International Journal of Computational Science and Engineering Research
ISSN: 3107 - 8605 (Online) , http://www.ijcser.com/
Regular Issue , Vol. 3, Issue. 1 (January – March) , 2026 , Pages: 44 - 51
Received: 12 January 2025 ; *Accepted:* 29 December 2025 ; *Published:* 19 January 2026
https://doi.org/10.63328/IJCSER-V3RI1P5
Original Paper : Applied Research – Experimental Research

Real-Time Stampede Risk Prediction from Crowd Videos Using YOLOv8 and Spatio-Temporal Modeling

Kuppireddy Krishna Reddy [1], T Ramya [2] , Yagireddi Ramesh [3] , J V G Prakasa Rao Pyla [4]

[1] Department of Computer Science and Engineering , Mother Theresa Institute of Engineering and Technology, Palamaner - 517408, Chittoor District, Andhra Pradesh ; krishnareddy206@mtieat@org
[2] Department of Computer Science and Engineering , Mohan Babu University, Tirupati, Andhra Pradesh , India ramyathenepalli@gmail.com
[3] Department of Computer Science and Engineering, Aditya Institute of Technology and Management , Andhra Pradesh , India ; rameshyegireddi@gmail.com
[4] Department of Computer Science and Engineering , Lendi Institute of Engineering and Technology , Vizianagaram Andhra Pradesh , India ; prakash.pyla@gmail.com

** Corresponding Author: Kuppireddy Krishna Reddy; krishnareddy206@mtieat@org*

Abstract: Stampedes pose a major risk to people's lives in densely populated settings, such as public gathering places, transit hubs, and places of worship. In order to avoid stampedes and fatalities, it is crucial to promptly recognize unstable crowds. In this paper, we offer a computer vision system that uses YOLOv8-based person detection to analyze crowd behavior and estimate the probability of a stampede. To ascertain how a crowd is acting collectively, the system makes use of data on time, motion patterns, and crowd density. Furthermore, we employed a spatial zone-based analytic technique to pinpoint regions of the scene with the highest potential risk of a stampede. Finally, we have included a way to build a temporal persistence on the population in order to reduce the incidence of false alarms and stabilize risk estimations over time. Through a real-time web-based interface, our technology generates visual analytics output and multi-level risk alerts. The suggested method can distinguish between crowds that are behaving normally, at the warning stage, and in a high-risk state, according to experimental results acquired from recorded videos of crowds. This creates a useful decision support tool for crowd monitoring applications.

Keywords: Computer Vision, Crowd Density, Motion Analysis, YOLOv8, Stampede Risk Detection, Crowd Analysis.

1. Introduction

Public safety professionals are very interested in the analysis of crowds during big public events (such as concerts, festivals, athletic events, and religious gatherings). These events frequently draw large numbers of people in cramped spaces, which can lead to uncontrollable crowd dynamics and potentially fatal incidents like stampedes. In the past, the majority of stampede incidents were not produced by a single trigger, but rather by a sequence of increasingly massive population rises followed by an abrupt shift in the direction of everyone's movement. The main method of crowd monitoring is usually manual surveillance. However, due to human reaction time limitations, weariness, and inattention, manual methods are ineffective. Continuous crowd condition monitoring is made possible by automated vision-based crowd monitoring systems, which may be able to warn of potentially hazardous situations before they materialize. However, problems like occlusion (i.e., people obstructing other people's view), varying crowd densities, and the difficulty of interpreting similar movements (collective behaviour) based only on visual data make it difficult to design automated systems for crowd monitoring.

People may now be detected in high-density or congested environments with substantially more reliability because to recent developments in Deep Learning-based Object Detection. Stampede danger involves a behavioral reaction (i.e., the conversion of physical space into motion) rather than just the quantity of people in a given area, yet identifying persons is an essential first step in preventing stampedes. Therefore, in order to design efficient early-warning systems for stampede occurrences, the features of

 © 2026, IJCSER, All Rights Reserved

ISSN: 3107 - 8605 (Online) , http://www.ijcser.com/ , Vol. 3, Issue 1, 2026, https://doi.org/10.63328/IJCSER-V3RI1P5

crowds must take into account not only their total number but also their spatial distribution, movement, and evolution over time. The method for evaluating stampede risk based on computer vision and crowd behaviour analysis is presented in this study. To enable early detection of stampede risk in crowds, the suggested approach integrates Real-Time Person Detection via YOLOv8, Crowd Density Estimation, Motion Analysis, Temporal Persistence, and Zone-Based Spatial Assessment. In the end, the findings of this study will offer a framework for proactive monitoring of crowded settings to reduce the likelihood of stampede-related accidents. The suggested method functions as a DDS that supports decision-making by enabling people to categorize crowds based on risk levels and identify areas prone to stampedes rather than attempting to predict significant incidents beforehand.

