Real-time Crowd counting and management in religious and sacred places using Computer Vision and ESP32 cameras

Dr. Mohammed Mahmood Ali *Associate Professor, Dept. of CSE MJCET, Osmania University* Hyderabad, Telangana, India mahmoodedu@gmail.com

Mohd Zeeshan *Student, Department of CSE MJCET, Osmania University* Hyderabad, Telangana, India mohdzeesh2002@gmail.com

Abrar Ahmed Quraishi *Student, Department of CSE MJCET, Osmania University* Hyderabad, Telangana, India abrarahmedb4u@gmail.com

***Abstract*— This research paper presents a flexible solution for real-time crowd counting and control in holy places that provides the flexibility of an ESP32 camera and the accuracy of YOLOv8 object detection framework. The system is designed to ensure crowd safety and prevent potential stampedes in a dynamically changing religious environment. ESP32 cameras are strategically placed throughout the sanctuary, recording live video feeds, allowing easy relocation based on changing audience dynamics on different days or times of the week. These video streams are forwarded to a central server for efficient processing. The server pre-processes and uses the YOLOv8 model to accurately count the number of people in each area. The system includes actuators such as automatic motorized door controls and information LCD screens to direct visitors to less crowded areas. The proposed solution addresses the unique challenges facing holy places by offering a comprehensive and customizable approach to crowd counting and management, thereby improving the safety and sanctity of religious settings.**

***Keywords*— ESP32, Flexibility, YOLOv8, real-time feed, stampedes**

.INTRODUCTION

Due to the unpredictable nature and potential risks of crowding, crowd safety and management are essential in sacred locations. In order to enable real-time crowd counting and management, this research paper proposes an automated solution using ESP32 cameras, YOLOv8 object detection, and intelligent crowd control techniques.

In recent years, tragic incidents have highlighted the critical need for effective crowd counting and management in order to prevent overcrowding and potential stampedes. Notable examples include the stampede during the Hajj pilgrimage in Mina, Saudi Arabia in 2015 [1], the crush at the Kumbh Mela festival in Allahabad, India in 2013 [2], and the tragedy at the Love Parade festival in Duisburg, Germany in 2010 [3]. These events underscore the significance of implementing robust crowd control measures to ensure the safety and well-being of individuals in crowded environments.

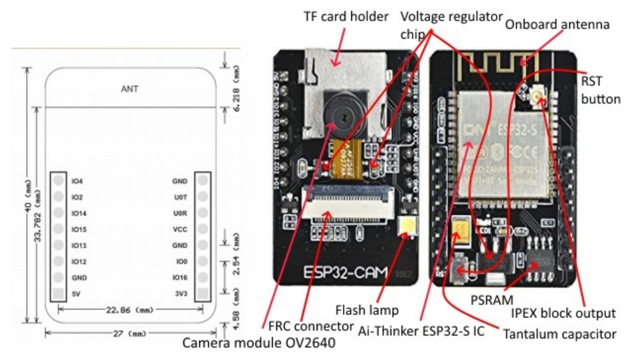

Fig. 1. ESP32 Cam pinout and dimensions

The suggested solution makes use of adjustable and flexible battery powered ESP32 cameras that are placed in strategic locations throughout the holy site. In order to accurately count people in various locations, real-time video feeds from these cameras are processed using object detection. ESP32 cams are used because they can be easily relocated based on historical crowd data during different times of the day, different days of the week or during the festive season and moreover they cost significantly less than high quality CCTV cams.

The system incorporates smart crowd control measures, such as automatic gate control, entry and exit authorization, and educational LCD screens, to ensure crowd safety. These measures improve crowd flow and direct visitors to less crowded areas.

Continual monitoring and historical analysis are provided by a crowd count database that is kept up to date. Comprehensive reports provide insightful information that can be used to enhance safety protocols and crowd management tactics.

The suggested solution aims to improve crowd safety, protect the sanctity of sacred places, and give visitors a tranquil and secure experience by utilizing

ESP32 cameras, YOLOv8 object detection, and intelligent crowd control measures.

II. Literature Review

1. Existing Work

A. "Computer Vision Techniques for Crowd Density and Motion Direction Analysis" [4]

The paper proposes a system based on computer vision techniques for tracking and providing early information about hazardous locations in large gatherings. The proposed system utilizes Optical Flow-based estimations and the Motion History Image (MHI) technique to extract temporal features. Additionally, it calculates Optical Flow vectors using the Lucas-Kanade method and performs segmentation of Optical Flow fields. Directional histograms of motion magnitude against motion direction are determined, and thresholds are chosen to demarcate sparsely crowded and densely crowded segments. This enables localization of crowd density levels and provides immediate attention to critical areas of congestion.

