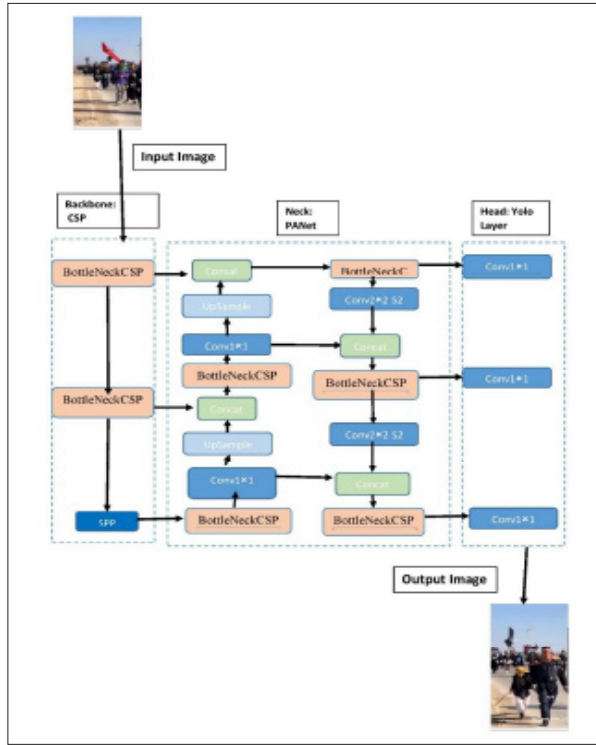


Figure 2: The architecture of the YOLOv5 model used for people detection and counting.

Results and Evaluation

In this section, the proposed people counting system that is based on the YOLOv5 will be tested on the video records of the fixed camera along the visitor pathway and the drone. Testing the accuracy of the system against the ziyarat Al- Arbaeen is done by comparing automatic counts to manual counts of different environmental conditions and camera angles.

1.Experimental Data Preparation:

The suggested system of counting people with the help of the YOLOv5 model was evaluated in the practical context, which simulated the situation found at the ziyarat Al-Arbaeen. Footages were captured at two different elevations to achieve a varying view of different population densities and angles.

1. Fixed camera: A Static camera was used to record high-definition video along a pilgrim's route to the shrine of Imam Hussein (peace be upon him) in Karbala. The camera has been set in a way that prevents people from distorting their view of visitors moving along the route.
2. Drone Camera: Videos filmed by a drone with a high-quality camera gave an aerial view and panorama of the way crowds moved through surrounding squares and streets, which enabled the study of spatial distribution and movement change in its flow.

2.Evaluation Mechanism:

To evaluate the system's accuracy in detecting and counting people, two data sets was used (video by fixed camera with a length of 5 seconds and Drone camera video with a length of 2 seconds. This simplifies the evaluation process and focuses on detection quality. The two videos, framed into images to test the model's performance under various shooting conditions. The system processes each image independently using a pre-trained YOLOv5 model, focusing exclusively on the "people" class. After the detection process is performed, the number of detected entities from this class is counted. To maintain simplicity and reduce visual noise in the resulting images, a direct display of the total number of people in the corner of the image was adopted, without showing bounding boxes around the individuals. This approach also speeds up the processing process and provides clearer output for the user, especially in cases where tracking or precise positioning of individuals is not required.

3.Experimental Results:

In Fig. 3, a frame taken from a video captured by a fixed-angle surveillance camera was selected, showing a group of people in a scene characterized by moderate density and a variety of body and clothing poses.

The system was able to detect 20 people within the frame, with overall performance considered good given the nature of the scene. The model demonstrated an effective ability to handle the horizontal angle without significantly impacting detection accuracy. It maintained a high level of counting accuracy, despite some overlap between individuals in the foreground. Homogeneous and consistent lighting helped improve the model's performance and reduced the potential for errors caused by shadows or high contrast. The weaknesses of the model included the difficulty distinguishing people overlapping in the far-away background, as the objects appeared too close together, causing some clusters to be counted as a single entity. This resulted in a slight decrease in overall counting accuracy, a known challenge in high-density scenes or those shot from narrow depth angles.



Figure 3: People counting using a fixed camera

Fig.4 shows A frame taken from a drone video, which shows a large crowd of people. The system detected 42 people in a single frame; the provided result reflects the model’s ability to handle large-scale scenes. The top-down perspective enabled the system to capture many individuals at once, without the need for multi-frame temporal analysis. Although there were some difficulties in detection in areas where crowds appeared closely spaced or overlapping, especially when objects were small due to distance, the system provided relatively accurate results that reflected the actual size of the crowd. This case demonstrates that the model maintains its effectiveness even under aerial conditions, making it suitable for applications such as aerial crowd monitoring and analyzing crowd density at public events or open spaces, without the need for complex tracking tools or additional equipment.



Figure 4: People counting using a drone-mounted camera

Performance Analysis:

The counts generated by the system were compared with those obtained manually by human experts, and the differences between the two counts were within $\pm 5\%$ in most cases, which is acceptable for real-world applications, especially in crowded environments, as shown in Table 4.1.