2. Literature Survey

Studying crowd behavior has become more popular due to its significance for security and public safety. Helbing and Molnar's Social Force Model, which used attracting and repulsive force vectors to mathematically depict how pedestrians interact with one another, was one of the earliest models of crowd behavior. The Social Force Model is impracticable for analyzing video footage from live video surveillance systems since it necessitates the manual determination of parameter values, even though it might be useful for theoretical analysis of crowd dynamics. Later, methods based on optical flow were developed to examine crowd motion patterns. Mehran et al. showed efficacy in moderately dense scenes by using optical flow to identify abnormal crowd behaviors. Nevertheless, these techniques perform poorly in densely populated areas due to their high sensitivity to camera motion and occlusions. The drawbacks of individual pedestrian detection were addressed by the proposal of regression-based crowd density estimation techniques. In order to directly estimate crowd density from image features, Chan and Vasconcelos used regression techniques. Even though these techniques are computationally efficient, they are less accurate in situations where the crowd is not evenly distributed.

Idrees et al. presented a multi-source framework that integrated a variety of manually created features for crowd counting in order to increase estimation accuracy. This method was computationally demanding and inappropriate for real-time deployment, even though it enhanced performance in complex scenes. Convolutional neural network (CNN)-based techniques greatly enhanced crowd analysis with the development of deep learning. In order to achieve high accuracy in dense and occluded scenes, Zhang et al. proposed density map estimation using CNNs. The significance of crowd behavior analysis in relation to public safety and surveillance makes it a crucial field of study. One of the first models to use Attractive and Repulsive Forces to mathematically depict pedestrian interactions was Helbing and Molnár's Social Force Model. Regretfully, even though these techniques have advanced significantly, they have only concentrated on counting the number of pedestrians in a crowd rather than doing a thorough examination of pedestrian behavior and risk prediction.

In order to apply Deep Learning techniques for the identification of unusual behaviors in crowds, Ravanbakhsh and associates used CNN-Based Models. Although the use of Deep Learning techniques has been helpful in identifying abnormal behaviors in the presence of novel and previously unseen behaviors, these techniques are not able to produce results that are interpretable for a safety-centric application. Recent studies have looked on classifying crowd behaviors and emotions using Deep Learning algorithms. By its very nature, crowd emotion recognition is subjective and difficult to generalize. As a result, while this type of study can shed light on a group of people and their behaviors, its application to real-world safety applications is constrained.

Security film may now be utilized to successfully detect persons thanks to real-time object detection algorithms like YOLO. While YOLO-based methods work well in scenarios with low to moderately dense crowds and are appropriate for real-time monitoring, they become noticeably unsuccessful when a crowd is extremely dense and badly obscured. The bulk of recent research focuses on discrete aspects of crowd dynamics, like motion analysis, anomaly detection, and counting. Additionally, the majority of existing approaches are neither reliable in extremely dense populations, nor are they able to offer a thorough risk assessment and a real-time response. This shortcoming emphasizes the necessity of creating a thorough framework that integrates motion analysis, density estimation, and real-time individual detection for efficient crowd monitoring and early stampede danger detection.

3. Existing System

In the past, human surveillance using closed circuit television (CCTV) technologies was used to monitor crowds and prevent stampedes. CCTV operators watched video feeds in real time to spot anomalous crowd behavior and take appropriate action. This manual method is still widely used in the majority of public spaces, but its application has been hampered by problems including operator fatigue, sluggish response times, and the incapacity to continuously visually monitor large regions. As computer vision has advanced, more automated

Jack Sparrow Publishers © 2026, IJCSER , All Rights Reserved
www.jacksparrowpublishers.com

ISSN: 3107 - 8605 (Online) , http://www.ijcser.com/ , Vol. 3, Issue 1, 2026, https://doi.org/10.63328/IJCSER-V3RI1P5

methods have also been created. Early computer vision methods mostly used hand-crafted features in conjunction with statistical modeling techniques to count people in a crowd and determine how densely packed an area was. These systems could estimate the number of people at a location, but they could not be trusted in highly crowded conditions because to their acute sensitivity to variations in lighting, frequent obstacles (people), and variations in overhead camera viewpoints. Deep learning methods, particularly those based on Convolutional Neural Networks (CNN), are used in many modern computer vision applications to calculate different crowd behaviors and emotions in films. CNN-based computer vision systems have been used to identify various crowd circumstances, including crowd density estimation, crowd anomaly detection, and emotion recognition. These systems are capable of learning intricate visual patterns. The majority of CNN-based methods still require a significant amount of computer processing power and computing time to train because CNN systems typically need a very large quantity of labelled training data (with a wide range of variations) to have high-quality generalizability (i.e., perform well across multiple crowd-event situations).