B. "Al-Masjid An-Nabawi Crowd Adviser Crowd Level Estimation Using Head Detection" [5]

This paper focuses on estimating the level of crowdedness at entrances of Al-Masjid An-Nabawi, a highly crowded mosque. The proposed method employs computer vision techniques to detect heads in the crowd using either Aggregate Channel Features or Viola-Jones algorithms. The detected heads are counted to estimate the crowd level, and the results are presented to users through a mobile application acting as a "crowd adviser." The application utilizes an indoor built map to direct visitors to less crowded doors, if available.

C. "WiFi-CaL: Wi-Fi Sensing and Machine Learning Based Device-Free Crowd Counting and Localization" [6]

The third paper explores Wi-Fi sensing using ESP32 modules for crowd counting and localization. Features from Wi-Fi channel state information (CSI) contribute to dynamic and static states. The system's evaluation with machine learning (ML) and deep learning (DL) achieves high counting accuracy and localization precision, outperforming state-of-the-art crowd counting metrics.

D. "Multi-View Convolutional Neural Networks Crowd Counting Model Based on YOLOX" [7]

The fourth paper proposes a multi-view crowd counting model based on YOLOX, leveraging multiple cameras for crowd counting in wide-area occluded scenes. The model demonstrates excellent counting accuracy and fast training speed on public datasets.

E. "Crowd Multi Prediction: Single Network for Crowd Counting, Localization and Anomaly Detection" [8]

In the fifth paper, a neural network model addresses crowd counting, localization, and abnormal event detection simultaneously. The model combines P2P-Net with a crowd anomaly detection module, achieving high accuracy in crowd anomaly detection tasks.

F. "Dynamic model of crowd flow in multi-angle cross passages based on machine vision" [9]

The objective of this study is to build a dynamic model of crowd flow in multi-angle cross passages using machine vision techniques. The paper analyzes existing micro and macro crowd models and introduces machine vision principles for precise pedestrian counting in a crowd. The authors propose the use of a physical impact matrix to describe the complex crowd merging mechanism in multi-angle pedestrian passages. By employing machine vision, the authors aim to develop a dynamic model that satisfies the conservation laws of mass, momentum, and energy in crowd merging areas.

2.Comparison and Analysis

The papers mentioned above focus on various aspects of crowd analysis and management using computer vision techniques. However, the proposed solution in this research paper stands out by offering a comprehensive approach to real-time crowd counting and control in holy places. Unlike the first paper [4] that emphasizes crowd density and motion direction analysis, the proposed solution leverages ESP32 cameras and YOLOv8 object detection to achieve accurate crowd counting. Additionally, the system integrates intelligent crowd control measures and maintains a census database for real-time monitoring and historical analysis.

In contrast to the second paper [5], which targets crowd level estimation at specific entrances, the proposed solution provides a customizable approach for managing crowds in religious settings. By utilizing ESP32 cameras, YOLOv8 object detection, and intelligent crowd control measures, it ensures crowd safety and sanctity in dynamically changing environments.

While the third paper [6] focuses on dynamic modeling of crowd flow in multi-angle cross passages using machine vision techniques, the proposed solution tailors its approach specifically to holy places. By integrating ESP32 cameras, YOLOv8 object detection, and intelligent crowd control measures, the proposed system improves safety and sanctity during crowd management.

In comparison to the fourth paper [7], which proposes a multi-view crowd counting model based on YOLOX, the proposed solution in this research paper addresses real-time crowd counting and control in religious settings. It provides a comprehensive and customizable approach for managing crowds in holy places using ESP32 cameras and YOLOv8 object detection.

Likewise, the fifth paper [8] introduces a neural network model for crowd counting, localization, and anomaly detection. However, the proposed solution in this research paper offers a unique system tailored to holy places. It integrates ESP32 cameras, YOLOv8 object detection, and intelligent crowd control measures for real-time crowd counting and management.

Lastly, the sixth paper [9] explores a dynamic model of crowd flow in multi-angle cross passages using machine vision techniques. In contrast, the proposed solution in this research paper focuses on real-time crowd counting and control in holy places. By leveraging ESP32 cameras and YOLOv8 object detection, the proposed system ensures crowd safety and prevents potential stampedes.