Table 4.1. Comparison of manual and system-based passenger counting

Frame Angle	Manual Count	System Count	Error Rate
Horizontal (Stabilized)	23	20	-13.0%
Overhead (Drone)	47	42	-10.6%

The automatic counting system offers several advantages that distinguish it from other conventional systems. It offers high accuracy in counting people, despite a slight negative error rate indicating an underestimation of the actual number. However, these rates remain within acceptable limits for practical use. The system is highly flexible in handling various shooting conditions, such as fixed horizontal shooting or overhead shooting using drones. This gives it a superior ability to adapt to a variety of scenes, something that is difficult to achieve with other systems that may be limited to a specific type of shooting. The system is also capable of processing large amounts of data quickly and efficiently, making it ideal for handling scenes that contain large numbers of people at once. This performance is particularly evident in high-angle angles, which require high precision and processing speed. The system relies on artificial intelligence technologies and continuous learning and training systems, allowing it to continuously improve its accuracy by learning new data and various challenges, such as overlapping objects or different sizes of people in images. These technologies also allow for changing lighting conditions and capturing images in different environments, enhancing system reliability and reducing the need for time-consuming and labor-intensive manual counting.

The use of AI saves human resources time and effort and helps obtain accurate and rapid results, enabling informed and immediate decisions. This is particularly important when managing large and complex crowds. The Arbaeen pilgrimage, one of the largest humanitarian gatherings, highlights the importance of this system is highlighted in its high-precision monitoring of visitor numbers, facilitating better planning of security, health, and logistical services. It also helps improve crowd flow management and reduce the risk of overcrowding in vital areas, ensuring visitors' safety. Furthermore, the system enables relevant authorities to take swift and effective action based on real-time data and contributes to future planning for the pilgrimage through data analysis across years, enhancing organizational sustainability and improving the overall visitor experience.

Conclusion

The study reflects an initial attempt to design an intelligent people counter counting system implemented by computer vision to guide the administration of large number of people in ziyarat Al- Arbaeen in Karbala, Iraq. The suggested device is based on the YOLOv5 model which is already trained to recognize people in the video shots recorded by stationary cameras and drones. The videos are then transferred to convert into the number of the total number of individuals in each frame without tracking the objects and displaying frames around those objects and no automatic alerts or warnings are also issued. The system tested footage recorded by both fixed cameras and drones and the output was presented in the form of images that provide only the numbers of the total people. These findings indicated adequate accuracy, as far as the given application is concerned, and revealed the potential of this kind of system as that which can offer preliminary assistance to decision-makers as large-scale events are being organized.

Even though simple in its present form, the system has tremendous prospects of growth and future expansion. Among the most significant potential future directions, there is a need to work on realizing the system functioning in a real-time so that it was possible to monitor the number and distribution of the crowds in real-time. It involves learning technical items related to the comprehension of anticipated frame rate, response time, and hardware requirements, including the deployment of powerful graphics processing units (GPUs), to supply the clean-cut performance of complex settings in real-world, crowded conditions. Moreover, it is also possible to incorporate a smart alarm system that would automatically notify people when the crowd density gets higher than the level that is unsafe, which is one of the essential steps towards making the crowd management process much safer and more efficient. Therefore, the offered system could be an easy and efficient method to control the number of visitors in difficult conditions and open further the method of applying AI technologies to the organization and protection of big religious and humanitarian events, which goes along with real-time concerns to make it be widely applicable.

References

1. Kadhim, H. T., Jaber, M. S., Salih, S. K., Hussein, H. A., & Abdalrazak, O. A. (2025, March). An intelligent automated traffic system for crowded Arbaeen's pilgrimage. *AL ARBA'IN*, 2(2).
2. Lin, H., Ma, Z., Hong, X., Shangguan, Q., & Meng, D. (2024). Gramformer: Learning crowd counting via graph-modulated transformer. *arXiv preprint arXiv:2401.03870*. <https://arxiv.org/abs/2401.03870>
3. Collini, E., Palesi, L. A. I., Nesi, P., et al. (2024, February). Flexible thermal camera solution for smart city people detection and counting. *Multimedia Tools and Applications*, 83, 20457–20485. <https://doi.org/10.1007/s11042-023-16374-x>
4. Ahmed, S., Hossain, M. F., Kaiser, M. S., Noor, M. B. T., Mahmud, M., & Chakraborty, C. (2021). Artificial intelligence and machine learning for ensuring security in smart cities. In C. Chakraborty, J. C.-W. Lin, & M. Alazab (Eds.), *Data-driven mining, learning and analytics for secured smart cities* (pp. xx–xx). Springer. https://doi.org/10.1007/978-3-030-72139-8_2
5. Elsis, M., Tran, M.-Q., Mahmoud, K., Lehtonen, M., & Darwish, M. M. F. (2021). Deep learning-based Industry 4.0 and Internet of Things towards effective energy management for smart buildings. *Sensors*, 21(4), 1038. <https://doi.org/10.3390/s21041038>
6. Shambour, M. K., & Gutub, A. (2022). Progress of IoT research technologies and applications serving Hajj and Umrah. *Arabian Journal for Science and Engineering*, 47, 1253–1273. <https://doi.org/10.1007/s13369-021-05838-7>
7. Wen, L., et al. (2021, May 6). Detection, tracking, and counting meets drones in crowds: A benchmark. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. <https://doi.org/10.48550/arXiv.2105.02440>
8. Lin, H., Ma, Z., Ji, R., Wang, Y., & Hong, X. (2022). Boosting crowd counting via multifaceted attention. *arXiv preprint arXiv:2203.02636*. <https://arxiv.org/abs/2203.02636>