Motion analysis techniques, like optical flow for irregular movements, are included into a number of current crowd surveillance technology. Motion-based techniques detect sudden changes in a crowd's behavior, but they mostly consider the scene as a whole, ignoring the specific region where congestion may arise. As a result, these devices don't show where in the crowd a stampede might occur. The present crowd surveillance systems are "siloed," concentrating on particulars like motion analysis, emotion detection, and counting. These solutions' interpretability and real-time capabilities are lacking. In order to overcome these shortcomings, a coordinated strategy that includes reliable human detection capabilities, the capacity to evaluate the temporal behavior of crowds, analyze the spatial distribution of crowds, and provide risk assessments for particular areas of an incident will need to be put into place in order to promptly issue warnings about potential stampede conditions and to encourage proactive crowd behavior management plans.

4. Proposed System

A novel automated support decision-making system that uses visual technologies to assess the likelihood of an approaching rush in a busy area is proposed. Unlike previous attempts to achieve this result using human observation or separate analytic methods, the proposed system integrates numerous components (such as individual identification, a crowd density evaluation, an analysis of individual movement, a detection of stability over time, and an assessment of the environment spatially) into a single complete process.This new method focuses on evaluating an area in real-time for predictive risk assessments rather than offering knowledge after a stampede has happened.

4.1. System Overview

In order to identify and locate people within the view frame, the new system uses video feeds from still photographs or from video surveillance, analyzing each frame using a Yolo v8 deep learning-based object detection model.

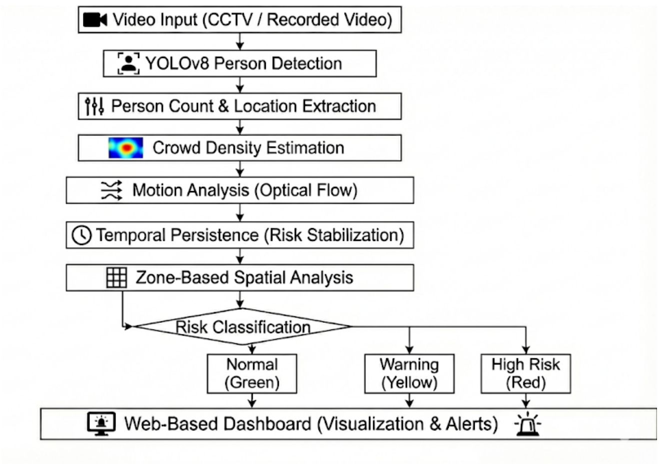

Fig. 1 Overall system architecture of the proposed YOLOv8-based crowd behavior and stampede risk monitoring system.

4.2. Person Detection Using YOLOv8

The YOLOv8 Real-time Object Detection Model, which provides effective and precise human detection, is the central component of this system. In addition to processing data quickly (single-stage), Yolo v8 can still detect objects accurately in moderately dense crowds. The center coordinates of each detected bounding box are extracted by the analyzed data from the bounding box that each detected individual formed. Bounding box detection eliminated the requirement for intricate scene mapping by enabling direct estimation of crowd density and crowd member locations.

4.3. Crowd Density Estimation

Researchers calculate the ratio of individuals spotted within a certain time period to the size of each frame examined in order to estimate crowd density. The density of the scene being captured at that precise moment is normalized by this metric. Density categories are then assigned by the estimator according to predetermined thresholds: low (less than X), medium (greater than X but below Y), or high (higher than Y). Because the likelihood of dangerous situations rises with population density, density is considered a significant predictor of risk.

Jack Sparrow Publishers © 2026, IJCSER , All Rights Reserved
www.jacksparrowpublishers.com

ISSN: 3107 - 8605 (Online) , http://www.ijcser.com/ , Vol. 3, Issue 1, 2026, https://doi.org/10.63328/IJCSER-V3RI1P5

4.4. *Motion Analysis*

In order to identify or at least approximate the amount of overall movement within that specific scene based on the magnitude of motion vectors, a method of optical flow-based motion estimation is utilized to generate subsequent video frames for the purpose of assessing collective crowd movement. Motion and crowd activity can be categorized into calm, moving, or chaotic states based on the optical flow method's subsequent comparison of video frames. Furthermore, any abrupt increase in motion intensity might be seen as a hint of potential instability or as a type of "danger," especially in an extremely crowded and urbanized setting.