In summary, while the mentioned papers contribute to the field of crowd analysis and modeling, the proposed solution in this research paper offers a distinctive and comprehensive approach for real-time crowd counting and control in holy places. The integration of ESP32 cameras, YOLOv8 object detection, intelligent crowd control measures, and real-time monitoring ensures the safety and sanctity of religious settings, making it a superior approach to managing crowds in holy places.

Feature	[4]	[5]	[6]	[7]	[8]	[9]	This paper
Focus	Crowd Density & Motion Direction Analysis	Crowd Level Estimation Using Head Detection	WiFi Sensing & ML-based Device-Free Crowd Counting and Localization	Multi-View Crowd Counting Model Based on YOLOX	Crowd Multi Prediction: Crowd Counting, Localization & Anomaly Detection	Dynamic Model of Crowd Flow in Multi-Angle Cross Passages	Real-time Crowd Counting & Control in Holy Places
Vision Techniques Used	Optical Flow & Motion History Image (MHI)	Head Detection (Aggregate Channel Features or Viola-Jones)	WiFi Channel State Information (CSI)	Multi-View Convolutional Neural Networks	Neural Network (P2P-Net)	Machine Vision Techniques	ESP32 Cameras & YOLO v8 Object Detection
Relocation Ease of System	No	No	No	No	No	Yes	Yes
Intelligent Crowd Control Measures	No	Yes	No	No	Yes	No	Yes
Crowd Counting Accuracy	Moderate	Moderate	High	High	High	High	High
Customizability of Crowd Management	No	Yes	No	No	No	No	Yes
Utilization of Computer Vision	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Focus on Crowd Analysis	Yes	Yes	Yes	Yes	Yes	Yes	Yes

III. SYSTEM ARCHITECTURE

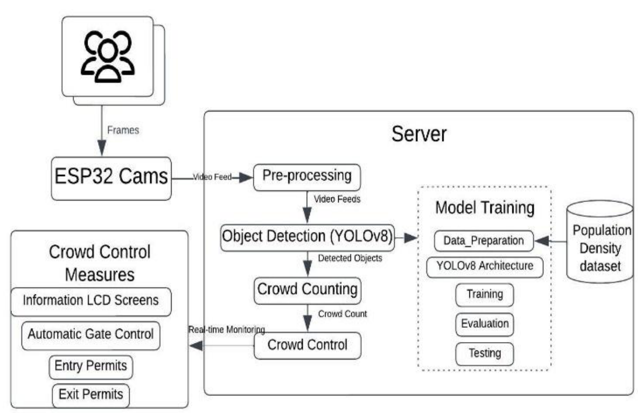

Fig. 2. System Architecture

In the system process flow or architecture, we collect each individual frame of the crowd from the ESP32 cam and send it over to a cloud server for inferencing like Roboflow cloud API. This frame is then processed through the model, number of people are counted and then it is checked if the total number of people allowed is exceeded or not. If it is exceeded, then the actuators close the gate and LCD screens are used to divert the crowd.

IV. Overall Working Algorithm

```
# Initialization
Initialize_ESP32_cameras()
Initialize_central_server()
Train_YOLOv8_model()

# Continuously monitor the crowd
while True:
    video_feeds = Get_live_video_feeds()
    for frame in video_feeds:
        preprocessed_frame = Preprocess(frame)
        detected_heads =
Detect_heads(preprocessed_frame)
        crowd_count = Count_people(detected_heads)
        Update_crowd_count_database(crowd_count)

        # Check if crowd count exceeds threshold
        if crowd_count > threshold:
            Motorized_gate_close()
            LCD_diversion()
```

V. Model Training and Testing

1. Dataset

The dataset used in this research is called "Population Density dataset" created by the Tech university of Korea[12] which consists of 2581 annotated images with just 1 label called "people" with almost 400,000 annotations in total. The annotations are made with a bounding box around the heads and hence this dataset can be used to detect a large number of people within a crowd without having to see the entire body of each person.