9. Pervaiz, M., Shorfuzzaman, M., Alsufyani, A., Jalal, A., Alsuhibany, S. A., & Park, J. (2022). Tracking and analysis of pedestrian's behavior in public places. *Computers, Materials and Continua*, 74(1), 841–853. <https://doi.org/10.32604/cmc.2023.029629>
10. Fan, Z., Zhang, H., Zhang, Z., Lu, G., Zhang, Y., & Wang, Y. (2022). A survey of crowd counting and density estimation based on convolutional neural networks. *Neurocomputing*, 472, 224–251. <https://doi.org/10.1016/j.neucom.2021.02.103>
11. Pervaiz, M., Jalal, A., & Kim, K. (2021). Hybrid algorithm for multi-people counting and tracking for smart surveillance. In *Proceedings of the 2021 International Bhurban Conference on Applied Sciences and Technologies (IBCAST)*. <https://doi.org/10.1109/IBCAST51254.2021.9393171>
12. Jasim, H. A., Abd-almajied, M. I., Alatta, H. J., & George, L. E. (2025, March). Analyzing and tracking the data of the millions sized gatherings for the Arba'in visit and proposing alternative ways to relieve congestion using spatial analysis algorithms. *Semi-Annual Scientific Journal*, 2(2, pt. 7).
13. Gong, V. X., Daamen, W., Bozzon, A., et al. (2021, December). Counting people in the crowd using social media images for crowd management in city events. *Transportation*, 48, 3085–3119. <https://doi.org/10.1007/s11116-020-10159-z>
14. Jin, F., Zhang, Z., Ning, Y., Lu, Y., Song, W., Qin, X., & Chen, J. (2024). Crowd counting and people density detection: An overview. In *Proceedings of the 3rd International Conference on Engineering Management and Information Science (EMIS 2024)* (pp. 434–454). https://doi.org/10.2991/978-94-6463-447-1_46
15. Gao, M., Souri, A., Zaker, M., et al. (2024, February). A comprehensive analysis for crowd counting methodologies and algorithms in Internet of Things. *Cluster Computing*, 27, 859–873. <https://doi.org/10.1007/s10586-023-03987-y>

16. Halboob, W., Altaheri, H., Derhab, A., & Almuhtadi, J. (2024). Crowd management intelligence framework: Umrah use case. *IEEE Access*, 12, 6752–6767. <https://doi.org/10.1109/ACCESS.2024.3350188>
17. Qaraqe, M., Elzein, A., Basaran, E., Yang, Y., Varghese, E. B., Costandi, W., Rizk, J., & Alam, N. (2024). PublicVision: A secure smart surveillance system for crowd behavior recognition. *IEEE Access*, 12, 26474–26491. <https://doi.org/10.1109/ACCESS.2024.3366693>
18. Jin, F., Zhang, Z., Ning, Y., Lu, Y., Song, W., Qin, X., & Chen, J. (2024). Crowd counting and people density detection: An overview. In T. Yao et al. (Eds.), *Proceedings of the 3rd International Conference on Engineering Management and Information Science (EMIS 2024)* (Vol. 111, pp. 434–454). *Advances in Computer Science Research*. https://doi.org/10.2991/978-94-6463-447-1_46
19. Shyaa, T. A. R., & Hashim, A. A. (2024). Enhancing real human detection and people counting using YOLOv8. *BIO Web of Conferences*, 97, Article 00061. <https://doi.org/10.1051/bioconf/20249700061>
20. Siva, P., Pujitha, G. B., Krishna, G. S., Hemanth, G., & Sai Teja, B. M. (2025). Smart surveillance systems using YOLOv8: A scalable approach for crowd and threat detection. *International Journal of Recent Advances in Engineering and Technology*, 14(1), 51–62. <https://journals.mriindia.com/index.php/ijraet/article/view/178>
21. V. M. R., S. S. M., R. S. Y., R. T. M., & V. S. M. (2025, June). Real-time passenger counting and occupancy-based announcement system for metro trains using YOLO-based deep learning. *Journal of Machine Learning and Signal Processing*, 1(1), 1–11.
22. Yue, Y., & Zhang, W. (2025). Detection and counting model of soybean at the flowering and podding stage in the field based on improved YOLOv5. *Agriculture*, 15(5), 528. <https://doi.org/10.3390/agriculture15050528>