4.5. *Temporal Persistence for Risk Stabilization*

Crowds' short-term volatility is utilized to gauge immediate hazards; conversely, using temporal persistence is suggested as a solution. This incorporates the latest risk assessment into a block of average recent assessments. The frequency of false alarms will be reduced by temporal persistence and the averaging of risk over time.Only when an unsafe environment persists will a risk assessment be conducted. The risk estimates are more reliable when temporal smoothing is used.

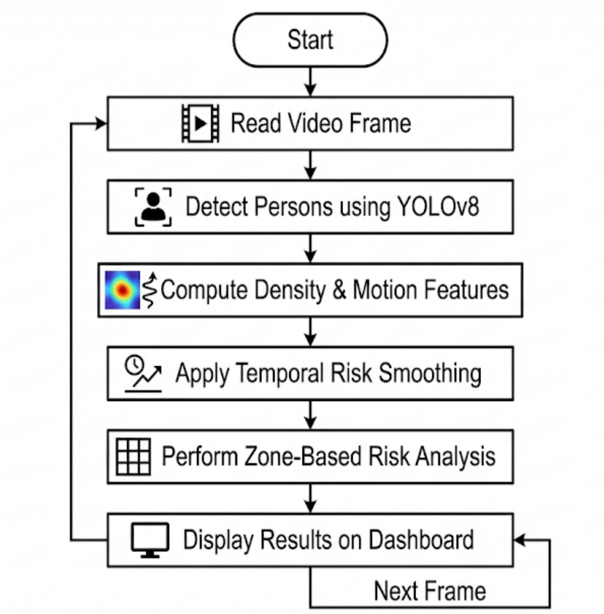

Fig. 2. Flow diagram illustrating the step-by-step processing of video frames for crowd risk estimation.

5. Methodology

The suggested methodology uses a methodical and modular approach to analyze crowd behavior using video footage in order to evaluate the risk of a stampede. Every video frame is processed by the system in order to extract significant temporal and spatial characteristics pertaining to regional congestion, motion dynamics, and crowd density. Reliable risk assessment is produced by combining these features with temporal stability and rule-based reasoning. Each methodological element is covered in detail in the ensuing subsections. The system's components are integrated to develop a dependable method of risk assessment using temporal stabilization and rule-based logic. Each element of this approach will be thoroughly explained in the next section.

5.1. *Crowd Density Estimation*

The danger of crowding rises with crowd density. As a result, one of the most crucial markers of possible danger in a location where people congregate, like concerts, is crowd density. Instead of estimating crowd density pixel-by-pixel, the suggested system estimates it person-by-person, which makes the analysis and results considerably more understandable and practical. The crowd density is calculated to be:

$$D = \frac{N}{A}$$

where N represents the number of detected individuals in a video frame and A denotes the total area of the frame. Based on predefined threshold values, the density is categorized into low, medium, or high levels to support risk estimation.

5.2. *Motion Intensity Analysis*

An examination of both density and collective motion behavior can help explain crowd stability. In extremely dense crowds (i.e., when a huge number of individuals are moving together), there is usually an increase in either panic-like or unstable-type behavior. The suggested approach is to use optical flow to assess each of these collective motion behaviors (i.e., how quickly or slow individuals are moving individually) between successive frames in order to record what is happening in both scenarios. The following formula can be used to determine the average motion's magnitude:

$$M = \frac{1}{K}\sum_{i=1}^{K}\sqrt{u_i^2 + v_i^2}$$

where u_i and v_i r indicate the optical flow's horizontal and vertical components at pixel i, and K signifies the total number of pixels. Using threshold-based categorization, the resulting motion intensity is divided into three states: calm, moving, and chaotic..

5.3. *Temporal Risk Stabilization*

Abrupt, transient disruptions in crowds might alter people's behavior at that particular moment (e.g. due to people running away from danger). Additionally, the suggested approach allows for temporal persistence of risk (or an estimate of risk at a specific

Jack Sparrow Publishers © 2026, IJCSER , All Rights Reserved
www.jacksparrowpublishers.com

ISSN: 3107 - 8605 (Online) , http://www.ijcser.com/ , Vol. 3, Issue 1, 2026, https://doi.org/10.63328/IJCSER-V3RI1P5

point in time), giving more time for changes to be detected and avoiding false alarms brought on by comparable behavioral variability. The following formula will be used to get the temporal persistence stabilization value:

$$R_{stable} = \frac{1}{W}\sum_{t=1}^{W} R_t$$

where W signifies the temporal window size and R_t represents the instantaneous risk level at time t. This method guarantees that risk alarms are only generated when dangerous conditions continue for several frames.