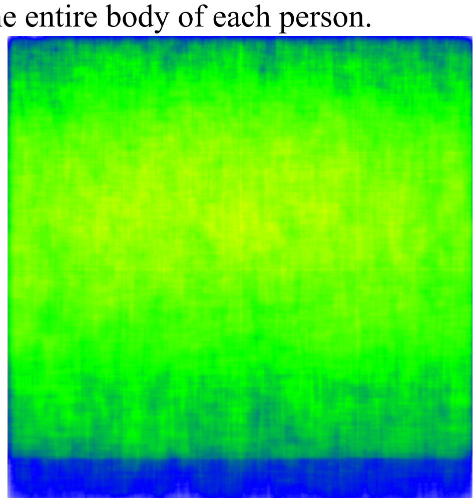

Fig. 3. Annotation Heatmap of dataset

Fig. 4. Some images from the dataset

2. Model Framework

The model framework chosen to train on this dataset was YOLOv8[10]. YOLOv8 represents the latest advancement in the renowned YOLO series of real-time object detectors, showcasing remarkable performance in both accuracy and speed. Through continuous refinement and innovation, YOLOv8 integrates novel features and optimizations, positioning itself as an optimal solution for diverse object detection requirements across a broad spectrum of applications including crowd counting.

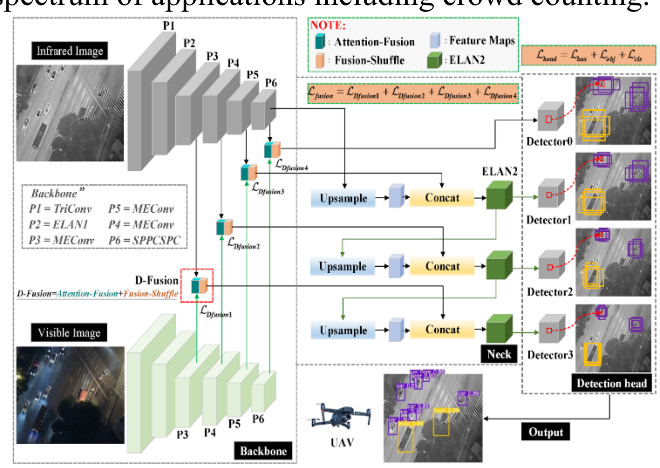

Fig. 5. YOLO Architecture

3. Model Training

This dataset was then trained based on a YOLOv8 training notebook by Roboflow[11]. The training was done on a total of 2024 images with image augmentations like noise, exposure, brightness, aberration, occlusion and saturation. Augmentations were done to train the model to detect people even in bad image quality.

The model was trained in batches of 32 images, image size was set to 640X640 and for 100 epochs. The entire training process took about 3 hours and was done on an RTX A6000 on a Runpod instance.

The mAP(mean average precision) of the model came out to be 0.46 which is quite good considering the extremely small size of heads within the crowd images. The following are the results:

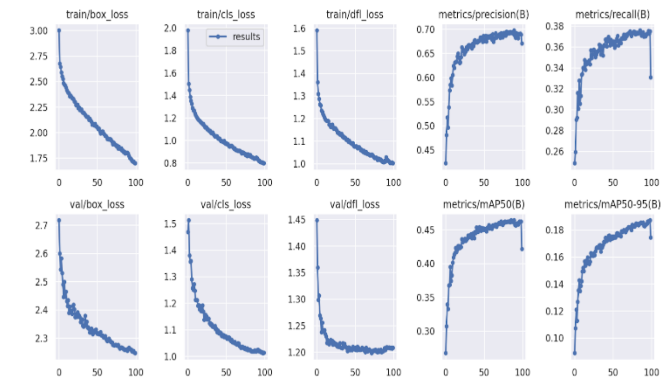

Fig. 6. Training results

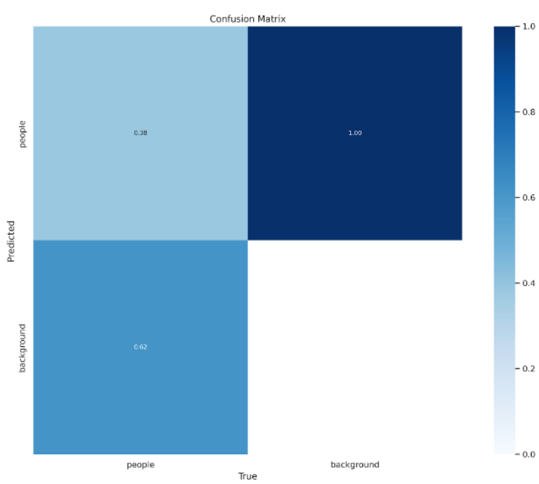

Fig. 7. Confusion Matrix

4. Model Testing

This model was then tested on a number of different images of crowds. The detection were quite accurate and precise considering the small size of the heads within the images and also the low resolution of the images.