5.4. *Zone-Based Spatial Risk Analysis*

The majority of crowd safety problems start and develop in a specific location rather than being distributed uniformly over the entire scene. Furthermore, the suggested method will review the framing of each unique video, divide it into several distinct zone areas, and measure the movement and density within each zone to determine the overall density in each zone.

$$Z_i = \frac{N_i}{N_{total}}$$

where N_{total} is the total number of identified individuals in the frame and N_t is the number of detected individuals in zone I. The localization of potentially hazardous areas is made possible by the assignment of zone-wise risk categories based on the relative crowd concentration.

5.5. *Risk Classification Logic*

A set of guidelines that take into account the density of the crowd and the speed of movement in that particular area at any given moment establish the risk rating related to crowds. For crowds, the total risk function can be computed as follows:

$$R = f(D, M, R_{stable}, Z)$$

where f(·) is a predefined risk classification function that assigns one of three risk classifications—normal, warning, or high risk—to the outcomes of the earlier computations. The system can produce understandable and useful risk evaluations that are appropriate for real-time monitoring thanks to this standardized methodology.

6. Result and Analysis

Images and videos of distinct crowd kinds (varying density levels, occlusions, and motion dynamics) were used to assess the suggested approach for tracking crowd stampede risk. The evaluation's goal was to determine the efficacy of a method that provides an accurate risk assessment of possible stampedes by combining density-based crowd modelling with object detection-based crowd estimating. The evaluation's findings show that using only one of the two estimation techniques is insufficient to accurately determine the likelihood of a stampede in a crowd, particularly in densely populated locations.

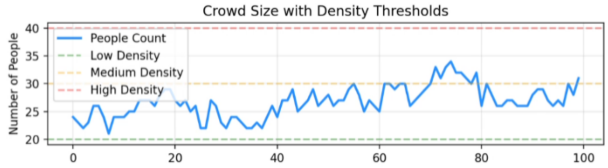

Fig. 4 Temporal Variation of detected crowd size

Figure 4 shows how the total crowd size changes over time and how it has been plotted using the predefined density thresholds. Since the time series accurately depicts the number of pedestrians entering and exiting the designated area, we would anticipate seeing this trend reflected in the number of persons tallied at a certain moment. This shows that the overall aggregate size was getting close to or beyond the medium density threshold throughout a number of these time intervals, which is a sign that localized congestion was developing in the overriding area. These variations would be crucial for real-time monitoring systems since a sharp rise in density has always suggested the potential for a potentially dangerous crowd incident to escalate quickly. Beyond this, our analyses unequivocally show that a simple count of a crowd does not give enough information about the possible risks associated with the crowd, especially if there are any occlusions taking place nearby.

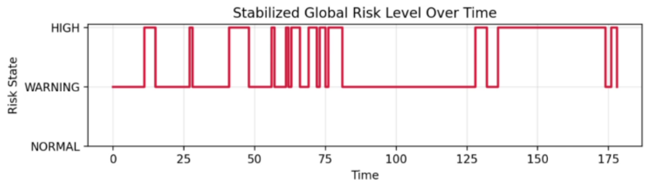

Fig. 5 Stabilized global risk level over time

The stable global risk level's evolution throughout time is seen in Figure 5. The suggested approach uses a temporal smoothing function that averages short-term observations across time to reduce immediate detection mistakes, in contrast to the instantaneous risk classification, which is highly vulnerable to such errors. In addition to putting this strategy into practice, the suggested method lessens the frequency of oscillations between risk levels brought on by temporary changes or partial occlusions, improving clarity and avoiding false alarms.

The system will show a WARNING State when there is moderate congestion for a long time, but a HIGH RISK State will only be created when the dangerous circumstances continue for a long time. One important prerequisite for the system's successful deployment for use in public safety is that it must stay in the WARNING state and only produce a HIGH RISK state when hazardous conditions persist.

Jack Sparrow Publishers © 2026, IJCSER , All Rights Reserved
www.jacksparrowpublishers.com

ISSN: 3107 - 8605 (Online) , http://www.ijcser.com/ , Vol. 3, Issue 1, 2026, https://doi.org/10.63328/IJCSER-V3RI1P5

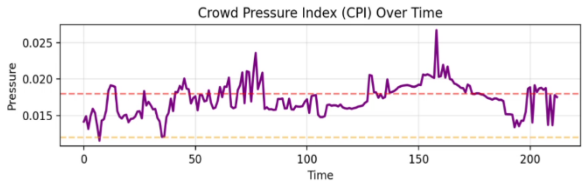

Fig. 6 Crowd Pressure Index (CPI)

In Figure 6, the Crowd Pressure Index (CPI) is calculated using estimations of the local population density as well as the separation between detected individuals. The graph shows that a few of the pressure peaks are significantly higher than the designated WARNING and DANGER thresholds. Crucially, the pressure surges may be seen even when the overall number of people in the crowd is comparatively constant, demonstrating that the crowd risk is not only influenced by the total number of people in the crowd. Since it is obvious that both spatial compression and mobility limitation are essential to the creation of a stampede, the pressure metrics must be established as essential elements of the suggested framework.

Fig. 7 Zone Wise Person Detection (HeatMap Based)

Fig. 7 shows the results of YOLOv8 for identifying individuals in the various zones. Because it properly recognizes people, YOLOv8 functions effectively in both moderately and sparsely populated areas, enabling localized risk assessment..

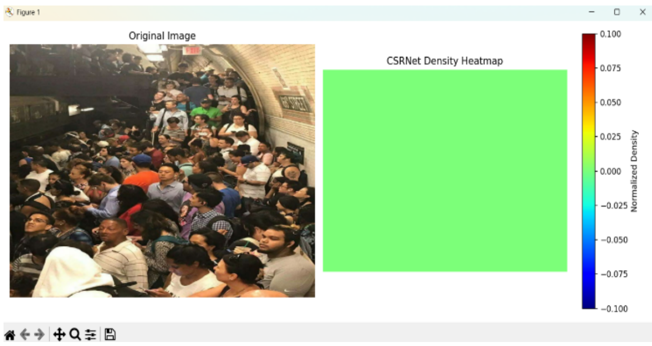

Fig. 8 CSRNet-based density estimation

However, YOLOv8's capacity to identify persons is greatly reduced in densely populated places where a large number of people are present due to heavy occlusion from overlapping individuals. Because YOLOv8 only detects a small number of persons who are visible, it underestimates the actual number of people in a given region. Bounding-box-based object detection techniques and their limitations in high-density crowd environments exacerbate the issue. Density estimate using CSRNet was created to mitigate YOLOv8's drawbacks in certain situations. As seen in Fig. 8, CSRNet does not need to detect every person in the crowd because it creates a continuous density map that depicts the spatial distribution of a crowd. No matter how many people are present and visible, the integrated density values from CSRNet produce accurate crowd estimations. According to experimental findings, in congested situations when YOLOv8 detects fewer than 10 people, CSRNet forecasts a crowd size of between 150 and 300. The aforementioned findings demonstrate the notable distinction between YOLOv8 modeling and density-based modeling for crowd detection in extremely dense environments.

Table. 1 Density estimate using CSRNet versus YOLO-based detection

Scenario Type	YOLO Count	CSRNet Estimated Count	Risk Level
Sparse Crowd	6–10	8–15	LOW
Medium Crowd	10–25	40–80	WARNING
Dense Crowd	1–5	300+	HIGH

The performance of CSRNet Density Estimation and YOLO Object Detection in different crowds is contrasted in Table 1. In high-density crowd scenarios, YOLO Object Detection can yield significantly lower overall density estimations than CSRNet. This system can operate across many density ranges by combining CSRNet Density Estimation with YOLO Object Detection. While CSRNet offers superior estimate capabilities in high-density crowd regions, YOLO offers good localization and tracking capabilities in nearly empty areas. Our method will produce a comprehensive, scalable solution for early identification of stampede-related danger by stabilizing temporal information through the use of CPI for pressure analysis and time intervals.

7. Discussion

The results of the studies show that in dense, highly obstructed situations, crowd risk assessments based solely on object-level detections will perform poorly. Crowding detection will no longer be a dependable source when working with these kinds of scenarios because crowded scenes contain a large number of people. For instance, while YOLO-based cameras can identify individual people in less crowded settings, their accuracy deteriorates significantly as crowd densities rise. Because YOLO cameras only detect a small percentage of people who are visible to the camera, they greatly underestimate the number of people present in extremely congested areas.Extreme inter-person occlusion, abrupt size shifts, and perspective distortion—all of which are common in

Jack Sparrow Publishers © 2026, IJCSER , All Rights Reserved
www.jacksparrowpublishers.com

ISSN: 3107 - 8605 (Online) , http://www.ijcser.com/ , Vol. 3, Issue 1, 2026, https://doi.org/10.63328/IJCSER-V3RI1P5

such big crowds—can be blamed for this failure. All of this happens when the camera operates outside the bounding-box-based detectors' ideal working parameters.