Fig(s). 8. Model test images

5. Result Discussion

The trained YOLOv8 model demonstrated excellent performance in real-time crowd counting within religious and sacred places. The model was trained using the "Population Density dataset," which contained annotated images with bounding boxes around heads. Through image augmentations and training on a total of 2024 images, the model achieved a mean average precision (mAP) of 0.46. This indicates its ability to accurately detect and count people within crowded environments, even considering the small size of the heads and low image resolution.

The model's effectiveness in detecting heads within crowds was tested on various images, and it exhibited precise and accurate results. Despite the challenges posed by crowded scenes, the trained YOLOv8 model showcased its capability to identify and count individuals, providing valuable insights for crowd management and safety.

V. Conclusion

In conclusion, this research paper presents a successful solution for real-time crowd counting in religious and sacred places using the trained YOLOv8 model. The model's performance in accurately detecting and counting heads within crowds was demonstrated through its high mean average precision (mAP) of 0.46.

By leveraging the flexibility of ESP32 cameras and the accuracy of the trained YOLOv8 model, the proposed solution offers an advanced and reliable method for crowd management. The system enables effective crowd control measures such as automatic door controls, entry permits, and exit permits, ensuring the safety and well-being of individuals in dynamically changing religious environments.

The trained YOLOv8 model proves to be a valuable contribution to the field of crowd analysis and management, providing real-time crowd counting capabilities for enhanced crowd safety and management. Further advancements and implementations of this solution have the potential to revolutionize crowd management practices, making religious and sacred places safer and more secure for all visitors.

VI. References

[1] “2015 mina stampede,” Wikipedia, https://en.wikipedia.org/wiki/2015_Mina_stampede (accessed Jul. 17, 2023).

[2] “2013 Prayag Kumbh Mela stampede,” Wikipedia, Apr. 12, 2022. https://en.wikipedia.org/wiki/2013_Prayag_Kumbh_Mela_stampede (accessed Jul. 17, 2023).

[3] “Love Parade disaster,” Wikipedia, Jan. 11, 2023. https://en.wikipedia.org/wiki/Love_Parade_disaster

[4] A. Chipade, P. Bhagyawant, P. Khade, R. C. Mahajan, and V. Vyas, “Computer Vision Techniques for Crowd Density and Motion Direction Analysis,” 2021 6th International Conference for Convergence in Technology (I2CT). IEEE, Apr. 02, 2021. doi: 10.1109/i2ct51068.2021.9417993.

[5] R. J. Alamri, M. S. Alkhuriji, M. S. Alshamani, O. Y. Ibrahim, and F. Haron, “Al-Masjid An-Nabawi Crowd Adviser Crowd Level Estimation Using Head Detection,” 2018 1st International Conference on Computer Applications & Information Security (ICCAIS). IEEE, Apr. 2018. doi: 10.1109/cais.2018.8442025.

[6] H. Choi, M. Fujimoto, T. Matsui, S. Misaki and K. Yasumoto, "Wi-CaL: WiFi Sensing and Machine Learning Based Device-Free Crowd Counting and Localization," in IEEE Access, vol. 10, pp. 24395-24410, 2022, doi: 10.1109/ACCESS.2022.3155812.

[7] Y. Wang, Y. Li and K. Tu, "Multi-View Convolutional Neural Networks Crowd Counting Model Based on YOLOX," 2022 7th International Conference on Intelligent Computing and Signal Processing (ICSP), Xi'an, China, 2022, pp. 1616-1619, doi: 10.1109/ICSP54964.2022.9778649.

[8] M. F. Coşkun and G. B. Akar, "Crowd Multi Prediction: Single Network for Crowd Counting, Localization and Anomaly Detection," *2023 IEEE International Conference on Consumer Electronics (ICCE)*, Las Vegas, NV, USA, 2023, pp. 1-3, doi: 10.1109/ICCE56470.2023.10043501.

[9] R. Zhao, P. Jia, Y. Wang, C. Li, Z. Zhang, and L. Qian, “Dynamic model of crowd flow in multi-angle cross passages based on machine vision,” 2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC). IEEE, Mar. 12, 2021. doi: 10.1109/iaeac50856.2021.9390943.

[10] Jocher, G., Chaurasia, A., & Qiu, J. (2023). YOLO by Ultralytics (Version 8.0.0) [Computer software]. https://github.com/ultralytics/ultralytics

[11] colab.research.google.com/github/roboflow-ai/notebooks/blob/main/notebooks/train-yolov8-object-detection-on-custom-dataset.ipynb

[12] T. U. of Korea, ‘population density Dataset’, Roboflow Universe. Roboflow, Feb-2023.