However, CSRNet density estimation is very resistant to these kinds of environmental difficulties. CSRNet uses continuous spatial density functions to depict crowd densities instead than relying on accurate individual person detections. Therefore, even when the individual members of the crowd are significantly concealed by occluding objects, CSRNet produces a very comprehensive overview of crowd distributions by using a spatial density function to model crowds. Crowds can be replicated as realistic crowd counts in reasonably high density (e.g., between 150 and 300 individuals) for environmental samples where YOLO could only detect less than ten individuals, as demonstrated by experimental tests using CSRNet.The differences in crowd density recognition between YOLO and CSRNet highlight the advantages of density estimation as a precise method of simulating crowds in settings (like large gatherings) when visual cues restrict the capacity to spatially distinguish individual objects.

Stampede risk cannot be determined solely by density evaluation. If there are no additional contributing variables in motion or pressure within the crowd, a large crowd by itself isn't always a dangerous condition. By measuring the Crowd Pressure Index (CPI) by looking at crowd spatial proximity patterns and how people are moving in any particular region, our method accounts for both of those factors. Figure 2 demonstrates that atypical increases in local crowd pressure, rather than just greater crowds, are directly linked to local increases in danger. In several cases, base crowd density levels were maintained while the CPI exceeded a predefined threshold, indicating "High Risk."These studies lead us to the conclusion that pressure-based metrics, which give users a real-time threshold of possible pressure-induced stampede risk in relation to the level of the static density of the crowd, can be the most dependable method of forecasting possible stampedes.

Our method incorporates temporal stability in addition to motion, proximity, and density measures. The system is built to guard against false alarms brought on by brief increases in detection output. To generate the final conclusion, the risk judgment module uses a moving average to aggregate the people's density, mobility, and pressure signals. Fewer false alarms from abnormally high density peaks or other transient circumstances are anticipated when the inputs of several detections are combined over time. A signal is only generated when there is a trend that suggests a persistent possibility of a stampede.

YOLO (You Only Look Once) and CSRNet (Crowd-Segmentation-Reconstruction Network), two complimentary technologies that serve as the basis for effective crowd monitoring, are integrated into the suggested hybrid architecture. YOLO can precisely identify an object in a picture and determine how many people are in each section of the picture. In order to give spatial risk advice, YOLO generates a fine-level of detail when segmenting out the at-risk spatial locations. Conversely, CSRNet concentrates on estimating the population in extremely crowded places, where YOLO may not be able to recognize individual items.

Time-smoothing technology combined with two hybrid models (crowd pressure index and time stamping) provides a consistent and uniform method of evaluating the accuracy of a particular hybrid model across time. The two hybrid models covered in this work will yield the best results when the hybrids' time-related variables or geographical locations (e.g. i. Over a predetermined amount of time, stationary or moving targets and static or dynamic environments can fluctuate and alter. The results demonstrate that the risk of a stampede cannot be reliably monitored by a stand-alone approach. Cue data from detection-based, density-based, and pressure-based analysis should be integrated into a single decision-making process throughout time in order to efficiently monitor the risk of stampedes. The system presented in this paper shows that such integration can take place in real-time and significantly improve the capacity to identify early risk potential, making it applicable to large public events, particularly in locations like transportation hubs, places of worship, and sizable gatherings.

8. Conclusion and Future Scope

This study describes a hybrid crowding behavior tracking system for identifying potential threats of stampedes at huge crowds in public spaces. The tracking system was developed by integrating motion analysis, tracking zones in space, monitoring stability risk over time, CSRNet for dense crowd size estimations, and YOLO for object-level identification. This hybrid approach uses both detection and density estimation to achieve high-performance tracking of crowds in three different types of crowd conditions, namely sparse, moderate, and dense with high density, whereas existing systems typically use either detection of the individuals or estimation of their density. The results of the experimental tests show that depending only on object detection has several drawbacks, including underestimating the number of people present. As a result, only a small portion of the total number of people—typically one person in a densely occluded environment—will be detected. On the other hand, CSRNet can replicate density distributions of a crowd using crowdsourced density estimation, which enables it to achieve precise size estimations even when a substantial portion of the crowd is

ISSN: 3107 - 8605 (Online) , http://www.ijcser.com/ , Vol. 3, Issue 1, 2026, https://doi.org/10.63328/IJCSER-V3RI1P5

suppressed or obscured by other objects. Accurate risk assessment under various crowding conditions results from the merging of these two sets of outcomes. In order to offer an early indication for the detection of localized instability prior to a catastrophic stampede, the Crowdsurge Pressure Index (CPI) measures the proximity of individuals to one another and the speed at which the crowd is moving towards the exit. The system relies on a temporal persistence and risk stabilization mechanism to improve system reliability by automating the detection of sustained elevated-risk situations and preventing false alarms caused by brief environmental disturbances or transient fluctuations in a sensor's signal output (camera noise).

The temporal risk timeline shows that, while reducing the amount of false alarms, high-risk scenarios that continue to be elevated were regularly and accurately identified. To sum up, our hybrid is a real-time risk-based crowd monitoring system that is scalable and comprehensible. As a result, the suggested framework has a wide range of potential uses at places that present high safety risks (i.e., high-impact locations), such as public transportation hubs, significant religious events, concerts, and major public events (gatherings, festivals, and gatherings with large numbers of people).

When it comes to using real-time crowd scanning to identify potential threats, the proposed method works very well.Nevertheless, there are other ways to expand the framework to improve its dependability and functionality.In order to increase performance across various lighting situations, camera angles, and crowd compositions, future research will involve training and optimizing density estimation models utilizing scene-specific data. Additionally, by using adaptive calibration techniques, the devices will be able to modify the pressure and density thresholds based on environmental conditions.

An intriguing field of study for large-scale venue monitoring is multi-camera and cross-camera crowd flows. In addition to decreasing the amount of blind spots associated with single-camera monitoring, the inclusion of multi-view spatial information enhances an overall comprehension of crowds. Predicting crowd trajectories will also make it possible to anticipate when and where crowds would congregate and create bottlenecks. In addition to a vision-based crowd analysis, future study will build on the utilization of additional data sources, including Bluetooth signals, Wi-Fi probe data, and other environmental sensors. Furthermore, the optimization of the decision-making process in a dynamic crowd environment will be facilitated by the application of machine learning risk classifiers and Reinforcement Learning (RL) based alert policies. It can develop into a comprehensive intelligent crowd safety platform with these other components.

References

[1]. Y. Li, X. Zhang, and D. Chen, “CSRNet: Dilated convolutional neural networks for understanding the highly congested scenes,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, UT, USA, Jun. 2018, pp. 1091–1100.

[2]. J. Redmon and A. Farhadi, “YOLOv3: An incremental improvement,” *arXiv* preprint arXiv:1804.02767, 2018.

[3]. A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, “YOLOv4: Optimal speed and accuracy of object detection,” *arXiv* preprint arXiv:2004.10934, 2020.

[4]. G. Jocher, A. Chaurasia, and J. Qiu, “YOLOv8,” Ultralytics, 2023. [Online]. https://github.com/ultralytics/ultralytics

[5]. G. Farnebäck, “Two-frame motion estimation based on polynomial expansion,” in *Proc. Scandinavian Conf. Image Analysis*, Halmstad, Sweden, 2003, pp. 363–370.

[6]. V. Lempitsky and A. Zisserman, “Learning to count objects in images,” in *Advances in Neural Information Processing Systems (NeurIPS)*, 2010, pp. 1324–1332.

[7]. Z. Zhang, M. Wang, and X. Geng, “Crowd counting in public scenes using density map estimation,” *IEEE Transactions on Image Processing*, vol. 28, no. 8, pp. 4124–4137, Aug. 2019.

[8]. C. C. Loy, K. Chen, S. Gong, and X. Xiang, “Crowd counting and profiling: Methodology and evaluation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 4, pp. 820–833, Apr. 2016.

[9]. D. Helbing and P. Molnár, “Social force model for pedestrian dynamics,” *Physical Review E*, vol. 51, no. 5, pp. 4282–4286, May 1995.

[10]. D. Helbing and A. Johansson, “Pedestrian, crowd and evacuation dynamics,” in *Encyclopedia of Complexity and Systems Science*. New York, NY, USA: Springer, 2009, pp. 6476–6495.

Declaration

Conflicts of Interest: The authors declare no conflict of interest.

Author Contribution: All authors wrote the main manuscript text and also consent to the submission.

Ethical approval: Not applicable.

Consent to Participate: All authors consent to participate.

Funding: Not applicable, and No funding was received

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Personal Statement: We declare with our best of knowledge that this research work is purely Original Work and No third party material used in this article drafting. If any such kind material found in further online publication, we are responsible only for any judicial and copyright issues.

Acknowledgements

We thank everyone who inspired our work.

Jack Sparrow Publishers © 2026, IJCSER , All Rights Reserved
www.jacksparrowpublishers.com